

確率論基礎

重川 一郎

平成 24 年 8 月 24 日

目次

第 1 章	確率空間と確率変数	5
1	確率空間	5
	可測空間	5
	確率空間	6
2	確率変数	8
	確率変数	8
	分布	8
	期待値	9
	モーメント, 分散, 標準偏差	10
3	独立性と条件付確率	12
	独立性	12
	条件付確率	15
	Bayes の公式	15
	Markov 連鎖	17
第 2 章	確率分布	21
1	離散分布	21
	2 項分布	21
	幾何分布	22
	ポアソン分布	23
2	連続分布	26
	一様分布	27
	指数分布	27
	ガンマ分布	27
	ベータ分布	28
	正規分布	28
3	多次元分布	30
	2 次元分布	30
	多次元確率分布	32
第 3 章	極限定理	35
1	大数の法則	35
	確率変数の収束	35

	大数の弱法則	35
	大数の強法則	36
2	特性関数	37
	特性関数	37
	テント関数	38
3	中心極限定理	44
	中心極限定理	44
	Notes	46
第 4 章	ランダム・ウォーク	47
1	単純ランダム・ウォーク	47
	単純ランダム・ウォーク	47
	再帰性, 非再帰性	47
	ウォリス (Wallis) の公式	49
	再帰確率	50
	2次元ランダム・ウォーク	51
	3次元ランダム・ウォーク	52
	Notes	54

第1章 確率空間と確率変数

余談から．確率概念は直感が働くと同時に，直感に騙されるということもある．慎重に考えないと間違った結論を出してしまうことも多いのである．「豪華乗用車とヤギ」(Car and Goat) という話がある．クイズに勝ち抜いた後で，挑戦者は賞品として車がもらえらるしよ．ただし，3つの扉があって，挑戦者はそのうちのひとつを選ぶのだが，車があるのは1つだけで，残りの扉の後ろにはヤギがいるだけ．挑戦者が選んだ後，司会者が残りの扉からひとつを選んで開ける．司会者はどこに車があり，ヤギがいるか知っているわけで，必ずヤギの居る扉を開ける．挑戦者はヤギの居る扉を一つ知らされた後，選んだ扉を変更するチャンスを与えられる．さて，この挑戦者は自分の選択を変更すべきだろうか．最善の戦略は？そしてそのときの車を獲得できる確率は？

残った二つのうちのひとつを選ぶのだから $1/2$ というのが一つの答え．だが正解は $2/3$ なのだ．挑戦者は最初でたらめに選んだ扉を放棄し，必ず残った扉を選ぶべきなのだ．そうすれば，最初にヤギを選んだときには，変更すれば必ず車が当たる．そして最初にヤギを選ぶ確率は $2/3$ なのだから．

1. 確率空間

確率論を数学的に述べるための，基本的な枠組みである確率空間について述べる． Ω を一般的な集合とする．

可測空間

定義 1.1. Ω の部分集合を要素とする集合族 \mathcal{F} が次の性質をみたすとき σ -集合体 (σ -field) という：

$$(1) \emptyset, \Omega \in \mathcal{F}.$$

$$(2) A \in \mathcal{F} \implies A^c \in \mathcal{F}$$

$$(3) A_n \in \mathcal{F}, n = 1, 2, \dots \implies \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$$

集合 Ω に σ -集合体 \mathcal{F} を付加した空間 (Ω, \mathcal{F}) を可測空間 という．一般に位相空間 S に対して開集合をすべて含む最小の σ -集合体が一意に定まる．これを Borel σ -集合体 (位相的 σ -集合体と呼ばれることも多い) とよび，以下 $\mathcal{B}(S)$ と記す． $(S, \mathcal{B}(S))$ は可測空間となる． S が位相空間の場合は特に断らなければ， σ -集合体として $\mathcal{B}(S)$ をとる． $S = \mathbb{R}, \mathbb{C}, \mathbb{R}^d$ などが典型的なものである．

命題 1.2. \mathcal{F} を σ -集合体とするとき，次のことが成り立つ：

- (1) $A, B \in \mathcal{F} \implies A \setminus B \in \mathcal{F}$.
- (2) $A_n \in \mathcal{F}, n = 1, 2, \dots \implies \bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$

証明 (1) : $A \setminus B = A \cap B^c$ より明らか .

(2) : 条件から

$$A_n^c \in \mathcal{F}, n = 1, 2, \dots \implies \bigcup_{n=1}^{\infty} A_n^c \in \mathcal{F} \implies \left(\bigcup_{n=1}^{\infty} A_n^c \right)^c \in \mathcal{F}$$

ここで de Morgan の法則を使って

$$\left(\bigcup_{n=1}^{\infty} A_n^c \right)^c = \bigcap_{n=1}^{\infty} (A_n^c)^c = \bigcap_{n=1}^{\infty} A_n$$

より，求める結果を得る . □

確率空間

基本的に σ -集合体では加算個の演算が自由にできる . 確率論では可測空間に，確率 P を付加したものを考える .

定義 1.3. 可測空間 (Ω, \mathcal{F}) 上の測度 P で $P(\Omega) = 1$ をみたすものを確率測度 (probability measure) という . すなわち次の条件がみたされる：

- (1) $P: \mathcal{F} \rightarrow [0, 1], P(\Omega) = 1$.
- (2) $A_n \in \mathcal{F}, n = 1, 2, \dots$ が互いに素 ($A_i \cap A_j = \emptyset, i \neq j$) であるとき，

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{i=1}^{\infty} P(A_n) \tag{1.1}$$

が成り立つ .

これらを組にした (Ω, \mathcal{F}, P) を確率空間 (probability space) という .

Ω を全事象，または標本空間 (sample space) という . Ω の要素 ω を根元事象 (elementary event) または標本 (sample) という . \mathcal{F} の要素 A を事象 (event) といい，その補集合 $A^c = \Omega \setminus A$ を余事象 (complementary event) という . $A \cap B$ を積事象， $A \cup B$ を和事象， \emptyset を空事象と呼ぶ .

例 1.1. サイコロ投げの場合

確率空間として次のものを準備すればよい.

$$\Omega = \{1, 2, \dots, 6\}^{\mathbb{N}} \ni \omega = (\omega_1, \omega_2, \dots).$$

ω_n は $1, 2, \dots, 6$ のいずれかで, n 回目に出た目を表す. 確率は $\eta_1, \eta_2, \dots, \eta_n$ を与えて

$$P(\omega_1 = \eta_1, \omega_2 = \eta_2, \dots, \omega_n = \eta_n) = \frac{1}{6^n}$$

と定めればよい. これが実際に σ -加法的に拡張できることは明らかではないが, Kolmogorov の拡張定理と呼ばれる定理により証明できる.

命題 1.4. 確率空間 (Ω, \mathcal{F}, P) において次のことが成り立つ:

- (1) $A \subseteq B \implies P(B \setminus A) = P(B) - P(A)$.
- (2) $P(A^c) = 1 - P(A)$
- (3) $A \subseteq B \implies P(A) \leq P(B)$.
- (4) 任意の $A_n \in \mathcal{F}, n = 1, 2, \dots$ に対し $P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{i=1}^{\infty} P(A_n)$.
- (5) $A_n \uparrow A$ (i.e., $A_1 \subseteq A_2 \subseteq \dots, A = \bigcup_{n=1}^{\infty} A_n$) のとき, $\lim_{n \rightarrow \infty} P(A_n) = P(A)$.
- (6) $A_n \downarrow A$ (i.e., $A_1 \supseteq A_2 \supseteq \dots, A = \bigcap_{n=1}^{\infty} A_n$) のとき, $\lim_{n \rightarrow \infty} P(A_n) = P(A)$.

証明 (1): $B = A + B \setminus A$ (disjoint union) より明らか.

(2): $A^c = \Omega \setminus A$ と $P(\Omega) = 1$ から明らか.

(3): (1) と確率の正値性から明らか.

(4): $B_1 = A_1, B_n = A_n \setminus \bigcup_{i=1}^{n-1} A_i$ ($n = 2, 3, \dots$) とおく. B_i は互いに素で

$$\sum_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i, \quad B_i \subseteq A_i.$$

よって, 完全加法性から

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

より, 求める結果を得る.

(5):

$$P(A) = P(A_n) + \sum_{k=n}^{\infty} P(A_{k+1} \setminus A_k).$$

収束性から $\sum_{k=n}^{\infty} P(A_{k+1} \setminus A_k) \rightarrow 0$ が成り立つので求める結果を得る.

(6): de Morgan の法則と (5) を用いればよい. □

系 1.5. 確率空間 (Ω, \mathcal{F}, P) において次のことが成り立つ :

$$(1) P(A_n) = 0, n = 1, 2, \dots \text{ ならば } P\left(\bigcup_{n=1}^{\infty} A_n\right) = 0.$$

$$(2) P(A_n) = 1, n = 1, 2, \dots \text{ ならば } P\left(\bigcap_{n=1}^{\infty} A_n\right) = 1.$$

証明 (1) : 命題 1.4 の (4) を用いればよい .

(2) : (1) の結果と de Morgan の法則を使う . □

2. 確率変数

確率変数

定義 2.1. (Ω, \mathcal{F}, P) を確率空間 , (S, \mathcal{S}) を可測空間とする . Ω から S への \mathcal{F}/\mathcal{S} 可測写像 $X: \Omega \rightarrow S$ を確率変数と呼ぶ . ここに X が \mathcal{F}/\mathcal{S} 可測写像であるとは , 任意の $B \in \mathcal{S}$ に対し , $X^{-1}(B) = \{\omega; X(\omega) \in B\} \in \mathcal{F}$ が成り立つことをいう .

多くの場合 S は位相空間で , このときは断らない限り , $\mathcal{S} = \mathcal{B}(S)$ とする . 特に $S = \mathbb{R}$ のとき , X を実確率変数 , $S = \mathbb{C}$ のとき , 複素確率変数 , $S = \mathbb{R}^d$ のとき , d 次元確率変数という .

分布

定義 2.2. (確率変数の分布) X を (S, \mathcal{S}) -値確率変数とするととき , (S, \mathcal{S}) 上に導入される確率測度 $P \circ X^{-1}$ (即ち $(P \circ X^{-1})(B) = P[X^{-1}(B)]$, $B \in \mathcal{S}$, で定義される (S, \mathcal{S}) 上の確率測度) を X の分布といい , P^X で表わす .

定義 2.3. 同じ値空間 (S, \mathcal{S}) をもつ 2 つの確率変数 X, Y ((必ずしも同一確率空間上で定義されている必要はない) に対し , $P^X = P^Y$ が成り立つとき , X と Y は同分布をもつ (同法則である) といい ,

$$X \stackrel{d}{=} Y, \text{ あるいは } X \stackrel{\mathcal{L}}{\approx} Y$$

と表わす .

定義 2.4. (分布関数) X を実確率変数 , P^X をその \mathbb{R} 上の分布とする . $F(x) = P(X \leq x) = P^X((-\infty, x])$, $x \in \mathbb{R}$, で定義される \mathbb{R} 上の関数 F を X の分布関数という .

分布関数 F は右連続 , 単調非減少で $\lim_{x \rightarrow -\infty} F(x) = 0$, $\lim_{x \rightarrow \infty} F(x) = 1$ が成り立つ . また逆にこの性質が満たされる関数が与えられれば , これから分布が定まる .

期待値

次に実確率変数 X の期待値 $E[X]$ を定義する．これは確率測度による積分

$$E[X] = \int_{\Omega} X(\omega)P(d\omega)$$

として定義されるものであるが，右辺の確率測度 P による積分は以下のように定義されるものである．

X が非負の単関数の場合，すなわち Ω の分割 $\Omega = \bigcup_{k=1}^N \Omega_k$ ($\Omega_k \in \mathcal{F}$) が存在し，

$$X(\omega) = \sum_{k=1}^N a_k 1_{\Omega_k}(\omega)$$

と表される場合，

$$\int_{\Omega} X(\omega)P(d\omega) = \sum_{k=1}^N a_k P(\Omega_k)$$

で定義する．次に非負確率変数 X が単関数の増加極限

$$X(\omega) = \lim_{n \rightarrow \infty} X_n(\omega), \quad X_n(\omega) \leq X_{n+1}(\omega), \quad n = 1, 2, \dots$$

となっているとき，

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

この極限は増加列 $\{X_n\}$ のとり方に依らない．この値が有限のとき X は P に関して可積分であるという． X_n の例として

$$X_n(\omega) = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} 1_{E_k}(\omega) + n 1_{F_n}(\omega) \quad (2.1)$$

がとれる．ここで

$$E_k = \left\{ \omega; \frac{k-1}{2^n} \leq X(\omega) < \frac{k}{2^n} \right\}, \quad k = 1, 2, \dots, n2^n,$$

$$F_n = \left\{ \omega; X(\omega) \geq n \right\}$$

である．従って

$$\int_{\Omega} X(\omega)P(d\omega) = \lim_{n \rightarrow \infty} \left\{ \sum_{k=1}^{n2^n} \frac{k-1}{2^n} P\left(\frac{k-1}{2^n} \leq X < \frac{k}{2^n}\right) + nP(X \geq n) \right\}$$

が成立している．右辺を X の P による積分と定義してもよい．

X が一般の場合は $|X|$ が可積分の場合に可積分と呼び

$$\int_{\Omega} X(\omega)P(d\omega) = \int_{\Omega} X_+(\omega)P(d\omega) - \int_{\Omega} X_-(\omega)P(d\omega)$$

で定義する．ただし $X_+ = X \vee 0$, $X_- = (-X) \vee 0$. 可積分関数全体を $L^1(P)$ で表す．また $p \geq 1$ に対し， $|X|^p$ が可積分なとき X は p -乗可積分であるといい，その全体を $L^p(P)$ とかく．

定義 2.5. $X \in L^1(P)$ のとき

$$E[X] = \int_{\Omega} X(\omega)P(d\omega) \quad (2.2)$$

を X の期待値 (平均) という.

平均に関して次のことは定義から容易に確かめられる.

命題 2.6. $X, Y \in L^1(P), \alpha, \beta \in \mathbb{R}$ に対し

$$\begin{aligned} X \geq 0 &\implies E[X] \geq 0, && \text{正值性} \\ E[\alpha X + \beta Y] &= \alpha E[X] + \beta E[Y], && \text{線形性} \end{aligned}$$

が成り立つ.

命題 2.7. (置換積分) X を (S, \mathcal{S}) に値をとる確率変数とする. また f を (S, \mathcal{S}) 上の実数値可測関数とする. 実確率変数 $f(X)$ が確率 P に関し可積分のとき, $f(x)$ は S 上 P^X に関し可積分で, 次の公式が成り立つ:

$$E[f(X)] = \int_{\Omega} f(X(\omega))P(d\omega) = \int_S f(x)P^X(dx). \quad (2.3)$$

右辺は確率測度 P^X による積分である.

証明 f が単関数の場合を示せばよい.

$$f(x) = \sum_{k=1}^n a_k 1_{B_k}(x)$$

とすると,

$$f(X) = \sum_{k=1}^n a_k 1_{B_k}(X) = \sum_{k=1}^n a_k 1_{X^{-1}(B_k)}.$$

よって

$$E[f(X)] = \sum_{k=1}^n a_k P(X^{-1}(B_k)) = \sum_{k=1}^n a_k P^X(B_k) = \int_S f(x)P^X(dx).$$

一般の場合は極限を取ればよい. □

モーメント, 分散, 標準偏差

定義 2.8. $X^n \in L^1(P)$ のとき $E[X^n]$ を n 次のモーメントという.

$X^2 \in L^1(P)$ のとき

$$V(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2 \quad (2.4)$$

を X の分散といい, $\sigma(X) = \sqrt{V(X)}$ を標準偏差という.

さて，積分に関連してよく使われる不等式を述べておく．

命題 2.9. (Chebyshev の不等式) $X \in L^p(P)$ ($p \geq 1$) に対し次が成り立つ：

$$P(|X| \geq k) \leq \frac{E[|X|^p]}{k^p}. \quad (2.5)$$

また $X \in L^2(P)$ に対し

$$P\left(\frac{|X - m|}{\sigma} \geq k\right) \leq \frac{1}{k^2} \quad (2.6)$$

が成り立つ．ここに m は平均， σ は標準偏差である．

証明 $|X|^p \geq k^p 1_{\{|X| \geq k\}}$ に注意すれば

$$E[|X|^p] \geq E[k^p 1_{\{|X| \geq k\}}] = k^p P(\{|X| \geq k\})$$

から (2.5) は明らか．

また

$$\begin{aligned} \sigma^2 &= E[|X - m|^2] \\ &\geq E[\sigma^2 k^2 1_{\{|X - m|^2 \geq \sigma^2 k^2\}}] \\ &= \sigma^2 k^2 P(\{|X - m|^2 \geq \sigma^2 k^2\}) \\ &= \sigma^2 k^2 P\left(\frac{|X - m|}{\sigma} \geq k\right) \end{aligned}$$

であるから，(2.6) が従う． □

最後に，平均の意味を分散と関連させて見てみよう． X を確率変数として，次の関数を考える：

$$f(x) = E[(X - x)^2].$$

この最小値を求めてみると， $m = E[X]$ として

$$\begin{aligned} f(x) &= E[(X - m + m - x)^2] = E[(X - m)^2 + 2(X - m)(x - m) + (m - x)^2] \\ &= V(X) + (m - x)^2 \end{aligned}$$

従って， $x = m$ のとき最小値 $V(X)$ を取ることが分かる． $f(x)$ は X を定数で近似するときの 2 乗平均誤差を表している．つまり平均は 2 乗平均誤差を最小とし，そのときの誤差が分散であることが分かる．このように 2 乗の平均で距離を測るといことはしばしば行われている．

3. 独立性と条件付確率

独立性

定義 3.1. 2つの事象 $A, B \in \mathcal{F}$ が独立 $\stackrel{\text{def}}{\iff} P(A \cap B) = P(A)P(B)$.

定義 3.2. 2つの sub σ -fields $\mathcal{F}_1, \mathcal{F}_2 \subseteq \mathcal{F}$ が独立 $\stackrel{\text{def}}{\iff} \forall A \in \mathcal{F}_1, \forall B \in \mathcal{F}_2: P(A \cap B) = P(A)P(B)$.

$A \in \mathcal{F}$ に対し, A を含む最小の σ -集合体を $\sigma(A)$ とかく. すなわち

$$\sigma(A) = \{\emptyset, \Omega, A, A^c\}.$$

この記法を用いれば, $A, B \in \mathcal{F}$ に対し

$$A, B \text{ が独立} \iff \sigma(A), \sigma(B) \text{ が独立}$$

であることが容易にわかる. たとえば A, B が独立のとき, $P(A^c \cap B) = P(A^c)P(B)$ は

$$P(A \cap B) + P(A^c \cap B) = P(B)$$

を用いて

$$P(A^c \cap B) = P(B) - P(A \cap B) = P(B) - P(A)P(B) = P(B)(1 - P(A)) = P(B)P(A^c)$$

より確かめられる.

定義 3.3. n 個の sub σ -fields $\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_n \subseteq \mathcal{F}$ が独立 $\stackrel{\text{def}}{\iff} \forall A_i \in \mathcal{F}_i, i = 1, 2, \dots, n: P\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n P(A_i)$.

注意 3.1. $\sigma(A), \sigma(B), \sigma(C)$ が独立のとき, A, B, C は独立という. 単に

$$P(A \cap B \cap C) = P(A)P(B)P(C)$$

が成り立つとき, A, B, C を独立と呼んではいけない.

定義 3.4. σ -fields $\mathcal{F}_\lambda \subseteq \mathcal{F}, \lambda \in \Lambda$ が独立 $\stackrel{\text{def}}{\iff}$ 任意の有限個の sub σ -fields が独立.

定義 3.5. X を (S, \mathcal{S}) に値をとる確率変数とするととき σ -集合体

$$\sigma(X) = \{A = X^{-1}(B); B \in \mathcal{S}\}$$

を X で生成される σ -集合体という.

確率変数の族 $\{X_\lambda; \lambda \in \Lambda\}$ が独立であるとは σ -集合体の族 $\{\sigma(X_\lambda); \lambda \in \Lambda\}$ が独立であるときと定義する.

独立確率変数に対して, 次の定理は重要である.

定理 3.6. X, Y を独立確率変数とする. $X, Y \in L^1(P)$ ならば $XY \in L^1(P)$ で

$$E[XY] = E[X]E[Y] \quad (3.1)$$

が成立する.

証明 X, Y が単関数のときを示す. Ω の分割 $\Omega = \sum_i \Omega_i$ と $\Omega = \sum_j \Omega'_j$ が存在して,

$$X = \sum_i a_i 1_{\Omega_i}, \quad Y = \sum_j b_j 1_{\Omega'_j}$$

と表されているとする.

$$\begin{aligned} E[XY] &= E\left[\left(\sum_i a_i 1_{\Omega_i}\right)\left(\sum_j b_j 1_{\Omega'_j}\right)\right] \\ &= E\left[\sum_{i,j} a_i b_j 1_{\Omega_i \cap \Omega'_j}\right] \\ &= \sum_{i,j} a_i b_j P(\Omega_i \cap \Omega'_j) \\ &= \sum_{i,j} a_i b_j P(\Omega_i)P(\Omega'_j) \\ &= \sum_i a_i P(\Omega_i) \sum_j b_j P(\Omega'_j) \\ &= E[X]E[Y]. \end{aligned}$$

一般の X, Y の場合は近似の列 X_n, Y_n を (2.1) のようにとればそれぞれ $\sigma(X), \sigma(Y)$ 可測になるから, 独立性が保存される. あとは極限をとればよい. \square

独立性は, いろいろなところで計算を簡略にする. 一つの例として分散を考えてみよう.

命題 3.7. X_1, X_2, \dots, X_n が独立のとき,

$$V(a_1 X_1 + \dots + a_n X_n) = a_1^2 V(X_1) + \dots + a_n^2 V(X_n) \quad (3.2)$$

が成立する.

証明 m_j を X_j の平均とするとき

$$\begin{aligned} V(a_1 X_1 + \dots + a_n X_n) &= E[(a_1 X_1 + \dots + a_n X_n - a_1 m_1 - \dots - a_n m_n)^2] \\ &= E\left[\left(\sum_j a_j (X_j - m_j)\right)^2\right] \\ &= \sum_{i,j} a_i a_j E[(X_i - m_i)(X_j - m_j)] \end{aligned}$$

$$\begin{aligned}
&= \sum_i a_i^2 E[(X_i - m_i)^2] + \sum_{i \neq j} a_i a_j E[(X_i - m_i)(X_j - m_j)] \\
&= \sum_i a_i^2 V(X_i) + \sum_{i \neq j} a_i a_j E[X_i - m_i] E[X_j - m_j] \\
&= \sum_i a_i^2 V(X_i).
\end{aligned}$$

これが示すべきことであった。 □

X を (S_1, \mathcal{S}_1) -値確率変数, Y を (S_2, \mathcal{S}_2) -値確率変数とし, P^X, P^Y をそれぞれの分布とする. X, Y を組にした確率変数 (X, Y) は $(S_1 \times S_2, \mathcal{S}_1 \times \mathcal{S}_2)$ -値確率変数となる. ここで $\mathcal{S}_1 \times \mathcal{S}_2$ は $A \times B$, の形の集合を含む最小の σ -集合体である. その分布を $P^{(X, Y)}$ とかく. X と Y が独立のとき, $A \in \mathcal{S}_1, B \in \mathcal{S}_2$ に対し

$$P^{(X, Y)}(A \times B) = P(X \in A, Y \in B) = P(X \in A)P(Y \in B) = P^X(A)P^Y(B)$$

が成り立つ. $P^{(X, Y)}(A \times B) = P^X(A)P^Y(B)$ がすべての A, B に対して成り立つとき, 測度 $P^{(X, Y)}$ を P^X, P^Y の直積測度と呼び, $P^X \times P^Y$ とかく. すなわち, 独立確率変数の同時分布は直積測度で与えられる.

次に \mathbb{R}^d の上の確率測度 μ, ν が与えられたとき, 確率測度 λ を

$$\lambda(A) = \int_{\mathbb{R}^d} \mu(A - x)\nu(dx), \quad A \in \mathcal{B}(\mathbb{R}^d)$$

で定めるとき, この λ を μ と ν の合成積と呼び $\mu * \nu$ とかく. 合成積は確率論的には, 独立確率変数の和の分布を意味している. すなわち \mathbb{R}^d -値確率変数 X, Y の分布がそれぞれ μ, ν であるとき, $X + Y$ の分布が $\mu * \nu$ で与えられる. このことは

$$\begin{aligned}
P(X + Y \in A) &= \int_{\mathbb{R}^d \times \mathbb{R}^d} 1_A(x + y)\mu(dx)\nu(dy) \\
&= \int_{\mathbb{R}^d} \nu(dy) \int_{\mathbb{R}^d} 1_A(x + y)\mu(dx) \\
&= \int_{\mathbb{R}^d} \mu(A - y)\nu(dy)
\end{aligned}$$

から明らかである.

分布が密度関数 f, g を持つ場合は, 合成積は

$$f * g(x) = \int_{-\infty}^{\infty} f(x - y)g(y) dy$$

で定義される. すなわち X, Y を独立な確率変数で, 密度関数 f, g を持つとするとき, $f * g$ は $X + Y$ の密度関数になっているのである. 実際

$$E[F(X + Y)] = \iint F(x + y)f(x)g(y) dx dy$$

$$\begin{aligned}
u &= x + y, v = y \\
\frac{\partial(x, y)}{\partial(u, v)} &= \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} = 1 \\
dx dy &= \left| \frac{\partial(x, y)}{\partial(u, v)} \right| du dv = du dv \\
&= \iint F(u)f(u-v)g(v) du dv \\
&= \int F(u)f * g(u) du.
\end{aligned}$$

条件付確率

定義 3.8. $A, B \in \mathcal{F}, P(B) \neq 0$ に対し

$$P(A|B) := \frac{P(A \cap B)}{P(B)} \quad (3.3)$$

を条件 B の下での A の条件付確率という.

命題 3.9.

$$P(A \cap B) = P(A|B)P(B) \quad (3.4)$$

が成立し,

$$A, B \text{ が独立} \iff P(A|B) = P(A) \quad (3.5)$$

である.

Bayes の公式

定理 3.10. (Bayes の公式)

$A_j, j = 1, \dots, n$ を $\sum_{j=1}^n A_j = \Omega$ となる排反事象とするとき

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^n P(A_j)P(B|A_j)}, \quad i = 1, \dots, n \quad (3.6)$$

が成立する.

証明

$$P(B) = \sum_{j=1}^n P(B \cap A_j) = \sum_{j=1}^n P(B|A_j)P(A_j)$$

であるから, 定理を示すには

$$P(A_i|B)P(B) = P(B|A_i)P(A_i)$$

が成り立つことを言えばよいが, 両辺ともに $P(A \cap B)$ に等しい.

□

この公式は、偶然現象において、次のような解釈のもとに応用される。事象 A_i は可能な原因の一つを表わし、原因 A_i から結果 B が起こる確率が $P(B|A_i)$ で、これは予め判っているとす。このとき原因 A_i の事前確率 $P(A_i)$ がなんらかの根拠に基づいて定め得たとすると、結果 B を観測したときの原因 A_i の事後確率 $P(A_i|B)$ が上の公式より求められる。このようにベイズの公式は、結果を観測してその原因を推測するという典型的な帰納的推論の方法を与える。

例 3.1. (3 囚人のディレンマ)

a, b, c 3 人の囚人がいて、1 人無罪、2 人有罪である。無罪である確率は 3 人とも等しく $\frac{1}{3}$ であるとする。囚人 a が看守に「囚人 b, c のいずれかは有罪なのだから、どちらが有罪か教えて欲しい」と頼み、b が有罪であると教えられた。a は a, c 2 人のうちのどちらかは無罪なのだから、無罪になる確率が $\frac{1}{3}$ から $\frac{1}{2}$ になったと喜んだ。これは正しい推論か？

以下この問題を Bayes の公式を用いて検討しよう。A, B, C をそれぞれ a, b, c が無罪である事象とする。G を看守が b が有罪であることを言明する事象とする。明らかに $G \subset B^c$ である。(看守は b が有罪であっても、b が有罪であると言明しない場合もある。即ち b, c ともに有罪であるが、c が有罪であると言明するときである。) さらに A, B, C は排反事象で $A \cup B \cup C = \Omega$ である。また次が成り立つ(とする)：

$$P(G|A) = \frac{1}{2}, \quad P(G|B) = 0, \quad P(G|C) = 1.$$

求める確率は $P(A|G)$ である。Bayes の公式からこれは次で与えられる：

$$P(A|G) = \frac{P(G|A)P(A)}{P(G|A)P(A) + P(G|B)P(B) + P(G|C)P(C)} = \frac{\frac{1}{2}}{\frac{1}{2} + 0 + 1} = \frac{1}{3}.$$

これは a が無罪である確率には変化がなく、 $\frac{1}{3}$ のままであることを意味している。即ち a の喜びは糠喜びでしかなかった。一方 c について考えれば、 $P(G|C) = 1$ だから

$$P(C|G) = \frac{P(G|C)P(C)}{P(G|A)P(A) + P(G|B)P(B) + P(G|C)P(C)} = \frac{1}{\frac{1}{2} + 0 + 1} = \frac{2}{3}$$

となる。即ち、c は無罪である確率が $\frac{2}{3}$ となったのであるから、c こそ喜ぶべきなのである。

例 3.2. (癌検診)

次のような条件の下で行う：

検出率	99%
擬陽性率	2%

ここで検出率は実際の癌患者に陽性の判定が出る割合であり、擬陽性率は実際には癌でないにもかかわらず陽性の判定が出る割合である。

癌は 1000 人に 1 人の割合であるとしよう。陽性の判定を受けたときに、実際に癌である確率はどうなるであろうか？

C	...	癌
P	...	陽性

として、条件は

$$P(P|C) = \frac{99}{100}, \quad P(P|C^c) = \frac{2}{100}, \quad P(C) = \frac{1}{1000}$$

で与えられるので、

$$\begin{aligned} P(C|P) &= \frac{P(P|C)P(C)}{P(P|C)P(C) + P(P|C^c)P(C^c)} \\ &= \frac{\frac{99}{100} \times \frac{1}{1000}}{\frac{99}{100} \times \frac{1}{1000} + \frac{2}{100} \times \frac{999}{1000}} = \frac{99}{2097} = 0.04721\dots \end{aligned}$$

Markov 連鎖

例 3.3. K 先生は大へん休講が好きで、1 回講義をすると次の講義を休講にする確率は 0.6 であり、1 度休講にすると、さすがに気が引けるのか、次の週が休講になる確率は 0.1 であるという。n 回目の講義が休講である確率を $p(n)$ とするとき、 $p(n)$ の極限を求めよ。

(注意：初回の確率は上の規則からは決まらないが、それに関係なく上の極限が存在することを示すこと)

この問題を次のように定式化する。確率変数 $\{X_n\}$ で n 回目の状態を

$$X_n = \begin{cases} 1, & \text{休講の場合} \\ 2, & \text{開講の場合} \end{cases}$$

のように表す。n 回目の状態が与えられたとき、n + 1 回目の状態の条件付確率を次のように置く：

$$p(i, j) = P(X_{n+1} = i | X_n = j), \quad i, j = 1, 2.$$

問題で与えられている条件は

$$\begin{pmatrix} p(1, 1) & p(1, 2) \\ p(2, 1) & p(2, 2) \end{pmatrix} = \begin{pmatrix} 0.1 & 0.6 \\ 0.9 & 0.4 \end{pmatrix}$$

ということであるが、特に具体的な数値は必要ないので一般形で述べる。

$$p(n) = P(X_n = 1), \quad q(n) = P(X_n = 2) = 1 - p(n)$$

とおくと

$$\begin{aligned} p(n+1) &= P(X_{n+1} = 1) = P(X_{n+1} = 1, X_n = 1) + P(X_{n+1} = 1, X_n = 2) \\ &= P(X_{n+1} = 1 | X_n = 1)P(X_n = 1) + P(X_{n+1} = 1 | X_n = 2)P(X_n = 2) \\ &= p(1, 1)p(n) + p(1, 2)(1 - p(n)) = (p(1, 1) - p(1, 2))p(n) + p(1, 2). \end{aligned}$$

ここで

$$x = (p(1, 1) - p(1, 2))x + p(1, 2)$$

の解を $x = \pi(1)$ とする：

$$\pi(1) = (p(1, 1) - p(1, 2))\pi(1) + p(1, 2). \quad (3.7)$$

従って

$$p(n+1) - \pi(1) = (p(1, 1) - p(1, 2))(p(n) - \pi(1)) = \cdots = (p(1, 1) - p(1, 2))^n (p(1) - \pi(1)).$$

ここで $|p(1, 1) - p(1, 2)| < 1$ であるから

$$\lim_{n \rightarrow \infty} p(n+1) = \pi(1).$$

これが求める結果である．

[別解] 行列を使って解く．上と同様に考えて $p(n+1), q(n+1)$ を $p(n), q(n)$ で表せば

$$\begin{pmatrix} p(n+1) \\ q(n+1) \end{pmatrix} = \begin{pmatrix} p(1, 1) & p(1, 2) \\ p(2, 1) & p(2, 2) \end{pmatrix} \begin{pmatrix} p(n) \\ q(n) \end{pmatrix}$$

ここで

$$A = \begin{pmatrix} p(1, 1) & p(1, 2) \\ p(2, 1) & p(2, 2) \end{pmatrix}$$

の固有値を計算する．固有方程式

$$\begin{vmatrix} x - p(1, 1) & -p(1, 2) \\ -p(2, 1) & x - p(2, 2) \end{vmatrix} = 0$$

を解いて

$$\begin{aligned} (x - p(1, 1))(x - p(2, 2)) - p(1, 2)p(2, 1) &= 0 \\ x^2 - (p(1, 1) + p(2, 2))x + p(1, 1)p(2, 2) - p(1, 2)p(2, 1) &= 0 \\ \begin{matrix} \downarrow & & \downarrow & & \downarrow \\ 1 - p(1, 2) & & 1 - p(1, 2) & & 1 - p(1, 1) \end{matrix} & \\ x^2 - (p(1, 1) - p(1, 2) + 1)x + p(1, 1)(1 - p(1, 2)) - p(1, 2)(1 - p(1, 1)) &= 0 \\ x^2 - (p(1, 1) - p(1, 2) + 1)x + p(1, 1) - p(1, 2) &= 0 \\ (x - 1)(x - p(1, 1) + p(1, 2)) &= 0 \\ x = 1, p(1, 1) - p(1, 2). & \end{aligned}$$

これで A の固有値が求まった．固有値 $x = 1$ に対する固有ベクトルを $\begin{pmatrix} \pi(1) \\ \pi(2) \end{pmatrix}$ として，

$\pi(1) + \pi(2) = 1$ の条件の下で求める：

$$\begin{pmatrix} \pi(1) \\ \pi(2) \end{pmatrix} = \begin{pmatrix} p(1, 1) & p(1, 2) \\ p(2, 1) & p(2, 2) \end{pmatrix} \begin{pmatrix} \pi(1) \\ \pi(2) \end{pmatrix} \quad \cdots \cdots \quad \text{この解} \begin{pmatrix} \pi(1) \\ \pi(2) \end{pmatrix} \text{を不変測度という}$$

より

$$\begin{aligned}\pi(1) &= p(1,1)\pi(1) + p(1,2)\pi(2) = p(1,1)\pi(1) + p(1,2)(1 - \pi(1)) \\ &= (p(1,1) - p(1,2))\pi(1) + p(1,2).\end{aligned}$$

これは (*) と同じであることを注意しよう．すなわち (*) は固有ベクトル (不変測度) を求める式だったわけである． $x = p(1,1) - p(1,2)$ に対する固有ベクトルは $\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ とする (具体的にもとめる必要はない)．2つの固有ベクトルは線型独立だから

$$\begin{pmatrix} p(1) \\ q(1) \end{pmatrix} = a \begin{pmatrix} \pi(1) \\ \pi(2) \end{pmatrix} + b \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

と表すことが出来るので

$$\begin{aligned}\begin{pmatrix} p(n+1) \\ q(n+1) \end{pmatrix} &= A^n \begin{pmatrix} p(1) \\ q(1) \end{pmatrix} = aA^n \begin{pmatrix} \pi(1) \\ \pi(2) \end{pmatrix} + bA^n \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \\ &= a \begin{pmatrix} \pi(1) \\ \pi(2) \end{pmatrix} + b(p(1,1) - p(1,2))^n \begin{pmatrix} \alpha \\ \beta \end{pmatrix}.\end{aligned}$$

よって

$$\lim_{n \rightarrow \infty} \begin{pmatrix} p(n) \\ q(n) \end{pmatrix} = a \begin{pmatrix} \pi(1) \\ \pi(2) \end{pmatrix}.$$

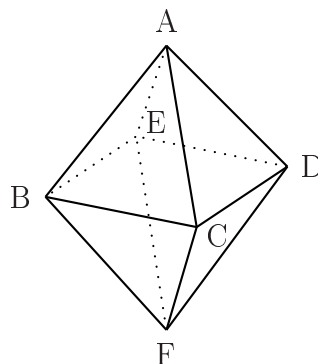
ここで $p(n) + q(n) = 1$, $\pi(1) + \pi(2) = 1$ から $a = 1$ が従う．

一般的にマルコフ連鎖は, 不変測度に収束することが知られているが, 今の場合は2状態なので具体的に計算できたわけである．

最後に (3.7) を与えられた問題の数値を用いて計算すれば $\pi(1) = 0.4$ が得られる．

問題

- 硬貨を3枚を投げて, 表の出た枚数を数える．この試行を繰り返し, 表の出た枚数を加えていく． n 回目に表の出た枚数の累積が初めて10枚以上になる確率を求めよ．
- 右の図のような正八面体 ABCDEF がある．時刻0にAを出発して, それぞれ隣りの頂点へ $1/4$ ずつの確率で移動していく運動を考える．従って時刻1では, B, C, D, Eのいずれかの点にそれぞれ $1/4$ の確率で移動する．時刻2ではさらに隣りの頂点に移動していく．この運動を繰り返していくとき, 時刻 n で初めて頂点Fに到達する確率を求めよ．



第2章 確率分布

この章でいろいろな分布を扱う。

1. 離散分布

\mathbb{R} の分布 μ が離散分布であるとは、高々可算集合 $\{a_i\} \subseteq \mathbb{R}$ と正数列 p_i で $\sum_i p_i = 1$ を満たすものが存在し、 $\mu = \sum_i p_i \delta_{a_i}$ と表わされる分布 (δ_a は、点 a における Dirac 測度) とあらわされることである。確率変数で言えば、 Ω の直和分割 $\Omega = \bigcup_{i=1}^{\infty} \Omega_i$ で $P(\Omega_i) = p_i$ となるものが存在し、 $X = \sum_{i=1}^{\infty} a_i 1_{\Omega_i}$ とあらわされることである。

2 項分布

2 項分布 $B(n, p)$ はパラメータとして $n = 1, 2, \dots, 0 < p < 1$ を持ち

$$\mu = \sum_{k=0}^n {}_n C_k p^k (1-p)^{n-k} \delta_k, \quad {}_n C_k = \binom{n}{k} = \frac{n!}{k!(n-k)!} \quad (1.1)$$

で与えられる分布で、 $n+1$ 個の点 $k = 0, 1, \dots, n$ の上においており、それぞれの確率が ${}_n C_k p^k (1-p)^{n-k}$ である。

今 X_1, X_2, \dots を $P(X_i = 1) = p, P(X_i = 0) = 1-p$ となる独立、かつ同分布な確率変数列 (簡単に、i.i.d. = independent identically distributed 確率変数列という) とするとき、

$$S_n = X_1 + \dots + X_n \quad (1.2)$$

の分布が 2 項分布 $B(n, p)$ である即ち、成功の確率が p であるような試行を何回も繰り返すとき、 n 回の試行における成功の回数の従う分布が 2 項分布 $B(n, p)$ である。このような試行をベルヌーイ (Bernoulli) 試行、また、確率変数列 X_1, X_2, \dots をベルヌーイ列という。

平均は np 、分散は $np(1-p)$ である。このことを確かめよう。2 項展開

$$(x+y)^n = \sum_{k=0}^n {}_n C_k x^k y^{n-k}$$

を x で微分して

$$n(x+y)^{n-1} = \sum_{k=0}^n k {}_n C_k x^{k-1} y^{n-k} \quad (1.3)$$

ここで $x = p, y = 1-p$ として両辺に p をかければ

$$np = \sum_{k=0}^n k {}_n C_k p^k (1-p)^{n-k} = \sum_{k=0}^n k p_k = E[S_n] \quad (1.4)$$

で平均が求まる．分散は (1.3) をさらに微分して

$$n(n-1)(x+y)^{n-2} = \sum_{k=0}^n k(k-1) {}_n C_k x^{k-2} y^{n-k} \quad (1.5)$$

ここで $x = p, y = 1 - p$ として両辺に p^2 をかければ

$$n(n-1)p^2 = \sum_{k=0}^n k(k-1) {}_n C_k p^k (1-p)^{n-k} = \sum_{k=0}^n k(k-1) p_k = E[S_n(S_n - 1)]. \quad (1.6)$$

分散 $V(S_n)$ は

$$\begin{aligned} V(S_n) &= E[S_n^2] - E[S_n]^2 = E[S_n(S_n - 1)] + E[S_n] - E[S_n]^2 \\ &= n(n-1)p^2 + np - (np)^2 = -np^2 + np = np(1-p) \end{aligned}$$

である．上の $E[S_n(S_n - 1)]$ の形の積分は階乗モーメントと呼ばれることがある．一般には $E[X(X-1)\cdots(X-n)]$ の形の積分である．

これらの計算は (1.3) の表現を用いた方が容易である． $E[X_k] = p, E[X_k^2] = p$ であるから

$$\begin{aligned} E[S_n] &= nE[X_1] = np \\ V[S_n] &= nV[X_1] = n(p - p^2) = np(1-p). \end{aligned}$$

幾何分布

幾何分布 (geometric distribution) $G(p)$ はパラメータ $0 < p < 1$ を持つ次の

$$\mu = \sum_{k=0}^{\infty} p(1-p)^k \delta_k \quad (1.7)$$

で与えられる分布で， $\{0, 1, 2, \dots\}$ の上の上のっている．ベルヌーイ試行において最初の成功が達成されるまでの待ち時間がこの分布に従う．これは初めて成功するまでの失敗の回数で，何回目に始めて成功したかという試行回数より，一つ少ない．

平均は $\frac{1-p}{p}$ ，分散は $\frac{1-p}{p^2}$ である．このことを確かめよう． X を幾何分布を持つ確率変数とする．幾何級数の等式

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k, \quad |x| < 1$$

を微分して

$$\frac{1}{(1-x)^2} = \sum_{k=0}^{\infty} kx^{k-1}. \quad (1.8)$$

ここで $x = 1 - p$ とおいて両辺に $p(1-p)$ を掛ければ

$$\frac{1-p}{p} = \sum_{k=0}^{\infty} kp(1-p)^k = E[X]. \quad (1.9)$$

これは平均が $\frac{1-p}{p}$ であることを意味する .

分散を計算するには , (1.8) をさらに微分して ,

$$\frac{2}{(1-x)^3} = \sum_{k=0}^{\infty} k(k-1)x^{k-2}. \quad (1.10)$$

$x = 1-p$ において , 両辺に $p(1-p)^2$ を掛ければ ,

$$\frac{2(1-p)^2}{p^2} = \sum_{k=0}^{\infty} k(k-1)p(1-p)^k = E[X(X-1)]. \quad (1.11)$$

分散 $V(X)$ は

$$\begin{aligned} V(X) &= E[X(X-1)] + E[X] - E[X]^2 = \frac{2(1-p)^2}{p^2} + \frac{1-p}{p} - \frac{(1-p)^2}{p^2} \\ &= \frac{2(1-p)^2 + p(1-p) - (1-p)^2}{p^2} = \frac{(1-p)(2-2p+p-1+p)}{p^2} = \frac{1-p}{p^2}. \end{aligned}$$

成功するのが時刻 n 以後である事象は $\{X \geq n\}$ であり , その確率は

$$P(X \geq n) = \sum_{k=n}^{\infty} p(1-p)^k = (1-p)^n$$

である . これから

$$\begin{aligned} P(X \geq m+n | X \geq m) &= \frac{P(X \geq m+n, X \geq m)}{P(X \geq m)} = \frac{P(X \geq m+n)}{P(X \geq m)} \\ &= \frac{(1-p)^{m+n}}{(1-p)^m} = (1-p)^n = P(X \geq n) \end{aligned}$$

この確率は n には関係していない . すなわち , ある時刻より前まで成功していないとき , その後にいつ成功するかということに影響しない . この性質を無記憶性という . 逆に無記憶性を持つ離散分布は幾何分布となる .

ポアソン分布

ポアソン分布 (Poisson distribution) $P(\lambda)$ はパラメーター $\lambda > 0$ をもつ次の分布

$$\mu = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} \delta_k \quad (1.12)$$

である .

平均 λ , 分散 λ であることをまず確かめておこう . X をポアソン分布に従う確率変数とする .

$$E[X] = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} \lambda \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} = \lambda \quad (1.13)$$

から平均が求まる．さらに

$$E[X(X-1)] = \sum_{k=0}^{\infty} k(k-1) \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=2}^{\infty} \lambda^2 \frac{\lambda^{k-2}}{(k-2)!} e^{-\lambda} = \lambda^2 \quad (1.14)$$

分散 $V(X)$ は

$$V(X) = E[X(X-1)] + E[X] - E[X]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

となる．

ポアソン分布は，2項分布 $B(n, p)$ で $np \rightarrow \lambda$ の条件のもとで $n \rightarrow \infty$ としたものに等しい．実際2項分布で k 回起こる確率は p_k は

$$\begin{aligned} p_k &= {}_n C_k p^k (1-p)^{n-k} = \frac{n!}{(n-k)!k!} \frac{(np)^k}{n^k} \left(1 - \frac{np}{n}\right)^{n-k} \\ &= \frac{n(n-1)\dots(n-k+1)}{k!n^k} (np)^k \left(1 - \frac{np}{n}\right)^n \left(1 - \frac{np}{n}\right)^{-k}. \end{aligned}$$

ここで $np \rightarrow \lambda$ から

$$\lim_{x \rightarrow \infty} \left(1 - \frac{np}{n}\right)^n = e^{-\lambda}$$

が成立することが，後の命題 1.1 から分かる．これを用いれば

$$\lim_{n \rightarrow \infty} p_k = \frac{\lambda^k}{k!} e^{-\lambda}$$

が容易に得られる．

命題 1.1. 複素数列 $\{\alpha_n\}$ が α に収束しているとする．このとき

$$\lim_{n \rightarrow \infty} \left(1 + \frac{\alpha_n}{n}\right)^n = e^\alpha \quad (1.15)$$

が成り立つ．

証明 複素数 $z_j, w_j, j = 1, \dots, n$ が $|z_j|, |w_j| \leq M$ をみたしているとする．このとき

$$\left| \prod_{j=1}^n z_j - \prod_{j=1}^n w_j \right| \leq M^{n-1} \sum_{j=1}^n |z_j - w_j| \quad (1.16)$$

が成立する．実際このことは

$$\begin{aligned} \left| \prod_{j=1}^n z_j - \prod_{j=1}^n w_j \right| &\leq \left| z_n \prod_{j=1}^{n-1} z_j - z_n \prod_{j=1}^{n-1} w_j \right| + \left| z_n \prod_{j=1}^{n-1} w_j - w_n \prod_{j=1}^{n-1} w_j \right| \\ &\leq M \left| \prod_{j=1}^{n-1} z_j - \prod_{j=1}^{n-1} w_j \right| + M^{n-1} |z_n - w_n| \end{aligned}$$

を用いて帰納法で証明出来る．さらに指数関数の Taylor 展開を用いて

$$\begin{aligned} |e^z - 1 - z| &= \left| \frac{z^2}{2!} + \frac{z^3}{3!} + \dots \right| = \frac{|z|^2}{2} \left| 1 + \frac{2!z}{3!} + \frac{2!z^2}{4!} \dots \right| \\ &\leq \frac{|z|^2}{2} \left(1 + |z| + \frac{|z|^2}{2!} \dots \right) \leq \frac{|z|^2}{2} e^{|z|} \end{aligned}$$

が成り立つことに注意しよう．

さて, (1.15) を示すには $z_j = 1 + \frac{\alpha_n}{n}, w_j = e^{\alpha_n/n}$ とおき, n を十分大きくとって $\alpha_n \leq |\alpha| + 1$ となるようにすれば $1 + \frac{\alpha_n}{n} \leq e^{\alpha_n/n} \leq e^{(|\alpha|+1)/n}$ に注意して (1.16) から

$$\begin{aligned} \left| \left(1 + \frac{\alpha_n}{n} \right)^n - (e^{\alpha_n/n})^n \right| &\leq (e^{(|\alpha|+1)/n})^{n-1} n \left| 1 + \frac{\alpha_n}{n} - e^{\alpha_n/n} \right| \\ &\leq e^{|\alpha|+1} n \frac{|\alpha_n|^2}{n^2} e^{|\alpha_n|/n} \\ &= e^{|\alpha|+1} \frac{|\alpha_n|^2}{n} e^{|\alpha_n|/n} \rightarrow 0. \end{aligned}$$

$(e^{\alpha_n/n})^n \rightarrow e^\alpha$ は明らかであるから求める結果を得る．

□

例 1.1. (丁半賭博)

勝つ確率 p

負ける確率 $1 - p$

勝つと掛け金の $\frac{1}{p}$ 倍もらえる

負けると掛け金を没収される

掛け金 x のとき, 期待値は

$$\frac{x}{p} \cdot p + 0 \cdot (1 - p) = x.$$

掛け金と期待値が等しいから公平なゲームである．

倍賭け法 (マルチンゲール)

勝つまで掛け金を $\frac{1}{1-p}$ 倍していく．

	掛け金	勝ち	負け	この回で終了する確率
①	1	p	$1 - p$	p
②	$\frac{1}{1-p}$	p	$1 - p$	$(1-p)p$
③	$\frac{1}{(1-p)^2}$	p	$1 - p$	$(1-p)^2 p$
④	$\frac{1}{(1-p)^3}$	p	$1 - p$	$(1-p)^3 p$
⋮	⋮	⋮	⋮	⋮

儲けを計算すると

$$\textcircled{1} \quad 1 \cdot \frac{1}{p} - 1$$

$$\textcircled{2} \quad \frac{1}{1-p} \cdot \frac{1}{p} - 1 - \frac{1}{1-p}$$

$$\textcircled{3} \quad \frac{1}{(1-p)^2} \cdot \frac{1}{p} - 1 - \frac{1}{1-p} - \frac{1}{(1-p)^2}$$

$$\textcircled{4} \quad \frac{1}{(1-p)^3} \cdot \frac{1}{p} - 1 - \frac{1}{1-p} - \frac{1}{(1-p)^2} - \frac{1}{(1-p)^3}$$

⋮

Ⓝ

$$\begin{aligned} & \frac{1}{(1-p)^{n-1}} \cdot \frac{1}{p} - 1 - \frac{1}{1-p} - \frac{1}{(1-p)^2} - \cdots - \frac{1}{(1-p)^{n-1}} \\ &= \frac{1}{(1-p)^{n-1}} \cdot \frac{1}{p} - \frac{1 - \frac{1}{(1-p)^n}}{1 - \frac{1}{1-p}} \\ &= \frac{1}{(1-p)^{n-1}} \cdot \frac{1}{p} + \frac{1-p - \frac{1}{(1-p)^{n-1}}}{p} = \frac{1-p}{p} \end{aligned}$$

従って、必ず $\frac{1-p}{p}$ 儲かっているから、必勝法といえる。

必要な投資額の平均：

$$\sum_{n=1}^{\infty} \frac{-1+p + \frac{1}{(1-p)^{n-1}}}{p} \times (1-p)^{n-1} p = \sum_{n=1}^{\infty} \{-(1-p)^n + 1\} = -\frac{1-p}{1-(1-p)} + \sum_{n=1}^{\infty} 1 = \infty.$$

従って、平均的に ∞ の資金を準備しておく必要がある。

2. 連続分布

実確率変数 X の分布関数 $F(x) = P(X \leq x)$ が連続であるとき、 X は連続分布を持つという。さらに連続関数 $p(x)$ が存在して、

$$F(x) = \int_{-\infty}^x p(y) dy$$

が成り立つとき、 $p(x)$ を確率密度関数と呼ぶ。これから

$$P(X \in (a, b]) = \int_a^b p(y) dy$$

が成り立つ。さらに一般に

$$P(X \in B) = \int_B p(y) dy \tag{2.1}$$

が任意の Borel 集合 B に対して成り立つ。従って (2.1) が密度関数の特徴づけであるといえる。

密度関数 p を持つ確率変数 X に関する積分は、次の公式により行う。

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)p(x) dx \quad (2.2)$$

この等式を確かめるには g が階段関数の場合を調べればよい。分割 $x_0 < x_1 < \dots < x_n$ をとり、

$$g(x) = \sum_{j=1}^n a_j 1_{(x_{j-1}, x_j]}(x)$$

とすると、

$$\begin{aligned} E[g(X)] &= E\left[\sum_j a_j 1_{(x_{j-1}, x_j]}(X)\right] \\ &= \sum_j a_j P(X \in (x_{j-1}, x_j]) \\ &= \sum_j a_j \int_{x_{j-1}}^{x_j} p(x) dx \\ &= \int_{-\infty}^{\infty} \sum_j a_j 1_{(x_{j-1}, x_j]}(x) p(x) dx \\ &= \int_{-\infty}^{\infty} g(x) p(x) dx. \end{aligned}$$

一般の場合は階段関数で近似して極限を取ればよい。

以下、応用上よく現れる密度関数の例を挙げる。

一様分布

一様分布 $U(a, b)$, $a < b$ は、区間 $[a, b]$ 上の密度関数 $f(x) = \frac{1}{b-a}$ をもつ分布である。平均は $\frac{a+b}{2}$, 分散は $\frac{(b-a)^2}{12}$ である。

指数分布

指数分布 $E(\lambda)$, $\lambda > 0$ は、半直線 $[0, \infty)$ 上の密度関数 $f(x) = \lambda e^{-\lambda x}$ をもつ分布である。平均は $\frac{1}{\lambda}$, 分散は $\frac{1}{\lambda^2}$ である。

ガンマ分布

ガンマ分布 $Ga(\alpha, \beta)$, $\alpha > 0, \beta > 0$ は半直線 $[0, \infty)$ 上に密度関数

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} \quad (2.3)$$

をもつ分布である。特に $Ga(1, \beta) = E(1/\beta)$ である。平均は $\alpha\beta$, 分散は $\alpha\beta^2$ である。

ベータ分布

ベータ分布 $Be(\alpha, \beta)$, $\alpha > 0, \beta > 0$ は区間 $[0, 1]$ 上で密度関数

$$f(x) = B(\alpha, \beta)^{-1} x^{\alpha-1} (1-x)^{\beta-1} \quad (2.4)$$

をもつ分布である．ここで $B(\alpha, \beta)$ はベータ関数

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx$$

である． $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ が成り立つ．これを用いれば平均 $\frac{\alpha}{\alpha+\beta}$, 分散 $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ が確かめられる．

正規分布

正規分布 $N(m, \sigma^2)$, $m \in \mathbf{R}, \sigma > 0$ は直線 $\mathbf{R} = (-\infty, \infty)$ 上の密度関数

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\} \quad (2.5)$$

をもつ分布である．平均は m , 分散は σ^2 である．さらにこのとき, $Z = (X-m)/\sigma$ は標準正規分布 $N(0, 1)$ に従う．

さて, 平均, 分散を実際に計算してみよう．その前にまず (2.5) の関数が実際に確率分布を定めていることを確かめよう．すなわち全区間での積分が 1 になることである．そのためにもまず次を示す．

命題 2.1. 次が成立する．

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}. \quad (2.6)$$

が成り立つ．

証明 2次元にして, 極座標を用いて計算する．

$$\begin{aligned} \left(\int_{-\infty}^{\infty} e^{-x^2} dx\right)^2 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-x^2-y^2} dx dy \\ & \quad x = r \cos \theta, \quad y = r \sin \theta \\ \frac{\partial(x, y)}{\partial(r, \theta)} &= \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{vmatrix} = r(\cos^2 \theta + \sin^2 \theta) = r \\ dx dy &= \left|\frac{\partial(x, y)}{\partial(r, \theta)}\right| dr d\theta = r dr d\theta \\ &= \int_0^{2\pi} d\theta \int_0^{\infty} e^{-r^2} r dr \end{aligned}$$

$$\begin{aligned}
&= 2\pi \int_0^\infty \frac{d}{dr} \left(-\frac{1}{2} e^{-r^2} \right) dr \\
&= 2\pi \left[-\frac{1}{2} e^{-r^2} \right]_0^\infty = \pi.
\end{aligned}$$

平方根をとれば (2.6) が得られる .

□

注意 2.1. 上の計算から

$$\begin{aligned}
\frac{\sqrt{\pi}}{2} &= \int_0^\infty e^{-x^2} dx \\
y &= x^2, \quad dy = 2x dx \longrightarrow dx = \frac{1}{2\sqrt{y}} dy \\
&= \int_0^\infty e^{-y} \frac{1}{2\sqrt{y}} dy = \frac{1}{2} \int_0^\infty e^{-y} y^{(1/2)-1} dy = \frac{1}{2} \Gamma\left(\frac{1}{2}\right).
\end{aligned}$$

結局ガンマ関数に対して $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ が証明されたことになる .

命題 2.1 を使うと

$$\begin{aligned}
\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-m)^2/2\sigma^2} dx &= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi\sigma^2}} e^{-y^2} \sqrt{2\sigma^2} dy \quad \left(y = \frac{x-m}{\sqrt{2\sigma^2}}, \quad dy = \frac{dx}{\sqrt{2\sigma^2}} \right) \\
&= \frac{1}{\sqrt{\pi}} \int_{-\infty}^\infty e^{-y^2} dy = 1.
\end{aligned}$$

平均は

$$\int_{-\infty}^\infty x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-m)^2/2\sigma^2} dx = \int_{-\infty}^\infty (x+m) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2} dx = m.$$

分散に関しては

$$\begin{aligned}
\int_{-\infty}^\infty (x-m)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-m)^2/2\sigma^2} dx &= \int_{-\infty}^\infty 2\sigma^2 y^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-y^2} \sqrt{2\sigma^2} dy \\
&\quad \left(y = \frac{x-m}{\sqrt{2\sigma^2}}, \quad dy = \frac{dx}{\sqrt{2\sigma^2}} \right) \\
&= \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^\infty y^2 e^{-y^2} dy \\
&= \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^\infty \left(-\frac{y}{2} \right) \frac{d}{dy} e^{-y^2} dy \\
&= \frac{2\sigma^2}{\sqrt{\pi}} \left[-\frac{y}{2} e^{-y^2} \right]_{-\infty}^\infty + \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^\infty \frac{1}{2} e^{-y^2} dy \\
&= \frac{2\sigma^2}{\sqrt{\pi}} \frac{\sqrt{\pi}}{2} = \sigma^2.
\end{aligned}$$

さて，密度関数の概形を調べよう． m は平行移動だけだから， $m = 0$ とし，また定数を無視して

$$f(x) = e^{-x^2/2\sigma^2}$$

を調べる．増減と凹凸を調べるために微分して

$$\begin{aligned} f'(x) &= -\frac{x}{\sigma^2}e^{-x^2/2\sigma^2} \\ f''(x) &= -\frac{1}{\sigma^2}e^{-x^2/2\sigma^2} + \frac{x^2}{\sigma^4}e^{-x^2/2\sigma^2} \\ &= \frac{1}{\sigma^4}(x^2 - \sigma^2)e^{-x^2/2\sigma^2} \\ &= \frac{1}{\sigma^4}(x + \sigma)(x - \sigma)e^{-x^2/2\sigma^2}. \end{aligned}$$

よって変局点が $x = \pm\sigma$ である，釣鐘状の関数であることが分かる．

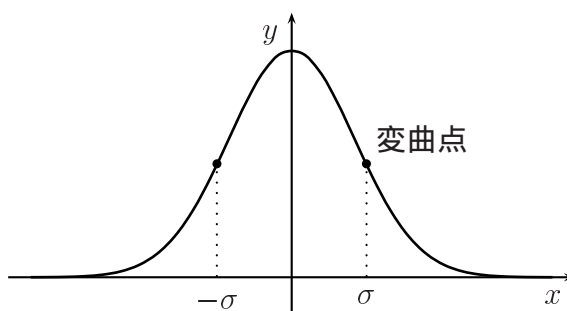


図 2.1: 正規分布 のグラフ

また正規分布は，ガウス (Gauss) 分布，あるいはガウス-ラプラス (Gauss-Laplace) 分布とも呼ばれる．

3. 多次元分布

2次元分布

2次元の確率変数は，実数値確率変数を X, Y を2つ並べた (X, Y) である．それぞれの平均，分散

$$E[X] = m_1, \quad V(X) = \sigma_1^2; \quad E[Y] = m_2, \quad V(Y) = \sigma_2^2$$

のほかに，共分散 $Cov(X, Y)$ を

$$Cov(X, Y) = E[(X - m_1)(Y - m_2)] = E[XY] - m_1m_2 \quad (3.1)$$

で定める $Cov(X, Y)$ を σ_{12} とかくこともある. X, Y の標準化 $Z_1 = \frac{X - m_1}{\sigma_1}$, $Z_2 = \frac{Y - m_2}{\sigma_2}$ の積の平均を相関係数 (correlation coefficient) といい, $Corr(X, Y)$ で表し, 記号 ρ を用いる:

$$\rho = Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{V(X)V(Y)}} = \frac{\sigma_{12}}{\sigma_1\sigma_2}. \quad (3.2)$$

$\rho > 0$ のとき正の相関, $\rho < 0$ のとき負の相関, $\rho = 0$ のときは無相関という.

定理 3.1. 2つの確率変数 X, Y の平均を m_1, m_2 , 分散を σ_1^2, σ_2^2 , 相関係数を ρ とするとき, 次が成り立つ.

(1) $|\rho| \leq 1$.

(2) $\rho = \pm 1$ となる必要十分条件は X, Y に次の線型関係が成り立つことである.

$$\frac{Y - m_2}{\sigma_2} = \pm \frac{X - m_1}{\sigma_1} \quad (\text{複合同順}) \quad (3.3)$$

(3) X, Y が独立のとき 共分散 $Cov(X, Y) = 0$ であり, 従って $\rho = 0$ である.

証明 (1) $Z_1 = \frac{X - m_1}{\sigma_1}$, $Z_2 = \frac{Y - m_2}{\sigma_2}$ とおき $W = Z_1 - tZ_2$ とすると

$$0 \leq E[W^2] = E[Z_1^2 - 2tZ_1Z_2 + Z_2^2] = 1 - 2\rho t + t^2.$$

これがすべての t に対して成り立つから判別式 $D = 4\rho^2 - 4 \leq 0$, すなわち $|\rho| \leq 1$.

(2) $\rho = \pm 1$ のとき, 上の式で $t = \pm 1$ のとき $W = 0$ だから (3.3) が成り立つ.

(3) X, Y が独立ならば,

$$Cov(X, Y) = E[(X - m_1)(Y - m_2)] = E[X - m_1]E[Y - m_2] = 0$$

より $\rho = 0$ となる. □

2つの確率変数 X, Y を考える. Y の値を X の関数 $\hat{Y} = \nu(X)$ によって推定することを Y を X に回帰させるという. このとき

$$E[(Y - \nu(X))^2]$$

を予測誤差 (prediction error) とよぶ. 予測誤差は小さいほど望ましいわけである.

X の線型関数 $\hat{Y} = a + bX$ による推定を線型回帰 (linear regression) という. このとき, 推定誤差を最小にする線型関数を最良線型回帰という.

定理 3.2. 2つの確率変数 X, Y の平均を m_1, m_2 , 分散を σ_1^2, σ_2^2 , 相関係数を ρ とするとき, 最良線型回帰 $\hat{Y} = a + bX$ は

$$a = m_2 - \rho \frac{\sigma_2}{\sigma_1} m_1, \quad b = \rho \frac{\sigma_2}{\sigma_1}$$

で与えられ, このとき最良線型回帰は

$$\frac{\hat{Y} - m_2}{\sigma_2} = \rho \frac{X - m_1}{\sigma_1} \quad (3.4)$$

で与えられる.

証明 予測誤差は

$$\begin{aligned}
 E[(Y-a-bX)^2] &= E[\{(Y-m_2)-b(X-m_1)+(m_2-a-bm_1)\}^2] \\
 &= E[(Y-m_2)^2 - 2b(Y-m_2)(X-m_1) + b^2(X-m_1)^2 + (m_2-a-bm_1)^2] \\
 &= \sigma_2^2 - 2b\rho\sigma_1\sigma_2 + b^2\sigma_1^2 + (m_2-a-bm_1)^2 \\
 &= \sigma_1^2 \left(b - \rho\frac{\sigma_2}{\sigma_1}\right)^2 + \sigma_2^2(1-\rho^2) + (m_2-a-bm_1)^2.
 \end{aligned}$$

これは

$$\begin{aligned}
 b &= \rho\frac{\sigma_2}{\sigma_1} \\
 a &= m_2 - bm_1 = m_2 - \rho\frac{\sigma_2}{\sigma_1}m_1
 \end{aligned}$$

のとき最小になる .

□

注意 3.1. X を Y に回帰させることも可能である . 実際このときの最良線型回帰は

$$\rho\frac{Y-m_2}{\sigma_2} = \frac{\hat{X}-m_1}{\sigma_1}$$

である . 相関係数 ρ が X, Y に関して対称だからこの直線は , (3.4) とは異なっていることに注意しよう .

多次元確率分布

n 次元確率変数 $X = (X_1, \dots, X_n)$ に対し , $m_i = E[X_i]$, $\sigma_{i,j} = Cov(X_i, X_j)$ とするとき ,

$$m = (m_1, \dots, m_n) , \quad V = (\sigma_{ij})$$

をそれぞれ , 平均ベクトル , 共分散行列と呼ぶ .

1次元の場合の (2.1) と同じように , 多次元の場合も

$$P(X \in B) = \int_B p(x) dx \tag{3.5}$$

をみたす関数 $p(x)$ を密度関数と呼ぶ .

n 次元確率分布の中で , もっとも重要なものは次の密度関数をもつ n 次元正規分布である .

$$p(x) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(V)}} \exp\left\{-\frac{1}{2}(x-m, V^{-1}(x-m))\right\} \tag{3.6}$$

ここで $m = (m_1, \dots, m_n) \in \mathbb{R}^n$, $V = (V_{ij})_{i,j=1,\dots,n}$ は正定値対称行列である . この密度を持つ分布を $N(m, V)$ とかく . $m = 0$, $V = E$ (単位行列) のときを n 次標準正規分布と呼ぶ .

$$Z = V^{-1/2}(X-m), \quad \text{すなわち} \quad X = m + V^{1/2}Z$$

とおくと,

$$\begin{aligned}
 E[f(Z)] &= \int_{\mathbb{R}^n} \frac{1}{(2\pi)^{n/2} \sqrt{\det(V)}} f(V^{-1/2}(x-m)) \exp\left\{-\frac{1}{2}(x-m, V^{-1}(x-m))\right\} dx \\
 &\quad z = V^{-1/2}(x-m), \quad x = m + V^{1/2}z \\
 &\quad \frac{\partial(x_1, \dots, x_n)}{\partial(z_1, \dots, z_n)} = \det(V^{1/2}) \\
 &= \int_{\mathbb{R}^n} \frac{1}{(2\pi)^{n/2} \sqrt{\det(V)}} f(z) \exp\left\{-\frac{1}{2}(z, z)\right\} \det(V^{1/2}) dz \\
 &= \int_{\mathbb{R}^n} \frac{1}{(2\pi)^{n/2}} f(z) \exp\left\{-\frac{1}{2}|z|^2\right\} dz
 \end{aligned}$$

であるから Z の分布は $N(0, E)$ に従う. $N(0, E)$ の場合は変数が分離しているから計算が容易で, 平均ベクトル 0 , 共分散行列 E を持つことが容易にわかる. $X = m + V^{1/2}Z$ であるから, X の平均ベクトルは m , 共分散行列が V であることが容易に確かめられる.

問題

1. ガンマ分布の平均と分散を計算せよ.

第3章 極限定理

この章では、確率論の中心的な話題である極限定理を扱う。

1. 大数の法則

確率変数の収束

定義 1.1. 確率変数列 $\{X_n\}$ が 確率変数 X に概収束するとは、

$$P\left(\lim_{n \rightarrow \infty} X_n = X\right) = 1 \quad (1.1)$$

が成り立つときをいう。

また任意の $\varepsilon > 0$ に対し

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0 \quad (1.2)$$

が成り立つとき $\{X_n\}$ は X に確率収束するという。

大数の弱法則

定理 1.2. $\{X_n\}$ を独立同分布の確率変数列で $E[X_1^2] < \infty$ を仮定する。このとき $m = E[X_1]$ として $\frac{X_1 + \cdots + X_n}{n}$ は m に確率概収束する。すなわち、任意の $\varepsilon > 0$ に対し

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{X_1 + \cdots + X_n}{n} - m\right| > \varepsilon\right) = 0.$$

証明 Chebyshev の不等式 (命題 2.9) を使う：

$$P\left(\frac{|X - m|}{\sigma} \geq k\right) \leq \frac{1}{k^2}.$$

m は X の平均、 σ^2 は分散であった。 $X = \frac{X_1 + \cdots + X_n}{n}$ とすれば、 $m = E[X_1]$ 、

$$\sigma^2 = V\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{1}{n^2} \sum_{j=1}^n V(X_j) = \frac{V(X_1)}{n}.$$

ここで任意の $\varepsilon > 0$ に対して $\varepsilon = \sigma k$ となるように k を決める：

$$k = \frac{\varepsilon}{\sigma} = \frac{\varepsilon}{\sqrt{V(X_1)/n}} = \frac{\sqrt{n}\varepsilon}{\sqrt{V(X_1)}}.$$

よって

$$\begin{aligned} P\left(\left|\frac{X_1 + \cdots + X_n}{n} - m\right| \geq \varepsilon\right) &= P\left(\frac{|X_1 + \cdots + X_n - m|}{\sigma} \geq k\right) \\ &\leq \frac{1}{k^2} = \frac{V(X_1)}{n\varepsilon^2} \rightarrow 0 \quad \text{as } n \rightarrow \infty. \end{aligned}$$

これが示すべきことであつた。 □

大数の強法則

定理 1.3. $\{X_n\}$ を独立同分布の確率変数列で $E[X_n^4] < \infty$ を仮定する。このとき $m = E[X_1]$ として $\frac{X_1 + \cdots + X_n}{n}$ は m に概収束する。すなわち

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

証明 X_n の代わりに $\tilde{X}_n = X_n - m$ を考えることにより, $m = 0$ として証明すればよい。

$$Y_n = \frac{X_1 + \cdots + X_n}{n}$$

とにおいて

$$\begin{aligned} E[Y_n^4] &= \frac{1}{n^4} E\left[\left(\sum_{j=1}^n X_j\right)^4\right] \\ &= \frac{1}{n^4} \sum_{i,j,k,l=1}^n E[X_i X_j X_k X_l] \\ &= \frac{1}{n^4} (nE[X_1^4] + 3n(n-1)E[X_1^2]^2) \end{aligned}$$

が成り立つ。実際 j, k, l, m が異なるならば, 独立性から $E[X_j^3 X_k], E[X_j^2 X_k X_l], E[X_i X_j X_k X_l]$ はすべて 0 になる。0 でないのは $E[X_j^4]$ の形のものが n 個, $E[X_j^2 X_k^2]$ の形のものが ${}_n C_2 \cdot {}_4 C_2 = 3n(n-1)$ 個ある。したがって,

$$E\left[\sum_{n=1}^{\infty} Y_n^4\right] = \sum_{n=1}^{\infty} E[Y_n^4] = \sum_{n=1}^{\infty} \frac{1}{n^4} (nE[X_1^4] + 3n(n-1)E[X_1^2]^2) < \infty$$

よって

$$P\left(\sum_{n=1}^{\infty} Y_n^4 < \infty\right) = 1.$$

これから

$$P\left(\lim_{n \rightarrow \infty} Y_n = 0\right) = 1$$

が従う。 □

2. 特性関数

特性関数

定義 2.1. 実確率変数 X に対し

$$\varphi_X(\xi) = E[e^{i\xi X}] \quad (2.1)$$

を X の特性関数 (characteristic function) という.

命題 2.2. 特性関数 φ_X は次の性質を持つ:

(1) φ_X は一様連続で $\varphi_X(0) = 1$.

(2) 任意の n と $\xi_1, \xi_2, \dots, \xi_n \in \mathbb{R}$ と $z_1, z_2, \dots, z_n \in \mathbb{C}$ に対して

$$\sum_{j=1}^n \sum_{k=1}^n \varphi_X(\xi_j - \xi_k) z_j \bar{z}_k \geq 0. \quad (2.2)$$

証明 (1) $\varphi_X(0) = 1$ は明らか. また

$$|\varphi(\xi + h) - \varphi(\xi)| = |E[e^{i\xi X}(e^{ihX} - 1)]| \leq E[|e^{ihX} - 1|]$$

右辺は 0 に収束するから, φ_X は一様連続であることが分かる.

(2) は

$$\sum_{j=1}^n \sum_{k=1}^n \varphi_X(\xi_j - \xi_k) z_j \bar{z}_k = E \left[\left| \sum_{j=1}^n e^{i\xi_j X} z_j \right|^2 \right] \geq 0.$$

□

命題 2.3. X_1, X_2, \dots, X_n が独立ならば

$$\varphi_{X_1 + \dots + X_n}(\xi) = \prod_{j=1}^n \varphi_{X_j}(\xi). \quad (2.3)$$

証明 独立性から

$$\varphi_{X_1 + \dots + X_n}(\xi) = E \left[\prod_{j=1}^n e^{i\xi X_j} \right] = \prod_{j=1}^n E[e^{i\xi X_j}] = \prod_{j=1}^n \varphi_{X_j}(\xi).$$

□

例 2.1. 一様分布 $U(a, b)$ の特性関数は

$$\varphi(\xi) = \int_a^b \frac{1}{b-a} e^{i\xi x} dx = \frac{1}{b-a} \left[\frac{1}{i\xi} e^{i\xi x} \right]_a^b = \frac{1}{(b-a)i\xi} (e^{i\xi b} - e^{i\xi a}).$$

正規分布 $N(m, \sigma^2)$ の特性関数は

$$\varphi(\xi) = e^{im\xi - \frac{\sigma^2}{2}\xi^2}$$

である．これは次を示せば，変数変換で容易に得られる：

$$\int_{-\infty}^{\infty} e^{i\xi x} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = e^{-\xi^2/2}$$

そこで左辺を $F(\xi)$ とおき， ξ で微分すると

$$\begin{aligned} F'(\xi) &= \int_{-\infty}^{\infty} ixe^{i\xi x} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} ie^{i\xi x} \frac{d}{dx} \left\{ -e^{-x^2/2} \right\} dx \\ &= \frac{1}{\sqrt{2\pi}} \left[-ie^{i\xi x} e^{-x^2/2} \right]_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} ie^{i\xi x} i\xi e^{-x^2/2} dx \\ &= -\frac{1}{\sqrt{2\pi}} \xi \int_{-\infty}^{\infty} e^{i\xi x} e^{-x^2/2} dx \\ &= -\xi F(\xi). \end{aligned}$$

すなわち $\frac{d}{d\xi} \log |F(\xi)| = -\xi$ であるから

$$\log |F(\xi)| = -\frac{1}{2}\xi^2 + C.$$

よって

$$F(\xi) = \pm e^C \exp\left\{-\frac{1}{2}\xi^2\right\}.$$

$F(0) = 1$ より，最終的に $F(\xi) = e^{-\xi^2/2}$ が得られる．

X の分布が $N(m_1, \sigma_1^2)$ Y の分布が $N(m_2, \sigma_2^2)$ で X, Y が独立のとき $X + Y$ の特性関数は

$$\varphi_{X+Y}(\xi) = e^{im_1\xi - \frac{\sigma_1^2}{2}\xi^2} e^{im_2\xi - \frac{\sigma_2^2}{2}\xi^2} = e^{i(m_1+m_2)\xi - \frac{\sigma_1^2 + \sigma_2^2}{2}\xi^2}$$

であるから，分布は $N(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$ で再び正規分布である．このように独立な和が再び同じ分布族に属するとき，再生性をもつという．

テント関数

次の関数 $T, T_{a,b,c}$ をテント関数いう：

$$T(x) = \begin{cases} x + 1, & -1 \leq x \leq 0 \\ -x + 1, & 0 \leq x \leq 1 \\ 0, & \text{elsewhere.} \end{cases}$$

$a < b, c > 0$ に対して

$$T_{a,b,c}(x) = cT\left(\frac{2}{b-a}\left(x - \frac{a+b}{2}\right)\right)$$

とおく .

また , $c > 1$ のテント関数を高さ 1 で切り取って , 台形関数 $D_{a,b,c}$ を

$$D_{a,b,c}(x) = T_{a,b,c}(x) \wedge 1 = T_{a,b,c}(x) - T_{a+(b-a)/2c, b-(b-a)/2c, c-1}(x)$$

と定める .

例 2.2. 確率変数 U, V はともに一様分布 $U(0, a)$ に従うとき , $U - V$ の分布を計算しよう . 分布関数 $F(t) = P(U - V \leq t)$ は $U - V$ の分布の対称性から $t \geq 0$ のとき計算すればよい . まず $t \geq a$ のときは明らかに $F(t) = 1$. まず $0 \leq t \leq a$ のとき

$$\begin{aligned} F(t) &= \frac{1}{a^2} \int_0^a dv \int_0^a 1_{(-\infty, t]}(u-v) du \\ &= \frac{1}{a^2} \int_0^a dv \int_0^a 1_{(-\infty, t+v]}(u) du \\ &= \frac{1}{a^2} \int_0^a (t+v) \wedge a dv \\ &= \frac{1}{a^2} \int_0^{a-t} (t+v) dv + \frac{1}{a^2} \int_{a-t}^a a dv \\ &= \frac{1}{a^2} \left[\frac{1}{2}(t+v)^2 \right]_0^{a-t} + \frac{t}{a} \\ &= \frac{1}{2a^2}(a^2 - t^2) + \frac{t}{a}. \end{aligned}$$

よってこのとき $F'(t) = \frac{1}{a^2}(a-t)$. 結局密度関数が $T_{-a, a, 1/a}$ であることが分かる . $U - V$ の特性関数 φ は

$$\begin{aligned} \varphi(\xi) &= \int_0^a du \int_0^a dv \frac{1}{a^2} e^{i\xi(u-v)} = \frac{e^{ia\xi} - 1}{ia\xi} \cdot \frac{e^{-ia\xi} - 1}{-ia\xi} \\ &= \frac{1}{a^2\xi^2}(2 - e^{ia\xi} - e^{-ia\xi}) = \frac{2}{a^2\xi^2}(1 - \cos a\xi) \end{aligned}$$

である .

命題 2.4. 確率変数 X の特性関数を φ_X とするとき ,

$$E[T(X)] = \frac{1}{\pi} \int_{-\infty}^{\infty} \varphi_X(\xi) \frac{1 - \cos \xi}{\xi^2} d\xi \quad (2.4)$$

$$E[T_{a,b,c}(X)] = \frac{2c}{\pi(b-a)} \int_{-\infty}^{\infty} \varphi_X(\xi) \exp\left\{-i\frac{(a+b)\xi}{2}\right\} \frac{1 - \cos \frac{(b-a)\xi}{2}}{\xi^2} d\xi \quad (2.5)$$

証明 例 2.2 から

$$\int_{-\infty}^{\infty} e^{i\xi x} T(x) dx = \frac{2(1 - \cos \xi)}{\xi^2}. \quad (2.6)$$

また

$$\int_{-\infty}^{\infty} e^{i\xi x} \frac{1 - \cos \xi}{\xi^2} d\xi = \pi T(x) \quad (2.7)$$

が成り立つ．これを見るには，まず $\frac{1 - \cos \xi}{\xi^2}$ が偶関数であるから上の積分は

$$\int_{-\infty}^{\infty} \cos \xi x \frac{1 - \cos \xi}{\xi^2} d\xi$$

に等しく，また三角公式から

$$\cos \xi x (1 - \cos \xi) = \cos \xi x - \frac{1}{2} \{ \cos \xi (x+1) + \cos \xi (x-1) \}$$

であり，さらに変数変換を用いて

$$\begin{aligned} (2.7) \text{ の左辺} &= \int_{-\infty}^{\infty} \cos \xi x \frac{1 - \cos \xi}{\xi^2} d\xi \\ &= \frac{1}{2} \int_{-\infty}^{\infty} \frac{1 - \cos \xi (x+1)}{\xi^2} d\xi + \frac{1}{2} \int_{-\infty}^{\infty} \frac{1 - \cos \xi (x-1)}{\xi^2} d\xi - \int_{-\infty}^{\infty} \frac{1 - \cos \xi x}{\xi^2} d\xi \\ &= \int_{-\infty}^{\infty} \frac{1 - \cos \xi}{\xi^2} d\xi \left(\frac{1}{2} |x+1| + \frac{1}{2} |x-1| - |x| \right). \end{aligned}$$

ここで

$$\int_{-\infty}^{\infty} \frac{1 - \cos \xi}{\xi^2} d\xi = \left[-(1 - \cos \xi) \frac{1}{\xi} \right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \frac{\sin \xi}{\xi} d\xi = \pi$$

に注意して，(2.7) が得られる．

(2.7) を用いて積分順序を交換すると

$$\begin{aligned} E[T(X)] &= \frac{1}{\pi} E \left[\int_{-\infty}^{\infty} e^{i\xi X} \frac{1 - \cos \xi}{\xi^2} d\xi \right] \\ &= \frac{1}{\pi} \int_{-\infty}^{\infty} E[e^{i\xi X}] \frac{1 - \cos \xi}{\xi^2} d\xi \\ &= \frac{1}{\pi} \int_{-\infty}^{\infty} \varphi_X(\xi) \frac{1 - \cos \xi}{\xi^2} d\xi \end{aligned}$$

となり，(2.4) が分かる．

さらに

$$\varphi_{\frac{2}{b-a}(X - \frac{a+b}{2})}(\xi) = \varphi_X \left(\frac{2\xi}{b-a} \right) e^{-i\frac{a+b}{b-a}\xi}$$

に注意すれば

$$E[T_{a,b,c}(X)] = \frac{c}{\pi} \int_{-\infty}^{\infty} \varphi_X \left(\frac{2\xi}{b-a} \right) \exp \left\{ -i\frac{a+b}{b-a}\xi \right\} \frac{1 - \cos \xi}{\xi^2} d\xi.$$

ここで変数変換を用いれば (2.5) が示される．

□

定理 2.5. 確率変数 X, Y の特性関数を φ_X, φ_Y とするとき,

$$\varphi_X(\xi) = \varphi_Y(\xi) \quad \forall \xi \in \mathbb{R}$$

が成り立つとき, X, Y の分布は一致する.

証明 命題 2.4 から, 任意のテント関数 $T_{a,b,c}$ に対して

$$E[T_{a,b,c}(X)] = E[T_{a,b,c}(Y)]$$

が成り立つから, 任意の台形関数 $D_{a,b,c}$ に対しても

$$E[D_{a,b,c}(X)] = E[D_{a,b,c}(Y)]. \quad (2.8)$$

ここで

$$\lim_{c \rightarrow \infty} D_{a,b,c}(x) = 1_{(a,b)}(x)$$

であるから, (2.8) で $c \rightarrow \infty$ として

$$P(a < X < b) = P(a < Y < b)$$

が成り立つ. a, b は任意だから二つの分布は一致する. □

定理 2.6. 確率変数 X と確率変数列 X_n の特性関数を φ_X, φ_{X_n} とする. このとき

$$\lim_{n \rightarrow \infty} \varphi_{X_n}(\xi) = \varphi_X(\xi) \quad \forall \xi \in \mathbb{R}$$

を仮定する. このとき, $P(X = a) = P(X = b) = 0$ を満たす $a < b$ に対して

$$\lim_{n \rightarrow \infty} P(a \leq X_n \leq b) = P(a \leq X \leq b) \quad (\text{等号がない場合も含める}) \quad (2.9)$$

が成り立つ.

証明 定理の仮定から

$$\begin{aligned} \lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} \varphi_{X_n}(\xi) \exp\left\{-i \frac{(a+b)\xi}{2}\right\} \frac{1 - \cos \frac{(b-a)\xi}{2}}{\xi^2} d\xi \\ = \int_{-\infty}^{\infty} \varphi_X(\xi) \exp\left\{-i \frac{(a+b)\xi}{2}\right\} \frac{1 - \cos \frac{(b-a)\xi}{2}}{\xi^2} d\xi. \end{aligned}$$

よって 命題 2.4 から任意のテント関数

$$\lim_{n \rightarrow \infty} E[T_{a,b,c}(X_n)] = E[T_{a,b,c}(Y)]$$

が成り立つから, 任意の台形関数 $D_{a,b,c}$ に対しても

$$\lim_{n \rightarrow \infty} E[D_{a,b,c}(X_n)] = E[D_{a,b,c}(X)]$$

が成り立つ．さらに $a < b, c > 1$ に対して

$$1_{[a+(b-a)/2c, b-(b-a)/2c]}(x) \leq D_{a,b,c} \leq 1_{(a,b)}$$

に注意して

$$\begin{aligned} \liminf_{n \rightarrow \infty} P(a < X_n < b) &\geq \lim_{n \rightarrow \infty} E[D_{a,b,c}(X_n)] \\ &= E[D_{a,b,c}(X)] \\ &\geq P\left(a + \frac{b-a}{2c} \leq X \leq b - \frac{b-a}{2c}\right). \end{aligned}$$

ここで $c \rightarrow \infty$ とすると

$$\liminf_{n \rightarrow \infty} P(a < X_n < b) \geq P(a < X < b). \quad (2.10)$$

さらに $b \rightarrow \infty$ とすれば

$$\liminf_{n \rightarrow \infty} P(X_n > a) \geq P(X > a). \quad (2.11)$$

また (2.10) で $a \rightarrow -\infty$ とし, b を a に取り直せば

$$\liminf_{n \rightarrow \infty} P(X_n < a) \geq P(X < a).$$

これから

$$\overline{\lim}_{n \rightarrow \infty} P(X_n \geq a) = 1 - \liminf_{n \rightarrow \infty} P(X_n < a) \leq 1 - P(X < a) = P(X \geq a).$$

以上で (2.11) と合わせて

$$P(X > a) \leq \liminf_{n \rightarrow \infty} P(X_n > a) \leq \overline{\lim}_{n \rightarrow \infty} P(X_n \geq a) \leq P(X \geq a).$$

ここで $P(X = a) = 0$ を仮定すれば, 両端が一致するから

$$\lim_{n \rightarrow \infty} P(X_n > a) = \lim_{n \rightarrow \infty} P(X_n \geq a) = P(X \geq a)$$

が成り立つ．これから (2.9) が得られることは明らかだろう． □

そのほかの同値条件を次にまとめておく．

定理 2.7. 確率変数 X と確率変数列 X_n の特性関数を φ_X, φ_{X_n} とする．このとき次は同値になる．

(1) $P(X = a) = P(X = b) = 0$ を満たす $a < b$ に対して

$$\lim_{n \rightarrow \infty} P(a \leq X_n \leq b) = P(a \leq X \leq b) \quad (\text{等号はなくても成り立つ}) \quad (2.12)$$

が成り立つ．

(2) 台が有界な任意の連続関数 f に対し

$$\lim_{n \rightarrow \infty} E[f(X_n)] = E[f(X)]. \quad (2.13)$$

(3) 任意の有界な連続関数 f に対し

$$\lim_{n \rightarrow \infty} E[f(X_n)] = E[f(X)]. \quad (2.14)$$

(4) 特性関数に対し

$$\lim_{n \rightarrow \infty} \varphi_{X_n}(\xi) = \varphi_X(\xi) \quad \forall \xi \in \mathbb{R} \quad (2.15)$$

証明 (1) \Rightarrow (2)

f を台が有界な任意の連続関数とする. $P(X = t)$ となる t は高々可算だから f の一様連続性から任意の ε に対し次のようにとる.

(i) $f(x) = 0, x \leq a$ または $x \geq b$.

(ii) $P(a \leq X \leq b) \geq 1 - \varepsilon$.

(iii) $a = t_0 \leq t_1 < \dots < t_n = b$ で $\sup\{f(x); t_{j-1} \leq x \leq t_j\} - \inf\{f(x); t_{j-1} \leq x \leq t_j\} < \varepsilon$

(iv) $P(X = t_j) = 0$

このとき

$$\varphi(x) = \sum_{j=1}^N 1_{[t_{j-1}, t_j)}(x) f(t_{j-1})$$

とおくと

$$|\varphi(x) - f(x)| \leq \varepsilon$$

が成り立っている. よって

$$\begin{aligned} |E[f(X_n)] - E[f(X)]| &\leq E[|f(X_n) - \varphi(X_n)|] + E[|\varphi(X_n) - \varphi(X)|] + E[|\varphi(X) - f(X)|] \\ &\leq \left| \sum_{j=1}^N f(t_{j-1}) P(t_{j-1} \leq X_n < t_j) - \sum_{j=1}^N f(t_{j-1}) P(t_{j-1} \leq X < t_j) \right| + 2\varepsilon \\ &\rightarrow 2\varepsilon \quad (\text{as } n \rightarrow \infty.) \end{aligned}$$

$\varepsilon > 0$ は任意だから $\lim_{n \rightarrow \infty} E[f(X_n)] = E[f(X)]$ が成り立つ.

(2) \Rightarrow (3)

f を任意の有界関数とする．任意の $\varepsilon > 0$ に対し K を大きくとれば $P(|X| \geq K) \leq \varepsilon$ とできる．このとき連続関数 φ を

$$\varphi(x) \begin{cases} = 1, & |x| \leq K, \\ \in [0, 1], & K \leq |x| \leq K+1, \\ = 0, & |x| \geq K+1 \end{cases}$$

とおくと, $1 - E[\varphi(X)] \leq \varepsilon$ が成り立つ．さらに

$$\begin{aligned} & |E[f(X_n)] - E[f(X)]| \\ & \leq E[|f(X_n) - \varphi(X_n)f(X_n)|] + |E[f(X_n)\varphi(X_n) - \varphi(X)f(X)]| + E[|\varphi(X)f(X) - f(X)|] \\ & \leq E[(1 - \varphi(X_n))\|f\|_\infty] + |E[f(X_n)\varphi(X_n) - \varphi(X)f(X)]| + E[(1 - \varphi(X))\|f\|_\infty] \\ & \leq (1 - E[\varphi(X_n)])\|f\|_\infty + |E[f(X_n)\varphi(X_n) - \varphi(X)f(X)]| + (1 - E[\varphi(X)])\|f\|_\infty \\ & \rightarrow (1 - E[\varphi(X)])\|f\|_\infty \quad (\text{as } n \rightarrow \infty) \\ & \leq 2\varepsilon\|f\|_\infty. \end{aligned}$$

$\varepsilon > 0$ は任意だから $\lim_{n \rightarrow \infty} E[f(X_n)] = E[f(X)]$ が成り立つ．

(3) \Rightarrow (4)

これは $e^{i\xi x}$ が有界連続関数であることに注意すればよい．

(4) \Rightarrow (1)

これは 定理 2.6 で既に示した． □

定義 2.8. 確率変数列 X_n が, 確率変数 X に法則収束することを, 任意の有界連続関数 f に対し

$$\lim_{n \rightarrow \infty} E[f(X_n)] = E[f(X)] \quad (2.16)$$

が成立することと定義する．

3. 中心極限定理

中心極限定理

中心極限定理を証明するために次の補題を準備する．

補題 3.1. $E[X] = 0, E[X^2] = 1$ をみたす確率変数 X に対して φ をその特性関数とするととき

$$\varphi\left(\frac{\xi}{\sqrt{n}}\right) = 1 - \frac{\xi^2}{2n} + o\left(\frac{1}{n}\right). \quad (n \rightarrow \infty) \quad (3.1)$$

証明

$$e^{i\xi} = 1 + i\xi - \frac{\xi^2}{2} + \xi^2 g(\xi)$$

となるように $g(\xi)$ を定めると,

$$|g(\xi)| \leq 1, \quad \lim_{\xi \rightarrow 0} g(\xi) = 0 \quad (3.2)$$

である. このことをまず示しておこう. Taylor の公式 (積分形) から

$$e^{ix} = 1 + ix - \frac{1}{2}x^2 - i \int_0^x \frac{(x-t)^2}{2} e^{it} dt$$

が成り立つ (x が負のときは $\int_0^x = -\int_x^0$ とする.) 剰余項を

$$R(x) = i \int_0^x \frac{(x-t)^2}{2} e^{it} dt \quad (3.3)$$

とにおいて評価しよう. 絶対値を取れば容易に

$$|R(x)| \leq \int_0^{|x|} \frac{(|x|-t)^2}{2} dt = \frac{1}{3!}|x|^3 \quad (3.4)$$

が得られる. 他方 (3.3) で部分積分を用いて

$$\begin{aligned} R(x) &= \left[\frac{(x-t)^2}{2} e^{it} \right]_0^x + \int_0^x (x-t) e^{it} dt \\ &= -\frac{1}{2}x^2 + \int_0^x (x-t) e^{it} dt \\ &= -\int_0^x (x-t) dt + \int_0^x (x-t) e^{it} dt \\ &= \int_0^x (x-t)(e^{it} - 1) dt. \end{aligned}$$

これから

$$|R(x)| \leq 2 \int_0^{|x|} (|x|-t) dt = |x|^2.$$

以上で $|R(x)| \leq \min\{|x|^2, |x|^3/3!\}$ が示せたので (3.2) が確かめられたことになる.

そこで

$$\exp\left\{\frac{i\xi X}{\sqrt{n}}\right\} = 1 + i\xi X - \frac{\xi^2 X^2}{2n} + \frac{\xi^2 X^2}{n} g\left(\frac{\xi X}{\sqrt{n}}\right)$$

において期待値を取れば

$$\varphi_X\left(\frac{\xi}{\sqrt{n}}\right) = 1 - \frac{\xi^2}{2n} + E\left[\frac{\xi^2 X^2}{n} g\left(\frac{\xi X}{\sqrt{n}}\right)\right].$$

(3.1) を示すためには

$$\lim_{n \rightarrow \infty} E\left[X^2 g\left(\frac{\xi X}{\sqrt{n}}\right)\right] = 0$$

を示せばよいが, これは g の性質から明らか.

□

定理 3.2. (中心極限定理) $\{X_n\}$ を独立同分布を持つ確率変数とし, 有限な期待値 $m = E[X_n]$ と分散 $\sigma^2 = V(X_n)$ を持つとする. このとき任意の $a < b$ に対して

$$\lim_{n \rightarrow \infty} P\left(a \leq \frac{\sum_{j=1}^n (X_j - m)}{\sigma\sqrt{n}} \leq b\right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

証明 $\tilde{X}_n = \frac{X_n - m}{\sigma}$ を考えれば, $m = 0, \sigma^2 = 1$ の場合に証明すればよい. また定理 2.6 から

$$Y_n = \frac{X_1 + \cdots + X_n}{\sqrt{n}}$$

の特性関数が $N(0, 1)$ の特性関数 $e^{-\xi^2/2}$ に収束することを示せばよい.

$$\varphi_{Y_n}(\xi) = E\left[\exp\left(\frac{i\xi}{\sqrt{n}} \sum_{j=1}^n X_j\right)\right] = \prod_{j=1}^n \varphi_{X_j}\left(\frac{\xi}{\sqrt{n}}\right) = \varphi_{X_1}\left(\frac{\xi}{\sqrt{n}}\right)^n.$$

ここで補題 3.1 を使えば

$$\lim_{n \rightarrow \infty} \varphi_{Y_n}(\xi) = \lim_{n \rightarrow \infty} \left(1 - \frac{\xi^2}{2n} + o\left(\frac{1}{n}\right)\right)^n = e^{-\xi^2/2}.$$

最後の等式は第 2 章, 命題 1.1 を使った. これで証明が終わる. □

Notes

- 第 3 節の中心極限定理の話は, 志賀 [3] § 5.5, 5.6 によった.

問題

1. Poisson 分布の特性関数を計算せよ. このことを用いて, 独立な確率変数 X, Y がともに Poisson 分布に従うとき, $X + Y$ も Poisson 分布に従うことを示せ.

第4章 ランダム・ウォーク

この章では、最も簡単な確率過程としてランダム・ウォークを扱う。

1. 単純ランダム・ウォーク

単純ランダム・ウォーク

定義 1.1. 時間 $t \in \mathbb{T}$ をパラメーターとして持つ確率変数の族 (X_t) を確率過程という。 \mathbb{T} として $[0, \infty)$, $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$ などがよく使われる。 $[0, \infty)$ のとき連続時間, \mathbb{Z}_+ のとき離散時間という。

以下では \mathbb{Z}_+ の場合のみを扱う。この場合は t の代わりに n を用いる。

定義 1.2. X_1, X_2, \dots を i.i.d. で各分布は

$$P(X_i = 1) = p, \quad P(X_i = -1) = q = 1 - p \quad (1.1)$$

で与えられているベルヌーイ列とする。1 が成功, -1 が失敗を表し, 成功の確率が p である。このとき

$$S_n = X_1 + X_2 + \dots + X_n \quad (1.2)$$

で定まる確率過程 (S_n) を単純ランダム・ウォークという。 $p = q = \frac{1}{2}$ のとき, 単純ランダム・ウォークは対称であるという。

再帰性, 非再帰性

定義 1.3. 単純ランダム・ウォーク S_n に対し

$$P(S_n = 0 \text{ となる } n \in \mathbb{Z}_+ \text{ が無限個存在する}) = 1 \quad (1.3)$$

が成り立つとき再帰的, 成り立たないとき非再帰的と呼ぶ。また確率 $P(S_n = 0 \text{ となる } n \in \mathbb{Z}_+ \text{ が存在する})$ を再帰確率と呼ぶ。

(1.3) で無限個存在するという条件が現れたが, これを i.o. (infinitely often) と略記する。従って $S_n = 0, \text{ i.o.}$ とかく。

再帰性の問題を考えるために, 確率変数の列 $\{\tau_n\}_{n=0}^{\infty}$ を帰納的に次のように定義する:

$$\tau_0 = 0, \quad \tau_n = \inf\{m > \tau_{n-1} : S_m = 0\}.$$

一度原点 0 に戻ると、過去とは独立に新たなランダム・ウォークが始まる。従って

$$P(\tau_n < \infty) = P(\tau_1 < \infty)^n \quad (1.4)$$

であることが容易にわかる。再帰確率は定義から $P(\tau_1 < \infty)$ に等しい。

定理 1.4. 単純ランダム・ウォークに対して、次は同値である

- (1) $P(\tau_1 < \infty) = 1$.
- (2) $P(S_n = 0, \text{i.o.}) = 1$.
- (3) $\sum_{n=1}^{\infty} P(S_n = 0) = \infty$.

証明 $P(\tau_1 < \infty) = 1$ ならば (1.4) から任意の n に対して $P(\tau_n < \infty)$ となり、従って $P(S_n = 0, \text{i.o.}) = 1$ が成り立つ。これは (1) \Rightarrow (2) を示している。

次に

$$V = \sum_{n=0}^{\infty} 1_{\{S_n=0\}} = \sum_{n=0}^{\infty} 1_{\{\tau_n < \infty\}}$$

と置けば、 V は時刻 0 のときも含めた 0 を訪問する回数を表す。よって

$$E[V] = \sum_{n=0}^{\infty} P(S_n = 0) = \sum_{n=0}^{\infty} P(\tau_n < \infty) = \sum_{n=0}^{\infty} P(\tau_1 < \infty)^n = \frac{1}{1 - P(\tau_1 < \infty)}. \quad (1.5)$$

(2) が成り立てば $V = \infty$ a.s. で $\infty = E[V] = \sum_{n=0}^{\infty} P(S_n = 0)$ から (2) \Rightarrow (3) がわかる。

また $\sum_{n=0}^{\infty} P(S_n = 0) = \frac{1}{1 - P(\tau_1 < \infty)}$ から $P(\tau_1 < \infty) < 1$ ならば $\sum_{n=0}^{\infty} P(S_n = 0) < \infty$ が成立する。つまり (1) が成り立たなければ (3) が成り立たないことがわかる。これは (3) \Rightarrow (1) を意味している。以上ですべての同値性が示せた。□

上の定理から非再帰的であれば、原点に無限回帰する確率は 1 より小さいことになる。 $E[X_n] = 2p - 1$ だから大数の強法則より、

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = 2p - 1$$

が成立する。よって $p \neq \frac{1}{2}$ であれば $S_n \rightarrow \infty$ または $S_n \rightarrow -\infty$ となり、原点に無限回帰する確率は 0 である。従ってこのときは非再帰的であることが容易にわかる。

次に $p = \frac{1}{2}$ のとき再帰的であることを示そう。定理 1.4 の (3) を確かめればよい。確率 $P(S_n = 0)$ は単純ランダム・ウォークの性質から n が奇数のときは 0 で $n = 2m$ のときは

$$P(S_{2m} = 0) = \binom{2m}{m} \frac{1}{2^{2m}}$$

となる。右辺を計算するために準備が必要となる。

ウォリス (Wallis) の公式

補題 1.5. (ウォリスの公式) 次の等式が成り立つ:

$$\lim_{n \rightarrow \infty} \frac{2^{2n}(n!)^2}{\sqrt{n}(2n)!} = \sqrt{\pi} \quad (1.6)$$

証明 定積分

$$I_n = \int_0^{\pi/2} \sin^n x \, dx$$

は

$$I_n = \frac{n-1}{n} I_{n-2}$$

の関係から

$$I_{2n} = \frac{2n-1}{2n} \frac{2n-3}{2n-2} \cdots \frac{1}{2} \frac{\pi}{2}, \quad (1.7)$$

$$I_{2n+1} = \frac{2n}{2n+1} \frac{2n-2}{2n-1} \cdots \frac{2}{3}. \quad (1.8)$$

したがって

$$\begin{aligned} I_{2n} I_{2n+1} &= \frac{\pi}{4n+2}, \\ I_{2n+1} \sqrt{\frac{I_{2n}}{I_{2n+1}}} &= \sqrt{\frac{\pi}{4n+2}}. \end{aligned} \quad (1.9)$$

一方 $0 < x < \frac{\pi}{2}$ では $0 < \sin x < 1$ だから

$$0 < I_{2n+1} < I_{2n} < I_{2n-1}, \quad 1 < \frac{I_{2n}}{I_{2n+1}} < \frac{I_{2n-1}}{I_{2n+1}} = \frac{2n+1}{2n}.$$

よって

$$\lim_{n \rightarrow \infty} \frac{I_{2n}}{I_{2n+1}} = 1.$$

これと (1.9) から

$$\frac{\sqrt{\pi}}{2} = \lim_{n \rightarrow \infty} \sqrt{n} I_{2n+1}.$$

また (1.8) から

$$I_{2n+1} = \frac{2^{2n}(n!)^2}{(2n+1)!}.$$

上の二つの式から (1.6) が得られる。□

以上の準備の下で次の定理を得る。

定理 1.6. 単純ランダム・ウォークは再帰的である.

証明 $P(S_{2n} = 0) = \binom{2n}{n} \frac{1}{2^{2n}}$ であったが, ウォリスの公式から

$$P(S_{2n} = 0) = \binom{2n}{n} \frac{1}{2^{2n}} \sim \frac{1}{\sqrt{n\pi}}$$

が従う. したがって

$$\sum_{n=0}^{\infty} P(S_{2n} = 0) = \infty$$

となり, 定理 1.4 から単純ランダム・ウォークが再帰的であることが分かる. \square

再帰確率

$p \neq \frac{1}{2}$ のときは非再帰的であることを見たが, 再帰確率を次に計算しておこう.

定理 1.7. 非対称単純ランダム・ウォークの再帰確率は

$$P(\tau_1 < \infty) = \begin{cases} 2(1-p), & p > \frac{1}{2} \\ 2p, & p < \frac{1}{2} \end{cases} \quad (1.10)$$

証明 まず

$$P(S_n = 0) = \binom{2n}{n} p^n q^n = \binom{2n}{n} 2^{-2n} (4pq)^n.$$

ここで, 一般の2項係数 $\binom{-1/2}{n}$ を用いる.

$$\begin{aligned} \binom{-1/2}{n} &= \frac{(-1/2)(-1/2-1)\dots(-1/2-n+1)}{n!} \\ &= \frac{1 \cdot 3 \cdots (2n-1)}{2^n (-1)^n n!} \\ &= \frac{(2n)!}{2 \cdot 4 \cdots (2n) \cdot 2^n (-1)^n n!} \\ &= \frac{(2n)!}{2^{2n} (-1)^n (n!)^2} \\ &= \binom{2n}{n} 2^{-2n} (-1)^n. \end{aligned}$$

よって

$$P(S_n = 0) = \binom{-1/2}{n} (-4pq)^n.$$

一方マクローリン展開

$$(1+x)^{-1/2} = \sum_{n=0}^{\infty} \binom{-1/2}{n} x^n$$

を用いれば

$$\sum_{n=0}^{\infty} P(S_n = 0) = \sum_{n=0}^{\infty} \binom{-1/2}{n} (-4pq)^n = (1 - 4pq)^{-1/2} = ((p+q)^2 - 4pq)^{-1/2} = \frac{1}{|p-q|}.$$

さらに (1.5) を用いれば

$$1 - P(\tau_1 < \infty) = |p - q|$$

が得られる．後は場合分けをして容易に (1.10) に到達する．

□

2次元ランダム・ウォーク

今までは1次元で考えてきたが，2次元でも同様な問題を考えることができる． (X_n) を \mathbb{Z}^2 に値をとる i.i.d の列とし，分布は

$$P(X_n = (1, 0)) = P(X_n = (-1, 0)) = P(X_n = (0, 1)) = P(X_n = (0, -1)) = \frac{1}{4}$$

で与えられるとし，1次元と同様に

$$S_n = X_1 + X_2 + \cdots + X_n$$

と定める．これは上下，右左にそれぞれ等確率の $\frac{1}{4}$ で動いていく運動を表している．これを2次元対称単純ランダム・ウォークという．

定理 1.8. 2次元対称単純ランダム・ウォークは再帰的である．

証明 2次元の場合でも定理 1.4 は同様に成り立つ．したがって $\sum_{n=0}^{\infty} P(S_{2n} = 0) = \infty$ を示せばよい． $S_{2n} = 0$ という事象は，上に k 回，下に k 回，右に $n - k$ 回，左に $n - k$ 回移動すればよい．ここで $k = 0, \dots, n$ である．よって

$$P(S_{2n} = 0) = \frac{1}{4^{2n}} \sum_{k=0}^n \frac{(2n)!}{(k!)^2 \{(n-k)!\}^2}. \quad (1.11)$$

さてここで自然数 a, b に対して

$$\sum_{k=0}^n \binom{a}{k} \cdot \binom{b}{n-k} = \binom{a+b}{n}$$

成り立つことに留意しよう．これは赤球 a 個，青球 b 個から n 個取り出す組み合わせの数を考えればよい．右辺がその総数であるが，赤球から k 個，青球から $n - k$ 個取り出すと考えると左辺の個数になる．

特に $a = b = n$ とすれば

$$\sum_{k=0}^n \binom{n}{k}^2 = \binom{2n}{n}.$$

さらに両辺に $\binom{2n}{n}$ をかければ

$$\binom{2n}{n} \binom{n}{k}^2 = \frac{(2n)!}{(n!)^2} \frac{(n!)^2}{(k!)^2 \{(n-k)!\}^2} = \frac{(2n)!}{(k!)^2 \{(n-k)!\}^2}$$

であるから

$$\sum_{k=0}^n \frac{(2n)!}{(k!)^2 \{(n-k)!\}^2} = \binom{2n}{n}^2$$

が得られる．ここで (1.11) にもどれば，

$$P(S_{2n} = 0) = \frac{1}{4^{2n}} \binom{2n}{n}^2.$$

再びウォリスの公式 (1.5) を使えば $\binom{2n}{n} \sim \frac{4^n}{\sqrt{\pi n}}$ だから

$$P(S_{2n} = 0) \sim \frac{1}{4^{2n}} \frac{4^{2n}}{\pi n} = \frac{1}{\pi n}$$

となり $\sum_{n=0}^{\infty} P(S_{2n} = 0) = \infty$ が示せた．以上でこの場合も再帰性が示せた． \square

3次元ランダム・ウォーク

3次元の場合を扱うには，ウォリスの公式よりも一般のスターリング (Stirling) の公式が必要になる．まずそれから述べよう．

定理 1.9. (スターリングの公式)

$$n! \sim \sqrt{2\pi n} n^n e^{-n} \quad (1.12)$$

証明 \log のテーラー展開から $|x| < 1$ に対し

$$\begin{aligned} \log(1+x) &= x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + \cdots, \\ \log(1-x) &= -x - \frac{1}{2}x^2 - \frac{1}{3}x^3 - \frac{1}{4}x^4 - \cdots. \end{aligned}$$

辺々引いて

$$\frac{1}{2x} \log\left(\frac{1+x}{1-x}\right) - 1 = \frac{1}{3}x^2 + \frac{1}{5}x^4 + \frac{1}{7}x^6 + \cdots \leq \frac{1}{3}x^2 + \frac{1}{3}x^4 + \frac{1}{3}x^6 + \cdots = \frac{x^2}{1-x^2}.$$

$y = x^{-1}$ とおくと

$$0 < c(y) := \frac{y}{2} \log\left(\frac{y+1}{y-1}\right) - 1 = \frac{1}{3(y^2-1)} = \frac{1}{6(y-1)} - \frac{1}{6(y+1)}.$$

さてここで $A_n = \frac{n!}{(n/e)^n \sqrt{n}}$, $a_n = \log A_n$ とおくと

$$a_n = \log n! - \left(n + \frac{1}{2}\right) \log n + n$$

であり, さらに

$$\begin{aligned}
 a_n - a_{n+1} &= -\left(n + \frac{1}{2}\right) \log n - \log(n+1) + \left(n + \frac{3}{2}\right) \log(n+1) - 1 \\
 &= \left(n + \frac{1}{2}\right) \log \frac{n+1}{n} - 1 \\
 &= \frac{1}{2}(2n+1) \log \frac{(2n+1)+1}{(2n+1)-1} - 1 \\
 &= c(2n+1) \\
 &\leq \frac{1}{6} \frac{1}{(2n+1)-1} - \frac{1}{6} \frac{1}{(2n+1)+1} \\
 &= \frac{1}{12n} - \frac{1}{12(n+1)}.
 \end{aligned}$$

これから $a_n - a_{n+1} > 0$ と $a_n - \frac{1}{12n} \leq a_{n+1} - \frac{1}{12(n+1)}$ がわかる. すなわち a_n は減少で, $a_n - \frac{1}{12n}$ は増加. 後の性質から $a_n - \frac{1}{12n} \geq a_1 - \frac{1}{12}$ で, 結局 a_n は下に有界で, 極限が存在することがわかる. $a = \lim_{n \rightarrow \infty} a_n$, $A = e^a$ とおくと

$$n! \sim A(n/e)^n \sqrt{n}$$

がわかる. $A = \sqrt{2\pi}$ であることを見るにはウォリスの公式から

$$\sqrt{\pi} = \lim_{n \rightarrow \infty} \frac{2^{2n} (n!)^2}{\sqrt{n} (2n)!} = \lim_{n \rightarrow \infty} \frac{2^{2n} A^2 n^{2n+1} e^{-2n}}{A(2n)^{2n+1/2} e^{-2n} \sqrt{n}} = \frac{A}{\sqrt{2}}$$

に注意すればよい. □

さて3次元の対称ランダム・ウォークは \mathbb{Z}^3 -値の i.i.d. 確率変数列 X_n でその分布が

$$P(X_n = (\pm 1, 0, 0)) = P(X_n = (0, \pm 1, 0)) = P(X_n = (0, 0, \pm 1)) = \frac{1}{6}$$

となるものを用いて $S_n = X_1 + \dots + X_n$ で定義される. このとき次が成り立つ.

定理 1.10. 3次元対称ランダム・ウォークは非再帰的である.

証明 この場合は, $\sum_{n=0}^{\infty} P(S_{2n} = 0) < \infty$ を示せばよい. 容易に

$$P(S_{2n} = 0) = \frac{1}{6^{2n}} \sum_{\substack{j,k,l \geq 0 \\ j+k+l=n}} \frac{(2n)!}{(j!)^2 (k!)^2 (l!)^2} = \frac{1}{2^{2n}} \binom{2n}{n} \sum_{\substack{j,k,l \geq 0 \\ j+k+l=n}} \left(\frac{1}{3^n} \frac{n!}{j!k!l!} \right)^2.$$

ここで

$$p_{j,k,l} = \frac{1}{3^n} \frac{n!}{j!k!l!}$$

とおくと3項展開の公式から

$$\sum_{\substack{j,k,l \geq 0 \\ j+k+l=n}} p_{j,k,l} = 1$$

だから,

$$\sum_{\substack{j,k,l \geq 0 \\ j+k+l=n}} p_{j,k,l}^2 \leq \max_{\substack{j,k,l \geq 0 \\ j+k+l=n}} p_{j,k,l}$$

で抑えることにすれば、 $p_{j,k,l}$ の最大値は $n = 3m$ の場合は $(j, k, l) = (m, m, m)$ のとき、 $n = 3m \pm 1$ の場合は $(j, k, l) = (m, m, m \pm 1)$ のときに実現する。ここでスターリングの公式を使えば

$$p_{m,m,m} \sim \frac{1}{3^{3m}} \frac{\sqrt{2\pi}(3m)^{3m+1/2} e^{-3m}}{(\sqrt{2\pi}m^{m+1/2} e^{-m})^3} = \frac{3\sqrt{3}}{2\pi n}$$

が成り立つ。 $p_{m,m,m \pm 1}$ についても同様である。 $\frac{1}{2^{2n}} \binom{2n}{n}$ についてはウォリスの公式を使えばよいので結局

$$P(S_{2n} = 0) = O\left(\frac{1}{\sqrt{\pi n}} \frac{1}{n}\right) = O(n^{-3/2}).$$

これから $\sum_{n=0}^{\infty} P(S_{2n} = 0) < \infty$ が従う。 □

Notes

- ランダム・ウォークの再帰性の話は Durrett [1] § 3.2, 福島 [2] § 5.2 などにある。

関連図書

- [1] R. Durrett, “*Probability: theory and examples*,” Second edition, Duxbury Press, Belmont, 1996.
- [2] 福島 正俊, 確率論, 裳華房, 東京, 1998.
- [3] 志賀 特造, ルベーク積分から確率論, 共立出版, 東京, 2000.
- [4] 吉田 伸生, 確率の基礎から統計へ, 遊星社, 東京, 2012.