

立教大学 数理統計学序論 1/ 数理統計学 1

今野 良彦

大阪公立大学

2026 年 10 月 10 日

- 前回の講義の記号の復習
- 1.5.1 同時確率関数 (p.m.f.) と同時密度関数 (p.d.f.)
- 1.5.2 周辺分布
- 1.5.3 (確率的) 独立性と条件付き分布
- 1.6 多次元分布と i.i.d. 標本
i.i.d.=identically and independently distributed
- 2.1 期待値
- 2.2 確率ベクトルの期待値
- 2.3 分散と共分散
- 2.4 条件付き期待値
- 2.5 積率母関数

記号の復習 (1)

- $(\Omega, \mathcal{A}, \Pr)$: 確率空間.
 - $\mathcal{A}$ は σ 加法族.
 - $\Pr : \mathcal{A} \rightarrow [0, 1]$ は確率測度.
- $\mathcal{A}$ は σ 加法族 $\Leftrightarrow$ (1) $\Omega \in \mathcal{A}$; (2) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$;
(3) $A_n \in \mathcal{A} (n = 1, 2, \dots) \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.
- $\Pr; \mathcal{A} \ni A \mapsto \Pr(A) \in [0, 1]$ は以下をみたす.
 - (1) $\Pr(\Omega) = 1$;
 - (2) $\forall A \in \mathcal{A}$ に対して, $\Pr(A^c) = 1 - \Pr(A)$;
 - (3) 互いに排反 な $\{A_n\}_{n=1}^{\infty} \subset \mathcal{A}$ に対して

$$\Pr\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \Pr(A_n).$$

記号の復習 (2)

- $X : \Omega \longrightarrow \mathbb{R}$ は確率変数
 $\Leftrightarrow \forall B \in \mathcal{B}(\mathbb{R})$ に対して

$$X^{-1}(B) = \{\omega \in \Omega; X(\omega) \in B\} \in \mathcal{A}.$$

$$\Leftrightarrow \forall r \in \mathbb{R} \text{ に対して}$$

$$\{X > r\} = \{\omega \in \Omega; X(\omega) > r\} \in \mathcal{A}.$$

$$\Leftrightarrow \forall r \in \mathbb{R} \text{ に対して}$$

$$\{X \leq r\} = \{\omega \in \Omega; X(\omega) \leq r\} \in \mathcal{A}.$$

ただし, $\mathcal{B}(\mathbb{R})$ は $\mathbb{R}$ の開集合族を含む最小の σ 加法族. これを Borel 集合族という.

記号の復習 (3)

- 確率変数 X の c.d.f.

$$\mathbf{F}^X(x) = \mathbf{Pr}(X \leq x) \quad (x \in \mathbb{R}).$$

- 分布関数 $\mathbf{F}^X$ の性質

- $\mathbf{F}^X$ は非減少 $\Leftrightarrow (x < y \Rightarrow \mathbf{F}^X(x) \leq \mathbf{F}^X(y))$.
- $\mathbf{F}^X$ は右連続 $\Leftrightarrow \lim_{\epsilon \downarrow 0} \mathbf{F}^X(x + \epsilon) = \mathbf{F}^X(x)$.
- $\lim_{x \rightarrow -\infty} \mathbf{F}^X(x) = 0$; $\lim_{x \rightarrow \infty} \mathbf{F}^X(x) = 1$.
- 不連続点は高々可算個.

記号の復習 (4)

- $F^X(\cdot)$ が階段関数のとき, X を離散型確率変数という.
 - 確率変数 X の p.m.f.

$$p^X(x) = F^X(x) - F^X(x-).$$

ただし, $F^X(x-) = \lim_{\epsilon \downarrow 0} F^X(x - \epsilon)$, すなわち, x における F^X の左極限.

- $\{x \in \mathbb{R}; p^X(x) > 0\}$ (F^X の不連続点) は高々可算個集合.
- $F^X(\cdot)$ が連続関数のとき, X を連続型確率変数という.
 - 確率変数 X の p.d.f.: $\mathbf{R}$ 上の非負値関数 p^X で

$$F^X(x) = \int_{-\infty}^x p^X(t) dt \quad (x \in \mathbb{R}).$$

1.5.1 同時確率関数 (p.m.f.) と確率密度関数 (p.d.f.)

確率空間 $(\Omega, \mathcal{A}, \mathbf{Pr})$ 上の離散型確率変数 (X, Y) の対に対して, 同時確率関数 (同時 p.m.f.) $\mathbf{p}(\cdot, \cdot)$ を

	$Y = 0$	$Y = 1$	
$X = 0$	$1/9$	$2/9$	$1/3$
$X = 1$	$2/9$	$4/9$	$2/3$
	$1/3$	$2/3$	

で定める. たとえば

$$\mathbf{p}(1, 1) = \mathbf{Pr}(X = 1, Y = 1) = \frac{4}{9}$$

である.

2次元の分布 (2)

定義 1.45 確率空間 $(\Omega, \mathcal{A}, \Pr)$ 上の確率変数 X, Y は連続型とする. $\mathbb{R}^2$ 上の非負値実数値関数 $\mathbf{p}$ が確率ベクトル (X, Y) の同時確率密度関数 (同時 p.d.f.) であるとは, 次の条件をみたすときをいう.

- (1) $\mathbf{p}(x, y) \geq 0 \ (\forall (x, y) \in \mathbb{R}^2)$.
- (2) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{p}(x, y) \, dx dy = 1$.
- (3) $\forall A \in \mathcal{B}(\mathbb{R}^2)$ に対して

$$\Pr((X, Y) \in A) = \int \int_A \mathbf{p}(x, y) \, dx dy.$$

ただし, $\mathcal{B}(\mathbb{R}^2)$ は $\mathbb{R}^2$ の開集合族を含む最小の σ 加法族.

2次元の分布 (3)

注意 1.48 (X, Y) の同時累積分布関数 (同時 c.d.f.) $\mathbf{F}(\cdot, \cdot)$ を

$$\begin{aligned}\mathbf{F}(x, y) &:= \Pr(X \leq x, Y \leq y) \\ &:= \Pr(\{\omega \in \Omega; X(\omega) \leq x, Y(\omega) \leq y\}) \quad (\forall (x, y) \in \mathbb{R}^2)\end{aligned}\tag{1}$$

で定義する. (X, Y) の同時分布 $\mathbf{P}$ を

$$\mathbf{P}(A) = \Pr((X, Y) \in A) \quad (\forall A \in \mathcal{B}(\mathbb{R}^2))\tag{2}$$

で定義する. どの確率変数の c.d.f. または確率分布であることを明示したいときには, $\mathbf{F}^{(X, Y)}$ または $\mathbf{P}^{(X, Y)}$ と記す.

細かなことであるが, (1) の $\mathbf{F}$ によって $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ 上の確率測度 $\mathbf{P}$ を一意的に定めていることができる. この点に関しては測度論の知識が必要となる. □

2次元の分布 (4)

例 1.47 連続型確率変数のベクトル (X, Y) は同時 p.d.f.

$$\mathbf{p}(x, y) = \begin{cases} x + y & (0 < x < 1, 0 < y < 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

を持つとする. このとき集合 $\{(x, y) \in \mathbb{R}^2; \mathbf{p}(x, y) > 0\}$ に注意すれば

$$\begin{aligned} \int_0^1 \int_0^1 \mathbf{p}(x, y) \, dx dy &= \int_0^1 \left[\int_0^1 x \, dx \right] dy + \int_0^1 \left[\int_0^1 y \, dy \right] dx \\ &= \int_0^1 \frac{1}{2} dy + \int_0^1 \frac{1}{2} dx = 1 \end{aligned}$$

となる.

□

1.5.2 周边分布 (1)

例

	$Y = 0$	$Y = 1$	
$X = 0$	$1/9$	$2/9$	$1/3$
$X = 1$	$2/9$	$4/9$	$2/3$
	$1/3$	$2/3$	

1.5.2 周辺分布 (2)

定義 1.48 (X, Y) を確率空間 $(\Omega, \mathcal{A}, \mathbf{Pr})$ 上で定義された確率ベクトルとする.

(1) (X, Y) が離散型で同時 p.m.f. $\mathbf{p}$ を持つとする. X の周辺確率関数 (周辺 p.m.f.) を

$$\mathbf{p}^X(x) = \mathbf{Pr}(X = x) = \sum_{y \in S_Y} \mathbf{Pr}(X = x, Y = y) = \sum_{y \in S_Y} \mathbf{p}(x, y) \quad (\forall x \in S_X)$$

で定義する. ただし

$$S_X := \{x \in \mathbb{R}; \text{ある } y \in \mathbb{R} \text{ に対して, } \mathbf{p}(x, y) > 0\}$$

$$S_Y := \{y \in \mathbb{R}; \text{ある } x \in \mathbb{R} \text{ に対して, } \mathbf{p}(x, y) > 0\}$$

である.

1.5.2 周辺分布 (3)

定義 1.48 の続き (2) (X, Y) は連続型とし, 同時 p.d.f. $\mathbf{p}$ を持つとする. このとき X の周辺確率密度関数 (周辺 p.d.f.) を

$$\mathbf{p}^X(x) = \int_{-\infty}^{\infty} \mathbf{p}(x, y) \, dy \quad (\forall x \in \mathbb{R})$$

で定義し, Y の周辺確率密度関数 (周辺 p.d.f.) を

$$\mathbf{p}^Y(y) = \int_{-\infty}^{\infty} \mathbf{p}(x, y) \, dx \quad (\forall y \in \mathbb{R})$$

で定義する.

□

1.5.2 周辺分布 (3)

注意 連続型確率ベクトル (X, Y) に対して

$$\begin{aligned} \mathbf{F}^X(x) &:= \Pr(X \leq x) = \int \int_{(s,t) \in \mathbb{R}^2; s \leq x} \mathbf{p}(s, t) \, ds dt \\ &= \int_{-\infty}^x \left\{ \int_{-\infty}^{\infty} \mathbf{p}(s, t) \, ds \right\} dt = \int_{-\infty}^x \mathbf{p}^X(s) \, ds \end{aligned}$$

となるので, X の周辺 p.d.f. と X の p.d.f. は高々可算個の点を除いて同じ値を取ることがわかる. □

1.5.3 (確率的) 独立性と条件付き分布 (1)

定義 1.51 確率空間 $(\Omega, \mathcal{A}, \mathbf{Pr})$ 上の 2 つの確率変数 X と Y は独立であるとは, $\forall A, B \in \mathcal{B}(\mathbb{R})$ に対して

$$\mathbf{Pr}(X \in A, Y \in B) = \mathbf{Pr}(X \in A)\mathbf{Pr}(Y \in B)$$

が成り立つときをいう. 独立でないとき X と Y は従属であるという.

1.5.3 (確率的) 独立性と条件付き分布 (2)

定理 1.52 $\mathbf{p}$ を確率ベクトル (X, Y) の同時 p.d.f./p.m.f. とし, $\mathbf{p}^X$ と $\mathbf{p}^Y$ を X と Y それぞれの周辺 p.d.f. とする. このとき

$$X \text{ と } Y \text{ は独立} \Leftrightarrow \mathbf{p}(x, y) = \mathbf{p}^X(x)\mathbf{p}^Y(y) \quad (\forall x, y \in \mathbb{R}).$$

となる.

証明 独立ならば, 同時 p.d.f. が周辺 p.d.f./p.m.f. の積で表現されることの証明は易しい. 逆はこの講義の範囲 (測度の拡張の議論が必要となる!) を超えるので, 信じることにする. $\square$

1.5.3 (確率的) 独立性と条件付き分布 (3)

注意 1.53 (1) 連続型の場合には, 測度 $\mathbf{0}$ の集合を除いて¹p.d.f. は定義されるので, 定理 1.52 の書き方はやや数学的な厳密性に欠ける.

□

¹この用語は定義されていない。「有限個の点を除いて」と理解してもこの講義の内容の範囲では問題ない.

1.5.3 (確率的) 独立性と条件付き分布 (4)

例 1.54 連続型確率変数 X と Y は独立で同じ p.d.f.

$$p(x) = \begin{cases} 2x & (0 < x < 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

を持つとする. このとき確率 $\Pr(X + Y \leq 1)$ を求めてみよう. 独立性より (X, Y) の同時 p.d.f. は

$$p^{(X,Y)}(x, y) = \begin{cases} 4xy & (0 < x < 1, 0 < y < 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

となる.

1.5.3 (確率的) 独立性と条件付き分布 (5)

例 1.54 の続き したがって

$$\begin{aligned}\Pr(X + Y \leq 1) &= \int \int_{x+y \leq 1} \mathbf{p}^{(X,Y)}(x, y) \, dx dy \\ &= 4 \int_0^1 x \left\{ \int_0^{1-x} y \, dy \right\} \, dx \\ &= 4 \int_0^1 x \frac{(1-x)^2}{2} \, dx \\ &= \frac{1}{6}\end{aligned}$$

となる.

□

1.5.3 (確率的) 独立性と条件付き分布 (6)

定義 1.58 (1) 離散型確率ベクトル (X, Y) は同時 p.m.f. $\mathbf{p}(x, y)$ を持つとする. $\mathbf{p}^Y(y) > 0$ なる y に対して, $Y = y$ を与えたときの X の条件付き確率関数 (条件付き p.m.f.) を

$$\begin{aligned}\mathbf{p}^{X|Y}(x|y) &= \Pr(X = x | Y = y) = \frac{\Pr(X = x, Y = y)}{\Pr(Y = y)} \\ &= \frac{\mathbf{p}(x, y)}{\mathbf{p}^Y(y)} \quad (x \in \mathbb{R})\end{aligned}$$

で定義する.

注意 (1) $\mathbf{p}^{X|Y}(x|y)$ は x が変数の $\mathbb{R}$ 上の関数であることに注意せよ.

(2) さらに $\{x \in \mathbb{R}; \mathbf{p}^{X|Y}(x|y) > 0\}$ は高々可算個の元からなる集合である.

1.5.3 (確率的) 独立性と条件付き分布 (7)

定義 1.58 の続き (2) 連続型確率ベクトル (X, Y) は同時 p.d.f. $\mathbf{p}(x, y)$ を持つとする. $\mathbf{p}^Y(y) > 0$ なる y に対して, $Y = y$ を与えたときの X の条件付き確率密度関数 (条件付き p.d.f.) を

$$\mathbf{p}^{X|Y}(x|y) = \frac{\mathbf{p}(x, y)}{\mathbf{p}^Y(y)} \quad (x \in \mathbb{R})$$

で定義する.

1.5.3 (確率的) 独立性と条件付き分布 (8)

注意 1.59 (1) (X, Y) が連続型確率ベクトルのとき, $\mathbf{p}^Y(y) > 0$ なる $y \in \mathbb{R}$ に対して, $Y = y$ を与えたときの事象 $\{X \in A\}$ の条件付き確率を

$$\Pr(X \in A | Y = y) := \int_A \mathbf{p}^{X|Y}(x|y) \, dx \quad (\forall A \in \mathcal{B}(\mathbb{R}))$$

と形式的に定義する.

(2) A を固定したとき, $\Pr(X \in A | Y = y)$ は y にのみ依存するので, 形式的に $h(y) := \Pr(X \in A | Y = y)$ と書いた²ときに, $h(y)$ に形式的に確率変数 Y を代入したものを

$$\Pr(X \in A | Y) := h(Y)$$

と書くと, $\Pr(X \in A | Y)$ も確率変数となる.

² h の可測性は自明ではないが, 信じることにする.

1.5.3 (確率的) 独立性と条件付き分布 (9)

注意 1.59 の続き (3) $A, B \in \mathcal{A}$ とし, $\Pr(B) > 0$ とする. A と B が独立のとき

$$\Pr(A|B) = \Pr(A)$$

が成立する.

(4) $\Pr(B) > 0$ のとき

$$\Pr(A \cap B) = \Pr(B)\Pr(A|B)$$

となる. これを乗法の公式という.

(5) $\Pr(B) > 0$ のとき, 関数 $\Pr(\cdot|B) : \mathcal{A} \rightarrow [0, 1]$ は定義 1.6(1) – (2) をみたす. すなわち $(\Omega, \mathcal{A})$ 上の確率測度である.

1.5.3 (確率的) 独立性と条件付き分布 (10)

注意 1.59 の続き (6) 「独立 $\Rightarrow$ 対独立」であるが, 逆の反例:
 $\Omega = \{b_1, b_2, b_3, b_4\}$ とし

$$\Pr(\{b_j\}) = \frac{1}{4} \quad (j = 1, 2, 3, 4)$$

とする. いま

$$A_j = \{b_j, b_4\} \quad (j = 1, 2, 3)$$

とする. すると

$A_1 \cap A_2 = A_1 \cap A_3 = A_2 \cap A_3 = A_1 \cap A_2 \cap A_3 = \{b_4\}$
である. したがって

$$\begin{aligned} \Pr(A_1 \cap A_2) &= \Pr(A_1)\Pr(A_2); \Pr(A_1 \cap A_3) = \Pr(A_1)\Pr(A_3); \\ \Pr(A_2 \cap A_3) &= \Pr(A_2)\Pr(A_3). \end{aligned}$$

1.5.3 (確率的) 独立性と条件付き分布 (11)

注意 1.59 の続き しかし

$$\frac{1}{4} = \mathbf{Pr}(A_1 \cap A_2 \cap A_3) \neq \mathbf{Pr}(A_1)\mathbf{Pr}(A_2)\mathbf{Pr}(A_3) = \frac{1}{8}$$

である.

□

1.5.3 (確率的) 独立性と条件付き分布 (12)

例 1.60 連続型確率ベクトル (X, Y) は同時 p.d.f.

$$p(x, y) = \begin{cases} x + y & (0 < x < 1, 0 < y < 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

を持つとする. このとき $\Pr(X \leq 1/4 | Y = 1/3)$ を求めてみよう. まず $0 < y < 1$ に対して

$$p^Y(y) = \int_{-\infty}^{\infty} p(x, y) dx = y + \int_0^1 x dx = y + \frac{1}{2}.$$

1.5.3 (確率的) 独立性と条件付き分布 (13)

例 1.60 の続き したがって

$$\mathbf{p}^Y(y) = \begin{cases} y + \frac{1}{2} & (0 < y < 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

となる. よって $0 < y < 1$ と $0 < x < 1$ に対して

$$\mathbf{p}^{X|Y}(x|y) = \frac{\mathbf{p}(x, y)}{\mathbf{p}^Y(y)} = \frac{x + y}{y + \frac{1}{2}}$$

となる.

1.5.3 (確率的) 独立性と条件付き分布 (14)

例 1.60 の続き これより $0 < y < 1$ のとき

$$p^{X|Y}(x|y) = \begin{cases} \frac{x+y}{y+\frac{1}{2}} & (0 < x < 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

となる. 注意 1.59 より

$$\begin{aligned} \Pr\left(X \leq \frac{1}{4} \middle| Y = \frac{1}{3}\right) &= \int_0^{1/4} p^{X|Y}\left(x \middle| \frac{1}{3}\right) dx = \int_0^{1/4} \frac{x + \frac{1}{3}}{\frac{1}{3} + \frac{1}{2}} dx \\ &= \frac{11}{80}. \end{aligned}$$

□

1.6 多次元の分布と i.i.d. 標本 (1)

$X_1, X_2, \dots, X_n$ を確率空間 $(\Omega, \mathcal{A}, \mathbf{Pr})$ 上の確率変数とし

$$X = (X_1, X_2, \dots, X_n)^\top = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$$

と書く. X を確率ベクトルという. $(\cdot)^\top$ は転置作用素.

注意 1.61 本講義録では, ベクトルは縦ベクトルとする.

□

1.6 多次元の分布と i.i.d. 標本 (2)

□

定義 1.62 $X_1, X_2, \dots, X_n$ が独立であるとは, すべての Borel 集合 $A_1, A_2, \dots, A_n \in \mathcal{B}(\mathbb{R})$ に対して

$$\Pr(X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n) = \prod_{j=1}^n \Pr(X_j \in A_j) \quad (3)$$

が成立するときである.

□

1.6 多次元の分布と i.i.d. 標本 (3)

□

注意 1.63 $X = (X_1, \dots, X_n)^\top$ の同時 p.d.f. を $\mathbf{p}(x_1, x_2, \dots, x_n)$ と書き, 各 X_j ($j = 1, 2, \dots, n$) の周辺 p.d.f. を $\mathbf{p}^{X_j}$ と書くことにする. 同時 p.d.f. が存在する場合には, (3) を確認するために

$$\mathbf{p}(x_1, x_2, \dots, x_n) = \prod_{j=1}^n \mathbf{p}^{X_j}(x_j) \quad (\forall x_1, x_2, \dots, x_n \in \mathbb{R})$$

を示せばよい.

□

1.6 多次元の分布と i.i.d. 標本 (4)

定義 1.64 $X_1, X_2, \dots, X_n$ は独立で, 各 X_j ($j = 1, 2, \dots, n$) は同じ c.d.f. $\mathbf{F}$ を持つとき, $X_1, X_2, \dots, X_n$ は 独立同一分布に従う (i.i.d. = identically and independently distributed) とい

$$X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mathbf{F}$$

と書く³. $X_1, X_2, \dots, X_n$ は累積分布関数 $\mathbf{F}$ からの標本の大きさが n のランダム標本 ともいう.

□

³p.d.f. $\mathbf{p}$ を使い

$$X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mathbf{p}$$

と書く. $\sim$ の右側には, 確率測度/確率分布/p.d.f./p.m.f./ $\mathbf{N}(\mathbf{0}, \mathbf{1})$ など分布を特定するものを書いてよいことにする.

2.1 期待値 (1)

X を確率空間 $(\Omega, \mathcal{A}, \mathbf{Pr})$ 上の確率変数とする. X が離散型るとき, その p.m.f. を $\mathbf{p}$ とし, X の取り得る値を $x_1, x_2, \dots$ とする. 連続型るとき, その p.d.f. も $\mathbf{p}$ と書くことにする.

2.1 期待値 (2)

定義 2.1 $g : \mathbb{R} \rightarrow \mathbb{R}$ を可測⁴関数とする. 確率変数 $g(X)$ の期待値 $\mathbf{E}[g(X)]$ を次のように定義する.

(1) $g(x) \geq 0 (x \in \mathbb{R})$ のとき

$$\mathbf{E}[g(X)] = \begin{cases} \sum_{n=1}^{\infty} g(x_n) \mathbf{p}(x_n) & (\text{離散型の場合}) \\ \int_{-\infty}^{\infty} g(x) \mathbf{p}(x) \, dx & (\text{連続型の場合}) \end{cases}$$

と定義する. 右辺は ∞ を許せば, 必ず存在する.

⁴関数 g は可測, $\Leftrightarrow \{x \in \mathbb{R}; f(x) > r\} \in \mathcal{B}(\mathbb{R}) \ (\forall r \in \mathbb{R})$.

2.1 期待値 (3)

定義 2.1 の続き

(2) 一般の可測関数 g に対して

$$g^+(x) = \max\{g(x), 0\}, \quad g^-(x) = \max\{-g(x), 0\}$$

と定義すれば, $g^+(x) \geq 0$, $g^-(x) \geq 0$ ($x \in \mathbb{R}$) となる.
 $\mathbf{E}[g^+(X)]$ または $\mathbf{E}[g^-(X)]$ のいずれかが有限ならば

$$\mathbf{E}[g(X)] := \mathbf{E}[g^+(X)] - \mathbf{E}[g^-(X)]$$

と定義する. $\mathbf{E}[g^+(X)] = \mathbf{E}[g^-(X)] = \infty$ のときは, $g(X)$ の期待値は定義されない. $\mathbf{E}[g^+(X)] < \infty$ かつ $\mathbf{E}[g^-(X)] < \infty$ のとき, $\mathbf{E}[g(X)]$ は有限となる.

□

2.1 期待値 (4)

補題 2.2 X を確率変数とする. 関数 $g, h : \mathbb{R} \rightarrow \mathbb{R}$ を可測とする.

(1) $h(X)$ は有限の期待値を持つ⁵ とする.

$0 \leq g(x) \leq h(x) (\forall x \in \mathbb{R})$ ならば

$$\mathbf{E}[g(X)] \leq \mathbf{E}[h(X)].$$

(2) $\mathbf{E}[|g(X)|] < \infty, \mathbf{E}[|h(X)|] < \infty$ を仮定する. $a, b \in \mathbb{R}$ に対して

$$\mathbf{E}[ag(X) + bh(X)] = a \mathbf{E}[g(X)] + b \mathbf{E}[h(X)].$$

⁵ g, h を共に非負値関数としているので, $\mathbf{E}[|h(X)|] < \infty$ だけを仮定すればよい. もし, g, h を可測関数とすると $\mathbf{E}[|g(X)|] < \infty$ の仮定も必要になる. なぜならば, $g(x) \leq h(x) (x \in \mathbb{R})$ と $\mathbf{E}[|h(X)|] < \infty$ だけだと $\mathbf{E}[h(X)]$ が定義できない場合があるからである. なので, 非負値関数に関する主張とした方が実用的である.

2.1 期待値 (5)

補題 2.3 X を非負値確率変数とする. このとき,

$$\mathbf{E}[X] = 0 \Rightarrow \mathbf{Pr}(X = 0) = 1$$

となる.

補題 2.3 の証明⁶ まず

$$\begin{aligned} A &:= \{\omega \in \Omega; X(\omega) > 0\} = \{X > 0\}, \\ A_n &:= \left\{ \omega \in \Omega; X(\omega) > \frac{1}{n} \right\} \quad (n = 1, 2, \dots) \end{aligned}$$

とおく. すると $A_1 \subset A_2 \subset \dots$ かつ $\bigcup_{n=1}^{\infty} A_n = A$ となる.

⁶ 当たり前の主張にみえるが, 証明はやや Technical である. 

2.1 期待値 (6)

証明の続き よって, 補題 1.7(5) から

$$\lim_{n \rightarrow \infty} \Pr(A_n) = \Pr\left(\bigcup_{n=1}^{\infty} A_n\right) = \Pr(A). \quad (4)$$

一方, $A_n \subset A$ なので, $0 \leq \mathbb{1}_{A_n}(\omega) \leq \mathbb{1}_A(\omega)$ ($\omega \in \Omega$) である. このことに注意すると

$$\begin{aligned} 0 &= \mathbf{E}[X] = \mathbf{E}\left[\underbrace{X\mathbb{1}_{A^c}}_{\geq 0} + X\mathbb{1}_A\right] \geq \mathbf{E}[X\mathbb{1}_A] \geq \mathbf{E}[X\mathbb{1}_{A_n}] \\ &\geq \frac{1}{n}\mathbf{E}[\mathbb{1}_{A_n}] = \frac{1}{n}\Pr(A_n). \end{aligned}$$

2.1 期待値 (7)

証明の続き したがって, $\Pr(A_n) = 0 \ (\forall n \in \mathbb{N})$ がわかる. (4) と合わせると $\Pr(A) = 0$ となる. よって

$$\begin{aligned} 1 &= \Pr(A^c) = \Pr\left(\left(\{X \leq 0\} \cap \{X \geq 0\}\right) \cup \left(\{X \leq 0\} \cap \{X < 0\}\right)\right) \\ &= \Pr(X = 0) + \underbrace{\Pr(X < 0)}_{=0} = \Pr(X = 0) \end{aligned}$$

が示せた.

□

2.1 期待値 (8)

補題 2.4 $0 < q < r$ に対して

$$\mathbf{E}[|X|^r] < \infty \Rightarrow \mathbf{E}[|X|^q] < \infty$$

となる.

証明

証明は Young の不等式を用いて行う.

□

2.1 期待値 (9)

定義 2.10

- (1) $k = 1, 2, \dots$ に対して, $\mathbf{E}[|X|^k] < \infty$ のとき, $\mathbf{E}[X^k]$ を X の k 次モーメント (または積率) という.
- (2) $\mathbf{E}[X]$ を X の平均値⁷ という.
- (3) $\mathbf{E}[X^2] < \infty$ のとき X の分散を

$$\text{Var}[X] := \mathbf{E}[\{X - \mathbf{E}[X]\}^2]$$

で定義する.

⁷簡単に「平均」ともいう.

2.1 期待値 (10)

定義 2.10 の続き

(4) 部分集合 $A \subset \mathbb{R}$ に対して

$$\mathbb{1}_A(x) = \begin{cases} 1 & (x \in A) \\ 0 & (x \notin A) \end{cases}$$

を A の指示関数⁸ という.

(5) $A \in \mathcal{B}(\mathbb{R})$ のとき

$$\mathbf{E}[\mathbb{1}_A(X)] = \Pr(X \in A)$$

となる.

注意 補題 2.3 から, $\mathbf{E}[X^2] < \infty$ のとき

$$\mathrm{Var}[X] = 0 \Rightarrow \Pr(X = \mathbf{E}[X]) = 1$$

となる.

⁸指示関数は $\mathbb{R}$ の任意の部分集合に定義ができることに注意をせよ.



2.1 期待値 (11)

注意 2.11 分散 $\text{Var}[X]$ は X の分布の平均 μ まわりの散らばりを測る量である. 分散がおおきいほど分布は広がっていることになる. $\square$

注意 2.12 (1) X は確率空間 $(\Omega, \mathcal{A}, \mathbf{Pr})$ 上の離散型確率変数とする. $\mathbf{p}^X$ を X の p.m.f. とし

$$S_X := \{x \in \mathbb{R}; \mathbf{p}^X(x) > 0\}$$

とする. 関数 $g: \mathbb{R} \rightarrow \mathbb{R}$ に対して

$$Y = g(X)$$

とおく.

2.1 期待値 (12)

注意 2.12 の続き $y \in \mathbb{R}$ に対して

$$A_y := \{x \in S_X; g(x) = y\}$$

とおき

$$S_Y := \{y \in \mathbb{R}; A_y \neq \emptyset\}$$

とする. $y \in S_Y$ に対して

$$\mathbf{p}^Y(y) := \Pr(Y = y) = \Pr(X \in A_y) = \sum_{x \in A_y} \Pr(X = x) = \sum_{x \in A_y} \mathbf{p}^X(x)$$

と書ける. 正項級数は項の順番を並び替えてもその値は変わらないので

2.1 期待値 (13)

注意 2.12 の続き

$$\begin{aligned}\sum_{y \in S_Y} |y| \mathbf{p}^Y(y) &= \sum_{y \in S_Y} |y| \sum_{x \in A_y} \mathbf{p}^X(x) = \sum_{y \in S_Y} \sum_{x \in A_y} |g(x)| \mathbf{p}^X(x) \\ &= \sum_{x \in S_X} |g(x)| \mathbf{p}^X(x)\end{aligned}$$

となる. さらにいずれかの和が有限ならば

$$\sum_{y \in S_Y} y \mathbf{p}^Y(y) = \sum_{y \in S_Y} y \sum_{x \in A_y} \mathbf{p}^X(x) = \sum_{y \in S_Y} \sum_{x \in A_y} g(x) \mathbf{p}^X(x) = \sum_{x \in S_X} g(x) \mathbf{p}^X(x)$$

となる.

2.1 期待値 (14)

注意 2.12 の続き よって

$$\mathbf{E}[Y] = \mathbf{E}[g(X)]$$

となる.

(2) 確率変数 X は連続型とする. $\mathbf{p}^X$ を X の p.d.f. とし

$$S_X := \{x \in \mathbb{R}; \mathbf{p}^X(x) > 0\}$$

とする. 関数 $g : S_X \rightarrow \mathbb{R}$ は狭義単調増加かつ C^1 級⁹ とする. このとき

$$Y = g(X), \quad S_Y := \{y \in \mathbb{R}; y = g(x) (\exists x \in S_X)\}$$

とする.

⁹1 回連続微分可能のこと.

2.1 期待値 (15)

注意 2.12 の続き ここで $g : S_X \rightarrow S_Y$ と制限¹⁰ すれば g の逆関数 $g^{-1} : S_Y \rightarrow S_X$ が存在し

$$\begin{aligned} F^Y(y) &:= \Pr(Y \leq y) = \Pr(g^{-1}(Y) \leq g^{-1}(y)) = \Pr(X \leq g^{-1}(y)) \\ &= F^X(g^{-1}(y)). \end{aligned}$$

さらに $g(g^{-1}(y)) = y$ より $\frac{d}{dy} g^{-1}(y) = \frac{1}{\dot{g}(g^{-1}(y))}$ となる. ただし $\dot{g}(y) = \frac{dg}{dy}(y)$ である.

¹⁰制限したものを同じ g を用いて表現するのは, 記号の乱用である. 記号が煩雑になるので, 記号を乱用した.

2.1 期待値 (16)

注意 2.12 の続き これらより, $y \in S_Y$ に対して

$$\begin{aligned} p^Y(y) &= \frac{d}{dy} F^Y(y) = \frac{d}{dy} F^X(g^{-1}(y)) = \dot{F}^X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y) \\ &= p^X(g^{-1}(y)) \frac{1}{\dot{g}(g^{-1}(y))} \end{aligned}$$

となる. よって

$$p^Y(y) = p^X(g^{-1}(y)) \frac{1}{\dot{g}(g^{-1}(y))}$$

となる.

2.1 期待値 (17)

注意 2.12 の続き g が狭義単調減少の場合もふくめると

$$p^Y(y) = p^X(g^{-1}(y)) \frac{1}{|\dot{g}(g^{-1}(y))|}$$

となる. 以上のことから

$$\begin{aligned} E[Y] &= \int_{S_Y} y p^Y(y) dy = \int_{S_Y} y p^X(g^{-1}(y)) \frac{1}{|\dot{g}(g^{-1}(y))|} dy \\ &= \int_{S_X} g(x) p^X(x) dx \\ &\quad (\because x = g^{-1}(y) \Leftrightarrow y = g(x); dy = |\dot{g}(x)| dx) \\ &= E[g(X)]. \end{aligned}$$

□

2.2 確率ベクトルの期待値 (1)

確率変数 X, Y を確率空間 $(\Omega, \mathcal{A}, \mathbf{Pr})$ 上で定義された確率変数とする. (X, Y) を確率ベクトルという.

確率ベクトル (X, Y) が離散型のときその同時 p.m.f. を $\mathbf{p}(x, y)$ とし, 連続型のときその同時 p.d.f. も $\mathbf{p}(x, y)$ と書くことにする.

2.2 確率ベクトルの期待値 (2)

定義 2.13 可測関数 $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ に対して $g(X, Y)$ の期待値を次のように定義する.

(1) $g \geq 0$ のとき

$$\mathbf{E}[g(X, Y)] = \begin{cases} \sum_{x,y} g(x, y) \mathbf{p}(x, y) & \text{(離散型の場合)} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \mathbf{p}(x, y) \, dx dy & \text{(連続型の場合)} \end{cases}$$

で定義する.

2.2 確率ベクトルの期待値 (3)

定義 2.13 の続き

(2) 一般の g に対して

$$g^+(x, y) := \max\{g(x, y), 0\}, \quad g^-(x, y) := \max\{-g(x, y), 0\}$$

と定義すれば $g^+, g^- \geq 0$ となる. $E[g^+(X, Y)]$ または $E[g^-(X, Y)]$ のいずれかが有限ならば

$$E[g(X, Y)] := E[g^+(X, Y)] - E[g^-(X, Y)]$$

で定義する. $E[g^+(X, Y)] = E[g^-(X, Y)] = \infty$ のときは, $g(X, Y)$ の期待値は定義されない. $E[g^+(X, Y)] < \infty$ かつ $E[g^-(X, Y)] < \infty$ のとき, $E[g(X, Y)]$ は有限の値を取る.

2.2 確率ベクトルの期待値 (4)

注意 2.14 3 つ以上の確率変数 $X_1, X_2, \dots, X_n$ に対しても期待値を定義 2.13 と同様に定義する.

定理 2.15 $X_1, X_2, \dots, X_n$ を確率変数とし, 各 X_j ($j = 1, 2, \dots, n$) の期待値は有限とする. $a_1, a_2, \dots, a_n$ を定数としたとき

$$\mathbf{E}\left[\sum_{j=1}^n a_j X_j\right] = \sum_{j=1}^n a_j \mathbf{E}[X_j]$$

となる.

2.2 確率ベクトルの期待値 (5)

定理 2.15 の証明 まず, $|\sum_{j=1}^n a_j x_j| \leq \sum_{j=1}^n a_j |x_j|$ なので, 補題 2.2(1) から $\sum_{j=1}^n a_j X_j$ は有限な期待値をもつ.

$X := (X_1, X_2, \dots, X_n)^\top$ とおき, X は同時 p.d.f. $\mathbf{p}^X$ を持つとする. すると定義 2.8 から 補題 2.2(2) と同様の期待値の線型性から

$$\begin{aligned}\mathbf{E}\left[\sum_{j=1}^n a_j X_j\right] &= \int_{\mathbb{R}^n} \left\{ \sum_{j=1}^n a_j x_j \right\} \mathbf{p}^X(x_1, x_2, \dots, x_n) \, dx_1 dx_2 \cdots dx_n \\ &= \sum_{j=1}^n a_j \int_{\mathbb{R}^n} x_j \mathbf{p}^X(x_1, x_2, \dots, x_n) \, dx_1 dx_2 \cdots dx_n = \sum_{j=1}^n a_j \mathbf{E}[X_j].\end{aligned}$$

□

2.2 確率ベクトルの期待値 (5)

例 2.16 $0 < p < 1$ と $j = 1, 2, \dots, n$ に対して, $X_j \sim \mathbf{Ber}(p)$ とする. このとき

$$\mathbf{E}[X_j] = \sum_{x=0}^1 x \Pr(X = x) = 0 \times (1 - p) + 1 \times p = p$$

である. $S = \sum_{j=1}^n X_j$ としたとき

$$\mathbf{E}[S] = \sum_{j=1}^n \mathbf{E}[X_j] = np$$

がわかる.

□

2.3 分散と共分散 (1)

X を確率変数とし, $E[X^2] < \infty$ とする. X の分散を

$$\text{Var}[X] := E[(X - \mu)^2]$$

で定義した. ただし $\mu = E[X]$ と書いた. さらに $\sqrt{\text{Var}[X]}$ を X の標準偏差 という.

2.3 分散と共分散 (2)

定理 2.17 以下の確率変数は有限の 2 次の積率を持つとする. このとき, 次が成立する.

- (1) $\text{Var}[X] = \mathbf{E}[X^2] - \{\mathbf{E}[X]\}^2$ となる. したがって, $\text{Var}[X] \geq 0$ から, $\mathbf{E}[X^2] \geq \{\mathbf{E}[X]\}^2$ となる.
- (2) 定数 $a, b \in \mathbb{R}$ に対して

$$\text{Var}[aX + b] = a^2 \text{Var}[X]$$

となる.

- (3) X と Y は独立で $\mathbf{E}[|XY|] < \infty$ とする. このとき

$$\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$$

となる.

2.3 分散と共分散 (3)

定理 2.17 の続き

- (4) $X_1, X_2, \dots, X_n$ は独立とし, $\mathbf{E}[X_j^2] < \infty$ ($j = 1, 2, \dots, n$) とする. $a_1, a_2, \dots, a_n$ は定数としたとき

$$\mathbf{Var}\left[\sum_{j=1}^n a_j X_j\right] = \sum_{j=1}^n a_j^2 \mathbf{Var}[X_j]$$

となる.

2.3 分散と共分散 (4)

例 2.18 $0 < p < 1$ と $j = 1, 2, \dots, n$ に対して, $X_j \sim \mathbf{Ber}(p)$ とし, $X_1, X_2, \dots, X_n$ は独立とする. すると定理 2.17 に注意すれば

$$\begin{aligned}\mathrm{Var}[X_j] &= \mathbf{E}[X_j^2] - \{\mathbf{E}[X_j]\}^2 = \sum_{x=0}^1 x^2 \Pr(X = x) - p^2 \\ &= 0 \times (1 - p) + 1 \times p - p^2 = p(1 - p)\end{aligned}$$

となる. これと定理 2.17(4) から

$$\mathrm{Var}[S] = \sum_{j=1}^n \mathrm{Var}[X_j] = np(1 - p).$$

□

2.3 分散と共分散 (5)

定理 2.19 $n \geq 2$ とし, $X_1, X_2, \dots, X_n$ は i.i.d. 確率変数列とし

$$\mathbf{E}[X_1] = \mu, \quad \mathbf{Var}[X_1] = \sigma^2$$

とする. ただし, $\mu \in \mathbb{R}$, $0 < \sigma < \infty$ である. $X_1, X_2, \dots, X_n$ に基づく標本平均を

$$\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j$$

で定義し, 標本 (不偏) 分散 を

$$S_n^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X}_n)^2$$

で定義する.

2.3 分散と共分散 (6)

定理 2.19 の続き このとき

$$(1) \mathbf{E}[\bar{X}_n] = \mu; \quad (2) \mathbf{Var}[\bar{X}_n] = \frac{\sigma^2}{n}; \quad (3) \mathbf{E}[S_n^2] = \sigma^2$$

となる.

2.3 分散と共分散 (7)

定理 2.19 の証明 (1), (2) は定理 2, 15, 2.17(4) よりわかる. (3) を証明するために

$$\sum_{j=1}^n (X_j - \bar{X}_n)^2 = \sum_{j=1}^n (X_j - \mu)^2 - n(\bar{X}_n - \mu)^2$$

に注意する. 定理 2.12(1) より

$$\mathbb{E}\left[\sum_{j=1}^n (X_j - \mu)^2\right] = n\sigma^2$$

がわかる. さらに, (1) と (2) より

$$\mathbb{E}[(\bar{X}_n - \mu)^2] = \text{Var}[\bar{X}_n] = \frac{\sigma^2}{n}.$$

2.3 分散と共分散 (8)

定理 2.19 の証明の続き

したがって

$$\begin{aligned}\mathbf{E}[S_n^2] &= \frac{1}{n-1} \left\{ \sum_{j=1}^n \mathbf{E}[(X_j - \mu)^2] - n\mathbf{E}[(\bar{X}_n - \mu)^2] \right\} \\ &= \frac{1}{n-1} \{n\sigma^2 - \sigma^2\} = \sigma^2\end{aligned}$$

がわかる.

□

2.3 分散と共分散 (9)

定義 2.20 X と Y は確率変数とし

$$\mathbf{E}[X] = \mu_X, \quad \mathbf{Var}[X] = \sigma_X^2, \quad \mathbf{E}[Y] = \mu_Y, \quad \mathbf{Var}[Y] = \sigma_Y^2$$

とする. ただし $\mu_X, \mu_Y \in \mathbb{R}$, $0 < \sigma_X, \sigma_Y < \infty$ とする. このとき
 X と Y の共分散 と X と Y の (Pearson) の相関係数 を

$$\begin{aligned} \mathbf{Cov}[X, Y] &:= \mathbf{E}[(X - \mu_X)(Y - \mu_Y)], \\ \rho &:= \rho[X, Y] := \frac{\mathbf{Cov}[X, Y]}{\sigma_X \sigma_Y} \end{aligned}$$

で定義する.

注意 $\mathbf{E}[X^2] < \infty, \mathbf{E}[Y^2] < \infty \Rightarrow \mathbf{E}[|XY|] < \infty$.

2.3 分散と共分散 (10)

定理 2.21 X, Y, Z は 2 次の積率が有限な確率変数とする.

(1) 共分散は

$$\mathbf{Cov}[X, Y] = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y]$$

と書き直せる.

(2) $\mathbf{Cov}[X, Y] = \mathbf{Cov}[Y, X]$ である.

(3) 定数 a, b に対して

$$\mathbf{Cov}[aX + bY, Z] = a\mathbf{Cov}[X, Z] + b\mathbf{Cov}[Y, Z]$$

となる.

(4) 相関係数は

$$-1 \leq \rho[X, Y] \leq 1$$

をみたす.

2.3 分散と共分散 (11)

定理 2.21 の続き (5) ある定数 a, b ($a \neq 0$) が存在して
 $Y = aX + b$ となったとき

$$a > 0 \Rightarrow \rho[X, Y] = 1,$$

$$a < 0 \Rightarrow \rho[X, Y] = -1$$

である.

(6) X と Y が独立のとき

$$\mathbf{Cov}[X, Y] = 0$$

となる.

2.3 分散と共分散 (12)

注意 2.22 $\mathbf{E}[X] = \mathbf{E}[Y] = \mathbf{0}$ の場合を考えると Cauchy-Schwarz の不等式

$$|\mathbf{E}[XY]| \leq \sqrt{\mathbf{E}[X^2]} \sqrt{\mathbf{E}[Y^2]}$$

と呼ばれる重要なものを得る.

□

注意 2.22 定理 2.21(6) の逆

$$\mathbf{Cov}[X, Y] = \mathbf{0} \Rightarrow X \quad \text{と} \quad Y \text{ は独立}$$

は一般に正しくない.

2.3 分散と共分散 (13)

定理 2.24 (1) X, Y は有限な 2 次の積率を持つ確率変数とする.
このとき $\mathbf{Var}[X + Y] = \mathbf{Var}[X] + \mathbf{Var}[Y] + 2\mathbf{Cov}[X, Y]$ となる.
(2) $d \geq 2$ とする. $X_1, X_2, \dots, X_d$ は有限な 2 次の積率を持つ確率変数とする. このとき

$$\mathbf{Var}\left[\sum_{j=1}^d a_j X_j\right] = \sum_{j=1}^d a_j^2 \mathbf{Var}[X_j] + 2 \sum_{j=1}^d \sum_{\ell=j+1}^d a_j a_\ell \mathbf{Cov}[X_j, X_\ell]$$

となる.

2.3 分散と共分散 (14)

定義 2.25 $X_1, X_2, \dots, X_d$ を有限な 2 次の積率を持つ確率変数とし

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_d \end{pmatrix}$$

と書く. このとき確率ベクトル X の期待値を

$$\mathbf{E}[X] := \begin{pmatrix} \mathbf{E}[X_1] \\ \mathbf{E}[X_2] \\ \vdots \\ \mathbf{E}[X_d] \end{pmatrix}$$

で定義する.

2.3 分散と共分散 (15)

定義 2.25 の続き X の共分散を

$$\mathbf{Var}[X] := \mathbf{E}[(X - \mu)(X - \mu)^T]$$

で定義する¹¹. ただし $\mu = (\mu_1, \mu_2, \dots, \mu_d)^T := \mathbf{E}[X]$ である. これは

$$\mathbf{Var}[X] = \begin{pmatrix} \mathbf{Var}[X_1] & \mathbf{Cov}[X_1, X_2] & \cdots & \mathbf{Cov}[X_1, X_d] \\ \mathbf{Cov}[X_2, X_1] & \mathbf{Var}[X_2] & \cdots & \mathbf{Cov}[X_2, X_d] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Cov}[X_d, X_1] & \mathbf{Cov}[X_d, X_2] & \cdots & \mathbf{Var}[X_d] \end{pmatrix}.$$

¹¹ 確率ベクトルと同様に確率変数を成分とする行列を確率行列という. 確率行列の期待値はそれぞれの成分の期待値を取ったものを配置した行列と定義している.

2.3 分散と共分散 (16)

注意 2.26 定義から $\mathbf{Var}[X]$ は半正定値対称行列となる. なぜならば, 期待値の線型性と補題 2.2(2) から, 任意の $a \in \mathbb{R}^d$ に対して

$$\begin{aligned} a^\top \mathbf{Var}[X] a &= \mathbf{E}[a^\top (X - \mu)(X - \mu)^\top a] \\ &= \mathbf{E}[\{a^\top (X - \mu)\}^2] \geq 0 \end{aligned}$$

となる. 最後の不等号は補題 2.2(2) からわかる. よって $\mathbf{Var}[X]$ は半正定値であることが示せた. $\square$

2.3 分散と共分散 (16)

補題 2.27 $X = (X_1, X_2, \dots, X_d)^\top$ は確率ベクトルで各成分は有限な 2 次の積率を持つとし

$$\mathbf{E}[X] = \mu, \quad \mathbf{Var}[X] = \Sigma$$

とする. ただし $\mu \in \mathbb{R}^d$, Σ は $d \times d$ の半正定値行列¹² である.

(1) 任意の定数ベクトル $a \in \mathbb{R}^d$ に対して

$$\mathbf{E}[a^\top X] = a^\top \mu, \quad \mathbf{Var}[a^\top X] = a^\top \Sigma a$$

となる.

¹² $d \times d$ の対称行列 A が半正定値であるとは, $\forall a \in \mathbb{R}^d$ に対して $a^\top A a \geq 0$ が成立するときをいう. また $d \times d$ の対称行列 A が正定値であるとは,

$\forall a \in \mathbb{R}^d (a \neq 0)$ に対して $a^\top A a > 0$ が成立するときをいう. 

2.3 分散と共分散 (17)

補題 2.27 の続き (2) $k \in \mathbb{N}$ とする. 任意の定数の $k \times d$ 行列 A に対して

$$\mathbf{E}[AX] = A\mathbf{E}[X], \quad \mathbf{Var}[AX] = A\mathbf{Var}[X]A^T$$

となる.

系 2.23 確率ベクトル X は, 任意の定数ベクトル $a(\neq 0)$ に対して

$$\Pr(a^T X = 0) = 0$$

をみたす. このとき, Σ は正定値である.

2.4 条件付き期待値 (1)

定義 2.30 (1) X と Y を確率変数とし, 条件付き p.d.f.(または条件付き p.m.f.) を $\mathbf{p}^{X|Y}$ (または $\mathbf{p}^{X|Y}$) とする. $Y = y$ を与えたときの X の条件付き期待値を

$$\mathbf{E}[X|Y = y] = \begin{cases} \sum_x x \mathbf{p}^{X|Y}(x|y) & (\text{離散型の場合}) \\ \int_{-\infty}^{\infty} x \mathbf{p}^{X|Y}(x|y) dx & (\text{連続型の場合}) \end{cases}$$

で定義する. ただし考えている $Y = y$ で条件付き p.d.f. または p.m.f. は定義され, さらに $\mathbf{E}[|X|] < \infty$ とする.

(2) (Borel 可測) 関数 $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ に対して, $g(X, Y)$ の条件付き期待値を

$$\mathbf{E}[g(X, Y)|Y = y] = \begin{cases} \sum_x g(x, y) \mathbf{p}^{X|Y}(x|y) & (\text{離散型の場合}) \\ \int_{-\infty}^{\infty} g(x, y) \mathbf{p}^{X|Y}(x|y) dx & (\text{連続型の場合}) \end{cases}$$

で定義する. ただし, 考えている $Y = y$ での条件付き p.d.f. または p.m.f. は定義され, $\mathbf{E}[|g(X, Y)|] < \infty$ とする.

注意 2.32 $\mathbf{E}[X]$ は定数であるが, $\mathbf{E}[X|Y = y]$ は一般に y の関数である. このことから

$$h(y) := \mathbf{E}[X|Y = y]$$

とおいたときに $h(y)$ に Y を代入したもの $h(Y)$ は確率変数¹³ になる. これを

$$\mathbf{E}[X|Y] := h(Y)$$

と記すことにする. したがって $\omega \in \Omega$ に対して, $y = Y(\omega)$ と書けば

$$\mathbf{E}[X|Y] : \Omega \ni \omega \mapsto \mathbf{E}[X|Y(\omega)] = \mathbf{E}[X|Y = y] \in \mathbb{R}$$

は可測となる.

測度論的確率論の教科書では, Radon-Nikodym の定理から条件付き期待値を定義する. これから条件付き p.d.f. を定義することになることが知られている. □

¹³ $h \mapsto h(y)$ の可測性は測度論の知識が必要となる. 

2.4 条件付き期待値 (3)

例 2.33 連続型確率変数 Y は p.d.f.

$$p^Y(y) = \begin{cases} 1 & (0 < y < 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

を持つとする. $Y = y$ ($0 < y < 1$) を観測したとき

$$X|Y = y \sim \text{Unif}(y, 1)$$

とする. すなわち $0 < y < 1$ のとき

$$p^{X|Y}(x|y) = \begin{cases} \frac{1}{1-y} & (y < x < 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

である. よって

$$E[X|Y = y] = \int_y^1 x p^{X|Y}(x|y) dx = \frac{1}{1-y} \int_y^1 x dx = \frac{1+y}{2}$$

となる. これより

$$E[X|Y] = \frac{1+Y}{2}$$

2.4 条件付き期待値 (4)

定理 2.34 (1) 有限な期待値を持つ確率変数 X, Y と確率変数 Z に対して

$$\mathbf{E}[X + Y | Z] = \mathbf{E}[X | Z] + \mathbf{E}[Y | Z]$$

となる.

(2) 有限な期待値を持つ確率変数 X, Y に対して

$$\mathbf{E}[\mathbf{E}[Y | X]] = \mathbf{E}[Y], \quad \mathbf{E}[\mathbf{E}[X | Y]] = \mathbf{E}[X]$$

となる.

(3) 一般の (Borel 可測) 関数 $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ を考える.

$\mathbf{E}[|g(X, Y)|] < \infty$ のとき

$$\mathbf{E}[\mathbf{E}[g(X, Y) | Y]] = \mathbf{E}[g(X, Y)]$$

となる.

(4) $\mathbf{E}[XY | Y] = Y\mathbf{E}[X | Y]$ となる.

例 2.30

$$\mathbf{E}[X] = \mathbf{E}\left[\frac{1+Y}{2}\right] = \frac{3}{4}$$

となる. 一方 $0 < y < x < 1$ に対して

$$\mathbf{p}(x, y) = \mathbf{p}^{X|Y}(x|y)\mathbf{p}^Y(y) = \frac{1}{1-y}$$

となる. よって

$$\begin{aligned}\mathbf{E}[X] &= \int_0^1 \left\{ \int_y^1 x \frac{1}{1-y} dx \right\} dy \\ &= \int_0^1 \frac{1}{1-y} \left[\frac{x^2}{2} \right]_y^1 dy \\ &= \int_0^1 \frac{1}{1-y} \frac{1-y^2}{2} dy = \left[\frac{(1+y)^2}{4} \right]_0^1 = \frac{3}{4}.\end{aligned}$$

2.5 積率母関数 (1)

定義 2.38 X を確率変数とし, ある $t_0 > 0$ が存在して,
 $\mathbf{E}[e^{tX}] < \infty$ ($\forall |t| < t_0$) とする. このとき, X の積率母関数 (Moment
Generating Function (m.g.f.)) を

$$\mathbf{m}^X(t) := \mathbf{E}[e^{tX}] \quad (-t_0 < t < t_0)$$

と定義する.

2.5 積率母関数 (2)

注意 2.39 確率変数 X の積率母関数 $\mathbf{m}^X(t)$ が存在するとき, 期待値と微分の記号の入れ替えが保証されること¹⁴が知られている. このことから

$$\begin{aligned}\dot{\mathbf{m}}^X(0) &= \dot{\mathbf{m}}^X(t)\Big|_{t=0} = \left[\frac{d}{dt}\mathbf{m}^X(t)\right]\Big|_{t=0} = \left[\frac{d}{dt}\mathbf{E}[e^{tX}]\right]\Big|_{t=0} = \mathbf{E}\left[\frac{d}{dt}e^{tX}\right]\Big|_{t=0} \\ &= \mathbf{E}[Xe^{tX}]\Big|_{t=0} = \mathbf{E}[X]\end{aligned}$$

となる. この議論を繰り返せば $k = 2, 3, \dots$ に対して

$$\{\mathbf{m}^X\}^{(k)}(0) = \mathbf{E}[X^k]$$

がわかる.

□

¹⁴たとえば, 清水 (2021) を参照のこと.

2.5 積率母関数 (3)

例 2.40 $X \sim \mathbf{Exp}(1)$ とする. $t < 1$ に対して

$$m^X(t) = E[e^{tX}] = \int_0^\infty e^{tx} e^{-x} dx = \int_0^\infty e^{-(1-t)x} dx = \frac{1}{1-t}$$

となる. $t \geq 1$ のときは, e^{tX} の期待値は発散する. したがって

$$m^X(t) = \frac{1}{1-t} \quad (t < 1)$$

となる. 簡単な計算から

$$\dot{m}^X(0) = 1, \quad \ddot{m}^X(0) = 2$$

なので

$$E[X] = 1, \quad E[X^2] = 2, \quad \text{Var}[X] = E[X^2] - \{E[X]\}^2 = 1$$

となる.

□

2.5 積率母関数 (4)

補題 2.41 (1) $a, b \in \mathbb{R} (a \neq 0)$ とする. $Y = aX + b$ としたとき

$$m^Y(t) = e^{tb} m^X(at)$$

となる.

(2) $X_1, X_2, \dots, X_d$ は独立とし $Y = \sum_{j=1}^d X_j$ とする. このとき

$$m^Y(t) = \prod_{j=1}^d m^{X_j}(t)$$

となる.

本日の講義のまとめ

- 前回の講義の記号の復習
- 1.5.1 同時確率関数 (p.m.f.) と同時密度関数 (p.d.f.)
- 1.5.2 周辺分布
- 1.5.3 (確率的) 独立性と条件付き分布
- 1.6 多次元分布と i.i.d. 標本
i.i.d.=identically and independently distributed
- 2.1 期待値
- 2.2 確率ベクトルの期待値
- 2.3 分散と共分散
- 2.4 条件付き期待値
- 2.5 積率母関数