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

2025 年 10 月 4 日

講義内容

- 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)

- $(\Omega, \mathcal{A}, \mathrm{Pr})$: 確率空間.
  - $\mathcal{A}$ は σ 加法族.
  - $\mathrm{Pr} : \mathcal{A} \longrightarrow [0, 1]$ 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) $A_n \in \mathcal{A}\,(n = 1, 2, \ldots) \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.
- Pr; $\mathcal{A} \ni A \mapsto \mathrm{Pr}(A) \in [0, 1]$ は以下をみたす.
  - (1) $\mathrm{Pr}(\Omega) = 1$;
  - (2) $\forall A \in \mathcal{A}$ に対して, $\mathrm{Pr}(A^c) = 1 - \mathrm{Pr}(A)$;
  - (3) <u>互いに排反な</u> $\{A_n\}_{n=1}^{\infty} \subset \mathcal{A}$ に対して

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

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

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

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

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

  $\Leftrightarrow$ $\forall r \in \mathbb{R}$ 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.

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

- 分布関数 F^X の性質
  - F^X は非減少
  - F^X は右連続.
  - $\lim_{x \to -\infty} \mathsf{F}^X(x) = 0$; $\lim_{x \to \infty} \mathsf{F}^X(x) = 1$.
  - 不連続点は高々加算個.

記号の復習 (4)

- $\mathsf{F}^X(\cdot)$ 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.

$$\mathsf{p}^X(x) = \mathsf{F}^X(x) - \mathsf{F}^X(x-).$$

ただし, $\mathsf{F}^X(x-) = \lim_{y \to x-0} \mathsf{F}^X(y)$.

- $\mathsf{F}^X(\cdot)$ 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.: 非負値関数 p^X で

$$\mathsf{F}^X(x) = \int_{-\infty}^{x} \mathsf{p}^X(t)\,\mathrm{d}t.$$

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 $(\Omega, \mathscr{A}, \mathrm{Pr})$ 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.) $\mathbf{p}(\cdot)$ 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

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

である.

2 次元の分布 (2)

定義 1.45 確率空間 $(\Omega, \mathcal{A}, \mathrm{Pr})$ 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 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) $\mathrm{p}(x, y) \geq 0 \ (\forall (x, y) \in \mathbb{R}^2)$.

(2) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathrm{p}(x, y)\, dxdy = 1$.

(3) $\forall A \in \mathcal{B}(\mathbb{R}^2)$ に対して

$$\mathrm{Pr}((X, Y) \in A) = \int\int_A \mathrm{p}(x, y)\, dxdy.$$

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.) F を

$$\mathsf{F}(x, y) := \Pr(X \le x, Y \le y) \tag{1}$$
$$:= \Pr(\{\omega \in \Omega;\ X(\omega) \le x, Y(\omega) \le y\}) \qquad (\forall (x, y) \in \mathbb{R}^2)$$

で定義する. (X, Y) の同時分布 P を

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

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かを明示したいときには, $\mathsf{F}^{(X,Y)}$ または $\mathsf{P}^{(X,Y)}$ と記す.
細かなことであるが, (??) の F によって $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ 上の確率測度 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. □

例 1.47 連続型確率変数のベクトル (X, Y) は同時 p.d.f.

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

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

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

となる. □

1.5.2 周辺分布 (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)

$\boxed{\text{定義 1.48}}$ (X, Y) を確率空間 $(\Omega, \mathscr{A}, \mathrm{Pr})$ 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. p を持つとする. X の周辺確率関数 (周辺 p.m.f.) を

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

で定義する. ただし

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

である.

$\boxed{\text{定義 1.48 の続き}}$ (2) (X, Y) は連続型とし, 同時 p.d.f. 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

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

で定義し, Y の周辺確率密度関数 (周辺 p.d.f.) を

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

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

$$\mathsf{F}^X(x) := \mathsf{Pr}(X \le x) = \iint_{(s,\,t)\in\mathbb{R}^2;\ s\le x} \mathsf{p}(s,\ t)\,\mathrm{d}s\mathrm{d}t$$

$$= \int_{-\infty}^{x} \left\{ \int_{-\infty}^{\infty} \mathsf{p}(s,\ t)\,\mathrm{d}s \right\} \mathrm{d}t = \int_{-\infty}^{x} \mathsf{p}^X(s)\,\mathrm{d}s$$

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 $(\Omega, \mathcal{A}, \mathrm{Pr})$ 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

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

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.

定理 1.52 $\mathbf{p}$ を確率ベクトル (X, Y) の同時 p.d.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. 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. □

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 **0** の集合を除いて [1]p.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.
(2) 離散型確率変数の場合には, p.d.f. を p.m.f. に置き換えればよい. □

[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.

例 1.54 連続型確率変数 X と Y は独立で同時 p.d.f.

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

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

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

となる.

1.5.3 (確率的) 独立性と条件付き分布 (5)

$\boxed{\text{例 1.54 の続き}}$ したがって

$$\mathrm{Pr}(X + Y \le 1) = \int\int_{x+y\le 1} \mathrm{p}^{(X,Y)}(x, y)\,\mathrm{d}x\mathrm{d}y = 4\int_0^1 x\left\{\int_0^{1-x} y\,\mathrm{d}y\right\}\mathrm{d}x$$
$$= 4\int_0^1 x\frac{(1-x)^2}{2}\,\mathrm{d}x = \frac{1}{6}$$

となる. □

定義 1.58 (1) 離散型確率ベクトル (X, Y) は同時 p.m.f. $\mathsf{p}(x, y)$ を持つとする. $\mathsf{p}_Y(y) > 0$ 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

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

で定義する.

1.5.3 (確率的) 独立性と条件付き分布 (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. $\mathsf{p}(x, y)$ を持つとする. $\mathsf{p}^Y(y) > 0$ 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

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

で定義する.

注意 1.59 (X, Y) が連続型確率ベクトルのとき, $\mathrm{p}^Y(y) > 0$ 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

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

と形式的に定義する. (1) $A, B \in \mathcal{A}$ とし, $\mathrm{Pr}(B) > 0$ とする. A と B が独立のとき

$$\mathrm{Pr}(A \,|\, B) = \mathrm{Pr}(A)$$

が成立する.

1.5.3 (確率的) 独立性と条件付き分布 (9)

注意 1.59 の続き (2) $\mathrm{Pr}(B) > 0$ のとき

$$\mathrm{Pr}(A \cap B) = \mathrm{Pr}(B)\mathrm{Pr}(A | B)$$

となる. これを乗法の公式という.
(3) $\mathrm{Pr}(B) > 0$ のとき, 関数 $\mathrm{Pr}(\cdot | B) : \mathscr{A} \to [0, 1]$ は定義 1.6(1) – (2) をみたす. すなわち $(\Omega, \mathscr{A})$ 上の確率測度である.

注意 1.59 の続き (4) $\Omega = \{b_1, b_2, b_3, b_4\}$ とし

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

とする. いま

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

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

$$\Pr(A_1 \cap A_2) = \Pr(A_1)\Pr(A_2); \ \Pr(A_1 \cap A_3) = \Pr(A_1)\Pr(A_3);$$
$$\Pr(A_2 \cap A_3) = \Pr(A_2)\Pr(A_3).$$

1.5.3 (確率的) 独立性と条件付き分布 (11)

$\boxed{\text{注意 1.59 の続き}}$ しかし

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

である. □

例 1.60 連続型確率ベクトル (X, Y) は同時 p.d.f.

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

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

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

例 1.60 の続き したがって

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

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

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

となる.

例 1.60 の続き　これより $0 < y < 1$ のとき

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

となる. 注意 1.59 より

$$\Pr\left(X \leq \frac{1}{4}\middle| Y = \frac{1}{3}\right) = \int_0^{1/4} p^{X|Y}\left(x\middle|\frac{1}{3}\right)\mathrm{d}x = \int_0^{1/4} \frac{x+\frac{1}{3}}{\frac{1}{3}+\frac{1}{2}}\mathrm{d}x = \frac{11}{80}$$

□

1.6 多次元の分布と i.i.d. 標本 (1)

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

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

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

$\boxed{\text{注意 1.61}}$ 本講義録では, ベクトルは縦ベクトルとする. □

□

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

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

が成立するときである. □

□

注意 1.63 X の同時 p.d.f. を $\mathsf{p}(x_1, x_2, \ldots, x_n)$ と書き, 各 X_j $(j = 1, 2, \ldots, n)$ の周辺 p.d.f. を p^{X_j} と書くことにする. (??) を示すためには

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

を示せばよい. □

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

$$X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{F}$$

と書く[2]. $X_1, X_2, \ldots, X_n$ は累積分布関数 F からの標本の大きさが n のランダム標本ともいう. □

[2] p.d.f. p を使い

$$X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{p}$$

と書く. ~ の右側には, 確率測度/確率分布/p.d.f./p.m.f./$\mathbf{N}(0, 1)$ 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 $(\Omega, \mathcal{A}, \mathrm{Pr})$ 上の確率変数とする. X が離散型のとき, その p.m.f. を $\mathbf{p}$ とし, X の取り得る値を $x_1, x_2, \ldots$ とする. 連続型のとき, その p.d.f. も $\mathbf{p}$ と書くことにする.

定義 2.1 $g:\mathbb{R}\to\mathbb{R}$ を可測[3] 関数とする. 確率変数 $g(X)$ の期待値 $\mathrm{E}[g(X)]$ を次のように定義する.

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

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

と定義する. 右辺は ∞ を許せば, 必ず存在する.

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

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. $\mathsf{E}[g^+(X)]$ または $\mathsf{E}[g^-(X)]$ のいずれかが有限ならば

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

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

□

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

(1) $h(X)$ は有限の期待値を持つ[4] とする.
$0 \le g(x) \le h(x)\ (\forall x\in\mathbb{R})$ ならば

$$\mathsf{E}[g(X)] \le \mathsf{E}[h(X)].$$

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

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

[4] $g,\ h$ を共に非負値関数としているので, $\mathsf{E}[|h(X)|]<\infty$ だけを仮定すればよい. もし, $g,\ h$ を可測関数とすると $\mathsf{E}[|g(X)|]<\infty$ の仮定も必要になる. なぜならば, $g(x)\le h(x)\ (x\in\mathbb{R})$ と $\mathsf{E}[|h(X)|]<\infty$ だけだと $\mathsf{E}[h(X)]$ 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のが実用的である.

補題 2.3 X を非負値確率変数とする. このとき,

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

となる.

証明 まず

$$A := \{\omega \in \Omega;\ X(\omega) > 0\} = \{X > 0\},$$

$$A_n := \left\{\omega \in \Omega;\ X(\omega) > \frac{1}{n}\right\} \qquad (n = 1, 2, \ldots)$$

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

2.1 期待値 (6)

$\boxed{\text{証明の続き}}$ よって, 補題 1.7(5) から

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

一方, $A_n \subset A$ なので, $\mathbb{1}_{A_n}(\omega) \le \mathbb{1}_A(\omega)\,(x \ge 0)\,(\omega \in \Omega)$ である.

2.1 期待値 (7)

$\boxed{\text{証明の続き}}$ このことに注意すると

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

2.1 期待値 (8)

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

$$
\begin{aligned}
1 &= \mathrm{Pr}(A^c) = \mathrm{Pr}\Big((\{X \le 0\} \cap \{X \ge 0\}) \cup (\{X \le 0\} \cap \{X < 0\})\Big) \\
&= \mathrm{Pr}(X = 0) + \underbrace{\mathrm{Pr}(X < 0)}_{=0} = \mathrm{Pr}(X = 0)
\end{aligned}
$$

が示せた. □

注意 2.4 $0 < q < r$ に対して

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

となる.

証明 Young の不等式

$$ab \le \frac{a^s}{s} + \frac{b^t}{t}; \quad a,\ b > 0; \quad s,\ t > 0; \quad \frac{1}{t} + \frac{1}{s} = 1$$

において

$$a = |x|^q;\ s = \frac{r}{q}\ (0 < s < 1) \quad b = 1;\ t = \frac{1}{1 - s/2}\ (0 < t < 1)$$

とおくと

2.1 期待値 (10)

証明の続き

$$|x|^q \le \frac{|x|^r}{s} + \frac{1}{t}$$

を得る. よって, 補題 2.2(1) からわかる. $\square$

定義 2.5

(1) $k = 1, 2, \ldots$ に対して, $\mathsf{E}[|X|^k] < \infty$ のとき, $\mathsf{E}[X^k]$ を X の k 次モーメント (または積率) という.

(2) $\mathsf{E}[|X|] < \infty$ のとき $\mathsf{E}[X]$ を X の平均値 [5] という.

(3) $\mathsf{E}[X^2] < \infty$ のとき X の分散を

$$\mathrm{Var}[X] := \mathsf{E}[\{X - \mathsf{E}[X]\}^2]$$

で定義する.

[5] 簡単に「平均」ともいう.

定義 2.5 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[6] という. $A \in \mathcal{B}(\mathbb{R})$ のとき

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

となる.

□

[6]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.

$\boxed{\text{注意 2.6}}$ 分散 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. □

$\boxed{\text{注意 2.7}}$ (1) X は確率空間 $(\Omega, \mathscr{A}, \mathrm{Pr})$ 上の離散型確率変数とする. p^X を X の p.m.f. とし

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

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

$$Y = g(X)$$

とおく.

注意 2.7 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

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

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

注意 2.7 の続き

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

となる. さらにいずれかの和が有限ならば

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

となる.

注意 2.7 の続き よって

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

となる.
(2) 確率変数 X は連続型とする. p^X を X の p.d.f. とし

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

とする. 関数 $g : S_X \to \mathbb{R}$ は狭義単調増加かつ C^1 級とする. 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.

注意 2.7 の続き　ここで $g: S_X \to S_Y$ と制限[7] すれば g の逆関数 $g^{-1}: S_Y \to S_X$ が存在し

$$\mathrm{F}^Y(y) := \Pr(Y \le y) = \Pr(g^{-1}(Y) \le g^{-1}(y)) = \Pr(X \le g^{-1}(y)) \\ = \mathrm{F}^X(g^{-1}(y)).$$

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

[7] 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.

注意 2.7 の続き　これらより, $y \in S_Y$ に対して

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

となる. よって

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

となる.

注意 2.7 の続き　g が狭義単調減少の場合もふくめると

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

となる. 以上のことから

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

□

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 2.8 可測関数 $g: \mathbb{R}^2 \to \mathbb{R}$ 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

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

で定義する.

定義 2.8 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. $\mathrm{E}[g^+(X, Y)]$ または $\mathrm{E}[g^-(X, Y)]$ のいずれかが有限ならば

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

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

2.2 確率ベクトルの期待値 (4)

注意 2.9 3 つ以上の確率変数 $X_1, X_2, \ldots, X_n$ に対しても期待値を定義 2.8 と同様に定義する.

定理 2.10 $X_1, X_2, \ldots, X_n$ を確率変数とし, 各 X_j $(j = 1, 2, \ldots, n)$ の期待値は有限とする. $a_1, a_2, \ldots, a_n$ を定数としたとき

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

となる.

証明　まず, $\left|\sum_{j=1}^{n} a_j x_j\right| \leq \sum_{j=1}^{n} a_j |x_j|$ 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. $X := (X_1, X_2, \ldots, X_n)^\top$ とおき, X は同時 p.d.f. 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

$$\mathsf{E}\Big[\sum_{j=1}^{n} a_j X_j\Big] = \int_{\mathbb{R}^n} \Big\{\sum_{j=1}^{n} a_j x_j\Big\} \mathsf{p}^X(x_1, x_2, \ldots, x_n)\, \mathrm{d}x_1 \mathrm{d}x_2 \cdots \mathrm{d}x_n$$

$$= \sum_{j=1}^{n} a_j \int_{\mathbb{R}^n} x_j \mathsf{p}^X(x_1, x_2, \ldots, x_n)\, \mathrm{d}x_1 \mathrm{d}x_2 \cdots \mathrm{d}x_n = \sum_{j=1}^{n} a_j \mathsf{E}[X_j].$$

□

例 2.11 $0 < p < 1$ と $j = 1, 2, \ldots, n$ に対して, $X_j \sim \mathrm{Ber}(p)$ とする. このとき

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

である. $S = \sum_{j=1}^{n} X_j$ としたとき

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

がわかる. □

2.3 分散と共分散 (1)

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

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

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

定理 2.12 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) $\mathsf{Var}[X] = \mathsf{E}[X^2] - \{\mathsf{E}[X]\}^2$ となる.

(2) 定数 $a, b \in \mathbb{R}$ に対して

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

となる.

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

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

となる.

2.3 分散と共分散 (3)

定理 2.12 の続き

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

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

となる.

例 2.13 $0 < p < 1$ と $j = 1, 2, \ldots, n$ に対して, $X_j \sim \mathsf{Ber}(p)$ とし, $X_1, X_2, \ldots, X_n$ は互いに独立とする. すると定理 2.12 に注意すれば

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

となる. これと定理 2.12(4) から

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

□

定理 2.14 $n \geq 2$ とし, $X_1, X_2, \ldots, X_n$ は i.i.d. 確率変数列とし

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

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

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

で定義し, 標本 (不偏) 分散を

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

で定義する.

定理 2.14 の続き　このとき

$$(1)\ \mathrm{E}[\overline{X}_n] = \mu; \qquad (2)\ \mathrm{Var}[\overline{X}_n] = \frac{\sigma^2}{n}; \qquad (3)\ \mathrm{E}[S_n^2] = \sigma^2$$

となる.

$\boxed{\text{証明}}$ (1), (2) は定理 ??, ??(4) よりわかる. (3) を証明するために

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

に注意する. 定理 ??(1) より

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

がわかる. さらに, (1) と (2) より

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

2.3 分散と共分散 (8)

証明の続き　したがって

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

がわかる. □

定義 2.15 X と Y は確率変数とし

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

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

$$\mathrm{Cov}[X, Y] := \mathsf{E}[(X - \mu_X)(Y - \mu_Y)],$$

$$\rho := \rho[X, Y] := \frac{\mathrm{Cov}[X, Y]}{\sigma_X \sigma_Y}$$

で定義する.

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

2.3 分散と共分散 (10)

定理 2.16 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

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

と書き直せる.
(2) $\mathrm{Cov}[X, Y] = \mathrm{Cov}[Y, X]$ である.
(3) 定数 a, b に対して

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

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)

$\boxed{\text{定理 2.16 の続き}}$ (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

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

となる.

2.3 分散と共分散 (12)

注意 2.17 $\mathsf{E}[X] = \mathsf{E}[Y] = \mathbf{0}$ の場合を考えると Cauchy-Schwarz の不等式

$$\left|\mathsf{E}[XY]\right| \leq \sqrt{\mathsf{E}[X^2]}\sqrt{\mathsf{E}[Y^2]}$$

と呼ばれる重要なものを得る. □

注意 2.18 定理 **2.16(6)** の逆は一般に正しくない.

2.3 分散と共分散 (13)

定理 2.19 (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 $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}[X, Y]$ となる．
(2) $d \geq 2$ とする．$X_1, X_2, , \ldots, X_d$ は有限な 2 次の積率を持つ確率変数とする．このとき

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

となる．

定義 2.20 $X_1, X_2, \ldots, X_d$ 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

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

で定義する.

定義 2.20 の続き　X の共分散を

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

で定義する[8]. ただし $\mu = (\mu_1, \mu_2, \ldots, \mu_d)^\top := \mathrm{E}[X]$ である. これは

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

[8]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.

注意 2.21 定義から $\mathrm{Var}[X]$ 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) から, 任意の $a \in \mathbb{R}^d$ に対して

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

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

補題 2.22 $X = (X_1, X_2, \ldots, X_d)^\top$ 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

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

とする. ただし $\mu \in \mathbb{R}^d$, Σ は $d \times d$ の半正定値行列[9] である.
(1) 任意の定数ベクトル $a \in \mathbb{R}^d$ に対して

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

となる.

[9] $d \times d$ 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 $a^\top \mathrm{A} a \geq 0$ 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 $a^\top \mathrm{A} a > 0$ が成立するときをいう.

補題 2.22 の続き (2) $k \in \mathbb{N}$ とする. 任意の定数の $k \times d$ 行列 $\mathbf{A}$ に対して

$$\mathsf{E}[\mathbf{A}X] = \mathbf{A}\mathsf{E}[X], \quad \mathsf{Var}[\mathbf{A}X] = \mathbf{A}\mathsf{Var}[X]\mathbf{A}^\top$$

となる.

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

$$\mathsf{Pr}(a^\top X = 0) = 0$$

をみたす. このとき, Σ は正定値である.

定義 2.25 (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 $\mathsf{p}^{X|Y}$(または $\mathsf{p}^{X|Y}$) とする. $Y = y$ を与えたときの X の条件付き期待値を

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

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 $\mathsf{E}[|X|] < \infty$ とする.
(2) (Borel 可測) 関数 $g : \mathbb{R}^2 \to \mathbb{R}$ に対して, $g(X, Y)$ の条件付き期待値を

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

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, $\mathsf{E}[|g(X,Y)|] < \infty$ とする.

注意 2.27 $\mathsf{E}[X]$ は定数であるが, $\mathsf{E}[X|\,Y = y]$ は一般に y の関数である. このことから

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

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[10]になる. これを

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

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

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

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. $\square$

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

例 2.28 連続型確率変数 Y は p.d.f.

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

を持つとする. $Y = y\ (0 < y < 1)$ を観測したとき

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

とする. すなわち $0 < y < 1$ のとき

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

である. よって

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

となる. これより

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

定理 2.29 (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

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

となる.
(2) 有限な期待値を持つ確率変数 X, Y に対して

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

となる.
(3) 一般の (Borel 可測) 関数 $g: \mathbb{R}^2 \to \mathbb{R}$ を考える.
$\mathsf{E}[|g(X,Y)|] < \infty$ のとき

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

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

例 2.30

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

となる. 一方 $0 < y < x < 1$ に対して

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

となる. よって

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

2.5 積率母関数 (1)

$\boxed{\text{定義 2.33}}$ X を確率変数とし, ある $t_0 > 0$ が存在して, $\mathrm{E}[e^{tX}] < \infty$ $(\forall |t| < t_0)$ とする. このとき, X の積率母関数 (Moment Generating Function (m.g.f.)) を

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

と定義する.

注意 2.34 確率変数 X の積率母関数 $\mathsf{m}^X(t)$ 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 [11] が知られている. このことから

$$\dot{\mathsf{m}}^X(0) = \dot{\mathsf{m}}^X(t)\Big|_{t=0} = \left[\frac{\mathrm{d}}{\mathrm{d}t} m^X(t)\right]\Bigg|_{t=0} = \left[\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{E}[e^{tX}]\right]\Bigg|_{t=0} = \mathsf{E}\left[\frac{\mathrm{d}}{\mathrm{d}t}e^{tX}\right]\Bigg|_{t=0}$$
$$= \mathsf{E}[Xe^{tX}]\Big|_{t=0} = \mathsf{E}[X]$$

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

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

がわかる. □

[11] たとえば, [?, pp.75-76] を参照のこと.

例 2.35 $X \sim \mathrm{Exp}(1)$ とする. $t < 1$ に対して

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

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

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

となる. 簡単な計算から

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

なので

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

となる. □

2.5 積率母関数 (4)

補題 2.36 (1) $a, b \in \mathbb{R}\,(a \neq 0)$ とする. $Y = aX + b$ としたとき

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

となる.

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

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

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