［先週の 28, 29 ページはこちらに換えて下さい］

2009-7-15

2-8. 同時分布 joint distribution

$(\Omega, \mathcal{F}, P)$ を確率空間とする

∘ ベクトル値確率変数（確率変数の組）
vector valued r.v.

$$\begin{array}{ccc} \Omega & \longrightarrow & \mathbb{R}^2 \\ \cup\!\!\!\shortmid & & \cup\!\!\!\shortmid \\ \omega & \longmapsto & (X(\omega), Y(\omega)) \end{array}$$

に対して

□ 同時分布関数

$$F_{XY}(x, y) := P(\{\omega \in \Omega \mid X(\omega) \le x, Y(\omega) \le y\})$$

□ 同時確率関数（離散型 r.v.）

$$P_{XY}(x, y) := P(\{\omega \in \Omega \mid X(\omega) = x, Y(\omega) = y\})$$

□ 同時確率密度関数（連続型 r.v.）

$$p(x, y) := \frac{\partial F_{XY}(x, y)}{\partial x \partial y}$$

が定義される

∘ 同時密度関数の意味

21 ページと同様に

□ $\Pr\{\omega \in \Omega \mid a < X(\omega) \le b, c < Y(\omega) \le d\}$

$$= \int_c^d \int_a^b p(x, y)\, dx\, dy$$

□ $\Pr\{\omega \in \Omega \mid x < X(\omega) \le x + \Delta x, y < Y(\omega) \le y + \Delta y\}$

$$\approx p(x, y)\, \Delta x\, \Delta y$$

（Δx, Δy が十分小さいとき）

∘ 3つ以上の確率変数についても同様

○ これ以降 主に 離散型 r.v. を扱う
基本的な考え方は 連続型でも同様

□ $\mathcal{X} := \{ X(\omega) \mid \omega \in \Omega \}$ 　X のとる値の集合
$\mathcal{Y} := \{ Y(\omega) \mid \omega \in \Omega \}$ 　Y 〃

□ $A_x := \{ \omega \in \Omega \mid X(\omega) = x \} \subset \Omega$
$X = x$ となる事象
$B_y := \{ \omega \in \Omega \mid Y(\omega) = y \} \subset \Omega$
$Y = y$ となる事象

○ これまでに，一変数確率関数
$P_X(x) = P(\{ \omega \in \Omega \mid X(\omega) = x \}) = P(A_x)$
$P_Y(y) = P(\{ \omega \in \Omega \mid Y(\omega) = y \}) = P(B_y)$
は習った，　$P_{XY}(x, y) = P(A_x \cap B_y)$
との関係は？

<u>Lem</u>

$$P_X(x) = \sum_{y \in \mathcal{Y}} P_{XY}(x, y) \quad \text{—— ①}$$

$$P_Y(y) = \sum_{x \in \mathcal{X}} P_{XY}(x, y) \quad \text{—— ②}$$

(証明)　$y_1, y_2 \in \mathcal{Y}$ について

$y_1 \neq y_2 \Rightarrow B_{y_1} \cap B_{y_2} = \emptyset$ 　となる

さらに　$\Omega = \bigcup_{y \in \mathcal{Y}} B_y$ 　（互いに素な分割）

よって，全確率の公式（8ページ）により

$$P(A_x) = \sum_{y \in \mathcal{Y}} P(A_x \cap B_y)$$

これは ① である． ② も同様

//

○ 用語の定義　terminology

$$P_{XY}(x, y) \longleftrightarrow P_X(x),\ P_Y(y)$$

同時確率関数　joint probability function　　周辺確率関数　marginal probability function

(同時分布 joint distribution)　　(周辺分布 marginal distribution)

$\boxed{\text{例}}$　$\mathcal{X} = \{1, 2\}$, $\mathcal{Y} = \{1, 2, 3\}$

X \ Y	1	2	3	
1	$\frac{2}{15}$	$\frac{3}{15}$	$\frac{4}{15}$	$\frac{9}{15}$
2	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{3}{15}$	$\frac{6}{15}$
	$\frac{3}{15}$	$\frac{5}{15}$	$\frac{7}{15}$	

$P_{XY}(x, y)$

$P_X(x)$

$P_Y(y)$

Def (条件付き確率関数)　conditional probability function

- $P_X(x) > 0$ のとき

$$P_{Y|X}(y|x) := \frac{P_{XY}(x, y)}{P_X(x)} = \frac{P(A_x \cap B_y)}{P(A_x)}$$

事象 $X = x$ が起こったもとで，事象 $Y = y$ が起きる確率

- Remark
  - □ 定義より　$P_{XY}(x, y) = P_X(x) P_{Y|X}(y|x)$
    chain rule
  - □ $P_{Y|X}(y|x)$ は x を固定すると確率関数
    すなわち　$\displaystyle\sum_{y \in \mathcal{Y}} \underset{\substack{\text{IV}\\ 0}}{P_{Y|X}(y|x)} = 1$

$\boxed{\text{例}}$ 続き

Y	1	2	3	
$P_{Y\mid X}(y\mid 1)$	$\frac{2}{9}$	$\frac{3}{9}$	$\frac{4}{9}$	← たすと1
$P_{Y\mid X}(y\mid 2)$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{3}{6}$	← たすと1

X	$P(x\mid 1)$	$P(x\mid 2)$	$P(x\mid 3)$
1	$\frac{1}{3}$	$\frac{3}{5}$	$\frac{4}{7}$
2	$\frac{2}{3}$	$\frac{2}{5}$	$\frac{3}{7}$
	↑	↑	↑

たすと1

Def（確率変数の独立性） independence of r.v.

○ X, Y が独立

$$\overset{\text{def}}{\iff} \quad \forall x \in \mathcal{X},\ \forall y \in \mathcal{Y},\quad P_{XY}(x, y) = P_X(x) P_Y(y)$$

○ 同様に

$X_1, X_2, \cdots, X_n$ が独立

$$\overset{\text{def}}{\iff} \quad \forall x_1, \forall x_2, \cdots, \forall x_n,$$

$$P_{X_1 X_2 \cdots X_n}(x_1, x_2, \cdots, x_n) = P_{X_1}(x_1) P_{X_2}(x_2) \cdots P_{X_n}(x_n)$$

2-9. 期待値と分散（続き）

○ $\boxed{\text{確率変数の関数は確率変数}}$

X が r.v. → $Y = f(X)$ は r.v.

$$\Omega \ni \omega \xrightarrow{X} X(\omega) \xrightarrow{f} Y(\omega) = f(X(\omega))$$

$Y = f \circ X$ （合成関数）

Lem $Y = f(X)$ のとき

$$E[Y] = \sum_{x \in \mathcal{X}} P_X(x) f(x) \quad \text{—— ①}$$

$\uparrow$

定義は $E[Y] = \sum_{y \in \mathcal{Y}} P_Y(y)\, y$

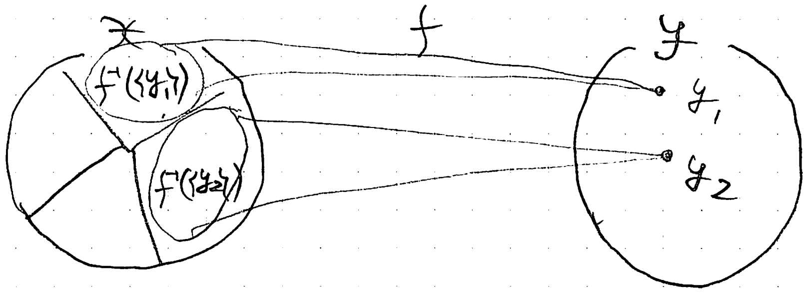
(証明)

$f: \mathcal{X} \to \mathcal{Y}$ で y の逆像

$$f^{-1}(\{y\}) = \{x \in \mathcal{X} \mid f(x) = y\}$$

は $\mathcal{X}$ の分割を与える．すなわち

$$\mathcal{X} = \bigcup_{y \in \mathcal{Y}} f^{-1}(\{y\}) \quad (\text{互いに素})$$



cf. $f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B)$

$A = \{y_1\}$, $B = \{y_2\}$　$y_1 \neq y_2$ とすると

$$\phi = f^{-1}(\underbrace{\{y_1\} \cap \{y_2\}}_{\phi}) = f^{-1}(\{y_1\}) \cap f^{-1}(\{y_2\})$$

よって ① 右辺

$$= \sum_{y \in \mathcal{Y}} \sum_{x \in f^{-1}(\{y\})} P_X(x) \underbrace{f(x)}_{y}$$

$$= \sum_{y} y \underbrace{\sum_{x:\, y = f(x)} P_X(x)}_{P_Y(y)}$$

$$= E[Y]$$

//

Lem （期待値の性質）

(1) $$E[aX+bY] = aE[X]+bE[Y] \quad a,b\in\mathbb{R}$$

期待値操作は線形 linear operation

特に $Y=1$（定数関数）とおくと

$$E[aX+b] = aE[X]+b$$

（X,Y が独立でなくても成立）

(2) X,Y が独立のとき

$$E[XY] = E[X]E[Y]$$

（証明）

(1) $f(X,Y) = aX+bY$ として 32ページ Lem を用いると

$$E[aX+bY] = \sum_x\sum_y p(x,y)(ax+by)$$

$$= a\sum_x \underbrace{\sum_y p(x,y)}_{=p(x)}\, x + b\sum_y \underbrace{\sum_x p(x,y)}_{=p(y)}\, y$$

$$= aE[X]+bE[Y]$$

(2) レポート

//

Lem（分散の性質）

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

(1) $V[X] \geq 0$

(2) $V[aX+b] = a^2 V[X]$ 　　$a, b \in \mathbb{R}$

(3) X と Y が独立なとき

$V[X+Y] = V[X] + V[Y]$

(4) $V[X] = E[X^2] - E[X]^2$ （レポート）

（証明）

(1) $V[X] = E[(X-\mu)^2] = \sum_x \underset{\underset{0}{\rotatebox{90}{$\leq$}}}{P(x)} \underset{\underset{0}{\rotatebox{90}{$\leq$}}}{(x-\mu)^2} \geq 0$

(2) $V[aX+b] = E[\{(aX+b) - (a\mu+b)\}^2]$

$\because$ $E[aX+b] = a\mu+b$

$= E[a^2(X-\mu)^2]$

$= a^2 E[(X-\mu)^2] = a^2 V[X]$

(3) $\mu_X = E[X]$，$\mu_Y = E[Y]$ とおくと

$E[X+Y] = \mu_X + \mu_Y$

$V[X+Y] = E[\{(X+Y) - (\mu_X+\mu_Y)\}^2]$

$= E[\underbrace{\{(X-\mu_X) + (Y-\mu_Y)\}^2}_{(X-\mu_X)^2 + 2(X-\mu_X)(Y-\mu_Y) + (Y-\mu_Y)^2}]$

$= E[(X-\mu_X)^2] + 2E[(X-\mu_X)(Y-\mu_Y)] + E[(Y-\mu_Y)^2]$

$= V[X] + 2E[(X-\mu_X)(Y-\mu_Y)] + V[Y]$

X, Y が独立なので

$E[(X-\mu_X)(Y-\mu_Y)] = 0$

$\because$ $E[(X-\mu_X)(Y-\mu_Y)] = E[XY - \mu_X Y - X\mu_Y + \mu_X\mu_Y]$

$= E[X]E[Y] - \mu_X E[Y] - E[X]\mu_Y + \mu_X\mu_Y$

$= \mu_X\mu_Y - \mu_X\mu_Y - \mu_X\mu_Y + \mu_X\mu_Y$

$= 0$

//

3. 大数の法則

Thm 大数の弱法則 weak law of large numbers

$X_1, X_2, \cdots, X_n$ は独立に同一の分布に従う
independently and identically distributed
確率変数とする
($X_1, X_2, \cdots, X_n \overset{\text{i.i.d.}}{\sim} P$ と書く)

□ $S_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ (算術平均)
arithmetic mean

□ $\mu = E[X_1] = E[X_2] = \cdots = E[X_n]$ とおくと

$\forall \varepsilon > 0$ に対して

$$\lim_{n\to\infty} \Pr\{ |S_n - \mu| > \varepsilon \} = 0$$

$$\left(\Longleftrightarrow \lim_{n\to\infty} \Pr\{ |S_n - \mu| \leq \varepsilon \} = 1 \right)$$

以下で証明

Lem (Markov の不等式)

Z を非負の値をとる r.v. とすると

$\forall a > 0$ に対して

$$\Pr\{ Z \geq a \} \leq \frac{E[Z]}{a}$$

(証明)

$$E[Z] = \sum_{z} P(z)\, z$$

$$= \sum_{z : z \geq a} P(z)\, z + \underbrace{\sum_{z : 0 \leq z < a} P(z)\, z}_{\geq 0}$$

$$\geq \sum_{z : z \geq a} P(z)\, a$$

$$= a \underbrace{\sum_{z : z \geq a} P(z)}_{= \Pr\{ Z \geq a \}}$$

両辺 a で割ればよい //

Lem（Chebyshev の不等式）

$\forall a > 0$ に対して

$$\Pr\{|X-\mu| \geq a\} \leq \frac{V[X]}{a^2}$$

ただし　$\mu = E[X]$

（証明）

$|X-\mu| \geq a \iff (X-\mu)^2 \geq a^2$ より

$$\Pr\{|X-\mu| \geq a\} = \Pr\{(X-\mu)^2 \geq a^2\}$$

$$\xrightarrow{\text{Markovの不等式}} \leq \frac{E[(X-\mu)^2]}{a^2} = \frac{V[X]}{a^2}$$

//

大数の法則の証明

Chebyshev の不等式より

$$\Pr\{|S_n-\mu| > \varepsilon\} \leq \frac{V[S_n]}{\varepsilon^2}$$

分散の性質より

$$V[S_n] = V\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]$$

$$= \frac{1}{n^2} V\left[\sum_{i=1}^{n} X_i\right] \overset{\text{独立性}}{=} \frac{1}{n^2}\sum_{i=1}^{n} \underset{V[X_1]}{\underset{\|}{V[X_i]}}$$

$$= \frac{1}{n} V[X_1]$$

よって

$$0 \leq \Pr\{|S_n-\mu| > \varepsilon\} \leq \frac{V[X_1]}{n\varepsilon^2} \longrightarrow 0 \quad (n \to \infty)$$

//

参考　中心極限定理　central limit theorem

$X_1, X_2, \cdots, X_n \overset{\text{i.i.d.}}{\sim} P$　のとき

$\mu = E[X_1]$,　$\sigma = \sqrt{V[X_1]}$　とおくと

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\,\sigma}$$ は $n \to \infty$ のとき 正規分布 $N(0,1)$ に近づく