

統計解析の備忘録 今野 良彦 (2011.01.13)

確率の公理 $(\Omega, \mathcal{A}, \mathbb{P})$

- (P1) 任意の事象 $A \in \mathcal{A}$ に対して, $0 \leq \mathbb{P}(A) \leq 1$.
 (P2) $\mathbb{P}(\Omega) = 1$.
 (P3) 互いに排反な $A_1, A_2, \dots \in \mathcal{A}$ ($A_i \cap A_j = \emptyset, i \neq j$) に対して,

$$\mathbb{P}(A_1 \cup A_2 \cup \dots) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \dots$$

確率 $(\Omega, \mathcal{A}, \mathbb{P})$ の性質

- (1) $\mathbb{P}(\emptyset) = 0$.
 (2) $A_1 \cap A_2 = \emptyset \implies \mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2)$.
 (3) $A \subset B \implies \mathbb{P}(B \cap A^c) = \mathbb{P}(B) - \mathbb{P}(A) \leq \mathbb{P}(B)$.
 (4) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
 (5) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

条件付き確率 $\mathbb{P}_B(A)$

$\mathbb{P}(B) \neq 0$ のとき,

$$\mathbb{P}(A|B) = \mathbb{P}_B(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \quad A \in \mathcal{A}.$$

条件付き確率と独立性

A と B は独立 $\iff \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$
 $\iff \mathbb{P}(A|B) = \mathbb{P}(A)$

ベイズの定理

$A_1 \cup A_2 = \Omega, A_1 \cap A_2 = \emptyset$ とする. $\mathbb{P}(B) \neq 0$ のとき,

$$\mathbb{P}(A_k|B) = \frac{\mathbb{P}(A_k)\mathbb{P}(B|A_k)}{\mathbb{P}(A_1)\mathbb{P}(B|A_1) + \mathbb{P}(A_2)\mathbb{P}(B|A_2)} \quad (k = 1, 2).$$

累積関数 $F_X(x) = \mathbb{P}(X \leq x)$ の性質

- (D1) $x_1 < x_2 \implies F_X(x_1) \leq F_X(x_2)$.
 (D2) $\lim_{h \rightarrow +0} F_X(x+h) = F_X(x)$.
 (D3) $\lim_{x \rightarrow -\infty} F_X(x) = 0; \lim_{x \rightarrow \infty} F_X(x) = 1$.

離散型確率変数 X の確率関数 $f_X(x) = \mathbb{P}(X = x)$ の定義

- (1) $f_X(x_k) \geq 0, k = 1, 2, \dots$
 (2) $\sum_{k=1}^{\infty} f_X(x_k) = 1$.

ただし, $\{x : f_X(x) > 0\} = \{x_1, x_2, \dots\}$ である.

連続型確率変数 X の確率密度関数 $f_X(x)$ の定義

$$F_X(x) = \int_{-\infty}^x f_X(t) dt.$$

連続型確率変数 X の確率密度関数 $f_X(x)$ の性質

- (1) $f_X(x) \geq 0$.
 (2) $\int_{-\infty}^{\infty} f_X(x) dx = 1$.
 (3) $F_X(x)$ が微分可能なとき, $f_X(x) = \frac{d}{dx} F_X(x)$.

確率変数 X の期待値の定義

$$\mathbb{E}[X] = \begin{cases} \sum_{k=1}^{\infty} x_k f_X(x_k) & (\text{離散型}) \\ \int_{-\infty}^{\infty} x f_X(x) dx & (\text{連続型}) \end{cases}$$

確率変数 $\varphi(X)$ の期待値の定義 ($\varphi: \mathbb{R} \rightarrow \mathbb{R}$)

$$\mathbb{E}[\varphi(X)] = \begin{cases} \sum_{k=1}^{\infty} \varphi(x_k) f_X(x_k) & (\text{離散型}) \\ \int_{-\infty}^{\infty} \varphi(x) f_X(x) dx & (\text{連続型}) \end{cases}$$

確率変数 X の期待値の性質

- (1) a, b は定数とする. このとき, $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$.

確率変数 X の分散の定義の性質

def: $\mu := \mathbb{E}[X]$ としたとき, $\text{VAR}(X) := \mathbb{E}[(X - \mu)^2]$.
 (1) $a \neq 0, b$ は定数としたとき, $\text{VAR}(aX + b) = a^2 \text{VAR}(X)$.
 (2) $\text{VAR}(X) = \mathbb{E}[X^2] - \{\mathbb{E}[X]\}^2$.

確率変数 X の基準化

$Z := \frac{X - \mathbb{E}[X]}{\sqrt{\text{VAR}[X]}}$ ($\text{VAR}[X] > 0$) としたとき,

$$\mathbb{E}[Z] = 1; \text{VAR}[Z] = 1.$$

チェビシエフの不等式

$0 < \text{VAR}[X] < \infty$ とする. このとき,

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \epsilon) \leq \frac{\text{VAR}[X]}{\epsilon^2}, \quad \forall \epsilon > 0.$$

確率ベクトル (X_1, X_2) の分布関数

$$F_{(X_1, X_2)}(x_1, x_2) = \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2), \quad (x_1, x_2) \in \mathbb{R}^2.$$

確率ベクトル (X_1, X_2) の分布関数の性質

- (1) 非減少性: $s_1 < t_1, s_2 < t_2$ ならば,

$$F_{(X_1, X_2)}(s_1, t_1) \leq F_{(X_1, X_2)}(s_2, t_2).$$

 (2) 右連続性: $F_{(X_1, X_2)}(s_2, t_2) = \lim_{h_1 \rightarrow +0, h_2 \rightarrow +0} F_{(X_1, X_2)}(x_1 + h_1, x_2 + h_2)$.
 (3) 有界性:

$$\lim_{x_2 \rightarrow -\infty} F_{(X_1, X_2)}(x_1, x_2) = 0,$$

$$\lim_{x_1 \rightarrow -\infty} F_{(X_1, X_2)}(x_1, x_2) = 0,$$

$$\lim_{x_1 \rightarrow \infty, x_2 \rightarrow \infty} F_{(X_1, X_2)}(x_1, x_2) = 1.$$

離散型確率ベクトル (X_1, X_2) の同時確率関数の定義

(X_1, X_2) の取る値が高々可算個の要素からなる集合 $D := \{(x_{1i}, x_{2j}), i, j = 1, 2, \dots\}$ のとき,

$$f_{(X_1, X_2)}(x_1, x_2) = \begin{cases} \mathbb{P}(X_1 = x_1, X_2 = x_2) & ((x_1, x_2) \in D); \\ 0 & ((x_1, x_2) \notin D) \end{cases}$$

離散型確率ベクトル (X_1, X_2) の同時確率関数の性質

- (1) $f_{(X_1, X_2)}(x_{1i}, x_{2j}) \geq 0, i = 1, 2, \dots; j = 1, 2, \dots$;
 (2) $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} f_{(X_1, X_2)}(x_{1i}, x_{2j}) = 1$.

連続型確率ベクトル (X_1, X_2) の同時確率密度関数の定義

$$F_{(X_1, X_2)}(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_{(X_1, X_2)}(t_1, t_2) dt_1 dt_2.$$

 特に, $f_{(X_1, X_2)}(x_1, x_2)$ の連続点において,

$$\frac{\partial^2}{\partial x_1 \partial x_2} F_{(X_1, X_2)}(x_1, x_2) = f_{(X_1, X_2)}(x_1, x_2).$$

連続型確率ベクトル (X_1, X_2) の同時確率密度関数の性質

- (1) $f_{(X_1, X_2)}(x_1, x_2) \geq 0$;
 (2) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{(X_1, X_2)}(x_1, x_2) dx_1 dx_2 = 1$.

周辺分布関数の定義

$$F_{X_1}(x_1) = \lim_{x_2 \rightarrow \infty} F_{(X_1, X_2)}(x_1, x_2);$$

$$F_{X_2}(x_2) = \lim_{x_1 \rightarrow \infty} F_{(X_1, X_2)}(x_1, x_2)$$

周辺確率関数の定義

$$f_{X_1}(x_1) = \sum_{j=1}^{\infty} f_{(X_1, X_2)}(x_1, x_{2j});$$

$$f_{X_2}(x_2) = \sum_{i=1}^{\infty} f_{(X_1, X_2)}(x_{1i}, x_2).$$

周辺確率密度関数の定義

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{(X_1, X_2)}(x_1, x_2) dx_2;$$

$$f_{X_2}(x_2) = \int_{-\infty}^{\infty} f_{(X_1, X_2)}(x_1, x_2) dx_1.$$

確率変数の独立性の定義

X_1 と X_2 は独立
 $\iff F_{(X_1, X_2)} = F_{X_1}(x_1)F_{X_2}(x_2) \quad \forall (x_1, x_2) \in \mathbb{R}^2$
 $\iff f_{(X_1, X_2)} = f_{X_1}(x_1)f_{X_2}(x_2) \quad \forall (x_1, x_2) \in \mathbb{R}^2$

確率変数 $\varphi(X_1, X_2)$ の期待値の定義 ($\varphi: \mathbb{R}^2 \rightarrow \mathbb{R}$)

$$E[\varphi(X_1, X_2)] = \begin{cases} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \varphi(x_{1i}, x_{2j}) f_{(X_1, X_2)}(x_{1i}, x_{2j}) & (\text{離散型}) \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(x_1, x_2) f_{(X_1, X_2)}(x_1, x_2) dx_1 dx_2 & (\text{連続型}) \end{cases}$$

確率変数 X の期待値の性質

- (1) $\mathbb{E}(X_1 + X_2) = \mathbb{E}(X_1) + \mathbb{E}(X_2)$;
 (2) $\text{VAR}(X_1 + X_2) = \text{VAR}(X_1) + \text{VAR}(X_2) + 2\text{COV}(X_1, X_2)$.
 (3) X_1 と X_2 が独立のとき, $\mathbb{E}(X_1 X_2) = \mathbb{E}(X_1)\mathbb{E}(X_2)$.

共分散との定義

$\mathbb{E}[X_1^2] < \infty, \mathbb{E}[X_2] < \infty$ のとき,
 $\text{COV}(X_1, X_2) = \mathbb{E}[\{X_1 - \mathbb{E}(X_1)\}\{X_2 - \mathbb{E}(X_2)\}]$.

共分散の性質

- (1) $\text{COV}(X_1, X_2) = \mathbb{E}(X_1 X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2)$.
- (2) 定数 a, b に対して,
 $\text{COV}(aX_1, bX_2) = ab\text{COV}(X_1, X_2)$.
- (3) X_1 と X_2 が独立のとき,
 $\text{COV}(X_1, X_2) = 0$.

相関係数の定義

$0 < \text{VAR}[X_1] < \infty, 0 < \text{VAR}[X_2] < \infty$ のとき,
 $\rho(X_1, X_2) = \frac{\text{COV}(X_1, X_2)}{\sqrt{\text{VAR}(X_1)}\sqrt{\text{VAR}(X_2)}}$.

相関係数の性質

- (1) $-1 \leq \rho(X_1, X_2) \leq 1$;
 X_1 と X_2 が独立ならば, $\rho(X_1, X_2) = 0$.

モーメント母関数の定義

$M_{(X_1, X_2)} = E[\exp(t_1 X_1 + t_2 X_2)]$.

モーメント母関数の性質

- (1) $\left. \frac{\partial M_{(X_1, X_2)}(t_1, t_2)}{\partial t_1} \right|_{(t_1, t_2)=(0,0)} = \mathbb{E}(X_1)$;
- (2) $\left. \frac{\partial^2 M_{(X_1, X_2)}(t_1, t_2)}{\partial t_1 \partial t_2} \right|_{(t_1, t_2)=(0,0)} = \mathbb{E}(X_1 X_2)$;
- (3) X_1 と X_2 が独立ならば,
 $M(X_1, X_2)(t_1, t_2) = M_{X_1}(t_1)M_{X_2}(t_2)$.

条件付確率 (密度) 関数の定義

$f_{X_1}(x_1) \neq 0$ なる x_1 に対して,
 $f_{X_2|X_1}(x_2|x_1) = \frac{f_{(X_1, X_2)}(x_1, x_2)}{f_{X_1}(x_1)}$.

確率変数列の収束のモード

- (1) 確率変数列 $\{X_n\}_{n=1}^{\infty}$ が定数 c に確率収束するとは,
任意の正数 ϵ に対して, つぎが成立することである:
 $\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - c| \geq \epsilon) = 0$.
これを $X_n \xrightarrow{P} c$ と書く.
- (2) 確率変数列 $\{X_n\}_{n=1}^{\infty}$ が分布関数 $F(x)$ に分布収束するとは, $F(x)$ のすべての連続点に対して, つぎが成立する:
 $\lim_{n \rightarrow \infty} \mathbb{P}(X_n \leq x) = F(x)$.
これを $X_n \xrightarrow{d} F$ と書く.

大数の弱法則

X_1, X_2, \dots は独立で同一の分布に従い, $\mu := \mathbb{E}(X_1)$ は存在するとき,
 $\frac{X_1 + X_2 + \dots + X_n}{n} \xrightarrow{P} \mu$.

中心極限定理

X_1, X_2, \dots は独立で同一の分布に従い, $0 < \sigma^2 := \text{VAR}(X_1)$ は存在するとき,
 $\mathbb{P}\left(\frac{X_1 + X_2 + \dots + X_n}{\sqrt{n}\sigma} \rightarrow \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt (n \rightarrow \infty)\right)$.

二項分布 $Bi(n, p)$ の確率関数

$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \dots, n$

ポアソン分布 $Po(\lambda)$ ($\lambda > 0$) の確率関数

$f_X(x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad x = 0, 1, \dots$

指数分布 $Exp(\lambda)$ ($\lambda > 0$) の確率密度関数

$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & (x > 0); \\ 0 & (x \leq 0). \end{cases}$

正規分布 $N(\mu, \sigma^2)$ ($\sigma > 0$) の確率密度関数

$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$.

正規分布の積率

- (1) $\mathbb{E}(X) = \mu$;
- (2) $\text{VAR}(X) = \sigma^2$;
- (3) $M_X(t) = \exp\left\{\mu t + \frac{\sigma^2}{2} t^2\right\}$.

確率変数の変数変換

$Y_1 = \varphi_1(X_1, X_2), Y_2 = \varphi_2(X_1, X_2)$ とし,
 $x_1 = \psi_1(y_1, y_2), x_2 = \psi_2(y_1, y_2)$ のとき,
 $f_{(Y_1, Y_2)}(y_1, y_2) = f_{(X_1, X_2)}(\psi_1(y_1, y_2), \psi_2(y_1, y_2)) |J|$. ただし,
 $J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}$

自由度 n の χ^2 分布の確率密度関数

$f_X(x) = \begin{cases} \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}, & (x > 0); \\ 0 & (x \leq 0). \end{cases}$

自由度 n の t 分布の確率密度関数

$f_X(x) = \begin{cases} \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}, & (x > 0); \\ 0 & (x \leq 0). \end{cases}$

自由度 (k_1, k_2) の F 分布の確率密度関数

$f_X(x) = \begin{cases} \frac{\left(\frac{k_1}{k_2}\right)^{k_1/2}}{B(k_1/2, k_2/2)} x^{k_1/2-1} \left(1 + \frac{k_1}{k_2} x\right)^{-(k_1+k_2)/2}, & (x > 0); \\ 0 & (x \leq 0). \end{cases}$

ガンマ関数とベータ関数の定義

- (1) $a > 0$ に対して, $\Gamma(a) = \int_0^{\infty} x^{a-1} e^{-x} dx$;
- (2) $a, b > 0$ に対して, $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

ガンマ関数の性質

- (1) $a > 0$ に対して, $\Gamma(a+1) = a\Gamma(a)$;
- (2) $n = 1, 2, \dots$ に対して, $\Gamma(n) = (n-1)!$;
- (3) $\Gamma(1/2) = \sqrt{\pi}$.