

確率・統計 A

若木宏文

wakaki@hiroshima-u.ac.jp

<http://home.hiroshima-u.ac.jp/wakaki/lecture/index.shtml>

2018.6.26

メモ用紙

2つの確率ベクトルの独立性

\mathbb{R}^n における任意の半開区間 $A = (a_{11}, a_{21}] \times \cdots \times (a_{1n}, a_{2n}]$
と

\mathbb{R}^m における任意の半開区間 $B = (b_{11}, b_{21}] \times \cdots \times (b_{1m}, b_{2m}]$
に対して

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

を満たすとき \mathbf{X} と \mathbf{Y} は独立であるという。

離散型分布と確率関数

確率変数 X のとりうる値が、有限個または可算無限個の値をとるとき、**離散型 (discrete type) 確率変数**であるといい、その分布を**離散型分布**という。

離散型分布は

$$f(x) = P(X = x) = P(\{\omega; X(\omega) = x\}), \quad x \in \mathbb{R}$$

によって決定される。 $f(x)$ を**確率関数 (probability function)**あるいは連続型の場合と同様に**確率密度関数 (probability density function), pdf**という。

離散型分布の例

2項分布 (Binomial distribution)
確率関数が

$$f(x) = P(X = x) = {}_n C_x \theta^x (1 - \theta)^{n-x}, \quad x = 0, 1, \dots, n$$

で与えられる分布をパラメータ $n, \theta \in [0, 1]$ の **2項分布** とい
い、 $B(n, \theta)$ で表す。

$D = \{0, 1, \dots, n\}$ とおくと

$$P(X \in D) = \sum_{x=0}^n P(X = x) = \{\theta + (1 - \theta)\}^n = 1 \quad (2 \text{項定理})$$

ポアソン分布 (Poisson distribution)

確率関数が

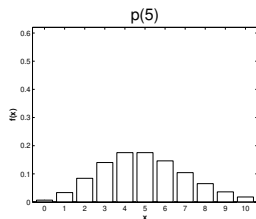
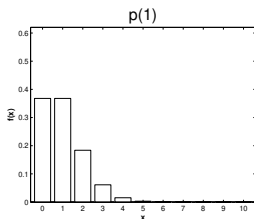
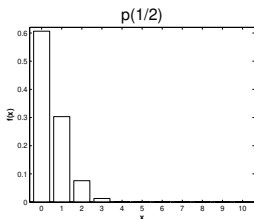
$$f(x) = P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}, \quad x = 0, 1, \dots$$

で与えられる分布をパラメータ $\lambda > 0$ のポアソン分布とい
い、 $p(\lambda)$ で表す。

$D = \{0, 1, \dots\}$ とおくと

$$P(X \in D) = \sum_{x=0}^{\infty} P(X = x) = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} \cdot e^{\lambda} = 1.$$

ポアソン分布の確率関数



ベルヌーイ試行 (1/2)

成功率 p , 反復回数 n のベルヌーイ試行 (Bernoulli trial) とは, 次の条件 (1), (2), (3) をみたす n 回の繰り返し試行のことである.

- (1) 各回の試行の結果は, A か A^c のどちらか一方しか起きない.
- (2) 各回の試行の結果は, 他の試行の結果と互いに独立である.
- (3) 事象 A が起こる確率 ($= p$) は毎回不変である.

事象 A を成功, 事象 A^c を失敗ともいう. また, このような試行の繰り返し試行を, **ベルヌーイ試行列**という.

ベルヌーイ試行 (2/2)

ある特定の事象 A の生起に着目する試行を繰り返す試行において、

第 i 回目の試行結果を表す確率変数 X_i を、

A が起これば 1 をとり、 A^c が起これば 0 をとるものとして定義する。

このとき、成功率 p 、反復回数 n のベルヌーイ試行は、次の (1), (2) をみたす確率変数列 X_1, \dots, X_n とみなすことができる。

(1) $P(X_i = 1) = p, P(X_i = 0) = 1 - p, i = 1, \dots, n.$

(2) 確率変数列 X_1, \dots, X_n は互いに独立である。

このような確率変数列を
成功率 p 、反復回数 n のベルヌーイ確率変数列という。

定理 2.8

確率変数列 X_1, \dots, X_n を成功率 p , 反復回数 n のベルヌーイ確率変数列とし, 成功の回数

$$S_n = X_1 + \dots + X_n$$

を考える. このとき, 次が成り立つ.

- (1) 確率変数 S_n は 2 項分布 $B(n, p)$ に従う.
- (2) 確率変数 S_n は 2 項分布 $B(n, p_n)$, $p_n = \lambda/n$ に従うとする. このとき

$$\lim_{n \rightarrow \infty} P(S_n = k) = p(k; \lambda), \quad k = 0, 1, \dots$$

ここで, $p(k; \lambda)$ はパラメータ λ のポアソン分布の確率関数である.

(証明は板書で)

メモ用紙

メモ用紙

その他の離散型分布

幾何分布 (Geometric distribution)

$$f(x) = P(X = x) = p(1 - p)^x, \quad x = 0, 1, 2, \dots$$

パラメータ p ($\in (0, 1)$) の幾何分布といい, $G(p)$ と表す.

負の2項分布 (Negative binomial distribution)

$$f(x) = P(X = x) = {}_{x+r-1}C_{r-1} p^r (1 - p)^x, \quad x = 0, 1, 2, \dots$$

パラメータ r, p ($\in (0, 1)$) の負の2項分布といい, $NB(r, p)$ と表す.

超幾何分布 (Hypergeometric distribution)

$$f(x) = P(X = x) = \frac{{}_M C_x {}_{N-M} C_{n-x}}{{}_N C_n}$$

x は非負の整数で, $\max\{0, n - (N - M)\} \leq x \leq \min\{n, M\}$.

パラメータ N, M, n の超幾何分布といい, $HG(N, M, n)$ と表す.

連続型分布と確率密度関数

身長, 温度, バスの待ち時間などのように X の取り得る値が連続する値となる場合で, とくに X の分布関数が

$$P(X \leq x) = F(x) = \int_{-\infty}^x f(t)dt, \quad x \in \mathbb{R}$$

と表されるとき, **連続型確率変数 (Continuous type random variable)** といい, その確率分布を**連続型分布**という.

\mathbb{R} 上の関数 $f(x)$ を**確率密度関数 (Probability density function, pdf)** という.

注. 分布関数 $F(x)$ が連続であるが, 連続型でない分布もある.

確率密度関数の性質

$$P(a < X \leq b) = \int_a^b f(x)dx$$

$f(x)$ の連続点で

$$\frac{d}{dx}F(x) = f(x)$$

(1) $f(x) \geq 0, \quad \forall x \in \mathbb{R}.$

(2) $\int_{-\infty}^{\infty} f(t)dt = 1.$

連続型分布の例

一様分布 (Uniform distribution)

確率密度関数が

$$f(x) = \begin{cases} \frac{1}{b-a}, & a < x \leq b \\ 0, & \text{その他} \end{cases}$$

で与えられる分布を区間 $(a, b]$ 上の**一様分布**といい, $U(a, b]$ で表す.

定理 2.9

連続型確率変数 X の分布関数 F は, $a < x < b$ において, 狭義単調増加で $F(a) = 0, F(b) = 1$ であるとする ($a = -\infty$ あるいは $b = \infty$ でもよい). このとき, 確率変数 $Y = F(X)$ は一様分布 $(0, 1]$ に従う. 逆に, Y が一様分布 $(0, 1]$ に従うとする. このとき, $X = F^{-1}(Y)$ の分布関数は F となる.

(証明は板書で)

メモ用紙

正規分布 (Normal distribution)

確率密度関数が

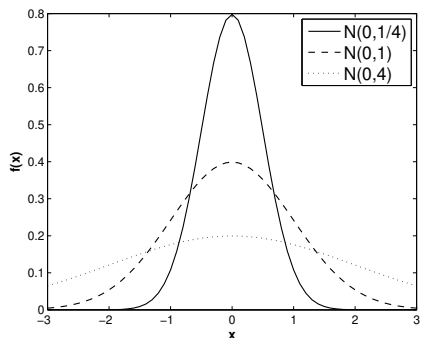
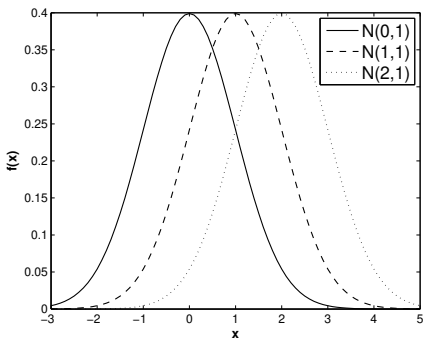
$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x - \mu)^2\right\}$$

で与えられる分布を平均 μ , 分散 σ^2 の**正規分布**といい,
 $N(\mu, \sigma^2)$ で表す. ここで, $-\infty < \mu < \infty, \sigma^2 > 0$.

$N(0, 1)$ を**標準正規分布 (Standard normal distribution)** とい
う. $N(0, 1)$ の pdf は

$$\phi(x) = f(x; 0, 1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$$

正規分布の確率密度関数



その他の連続型分布

指数分布 (Exponential distribution)

確率密度関数が

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

で与えられる分布を**指数分布**といい, $Ex(\lambda)$ で表す. ここで, λ は正の実数である.

ガンマ分布 (Gamma distribution)

確率密度関数が

$$f(x) = \begin{cases} \frac{1}{\Gamma(r)} \lambda^r e^{-\lambda x} x^{r-1}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

で与えられる分布を**ガンマ分布**といい, $Ga(r, \lambda)$ で表す. ここで, r, λ は正の実数である.

2.5.2. 2次元の場合

2次元離散型分布と連続型分布

2次元確率変数 (X, Y) の取り得る値の集合 D が, \mathbb{R}^2 の高々可算集合であるとき, (X, Y) は **離散型** 確率変数, また, (X, Y) の分布を **離散型分布** という.

離散型分布は, 確率関数

$$f(x, y) = P(X = x, Y = y), \quad (x, y) \in \mathbb{R}^2$$

によって決定される.

2次元確率変数 (X, Y) 分布関数が

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(s, t) ds dt$$

と表されるとき, (X, Y) は **連続型** 確率変数, また, (X, Y) の分布を **連続型分布** という.

$f(x, y)$ を (X, Y) または, (X, Y) の分布の **確率密度関数** と
いう.

2次元確率関数と確率密度関数の性質

- 確率関数 (離散型)

X の取り得る値を $x_1, x_2, \dots,$

Y の取り得る値を y_1, y_2, \dots とすると

$$(1) f(x_i, y_j) \geq 0, \quad i, j = 1, 2, \dots \quad (2) \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} f(x_i, y_j) = 1$$

$$\text{同時分布関数: } F(x, y) = \sum_{x_i \leq x} \sum_{y_j \leq y} f(x_i, y_j)$$

- 確率密度関数 (連続型)

$$(1) f(x, y) \geq 0, \quad \forall (x, y) \in \mathbb{R}^2 \quad (2) \int_{\mathbb{R}^2} f(x, y) dx dy = 1$$

$$\text{同時分布関数: } F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(s, t) ds dt$$

$$\left(\frac{\partial^2}{\partial x \partial y} F(x, y) = f(x, y) \right)$$

3項分布

1回の試行で事象 A, B, C のいずれかが起こりうる。

A が起こる確率を p_1 , B が起こる確率を p_2 とする。

事象 C が起こる確率は, $1 - p_1 - p_2$ である。

n 回の独立な試行で, A が起こる回数を X , B が起こる回数を Y とすれば

$$\begin{aligned} f(x, y) &= P(X = x, Y = y) \\ &= \frac{n!}{x!y!(n-x-y)!} p_1^x p_2^y (1-p_1-p_2)^{n-x-y} \\ &\quad ((x, y) \in D; 0 \leq p_1, 0 \leq p_2, p_1 + p_2 \leq 1) \end{aligned}$$

ただし, $D = \{(x, y); x, y \text{ は非負整数}, x + y \leq n\}$.

このような分布を, パラメータ $n, (p_1, p_2)$ をもつ **3項分布** といい, $M_3(n, (p_1, p_2))$ と表す。

2次元正規分布

(X, Y) の確率密度関数が

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left[-\frac{1}{2(1-\rho^2)} \left\{ \left(\frac{x-\mu_1}{\sigma_1} \right)^2 - 2\rho \left(\frac{x-\mu_1}{\sigma_1} \right) \left(\frac{y-\mu_2}{\sigma_2} \right) + \left(\frac{y-\mu_2}{\sigma_2} \right)^2 \right\} \right]$$

$$\mu_1, \mu_2 \in \mathbb{R}, \sigma_1^2 > 0, \sigma_2^2 > 0, -1 < \rho < 1$$

で表されるとき、**2次元正規分布**という。

$$\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

とおくと

$$f(x, y) = (2\pi)^{-1} |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}$$

と表せるので、 (X, Y) の分布を $N_2(\boldsymbol{\mu}, \Sigma)$ と表す。

周辺分布

- (X, Y) が離散型で、その確率関数を $f(x_i, y_j)$, $i, j = 1, 2, \dots$ とすると、 X は離散型で、その確率関数は

$$f_X(x_i) = P(X = x_i, Y \in D_y) = \sum_{y_j \in D_y} f(x_i, y_j),$$

$$D_y = \{y_1, y_2, \dots\}$$

- (X, Y) が連続型で、分布関数を $F(x, y)$, 確率密度関数を $f(x, y)$ とすると、 X は連続型で、確率密度関数は

$$F(x, \infty) = \int_{-\infty}^x \left\{ \int_{-\infty}^{\infty} f(s, t) dt \right\} ds \text{ より}$$

$$f_X(x) = \int_{-\infty}^{\infty} f(x, t) dt$$

定理 2.10

- (1) (X, Y) が 3 項分布 $M_3(n, (p_1, p_2))$ に従うとする. このとき, X, Y はそれぞれ 2 項分布 $B(n, p_1), B(n, p_2)$ に従う.
- (2) (X, Y) が 2 次元正規分布 $N_2(\boldsymbol{\mu}, \Sigma)$ に従うとする. このとき, X, Y はそれぞれ正規分布 $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$ に従う.

(証明は板書で)

メモ用紙



メモ用紙

メモ用紙

定理 2.11

(X, Y) は離散型あるいは連続型.

$f_{X,Y}(x, y)$: 同時確率密度関数

$f_X(x), f_Y(y)$: および周辺確率密度関数

$\Rightarrow X$ と Y が独立であることと

$$f_{X,Y}(x, y) = f_X(x)f_Y(y), \quad (x, y) \in D = \{(x, y); f_{X,Y}(x, y) > 0\}$$

が成り立つことは同値である.

(証明は板書で)

メモ用紙

メモ用紙