

第3章「2次元確率ベクトルの分布」

1. 同時分布関数 (Joint Distribution Function)

2つの確率変数 X, Y に対して,

$$F(x, y) = P(X \leq x, Y \leq y)$$

を (X, Y) の同時分布関数という。このとき,

$$F_1(x) = F(x, \infty) = P(X \leq x), \quad F_2(y) = F(\infty, y) = P(Y \leq y)$$

を, それぞれ, X, Y の周辺分布関数 (1次元の場合の分布関数) という。

独立性: 2つの確率変数 X と Y が互いに独立であるとは, 任意の実数 x, y に対して,

$$F(x, y) = F_1(x) F_2(y) \quad \Leftrightarrow \quad P(X \leq x, Y \leq y) = P(X \leq x) P(Y \leq y)$$

が成り立つことをいう。

2. 同時確率分布 (Joint Probability Distribution)

・離散型の場合

・同時分布 $P(X = x, Y = y)$

・周辺分布 $P(X = x) = \sum_y P(X = x, Y = y), \quad P(Y = y) = \sum_x P(X = x, Y = y)$

・条件付き分布 $P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$

・条件付き平均 $E(X|Y = y) = \sum_x x P(X = x|Y = y)$

・条件付き分散 $V(X|Y = y) = \sum_x (x - E(X|Y = y))^2 P(X = x|Y = y)$

(問1) 2個のサイコロを投げて出る目を A, B とする。そして, $X = A + B, Y = |A - B|$ とするとき, 上で定義した同時分布, \dots , 条件付き分散を求めよ。

・連続型の場合

・同時密度関数 $f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y)$

・周辺密度関数 $f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy, \quad f_2(y) = \int_{-\infty}^{\infty} f(x, y) dx$

・条件付き密度関数 $f_1(x|y) = \frac{f(x, y)}{f_2(y)}$

・条件付き平均 $E(X|Y = y) = \int_{-\infty}^{\infty} x f_1(x|y) dx$

・条件付き分散 $V(X|Y = y) = \int_{-\infty}^{\infty} (x - E(X|Y))^2 f_1(x|y) dx$

(注) 連続型の場合は, これらを手計算で行うことは一般に困難である.

・積率母関数

$$M(s, t) = E\{\exp(sX + tY)\}, \quad E(XY) = \left. \frac{\partial^2 M(s, t)}{\partial s \partial t} \right|_{s=t=0}$$

(注) 三項分布 (4節参照) の共分散を求める場合に使う.

3.2 次元確率ベクトルの関数の期待値

(X, Y) の関数 $h(X, Y)$ の期待値は,

$$E(h(X, Y)) = \begin{cases} \sum_x \sum_y h(x, y) P(X = x, Y = y) & (\text{離散型の場合}) \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) f(x, y) dx dy & (\text{連続型の場合}) \end{cases}$$

(問2) 問1のサイコロ投げにおいて, $E(XY)$ を求めよ.

・共分散 $Cov(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$

(問3) 問1のサイコロ投げにおいて, $Cov(X, Y)$ を求めよ.

・相関係数 $Corr(X, Y) = Cov(X, Y)/(SD(X) SD(Y))$

(問4) 問1のサイコロ投げにおいて, $Corr(X, Y)$ を求めよ.

・相関係数の性質

(1) $ac > 0$ ならば, $Corr(X, Y) = Corr(aX + b, cY + d)$. 特に,

$$Corr(X, Y) = E \left[\left(\frac{X - E(X)}{SD(X)} \right) \left(\frac{Y - E(Y)}{SD(Y)} \right) \right]$$

(2) 常に $|Corr(X, Y)| \leq 1$ である. 等号が成立するのは,

$$Corr(X, Y) = \begin{cases} 1 & Y = E(Y) + (SD(Y)/SD(X))(X - E(X)) \quad \text{のとき} \\ -1 & Y = E(Y) - (SD(Y)/SD(X))(X - E(X)) \quad \text{のとき} \end{cases}$$

(3) X と Y が互いに独立ならば $Corr(X, Y) = 0$ (無相関).

(問5) 上で述べた相関係数の性質 (1), (2), (3) を証明せよ.

(注) 無相関でも, 必ずしも独立ではない. 例えば, $W \sim U(0, 1)$ のとき,

$$X = \cos 2\pi W, \quad Y = \sin 2\pi W$$

は無相関である(チェック). しかし, $X^2 + Y^2 = 1$ であることから, X と Y は独立ではない.

・期待値の繰り返しの公式

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

[証明]

$$\begin{aligned} E[XY] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x, y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf_1(x|y)f_2(y) dx dy \\ &= \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} xf_1(x|y) dx \right\} yf_2(y) dy \\ &= E[Y E[X|Y]] \end{aligned}$$

4.2 次元確率分布の例

・三項分布 $T_N(n; p, q)$

各回の結果が「勝ち」「負け」「引き分け」の3通りのいずれかで, 各々の確率が $p, q, r = 1 - p - q$ である独立試行を n 回行ったとき, 「勝ち」の回数 X と「負け」の回数 Y の確率分布は,

$$P(X = x, Y = y) = \frac{n!}{x! y! (n - x - y)!} p^x q^y (1 - p - q)^{n-x-y}$$

となる. このとき, (X, Y) は三項分布 $T_N(n; p, q)$ に従うという. 周辺分布は

$$X \sim B_N(n, p) \quad Y \sim B_N(n, q)$$

(問6) $n = 4$ の場合の三項分布を明示的に表せ. また, その場合の周辺分布が二項分布になることを確かめよ.

三項分布の積率母関数と共分散, 相関係数

$$M(s, t) = E(\exp(sX + tY)) = (pe^s + qe^t + 1 - p - q)^n$$

$$E(XY) = \left. \frac{\partial^2 M(s, t)}{\partial s \partial t} \right|_{s=t=0} = n(n-1)pq$$

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = n(n-1)pq - npnq = -npq$$

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{SD(X)SD(Y)} = \frac{-npq}{\sqrt{np(1-p)nq(1-q)}} = -\sqrt{\frac{p}{1-p} \frac{q}{1-q}}$$

• 2次元正規分布 $N_2(\mu, \Sigma)$

2次元確率ベクトル $Z = (X, Y)'$ の同時密度が

$$\begin{aligned} f(x, y) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \\ &\times \exp \left[-\frac{1}{2(1-\rho^2)} \left\{ \left(\frac{x-\mu_1}{\sigma_1} \right)^2 - 2\rho \left(\frac{x-\mu_1}{\sigma_1} \right) \left(\frac{y-\mu_2}{\sigma_2} \right) + \left(\frac{y-\mu_2}{\sigma_2} \right)^2 \right\} \right] \\ &= \frac{1}{2\pi|\Sigma|^{1/2}} \exp \left(-\frac{1}{2}(z-\mu)' \Sigma^{-1} (z-\mu) \right) \end{aligned}$$

と表されるとき, Z は平均 μ , 分散共分散行列 Σ の2次元正規分布に従うといい, $Z \sim N_2(\mu, \Sigma)$ と表す. ここで,

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

同時密度 $f(x, y)$ は,

$$f(x, y) = f_1(x) \times f_2(y|x)$$

と表現される. ここで,

$$f_1(x) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp \left[-\frac{1}{2} \left(\frac{x-\mu_1}{\sigma_1} \right)^2 \right] := N(\mu_1, \sigma_1^2) \quad \text{の密度関数}$$

$$f_2(y|x) := N \left(\mu_2 + \rho \frac{\sigma_2}{\sigma_1} (x - \mu_1), (1 - \rho^2)\sigma_2^2 \right) \quad \text{の密度関数}$$

• 条件付き期待値 $E(Y|X) = \mu_2 + \rho \frac{\sigma_2}{\sigma_1} (x - \mu_1)$

• 条件付き分散 $V(Y|X) = (1 - \rho^2)\sigma_2^2$

• 共分散

$$\begin{aligned} Cov(X, Y) &= E[(X - \mu_1)(Y - \mu_2)] = E[E[(X - \mu_1)(Y - \mu_2)|X]] \\ &= E[(X - \mu_1)E[(Y - \mu_2)|X]] \\ &= E \left[(X - \mu_1) \left[\rho \frac{\sigma_2}{\sigma_1} (X - \mu_1) \right] \right] = \rho \frac{\sigma_2}{\sigma_1} E[(X - \mu_1)^2] = \rho\sigma_1\sigma_2 \end{aligned}$$

• 相関係数 $Cov(X, Y) = \rho\sigma_1\sigma_2$ の関係から, 明らかに $Corr(X, Y) = \rho$

• 無相関と独立性 2次元正規分布においては, 同値である.

・積率母関数

$$\begin{aligned}
 M(s, t) &= E[\exp(sX + tY)] = E[E[\exp(sX + tY)|X]] = E[\exp(sX) E[\exp(tY)|X]] \\
 &= E \left[\exp(sX) \exp \left[t \left\{ \mu_2 + \rho \frac{\sigma_2}{\sigma_1} (X - \mu_1) \right\} + \frac{1}{2} t^2 (1 - \rho^2) \sigma_2^2 \right] \right] \\
 &= \exp \left\{ s\mu_1 + t\mu_2 + \frac{1}{2} t^2 (1 - \rho^2) \sigma_2^2 \right\} E \left[\exp \left\{ \left(s + t\rho \frac{\sigma_2}{\sigma_1} \right) (X - \mu_1) \right\} \right] \\
 &= \exp \left\{ s\mu_1 + t\mu_2 + \frac{1}{2} t^2 (1 - \rho^2) \sigma_2^2 \right\} \exp \left\{ \frac{1}{2} \left(s + t\rho \frac{\sigma_2}{\sigma_1} \right)^2 \sigma_1^2 \right\} \\
 &= \exp \left\{ s\mu_1 + t\mu_2 + \frac{1}{2} s^2 \sigma_1^2 + st\rho\sigma_1\sigma_2 + \frac{1}{2} t^2 \sigma_2^2 \right\}
 \end{aligned}$$

5 . 2 つの独立な確率変数の和の分布

X と Y を互いに独立な連続的確率変数であるとする．また， $f(x)$ を X の密度関数， $g(y)$ を Y の密度関数とする．このとき， $Z = X + Y$ の密度関数 $h(z)$ は

$$h(z) = \int_{-\infty}^{\infty} f(z - y)g(y) dy = \int_{-\infty}^{\infty} f(x)g(z - x) dx$$

となる．これを分布のたたみ込み (convolution) という．

[証明]

$$\begin{aligned}
 P(Z \leq z) &= P(X + Y \leq z) = \int \int_{x+y \leq z} f(x)g(y) dx dy \\
 &= \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{z-y} f(x) dx \right\} g(y) dy = \int_{-\infty}^{\infty} P(X \leq z - y) g(y) dy
 \end{aligned}$$

両辺を z で微分することにより，結論を得る．

(問) X と Y が互いに独立に $[0, 1]$ 上の一様分布に従うとき， $X + Y$ の密度関数，および $X - Y$ の密度関数を求めよ．

6 . 分布の再生性

・分布族 パラメータ θ により特徴付けられた分布の集まり (パラメータは異なるが，同一の確率分布形からなる分布の集まり)．例えば，二項分布族，正規分布族，など．

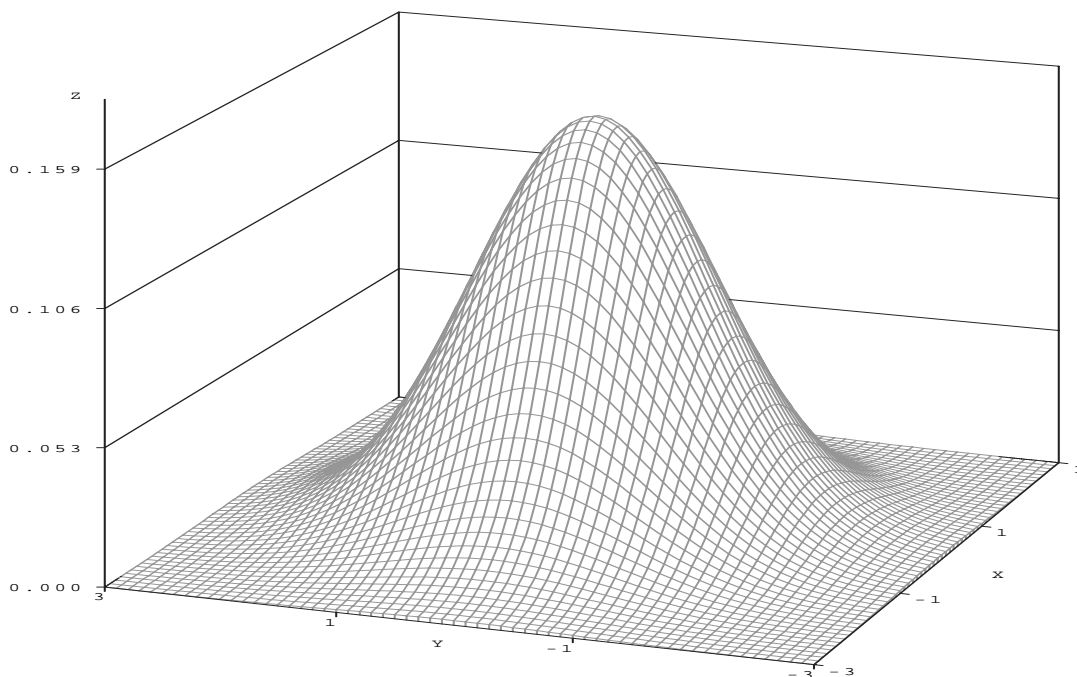
・再生性 確率変数 X と Y が互いに独立で同一の分布族の中の分布に従うとき， $Z = X + Y$ も同じ分布族の分布に従う性質をいう．

・再生性をもつ分布 二項分布，ポアソン分布，正規分布，ガンマ分布など (証明は積率母関数を使う)

・再生性をもたない分布 ベータ分布，一様分布など

2次元正規分布の密度関数

2-D STANDARD NORMAL DISTRIBUTION



上の密度関数の断面図

Contour of standard 2-d normal distribution

