2 次元正規分布

樋口さぶろお https://hig3.net

龍谷大学大学院 理工学研究科 数理情報学専攻

理論物理学特論　L02 (2022-09-28 Wed)

最終更新: Time-stamp: "2022-09-28 Wed 13:44 JST hig"

今日の目標

- 2 次元正規分布の, *母平均値母分散共分散*と, 確率密度関数が互いに求められる.
- 2 次元正規分布の周辺分布を求められる
- 正規分布が再生性を持つことを説明できる

略解 I

L01-Q1

Quiz 解答: 指数分布の確率

確率密度関数は $f(x) = \begin{cases} 2\mathrm{e}^{-2x} & (x \geq 0) \\ 0 & (x < 0) \end{cases}$.

1. 確率 $P(X \leq 3) = \int_{-\infty}^{3} f(x)\mathrm{d}x = F(3) = 1 - \mathrm{e}^{-6} = 0.997521$.
2. $F(x) = \int_{-\infty}^{x} f(x')\mathrm{d}x = \begin{cases} 0 & (x \leq 0) \\ 1 - \mathrm{e}^{-2x} & (x > 0) \end{cases}$.
3. $F^{-1}(p) = -\frac{1}{2}\log(1-p)$.
4. 確率 $P(3 < X \leq 4) = \int_{3}^{4} f(x)\mathrm{d}x = F(4) - F(3) = \mathrm{e}^{-6} - \mathrm{e}^{-8} = 0.00214329$.
5. 確率 $P(X \geq 4) = \int_{4}^{+\infty} f(x)\mathrm{d}x = F(\infty) - F(4) = \mathrm{e}^{-8} = 0.000335463$.

略解 II

6. $F(x) = 3/4$ より, $F^{-1}(3/4)$. $\mathrm{e}^{-2x} = 1/4$ より,
   $x = -\frac{1}{2}\log\frac{1}{4} = \log 2 = 0.693147$.

L01-Q2

Quiz 解答:2 次元の確率密度関数

X, Y は独立だが, これらは, 1 変数の量だけからは求められない.

1. $P(X \leq 1 \text{または} Y \leq 5) = 1 - P(X > 1 \text{かつ} Y > 5) =$
   $1 - (1 - P(X \leq 1))(1 - P(Y \leq 5) = 1 - \mathrm{e}^{-2\cdot 1}\mathrm{e}^{-3\cdot 5}$.
2. $P(Y \leq 5X) = 6\int_0^\infty \mathrm{d}x\, \mathrm{e}^{-2x} \int_0^{5x} \mathrm{d}y\mathrm{e}^{;}\mathrm{e}^{-3y} = \frac{2}{17}$.
3. $\mathrm{E}[3 + \mathrm{e}^{x+y}] = 3 + 6\int_0^\infty \mathrm{e}^{(-2+1)x\, \mathrm{d}x \times \int_0^\infty (-3+1)y}\mathrm{d}y = 3 + \frac{6}{1\cdot 2} = 6$.

L01-Q3

Quiz 解答: 独立と限らない多次元の連続型確率変数

略解 III

1. $\dfrac{459}{40}$
2. $\dfrac{7}{30}$

ここまで来たよ

1 多次元の確率変数

2 2 次元正規分布
- 一般の 1 次元正規分布
- 2 次元正規分布 (X, Y が独立)
- 2 次元正規分布 (X, Y が独立でない)
- 周辺分布
- 和の分布と再生性

一般の 1 次元正規分布 normal distribution $\mathrm{N}(\mu, \sigma^2)$

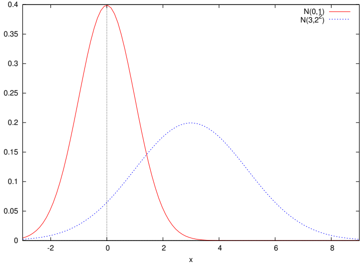
定義 (一般の 1 次元正規分布 $\mathrm{N}(\mu, \sigma^2)$ 岩薩林 確率・統計 (4.23))

母平均値 $\mathrm{E}[X] = \mu$, 母分散 $\mathrm{V}[X] = \sigma^2$ の正規分布 $\mathrm{N}(\mu, \sigma^2)$ の確率密度関数は,

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

scipy.stats.norm(loc=mu,scale=σ)

標準正規分布は, $\mu = 0, \sigma^2 = 1$ のときのこと. $Z \sim \mathrm{N}(0, 1^2)$.



きょう繰り返し使う積分

$Z \sim \mathrm{N}(0, 1^2)$ に対して, $X = \sigma Z + \mu$ は, と $X \sim \mathrm{N}(\mu, \sigma^2)$.

$$\mathrm{E}[X^0] = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{1}{2\sigma^2}(x-\mu)^2} \, \mathrm{d}x = 1,$$

$$\int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{1}{2}ax^2} \, \mathrm{d}x = \sqrt{\frac{2\pi}{a}},$$

$$\int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{1}{2}ax^2 + bx} \, \mathrm{d}x = \sqrt{\frac{2\pi}{a}} \cdot \mathrm{e}^{-\frac{b^2}{2a}}.$$

$a = \frac{1}{\sigma^2}$, $\frac{b}{a} = \mu$.

ここまで来たよ

1. 多次元の確率変数
2. 2 次元正規分布
   - 一般の 1 次元正規分布
   - 2 次元正規分布 (X, Y が独立)
   - 2 次元正規分布 (X, Y が独立でない)
   - 周辺分布
   - 和の分布と再生性

2 次元正規分布 (X, Y が独立でない) をプレビュー

定義 (一般の 2 次元正規分布)

連続型確率変数 X, Y が 次の形 (指数関数の肩が x, y の 2 次式) の同時確率密度関数を持つとき, X, Y は 2 次元正規分布にしたがうという.

$$\begin{aligned} f(x, y) =& C \times \mathrm{e}^{-a(x-\mu_{\mathrm{X}})^2 - b(y-\mu_{\mathrm{Y}})^2 + cxy} \\ =& C' \times \mathrm{e}^{-ax^2 - by^2 + cxy + px + qy} \end{aligned}$$

$a, b > 0, c$ と C, C' は, ある条件を満たす定数.

あとから行列とベクトルで美しい形で書く

2 次元正規分布で $c = 0$(X, Y が独立) 岩薩林 確率・統計 §4.6, 例題 4.12

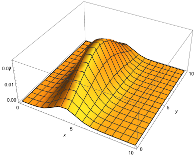
確率変数 X, Y に対して, 次の同時確率密度関数を考える.

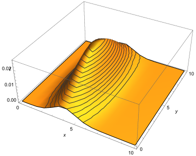
$$f(x,y) = \frac{1}{(2\pi\sigma_{\mathrm{X}}^2)^{1/2}} \mathrm{e}^{-\frac{(x-\mu_{\mathrm{X}})^2}{2\sigma_{\mathrm{X}}^2}} \times \frac{1}{(2\pi\sigma_{\mathrm{Y}}^2)^{1/2}} \mathrm{e}^{-\frac{(y-\mu_{\mathrm{Y}})^2}{2\sigma_{\mathrm{Y}}^2}}.$$

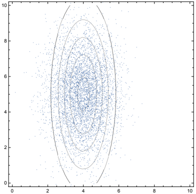
$c = 0$ のとき, X, Y の周辺分布は… よって X, Y は独立.

$$\mathrm{E}[X] = \mu_{\mathrm{X}}, \quad \mathrm{E}[Y] = \mu_{\mathrm{Y}}, \quad \mathrm{V}[X] = \sigma_{\mathrm{X}}^2, \quad \mathrm{V}[Y] = \sigma_{\mathrm{Y}}^2, \mathrm{E}[XY] = \mu_{\mathrm{X}} \cdot \mu_{\mathrm{Y}}$$

$$\text{母共分散}\mathrm{Cov}[X,Y] = 0, \text{母相関係数} r_{\mathrm{XY}} = \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{V}[X]}\sqrt{\mathrm{V}[Y]}} = 0$$







f の等高線は $(\mu_{\mathrm{X}}, \mu_{\mathrm{Y}})$ を中心とする楕円.

L02-Q1

Quiz(2 次元正規分布)

次の同時確率密度関数は 2 次元正規分布を定める.

$$f(x, y) = C \cdot \mathrm{e}^{-x^2-4x-2y^2+12y-5}.$$

1. X, Y の母平均値, 母分散, 母共分散を求めよう.
2. $\mathrm{E}[1] = 1$ が満たされるように定数 C を定めよう.

2 次形式の標準化 (線形代数 I)

L02-Q2

Quiz(2 次元正規分布)

次の 2 変数確率密度関数は 2 次元正規分布を定める.

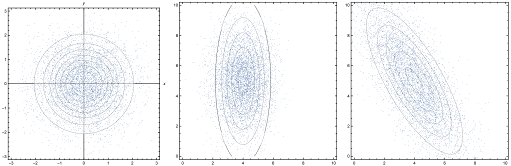
$$f(x,y) = C \cdot \mathrm{e}^{-4x^2 - \frac{1}{6}y^2 + 2y}.$$

1. X, Y の母平均値, 母分散, 母共分散を求めよう.
2. $\mathrm{E}[1] = 1$ が満たされるように定数 C を定めよう.

ここまで来たよ

1. 多次元の確率変数
2. 2 次元正規分布
   - 一般の 1 次元正規分布
   - 2 次元正規分布 $(X, Y$ が独立)
   - 2 次元正規分布 $(X, Y$ が独立でない)
   - 周辺分布
   - 和の分布と再生性

等高線が傾いた楕円の分布も作りたい!



考え方 1

等高線が円 $D = x^2 + y^2$

等高線が傾いた楕円 $D = x^2 + cxy + 2y^2$　　2 次形式の標準化 (線形代数 I,II)

考え方 2 独立じゃなくしたいから, 指数関数の引数に xy の項でもいれておけば?

一般の 2 次元正規分布

$$f(x,y) = C \times \mathrm{e}^{-ax^2-by^2+cxy+px+qy}$$

命題 (一般の 2 次元正規分布の同時確率密度関数)

2 次元正規分布 $\mathrm{N}(\boldsymbol{\mu}, \Sigma)$ の同時確率密度関数は, 確率変数を $\boldsymbol{X} = {}^{\mathrm{t}}(X, Y)$ とするとき,

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^2 \det \Sigma}} \mathrm{e}^{-\frac{1}{2}{}^{\mathrm{t}}(\boldsymbol{x}-\boldsymbol{\mu})\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$$

$$\text{母平均値 (ベクトル)}\boldsymbol{\mu} = \begin{pmatrix} \mu_{\mathrm{X}} \\ \mu_{\mathrm{Y}} \end{pmatrix} \overset{\text{要証明}}{=} \begin{pmatrix} \mathrm{E}[X] \\ \mathrm{E}[Y] \end{pmatrix}$$

$$\text{母共分散行列}\Sigma = \begin{pmatrix} \sigma_{\mathrm{X}}^2 & C_{\mathrm{XY}} \\ C_{\mathrm{XY}} & \sigma_{\mathrm{Y}}^2 \end{pmatrix} \overset{\text{要証明}}{=} \begin{pmatrix} \mathrm{V}[X] & \mathrm{Cov}[X,Y] \\ \mathrm{Cov}[X,Y] & \mathrm{V}[Y] \end{pmatrix}$$

- たしかに正規分布
- $\mathrm{Cov}[X, Y] = 0$ のとき独立なケースに帰着.

scipy.stats.multivariate_normal(mean=[μ_x,μ_y],cov=[[σ_x^2,σ_{xy}],[σ_{xy},σ_y^2]])

multivariate analysis $=$ 多変量解析 $\neq$ 多変数の解析学 $=$ multivariable calculus

L02-Q3

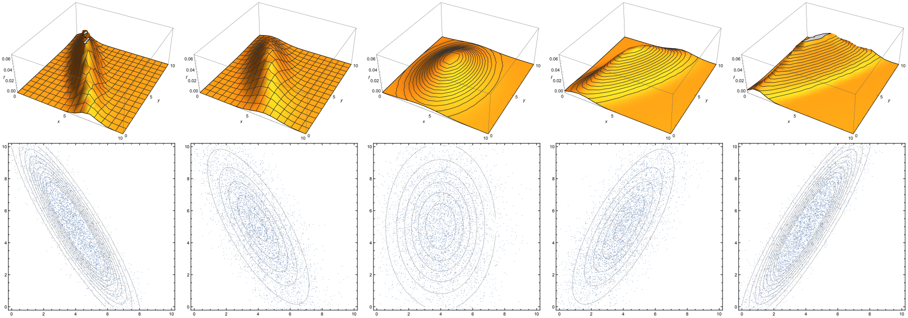
Quiz(2 次元正規分布の確率密度関数)

確率変数 X, Y は 2 次元正規分布にしたがい, $\mathrm{E}[X] = -4, \mathrm{E}[Y] = 5$, $\mathrm{V}[X] = 2, \mathrm{V}[Y] = 3, \mathrm{Cov}[X, Y] = -1$ を満たす.
同時確率密度関数 $f(x, y)$ を (行列やベクトルを使わずに) 具体的に求めよう.

母相関係数 r, 共分散行列 Σ. $c = \frac{9\sqrt{21}}{10}$

$r = -0.90 \qquad -0.76 \qquad 0 \qquad +0.76 \qquad +0.90$

$$\Sigma = \begin{pmatrix} 3 & -c \\ -c & 7 \end{pmatrix}, \begin{pmatrix} 3 & -2\sqrt{3} \\ -2\sqrt{3} & 7 \end{pmatrix}, \begin{pmatrix} 3 & 0 \\ 0 & 7 \end{pmatrix}, \begin{pmatrix} 3 & +2\sqrt{3} \\ +2\sqrt{3} & 7 \end{pmatrix}, \begin{pmatrix} 3 & +c \\ +c & 7 \end{pmatrix}$$



確率密度関数から母ナントカ I

L02-Q4

Quiz(2 次元正規分布)

次の 2 変数確率密度関数は 2 次元正規分布を定める.

$$f(x, y) = C \cdot \mathrm{e}^{-2x^2+6xy-7y^2+26x-54y-7}.$$

1. 母平均値, 母共分散行列を求めよう.
2. $\mathrm{E}[1] = 1$ が満たされるように定数 C を定めよう.

L02-Q5

Quiz(2 次元正規分布)

次の 2 変数確率密度関数は 2 次元正規分布を定める.

$$f(x, y) = C \cdot \mathrm{e}^{-x^2+4xy-7y^2}.$$

1. 母共分散行列を求めよう.
2. $\mathrm{E}[1] = 1$ が満たされるように定数 C を定めよう.

2 次形式の標準化 (線形代数 I)

ここまで来たよ

1 多次元の確率変数

2 2 次元正規分布
- 一般の 1 次元正規分布
- 2 次元正規分布 (X, Y が独立)
- 2 次元正規分布 (X, Y が独立でない)
- 周辺分布
- 和の分布と再生性

周辺分布

定義 (周辺分布)

$f(x, y)$：連続型確率変数 X, Y の同時確率密度関数とする
次を X の周辺分布の確率密度関数は次で与えられる.

$$f_{\mathrm{X}}(x) = \int_{-\infty}^{+\infty} f(x, y)\mathrm{d}y$$

2 次元正規分布の周辺分布

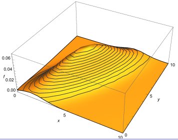
命題 (2 次元正規分布の周辺分布)

X, Y が (独立とはかぎらない)2 次元正規分布にしたがうとき，
X の周辺分布は 1 次元正規分布である

2 次形式の標準化 (線形代数 I)

簡単のため,1 次の項がない場合で示す.

$$
\begin{aligned}
&\int_{-\infty}^{+\infty} C\mathrm{e}^{-ax^2-by^2+cxy}\,\mathrm{d}y \\
=&\int_{-\infty}^{+\infty} C\mathrm{e}^{-ax^2-b(y-\frac{c}{2b}x)^2+\frac{c^2}{4b}x^2}\,\mathrm{d}y \\
=&C'\,\mathrm{e}^{-(a-\frac{c^2}{4b})x^2}
\end{aligned}
$$



L02-Q6

Quiz(2 次元正規分布の周辺分布)

次の 2 変数確率密度関数は (X, Y) の 2 次元正規分布を定める.

$$f(x, y) = C \cdot \mathrm{e}^{-x^2+4xy-7y^2}.$$

1. X の周辺分布の確率密度関数 $f_{\mathrm{X}}(x)$ を求めよう.
2. Y の周辺分布の確率密度関数 $f_{\mathrm{Y}}(y)$ を求めよう.

ここまで来たよ

1 多次元の確率変数

2 2 次元正規分布
- 一般の 1 次元正規分布
- 2 次元正規分布 (X, Y が独立)
- 2 次元正規分布 (X, Y が独立でない)
- 周辺分布
- 和の分布と再生性

2 次元連続型確率変数の和

$S = X + Y$ の確率密度関数 $f_{\mathrm{S}}(s)$ は?
確率 $P(c \leq S < d) = \int_{-\infty}^{+\infty} \mathrm{I}_{[c \leq s < d]}(s) f_{\mathrm{S}}(s)\,\mathrm{d}s$ となるべき.

$$\text{左辺} \overset{\text{定義}}{=} \iint_{-\infty}^{+\infty} \mathrm{I}_{[c \leq x+y < d]}(x, y) f(x, y)\,\mathrm{d}y\mathrm{d}x$$

$$= \iint_{-\infty}^{+\infty} \mathrm{I}_{[c \leq s < d]}(s) f(t, s-t)\,\mathrm{d}s\mathrm{d}t \quad \text{変換 } s = x + y,\ t = x.\ |J| = 1$$

よって, $f_{\mathrm{S}}(s) = \int_{-\infty}^{+\infty} f(t, s-t)\mathrm{d}t$.
($c = -\infty$ とおいて累積分布関数 $F(d) = P(S < d)$ で話すことも)

命題 (和の確率密度関数)

連続型 X, Y の確率密度関数:$f(x, y)$. 和 $S = X + Y$ の確率密度関数は,

$$f_{\mathrm{S}}(s) = \int_{-\infty}^{+\infty} f(x, s-x)\mathrm{d}x.$$

たたみ込み convolution

命題 (和の確率密度関数)

連続型確率密度関数が独立で, X, Y の確率密度関数 $f(x,y)$ が $f(x,y) = f_{\mathrm{X}}(x) f_{\mathrm{Y}}(y)$ と書けるとき, 和 $S = X + Y$ の確率密度関数 $f_{\mathrm{S}}(s)$ は,

$$f_{\mathrm{S}}(s) = \int_{-\infty}^{+\infty} f_{\mathrm{X}}(x) f_{\mathrm{Y}}(s-x) \mathrm{d}x.$$

この右辺の積分を, 2 つの関数 $f(x), g(y)$ のたたみ込み合成積 convolution といい, $(f * g)(s)$ と書くことがある.
合成積 $(f * g)$ のフーリエ変換は, f, g のフーリエ変換の積　フーリエ解析

L02-Q7

Quiz(正規分布の和の分布)

連続型確率変数 X_1, X_2 は, 独立同分布 $\mathrm{N}(0, 1^2)$ にしたがう.

1. 連続型確率変数 $S = X_1 + X_2$ のしたがう分布を答えよう.
2. 連続型確率変数 $D = 2X_1$ のしたがう分布を答えよう.

再生性 reproductive property

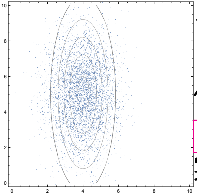
命題 (正規分布の再生性 岩薩林 確率・統計 定理 4.3)

独立な連続型確率変数 $X \sim \mathrm{N}(\mu_{\mathrm{X}}, \sigma_{\mathrm{X}}^2)$, $Y \sim \mathrm{N}(\mu_{\mathrm{Y}}, \sigma_{\mathrm{Y}}^2)$ に対して,

$$S = aX + bY \sim \mathrm{N}(a\mu_{\mathrm{X}} + b\mu_{\mathrm{Y}}, a^2\sigma_{\mathrm{X}}^2 + b^2\sigma_{\mathrm{Y}}^2)$$

特に,

$$S = X + Y \sim \mathrm{N}(\mu_{\mathrm{X}} + \mu_{\mathrm{Y}}, \sigma_{\mathrm{X}}^2 + \sigma_{\mathrm{Y}}^2)$$



再生性とは, 独立な確率変数 X, Y の和 $S = X + Y$ が (パラメタは違うけど) 同じ分布 (今なら正規分布) にしたがうという, 稀なよい性質.
他に, 二項分布, ポアソン分布など.
岩薩林 確率・統計 例題 4.14 特定の a, b, μ, σ での証明
実は, 一般の (独立とかぎらない)2 次元正規分布 (X, Y) に対しても $aX + bY$ は 1 次元正規分布にしたがう).

L02-Q8

Quiz(正規分布の和の分布)

連続型確率変数 X_i $(i = 1, \ldots, n)$ は, 独立同分布 $\mathrm{N}(3, 2^2)$ にしたがう. 連続型確率変数 $S = \sum_{i=1}^{n} X_i$ のしたがう分布を答えよう.