# 船津・小寺研究室 新人向け事前課題 (解答)

井上 貴央

2019 年 4 月 15 日

**問題 1 (1 変数関数の微分)**
以下の関数の導関数を求めてください.

(1) $f_1(x) = x\log x \quad (x > 0)$.
(2) $f_2(x) = \exp\left(-\frac{1}{2}(x-m)^2\right) \quad (m \in \mathbb{R}$ は定数).
(3) $f_3(x) = \arctan x$.
(4) $f_4(x) = \frac{1}{1+\exp(-ax)} \quad (a \in \mathbb{R}$ は定数).

**問題の解答 1**

(1) 積の微分公式を使って,

$$f_1'(x) = \log x + x \cdot \frac{1}{x} = \log x + 1.$$

(2) 合成関数の微分公式を使って,

$$f_2'(x) = \exp\left(-\frac{1}{2}(x-m)^2\right) \cdot (-(x-m)) = -(x-m)\exp\left(-\frac{1}{2}(x-m)^2\right).$$

(3) $y = \arctan x$ とおく. ここで, arctan は tan の逆関数で $x = \tan y$ であるから,

$$\frac{\mathrm{d}x}{\mathrm{d}y} = \frac{1}{\cos^2 y} = 1 + \tan^2 y = 1 + (\tan(\arctan x))^2 = 1 + x^2.$$

よって, 逆関数の微分公式より

$$f_3'(x) = \frac{\mathrm{d}y}{\mathrm{d}x} = \left(\frac{\mathrm{d}x}{\mathrm{d}y}\right)^{-1} = \frac{1}{1+x^2}.$$

(4) 合成関数の微分公式を使って,

$$f_4'(x) = -\frac{1}{(1+\exp(-ax))^2} \cdot \exp(-ax) \cdot (-a) = \frac{a\exp(-ax)}{(1+\exp(-ax))^2}.$$

□

**問題 2 (多変数関数の偏微分)**
以下の関数の $x_1$ に関する偏導関数を求めてください.

(1) $f_1(x_1, x_2) = x_1 \cos x_2$.
(2) $f_2(x_1, x_2) = \frac{2x_1x_2}{x_2^2 - x_1^2}$.
(3) $f_3(x_1, x_2, x_3) = \sqrt{x_1^2 + x_2^2 + x_3^2}$.
(4) $f_4(x_1, x_2) = \arctan \frac{x_2}{x_1} \quad (x_1 \neq 0)$.

**問題の解答 2**

(1) 単純に微分するだけ.

$$\frac{\partial}{\partial x_1} f_1(x_1, x_2) = \cos x_2.$$

(2) 商の微分公式を使えばよく,

$$\frac{\partial}{\partial x_1} f_2(x_1, x_2) = 2x_2 \cdot \frac{(x_2^2 - x_1^2) - x_1(-2x_1)}{(x_2^2 - x_1^2)^2} = \frac{2x_2(x_1^2 + x_2^2)}{(x_2^2 - x_1^2)^2}.$$

(3) 合成関数の微分公式を使って,

$$\frac{\partial}{\partial x_1} f_3(x_1, x_2, x_3) = \frac{1}{2} \frac{2x_1}{\sqrt{x_1^2 + x_2^2 + x_3^2}} = \frac{x_1}{\sqrt{x_1^2 + x_2^2 + x_3^2}}.$$

(4) 合成関数の微分公式と, **問題 1** の (3) の結果を利用すれば,

$$\frac{\partial}{\partial x_1} f_4(x_1, x_2) = \frac{1}{1 + (x_2/x_1)^2} \left( -\frac{x_2}{x_1^2} \right) = -\frac{x_2}{x_1^2 + x_2^2}.$$

□

**発展問題 1 (Gauss 積分)**
Gauss 積分について

$$\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{\lambda}{2}(x-\mu)^2}\, \mathrm{d}x = \sqrt{\frac{2\pi}{\lambda}} \quad (\mu \in \mathbb{R},\ \lambda > 0 \text{ は定数})$$

を証明してください.

**発展問題の解答 1**
求める積分を

$$I = \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{\lambda}{2}(x-\mu)^2}\, \mathrm{d}x$$

とおく. $t = \sqrt{\frac{\lambda}{2}}(x-\mu)$ とすると, $\frac{\mathrm{d}t}{\mathrm{d}x} = \sqrt{\frac{\lambda}{2}}$ なので,

$$I = \int_{-\infty}^{\infty} \mathrm{e}^{-t^2} \frac{\mathrm{d}x}{\mathrm{d}t}\,\mathrm{d}t = \sqrt{\frac{2}{\lambda}} \int_{-\infty}^{\infty} \mathrm{e}^{-t^2}\,\mathrm{d}t.$$

以下, 積分

$$I' = \int_{-\infty}^{\infty} \mathrm{e}^{-t^2}\,\mathrm{d}t$$

を計算しよう. 両辺を 2 乗すると

$$(I')^2 = \left(\int_{-\infty}^{\infty} \mathrm{e}^{-t^2}\,\mathrm{d}t\right)^2 = \left(\int_{-\infty}^{\infty} \mathrm{e}^{-s^2}\,\mathrm{d}s\right)\left(\int_{-\infty}^{\infty} \mathrm{e}^{-t^2}\,\mathrm{d}t\right) = \int_{\mathbb{R}^2} \mathrm{e}^{-(s^2+t^2)}\,\mathrm{d}s\mathrm{d}t$$

である[*1]. ここで, 極座標変換

$$\begin{cases} s = r\cos\theta \\ t = r\sin\theta \end{cases}$$

を行う $(r > 0,\ \theta \in [0, 2\pi))$. この変換の Jacobi 行列は,

$$\frac{\partial(s,t)}{\partial(r,\theta)} = \begin{pmatrix} \dfrac{\partial s}{\partial r} & \dfrac{\partial s}{\partial \theta} \\ \dfrac{\partial t}{\partial r} & \dfrac{\partial t}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$$

なので, Jacobian は

$$J = \det\left(\frac{\partial(s,t)}{\partial(r,\theta)}\right) = r.$$

よって,

$$\begin{aligned}
(I')^2 &= \int_0^{\infty}\int_0^{2\pi} \mathrm{e}^{-r^2}|J|\,\mathrm{d}\theta\mathrm{d}r \\
&= \left(\int_0^{2\pi} \mathrm{d}\theta\right)\left(\int_0^{\infty} r\mathrm{e}^{-r^2}\,\mathrm{d}r\right) \\
&= 2\pi\left[-\frac{1}{2}\mathrm{e}^{-r^2}\right]_0^{\infty} \\
&= \pi.
\end{aligned}$$

積分 $I'$ の被積分関数は $e^{-t^2} > 0$ なので, $I' > 0$. これより,

$$I' = \sqrt{\pi}.$$

ゆえに,

$$I = \sqrt{\frac{2}{\lambda}} I' = \sqrt{\frac{2\pi}{\lambda}}.$$

□

[*1] 広義積分を取り扱っているので, 本当は最後の等号が成立するかどうかについて吟味する必要があるが, ここでは事実として成立することを利用した.

**問題 3 (凸関数の制約なし最適化)**
以下の目的関数の最小解とそのときの最小値を求めてください．なお, その際に以下の関数が凸関数になっていることを用いても構いません．

(1) $f_1(x) = ax^2 + bx + c \quad (a, b, c \in \mathbb{R}$ は定数, $a > 0)$.
(2) $f_2(x) = x \log x \quad (x > 0)$.
(3) $f_3(x_1, x_2, x_3) = 4x_1^2 + 4x_2^2 + 4x_3^2 - 2x_1x_2 - 2x_2x_3 + 2x_3x_1$.
(4) $f_4(x) = |x - a| \quad (a \in \mathbb{R}$ は定数$)$.

**問題の解答 3**
関数 $f_1, f_2, f_3, f_4$ はすべて凸関数で制約条件がないので, 停留点を求めればそれが最小解になる.

(1) 停留点は, $f_1'(x) = 2ax + b = 0$ となる点．これより, $x = -\frac{b}{2a}$．ゆえに, $x = -\frac{b}{2a}$ で関数 $f_1(x)$ は最小値 $-\frac{b^2}{4a} + c$ をとる[*2].
(2) 停留点は, $f_2'(x) = \log x + 1 = 0$ となる点．これより, $x = \mathrm{e}^{-1}$．ゆえに, $x = \mathrm{e}^{-1}$ で関数 $f_2(x)$ は最小値 $-\mathrm{e}^{-1}$ をとる．
(3) 停留点は,

$$\begin{cases} \dfrac{\partial f_3}{\partial x_1} = 8x_1 - 2x_2 + 2x_3 = 0 \\[2ex] \dfrac{\partial f_3}{\partial x_2} = -2x_1 + 8x_2 - 2x_3 = 0 \\[2ex] \dfrac{\partial f_3}{\partial x_3} = 2x_1 - 2x_2 + 8x_3 = 0 \end{cases}$$

をとなる点．この連立方程式を行列の形で書くと,

$$\begin{pmatrix} 4 & -1 & 1 \\ -1 & 4 & -1 \\ 1 & -1 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

となるが, 行列

$$\begin{pmatrix} 4 & -1 & 1 \\ -1 & 4 & -1 \\ 1 & -1 & 4 \end{pmatrix}$$

[*2] ちなみに, この問題は平方完成でも解ける．

$$f_1(x) = a\left(x^2 + \frac{b}{a}x\right) + c = a\left(x + \frac{b}{2a}\right)^2 + c - \frac{b^2}{4a}$$

となり, 答えは一致する．

は正則であるから逆行列をもち,

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

が分かる. ゆえに, $(x_1, x_2, x_3) = (0, 0, 0)$ で関数 $f_3(x_1, x_2, x_3)$ は最小値 0 をとる.

(4) 関数 $f_4(x)$ は,

$$f_4(x) = \begin{cases} -x + a & (x < a) \\ x - a & (x \geq a) \end{cases}$$

となっており, $x < a$ で狭義単調減少, $x \geq a$ で狭義単調増加であるから, $x = a$ で関数 $f_4(x)$ は最小値 0 をとる. □

---

**発展問題 2 (Lagrange の未定乗数法)**

等式制約 $x_1 + x_2 + x_3 = 1$ のもとで, 関数 $f(x_1, x_2, x_3) = 4x_1^2 + 4x_2^2 + 4x_3^2 - 2x_1x_2 - 2x_2x_3 + 2x_3x_1$ の最小解 $x_1, x_2, x_3$ とそのときの最小値を求めてください. なお, 目的関数が凸関数になっていることを利用しても構いません.

---

**発展問題の解答 2**

目的関数が凸関数で, 制約条件が 1 次式なので, Lagrange 関数

$$L(x_1, x_2, x_3, \lambda) := f(x_1, x_2, x_3) - \lambda(x_1 + x_2 + x_3 - 1)$$

の停留点を求めればそれが最小解になる. 停留点は,

$$\begin{cases} \dfrac{\partial f}{\partial x_1} = 8x_1 - 2x_2 + 2x_3 - \lambda = 0 \\ \dfrac{\partial f}{\partial x_2} = -2x_1 + 8x_2 - 2x_3 - \lambda = 0 \\ \dfrac{\partial f}{\partial x_3} = 2x_1 - 2x_2 + 8x_3 - \lambda = 0 \\ \dfrac{\partial f}{\partial \lambda} = -x_1 - x_2 - x_3 + 1 = 0 \end{cases}$$

を満たす点である．整理して，

$$
\begin{cases}
\dfrac{\partial f}{\partial x_1} = 4x_1 - x_2 + x_3 = \dfrac{\lambda}{2} & (1) \\[2ex]
\dfrac{\partial f}{\partial x_2} = -x_1 + 4x_2 - x_3 = \dfrac{\lambda}{2} & (2) \\[2ex]
\dfrac{\partial f}{\partial x_3} = x_1 - x_2 + 4x_3 = \dfrac{\lambda}{2} & (3) \\[2ex]
\dfrac{\partial f}{\partial \lambda} = x_1 + x_2 + x_3 = 1 & (4)
\end{cases}
$$

を満たす点を求めれば良い．式 (3) から式 (1) を辺々引くことで，

$$x_1 = x_3 \tag{5}$$

が分かる．これと式 (4) より，

$$2x_1 + x_2 = 1. \tag{6}$$

また，式 (1) と式 (2) を辺々足すことで，

$$x_1 + x_2 = \frac{\lambda}{3} \tag{7}$$

が分かる．ここで，$\lambda$ を定数だと思って式 (6), (7) を解くと，

$$x_1 = 1 - \frac{\lambda}{3}, \quad x_2 = \frac{2}{3}\lambda - 1 \tag{8}$$

これと式 (5) をあわせて，式 (1) に代入して $\lambda$ を求めると，

$$\lambda = \frac{36}{17}$$

となる．これと式 (5), (8) より，

$$x_1 = \frac{5}{17}, \quad x_2 = \frac{7}{17}, \quad x_3 = \frac{5}{17}.$$

よって，$x_1 + x_2 + x_3 = 1$ のもとでは，$(x_1, x_2, x_3) = (\frac{5}{17}, \frac{7}{17}, \frac{5}{17})$ のとき最小値 $\frac{18}{17}$ となる． □

**発展問題 3 (一般の関数の制約なし最適化)**
関数 $f(x_1, x_2) = x_1^4 + x_2^4 - (x_1 + x_2)^2$ の最小解 $x_1, x_2$ とそのときの最小値を求めてください．

**発展問題の解答 3**

まずは, 停留点をすべて求める. 停留点は,

$$\begin{cases} \dfrac{\partial f}{\partial x_1} = 4x_1^3 - 2(x_1 + x_2) = 0 \\ \dfrac{\partial f}{\partial x_2} = 4x_2^3 - 2(x_1 + x_2) = 0 \end{cases}$$

を満たす点. すなわち,

$$\begin{cases} \dfrac{\partial f}{\partial x_1} = 4x_1^3 - 2(x_1 + x_2) = 0 & (1) \\ \dfrac{\partial f}{\partial x_2} = 4x_2^3 - 2(x_1 + x_2) = 0 & (2) \end{cases}$$

を解けば良い. 式 (1) から式 (2) を辺々引いて両辺を 2 で割ると,

$$x_1^3 - x_2^3 = 0$$

すなわち,

$$(x_1 - x_2)(x_1^2 + x_1x_2 + x_2^2) = 0 \tag{3}$$

となる. これより, $x_1 = x_2$ または $x_1^2 + x_1x_2 + x_2^2 = 0$ であるが,

$$x_1^2 + x_1x_2 + x_2^2 = \left(x_1 + \frac{x_2}{2}\right)^2 + \frac{3}{4}x_2^2$$

なので, $x_1^2 + x_1x_2 + x_2^2 = 0 \iff x_1 = x_2 = 0$. 結局, $x_1 = x_2$ である. このとき, 式 (1) より

$$x_1^3 - x_1 = 0 \iff x_1(x_1 - 1)(x_1 + 1) = 0.$$

よって, $(x_1, x_2) = (0, 0), (\pm 1, \pm 1)$ である (複号同順).

次に, 点 $(x_1, x_2)$ における Hesse 行列 $H(x_1, x_2)$ を求める.

$$H(x_1, x_2) := \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} \\ \dfrac{\partial^2 f}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} \end{pmatrix} = 2\begin{pmatrix} 6x_1^2 - 1 & -1 \\ -1 & 6x_2^2 - 1 \end{pmatrix}.$$

あとは, 各停留点に対して, Hesse 行列が正定値であるかどうかを判定する. これは, Hesse 行列のすべての首座小行列式を確認すれば良い (Sylvester の判定法).

**点 $(0, 0)$ について**. Hesse 行列は

$$\begin{pmatrix} -2 & -2 \\ -2 & -2 \end{pmatrix}.$$

であり, 首座小行列式は

$$\det(-2) = -2, \quad \det\begin{pmatrix} -2 & -2 \\ -2 & -2 \end{pmatrix} = 0$$

となる．この Hesse 行列は半負定値であるからこの点で極値をとるかどうかはわからないが, Hesse 行列が半正定値でないならば極小解ではないので, 極大解か鞍点のいずれかである．ゆえに, 最小解にはなりえない．(実は, この点は鞍点である．実際に, 点 $(0,0)$ での関数 $f$ の値は 0 であり, 点 $(0,0)$ から微小に $(\varepsilon_1, \varepsilon_2)$ だけずらした点での値は,

$$\varepsilon_1^4 + \varepsilon_2^4 - (\varepsilon_1 + \varepsilon_2)^2$$

である．もし $-\sqrt{2} < \varepsilon_1 = \varepsilon_2 < \sqrt{2}$ であれば,

$$2\varepsilon_1^4 - 4\varepsilon_1^2 = 2\varepsilon_1^2(\varepsilon_1^2 - 2) < 0$$

となるが, $\varepsilon_1 = -\varepsilon_2$ であれば,

$$2\varepsilon_1^4 > 0$$

となり, 鞍点であることが分かる.)

**点 $(\pm 1, \pm 1)$ について**．Hesse 行列は

$$\begin{pmatrix} 10 & -2 \\ -2 & 10 \end{pmatrix}.$$

であり, 首座小行列式は

$$\det(10) = 10, \quad \det \begin{pmatrix} 10 & -2 \\ -2 & 10 \end{pmatrix} = 96 > 0$$

となる．よって, Hesse 行列は正定値となり, 関数 $f$ は点 $(\pm 1, \pm 1)$ で極小値をとる．そのときの値はいずれも, $-2$ である.

以上より, 点 $(\pm 1, \pm 1)$ で関数 $f$ は最小値 $-2$ をとる. □

**問題 4 (行列演算)**
行列 $A, B, C$ が

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 2 & 0 & 2 \end{pmatrix}, \quad C = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix},$$

ベクトル $\boldsymbol{x}, \boldsymbol{y}$ が

$$\boldsymbol{x} = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \quad \boldsymbol{y} = \begin{pmatrix} 2 \\ 1 \\ 0 \\ 1 \end{pmatrix}$$

となっています．以下の計算をしてください.

(1) $(A^\top + 2B)\boldsymbol{y}$.

(2) $C^{-1}(C+I_3)C$.

(3) $AB$.

(4) $B^\top A^\top B^\top$.

(5) $\boldsymbol{x}^\top A^\top A\boldsymbol{x}$.

**問題の解答 4**

(1) まず,

$$A^\top = \begin{pmatrix} 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

である．$(A^\top + 2B)\boldsymbol{y} = A^\top\boldsymbol{y} + 2B\boldsymbol{y}$ なので, 各項を計算する．

$$A^\top\boldsymbol{y} = \begin{pmatrix} 4 \\ 2 \\ 2 \end{pmatrix}, \quad 2B\boldsymbol{y} = \begin{pmatrix} 8 \\ 6 \\ 8 \end{pmatrix}$$

なので,

$$(A^\top + 2B)\boldsymbol{y} = \begin{pmatrix} 4 \\ 2 \\ 2 \end{pmatrix} + \begin{pmatrix} 8 \\ 6 \\ 8 \end{pmatrix} = \begin{pmatrix} 12 \\ 8 \\ 10 \end{pmatrix}.$$

(2) まずは展開する．$C^{-1}C = I_3$ や $I_3C = C$ に気をつければ, $C^{-1}(C+I_3)C = C^{-1}CC + C^{-1}I_3C = I_3C + C^{-1}C = C + I_3$ となる．よって,

$$C^{-1}(C+I_3)C = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 1 & 0 \\ 1 & 3 & 1 \\ 0 & 1 & 3 \end{pmatrix}.$$

(3)

$$AB = \begin{pmatrix} 2 & 2 & 0 & 2 \\ 5 & 1 & 1 & 0 \\ 4 & 4 & 2 & 2 \\ 1 & 1 & 1 & 0 \end{pmatrix}.$$

(4) $B^\top A^\top B^\top = (BAB)^\top$ である．(3) の結果を使うと,

$$BAB = B(AB) = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 2 & 0 & 2 \end{pmatrix} \begin{pmatrix} 2 & 2 & 0 & 2 \\ 5 & 1 & 1 & 0 \\ 4 & 4 & 2 & 2 \\ 1 & 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 4 & 4 & 0 & 4 \\ 11 & 7 & 3 & 4 \\ 12 & 4 & 4 & 0 \end{pmatrix}.$$

よって,

$$B^\top A^\top B^\top = \begin{pmatrix} 4 & 4 & 0 & 4 \\ 11 & 7 & 3 & 4 \\ 12 & 4 & 4 & 0 \end{pmatrix}^\top = \begin{pmatrix} 4 & 11 & 12 \\ 4 & 7 & 4 \\ 0 & 3 & 4 \\ 4 & 4 & 0 \end{pmatrix}.$$

(5) $\boldsymbol{x}^\top A^\top A\boldsymbol{x} = (A\boldsymbol{x})^\top A\boldsymbol{x} (= \|A\boldsymbol{x}\|)$ なので, $A\boldsymbol{x}$ を計算する.

$$A\boldsymbol{x} = \begin{pmatrix} 3 \\ 2 \\ 3 \\ 0 \end{pmatrix}$$

なので,

$$\boldsymbol{x}^\top A^\top A\boldsymbol{x} = 22$$

となる[*3]. □

---

**問題 5 (内積とノルム)**
ベクトル $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3$ が

$$\boldsymbol{x}_1 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \quad \boldsymbol{x}_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad \boldsymbol{x}_3 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$$

となっています. 以下の計算をしてください.

(1) $\ell_1 = \|\boldsymbol{x}_1\|$.
(2) $\boldsymbol{v}_1 = \ell_1^{-1}\boldsymbol{x}_1$.
(3) $\ell_2 = \|\boldsymbol{x}_2 - \langle \boldsymbol{x}_2, \boldsymbol{v}_1\rangle \boldsymbol{v}_1\|$.
(4) $\boldsymbol{v}_2 = \ell_2^{-1}(\boldsymbol{x}_2 - \langle \boldsymbol{x}_2, \boldsymbol{v}_1\rangle \boldsymbol{v}_1)$.
(5) $\ell_3 = \|\boldsymbol{x}_3 - \langle \boldsymbol{x}_3, \boldsymbol{v}_1\rangle \boldsymbol{v}_1 - \langle \boldsymbol{x}_3, \boldsymbol{v}_2\rangle \boldsymbol{v}_2\|$.
(6) $\boldsymbol{v}_3 = \ell_3^{-1}(\boldsymbol{x}_3 - \langle \boldsymbol{x}_3, \boldsymbol{v}_1\rangle \boldsymbol{v}_1 - \langle \boldsymbol{x}_3, \boldsymbol{v}_2\rangle \boldsymbol{v}_2)$.
(7) $\langle \boldsymbol{v}_1, \boldsymbol{v}_2\rangle, \quad \langle \boldsymbol{v}_2, \boldsymbol{v}_3\rangle, \quad \langle \boldsymbol{v}_3, \boldsymbol{v}_1\rangle$.

---

**問題の解答 5**

(1) $\ell_1 = \sqrt{1+1+1} = \sqrt{3}$.
(2)

$$\boldsymbol{v}_1 = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}.$$

(3) $\langle \boldsymbol{x}_2, \boldsymbol{v}_1\rangle = 0$ なので, $\boldsymbol{x}_2 - \langle \boldsymbol{x}_2, \boldsymbol{v}_1\rangle \boldsymbol{v}_1 = \boldsymbol{x}_2$. よって, $\ell_2 = \|\boldsymbol{x}_2\| = \sqrt{1+1+0} = \sqrt{2}$.
(4)

$$\boldsymbol{v}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}.$$

[*3] 本来は (22) と $1 \times 1$ 行列になるが, ふつうはこれを実数とみなす (同一視という).

(5) $\langle \boldsymbol{x}_3, \boldsymbol{v}_1 \rangle = 0$, $\langle \boldsymbol{x}_3, \boldsymbol{v}_2 \rangle = \frac{1}{\sqrt{2}}$ なので,

$$\ell_3 = \| \boldsymbol{x}_3 - \langle \boldsymbol{x}_3, \boldsymbol{v}_2 \rangle \boldsymbol{v}_2 \| = \left\| \begin{pmatrix} 1/2 \\ -1/2 \\ -1 \end{pmatrix} \right\| = \sqrt{\frac{1}{4} + \frac{1}{4} + 1} = \frac{\sqrt{6}}{2}.$$

(6)

$$\boldsymbol{v}_3 = \frac{2}{\sqrt{6}} \begin{pmatrix} 1/2 \\ -1/2 \\ -1 \end{pmatrix} = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ -1 \\ -2 \end{pmatrix}.$$

(7) $\langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle = \langle \boldsymbol{v}_2, \boldsymbol{v}_3 \rangle = \langle \boldsymbol{v}_3, \boldsymbol{v}_1 \rangle = 0.$ □

**問題 6 (対角化)**

行列 $A$ が

$$A = \begin{pmatrix} 4 & -1 & 1 \\ -1 & 4 & -1 \\ 1 & -1 & 4 \end{pmatrix}$$

となっています. 以下の問に答えてください.

(1) 行列 $A$ のすべての固有値を求めてください.

(2) (1) で求めたそれぞれの固有値に対応する固有ベクトル $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ を互いに正規直交するように選んでください.

(3) 行列 $P = (\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ が直交行列であることを確認してください.

(4) 行列 $A$ を直交行列 $P$ によって対角化してください.

**問題の解答 6**

(1) 行列 $A$ の固有方程式を求める. 行列の各行/列を定数倍して別の行/列に加えても行列式が変わらないことと, Sarrus の公式を利用して,

$$\begin{aligned} \det \begin{pmatrix} 4-\lambda & -1 & 1 \\ -1 & 4-\lambda & -1 \\ 1 & -1 & 4-\lambda \end{pmatrix} &= \det \begin{pmatrix} 3-\lambda & 3-\lambda & 0 \\ -1 & 4-\lambda & -1 \\ 0 & 3-\lambda & 3-\lambda \end{pmatrix} \\ &= (3-\lambda)^2(4-\lambda) + (3-\lambda)^2 + (3-\lambda)^2 \\ &= (3-\lambda)^2(6-\lambda). \end{aligned}$$

よって, 固有方程式は

$$(3-\lambda)^2(6-\lambda) = 0$$

となり, 固有値 $\lambda = 6, 3$ を得る.

(2) 解くべき連立方程式は, $(A - \lambda I_3)\boldsymbol{v} = \boldsymbol{0}$ である.

**$\lambda=6$ に対応する固有ベクトル** 連立方程式の係数行列は

$$A-\lambda I_3=\begin{pmatrix}-2 & -1 & 1\\ -1 & -2 & -1\\ 1 & -1 & -2\end{pmatrix}$$

なので, 行に関する基本変形[*4]を行えば

$$\begin{pmatrix}-1 & 0 & 1\\ 1 & 1 & 0\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}v_1\\ v_2\\ v_3\end{pmatrix}=\begin{pmatrix}0\\ 0\\ 0\end{pmatrix}$$

を解けばよいことが分かる. これより, $v_1=v_3,\quad v_1=-v_2$ となるので, 例えば

$$\tilde{\boldsymbol{v}}_1=\begin{pmatrix}1\\ -1\\ 1\end{pmatrix}$$

などととる.

**$\lambda=3$ に対応する固有ベクトル** 連立方程式の係数行列は

$$A-\lambda I_3=\begin{pmatrix}1 & -1 & 1\\ -1 & 1 & -1\\ 1 & -1 & 1\end{pmatrix}$$

なので, 行に関する基本変形を行えば

$$\begin{pmatrix}1 & -1 & 1\\ 0 & 0 & 0\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}v_1\\ v_2\\ v_3\end{pmatrix}=\begin{pmatrix}0\\ 0\\ 0\end{pmatrix}$$

を解けばよいことが分かる. これより, $v_1-v_2+v_3=0$ となるので, 例えば

$$\tilde{\boldsymbol{v}}_2=\begin{pmatrix}1\\ 1\\ 0\end{pmatrix},\quad \tilde{\boldsymbol{v}}_3=\begin{pmatrix}1\\ 0\\ -1\end{pmatrix}$$

などととる.

最後に, これらのベクトルを Gram–Schmidt の正規直交化法により正規直交ベクトルに変形すればよいが, 実は既に**問題 5** で計算していた! よって,

$$\boldsymbol{v}_1=\frac{1}{\sqrt{3}}\begin{pmatrix}1\\ -1\\ 1\end{pmatrix},\quad \boldsymbol{v}_2=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\ 1\\ 0\end{pmatrix},\quad \boldsymbol{v}_3=\frac{1}{\sqrt{6}}\begin{pmatrix}1\\ -1\\ -2\end{pmatrix}.$$

(3) 行列積を明示的に計算してもよいが, ここでは $\boldsymbol{v}_1,\boldsymbol{v}_2,\boldsymbol{v}_3$ が正規直交ベクトルであることを利用しよう.

$$P^\top P=\begin{pmatrix}\boldsymbol{v}_1^\top\\ \boldsymbol{v}_2^\top\\ \boldsymbol{v}_3^\top\end{pmatrix}(\boldsymbol{v}_1,\boldsymbol{v}_2,\boldsymbol{v}_3)=\begin{pmatrix}\boldsymbol{v}_1^\top\boldsymbol{v}_1 & \boldsymbol{v}_1^\top\boldsymbol{v}_2 & \boldsymbol{v}_1^\top\boldsymbol{v}_3\\ \boldsymbol{v}_2^\top\boldsymbol{v}_1 & \boldsymbol{v}_2^\top\boldsymbol{v}_2 & \boldsymbol{v}_2^\top\boldsymbol{v}_3\\ \boldsymbol{v}_3^\top\boldsymbol{v}_1 & \boldsymbol{v}_3^\top\boldsymbol{v}_2 & \boldsymbol{v}_3^\top\boldsymbol{v}_3\end{pmatrix}=\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}=I_3.$$

[*4] ここでの目的は連立方程式を解くことなので, 列に関する基本変形は行ってはいけないことに注意する.

よって, $P$ は確かに直交行列.

(4) $A$ を $P$ により対角化する.

$$AP = P\begin{pmatrix} 6 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

なので, 左から $P^\top$ をかけて,

$$P^\top AP = \begin{pmatrix} 6 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

□

**問題 7 (二項分布, 幾何分布)**

表が $\frac{1}{3}$ の確率で, 裏が $\frac{2}{3}$ の確率で出る歪んだコインがあります. 以下の問いに答えてください.

(1) コイン投げを $n$ 回行って, 表が出る回数を $X$ とします. 確率変数 $X$ の確率質量関数 $p_X(x)$ を求めてください.

(2) コイン投げを表が初めて出るまで行って, 表が出るまでにコインを投げた回数を $Y$ とします. 確率変数 $Y$ の確率質量関数 $p_Y(y)$ を求めてください.

**問題の解答 7**

(1) $n$ 回の試行のうち, $x$ 回が表, $n-x$ 回が裏になる場合の数は,

$$\binom{n}{x} \; (= {}_nC_x)$$

通りである[*5]. そのそれぞれの場合の状況が起きる確率は,

$$\left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{n-x}$$

であるから, 求める確率密度関数は

$$p_X(x) = \binom{n}{x}\left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{n-x}.$$

(2) はじめ $y-1$ 回が連続で裏になり, 次の 1 回が表になる確率が求める確率密度関数であるので

$$p_Y(y) = \left(\frac{1}{3}\right)\left(\frac{2}{3}\right)^{y-1}.$$

□

[*5] 高校では ${}_nC_x$ の表記をしていたと思いますが, $\binom{n}{x}$ の表記が一般的のようです.

**問題 8 (指数分布)**
分布関数が, $\lambda > 0$ を定数として

$$F(x) = \begin{cases} 1 - \mathrm{e}^{-\lambda x} & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$

で与えられる分布の確率密度関数 $f(x)$ を求めてください.

**問題の解答 8**
確率密度関数は分布関数を微分すればよく,

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

となる. □

**問題 9 (期待値と分散)**
以下の問に答えてください.

(1) 一般の確率変数 $X$ の分散について, $\mathbb{V}[X] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$ が成立することを証明してください.

(2) 確率変数 $X$ が, 確率密度関数

$$f_X(x) = \begin{cases} 1 & (0 \leq x \leq 1) \\ 0 & (x < 0,\ 1 < x) \end{cases}$$

の分布に従っているとします. 確率変数 $X$ の期待値と分散を求めてください.

(3) 確率変数 $Y$ が, 確率密度関数

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right)$$

の分布に従っているとします. 確率変数 $Y$ の期待値と分散を求めてください.

**問題の解答 9**

(1) 分散の定義式の展開をして, 期待値の線形性を利用すればよい. 確率変数 $X$ の平均を

$\mu = \mathbb{E}[X]$ とおく.

$$
\begin{aligned}
\mathbb{V}[X] &= \mathbb{E}\left[(X-\mu)^2\right] \\
&= \mathbb{E}\left[X^2 - 2\mu X + \mu^2\right] \\
&= \mathbb{E}\left[X^2\right] - 2\mu\mathbb{E}[X] + \mu^2 \\
&= \mathbb{E}\left[X^2\right] - 2\mu^2 + \mu^2 \\
&= \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 .
\end{aligned}
$$

(2) 定義どおり計算すればよい. 平均は,

$$
\mathbb{E}[X] = \int_0^1 x \cdot 1 \,\mathrm{d}x = \left[\frac{1}{2}x^2\right]_0^1 = \frac{1}{2}.
$$

分散は,

$$
\mathbb{V}[X] = \int_0^1 \left(x - \frac{1}{2}\right)^2 \cdot 1 \,\mathrm{d}x = \left[\frac{1}{3}\left(x - \frac{1}{2}\right)^3\right]_0^1 = \frac{1}{12}.
$$

(3) これも定義どおり計算すればよい. 変数変換 $t = \frac{1}{2}y^2$ を利用して,

$$
\begin{aligned}
\mathbb{E}[Y] &= \int_{-\infty}^{\infty} y \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right) \mathrm{d}y \\
&= \frac{1}{\sqrt{2\pi}} \left(\int_{-\infty}^{0} y \exp\left(-\frac{1}{2}y^2\right) \mathrm{d}y + \int_0^{\infty} y \exp\left(-\frac{1}{2}y^2\right) \mathrm{d}y\right) \\
&= \frac{1}{\sqrt{2\pi}} \left(-\int_0^{\infty} \mathrm{e}^{-t} \,\mathrm{d}y + \int_0^{\infty} \mathrm{e}^{-t} \,\mathrm{d}y\right) \\
&= \frac{1}{\sqrt{2\pi}} (-1+1) \\
&= 0.
\end{aligned}
$$

分散は,

$$
\mathbb{V}[Y] = \int_{-\infty}^{\infty} y^2 \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right) \mathrm{d}y = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 \exp\left(-\frac{1}{2}y^2\right) \mathrm{d}y.
$$

ここで, 最終項の積分は部分積分と Gauss 積分を利用して,

$$
\begin{aligned}
\int_{-\infty}^{\infty} y^2 \exp\left(-\frac{1}{2}y^2\right) \mathrm{d}y &= -\int_{-\infty}^{\infty} y \left(\exp\left(-\frac{1}{2}y^2\right)\right)' \mathrm{d}y \\
&= -\left[y \exp\left(-\frac{1}{2}y^2\right)\right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}y^2\right) \mathrm{d}y \\
&= -0 + \sqrt{2\pi} \\
&= \sqrt{2\pi}.
\end{aligned}
$$

よって,

$$
\mathbb{V}[Y] = \frac{1}{\sqrt{2\pi}} \cdot \sqrt{2\pi} = 1.
$$

□

**問題 10 (独立性, 標本平均と標本分散)**
確率変数 $X_1, \dots, X_n$ は i.i.d. であるとします. 次の問に答えてください.

(1) 異なる確率変数 $X_i, X_j$ $(i \neq j)$ の共分散について, $\mathrm{Cov}\left[X_i, X_j\right] = \mathbb{E}\left[X_i X_j\right] - \mathbb{E}\left[X_i\right]\mathbb{E}\left[X_j\right]$ であることを証明してください.

(2) 確率変数の積の期待値について, $\mathbb{E}\left[X_1 \cdots X_n\right] = \mathbb{E}\left[X_1\right] \cdots \mathbb{E}\left[X_n\right]$ であることを証明してください.

(3) 確率変数の和の分散について, $\mathbb{V}\left[X_1 + \cdots + X_n\right] = \mathbb{V}\left[X_1\right] + \cdots + \mathbb{V}\left[X_n\right]$ であることを証明してください.

(4) 確率変数 $X_1$ の平均と分散が $\mathbb{E}\left[X_1\right] = \mu$, $\mathbb{V}\left[X_1\right] = \sigma^2 > 0$ のとき, 確率変数 $X_1, \dots, X_n$ の**標本平均** (sample mean)

$$\bar{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$$

の平均 $\mathbb{E}\left[\bar{X}\right]$ と分散 $\mathbb{V}\left[\bar{X}\right]$ を求めてください.

(5) 確率変数 $X_1$ の平均と分散が $\mathbb{E}\left[X_1\right] = \mu$, $\mathbb{V}\left[X_1\right] = \sigma^2 > 0$ のとき, 確率変数 $X_1, \dots, X_n$ の**標本分散** (sample variance)

$$S^2 := \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})^2$$

の平均 $\mathbb{E}\left[S^2\right]$ を求めてください.

**問題の解答 10**

(1) 定義の式を展開し, 期待値の線形性を利用して整理すればよい. $\mu_i = \mathbb{E}\left[X_i\right]$, $\mu_j = \mathbb{E}\left[X_j\right]$ とすれば, $\mu_i, \mu_j$ が定数であることに注意して,

$$\begin{aligned}
\mathrm{Cov}\left[X_i, X_j\right] &= \mathbb{E}\left[(X_i - \mu_i)(X_j - \mu_j)\right] \\
&= \mathbb{E}\left[X_i X_j - X_i \mu_j - \mu_i X_j + \mu_i \mu_j\right] \\
&= \mathbb{E}\left[X_i X_j\right] - \mathbb{E}\left[X_i\right]\mu_j - \mu_i \mathbb{E}\left[X_j\right] + \mu_i \mu_j \quad \text{(期待値の線形性)} \\
&= \mathbb{E}\left[X_i X_j\right] - \mu_i \mu_j - \mu_i \mu_j + \mu_i \mu_j \\
&= \mathbb{E}\left[X_i X_j\right] - \mathbb{E}\left[X_i\right]\mathbb{E}\left[X_j\right].
\end{aligned}$$

(2) 期待値の定義と, 確率変数 $X_1, \ldots, X_n$ が独立であることを利用すると,

$$
\begin{aligned}
\mathbb{E}\left[X_1 \cdots X_n\right] &= \int_{\mathbb{R}^n} x_1 \cdots x_n f_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \\
&= \int_{\mathbb{R}^n} x_1 \cdots x_n f_{X_1}(x_1) \cdots f_{X_n}(x_n)\,\mathrm{d}\boldsymbol{x} \quad (\text{独立の定義より}) \\
&= \left(\int_{-\infty}^{\infty} x_1 f_{X_1}(x_1)\,\mathrm{d}x_1\right) \cdots \left(\int_{-\infty}^{\infty} x_n f_{X_n}(x_n)\,\mathrm{d}x_n\right) \\
&= \mathbb{E}\left[X_1\right] \cdots \mathbb{E}\left[X_n\right].
\end{aligned}
$$

(3) まず, $X_1 + \cdots + X_n$ の平均を求める. $\mu_i = \mathbb{E}\left[X_i\right]$ とすると, 期待値の線形性より

$$
\mathbb{E}\left[X_1 + \cdots + X_n\right] = \sum_{i=1}^{n} \mu_i.
$$

また, 確率変数 $X_1, \ldots, X_n$ が独立であることから (1)(2) より,

$$
\mathrm{Cov}\left[X_i, X_j\right] = 0 \quad (i \neq j)
$$

が成立する. これより分散は,

$$
\begin{aligned}
& \mathbb{V}\left[X_1 + \cdots + X_n\right] \\
=\ & \mathbb{E}\left[\left(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i\right)^2\right] \\
=\ & \mathbb{E}\left[\left(\sum_{i=1}^{n} (X_i - \mu_i)\right)^2\right] \\
=\ & \mathbb{E}\left[\sum_{i=1}^{n} (X_i - \mu_i)^2 + 2\sum_{i<j} (X_i - \mu_i)(X_j - \mu_j)\right] \\
=\ & \sum_{i=1}^{n} \mathbb{E}\left[(X_i - \mu_i)^2\right] + 2\sum_{i<j} \mathbb{E}\left[(X_i - \mu_i)(X_j - \mu_j)\right] \quad (\text{期待値の線形性}) \\
=\ & \sum_{i=1}^{n} \mathbb{V}\left[X_i\right] + 2\sum_{i<j} \mathrm{Cov}\left[X_i, X_j\right] \\
=\ & \sum_{i=1}^{n} \mathbb{V}\left[X_i\right].
\end{aligned}
$$

なお, 記号 $\sum_{i<j}$ は $1 \leq i < j \leq n$ でとりうる $i, j$ 全てにわたる和をとることを表す.

(4) 平均は, 期待値の線形性により

$$
\mathbb{E}\left[\bar{X}\right] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[X_i\right] = \frac{1}{n} \cdot n\mu = \mu.
$$

分散は,

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

であることと (3) の結果を使うと,

$$\begin{aligned}
\mathbb{V}\left[\bar{X}\right] &= \frac{1}{n^2}\mathbb{V}\left[\sum_{i=1}^{n} X_i\right] \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}[X_i] \quad (\text{(3) の結果を利用した}) \\
&= \frac{1}{n^2}\cdot n\sigma^2 \\
&= \frac{\sigma^2}{n}.
\end{aligned}$$

(5) 期待値の線形性より,

$$\mathbb{E}\left[S^2\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[(X_i-\bar{X})^2\right] = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{E}\left[X_i^2\right] - 2\mathbb{E}\left[X_i\bar{X}\right] + \mathbb{E}\left[\bar{X}^2\right]\right).$$

まず,

$$\mathbb{V}[X_i] = \mathbb{E}\left[X_i^2\right] - \mathbb{E}[X_i]^2$$

より,

$$\mathbb{E}\left[X_i^2\right] = \mathbb{V}[X_i] + \mathbb{E}[X_i]^2 = \sigma^2 + \mu^2.$$

同様に,

$$\mathbb{V}\left[\bar{X}\right] = \mathbb{E}\left[\bar{X}^2\right] - \mathbb{E}\left[\bar{X}\right]^2$$

なので, (4) の結果より

$$\mathbb{E}\left[\bar{X}^2\right] = \mathbb{V}\left[\bar{X}\right] + \mathbb{E}\left[\bar{X}\right]^2 = \frac{\sigma^2}{n} + \mu^2.$$

そして, $\mathbb{E}\left[X_i\bar{X}\right]$ については,

$$
\begin{aligned}
\mathbb{E}\left[X_i\bar{X}\right] &= \frac{1}{n}\mathbb{E}\left[X_i\sum_{j=1}^{n}X_j\right] \\
&= \frac{1}{n}\mathbb{E}\left[X_i^2+\sum_{j\neq i}X_iX_j\right] \\
&= \frac{1}{n}\left(\mathbb{E}\left[X_i^2\right]+\sum_{j\neq i}\mathbb{E}\left[X_iX_j\right]\right) \quad \text{(期待値の線形性)} \\
&= \frac{1}{n}\left(\mathbb{E}\left[X_i^2\right]+\sum_{j\neq i}\mathbb{E}\left[X_i\right]\mathbb{E}\left[X_j\right]\right) \quad \text{((2) の結果を利用)} \\
&= \frac{1}{n}\left(\sigma^2+\mu^2+(n-1)\mu^2\right) \\
&= \frac{\sigma^2}{n}+\mu^2.
\end{aligned}
$$

以上より,

$$
\begin{aligned}
\mathbb{E}\left[S^2\right] &= \frac{1}{n}\sum_{i=1}^{n}\left(\sigma^2+\mu^2-\frac{2\sigma^2}{n}-2\mu^2+\frac{\sigma^2}{n}+\mu^2\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{(n-1)\sigma^2}{n}\right) \\
&= \frac{n-1}{n}\sigma^2.
\end{aligned}
$$

□

**発展問題 4 (標本分散の分散)**
確率変数 $X_1,\ldots,X_n$ は i.i.d. であり, 確率変数 $X_1$ の平均と分散が $\mathbb{E}\left[X_1\right]=\mu$, $\mathbb{V}\left[X_1\right]=\sigma^2>0$ とします. 確率変数 $X_1,\ldots,X_n$ の標本平均を

$$
\bar{X} := \frac{1}{n}\sum_{i=1}^{n}X_i
$$

とし, 標本分散を

$$
S^2 := \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2
$$

と書きます. 標本分散の分散を計算しましょう. 次の問に答えてください.

(1) **ガンマ関数** (gamma function) は,

$$\Gamma(x) := \int_0^\infty t^{x-1}\mathrm{e}^{-t}\,\mathrm{d}t$$

で定義されます．ガンマ関数の性質として，0 と負整数以外の実数 $x$ について $\Gamma(x+1) = x\Gamma(x)$ であることを証明してください．

(2) 自由度 $k$ の**カイ二乗分布** (chi-squared distribution) は, 確率密度関数が

$$f(x) = \begin{cases} \frac{1}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})} x^{\frac{k}{2}-1}\mathrm{e}^{-\frac{x}{2}} & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$

で与えられます．自由度 $k$ のカイ二乗分布に従う確率変数 $X$ の平均 $\mathbb{E}[X]$ と分散 $\mathbb{V}[X]$ を計算してください．

(3) 事実として, $\frac{nS^2}{\sigma^2}$ が自由度 $n-1$ のカイ二乗分布に従うことが知られています．これを用いて, 標本分散 $S^2$ の分散 $\mathbb{V}\left[S^2\right]$ を求めてください．

**発展問題の解答 4**

(1) 部分積分をすることにより,

$$\begin{aligned}
\Gamma(x+1) &= \int_0^\infty t^x\mathrm{e}^{-t}\,\mathrm{d}t \\
&= \left[-t^{x+1-1}\mathrm{e}^{-t}\right]_0^\infty + \int_0^\infty xt^{x-1}\mathrm{e}^{-t}\,\mathrm{d}t \\
&= 0 + x\int_0^\infty t^{x-1}\mathrm{e}^{-t}\,\mathrm{d}t \\
&= x\Gamma(x).
\end{aligned}$$

(2) 定義どおり計算する．簡略化のため,

$$C := \frac{1}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)}$$

とおく. 平均は,

$$
\begin{aligned}
\mathbb{E}[X] &= \int_0^\infty x \cdot C x^{\frac{k}{2}-1} \mathrm{e}^{-\frac{x}{2}} \, \mathrm{d}x \\
&= C \int_0^\infty x^{\frac{k}{2}} \mathrm{e}^{-\frac{x}{2}} \, \mathrm{d}x \\
&= C \int_0^\infty (2t)^{\frac{k}{2}} \mathrm{e}^{-t} \cdot 2 \, \mathrm{d}t \quad \left(t = \frac{x}{2} \text{と置換}\right) \\
&= 2^{\frac{k}{2}+1} C \int_0^\infty t^{\frac{k}{2}} \mathrm{e}^{-t} \, \mathrm{d}t \\
&= 2 \left(\Gamma\left(\frac{k}{2}\right)\right)^{-1} \Gamma\left(\frac{k}{2}+1\right) \\
&= 2 \left(\Gamma\left(\frac{k}{2}\right)\right)^{-1} \cdot \frac{k}{2} \Gamma\left(\frac{k}{2}\right) \quad \text{((1) の結果を利用)} \\
&= k.
\end{aligned}
$$

分散は, $\mathbb{V}[X] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$ を利用して求める. $\mathbb{E}\left[X^2\right]$ は,

$$
\begin{aligned}
\mathbb{E}\left[X^2\right] &= \int_0^\infty x^2 \cdot C x^{\frac{k}{2}-1} \mathrm{e}^{-\frac{x}{2}} \, \mathrm{d}x \\
&= C \int_0^\infty x^{\frac{k}{2}+1} \mathrm{e}^{-\frac{x}{2}} \, \mathrm{d}x \\
&= C \int_0^\infty (2t)^{\frac{k}{2}+1} \mathrm{e}^{-t} \cdot 2 \, \mathrm{d}t \quad \left(t = \frac{x}{2} \text{と置換}\right) \\
&= 2^{\frac{k}{2}+2} C \int_0^\infty t^{\frac{k}{2}+1} \mathrm{e}^{-t} \, \mathrm{d}t \\
&= 2^2 \left(\Gamma\left(\frac{k}{2}\right)\right)^{-1} \Gamma\left(\frac{k}{2}+2\right) \\
&= 2^2 \left(\Gamma\left(\frac{k}{2}\right)\right)^{-1} \cdot \left(\frac{k}{2}+1\right) \cdot \frac{k}{2} \Gamma\left(\frac{k}{2}\right) \quad \text{((1) の結果を繰り返し利用)} \\
&= k(k+2) \\
&= k^2 + 2k.
\end{aligned}
$$

よって, 分散は

$$\mathbb{V}[X] = k^2 + 2k - k^2 = 2k.$$

(3) (2) の結果を利用して,

$$\mathbb{V}\left[\frac{nS^2}{\sigma^2}\right] = 2(n-1).$$

分散の性質より,

$$\mathbb{V}\left[\frac{nS^2}{\sigma^2}\right] = \frac{n^2}{\sigma^4} \mathbb{V}\left[S^2\right]$$

なので,

$$\mathbb{V}\left[S^2\right] = \frac{2(n-1)\sigma^4}{n^2}.$$

□

**問題 11 (単回帰)**
ある変数 $x \in \mathbb{R}$ と別の変数 $y \in \mathbb{R}$ には $y = w_1 x + w_0$ という関係があることがわかっており, 係数 $w_1, w_0$ を統計的に決定するために, $n$ 個のサンプルからなるデータセット

$$\mathcal{D} = \{\, (x_1, y_1), \ldots, (x_n, y_n) \,\}$$

を取得しました. 最小二乗法により, 係数 $w_1, w_0$ を決定してください.

**問題の解答 11**
最小化すべき関数は,

$$E_2(w_1, w_0) = \sum_{i=1}^{n} (y_i - w_1 x_i - w_0)^2$$

であり, これは $w_1, w_0$ について凸関数となっている. よって, 停留点を求めればそれが最小解になる.

$$\begin{aligned}
\frac{\partial}{\partial w_1} E_2(w_1, w_0) &= \sum_{i=1}^{n} \frac{\partial}{\partial w_1} (y_i - w_1 x_i - w_0)^2 \\
&= \sum_{i=1}^{n} 2(y_i - w_1 x_i - w_0) \cdot (-x_i) \\
&= -2 \left( \sum_{i=1}^{n} x_i y_i - \left( \sum_{i=1}^{n} x_i^2 \right) w_1 - \left( \sum_{i=1}^{n} x_i \right) w_0 \right). \\
\frac{\partial}{\partial w_0} E_2(w_1, w_0) &= \sum_{i=1}^{n} \frac{\partial}{\partial w_0} (y_i - w_1 x_i - w_0)^2 \\
&= \sum_{i=1}^{n} 2(y_i - w_1 x_i - w_0) \cdot (-1) \\
&= -2 \left( \sum_{i=1}^{n} y_i - \left( \sum_{i=1}^{n} x_i \right) w_1 - n w_0 \right).
\end{aligned}$$

ここで簡略化のため, 記法

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (x_i \text{ の標本平均})$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad (y_i \text{ の標本平均})$$

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2 \quad (x_i \text{ の標本分散})$$

$$s_{xy} = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \quad (x_i, y_i \text{ の標本共分散})$$

を導入する. これより,

$$\sum_{i=1}^{n} x_i = n\bar{x}$$

$$\sum_{i=1}^{n} y_i = n\bar{y}$$

$$\sum_{i=1}^{n} x_i^2 = ns_x^2 + n\bar{x}^2$$

$$\sum_{i=1}^{n} x_i y_i = ns_{xy} + n\bar{x}\bar{y}$$

と書ける. 解くべき方程式は,

$$\begin{cases} (s_x^2 + \bar{x}^2)w_1 + \bar{x}w_0 & = & s_{xy} + \bar{x}\bar{y} \\ \bar{x}w_1 + w_0 & = & \bar{y} \end{cases}$$

である. 第 2 式より $w_0 = \bar{y} - \bar{x}w_1$ を第 1 式に代入して整理すれば,

$$w_1 = \frac{s_{xy}}{s_x^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

これより,

$$w_0 = \bar{y} - \frac{s_{xy}}{s_x^2}\bar{x} = \frac{1}{n}\left(\sum_{i=1}^{n} y_i - \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sum_{i=1}^{n} x_i\right).$$

□

**発展問題 5 (重回帰)**

ある変数 $\boldsymbol{x} \in \mathbb{R}^d$ と別の変数 $y \in \mathbb{R}$ には $y = \boldsymbol{w}^\top \boldsymbol{x}$ という関係があることがわかっており, 係数 $\boldsymbol{w}$ を統計的に決定するために, $n$ 個のサンプルからなるデータセット

$$\mathcal{D} = \{\, (\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n) \,\}$$

を取得しました. データを並べて得られる行列とベクトルを

$$X = \begin{pmatrix} \boldsymbol{x}_1^\top \\ \vdots \\ \boldsymbol{x}_n^\top \end{pmatrix}, \quad \boldsymbol{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

とし, 係数 $\boldsymbol{w}$ を

$$E_2(\boldsymbol{w}) = \|\boldsymbol{y} - X\boldsymbol{w}\|^2$$

を最小化することで決定します (最小二乗法). 次の問に答えてください.

(1) 行列 $X^\top X$ が対称行列であることを確認してください.

(2) 目的関数 $E_2(\boldsymbol{w})$ を最小にする $\boldsymbol{w}^*$ が**正規方程式** (normal equation)

$$X^\top X\boldsymbol{w}^* = X^\top \boldsymbol{y}$$

を満たすことを示してください. さらに, 行列 $X^\top X$ が正則であるとき, $\boldsymbol{w}^*$ はどのように表されるでしょうか.

(3) 目的関数 $E_2(\boldsymbol{w})$ の代わりに, $\lambda > 0$ を定数として

$$\tilde{E}_2(\boldsymbol{w}) = \|\boldsymbol{y} - X\boldsymbol{w}\|^2 + \lambda \|\boldsymbol{w}\|^2$$

を最小化することで $\boldsymbol{w}$ を推定することを考えます. このとき, 目的関数 $\tilde{E}_2(\boldsymbol{w})$ を最小にする $\hat{\boldsymbol{w}}$ はどのような関係式を満たすでしょうか. なお, 関数 $\tilde{E}_2(\boldsymbol{w})$ が凸関数であることを証明なしに利用しても構いません.

**発展問題の解答 5**

(1) 定義どおり $(X^\top X)^\top = X^\top X$ であることを確認すればよい. 転置の性質を利用すれば,

$$(X^\top X)^\top = X^\top (X^\top)^\top = X^\top X.$$

(2) $E_2(\boldsymbol{w})$ は $\boldsymbol{w}$ に関して凸関数であるから, これを最小化する $\boldsymbol{w}^*$ は $E_2(\boldsymbol{w})$ の停留点である. $E_2(\boldsymbol{w})$ を展開すると,

$$\boldsymbol{y}^\top X\boldsymbol{w} = \boldsymbol{y}^\top (X\boldsymbol{w}) = (X\boldsymbol{w})^\top \boldsymbol{y} = \boldsymbol{w}^\top X^\top \boldsymbol{y}$$

や微分公式を利用することで,

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{w}} E_2(\boldsymbol{w}) &= \frac{\partial}{\partial \boldsymbol{w}} \left((\boldsymbol{y} - X\boldsymbol{w})^\top (\boldsymbol{y} - X\boldsymbol{w})\right) \\
&= \frac{\partial}{\partial \boldsymbol{w}} \left(\boldsymbol{y}^\top \boldsymbol{y} - \boldsymbol{y}^\top X\boldsymbol{w} - \boldsymbol{w}^\top X^\top \boldsymbol{y} + \boldsymbol{w}^\top X^\top X\boldsymbol{w}\right) \\
&= \frac{\partial}{\partial \boldsymbol{w}} \left(\boldsymbol{y}^\top \boldsymbol{y}\right) - 2\frac{\partial}{\partial \boldsymbol{w}} \left(\boldsymbol{y}^\top X\boldsymbol{w}\right) + \frac{\partial}{\partial \boldsymbol{w}} \left(\boldsymbol{w}^\top X^\top X\boldsymbol{w}\right) \\
&= -2X^\top \boldsymbol{y} + 2X^\top X\boldsymbol{w}.
\end{aligned}$$

よって, $\boldsymbol{w}^*$ は正規方程式

$$X^\top X\boldsymbol{w}^* = X^\top \boldsymbol{y}$$

を満たす. さらに, $X^\top X$ が正則ならば $(X^\top X)^{-1}$ が存在し,

$$\boldsymbol{w}^* = (X^\top X)^{-1}X^\top \boldsymbol{y}$$

と求まる.

(3) 関数 $\tilde{E}_2(\boldsymbol{w}) = E_2(\boldsymbol{w}) + \lambda \|\boldsymbol{w}\|^2$ も凸関数なので, これを最小化する $\hat{\boldsymbol{w}}$ は $\tilde{E}_2(\boldsymbol{w})$ の停留点である.

$$\frac{\partial}{\partial \boldsymbol{w}}\tilde{E}_2(\boldsymbol{w}) = \frac{\partial}{\partial \boldsymbol{w}}E_2(\boldsymbol{w}) + \lambda\frac{\partial}{\partial \boldsymbol{w}}\left(\boldsymbol{w}^\top I_d\boldsymbol{w}\right) = -2X^\top\boldsymbol{y} + 2X^\top X\boldsymbol{w} + 2\lambda\boldsymbol{w}.$$

ゆえに, $\hat{\boldsymbol{w}}$ は

$$(X^\top X + \lambda I_d)\hat{\boldsymbol{w}} = X^\top \boldsymbol{y}$$

を満たす. 行列 $X^\top X$ が非負定値対称行列であることから, すべての固有値は 0 以上である. この行列に $\lambda I_d$ を加えると, $\lambda > 0$ であることからすべての固有値が 0 より大きくなり, $\det(X^\top X + \lambda I_d) > 0$ が言える. よって, 実は行列 $X^\top X + \lambda I_d$ はいつでも正則であり, 逆行列が存在する. したがって,

$$\hat{\boldsymbol{w}} = (X^\top X + \lambda I_d)^{-1}X^\top \boldsymbol{y}$$

となる. □