

数学的準備

多変量解析 - 第2講

村田 昇

講義の内容

- 確率
 - 確率分布
 - 確率質量関数・確率密度関数
 - 正規分布 (χ^2 分布, t 分布, F 分布)
- 統計
 - 統計量 (標本平均, 不偏分散・共分散, 相関係数)
 - 最尤法 (尤度関数)
 - Bayes の定理
- 関数の微分
 - ベクトルによる微分
 - 行列による微分

確率

確率分布

- 定義

注目する事象 (標本空間の部分集合) に対して, それが起きる確率 (区間 $[0, 1]$ の実数) を返す関数

$$P(\text{事象}) = \text{確率値}$$

を **確率分布** という.

- 応用上重要な分布
 - 離散分布
 - 連続分布 (絶対連続な分布)

確率質量関数

- 離散分布を記述する方法

1つの見本点 x からなる事象 (根元事象という) を $A = \{x\}$ とする. 事象 A の起きる確率

$$P(A = \{x\}) = p(x)$$

を表す関数 p を **確率質量関数** という.

確率密度関数

- 連続分布を記述する方法

事象 A が起きる確率は**確率密度関数** p の積分

$$P(A) = \int_A p(x) dx$$

で表される.

- 事象 A が十分小さな集合の場合

A に含まれる適当な点を x とし, A の大きさ (考える空間により体積や面積に相当) を $|A|$ と書くことにすれば, 事象 A の起きる確率を密度と事象の大きさの積

$$P(A) = \int_A p(x) dx \simeq p(x) \cdot |A|$$

で近似することができる.

正規分布 (normal/Gaussian distribution)

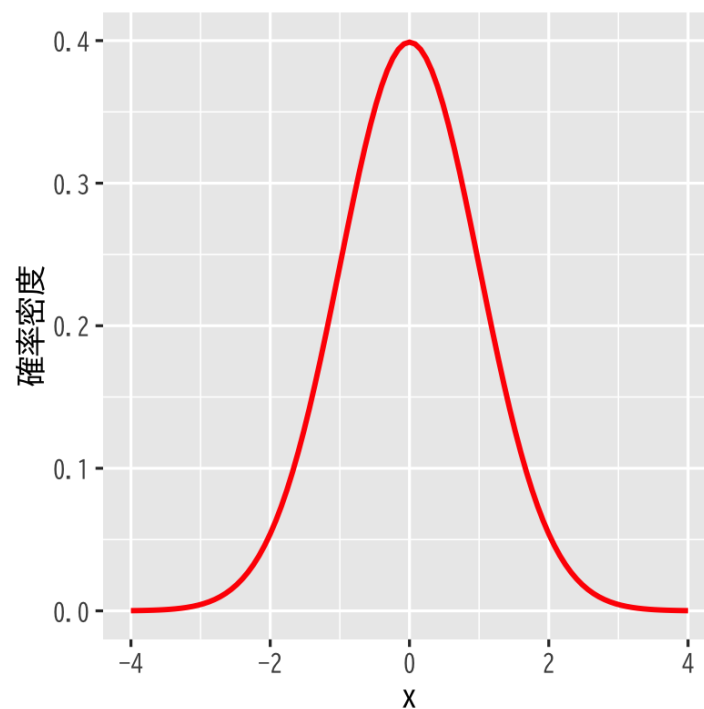


図 1: 正規分布 (平均 0, 分散 1)

- 標本空間: $(-\infty, \infty)$
- 母数: 平均 μ , 分散 σ^2
- 密度関数:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- 備考: $\mu = 0, \sigma = 1$ のとき**標準正規分布**と呼ぶ.

多次元正規分布

- 標本空間: \mathbb{R}^P
- 母数: 平均 μ , 分散共分散行列 Σ
- 密度関数:

$$p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^P |\Sigma|}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^\top \Sigma^{-1}(\mathbf{x}-\mu)}$$

正規分布の特徴付け

- 性質
 - さまざまな誤差の集積は正規分布となる (中心極限定理)
 - 同じ分散 (ばらつき) を持つ分布の中で最も情報量 (エントロピー) が大きい
- 用途
 - 誤差の分布に関する知識がないときには正規分布だと考えておくとなかなか安全なことが多い
 - 多変量解析の多くの手法は誤差の分布に関して正規性を仮定して導出する

χ^2 分布 (χ^2 -distribution)

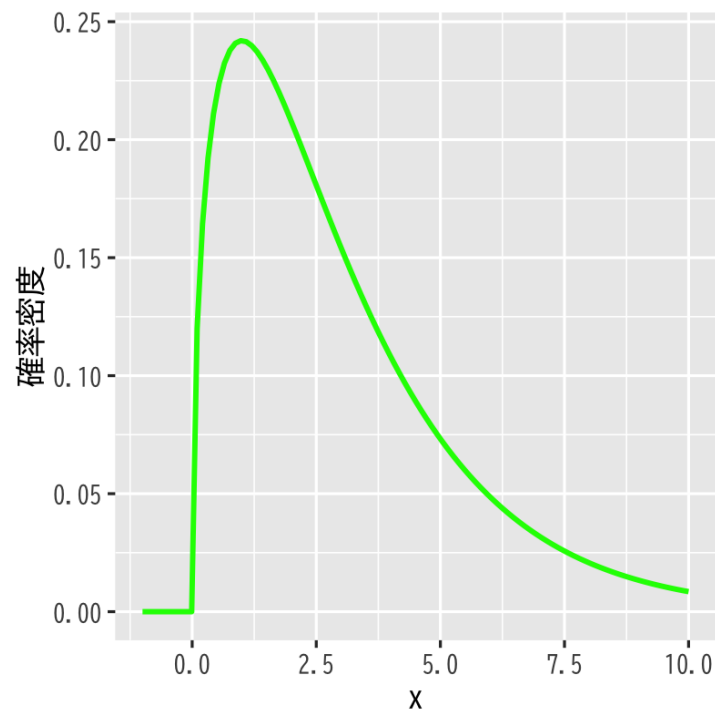


図 2: χ^2 分布 (自由度 3)

- 標本空間: $[0, \infty)$
- 母数: 自由度 ν
- 密度関数:

$$p(x) = \frac{1}{2^{\nu/2} \Gamma(\frac{\nu}{2})} x^{\nu/2-1} e^{-x/2}$$
$$\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt$$

- 備考：標準正規分布に従う独立な確率変数の 2 乗和の分布

χ^2 分布の特徴付け

- 性質
 - 標準正規分布に従う独立な確率変数の 2 乗和の分布

$$Z = \sum_{i=1}^{\nu} X_i^2 \sim \chi^2(\nu) \quad (\text{自由度 } \nu \text{ の } \chi^2 \text{ 分布})$$

$$X_i \sim \mathcal{N}(0, 1) \quad (i = 1, \dots, \nu) \quad (\text{標準正規分布})$$

- 正規分布に従う確率変数の不偏分散は χ^2 分布に従う確率変数の定数倍となる
- 用途
 - 分散の大きさに関する検定に用いられる
 - 独立性検定・適合度検定などの検定でも用いられる

t 分布 (Student's t -distribution)

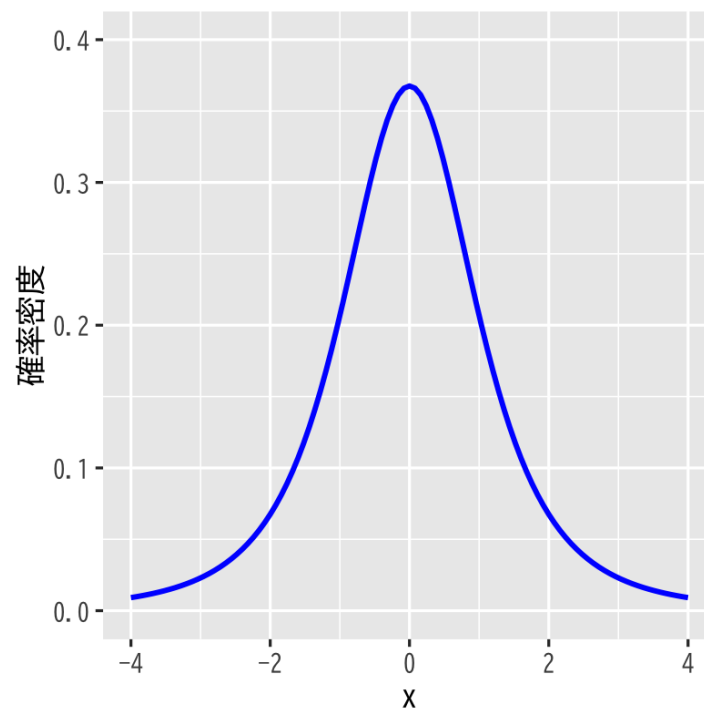


図 3: t 分布 (自由度 3)

- 標本空間: $(-\infty, \infty)$
- 母数: 自由度 ν
- 密度関数:

$$p(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{1}{2}(\nu+1)}$$

- 備考：標準正規分布と χ^2 分布に従う独立な確率変数の比の分布

t 分布の特徴付け

- 性質
 - 標準正規分布と χ^2 分布に従う独立な確率変数の比の分布

$$Z = \frac{X}{\sqrt{Y/\nu}} \sim \mathcal{T}(\nu) \quad (\text{自由度 } \nu \text{ の } t \text{ 分布})$$
$$X \sim \mathcal{N}(0, 1), \quad Y \sim \chi^2(\nu)$$

- 正規分布に従う確率変数の標本平均と真の値の差は不偏分散で正規化すると t 分布に従う確率変数となる
- 用途
 - 平均値を推定する問題の検定に利用される
 - 信頼区間の構成に利用される

F 分布 (F -distribution)

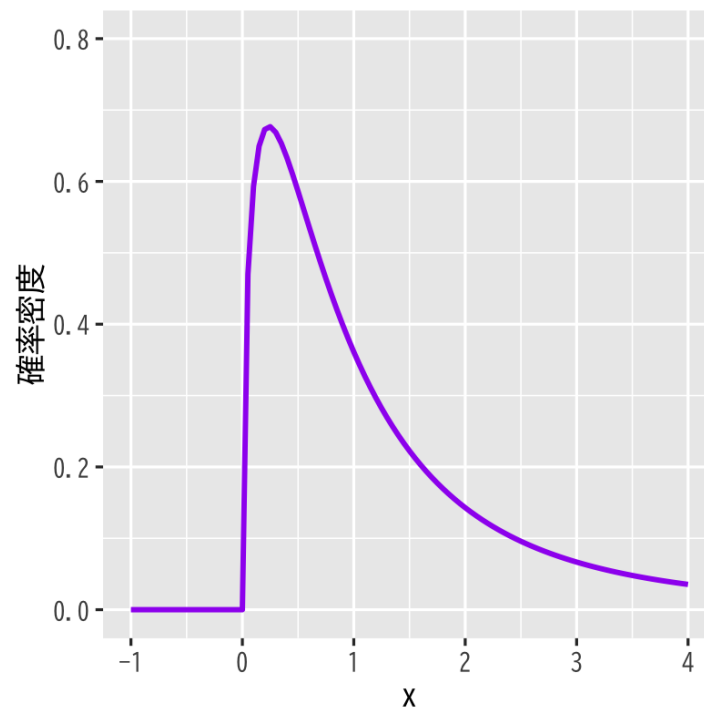


図 4: F 分布 (自由度 3, 5)

- 標本空間: $[0, \infty)$
- 母数: 自由度 ν_1, ν_2
- 密度関数:

$$p(x) = \frac{(\nu_1/\nu_2)^{\nu_1/2}}{B(\nu_1/2, \nu_2/2)} \frac{x^{\nu_1/2-1}}{(1 + \nu_1 x/\nu_2)^{(\nu_1+\nu_2)/2}}$$
$$B(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt$$

- 備考: χ^2 分布に従う独立な確率変数の比の分布

F 分布の特徴付け

- 性質
 - χ^2 分布に従う独立な確率変数の比の分布

$$Z = \frac{Y_1/\nu_1}{Y_2/\nu_2} \sim \mathcal{F}(\nu_1, \nu_2) \quad (\text{自由度 } \nu_1, \nu_2 \text{ の } F \text{ 分布})$$
$$Y_i \sim \chi^2(\nu_i) \quad (i = 1, 2)$$

- 正規分布に従う確率変数の 2 つの独立な標本の不偏分散の比は F 分布に従う確率変数となる
- 用途
 - 分散を推定する問題の検定に利用される
 - 分散の信頼区間の構成に利用される

演習

問題

- 標準正規分布の密度関数

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

を \mathbb{R} 上で積分すると 1 となることを確かめよ

- 標準正規分布に従う確率変数を X とする. X が 0 付近の値をとる確率と 1 付近の値をとる確率の比を求めよ

解答例

- 2 つの標準正規分布の積を考えて, 重積分を極座標に変換すればよい

$$\begin{aligned} & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-x^2/2} e^{-y^2/2} dx dy \\ &= \int_0^{2\pi} \int_0^{\infty} e^{-r^2/2} r dr d\theta = 2\pi \int_0^{\infty} e^{-z} dz = 2\pi \end{aligned}$$

- 微小な区間 Δ を考えて, 密度を用いた近似計算を利用すればよい

$$\begin{aligned} & \frac{P(X \text{ が } 0 \text{ 付近})}{P(X \text{ が } 1 \text{ 付近})} \\ & \simeq \frac{p(0)|\Delta|}{p(1)|\Delta|} = \frac{\exp(0)}{\exp(-1/2)} = \sqrt{e} \simeq 1.65 \end{aligned}$$

統計

記述統計量

- 記述統計量 : (または要約統計量・基本統計量)
 - データを簡潔に要約して表すための統計値

- その集団全体の特徴を表す重要な指標
- 一般に確率分布は未知
- 手に入る少数のサンプル (観測データ) から **推定**

$$X_1, X_2, \dots, X_n$$

- 真の値と観測データによる推定には差が存在

平均

- データの代表値を表す記述統計量
- **平均** (mean)

$$\mu = \mathbb{E}[X] = \begin{cases} \sum_{x \in \Omega} x p(x), & (\text{離散分布の場合}) \\ \int_{x \in \Omega} x p(x) dx, & (\text{連続分布の場合}) \end{cases}$$

- **標本平均** (sample mean)

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{X_1 + \dots + X_n}{n}$$

分散

- データのばらつき具合を表す記述統計量
- **分散** (variance)

$$\text{Var}(X) = \sigma^2 = \mathbb{E}[(X - \mu)^2]$$

- **標本分散** (sample variance)

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{(X_1 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{n}$$

標本平均・分散の不偏性

- 標本平均は μ の **不偏推定量**である

$$\mathbb{E}[\bar{X}] = \mu$$

- 標本分散は σ^2 の **不偏推定量**ではない

$$\mathbb{E}[S^2] = \frac{n-1}{n} \sigma^2$$

- 標本分散は平均的には真の分散を **過小推定** する

不偏分散

- 不偏性を担保した分散の推定量
- バイアス補正：標本分散に $n/(n-1)$ を乗じたもの

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

は σ^2 の不偏推定量となる

標準偏差

- 標準偏差 (standard deviation)

$$\sigma = \sqrt{\text{Var}(X)} = (\mathbb{E}[(X - \mu)^2])^{1/2}$$

– 分散の平方根

- 標本標準偏差 (sample standard deviation)

$$\hat{\sigma} = s$$

– 通常、不偏分散の平方根 s を用いる

- 一般に s は標準偏差 σ の **不偏推定量ではない**

標準化

- 複数データの分析のために単位や基準を揃える方法
- データ X_1, X_2, \dots, X_n の標準化 (standardization)

$$Z_i = \frac{X_i - \bar{X}}{s} \quad (i = 1, 2, \dots, n)$$

– s の代わりに S で割って定義する文献もある

- 定義から Z_1, Z_2, \dots, Z_n の標本平均は 0, 不偏分散は 1 に規格化される
- Z_i : **標準得点** (standard score) あるいは **Z スコア** (Z-score) と呼ばれる

偏差値

- 別の基準での標準化
 - 教育学や心理学では、平均 50, 標準偏差 10 が好まれる
- 標本平均 50, 標準偏差 10 への線形変換

$$T_i = 10Z_i + 50 \quad (i = 1, \dots, n)$$

- T_i : **偏差値得点** あるいは **T スコア** (T-score) と呼ばれる

共分散

- 複数のデータ間の関係を知るための記述統計量
- **共分散** (covariance)

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

- **標本共分散** (sample covariance)

$$\text{Cov}(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n - 1}$$

相関

- 2 種類のデータ間の比例関係の大きさ
- **相関** (correlation)

$$\rho = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

- **標本相関** (sample correlation)

$$\rho = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

離散分布の尤度関数

- $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ の同時確率
 - 確率質量関数: $f_{\theta}(x)$
 - 確率質量関数のパラメタ: $\theta = (\theta_1, \dots, \theta_p)$
 - 独立な確率変数の同時確率

$$\begin{aligned} P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) &= \prod_{i=1}^n P(X_i = x_i) \\ &= \prod_{i=1}^n f_{\theta}(x_i) = f_{\theta}(x_1) \cdot f_{\theta}(x_2) \cdots f_{\theta}(x_n) \end{aligned}$$

- 定義

パラメタ θ に対して観測データ X_1, X_2, \dots, X_n が得られる理論上の確率

$$L(\theta) = \prod_{i=1}^n f_{\theta}(X_i)$$

を θ の **尤度** と言い、 θ の関数 L を **尤度関数** と呼ぶ。

- 観測データ X_1, X_2, \dots, X_n が現れるのにパラメタ θ の値がどの程度尤もらしいかを測る尺度となる

連続分布の尤度関数

- $x_1 \leq X_1 \leq x_1 + \delta, \dots, x_n \leq X_n \leq x_n + \delta$ の同時確率
 - 確率密度関数: $f_{\theta}(x)$
 - 確率密度関数のパラメタ: $\theta = (\theta_1, \dots, \theta_p)$
 - 独立な確率変数の同時確率

$$\begin{aligned} P(x_1 \leq X_1 \leq x_1 + \delta, \dots, x_n \leq X_n \leq x_n + \delta) &= \prod_{i=1}^n P(x_i \leq X_i \leq x_i + \delta) \\ &\simeq \prod_{i=1}^n f_{\theta}(x_i) \delta = f_{\theta}(x_1) \cdot f_{\theta}(x_2) \cdots f_{\theta}(x_n) \delta^n \end{aligned}$$

- 定義

パラメタ θ に対して観測データ X_1, X_2, \dots, X_n が得られる理論上の確率密度

$$L(\theta) = \prod_{i=1}^n f_{\theta}(X_i)$$

を θ の **尤度** と言い, θ の関数 L を **尤度関数** と呼ぶ.

- 確率ではないので 1 より大きな値となることもある

最尤法

- 最尤法

観測データに対して「最も尤もらしい」パラメタ値を θ の推定量として採用する方法を最尤法という.

- 最尤推定量

Θ を尤度関数の定義域として, 尤度関数を最大とする $\hat{\theta}$

$$L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta).$$

を θ の **最尤推定量** という.

Bayes の定理 (基本形)

- 定理

条件付確率では次の等式が成り立つ.

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}.$$

- 左辺と右辺で事象 A, B の役割が異なる

Bayes の定理 (一般形)

- 定理

$\Omega = A_1 + A_2 + \cdots + A_n$ のとき

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{k=1}^n P(A_k)P(B|A_k)}$$

が成り立つ.

演習

問題

- 以下の問に答えなさい

A 先生は大の野球ファンで、球団 H の勝敗で翌日の機嫌が左右されるとしよう。よくよく調べた結果

- 球団 H が勝つと 90%の確率で機嫌が良い
- 球団 H が負けると 70%の確率で機嫌が悪い

が成り立っているとする。

また球団 H の勝率は現在のところ

- 球団 H は 60%の確率で勝つ
- 球団 H は 40%の確率で負ける

となっているとする。

- A 先生が機嫌が良いときに球団 H が勝った確率は？
- A 先生が機嫌が悪いときに球団 H が負けた確率は？

解答例

- まず事象を定義する

A : 先生の機嫌が良い

H : 球団が勝つ

A^c : 先生の機嫌が悪い

H^c : 球団が負ける

- 条件を書き下す

$$P(A|H) = 0.9$$

$$P(A^c|H^c) = 0.7$$

$$P(H) = 0.6$$

$$P(A^c|H) = 0.1$$

$$P(A|H^c) = 0.3$$

$$P(H^c) = 0.4$$

- A 先生が機嫌が良いときに球団 H が勝った確率は？

$$\begin{aligned} P(H|A) &= \frac{P(A, H)}{P(A)} \\ &= \frac{P(A, H)}{P(A, H) + P(A, H^c)} \\ &= \frac{P(A|H)P(H)}{P(A|H)P(H) + P(A|H^c)P(H^c)} \\ &= \frac{0.9 \times 0.6}{0.9 \times 0.6 + 0.3 \times 0.4} \\ &= \frac{9}{11} \approx 0.818 \end{aligned}$$

- A 先生が機嫌が悪いときに球団 H が負けた確率は？

$$\begin{aligned}
P(H^c|A^c) &= \frac{P(A^c, H^c)}{P(A^c)} \\
&= \frac{P(A^c, H^c)}{P(A^c, H) + P(A^c, H^c)} \\
&= \frac{P(A^c|H^c)P(H^c)}{P(A^c|H)P(H) + P(A^c|H^c)P(H^c)} \\
&= \frac{0.7 \times 0.4}{0.1 \times 0.6 + 0.7 \times 0.4} \\
&= \frac{14}{17} \simeq 0.824
\end{aligned}$$

関数の微分

ベクトルによる微分

- d 次元ベクトル

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_d \end{pmatrix} = (a_1, a_2, \dots, a_d)^\top$$

- ベクトル \mathbf{a} による関数 $f(\mathbf{a})$ の微分の定義

$$\frac{\partial f}{\partial \mathbf{a}} = \left(\frac{\partial f}{\partial a_1}, \frac{\partial f}{\partial a_2}, \dots, \frac{\partial f}{\partial a_d} \right)^\top$$

ベクトルによる微分 (例題)

- 問題

d 次元ベクトル \mathbf{a} と \mathbf{b} を用いて定義される関数 $f(\mathbf{a}) = \mathbf{b}^\top \mathbf{a} = \mathbf{a}^\top \mathbf{b}$ の \mathbf{a} による微分を求めよ.

- 解答例

各成分で考えると以下のように計算される.

$$\frac{\partial f}{\partial a_i} = \frac{\partial}{\partial a_i} (a_1 b_1 + \dots + a_i b_i + \dots + a_d b_d) = b_i.$$

したがって

$$\frac{\partial f}{\partial \mathbf{a}} = (b_1, b_2, \dots, b_d)^\top = \mathbf{b}$$

となる.

- 注意

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{a}} (\mathbf{a}^\top \mathbf{b}) &= \mathbf{b} \\
\frac{\partial}{\partial \mathbf{a}} (\mathbf{b}^\top \mathbf{a}) &= (\mathbf{b}^\top)^\top = \mathbf{b}
\end{aligned}$$

というルールがあることがわかる.

行列による微分

- $d \times d$ 行列

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1d} \\ a_{21} & a_{22} & \cdots & a_{2d} \\ \vdots & & \ddots & \vdots \\ a_{d1} & a_{d2} & \cdots & a_{dd} \end{pmatrix}$$

- 行列 A による関数 $f(A)$ の微分の定義

$$\frac{\partial f}{\partial A} = \begin{pmatrix} \frac{\partial f}{\partial a_{11}} & \frac{\partial f}{\partial a_{12}} & \cdots & \frac{\partial f}{\partial a_{1d}} \\ \frac{\partial f}{\partial a_{21}} & \frac{\partial f}{\partial a_{22}} & \cdots & \frac{\partial f}{\partial a_{2d}} \\ \vdots & & \ddots & \vdots \\ \frac{\partial f}{\partial a_{d1}} & \frac{\partial f}{\partial a_{d2}} & \cdots & \frac{\partial f}{\partial a_{dd}} \end{pmatrix}$$

行列による微分 (例題 1)

- 問題

行列 A と d 次元ベクトル \mathbf{b} を用いて定義される関数

$$f(A) = \mathbf{b}^\top A \mathbf{b} = \sum_{i,j=1}^d b_i a_{ij} b_j$$

の行列 A による微分を求めよ.

- 解答例

成分で考えると

$$\frac{\partial f}{\partial a_{ij}} = \frac{\partial}{\partial a_{ij}} \sum_{i',j'=1}^d b_{i'} a_{i'j'} b_{j'} = b_i b_j$$

となるので,

$$\frac{\partial}{\partial A} \mathbf{b}^\top A \mathbf{b} = \begin{pmatrix} b_1 b_1 & b_1 b_2 & \cdots & b_1 b_d \\ b_2 b_1 & b_2 b_2 & \cdots & b_2 b_d \\ \vdots & & \ddots & \vdots \\ b_d b_1 & b_d b_2 & \cdots & b_d b_d \end{pmatrix} = \mathbf{b} \mathbf{b}^\top$$

と書くことができる.

行列による微分 (例題 2)

- 問題

$d \times d$ 行列 A と B を用いて定義される関数

$$f(A) = \text{tr} AB = \sum_{i,j=1}^d a_{ij} b_{ji}$$

の行列 A による微分を求めよ.

- 解答例

成分では

$$\frac{\partial f}{\partial a_{ij}} = b_{ji}$$

となるので,

$$\frac{\partial}{\partial A} \text{tr} AB = \begin{pmatrix} b_{11} & b_{21} & \dots & b_{d1} \\ b_{12} & b_{22} & \dots & b_{d2} \\ \vdots & & \ddots & \vdots \\ b_{1d} & b_{2d} & \dots & b_{dd} \end{pmatrix} = B^T$$

と書くことができる.

• 注意 1

行列のトレースの性質

$$\text{tr} AB = \text{tr} BA, \quad \text{tr} AB = \text{tr}(AB)^T = \text{tr} B^T A^T$$

より

$$\frac{\partial}{\partial A} \text{tr} AB = \frac{\partial}{\partial A} \text{tr} BA = \frac{\partial}{\partial A} \text{tr} A^T B^T = \frac{\partial}{\partial A} \text{tr} B^T A^T = B^T$$

ことが容易に確かめられる.

• 注意 2

$$b^T A b = \text{tr} b^T A b = \text{tr} A b b^T$$

となることから

$$\frac{\partial}{\partial A} b^T A b = \frac{\partial}{\partial A} \text{tr} A b b^T = (b b^T)^T = b b^T$$

となり, 2つの例での計算結果が矛盾しないことが確かめられる.

演習

問題

- 行列 (正方行列に限らない) のトレースに関して

$$\text{tr} AB = \text{tr} B^T A^T = \text{tr} BA = \text{tr} A^T B^T$$

が成り立つことを示せ

- d 次元ベクトル a と $d \times d$ 行列 A で定義される関数 $f(a) = a^T A a$ のベクトル a による微分を求めよ
- 行列 A の行列式を $|A|$ と書くとき行列 A による行列式 $|A|$ の微分を求めよ
 - ヒント: 余因子展開を利用すると容易に求められる

解答例

- $\text{tr} AB$ より, 行列 A, B の積は正方行列になることから, A が $n \times m$ 行列とすれば, B は $m \times n$ 行列となる. したがって

$$\text{tr} AB = \sum_{i=1}^n \sum_{j=1}^m a_{ij} b_{ji}$$

と書くことができる。他の式も同様に書けることを確認すればよい。

- 微分における積の法則 (Leibniz 則) を用いればよい。

$$\begin{aligned}\frac{\partial f(\mathbf{a})}{\partial \mathbf{a}} &= \frac{\partial}{\partial \mathbf{a}} \mathbf{a}^\top A \mathbf{b} \big|_{\mathbf{b}=\mathbf{a}} + \frac{\partial}{\partial \mathbf{a}} \mathbf{b}^\top A \mathbf{a} \big|_{\mathbf{b}=\mathbf{a}} \\ &= A\mathbf{a} + (\mathbf{a}^\top A)^\top \\ &= (A + A^\top)\mathbf{a}\end{aligned}$$

- 行列 A の (i, j) 成分に関する余因子を Δ_{ij} とする。行列式 $|A|$ と逆行列 A^{-1} の (i, j) 成分はそれぞれ

$$|A| = \sum_{j=1}^d a_{ij} \Delta_{ij}, \quad \forall i \quad (A^{-1})_{ij} = \frac{\Delta_{ji}}{|A|}$$

と書くことができる。したがって

$$\frac{\partial |A|}{\partial A} = |A| (A^{-1})^\top$$

となる。

次回の予定

- 第1日：回帰モデルの考え方と推定
- 第2日：モデルの評価
- 第3日：モデルによる予測と発展的なモデル