クレジット:
UTokyo Online Education 統計データ解析Ⅱ 2018 小池祐太

統計データ解析 (II) 第 7 回

小池祐太

2018 年 5 月 29 日

1. 多次元確率変数と多変量分布
   - 多変量正規分布
2. 重回帰分析
   - 目的
   - 回帰係数の推定
   - R での実行
   - 最小二乗法
   - 最小二乗法による線形回帰式の推定の幾何学的解釈
   - 線形回帰式と標本平均

多次元確率変数と多変量分布

- 本講義では多変量データを扱うので, 確率変数の多次元版を考える必要がある
- 値がランダムに決定される d 次元ベクトルで, 各座標が確率変数であるようなものを d **次元確率変数 (**d**-dimensional random variable)** または d **次元確率ベクトル (**d**-dimensional random vector)** と呼ぶ.[1]

[1] 以下特に断らない限り, ベクトルは列ベクトルとみなす.

多次元確率変数と多変量分布

- 1 次元の場合の多次元化として, 多次元確率変数の分布を以下のようにして定義する
- d 次元確率変数 $X = (X_1, X_2, \ldots, X_d)^\top$ に対して, d 次元長方形 $\{(x_1, \ldots, x_d) : a_i \le x_i \le b_i \ (i = 1, \ldots, d)\}$ $(a_i \le b_i, \ i = 1, \ldots, d)$ と, X がこの d 次元長方形に含まれる確率

$$P(a_1 \le X_1 \le b_1, a_2 \le X_2 \le b_2, \ldots, a_d \le X_d \le b_d)$$

との対応を示したものを, X の **(d 変量) 確率分布**または単に **(d 変量) 分布**といい, X はこの分布に**従う**という

多変量正規分布

- 1 変量の場合と同様に, 多変量の場合も連続分布が定義される
- d 次元確率変数 X(の分布) が**連続型**であるとは, ある d 個の変数をもつ非負値関数 $f(x_1, \ldots, x_d)$ が存在して, $a_i \leq b_i$ $(i = 1, 2, \ldots, d)$ なる任意の実数 $a_1, b_1, a_2, b_2, \ldots, a_d, b_d$ に対して,

$$
\begin{aligned}
&P(a_1 \leq X_1 \leq b_1, a_2 \leq X_2 \leq b_2, \ldots, a_d \leq X_d \leq b_d) \\
&\qquad = \int_{a_1}^{b_1} \int_{a_2}^{b_2} \cdots \int_{a_d}^{b_d} f(x_1, x_2, \ldots, x_d) dx_1 dx_2 \cdots dx_d
\end{aligned}
$$

  が成り立つことをいう
- 1 変量の場合と同様 X の分布は関数 f によって完全に決定されることが知られており, この関数 f を X(の分布) の **(確率) 密度 (関数)** と呼ぶ

多変量正規分布

- 多変量連続分布の最も重要な例は多変量正規分布である
- $\boldsymbol{\mu}$ を d 次元ベクトル, Σ を d 次正定値対称行列[2] とするとき, 確率密度関数が

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}\sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)$$

で与えられる連続型 d 変量分布を, 平均ベクトル $\boldsymbol{\mu}$, 共分散行列 Σ の d **変量正規分布 (*d*-dimensional normal distribution)** と呼ぶ

- Σ の行列式は Σ の固有値の積で与えられるから, $\det \Sigma > 0$ であり, 特に Σ は正則である
- d 次元ベクトル $\boldsymbol{x}$ は列ベクトルとみなしている

[2] d 次対称行列 Σ が**正定値 (positive definite)** であるとは, Σ の固有値がすべて正であることをいう.

多変量正規分布

- 平均ベクトルが零ベクトルで共分散行列が単位行列の d 変量正規分布を d **変量標準正規分布 (*d*-dimensional standard normal distribution)** と呼ぶ
- R にはデフォルトでは多変量正規分布に従う乱数を生成するための関数は用意されていないため, 自作する必要がある
  - パッケージをインストールする方法もある (後述)
- 多変量正規分布のシミュレーションを行うためには, 次の命題が有用である:

多変量正規分布

命題 1

(a) $X_1, \ldots, X_d$ を標準正規分布に従う独立な d 個の確率変数とする．このとき，d 次元確率変数 $X = (X_1, \ldots, X_d)^\top$ は d 変量標準正規分布に従う．
(b) X を平均ベクトル $\boldsymbol{\mu}$, 共分散行列 Σ の d 変量正規分布に従う d 変量確率変数とする．このとき，A が d 次正則行列，$\boldsymbol{b}$ が d 次元列ベクトルならば，d 次元確率変数 $AX + \boldsymbol{b}$ は平均ベクトル $A\boldsymbol{\mu} + \boldsymbol{b}$, 共分散行列 $A\Sigma A^\top$ の d 変量正規分布に従う．

多変量正規分布

- 命題 1 より, d 次元ベクトル $\boldsymbol{\mu}$ および d 次元正定値対称行列 Σ が与えられたとき, 以下の手順によって平均ベクトル $\boldsymbol{\mu}$, 共分散行列 Σ の d 変量正規分布に従う d 次元確率変数 X を生成できる:
  (1) d 個の標準正規乱数 $Z_1, \dots, Z_d$ を生成し, $Z = (Z_1, \dots, Z_d)^\top$ とおく. 命題 1(a) より Z は d 変量標準正規分布に従う.
  (2) d 次正方行列 A で $\Sigma = AA^\top$ を満たすものを計算し, $X = \boldsymbol{\mu} + AZ$ とおく. 命題 1(b) より X は平均ベクトル $\boldsymbol{\mu}$, 共分散行列 Σ の d 変量正規分布に従う.
- 上の手順のうち, (1) における標準正規乱数の生成は関数 `rnorm()` によって実行できる
- 手順 (2) における行列 A の計算にはいくつか方法があるが, ここでは固有値分解を用いる方法を説明する.[3]

[3] 別の方法としては, 例えば配布資料 2.7.3 節で説明したコレスキー分解を使う方法があり, より直接的である.

多変量正規分布

- Σ は対称行列だから, ある d 次正則行列 V によって対角化できる:

$$V^{-1}\Sigma V = \Lambda$$

  - Λ は Σ の固有値を対角成分とする対角行列
- さらに, V を直交行列, すなわち $V^{-1} = V^\top$ となるようにとることができることが知られている
- いま, Σ は正定値であったから, Λ の対角成分 $\lambda_1, \ldots, \lambda_d$ はすべて正
- 従って, $\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_d}$ を対角成分とする対角行列 D を考えることができる

多変量正規分布

- このとき $A := VDV^\top$ とおくと,

$$AA^\top = VDV^\top(VDV^\top)^\top = VDV^\top VDV^\top = VD^2V^\top = V\Lambda V^\top = \Sigma$$

  となるので, この行列 A が求めるべきものである
- 多変量正規分布に従う乱数を発生させるための関数を実装しているパッケージはいくつか存在する
  - パッケージ MASS には関数 `mvrnorm()` が, パッケージ `mvtnorm` には関数 `rmvnorm()` がそれぞれ多変量正規分布に従う乱数を発生させるための関数として実装されている
- 実行例 `rmvnorm2.r`

重回帰分析: 目的

- **回帰分析 (regression analysis)**
  - ある 1 種類の変数/データを別の変数/データ (1 種類もしくは複数) によって説明もしくは予測するための関係式 (**回帰 (方程) 式 (regression equation)**) を構成することを目的とする分析法
- 説明される側のデータは, 目的変数, 被説明変数, 従属変数, 応答変数などと呼ばれる
- 説明する側のデータは, 説明変数, 独立変数, 共変量などと呼ばれる
- 説明変数が 1 種類の場合を**単回帰 (simple regression)**, 複数の場合を**重回帰 (multiple regression)** と呼ぶ

目的

- 以下, 目的変数を Y, 説明変数を $X_1, \ldots, X_p$ で表すことにし, 組 $(Y, X_1, \ldots, X_p)$ に対する n 個の観測データ

$$\{(y_i, x_{i1}, \ldots, x_{ip})\}_{i=1}^{n} \tag{1}$$

が得られている状況を考える

- Y を $X_1, \ldots, X_p$ で説明するための関係式は, 一般にはある p 変数関数 f を使って,

$$Y = f(X_1, \ldots, X_p) \tag{2}$$

と書ける

目的

- この形では一般的すぎて分析に不向きのため，多くの場合 f として 1 次関数のみを考える
- すなわち，ある定数 $\beta_0, \beta_1, \ldots, \beta_p$ によって

$$f(x_1, \ldots, x_p) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

  と書ける場合を考察する
  - この場合を**線形回帰 (linear regression)** と呼ぶ (一般の場合は**非線形回帰 (nonlinear regression)** と呼ばれる)
- 注意
  - 例えば X_j^2 や $X_j X_k$ といった 2 次式や，$\log X_j$ などの適切な非線形関数で変換した説明変数を新たな説明変数として加えることにすれば，説明変数と目的変数の間のある程度非線形な関係を線形回帰モデルでも表すことができる

目的

- 線形回帰の場合, 回帰式 (2) は

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p \tag{3}$$

  という形になる
- パラメーター $\beta_0, \beta_1, \ldots, \beta_p$ は**回帰係数 (regression coefficients)** と呼ばれ, 特に β_0 を**定数項 (constant term)** と呼ぶ
- 回帰係数は未知なので, 回帰式の構成にはそれらをデータから決定することが必要となる

回帰係数の推定

- 一般にあるモデルを考えたとき, そのモデルに含まれる未知パラメーターを観測データから (何らかの意味で) 決定する作業を**推定 (estimation)** と呼ぶ
- 我々の目標は, 回帰式 (3) に含まれるパラメーター $\beta_0, \beta_1, \ldots, \beta_p$ を観測データ (1) から推定することである

回帰係数の推定

- 実際のデータは観測誤差などのランダムな変動を含むと考えられるため, 回帰式 (3) が観測データに対してそのまま成立するとは期待しづらい
- そのため, 統計学では, データのばらつきを表す項を ϵ_i として, 以下の形の確率モデルを分析することを考える:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i, \qquad i = 1, \ldots, n. \tag{4}$$

  - $\epsilon_1, \ldots, \epsilon_n$ は**誤差項 (error term)** もしくは**撹乱項 (disturbance term)** と呼ばれる
  - 誤差項は確率変数であり, しばしば平均 0 の正規乱数列と仮定される

R での実行

- R では線形回帰分析を実行するための関数 `lm( )` が用意されている
  - モデル (4) において, 目的変数 Y および説明変数 $X_1, \ldots, X_p$ の観測データに対応するベクトルがそれぞれ `y` および `x1,...,xp` で与えられているとする
  - このとき, モデル (4) の回帰係数の推定は, コマンド

    `lm(y ~ x1 + ··· + xp)`

    で実行できる

R での実行

- 実際のデータを使って解析する際は, データセットの一部の変数を目的変数および説明変数として回帰分析をすることが多い
- そのような場合, データセットに対応するデータフレームを `dat` とすれば, 以下のコマンドで回帰係数の推定を実行できる:

  `lm(`Y の変数名 `~` X_1 の変数名 `+` $\cdots$ `+` X_p の変数名`, data = dat)`

  - ここで, `dat` は列が各変数に対応するような形式になっている必要がある
- 実行例 `lse.r`

最小二乗法

- 回帰係数の推定には通常**最小二乗法 (least squares)** が用いられる
- 最小二乗法の考え方は以下の通りである
  - 回帰係数 $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^\top$ を 1 つ決めたとき, 回帰式では説明できない目的変数の変動は,

$$e_i(\boldsymbol{\beta}) = y_i - (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}), \qquad i = 1, \ldots, n$$

    で与えられる
  - これらの変動 $e_1(\boldsymbol{\beta}), \ldots, e_n(\boldsymbol{\beta})$ はいずれも絶対値が小さいほど当てはまりがよいと考えられる
  - そこで, 最小二乗法では, $e_1(\boldsymbol{\beta}), \ldots, e_n(\boldsymbol{\beta})$ の平方和

$$S(\boldsymbol{\beta}) := \sum_{i=1}^{n} e_i(\boldsymbol{\beta})^2$$

    を最小にするように回帰係数 $\boldsymbol{\beta}$ を決定する

最小二乗法

- $S(\beta)$ は**残差平方和 (residual sum of squares)** と呼ばれ, $S(\beta)$ を最小にする β は**最小二乗推定量 (least squares estimator)** と呼ばれる
- 最小二乗推定量はしばしば記号 $\hat{\beta}$ で表される

線形回帰式の行列による表現

- 最小二乗推定量の計算のために，モデル (4) を行列を用いて表現する

$$
\boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix},
$$

$$
\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \qquad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \qquad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}
$$

とおくと，モデル (4) は

$$
\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}
$$

と表すことができる

線形回帰式の行列による表現

- 行列 $\boldsymbol{X}$ は**デザイン行列 (design matrix)** と呼ばれることがある
- 一般に列ベクトル $\boldsymbol{a} = (a_1, \ldots, a_n)^\top$ に対して, $\boldsymbol{a}^\top \boldsymbol{a}$ は $\boldsymbol{a}$ の各成分の二乗の総和 $\sum_{i=1}^n a_i^2$ に一致することに注意すれば, 残差平方和は

$$S(\boldsymbol{\beta}) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$$

と表せる

正規方程式

- もし $\boldsymbol{\beta}$ が最小二乗推定量ならば, 微分積分学の理論から, 点 $\boldsymbol{\beta}$ における残差二乗和の勾配は零ベクトルとなる必要がある[4]:

$$\nabla S(\beta) := \left(\frac{\partial S}{\partial \beta_0}(\boldsymbol{\beta}), \frac{\partial S}{\partial \beta_1}(\boldsymbol{\beta}), \dots, \frac{\partial S}{\partial \beta_p}(\boldsymbol{\beta})\right)^\top = \mathbf{0}. \tag{5}$$

- 各 $j = 0, 1, \dots, p$ について偏微分を計算すると,

$$\frac{\partial S}{\partial \beta_j}(\boldsymbol{\beta}) = -2\sum_{i=1}^{n}\left(y_i - \sum_{k=0}^{p}\beta_k x_{ik}\right)x_{ij}$$

となる. 但し, $x_{i0} = 1$ $(i = 1, \dots, n)$ とおいた

[4]例えば杉浦光夫著「解析入門 I」(東京大学出版会) の第 II 章定理 8.1 参照

正規方程式

- 従って方程式 (5) は,

$$\sum_{i=1}^{n} x_{ij} \left(\sum_{k=0}^{p} x_{ik}\beta_k \right) = \sum_{i=1}^{n} x_{ij} y_i \qquad (j = 0, 1, \ldots, p)$$

  と書き直せる
- x_{ij} が行列 $\boldsymbol{X}$ の (i, j) 成分に対応していることに注意すれば, 上の方程式は

$$\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}^\top \boldsymbol{y} \tag{6}$$

  と書ける
- 方程式 (6) を**正規方程式 (normal equation)** と呼ぶ

正規方程式

- 正規方程式の性質
  - 正規方程式は必ず解をもつ (資料の定理 5.2)
  - 正規方程式の任意の解は最小二乗推定量である (資料の定理 5.3)
    - ★ 一般に, $\boldsymbol{\beta}$ が方程式 (5) の解であることは, $\boldsymbol{\beta}$ が $S(\boldsymbol{\beta})$ を最小にするための必要条件であって十分条件であるとは限らないことに注意
  - 正規方程式の解が一意的であるための必要十分条件は, $(p+1)$ 次正方行列 $\boldsymbol{X}^\top\boldsymbol{X}$ が正則であることであり, このとき正規方程式の解は

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y} \tag{7}$$

    で与えられる (資料の定理 5.1)
    - ★ $\boldsymbol{X}^\top\boldsymbol{X}$ は $\boldsymbol{X}$ の **Gram 行列 (Gram matrix)** と呼ばれることがある
- 実行例 `lse.r`

正規方程式

- 実際のデータ解析をする上では，最小二乗推定量がただ一つだけ存在する状況が好ましい (例えば推定量の選択によって分析結果が変化してしまうことが避けられる)
  - 前頁の正規方程式の性質から，これは $\boldsymbol{X}^\top \boldsymbol{X}$ が正則であることと同値
  - $\boldsymbol{X}^\top \boldsymbol{X}$ が正則であることは，$\boldsymbol{X}$ の列ベクトルが 1 次独立であることと同値であるが証明できる (資料参照)
  - $\boldsymbol{X}$ の各列は各説明変数の観測データからなるベクトルで与えられているから，これは回帰式の構築に利用する説明変数たちが互いに異なる情報をもつという意味だと解釈できる
- 説明変数の間の関係の 1 次従属関係への近さの度合いは**多重共線性 (multicollinearity)** と呼ばれる
- 前頁で述べた理由から，データ解析をする上では説明変数は多重共線性があまり強くならないように選択すべきである
  - 例えば，似たような動きをする説明変数を重複して利用することは避けるべき

最小二乗法による線形回帰式の推定の幾何学的解釈

- 一般に $\hat{\beta}$ を回帰係数の推定値とすると，目的変数の観測データ $\boldsymbol{y}$ から観測誤差の影響を除いた目的変数の真の値が，

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\beta}$$

によって推定できる
  - $\hat{\boldsymbol{y}}$ を**あてはめ値 (fitted values)** または**予測値 (predicted values)** と呼ぶ
- 特に $\hat{\beta}$ が最小二乗推定量のとき，デザイン行列 $\boldsymbol{X}$ の列ベクトルたちで張られる $\mathbb{R}^n$ の部分線形空間 (超平面) を $L[\boldsymbol{X}]$ と書くことにすれば，$\hat{\boldsymbol{y}}$ はベクトル $\boldsymbol{y}$ の $L[\boldsymbol{X}]$ への直交射影となることがわかる (資料の定理 5.3)
  - すなわち，$\hat{\boldsymbol{y}}$ は，$L[\boldsymbol{X}]$ に属するベクトル $\boldsymbol{z}$ で $\boldsymbol{y} - \boldsymbol{z}$ が $L[\boldsymbol{X}]$ に属するすべてのベクトルに直交するような唯一のもの

最小二乗法による線形回帰式の推定の幾何学的解釈

- 幾何学的には, ベクトル $\boldsymbol{y}$ からデザイン行列 $\boldsymbol{X}$ の列ベクトルによって張られる (超) 平面に垂線を下ろした際の垂線の足が $\hat{\boldsymbol{y}}$ となる
- 次頁に $n = 3$, $p + 1 = 2$ の場合に最小二乗法による推定の様子を図示したものを示す

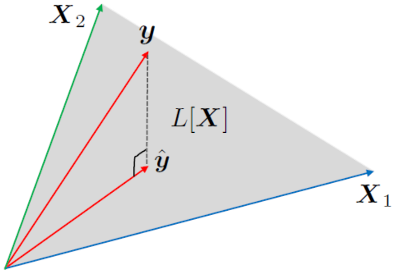

図 1: $n = 3$, $p + 1 = 2$ の場合の最小二乗法による推定の図示. $\boldsymbol{X}_1$, $\boldsymbol{X}_2$ はそれぞれデザイン行列 $\boldsymbol{X}$ の第 1 列, 第 2 列からなるベクトルに対応している. グレーの平面が $L[\boldsymbol{X}]$ に対応.

最小二乗法による線形回帰式の推定の幾何学的解釈

- 特に, ベクトル $\hat{\epsilon} := \boldsymbol{y} - \hat{\boldsymbol{y}}$ はあてはめ値 $\hat{\boldsymbol{y}}$ に直交する:

$$\hat{\epsilon} \cdot \hat{\boldsymbol{y}} = 0.$$

- $\hat{\epsilon}$ は**残差 (residuals)** と呼ばれ, 回帰式による目的変数のあてはめ値と実際の観測値とのずれを表す
- 実行例 `lse.r`

線形回帰式と標本平均

- 各 $i = 1, \ldots, n$ について，説明変数の i 番目の観測データに対応するベクトル $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})^\top$ を導入する
- 説明変数および目的変数の標本平均を考える

$$\bar{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i, \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i,$$

- このとき，$\hat{\beta}$ を最小二乗推定量とすれば，$\bar{y} = (1, \bar{\boldsymbol{x}}^\top)\hat{\beta}$ が成り立つことが確認できる (資料参照)
- 幾何学的には，方程式 $y = (1, \boldsymbol{x}^\top)\hat{\beta}$ によって定まる超平面は常に点 $(\bar{\boldsymbol{x}}^\top, \bar{y})$ を通るということである
- 実行例 `lse.r`