大学院計量経済学講義ノートI

劉　慶豊 (Qingfeng Liu)*

平成21年10月9日

小樽商科大学商学部

*qliu@res.otaru-uc.ac.jp

1　線形数学入門

1.1　行列表記

- $n \times K$ 行列

$$\mathbf{A} = [a_{ik}] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1K} \\ a_{21} & a_{22} & \cdots & a_{1K} \\ & & \cdots & \\ a_{n1} & a_{n2} & \cdots & a_{nK} \end{bmatrix} \tag{1}$$

- ベクトルは $n \times 1$ の行列（列ベクトル）または $1 \times K$ の行列と考えられる (行ベクトル)。例：

$$\text{列ベクトル} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \tag{2}$$

$$\text{行ベクトル} \begin{bmatrix} a_1 & a_2 & \cdots & a_K \end{bmatrix} \tag{3}$$

- 行列の次元 $n \times K$
- 正方行列　$n = K$
- 対称行列　$a_{ik} = a_{ki}$
- 対角行列、スカラー
- 単位行列 $\mathbf{I}$：対角要素が全て 1 それ以外は全部 0

$$\mathbf{I}_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{4}$$

- 三角行列: 対角線の下または上の要素が全て 0
- $\mathbf{A} = \mathbf{B}$ if and only if $a_{ik} = b_{ik}$ for all i and k

1.2　行列四則演算

- 転置 $\mathbf{A}$ の転置 $\mathbf{A}'$ の第 k 行が $\mathbf{A}$ の第 k 列。a_{ik} を $\mathbf{A}$ の要素とし、b_{ik} を $\mathbf{A}'$ の要素とする、$b_{ik} = a_{ki}$ となります。例、

$$\mathbf{A} = \begin{bmatrix} 0 & 2 & 6 & 4 \\ 5 & 3 & 9 & 8 \end{bmatrix}$$

$$
\mathbf{A}' = \begin{bmatrix} 0 & 5 \\ 2 & 3 \\ 6 & 9 \\ 4 & 8 \end{bmatrix} \tag{5}
$$

1. 転置が以下の性質がある、$\mathbf{A}$ が対称行列であれば、$\mathbf{A} = \mathbf{A}\prime$
2. $(\mathbf{A}\prime)' = \mathbf{A}$

- 行列の足し算、引き算：同じ次元の行列（ベクトル）だけ足し算と引き算の計算ができる

$$
\begin{aligned}
\mathbf{A} + \mathbf{B} &= [a_{ik} + b_{ik}] \\
\mathbf{A} - \mathbf{B} &= [a_{ik} - b_{ik}]
\end{aligned}
$$

  以下の性質がある

  1. $\mathbf{A} \pm \mathbf{B} = \mathbf{B} \pm \mathbf{A}$
  2. $(\mathbf{A} \pm \mathbf{B}) \pm \mathbf{C} = \mathbf{A} \pm (\mathbf{B} \pm \mathbf{C})$
  3. $(\mathbf{A} \pm \mathbf{B})' = \mathbf{A}' \pm \mathbf{B}'$

- ベクトルの内積

$$
\mathbf{a}'\mathbf{b} = \mathbf{b}'\mathbf{a} = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n
$$

  注：通常ボールド体の小文字のアルファベットで**列ベクトル**を表記する。

- 行列の掛け算

  $\mathbf{A}$ は $n \times K$ の行列、$\mathbf{B}$ を $K \times m$ の行列とする。$\mathbf{C} = [c_{ij}] = \mathbf{AB}$ は $n \times m$ の行列になる。その要素 $c_{ij} = a_i' b_j$ は $\mathbf{A}$ の第 i 行と $\mathbf{B}$ の j 列の内積になる。掛け算ができるのは前の行列の列数必ず後ろの行列の行数と同じであるときだけ。

$$
\mathbf{A} = \begin{bmatrix} \quad 1 \quad 2 \quad 1 \quad 2 \\ \begin{bmatrix} 3 & 7 & 6 & 9 \end{bmatrix} \end{bmatrix}
$$

$$
\mathbf{B} = \begin{bmatrix} \begin{bmatrix} 3 \\ 0 \\ 4 \\ 9 \end{bmatrix} & \begin{matrix} 5 & 2 \\ 6 & 4 \\ 5 & 3 \\ 7 & 8 \end{matrix} \end{bmatrix}
$$

$$
\mathbf{C} = \mathbf{AB} = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \end{bmatrix}
$$

$$
\begin{aligned}
c_{21} &= a_2' b_1 \\
&= \left[\begin{array}{cccc} 3 & 7 & 6 & 9 \end{array}\right] \left[\begin{array}{c} 3 \\ 0 \\ 4 \\ 9 \end{array}\right] \\
&= 114
\end{aligned}
$$

c がスカラーの時

$$c\mathbf{A} = [ca_{ik}]$$

$$\left[\begin{array}{c} 5 \\ 4 \\ 1 \end{array}\right] = \left[\begin{array}{ccc} 4 & 2 & 1 \\ 2 & 6 & 1 \\ 1 & 1 & 0 \end{array}\right] \left[\begin{array}{c} a \\ b \\ c \end{array}\right]$$

は

$$
\begin{aligned}
5 &= 4a + 2b + 1c \\
4 &= 2a + 6b + 1c \\
1 &= a + b + 0c
\end{aligned}
$$

の簡単な表現。

行列の掛け算の法則

1. $(\mathbf{AB})\,\mathbf{C} = \mathbf{A}(\mathbf{BC})$
2. $\mathbf{A}\,(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$
3. $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$
4. $(\mathbf{ABC})' = \mathbf{C}'\mathbf{B}'\mathbf{A}'$

1.3 冪等行列

- $\mathbf{AA} = \mathbf{A}$ となる正方行列が冪等行列である。

$$
\begin{aligned}
\mathbf{i} &= \left[\begin{array}{cccc} 1 & 1 & \cdots & 1 \end{array}\right] \\
\mathbf{M}^0 &= \mathbf{I} - \frac{1}{n}\mathbf{ii}' \\
\mathbf{i}'\mathbf{x} &= \sum_{i=1}^{n} x_i \\
\frac{1}{n}\mathbf{ii}'\mathbf{x} &= \left[\begin{array}{cccc} \bar{x} & \bar{x} & \cdots & \bar{x} \end{array}\right]' \\
\mathbf{M}^0\mathbf{x} &= \left[\begin{array}{cccc} x_1 - \bar{x} & x_2 - \bar{x} & \cdots & x_n - \bar{x} \end{array}\right]'
\end{aligned}
$$

$\mathbf{M}^0$ が対称な冪等行列

$$\mathbf{M}^0\mathbf{M}^0 = \mathbf{M}^{0\prime}\mathbf{M}^0 = \mathbf{M}^0$$

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = \mathbf{x}'\mathbf{M}^{0\prime}\mathbf{M}^0\mathbf{x} = \mathbf{x}'\mathbf{M}^0\mathbf{x}$$

$$\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \mathbf{x}'\mathbf{M}^0\mathbf{y}$$

1.4 ベクトル空間

- 線形従属,Linear Dependent

  一組のベクトルの中に他のベクトルの線形結合で表現できるようなベクトルが存在する場合、これらのベクトルが線形従属, Linear Dependent という。

- 線形独立, Linear Independent
  方程式 $\alpha_1\mathbf{a}_1 + \alpha_2\mathbf{a}_2 + \cdots + \alpha_K\mathbf{a}_K = 0$ の解は $\alpha_1 = \alpha_2 = \cdots = \alpha_K = 0$ しかない場合、$\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_K$ が線形独立,Linear Independent という。

- ベクトル空間

  K 次元のベクトル全体が作るベクトルの集合が K 次元ベクトル空間となる。ベクトル空間はスカラー積及び足し算に関して閉じている。

- ベクトル空間の基底
  K 次元のベクトル空間の中の任意の K 個の線形独立なベクトル。

- 行列のランク

  1. 行列のランクは行列の列ランクであって線形独立な列の数である。同じ行列の列ランクと行ランクが同じである。
  2. $m \times n$ の行列 $\mathbf{A}$ の場合、rank($\mathbf{A}$) =rank($\mathbf{A}'$) $\leq \min(m, n)$。特に $n \times n$ の正方行列 $\mathbf{A}$ のランクは n の時フルランクと言う。このとき $\mathbf{A}$ が非特異行列であると言う。

1.5 正方行列の行列式

- **定義**

$$\det(\mathbf{A}) = |\mathbf{A}| = \sum sign(i_1, i_2, \cdots, i_m)\, a_{1i_1}a_{2i_2}\cdots a_{mi_m} \tag{6}$$

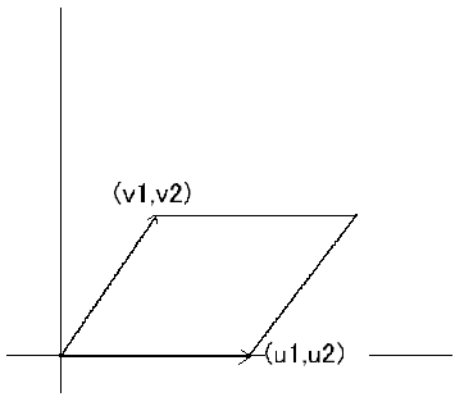
  ただし、$sign(i_1, i_2, \cdots, i_m)$ は $i_1, i_2, \cdots, i_m$ が偶順列のとき 1, 奇順列のときは -1 を取る符号関数である。

- 定義を使って $\mathbf{A}=\begin{bmatrix} u_1 & v_1 \\ u_2 & v_2 \end{bmatrix}$ の行列式を導出しなさい。
- 幾何的な意味合い

$$\mathbf{A}=\begin{bmatrix} u_1 & v_1 \\ u_2 & v_2 \end{bmatrix} \tag{7}$$

$$|\mathbf{A}| = u_1v_2 - u_2v_1$$

行列 $\mathbf{A}$ の行列式は二つのベクトルで囲まれた平行四辺形の面積になっている。面積の値はベクトルの位置関係によって符号が変わることを注意しなさい。



- ベクトルの回転

ベクトル (x, y) を角度 ϕ を回転して作った (x', y') の座標は

$$(x\cos(\phi) - y(\sin(\phi), x\sin(\phi) + y\cos(\phi)))$$

となる。

- ベクトルの回転の公式を利用して、上の図を参考に一本のベクトルは x 軸上にあるような特殊ケースから出発して、任意の 2×2 の行列の行列式の絶対値がその行列の二つの列ベクトルで囲まれた平行四辺形の面積になることを示しなさい。
- 行列式の性質 $n\times n$ 正方行列 $\mathbf{A}$ に関して

1. $|\mathbf{A}| = |\mathbf{A}'|$
2. $|c\mathbf{A}| = c^n |\mathbf{A}|$
3. $|\mathbf{AB}| = |\mathbf{A}|\,|\mathbf{B}|$
4. $\mathbf{A}$ が対角または三角行列のとき $|\mathbf{A}| = \prod_{i=1}^{n} a_{ii}$
5. $\mathbf{A}$ がフルランクではない（特異行列）とき $|\mathbf{A}| = 0$

1.6 逆行列

- **定義**

  非特異行列$\mathbf{A}$ にたいして
$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I} \tag{8}$$
  を満たす行列 $\mathbf{A}^{-1}$ が $\mathbf{A}$ の逆行列と言う。

- 方程式を解く
$$\begin{aligned} \mathbf{A}\mathbf{x} &= \mathbf{b} \\ \mathbf{x} &= \mathbf{A}^{-1}\mathbf{b} \end{aligned} \tag{9}$$

- $(\mathbf{AB})^{-1} = \mathbf{A}^{-1}\mathbf{B}^{-1}$

- $|\mathbf{A}^{-1}| = |\mathbf{A}|^{-1}$

- $\begin{vmatrix} a & b \\ c & d \end{vmatrix}^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

- $(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{BC}\left(\mathbf{C} + \mathbf{CDA}^{-1}\mathbf{BC}\right)^{-1}\mathbf{CDA}^{-1}$

- 分割行列
$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \tag{10}$$
$$\begin{vmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{vmatrix} = |\mathbf{A}_{22}| \cdot \left|\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right| = |\mathbf{A}_{11}| \cdot \left|\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\right| \tag{11}$$
$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22}^{-1} \end{bmatrix} \tag{12}$$
$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}_{11}^{-1}\left(\mathbf{I} + \mathbf{A}_{12}\mathbf{F}_2\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\right) & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{F}_2 \\ -\mathbf{F}_2\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{F}_2 \end{bmatrix} \tag{13}$$
  ただし、
$$\begin{aligned} \mathbf{F}_1 &= \left(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right)^{-1} \\ \mathbf{F}_2 &= \left(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\right)^{-1} \end{aligned} \tag{14}$$

1.7　トレス

- **定義**

$$\mathrm{tr}\,(\mathbf{A}) = \sum_{i=1}^{n} a_{ii} \tag{15}$$

- 性質

1. $\mathrm{tr}(c\mathbf{A}) = c\mathrm{tr}(\mathbf{A})$
2. $\mathrm{tr}(\mathbf{A}')$ $=\mathrm{tr}(\mathbf{A})$
3. $\mathrm{tr}(\mathbf{A}+\mathbf{B})$ $=\mathrm{tr}(\mathbf{A})$ $+\mathrm{tr}(\mathbf{B})$
4. $\mathrm{tr}(\mathbf{AB})$ $=\mathrm{tr}(\mathbf{AB})$、示しなさい。

- ノルム　$\|\mathbf{a}\| = \sqrt{\mathbf{a}'\mathbf{a}}$　直交　$\mathbf{a}'\mathbf{b} = \mathbf{0}$。

1.8　クロネッカー積

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1K}\mathbf{B} \\ a_{21}\mathbf{B} & a_{21}\mathbf{B} & \cdots & a_{2K}\mathbf{B} \\ & & \cdots & \\ a_{n1}\mathbf{B} & a_{n2}\mathbf{B} & \cdots & a_{nK}\mathbf{B} \end{bmatrix}$$

性質は教科書を参照ください。

1.9　固有値と固有ベクトル

- 固有値と固有ベクトルの定義

正方行列 $\mathbf{A}$ に関して、方程式

$$\mathbf{Ac} = \lambda\mathbf{c} \tag{16}$$

を満たすスカラー λ とベクトル $\mathbf{c}$ がそれぞれ $\mathbf{A}$ の固有値と固有ベクトルである。定数倍した $\mathbf{c}$ も解になり、混乱が生じるため、通常 $\mathbf{c}'\mathbf{c} = 1$ になるように基準化する。注意してほしいが固有値と固有ベクトルの解は一組とは限らない。

- 固有方程式

(16) より、

$$\mathbf{Ac} = \lambda\mathbf{Ic} \tag{17}$$

$$(\mathbf{A} - \lambda\mathbf{I})\,\mathbf{c} = \mathbf{0} \tag{18}$$

すべての要素が全部ゼロにならないような $\mathbf{c}$ の解が存在するために行列 $(\mathbf{A}-\lambda\mathbf{I})$ の列は線形従属ではないといけない。すなわち、$(\mathbf{A}-\lambda\mathbf{I})$ が特異行列で

$$|\mathbf{A}-\lambda\mathbf{I}| = 0 \tag{19}$$

になる。上の式は固有方程式と呼ばれる。

$$\mathbf{A} = \begin{bmatrix} 2 & 6 \\ 6 & 7 \end{bmatrix}$$

$$\begin{vmatrix} 2-\lambda & 6 \\ 6 & 7-\lambda \end{vmatrix} = 0$$

を解いて、$\lambda = -2, 11$。それぞれ $\mathbf{Ac}-\lambda\mathbf{c} = \mathbf{0}$ に代入して $\mathbf{c}$ を求める。$\lambda = -2$ の時

$$\begin{bmatrix} 2 & 6 \\ 6 & 7 \end{bmatrix}\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} - \lambda\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = 0$$

より

$$2c_1 + 6c_2 + 2c_1 = 0$$
$$6c_1 + 7c_2 + 2c_2 = 0$$

$$2c_1 = 3c_2$$

がわから。固有ベクトルは

$$\begin{bmatrix} 3a \\ -2a \end{bmatrix}$$

になるが、a が任意の実数。さらに標準化すれば $\left(3\frac{1}{\sqrt{13}}, -2\frac{1}{\sqrt{13}}\right)'$ になる。

問題：$\mathbf{A}$ の固有値と固有ベクトルを求めなさい。注意してほしいが、固有値が虚数になる場合もある。対称行列の場合は固有値が実数になる。

- 固有ベクトルの性質

1. 対称行列の相異なる固有値に対応する固有ベクトルは互いに直交する。
   $\mathbf{Ac}_1{=}\lambda_1\mathbf{c}_1, \mathbf{Ac}_2{=}\lambda_2\mathbf{c}_2$ と対称行列の性質を利用して $\mathbf{c}_1'\mathbf{Ac}_2{=}\,\mathbf{c}_2'\mathbf{Ac}_1$ ゆえに $\lambda_2\mathbf{c}_1'\mathbf{c}_2 = \lambda_1\mathbf{c}_2'\mathbf{c}_1$ が分かる。$\lambda_1 \neq \lambda_2, \mathbf{c}_1'\mathbf{c}_2 = \mathbf{c}_2'\mathbf{c}_1$ なので $(\lambda_1 - \lambda_2)\,\mathbf{c}_1'\mathbf{c}_2 = 0, \mathbf{c}_1'\mathbf{c}_2 = 0$。Greene の説明は「相異なる固有値に対応する」という表現をないため、混乱を招く。
2. $n \times n$ の**対称行列A に関して、**$\mathbf{P}'\mathbf{AP} = \mathbf{\Lambda}$ となる直交行列が存在する。$\mathbf{\Lambda}$ を求める操作は**三角化という**。その逆計算となっている

$$\mathbf{A} = \mathbf{P\Lambda P}' = \sum_{i=1}^{n} \lambda_i \mathbf{p}_i \mathbf{p}_i$$

はスペクトル分解という。

$n \times n$ の行列 $\mathbf{A}$ には n 個の固有値があるとする。以下の三角化

$$\mathbf{C}'\mathbf{A}\mathbf{C} = \mathbf{\Lambda} \tag{20}$$

と

$$\mathbf{A} = \mathbf{C}\mathbf{\Lambda}\mathbf{C}' = \sum_{i=1}^{n} \lambda_i \mathbf{c}_k \mathbf{c}_k \tag{21}$$

を満たす n 個の固有値に対応する固有ベクトルで構成する $\mathbf{C} = (\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_n)$ とが存在する。ただし、

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \tag{22}$$

である。(21) 式はスペクトル分解同時に**固有値分解**という。

Remark：本講義でのほとんどの場合は n 個の固有値がすべて異なる。そのとき、$\mathbf{C}$ を構成する固有ベクトルの符号の違いを無視すれば $\mathbf{C}$ が唯一に決まる。本講義の範囲を超えるが、固有値に重根が存在する場合は、**重根に対応する固有ベクトル（複数の解がある）を**(20) と (21) 式を満たすように選ぶことが必要である。その例としては、一番単純で単位行列が考えられる。

3. $n \times n$ の正方行列 $\mathbf{A}$ に関して、$\lambda_i, i = 1, 2, \cdots, n$ を固有値とする（重根の場合も含む)。以下の性質はすべての正方行列に関して成り立つ。
   (a) $\mathrm{tr}(\mathbf{A}) = \sum_{i=1,n} \lambda_i$
   (b) $|\mathbf{A}| = \prod_{i=1,n} \lambda_i$
   (c) $\mathbf{A}$ のランクはゼロではない固有値の数に等しい
4. 非特異な対称行列 $\mathbf{A}$ に関して、$\mathbf{A}^K$ の各固有値は λ_i^K になる。ただし、K は任意の整数である。$\mathbf{A}$ と $\mathbf{A}^K$ の固有ベクトルが同じである。$\mathbf{A} = \mathbf{C}\mathbf{\Lambda}\mathbf{C}'$、$\mathbf{A}^K = \mathbf{C}\mathbf{\Lambda}^K\mathbf{C}'$。

1.10 二次形式と定符号行列

- 二次形式：

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i a_{ij} x_j$$

- 任意のゼロではないベクトル $\mathbf{x}$ に関して、$\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$ なら非負値定符号（行列式と固有値が 0 か 0 より大きい)、$\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ 正値定符号（行列式とすべての固有値が 0 より大きい)、$\mathbf{x}'\mathbf{A}\mathbf{x} \leq 0$ 非正値定符号（行列式と固有値が 0 か 0 より小さい)、$\mathbf{x}'\mathbf{A}\mathbf{x} < 0$ 負値定符号（行列式とすべての固有値が 0 より小さい)。

1.11 行列の解析

- 一次微分、グラジエント

$$y = f(x_1, x_2, \cdots, x_n)$$

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \begin{bmatrix} \partial y/\partial x_1 \\ \partial y/\partial x_2 \\ \vdots \\ \partial y/\partial x_n \end{bmatrix} = [f_i]$$

- 二次微分、ヘッシアン

$$\frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x}\partial \mathbf{x}'} = \begin{bmatrix} \partial^2 y/\partial x_1\partial x_1 & \partial^2 y/\partial x_1\partial x_2 & \cdots & \partial^2 y/\partial x_1\partial x_n \\ \partial^2 y/\partial x_2\partial x_1 & \partial^2 y/\partial x_2\partial x_2 & \cdots & \partial^2 y/\partial x_2\partial x_n \\ & & \cdots & \\ \partial^2 y/\partial x_n\partial x_1 & \partial^2 y/\partial x_n\partial x_2 & \cdots & \partial^2 y/\partial x_n\partial x_n \end{bmatrix} = [f_{ij}]$$

- ベクトル間の微分、ヤコビアン

$$\left|\frac{\partial \mathbf{z}}{\partial \mathbf{x}'}\right| = |[a_{ij}]| = \left|\left[\frac{\partial z_i}{\partial x_j}\right]\right|$$

ただし、$|\cdot|$ は行列式を意味する。

2 確率論入門

2.1 集合

集合と記号集合論で使われる記号 $\cup, \cap$。
集合は物の集まり。

例 1 クラスを表す集合で、A をクラス *1* のメンバーとする。$A = \{$ 酒井さん、中村さん、田中さん、...。$\}$。

例 2 整数の集合、$A = \{2, 1, 5\}, B = \{3, 6, 5\}$。

集合
$A \cup B$ は A と B の和集合を表す。上の例の場合 $A \cup B = \{2, 1, 5, 3, 6\}$。
$A \cap B$ は A と B の積集合で両方がともに持つ要素より構成される。同じ例で $A \cap B = \{5\}$。
$a \in A$ は要素 a が集合 A に属することを表す。例：$2 \in A$。
$C \subset A$ は集合 C が集合 A の部分集合であることをあらわす。C の要素がすべて A の要素であることが必要十分条件である。
A^c は A の補集合を現す。$A \subset \Omega$ とする、Ω に関する A の補集合 A^c は A の要素ではないすべての Ω の要素によって構成される集合である。$A^c = \{x : x \in \Omega \text{ and } x \notin A\}$。

2.1.1 性質

1. $A \cup B = B \cup A$
2. $A \cap B = B \cap A$
3. $A \cup (B \cup C) = (A \cup B) \cup C = (A \cup C) \cup B...,\qquad A \cap (B \cap C) = (A \cap B) \cap C$
4. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)\,,\qquad A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
5. $(A \cup B)^c = A^c \cap B^c,\quad (A \cap B)^c = A^c \cup B^c$

2.2 確率

日常生活でよく気軽に「確率」という言葉を使っている。たとえば、

- 今日雨が降っているから、彼女遅刻する確率が高いね；
- 両端の車両は席がたくさん空いて、座れる確率が高い；
- 男の子と女の子生まれてくる確率が同じ

などなど。今日はこのような言葉で表したあいまいな確率と違って、数学の表現を利用してより厳密的に確率の意味を説明する。

2.3 確率に関する例

定義を与える前に、幾つかの例で印象を付けてもらう。

例 3 コインを投げて100回、1000回を投げれば、大体表と裏を同じ回数５０回、５００回ぐらい出る。経験的に、コインを投げる場合、表と裏をでる確率は同じで *1/2* である。

例 4 正確に作ったサイコロを投げる。*1* が出る確率は *1/6* である。

例 5 太陽は西から沈む確率は *100*％（１）である。

2.4 確率と確率に関連する幾つかの概念

試行 実験、行動、操作、自然現象など。

標本空間 試行のありうるすべて結果の集まり。Ω と記する。

空集合 空である集合。ϕ と記する。ϕ は任意の集合の部分集合である

事象 試行によって生じた結果の集まりで、Ω 部分集合である。A, B などで表記する。$A \subset \Omega$。

主観的確率 経験したことのない事象の起きる可能性を主観的に判断して、確信度としての確率。たとえば、柔道の選手が「僕明日勝つ確率 90 % ある！」と言う。明日の試合は以前の試合と同じではない、選手の話は自分の確信度を表しているしかない。

統計的確率 仮にある試行を同じ条件で互いに無関係に n 回繰り返すことができるとする。n 回の中である事象 A が m 回起きたとする。n が無限大に近づくにつれ、m/n はある定数 $p = P(A)$ に収束する。この事象が起こる確率は p となる。p は統計的確率である。

2.5 確率の公理

どんなふうに確率を定義しても以下の三つの公理を満たさなければならない。

1. 任意の事象 A の確率 $P(A)$ は $0 \leq P(A) \leq 1$ である。
2. 確実に起こる事象 Ω の確率 $P(\Omega) = 1$。$P(\phi) = 0$。
3. 事象 A と B が必ず同時に起こらない場合、A または B 何れか起こる確率は $P(A \cup B) = P(A) + P(B)$。この場合、事象 A と B が排反するという。

2.6 厳密な測度論的な定義

定義 6 **(σ−algebra)** 集合体（集合の集合）$\mathcal{B}$ が以下の条件を満たすとき $\sigma-$ 集合体と言う。a, 空集合 $\phi \in \mathcal{B}$、b, $A \in \mathcal{B}$ であれば必ず $A^c \in \mathcal{B}$、3, $A_1, A_2, \cdots \in \mathcal{B}$ であれば $\cup_{i=1}^{\infty} A_i \in \mathcal{B}$。

定義 7 Ω の部分集合の族が $\mathcal{A}$ 与えられたとき、$\mathcal{A}$ を含む最小な $\sigma-$ 集合体 $\mathcal{B}$ はに $\mathcal{A}$ より生成した $\sigma-$ 集合体という。

定義 8 Ω と Ω の部分集合の族 $\mathcal{A}$ より生成した $\sigma-$ 集合体 $\mathcal{B}$ が与えられたとき、確率測度 $\mathbf{P}$ は以下の条件を満たす確率関数である。a, $\mathbf{P}(\Omega) = 1$,b, $\mathbf{P}(\phi) = 0$、c, $A_1, A_2, \cdots \in \mathcal{B}$ でかつ互いに背反であれば $\mathbf{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbf{P}(A_i)$。

2.7 確率の性質

1. $0 \leq P(A) \leq 1$
2. $P(A) = 1 - P(A^c)$

3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

4. もし $A \subset B$ ならば、$P(A) \leq P(B)$

2.8　確率に関する計算例

例題 9 コインを 2 回投げる、2 回表が出る確率はいくら。

解答 10 1 回投げて表と裏が出る確率が同じ 0.5 とする。2 回投げて、出るかの性のあるすべてのパターンを考える。

$$\begin{cases} 表 \begin{cases} 表 \\ 裏 \end{cases} \\ 裏 \begin{cases} 表 \\ 裏 \end{cases} \end{cases}$$

上の表で見れば、表表、表裏、裏表、裏裏 4 パターンの中の 1 パターンだから、明らかに確率は 1/4。

例題 11 上の例の続きで $A = \{$ 表が 2 回 $\}, B = \{$ 裏が一回以上 $\}$ とする。$P\{A \cup B\}$ はいくらになる?

解答 12 $\{A \cup B\} = \{$ 表が 2 回または裏が一回以上 $\}$、調べてみれば全部である。$P\{A \cup B\} = 1$。

例題 13 $C = \{$ 表が 2 回 $\}$ かつ $D = \{$ 裏が 2 回 $\}$ の確率求めてください。

解答 14 求める確率は $P(C \cap D)$ で、$C = \{$ 表表 $\}$、$D = \{$ 裏裏 $\}$、C と D には共通要素がないため、$C \cap D = \phi$、ただし ϕ は空を表す、空集合と呼ぶ。$P\{\phi\} = 0$。言葉で言うと、表が 2 回と裏が 2 回が同時に起きる可能性はあり得ない。

練習 15 例題の例の続きで、$P\{B\}$ を計算して、そして $P\{A \cup B\} = P\{A\} + P\{B\}$ を確かめよう。

練習 16 コインを 3 回投げる、2 回表が出る確率はいくら?{ 表 1 回以上または裏 2 回 } の確率はいくら?{ 表 1 回以上かつ裏 2 回 } の確率はいくら?

2.9　確率変数

確率変数が取る値はおのおのに一定な確率が対応している。

前回で説明した、コイン投げの実験では表を 1 とし裏を 0 としたら、コイン投げの結果を確率変数 X のとる値とそれに対応している確率は $X = 0$ の確率は $0.5, X = 1$ の確率も 0.5 である。サイコロの場合は $x = 1$ の確率は $\frac{1}{6}, x = 2$ の確率も $\frac{1}{6}, \ldots$。

2.9.1 離散確率変数

上述した二つの確率変数の例は確率変数 X は 0 と 1 または $0, 1, 2, 3, 4, 5, 6$ と飛び飛びとした値しかとらない、たとえば 0 と 1 の間の値を取ることがない、この場合は離散確率変数と呼ぶ。

確率変数が取る値とそれに対応する確率との関係は以下の表で表せる。

X が取る値	x_1	x_2	x_3	$\cdots$	x_4
対応する確率 $P(x)$	$P(x_1)$	$P(x_2)$	$P(x_3)$	$\cdots$	$P(x_4)$

ただしここでは x_i が昇順で並べてられているとする。ここでは $P(x)$ は**確率度数関数**と呼ばれる。$P(x_1), P(x_2)...$ は x の具体的な値に対応する確率を表す。

分布関数：分布関数を $F(x)$ で表すと集合 $\{X \leq x\}$ の確率で $F(x) = P(\{X \leq x\})$ となる。離散確率変数の場合は

$$
\begin{aligned}
F(x) &= P(\{X \leq x\}) \\
&= \sum_{x_i < x} P(x_i)
\end{aligned}
$$

上式の和記号の意味は x より小さい x_i に関して和を取る意味である。たとえば上の表の例では、$F(x_3) = P(X \leq x_3) = P(x_1) + P(x_2) + P(x_3)$。ただしここでは x_i が昇順で並べてられているとする。

確率変数の**期待値**：確率変数 X の期待値を $E(X)$ で表す。離散確率変数の場合

$$
E(X) = \sum_{i=1}^{n} x_i P(x_i).
$$

確率変数の**分散**：確率変数 X の分散を $V(X)$ で表して、離散確率変数の場合

$$
V(X) = E\left[(X - E(X))^2\right] = \sum_{i=1}^{n} \left[(x_i - E(X))^2 P(x_i)\right].
$$

定義 17 (確率分布) ある確率変数 X の取る値に対応する確率がある関数 $P(x)$ に対応するとき、この確率変数 X の**確率分布**は $P(x)$ に従うという。

2.9.2 離散確率分布の例

1. ベルヌーイ分布：

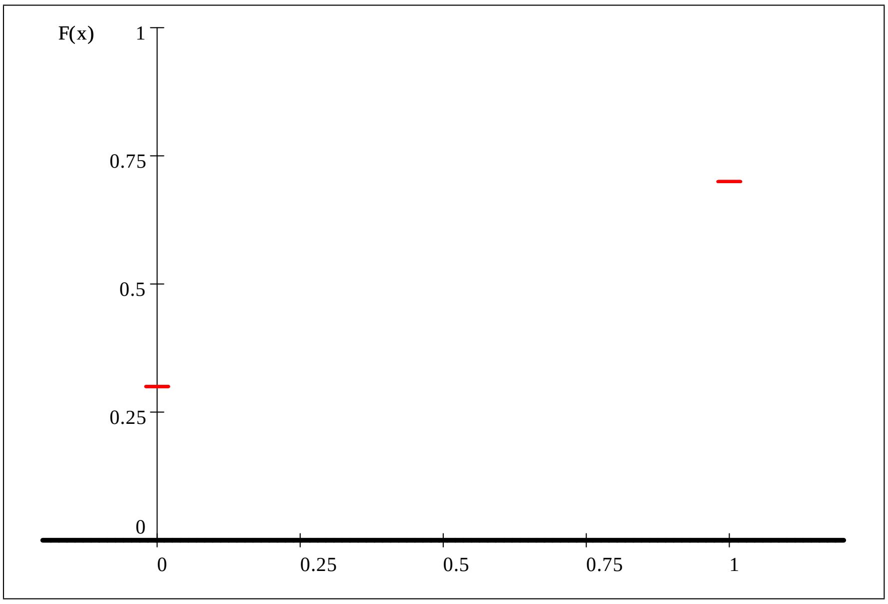
   確率度数関数

$$
\begin{cases}
P(x) = p & x = 1 \\
P(x) = 1 - p & x = 0
\end{cases}
$$

期待値 $E\left(X\right)=1\times p+0\times\left(1-p\right)=p.$ 分散は

$$\begin{aligned}V\left(X\right)&=\left(0-p\right)^2\times\left(1-p\right)+\left(1-p\right)^2\times p\\&=p^2-p^3+p-2p^2+p^3=p-p^2\end{aligned}$$

確率度数関数のグラフ

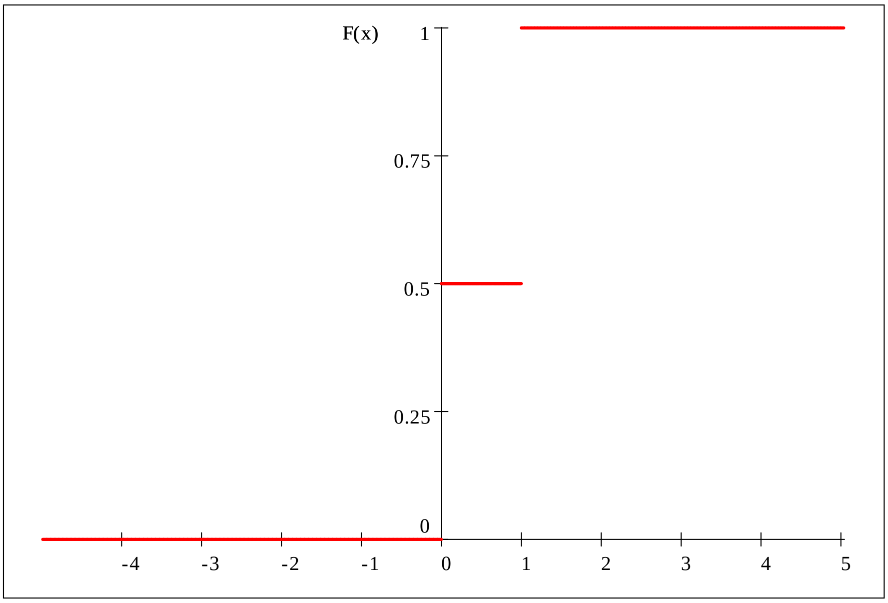


ベルヌーイの確率度数関数　$p=0.7$

確率分布関数

$$F\left(x\right)=\begin{cases}0 & x<0\\1-p & 0\leq x<1\\1 & x\geq 1\end{cases}$$

確率分布関数のグラフ



ベルヌーイ分布の分布関数 $p = 0.5$

コイン投げの結果はベルヌーイ分布のひとつの例になる。コイン一回投げた場合の出方の分布は $p = 0.5$ のベルヌーイ分布である。

2. 二項分布：

確率度数関数

$$P(x) = C_n^x p^x (1-p)^{n-x} \qquad x = 0, 1, 2, ..., n.$$

Y_i を 0 になる確率が p のベルヌーイ分布に従うとする。$X = \sum_{i=1}^n Y_i$

$$E(X) = E\left(\sum_{i=1}^n Y_i\right) = np$$
$$V(X) = V\left(\sum_{i=1}^n Y_i\right) = n(p - p^2)$$

ただしここでは C_n^x は組み合わせの計算を表している。

$$C_n^x = \frac{P_n^x}{P_x^x} = \frac{n!/(n-x)!}{x!}$$
$$= \frac{n(n-1)(n-2)\cdots(n-x+1)}{x(x-1)(x-2)\cdots 3 \times 2}$$

確率分布関数

$$F(x) = \sum_{i=1}^{x} P(x_i)$$

コイン投げはひとつ二項分布の例になる。コインを n 回を投げて x 回表が出る確率は $p=0.5$ の二項分布になる。

3. ポアソン分布

確率度数関数

$$P\left(x\right)=\frac{e^{-\lambda}\lambda^{x}}{x!}$$

ポアソン分布の度数関数はに $p=\lambda/n$ の二項分布の度数関数の極限として導出できる。

証明. $p=\lambda/n$ として、証明する。

$$\begin{aligned}C_{n}^{x}p^{x}\left(1-p\right)^{n-x}&=\frac{n!/\left(n-x\right)!}{x!}\left(\frac{m}{n}\right)^{x}\left(1-\frac{m}{n}\right)^{n-x}\\&=\frac{\frac{n}{n}\frac{(n-1)}{n}\frac{(n-2)}{n}\cdots\frac{(n-(x-1))}{n}}{x!}m^{x}\left(1-\frac{m}{n}\right)^{n}\left(1-\frac{m}{n}\right)^{-x}\end{aligned}$$

ここで $\frac{n}{n}\frac{(n-1)}{n}\frac{(n-2)}{n}\cdots\frac{(n-(x-1))}{n}\to1,\left(1-\frac{m}{n}\right)^{-x}\to1$ と $\left(1-\frac{m}{n}\right)^{n}\to e^{-m}$ を利用すれば、ポアソン分布の度数関数 $P\left(x\right)=\frac{e^{-\lambda}\lambda^{x}}{x!}$ が導かれる。 □

$$\begin{aligned}E\left(X\right)&=\lambda.\\V\left(X\right)&=\lambda.\end{aligned}$$

確率分布関数

$$F\left(x\right)=\sum_{i=1}^{x}P\left(x_{i}\right)$$

2.9.3 ポアソン分布の応用例

ポアソン分布はパラメーター n を無限大にして確率 p を極めて小さくした二項分布の極限として考えられるため、起きる確率が極めて小さい出来事を分析するとき使われる。たとえば、車の事故、地震が起きる確率など。たとえば、10 年間に平均的に 5 度強の地震は 10 回起きるとする。次の十年間に 5 度強の地震が 2 回以下起きる確率は

$$\begin{aligned}P\left(X\leq2\right)=F\left(2\right)&=\sum_{i=0}^{2}\frac{e^{-\lambda_{i}}\lambda^{x_{i}}}{x_{i}!}\\&=\frac{e^{-\lambda}\lambda^{0}}{0!}+\frac{e^{-\lambda}\lambda^{1}}{1!}+\frac{e^{-\lambda}\lambda^{2}}{2!}\end{aligned}$$

ここで $\lambda=10$ を代入して

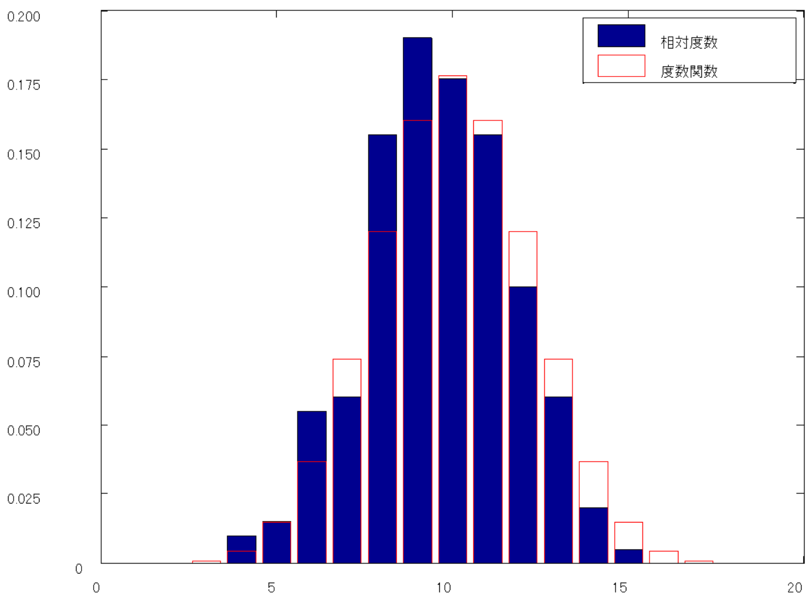
$$\begin{aligned}&=\frac{e^{-10}10^{0}}{0!}+\frac{e^{-10}10^{1}}{1!}+\frac{e^{-10}10^{2}}{2!}\\&=0.0028.\end{aligned}$$

二回以下起きる確率は 0.0028 である。

2.10 ヒストグラムと確率度数（密度）関数（Probability Density Funtion PDF）二項分布を例に

2.10.1 二項分布のヒストグラムと確率度数関数のグラフ

20 回コインを投げて、表が出た回数を記録し x とする、この実験を繰り返し 1000 回を行った、1000 個の x の値が得られる。x の 1000 個のデータを利用してヒストグラム（相対度数で描いたもの）を作成して、その上に二項分布の確率度数関数を重ね合わせたグラフ。



二項分布

二項分布の確率度数関数

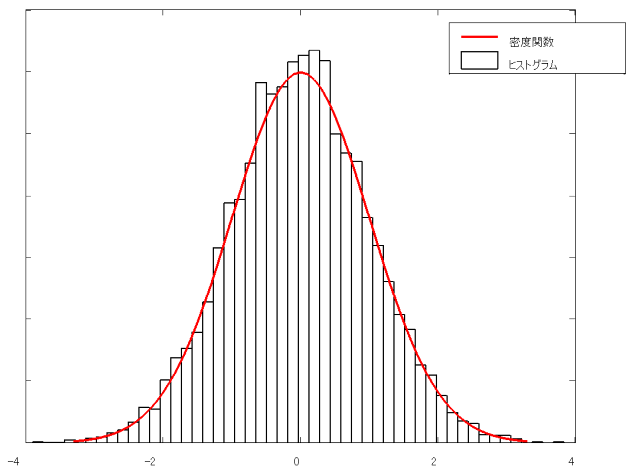
$$P\left(c\right)=C_{n}^{c}p^{c}\left(1-p\right)^{n}-c \qquad x=0,1,2,...,n. \tag{23}$$

理解しやすくするため確率度数関数、密度関数、累積分布関数を $P\left(c\right),f\left(c\right),F\left(c\right)$ などで表しているが、統計学の慣習では $P\left(x\right),f\left(x\right),F\left(x\right)$ で表記する。

2.10.2 正規分布のヒストグラムと確率度数関数のグラフ

シミュレーションで作った標準化した身長のデータのヒストグラム（相対度数で描いたもの）を作成して、その上に標準正規分布の確率密度関数を重ね合わせたグラフ。

正規分布の確率密度関数のグラフは鐘のまたは富士山の形をしている。



正規分布

正規分布の確率密度関数

$$f(c) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(c-\mu)^2}{2\sigma^2}} \tag{24}$$

2.10.3 ヒストグラムと度数関数（密度関数）の意味合いの比較

ヒストグラム

ヒストグラムから確率変数の実現値（実験の結果）の度数が占める全体の度数の割合を示している。

度数関数（密度関数） $f(x)$

度数関数や密度関数のグラフは確率の大きさを示している。

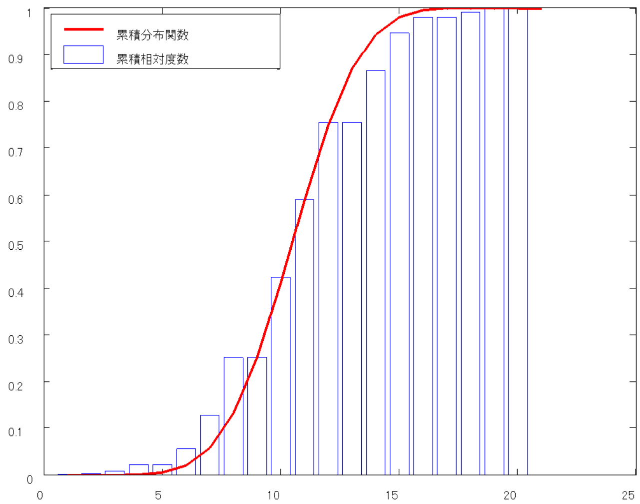
確率変数 X のある値 c での度数関数と密度関数の値が大きければ大きいほど、c に対応している確率大きい。

図 1 で言えば、値 10 に対応している確率がもっとも大きい。

図 2 の場合だと、0 に対応している確率密度がもっとも高い。0 近辺の値が起きやすい。（連続確率関数の場合、特定の値に確率を対応させることができなく、区間に対して確率が振り分けられている。特定の値に対応させているのは確率密度である。）

3 累積相対度数と累積分布関数 (Cumulative Distribution Function CDF)

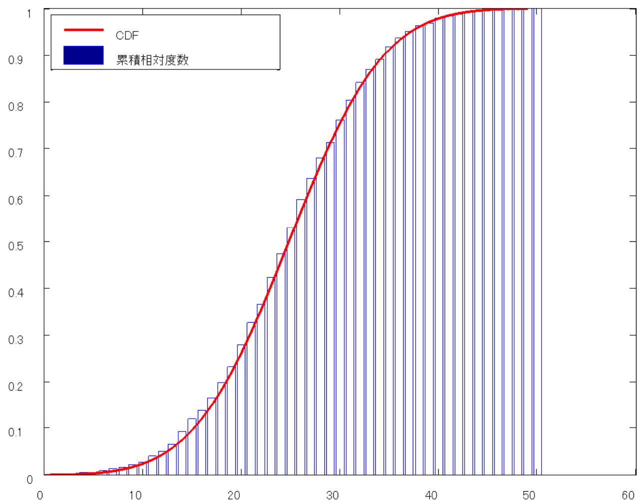
3.0.4 二項分布の累積相対度数と累積分布関数のグラフ



二項分布の累積分布関数

$$F(c) = P(X \leq c) = \sum_{i=1}^{c} P(x_i) = \sum_{i=1}^{c} C_n^{x_i} p^{x_i} (1-p)^{n-x_i} \qquad x = 0, 1, 2, ..., n. \quad (25)$$

3.0.5 標準正規分布の累積相対度数と累積分布関数のグラフ



正規分布の累積分布関数

$$F(c) = P(X \leq c) = \int_{-\infty}^{c} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \tag{26}$$

3.0.6 累積相対度数関数と累積分布関数の意味合いの比較

累積相対度数

k 番目の階級の累積相対度数は k 番目の階級より低い階級（k を含む）の 度数の総和 である。

累積分布関数 $F(x)$：

ある値 c に対応する累積分布関数の値は c より小さい値（c を含む）に対応する 確率の総和 である。

3.0.7 確率度数関数 $P(c)$（確率密度関数 $f(c)$）と累積分布関数 $F(c)$ の関係

離散確率変数：ある二つの値 a と b（$a < b$）が存在して、確率変数 X がこの二つの値 a と b を取れるが、a と b の間の値を取れない場合、X が離散確率変数であるという。例：コイン投げの結果、サイコロの出る目。

連続確率変数：確率変数 X が取れる値の中から異なる任意の二つの値 c と d を取り出して、c と d の間の任意の値を取れるなら、X が連続確率変数である。例：人の身長、リンゴの重さ。

離散確率変数の場合

累積分布関数は確率密度関数の和の形になっている。数式で表すと

$$F(c) = P(X \leq c) = \sum_{i=1}^{c} P(x_i) \tag{27}$$

。

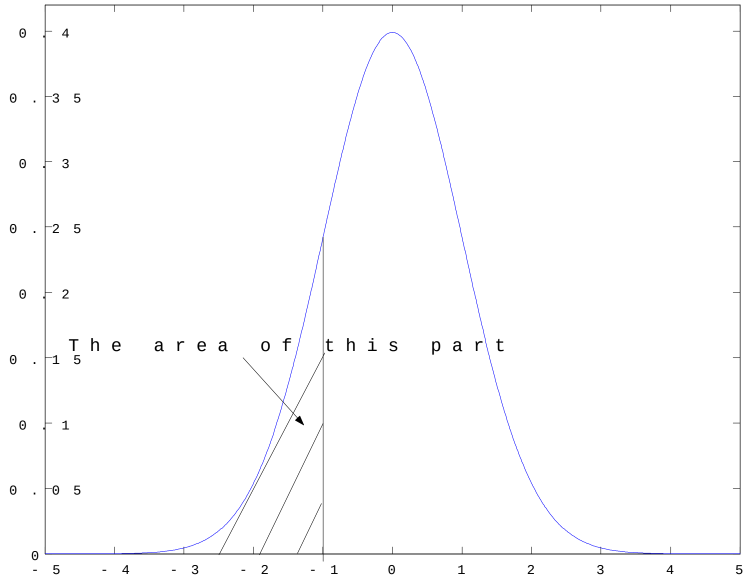
連続確率変数の場合

累積分布関数は確率密度関数の積分の形になっている。数式で表すと

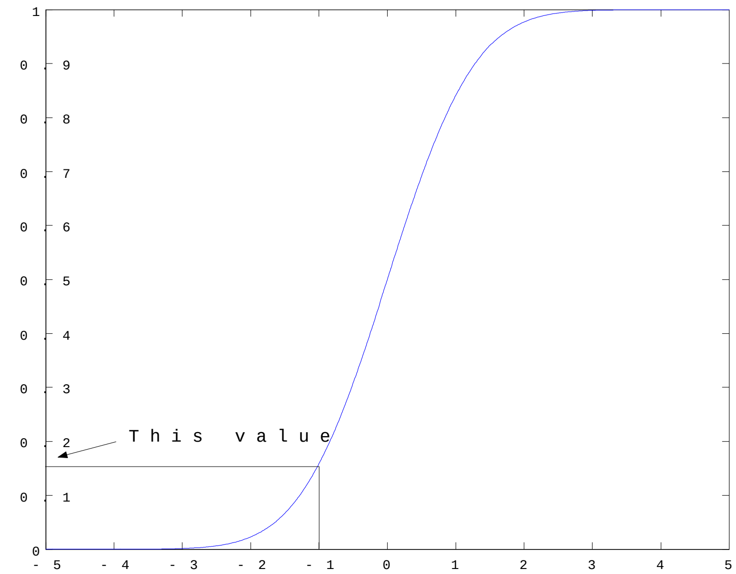
$$F(c) = P(X \leq c) = \int_{-\infty}^{c} f(x)\,dx \tag{28}$$

積分が図形の**面積**の計算に対応していることから、連続確率変数の場合に関して、標

準正規分布を例にグラフに描けば以下のようになる：



$F(-1)$ の値を密度関数のグラフで表す



$F(-1)$ の値を累積分布関数のグラフで表す

3.0.8 期待値と分散

前回は既に離散確率変数の期待値と分散に関して説明した。離散確率変数の場合

$$E(X) = \sum_{i=1}^{n} x_i P(x_i).$$

確率変数の**分散**：確率変数 X の分散を $V(X)$ で表して、離散確率変数の場合

$$V(X)=E\left[(X-E(X))^2\right]=\sum_{i=1}^{n}\left[(x_i-E(X))^2P(x_i)\right].$$

連続関数の場合は

$$E(X)=\int_{-\infty}^{\infty}xf(x)\,dx.$$

$$V(X)=E\left[(X-E(X))^2\right]=\int_{-\infty}^{\infty}(x-E(X))^2f(x)\,dx.$$

統計学用語のまとめ

標本の場合	度数	度数分布	累積相対度数	平均	分散（データから）
母集団の場合	確率	確率密度関数	累積分布関数	期待値	分散（密度関数から）

3.1 二項分布と正規分布のまとめ

1. 二項分布：

   確率度数関数

$$P(x)=C_n^xp^x(1-p)^{n-x}\qquad x=0,1,2,...,n. \tag{29}$$

   累積分布関数

$$F(x)=\sum_{i=1}^{x}P(x_i) \tag{30}$$

   期待値と分散

$$E(X)=E\left(\sum_{i=1}^{n}Y_i\right)=np \tag{31}$$

$$V(X)=V\left(\sum_{i=1}^{n}Y_i\right)=n(p-p^2). \tag{32}$$

2. 正規分布：

   確率密度関数

$$f(c)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(c-\mu)^2}{2\sigma^2}} \tag{33}$$

   累積分布関数

$$F(c)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{c}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx \tag{34}$$

期待値と分散

正規分布の期待値はμ **分散が**σ^2：

$$\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} xf(x)\,dx \\
&= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{c} xe^{-\frac{(x-\mu)^2}{2\sigma^2}}dx \\
&= \mu \\
V(X) &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx \\
&= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{c} (x-\mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx \\
&= \sigma^2
\end{aligned}$$

3.2 確率ベクトル

確率変数の測度論的な定義 d 次元確率ベクトル $\mathbf{X} = (x_1, \cdots, x_d)'$ は確率空間 $(\boldsymbol{\Omega}, \mathcal{F}, \mathbf{P})$ にある $\Omega \to \mathbb{R}^d$ の可測関数である。

直感的な理解 d 次元確率ベクトル $\mathbf{X} = (x_1, \cdots, x_d)'$ は取る値はおのおのに一定な確率が対応している。

例 離散の二次元確率ベクトルの場合、二回コインを投げて最初の一回が表で 2 回目が裏の確率

$$P\left(\mathbf{X} = (0,1)'\right) = P\left((x_1, x_2) = (0,1)\right) = 0.25$$

表記を簡単にするため以下から二変量確率ベクトル $X = (X_1, X_2)'$ について述べる。

分布関数 CDF $F(X_1, X_2) = \mathbf{P}(X_1 \leq x_1, X_2 \leq x_2)$

連続の CDF に対応する PDF

$$f(x_1, x_2) = \frac{\partial^2 F(x_1, x_2)}{\partial x_1 \partial x_2}$$

有用な公式

$$\mathbf{P}(x \in \mathbf{A}) = \int\int_A f(x_1, x_2)\,dx_1 d_{x_2}$$

$$\mathbf{E}(g(\mathbf{X})) = \int_R\int_R g(x_1, x_2) f(x_1, x_2)\,dx_1 d_{x_2}$$

周辺分布

- 分布関数

$$
\begin{aligned}
F_{X_1}(x_1) &= \mathbf{P}(X_1 \leq x_1) \\
&= \lim_{x_2 \to \infty} F(x_1, x_2) \\
&= \int_{-\infty}^{x_1} \int_{-\infty}^{\infty} f(x_1, x_2)\, dx_2 dx_1
\end{aligned}
$$

- 密度関数

$$
f_{X_1}(x_1) = \frac{\partial F_{X_1}(x_1)}{\partial x_1} = \int_{-\infty}^{\infty} f(x_1, x_2)\, dx_2
$$

3.3 条件付分布と条件付期待値

定義 18 条件付分布 $X = x$ の条件のもとでの確率変数 Y の条件付密度関数

$$
f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)}
$$

ただし、$f(x, y)$ は X と Y の同時密度関数、$f_X(x)$ は X の周辺密度関数である。

$$
\begin{aligned}
f_{Y|X}(y|x) &= \frac{\partial}{\partial y} \lim_{\varepsilon \to 0} \mathbf{P}(Y \leq y | x \leq X \leq x + \varepsilon) \\
&= \frac{\partial}{\partial y} \lim_{\varepsilon \to 0} \frac{\mathbf{P}(\{Y \leq y\} \cap \{x \leq X \leq x + \varepsilon\})}{\mathbf{P}(\{x \leq X \leq x + \varepsilon\})} \\
&= \frac{\partial}{\partial y} \lim_{\varepsilon \to 0} \frac{F(x + \varepsilon, y) - F(x, y)}{f_X(x)} \\
&= \frac{\partial}{\partial y} \lim_{\varepsilon \to 0} \frac{\frac{\partial}{\partial x} F(x + \varepsilon, y)}{f_X(x + \varepsilon)} \\
&= \lim_{\varepsilon \to 0} \frac{\frac{\partial^2}{\partial x \partial y} F(x + \varepsilon, y)}{f_X(x + \varepsilon)} \\
&= \frac{f(x, y)}{f_X(x)}
\end{aligned}
$$

定義 19 (条件付期待値)

$$
E(Y|X = x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)\, dy
$$

定義 20 (条件付分散)

$$
\begin{aligned}
& var(Y|X = x) \\
&= E\left([Y - E(Y|X = x)]^2 \,|X = x \right) \\
&= E\left(Y^2 | X = x \right) - (E(Y|X = x))^2
\end{aligned}
$$

Remark 条件付期待値および条件付き分散は X の関数で確率変数になる。

定理 21 (Simple Law of Iterated Expections)

$$E\left(E\left(Y|X\right)\right)=E\left(Y\right)$$

証明.

$$\begin{aligned}
E\left(E\left(Y|X\right)\right) &= E\left(\int_{-\infty}^{\infty} y f_{Y|X}\left(y|x\right)dy\right) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{Y|X}\left(y|x\right)dy f_X\left(x\right)dx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\frac{f\left(x,y\right)}{f_X\left(x\right)}f_X\left(x\right)dxdy \\
&= \int_{-\infty}^{\infty} y f\left(x,y\right)dxdy \\
&= E\left(Y\right)
\end{aligned}$$

□

定理 22 (Law of Iterated Expections)

$$E\left(E\left(Y|X,Z\right)|X\right)=E\left(Y|X\right)$$

証明.

$$\begin{aligned}
E\left(E\left(Y|X,Z\right)|X\right) &= E\left(\int_{-\infty}^{\infty} y f_{Y|X,Z}\left(y|x,z\right)dy|X\right) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{Y|X,Z}\left(y|x,z\right)dy f_{Z|X}\left(z\right)dz \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\frac{f\left(x,y,z\right)}{f_{X,Z}\left(x,z\right)}dy f_{Z|X}\left(z\right)dz \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\frac{f\left(x,y,z\right)}{f\left(x,z\right)}\frac{f\left(z,x\right)}{f_X\left(x\right)}dydz \\
&= \int_{-\infty}^{\infty} y\frac{f\left(x,y\right)}{f_X\left(x\right)}dy \\
&= E\left(Y|X\right)
\end{aligned}$$

□

定理 23 (Conditioning Theorem) 任意の関数 $g\left(x\right)$ に対して

$$E\left(g\left(X\right)Y|X\right)=g\left(X\right)E\left(Y|X\right) \tag{35}$$

3.4 確率変数の関数の分布（Change-of-vatiables）

3.4.1 一変量の場合

X の分布が分かった場合、$Y = g(X)$ の分布を考える。ただし、$g(x)$ が一対一で微分可逆な関数である。ここで $h(y)$ が $g(x)$ の逆関数とする。分布関数は

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(g(X) \leq y) \\ &= P(X \leq h(y)) \\ &= F_X(h(y)). \end{aligned}$$

密度関数は

$$\begin{aligned} f_Y(y) &= \frac{\partial}{\partial y} F_X(h(y)) \\ &= \frac{\partial}{\partial h(y)} F_X(h(y)) \frac{\partial h(y)}{\partial y} \\ &= f_X(h(y)) \frac{\partial h(y)}{\partial y} \end{aligned}$$

多変量の場合に一般化して

$$f_Y(y) = f_X(h(y)) |J(y)| \tag{36}$$

ただし、$|J(y)| = \left\| \frac{\partial \mathbf{h}(\mathbf{y})}{\partial \mathbf{y}'} \right\|$。$\|\cdot\|$ は行列式の絶対値を意味する。$J(y)$ はヤコビアンである。

4 統計入門

4.1 特性関数（Characristic Function）

確率変数 X の特性関数以下のように定義される

$$\psi(t, X) = E\left(e^{it'X}\right) \tag{37}$$

ただし $i = \sqrt{-1}$ は虚数単位である。X が如何なる分布に従うとしてもその特性関数が存在する。確率変数の特性関数と分布は一対一の関係を持つ。すなわち、二つの確率変数がある場合特性関数が同じであれば分布も同じである。その逆も成り立つ。

定理 24 (Euler's formula オイラーの公式)

$$e^{it} = \cos(t) + i \cdot \sin(t) \tag{38}$$

特性関数に関してオイラーの公式を適用して、使いやすい変形を導出できる

$$\begin{aligned} \psi(t, X) &= E\left(e^{it'X}\right) \\ &= \cos\left(t'X\right) + i \cdot \sin\left(t'X\right) \end{aligned}$$

4.2 正規分布の性質

- 多次元正規分布の密度関数

$$f_X(X) = (2\pi)^{-2n} |\Sigma|^{-1/2} \exp\left(1/2 [X-\mu]' \Sigma^{-1} [X-\mu]\right) \tag{39}$$

- 多次元正規分布の特性関数

$$\begin{aligned}\phi_X(t) &= E\left(e^{it'X}\right) \\ &= \exp\left(it'\mu - 1/2t'\Sigma t\right)\end{aligned}$$

- $X \sim N(\mu, \Sigma)$ ならば、$AX + b \sim N(A\mu + b, A\Sigma A')$

証明その一. A が正値定符号の場合だけを証明する。$Y{=}AX + b$ とする。$X = A^{-1}(Y-b)$

$$\begin{aligned}f_Y(Y) &= f_X\left(A^{-1}(Y-b)\right) \left|\left|A^{-1}\right|\right| \\ &= (2\pi)^{-2n} |\Sigma|^{-1/2} \left|\left|A^{-1}\right|\right| \exp\left(1/2 \left[A^{-1}(Y-b)-\mu\right]' \Sigma^{-1} \left[A^{-1}(Y-b)-\mu\right]\right) \\ &= (2\pi)^{-2n} |A\Sigma A'|^{-1/2} \exp\left(1/2 \left[A^{-1}(Y-b)-\mu\right]' A' (A')^{-1} \Sigma^{-1} (A)^{-1} A \left[A^{-1}(Y-b)-\mu\right]\right) \\ &= (2\pi)^{-2n} |A\Sigma A'|^{-1/2} \exp\left(1/2 [Y-(A\mu+b)]' (A\Sigma A')^{-1} [Y-(A\mu+b)]\right)\end{aligned}$$

□

証明その二. 正規分布の特性関数が

$$\begin{aligned}\phi_X(t) &= E\left(e^{it'X}\right) \\ &= \exp\left(it'\mu - 1/2t'\Sigma t\right)\end{aligned}$$

であることと特性関数が同一であれば分布も同じであることを利用して証明する。

$Y{=}AX + b$ とあらわす。

$$\begin{aligned}\phi_Y(t) &= E\left(e^{it'Y}\right) \\ &= E\exp\left(it'(AX+b)\right) \\ &= E\left[\exp\left(it'AX\right)\right]\exp\left(it'b\right) \\ &= E\left[\exp\left(i(A't)'X\right)\right]\exp\left(it'b\right) \\ &= \exp\left(i(A't)'\mu - 1/2(A't)'\Sigma(A't)\right)\exp\left(it'b\right) \\ &= \exp\left(it'(A\mu+b) - 1/2t'A\Sigma A't\right)\end{aligned}$$

□

練習 25 以下を証明しなさい、*If* $X \sim N(\mu, \Sigma)$,*then*$\Sigma^{-1/2}(X-\mu) \sim N(\mathbf{0}, \mathbf{1})$.

- If $X \sim N(\mu, \Sigma)$,then$(X-\mu)' \Sigma^{-1} (X-\mu) \sim \chi^2_n$. χ^2_n は自由度 n のカイ二乗分布。???? カイ二乗分布の説明。

4.3 Asymptotic Theory

定理 26 (Markov's Inequality) y を非負な確率変数とする。任意の正の実数 δ に関して以下の式が成り立つ

$$P\left(y>\delta\right)\leq\frac{E\left(y\right)}{\delta}.$$

証明.

$$E\left(y\right)=E\left(y|y>\delta\right)P\left(y>\delta\right)+E\left(y|y\leq\delta\right)P\left(y\leq\delta\right)$$

である。$E\left(y|y>\delta\right)>\delta$ なので、

$$E\left(y\right)\geq\delta P\left(y>\delta\right)+E\left(y|y\leq\delta\right)P\left(y\leq\delta\right)$$

がわかる。さらに、y が非負なので、$E\left(y|y\leq\delta\right)P\left(y\leq\delta\right)$ が非負である。ゆえに

$$E\left(y\right)\geq\delta P\left(y>\delta\right)$$

証明完了。 □

定理 27 (Chebychev's Inequality) x を確率変数とする。任意の実数 c と ε に関して以下の式が成り立つ

$$P\left(|x-c|>\varepsilon\right)\leq\frac{E\left[\left(x-c\right)^2\right]}{\varepsilon^2}.$$

練習 28 *Markov's Inequality* を利用して *Chebychev's Inequality* を証明しなさい。

4.3.1 収束の定義

定義 29 (確率収束) $^{\forall}\varepsilon>0$

$$\lim P\left(|x_n-x|>\varepsilon\right)=0$$

$$\operatorname{plim} x_n=x$$

x_n が x に確率収束するという。

定義 30 (平均二乗収束) x_n が x に平均二乗収束であるとは

$$\lim E\left(x_n-x\right)^2=0$$

定義 31 (定数に Almost Sure Convergence 概収束)

$$P\left(\lim x_n=c\right)=1 \tag{40}$$

定義 32 (分布収束) $F\left(x\right)$ のすべての連続の点において

$$\lim\left|F_n\left(x_n\right)-F\left(x\right)\right|=0 \tag{41}$$

$$x_n\xrightarrow{d}x \tag{42}$$

x_n が x に分布収束するという。

問題 33 平均二乗収束であれば確率収束であることを示せ。
　定義から、x_n が x に平均二乗収束であるとは

$$\lim E\left(x_n - x\right)^2 = 0 \tag{43}$$

である。任意の正の実数 ε に関して *Chebychev's Inequality* から

$$P\left(|x_n - x| > \varepsilon\right) \leq \frac{E\left(x_n - x\right)^2}{\varepsilon^2} \tag{44}$$

両辺の $\lim$ を取り

$$\lim P\left(|x_n - x| > \varepsilon\right) \leq \frac{\lim E\left(x_n - x\right)^2}{\varepsilon^2} \tag{45}$$

$\lim P\left(|x_n - x| > \varepsilon\right) = 0$ が分かる。証明完了。

練習 34 以下の分布を持つ確率変数 x_n が 0 に確率収束することを証明しなさい

$$\begin{cases} P\left(x_n = 0\right) = \frac{n-1}{n} \\ P\left(x_n = 1\right) = \frac{1}{n} \end{cases}$$

定義 35 (Consistent Estimator 一致推定量) パラメーター θ とその推定量 $\hat{\theta}$ に関して

$$\operatorname*{plim}_{n\to\infty} \hat{\theta} = \theta$$

を満たすとき、$\hat{\theta}$ が θ の一推定量である。

定理 36 (標本平均の一致性) 有限な母平均 μ と母分散 σ^2 を持つ確率変数 x のランダムなサンプル（無作為標本）の標本平均 $\bar{x}$ が μ の一致推定量である。

証明. $\bar{x}$ が μ に平均二乗収束することを証明していく。

$$\begin{aligned}
&\lim E\left[(\bar{x} - \mu)^2\right] \\
&= \lim \left\{E\left(\bar{x}^2\right) - 2\mu^2 + \mu^2\right\} \\
&= \lim \left\{E\left(\bar{x}^2\right) - \mu^2\right\} \\
&= \lim \left\{E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2 - \mu^2\right\} \\
&= \lim \frac{1}{n^2}\left\{\sum_{i=1}^{n} E(x_i^2) + \sum_{i=1, i\neq j}^{n} E\left(x_i x_j\right) - n^2\mu^2\right\} \\
&= \lim \frac{1}{n^2}\left\{nE(x_i^2) + \left(n^2 - n\right)\mu^2 - n^2\mu^2\right\} \\
&= \lim \frac{1}{n}\left(Ex_i^2 - \mu^2\right) \\
&= \lim \frac{1}{n}\sigma^2 = 0
\end{aligned}$$

最後の等号は分散 σ^2 が有限であるため。平均二乗収束が証明された。平均二乗収束は確率収束を意味するため、

$$\operatorname*{plim}_{n\to\infty} \bar{x} = \mu$$

となり、$\bar{x}$ が一推定量である。 □

定理 37 (Khinchine's Weak Law of Large Number (i.i.d. のケース) 大数の法則) *i.i.d.*（独立同一分布）の標本 $x_i, i = 1, \cdots, n$ が有限の平均 $E(x_i) = \mu$ を持っていれば、$\text{plim}\,\bar{x}_n = \mu.$

証明. http://www.ssc.wisc.edu/~bhansen/econometrics/Econometrics.pdf　の Theorem C.2.1 の証明を参照せよ。 □

定理 38 (Lindeberg-Levy Central Limit Theroem, Lindeberg-Levy の中心極限定理) $x_1, x_2, \cdots, x_n$ が *i.i.d.* でかつ有限な平均 μ と分散 σ^2 を持つ場合、標本平均 $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$ が以下の性質を持つ

$$\sqrt{n}\left(\bar{x}_n - \mu\right) \xrightarrow{d} N\left(0, \sigma^2\right).$$

証明. http://www.ssc.wisc.edu/~bhansen/econometrics/Econometrics.pdf　の Theorem C.3.1. の証明を参照せよ。 □

5　最小二乗法の統計学的性質

5.1　線形回帰モデルに関する条件 (強いバージョン)

線形回帰モデルの最小二乗（ＯＬＳ）推定量がいい性質を持つために様々な前提条件を必要とする。一般的なテキストでは条件ではなく仮定と言うが、上記の意味合いの文脈では条件と呼ぶほうが適切である。このような条件は強さによって様々なバージョンがあるが、ここではまずその中の一つ強いバージョンを紹介する。ここでの強いというのは普通の経済データはなかなかこのような条件を満たさないことを意味する。だが、データが以下の条件を満たされれば、単純な最小二乗法で分析できる。

A1. 線形性 y と x が以下の線形関係を持つ

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{iK}\beta_K + \varepsilon_i \tag{46}$$

y : 非説明変数、$x_{i1}, x_{i2}, \cdots$: 定数項を含む説明変数、ε_i 撹乱項または誤差項。行列表現

$$y = X\beta + \varepsilon \tag{47}$$

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{K1} \\ x_{12} & x_{22} & \cdots & x_{K2} \\ \vdots & \vdots & \cdots & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{Kn} \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

A2. フルランク (Full rank) $rank\,(X) = K$。ただし、K は説明変数の数で X の列数である。

A3. $E\,(\varepsilon|X) = 0$、X と ε が無相関であることを意味する。

A4. $Var\,(\varepsilon|X) = \sigma^2 I$、ただし、$Var\,(\varepsilon|X) = E\left[(\varepsilon - E\,(\varepsilon|X))\,(\varepsilon - E\,(\varepsilon|X))'\,|X\right]$。

A5. X が非確率変数または外生変数。すなわち、X は ε とまったく無関係である。

A.6 正規性 $\varepsilon \sim N\,(0, \sigma^2 I)$。

5.2 各条件の意味合い

A1. 線形性 $y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{iK}\beta_K + \varepsilon_i$。この線形性必ずしも成り立たないときが少なくない。あくまでも経済指標などの間の関係の線形近似である。たとえば、生産関数の規模と収穫の関係に関しては、収穫逓増や逓減のとき線形で近似しきれない場合がある。その時非線形のモデルを考える。これについでは後の章で説明する。

A2. フルランク (Full rank) $rank\,(X) = K$。ただし、K は説明変数の数で X の列数である。この条件が満たされなければ最小二乗推定量 $\hat{\beta}$ が一意に決まらず、推定できない。(**??**) の導出を参照せよ。

練習 39 *Example* 2.5 を説明しなさい。

A3. $E\,(\varepsilon|X) = 0$、X と ε が無相関であることを意味する。$E\,(\varepsilon) = 0$、$Cov\,(\varepsilon_i, X) = 0$ を意味する。

問題 40 $E\,(\varepsilon) = 0$、$Cov\,(\varepsilon_i, X) = 0$ を証明しなさい。*Conditioning Theorem* を参照。

A4. $Var\,(\varepsilon|X) = \sigma^2 I$、ただし、$Var\,(\varepsilon|X) = E\left[(\varepsilon - E\,(\varepsilon|X))\,(\varepsilon - E\,(\varepsilon|X))'\,|X\right]$。異なった i と j に関しては $Cov\,(\varepsilon_i, \varepsilon_j) = 0$、$Var\,(\varepsilon) = \sigma^2 I$ を意味する。すなわち、分散均一で系列相関がないのである。講義中グラフを持って説明する。

練習 41 $Var\,(\varepsilon) = \sigma^2 I$ を証明しなさい。

A5. X が非確率変数または外生変数。すなわち、X は ε とまったく無関係である。$A5$ は仮定 $A3$ 及び $A4$ と意味合い上、重複するところがある。

A.6 正規性 $\varepsilon \sim N\,(0, \sigma^2 I)$。$A5$ と $A6$ により、最小二乗法の推定量の性質をより簡単に導出することができる。$\hat{\beta}$ の推定量としての性質が優れていることを保障してくれる。その性質を利用して $\hat{\beta}$ の信頼性を検定できる。以下からこれらの条件の下で最小二推定量の性質を導出していく。

5.3 不偏推定量

定義 42 β の推定量 b が以下の性質を満たすとき、b が β の不偏推定量という。

$$E\left(b|X\right)=E\left(b\right)=\beta \tag{48}$$

性質 1 条件 $A1-A3$ が満たされるとき、最小二乗推定量 b が β の不偏推定量である。

証明.

$$b=\left(X'X\right)^{-1}X'y=\left(X'X\right)^{-1}X'\left(X\beta+\varepsilon\right)=\beta+\left(X'X\right)^{-1}X'\varepsilon \tag{49}$$

$$E\left(b\right)=E\left[\beta+\left(X'X\right)^{-1}X'\varepsilon\right]=\beta+E\left(E\left(\left(X'X\right)^{-1}X'\varepsilon|X\right)\right)=\beta \tag{50}$$

$$E\left(b|X\right)=E\left[\beta+\left(X'X\right)^{-1}X'\varepsilon|X\right]=\beta+E\left(\left(X'X\right)^{-1}X'\varepsilon|X\right)=\beta \tag{51}$$

□

5.4 最小二乗推定量の分散

$$\begin{aligned}
V\left(b\right)&=E\left(b-\beta\right)\left(b-\beta\right)'\\
&=E\left(\left(X'X\right)^{-1}X'\varepsilon\right)\left(\left(X'X\right)^{-1}X'\varepsilon\right)'\\
&=E\left(\left(X'X\right)^{-1}X'\varepsilon\varepsilon'X\left(X'X\right)^{-1}\right)\\
&=E\left[E\left(\left(X'X\right)^{-1}X'\varepsilon\varepsilon'X\left(X'X\right)^{-1}|X\right)\right]\\
&=E\left[\left(X'X\right)^{-1}X'E\left(\varepsilon\varepsilon'|X\right)X\left(X'X\right)^{-1}\right]\\
&=E\left[\left(X'X\right)^{-1}X'\sigma^2IX\left(X'X\right)^{-1}\right]\\
&=\sigma^2E\left[\left(X'X\right)^{-1}\right]
\end{aligned} \tag{52}$$

第 6 番の等式を導出するために性質 A4 が使われた。

性質 2 従って、X が非確率変数の時

$$V\left(b\right)=\sigma^2\left(X'X\right)^{-1} \tag{53}$$

X が確率変数の時

$$\begin{aligned}
V\left(b|X\right)&=E\left[\left(b-\beta\right)\left(b-\beta\right)'|X\right]\\
&=E\left(\left(X'X\right)^{-1}X'\varepsilon\varepsilon'X\left(X'X\right)^{-1}|X\right)\\
&=\left(X'X\right)^{-1}X'E\left(\varepsilon\varepsilon'|X\right)X\left(X'X\right)^{-1}\\
&=\left(X'X\right)^{-1}X'\sigma^2IX\left(X'X\right)^{-1}\\
&=\sigma^2\left(X'X\right)^{-1}
\end{aligned} \tag{54}$$

5.5 最小二乗推定量の分布

$b=\beta+\left(X'X\right)^{-1}X'\varepsilon$ であるのと $\varepsilon\sim N\left(0,\sigma^2 I\right)$ であることに注意して、条件 A1-A6 の下で、以上の結果と正規分布の線形結合の性質を利用すれば、以下の結果が得られる

性質 3 X が非確率変数の場合、

$$b\sim N\left(\beta,\sigma^2\left(X'X\right)^{-1}\right)$$

X が確率変数の場合

$$b|X\sim N\left(\beta,\sigma^2\left(X'X\right)^{-1}\right)$$

5.6 Gauss-Markov Theorem(BLUE, Best Linear Unbiased Estimator)

条件 A1-A6 が満たされれば、最小二乗法は最小分散線形不偏推定量（ＢＬＵＥ）となる。

すなわち、最小二乗推定量が b とする、それと異なる線形不偏推定量を $b_0=Cy$ とする。

必ず

$$V\left(b\right)\leq V\left(b_0\right)$$

となる。

ここでの線形の意味は推定量は撹乱項 ε の線形係数であることです。

証明. b_0 が不偏推定量から

$$\begin{aligned}E\left(b_0|X\right)&=\beta\\E\left(CX'\beta+C\varepsilon|X\right)&=\beta\\CX'\beta&=\beta\\CX'&=I\end{aligned}$$

次に

$$\begin{aligned}V\left(b_0|X\right)&=E\left[\left(Cy-\beta\right)\left(Cy-\beta\right)'|X\right]\\&=E\left[C\varepsilon\left(C\varepsilon\right)'|X\right]\\&=\sigma^2CC'\\&=\sigma^2\left[\left(\left[C-\left(X'X\right)^{-1}X\right]+\left(X'X\right)^{-1}X\right)\left(\left[C-\left(X'X\right)^{-1}X\right]+\left(X'X\right)^{-1}X\right)'\right]\end{aligned}$$

$CX'=I$ から $\left[C-\left(X'X\right)^{-1}X\right]\times\left[\left(X'X\right)^{-1}X\right]'=0$ なので

$$V\left(b_0|X\right)=\sigma^2\left(X'X\right)^{-1}+\sigma^2\left(\left[C-\left(X'X\right)^{-1}X\right]\left[C-\left(X'X\right)^{-1}X\right]'\right)$$

さらに任意の次元が合うベクトルqに関して$q'\left[C-(X'X)^{-1}X\right]\left[C-(X'X)^{-1}X\right]'q=z'z\geq 0$なので$\left[C-(X'X)^{-1}X\right]\left[C-(X'X)^{-1}X\right]$は非負値定符号行列となる。証明完了。□

5.7 Minimum Mean Squared Error（最小ＭＳＥ予測値）

定義 43 $MSE=E\left(\hat{\theta}-\theta\right)^2$

yのXによる線形予測を$\hat{y}=X'\beta$とする。予測MSEは

$$MSE=E\left(y-X'\beta\right)^2$$

このMSEの最小化条件は

$$\begin{aligned}\frac{\partial MSE}{\partial\beta}&=0\\E\left(Xy\right)&=E\left(XX'\right)\beta\end{aligned}$$

となる。これの標本版は最小二乗法の正規方程式

$$\frac{1}{n}\sum_{i=1}^{n}X_iy_i=\left(\frac{1}{n}\sum_{i=1}^{n}X_iX_i'\right)b$$

になっている。ゆえに、大数の法則が満たされる場合、bは予測MSEを最小にするような予測値$X'\beta$の中の係数の推定量であると考えられる。

5.8 分散σ^2の推定

推定量として

$$\hat{\sigma}^2=\frac{1}{n}e'e$$

が考えられるが、しかし$\hat{\sigma}^2$は不偏推定量ではない。確認していく。残差メーカー$M=I-X\left(X'X\right)^{-1}X'$として

$$\begin{aligned}E\left(\hat{\sigma}^2\right)&=E\left(\frac{1}{n}e'e\right)\\&=E\left(\frac{1}{n}\left(My\right)'\left(My\right)\right)\\&=E\left(\frac{1}{n}\left[M\left(X\beta+\varepsilon\right)\right]'\left[M\left(X\beta+\varepsilon\right)\right]\right)\end{aligned}$$

$MX = \left(I - X\left(X'X\right)^{-1}X'\right)X = 0$ なので、上式は

$$
\begin{aligned}
&= E\left(\frac{1}{n}\left(M\varepsilon\right)' M\varepsilon\right) \\
&= E\left(\frac{1}{n}\varepsilon' M\varepsilon\right) \\
&= E\left(\frac{1}{n}tr\left(\varepsilon' M\varepsilon\right)\right) \\
&= E\left(\frac{1}{n}tr\left(\varepsilon\varepsilon' M\right)\right) \\
&= E\left\{E\left(\frac{1}{n}tr\left(\varepsilon\varepsilon' M\right)|X\right)\right\} \\
&= E\left\{\frac{1}{n}tr\left(E\left(\varepsilon\varepsilon'|X\right)M\right)\right\} \\
&= E\left\{\frac{1}{n}tr\left(M\right)\right\}\sigma^2 \\
&= E\left\{\frac{1}{n}tr\left(I - X\left(X'X\right)^{-1}X'\right)\right\}\sigma^2 \\
&= E\left\{\frac{1}{n}tr\left(I\right) - tr\left(X'X\left(X'X\right)^{-1}\right)\right\}\sigma^2 \\
&= \frac{n-K}{n}\sigma^2
\end{aligned}
$$

以上の結果を踏まえて

$$s^2 = \frac{n}{n-K}\frac{1}{n}e'e = \frac{1}{n-K}e'e$$

が σ^2 の不偏推定量であることがわかる。

そして、b の分散 $V\left(b|X\right) = \sigma^2\left(X'X\right)^{-1}$ の推定量は $s^2\left(X'X\right)^{-1}$ となる。s は回帰分析の標準誤差（standard error）で、$s^2\left(X'X\right)^{-1}$ の第 (k,k) 要素は b_k の標準誤差と呼ぶ。

5.9 検定

以上で導出した推定量 b の分布を利用して幾つかの検定統計量を構成できる。

5.9.1 t 統計量

$$t_k = \frac{b_k - \beta_k}{s_{b_k}} = \frac{b_k - \beta_k}{s\sqrt{\left[\left(X'X\right)^{-1}\right]_{kk}}}$$

ただし、$\left[\left(X'X\right)^{-1}\right]_{kk}$ は $\left(X'X\right)^{-1}$ の第 (k,k) 要素。t_k は自由度が $(n-K)$ の t 分布に従う。

有意水準が $a\%$ として、帰無仮説 $H_0 : \beta_k = 0$ 対立仮説 $H_a : \beta_k > 0$ の片側検定を行うとき、まず

$$t_k = \frac{b_k - 0}{s_{b_k}}$$

を計算して、そして、t_k の値を自由度 $(n-K)$ の $1-a\%$ の分位点の値と比べる。t_k の方が大きい場合、帰無仮説を棄却する、逆の場合、帰無仮説を採択する。

5.9.2 t 統計量の導出

X が非確率変数の場合、ベクトル z が標準正規分布に従うとする。M がべき等行列の場合、$z'Mz$ が自由度が $rank\,(M)$ の χ^2 分布に従う。

互いに独立な標準正規分布と χ^2_k 分布に従う二つの確率変数 x, y の比 $x/\sqrt{y/k}$ が自由度 k の t 分布に従う。

$z_k = (b_k - \beta_k) \Big/ \sqrt{\left[\sigma^2 \left(X'X\right)^{-1}\right]_{kk}}$ とする。b の分布から $z_k \sim N\,(0,1)$。さらに $v = (n-K)\,s^2\,/\sigma^2$ とする。

$$\begin{aligned} v &= e'e/\sigma^2 \\ &= \left(\frac{\varepsilon}{\sigma}\right)' M \left(\frac{\varepsilon}{\sigma}\right) \end{aligned}$$

$\varepsilon/\sigma \sim N\,(0,1)$ と M がべき等行列なので、v が自由度が $rank\,(M) = n-K$ の χ^2 分布に従う。

さらに

$$t_k = \frac{z_k}{\sqrt{v/\,(n-K)}} = \frac{b_k - \beta_k}{s_{b_k}}$$

となることがわかる。t_k が t 分布に従うことを証明するためには、z_k と v が独立であることを証明する必要がある。定理 $B.12$ から、$(X'X)^{-1}\,XM = 0$ であるので、独立が証明される。ゆえに t_k は互いに独立な標準正規分布と $\chi^2_{(n-K)}$ 分布に従う二つの確率変数 z_k, v の比 $z_k/\sqrt{v/\,(n-K)}$ となり、自由度 $(n-K)$ の t 分布に従う。

$X \in D$ が確率変数の場合、以上の結果から

$$t_k|X \sim t\,(n-K)$$

となることがわかる。さらに $f_{t_k}\,(t_k) = \int_D f_{t_k|X}\,(t_k)\,f_X\,(X)\,dX$ だが、$f_{t_k|X}\,(t_k)$ が X と無関係で自由度 $(n-K)$ の t 分布に従うことより $f_{t_k}\,(t_k) = f_{t_k|X}\,(t_k) \int_D f_X\,(X)\,dX = f_{t_k|X}\,(t_k)$ がわかる。すなわち

$$t_k \sim t\,(n-K)$$

条件有りと無しに関係なく t_k は自由度 $(n-K)$ の t 分布に従うがわかる。

5.10 F 統計量

定数項以外すべての係数が 0 である、すなわち回帰モデルはまったく説明力がないという仮説を検定するとき、

$$F = \frac{R^2 / (K-1)}{(1-R^2) / (n-K)}$$

は自由度 $(K-1, n-K)$ の F 分布に従う。t_k と同じように、F 値も条件有りと無しに関係なく同じ自由度 $(K-1, n-K)$ の F 分布に従う。

6 最小二乗法の大標本の性質（Large sample properties）

6.1 追加条件

A5a. 観測値 (X_i, ε_i) は i に関して独立、

$$p\lim\left(\frac{X'X}{n}\right) = Q$$

ただし Q が正値定符号。$E(X_i X_i') = Q_i < \infty$。

6.2 一致性

$$p\lim b = \beta$$

証明.

$$b = \beta + \left(\frac{X'X}{n}\right)^{-1}\left(\frac{X'\varepsilon}{n}\right)$$
$$p\lim(b) = \beta + Q^{-1} p\lim\left(\frac{X'\varepsilon}{n}\right)$$

[1]

$$\begin{aligned}
&\lim E\left(\frac{X'\varepsilon}{n}\frac{\varepsilon' X}{n}\right) \\
&= \lim E\left[E\left(\frac{X'\varepsilon}{n}\frac{\varepsilon' X}{n}|X\right)\right] \\
&= \lim E\left[\frac{1}{n^2}X'E\left(\varepsilon\varepsilon'|X\right)X\right] \\
&= \lim\left(\frac{\sigma^2}{n}E\left(\frac{X'X}{n}\right)\right) \\
&= 0\cdot Q = \mathbf{0}
\end{aligned}$$

[2] ゆえに $\frac{X'\varepsilon}{n}$ の各要素が 0 に平均二乗収束、すなわち $\lim E\left(\frac{1}{n}X_j'\varepsilon\right)^2 = 0$、ゆえに 0 に確率収束する、$p\lim\left(\frac{X'\varepsilon}{n}\right) = \mathbf{0}$。$p\lim(b) = \beta$ がわかる。

さらに、(X_i, ε_i) が i に関して $i.i.d.$ であれば、大数の法則より仮定 $E(\varepsilon|X) = 0$ より、$p\lim\left(\frac{X'\varepsilon}{n}\right) = p\lim\left(\frac{1}{n}\sum_{i=1}^{n}X_i\varepsilon_i\right) = E(X_i\varepsilon_i) = 0$。線形回帰の最小二乗推定量の場合、大数の法則で証明する方法はよく利用される。（Cameron and Trivedi 2005 を参照。）
□

6.3　漸近正規性

定理 44 *Lindberg-Feller* 中心極限定理

$x_i, i = 1, 2, \cdots, n,$ が i に関して独立である。$E(x_i) = \mu_i < \infty$、$Var(x_i) = \sigma_i^2 < \infty$。$\bar{\mu} = \frac{1}{n}\sum_{i=1}^{n}\mu_i, \bar{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\sigma_i^2$ とする。もし $\lim\max(\sigma_i)/(n\bar{\sigma}_n) = 0$、かつ $\lim\bar{\sigma}_n^2 = \bar{\sigma}^2$ であれば

$$\sqrt{n}\left(\bar{x}_n - \bar{\mu}_n\right) \overset{a}{\sim} N\left(0, \bar{\sigma}^2\right)$$

定理 45 多変量 *Lindberg-Feller* 中心極限定理（テキストを参照）

定理 46 漸近正規性

$$\sqrt{n}\left(b-\beta\right) \overset{a}{\sim} N\left(0, \sigma^2 Q^{-1}\right)$$

証明.

$$\sqrt{n}\left(b-\beta\right) = \left(\frac{X'X}{n}\right)^{-1}\left(\frac{X'\varepsilon}{\sqrt{n}}\right)$$

$$Var\left(X_i\varepsilon_i\right) = \sigma^2 E\left(X_iX_i'\right) = \sigma^2 Q_i$$

---

[1] $p\lim(X_n) = a$、 $p\lim(Y_n) = b$ であれば、$p\lim(X_nY_n) = ab$。
[2] 条件

A4. $Var(\varepsilon|X) = \sigma^2 I$、ただし、$Var(\varepsilon|X) = E\left[\left(\varepsilon - E(\varepsilon|X)\right)\left(\varepsilon - E(\varepsilon|X)\right)'|X\right]$ と $E(\varepsilon|X) = 0$ は $E(\varepsilon\varepsilon'|X) = \sigma^2 I$ の十分条件。

$$
\begin{aligned}
Var\left(\frac{X'\varepsilon}{\sqrt{n}}\right) &= \sigma^2 \frac{Q_1 + Q_2 + \cdots + Q_n}{n} \\
&= \sigma^2 \bar{Q}_n
\end{aligned}
$$

条件 $A5a$ から

$$
\lim Var\left(\frac{X'\varepsilon}{\sqrt{n}}\right) = \lim \sigma^2 \bar{Q}_n = \sigma^2 Q
$$

さらに $E\left(X_i X_i'\right) = Q_i < \infty$ なので

$$
\lim \left(n\bar{Q}_n\right)^{-1} Q_i = 0
$$

多変量 Lindberg-Feller 中心極限定理の条件を満たす。

$$
\frac{X'\varepsilon}{\sqrt{n}} \xrightarrow{d} N\left(0, \sigma^2 Q\right)
$$

さらに $p\lim\left(\frac{X'X}{n}\right) = Q$ のため[3]

$$
\sqrt{n}\left(b - \beta\right) \xrightarrow{d} Q^{-1} \times N\left(0, \sigma^2 Q\right) = N\left(0, \sigma^2 Q^{-1}\right)
$$

□

7 検定

例 47 投資関数の検定を考えよう。テキスト *pp*.93。

$$
\ln I_t = \beta_1 + \beta_2 i_t + \beta_3 \triangle p_t + \beta_4 \ln Y_t + \beta_5 t + \varepsilon_t \tag{55}
$$
$$
\ln I_t = \beta_1 + \beta_2 \left(i_t - \triangle p_t\right) + \beta_3 \triangle p_t + \beta_4 \ln Y_t + \beta_5 t + \varepsilon_t \tag{56}
$$
$$
\ln I_t = \beta_1 + \beta_2 \left(i_t - \triangle p_t\right) + \beta_4 \ln Y_t + \beta_5 t + \varepsilon_t \tag{57}
$$

ただし、 I_t *investment,* i_t *nominal interest,* $\triangle p_t$ *the rate of inflation,* Y_t *GDP,* t *time trend.*

係数に関する制約を行列を利用して表現する

$$
\begin{aligned}
H_0 &: R\beta - q = 0 \\
H_1 &: R\beta - q \neq 0
\end{aligned}
$$

R と q をいろいろな値を取らせるより、検定したい仮定を表現する。

例 48 $H_0 : \beta_2 = 0$

$$
\begin{aligned}
R &= \left[\begin{array}{ccccc} 0 & 1 & 0 & \cdots & 0 \end{array}\right] \\
q &= 0
\end{aligned}
$$

[3] $X_n \xrightarrow{d} X$、$Y_n \xrightarrow{p} c$ である場合、$X_n Y_n \xrightarrow{d} cX$。

例 49 $H_0 : \beta_2 = \beta_3$

$$R = \begin{bmatrix} 0 & 1 & -1 & \cdots & 0 \end{bmatrix}$$
$$q = 0$$

例 50 $H_0 : \beta_1 + \beta_2 + \beta_3 = 2, \beta_4 = \beta_5$

$$R = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$
$$q = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

例 51 H_0 : 定数項以外の係数がすべて 0

$$R = [\mathbf{0} \quad I_{K-1}]$$
$$q = \mathbf{0}$$

χ^2 **検定** $m = Rb - q$ とする。b が正規分布に従うので m も正規分布に従う。帰無仮説の元で $R\beta - q = 0$ なので（帰無仮説が正しなら）

$$E(m|X) = RE(b|X) - q = R\beta - q = 0$$
$$Var(m|X) = \sigma^2 R(X'X)^{-1} R'$$

$W = m'(Var(m|X))^{-1} m$ とする。$W^{1/2} = m'(Var(m|X))^{-1/2}$ が標準正規分布に従う。ゆえに

$$W = (Rb - q)' \left[\sigma^2 R(X'X)^{-1} R'\right]^{-1} (Rb - q) \sim \chi^2_J$$

ただし J が R の行数で制約の数となる。この検定等計量 W を使えば χ^2 検定ができるが、しかし、σ が未知なので実行不可能である。

F **検定** σ が未知なので s を σ の代わりに使って検定を行う、そうすると誤差項が正規分布の仮定下では、検定等計量は F 分布になる。

$$\begin{aligned} F(J, n-K) &= \frac{(Rb - q)' \left[\sigma^2 R(X'X)^{-1} R'\right]^{-1} (Rb - q) \big/ J}{[(n-K)\, s^2/\sigma^2])/(n-K)} \\ &= (Rb - q)' \left[s^2 R(X'X)^{-1} R'\right]^{-1} (Rb - q) \Big/ J \end{aligned}$$

自由度 $(J, n-K)$ の F 分布[4]に従う。

[4]互いに独立な二つの χ^2 分布に従う確率変数それぞれの自由度で割った上での比が F 分布に従う。その自由度は分子と分母の自由度となる。

R が一行だけの場合、すなわち制約が一つの場合、$\hat{q} = Rb$ とする。

$$t = \frac{\hat{q} - q}{se\left(\hat{q}\right)}$$

自由度が $(n-K)$ の t 分布に従う。

$$t^2 = \frac{\left(\hat{q} - q\right)^2}{Var\left(\hat{q} - q|X\right)} = (Rb - q)' \left[s^2 R \left(X'X\right)^{-1} R'\right]^{-1} (Rb - q) \Big/ 1$$

8　制約付き最小二乗法

帰無仮説 $R\beta = q$ を制約条件として考える。制約 $Rb_0 = q$ のもとで

$$S\left(b_0\right) = \left(y - Xb_0\right)' \left(y - Xb_0\right)$$

の最小化を行うことによって、推定量 b_0 を求める。ラグランジュ乗数法を利用する。

$$L^*\left(b_0, \lambda\right) = \left(y - Xb_0\right)' \left(y - Xb_0\right) + 2\lambda' \left(Rb_0 - q\right)$$

一次条件は

$$\begin{aligned} \frac{\partial L^*}{\partial b_0} &= -2X' \left(y - Xb_0\right) + 2R\lambda = 0 \\ \frac{\partial L^*}{\partial \lambda} &= 2\left(Rb_0 - q\right) = 0 \end{aligned}$$

解くと

$$\begin{bmatrix} X'X & R' \\ R & 0 \end{bmatrix} \begin{bmatrix} b_0 \\ \lambda \end{bmatrix} = \begin{bmatrix} X'y \\ q \end{bmatrix}$$

分割行列の逆行列の公式を利用すれば

$$\begin{aligned} b_0 &= b - \left(X'X\right)^{-1} R \left[R \left(X'X\right)^{-1} R'\right]^{-1} (Rb - q) \\ \lambda &= \left[R \left(X'X\right)^{-1} R'\right]^{-1} (Rb - q) \end{aligned}$$

ただし、b は制約なしの普通の最小二乗推定量。更に

$$\begin{aligned} Var\left(b_0|X\right) &= \sigma^2 \left(X'X\right)^{-1} - \sigma^2 \left(X'X\right)^{-1} R' \left[R \left(X'X\right)^{-1} R'\right]^{-1} R \left(X'X\right)^{-1} \\ &= Var\left(b|X\right) - \sigma^2 \left(X'X\right)^{-1} R' \left[R \left(X'X\right)^{-1} R'\right]^{-1} R \left(X'X\right)^{-1} \end{aligned}$$

8.1 制約付き最小二乗法の結果を利用する F 検定

制約付き最小二乗法の残差を計算する

$$\begin{aligned} e_0 &= y - Xb - X\left(b_0 - b\right) = e - X\left(b_0 - b\right) \\ e_0' e_0 &= e'e + \left(b_0 - b\right)' X'X \left(b_0 - b\right) \geq e'e \\ e_0' e_0 - e'e &= \left(b_0 - b\right)' X'X \left(b_0 - b\right) \\ &= \left[\left(X'X\right)^{-1} R \left[R\left(X'X\right)^{-1} R'\right]^{-1} \left(Rb - q\right)\right]' X'X \left(X'X\right)^{-1} R \left[R\left(X'X\right)^{-1} R'\right]^{-1} \left(Rb - q\right) \\ &= \left(Rb - q\right)' \left[R\left(X'X\right)^{-1} R'\right]^{-1} \left(Rb - q\right) \end{aligned}$$

以上の結果と以前定義した

$$F\left(J, n-K\right) = \left(Rb - q\right)' \left[s^2 R \left(X'X\right)^{-1} R'\right]^{-1} \left(Rb - q\right) \Big/ J$$

から

$$F\left(J, n-K\right) = \frac{\left(e_0' e_0 - e'e\right) / J}{e'e / \left(n-K\right)}$$

分母と分子両方を $\sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2$ で割って F 統計量の別の表現が求まる

$$F\left(J, n-K\right) = \frac{\left(R^2 - R_0^2\right)}{\left(1 - R^2\right) / \left(n-K\right)} \tag{58}$$

ただし R_0^2 は制約付き最小二乗法の決定係数である。

8.2 正規仮定がない場合の検定

誤差項が正規分布に従うと仮定すれば

$$F\left(J, n-K\right) = \left(Rb - q\right)' \left[s^2 R \left(X'X\right)^{-1} R'\right]^{-1} \left(Rb - q\right) \Big/ J$$

が自由度 $(J, n-K)$ の F 分布に従う。しかし、誤差項が正規分布ではない場合、この結論は成り立たない。e_i が正規分布に従わず、$s^2 = e'e/\left(n-K\right)$ が χ^2 分布にならないためである。

大標本すなわち n が無限大になって場合、$F\left(J, n-K\right)$ を変形して

$$F\left(J, n-K\right) = \frac{\left(Rb - q\right)' \left[\sigma^2 R \left(X'X\right)^{-1} R'\right]^{-1} \left(Rb - q\right)}{\left(s^2/\sigma^2\right) J}$$

が分かる。さらに、$p\lim s^2 = \sigma^2$ のため

$$p\lim F\left(J, n-K\right) = \left(Rb - q\right)' \left[\sigma^2 R \left(X'X\right)^{-1} R'\right]^{-1} \left(Rb - q\right) / J$$

を利用して、J 倍の F 統計量

$$\begin{aligned} W &= J \times F\left(J, n-K\right) \\ &= \left(Rb-q\right)' \left[s^2 R \left(X'X\right)^{-1} R'\right]^{-1} \left(Rb-q\right) \\ &= \left(Rb-q\right)' \left[s^2 R \left(X'X\right)^{-1} R'\right]^{-1} \left(Rb-q\right) \end{aligned}$$

そこで仮定

$$p \lim \frac{X'X}{n} = Q$$

から

$$p \lim s^2 R \left(X'X\right)^{-1} R' = \left(\sigma^2/n\right) R Q^{-1} R'$$

なので W が漸近的に

$$\sqrt{n}\left(Rb-q\right)' \left[R\left(\sigma^2 Q^{-1}\right) R'\right]^{-1} \sqrt{n}\left(Rb-q\right)$$

に分布収束する。さらに

$$\sqrt{n}\left(Rb-q\right) = \sqrt{n} R \left(b-\beta\right) \xrightarrow{d} N\left(0, R\left(\sigma^2 Q^{-1}\right) R'\right)$$

なので、

$$\left[R\left(\sigma^2 Q^{-1}\right) R'\right]^{-1/2} \left[\sqrt{n}\left(Rb-q\right)\right] \xrightarrow{d} N\left(0, I\right)$$

従って W が漸近的互いに独立な正規分布の二乗和となり、自由度 J の χ^2 分布に分布収束する。