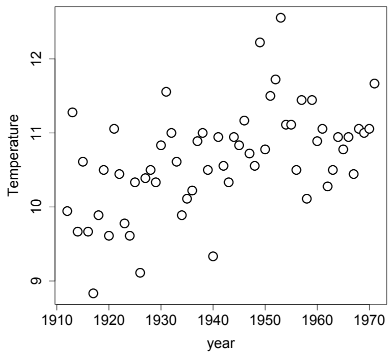
数理情報学９ (数理統計学)　4.　回帰分析

解説

ペアになっているデータ $(x_1, Y_1), \ldots, (x_n, Y_n)$ が得られているとして，x と Y とのあいだの関係を知りたいという状況を考える．このときによく使われる手法が線形回帰モデルによる推定方法である．ここでは $x_1, \ldots, x_n$ は定数であり $Y_1, \ldots, Y_n$ は確率変数という状況を扱う．

例 **1.** ある地域の各年の平均気温データ．x は年 (西暦)，Y はその年の平均気温．x は定数であり，Y を確率変数と考える．このデータから，温暖化していると言えるか？

The mean annual temperature in degrees Celsius in New Haven, Connecticut, from 1912 to 1971.
McNeil, D. R. (1977) Interactive Data Analysis. New York: Wiley.



1　線形回帰モデル

観測データ x_i と Y_i とのあいだに

$$\underline{\text{線形回帰モデル}}:\quad Y_i = \theta_0 + \theta_1 x_i + \varepsilon, \quad E[\varepsilon_i] = 0$$
$$\varepsilon\ :\ Y_i\text{の観測に対するノイズ}$$

という関係を仮定する．x_i と Y_i との間の関数関係がパラメータ θ_0, θ_1 に関する線形式よって定められている (図 1)．このとき x_i, Y_i をそれぞれ

$$x_i:\ \text{独立変数},\quad Y_i:\ \text{従属変数}$$

とよぶ．ここで独立変数 x_i は定数 (既知) とみなしているので小文字で表記し，従属変数 Y_i は確率変数と考えて大文字で書いている．独立変数も確率変数になるような状況は，演習では扱わない．

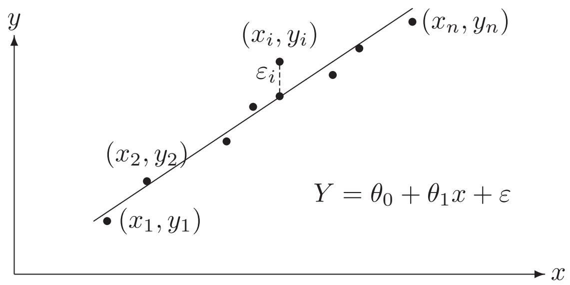


図 1: 線形回帰モデルによる推定

線形回帰モデルは $\theta_0 + \theta_1 x$ がパラメータ θ_0, θ_1 に関して線形式になっている．これが線形回帰モデルと呼ばれる所以である．独立変数 x に関する 2 次式も

$$Y_i = \theta_0 + \theta_1 x_i + \theta_2 x_i^2 + \varepsilon$$

とおくと $\theta_0 + \theta_1 x_i + \theta_2 x_i^2$ がパラメータ $(\theta_0, \theta_1, \theta_2)$ について線形式なので，線形回帰モデルとして扱うことができる．

2　最小二乗推定量

観測したデータ $(x_1, Y_1), \ldots, (x_n, Y_n)$ に対して線形回帰モデル

$$Y_i = \theta_0 + \theta_1 x_i + \varepsilon_i, \quad i = 1 \ldots, n$$

を仮定する．このとき θ_0, θ_1 をどのように推定したらよいだろうか？　ここでは二乗誤差を最小にするようにパラメータ θ_0, θ_1 を決める方法について解説する．

最小二乗推定量 $(\widehat{\theta}_0, \widehat{\theta}_1)$ は次のように定義される：

$$\sum_{i=1}^{n} \underbrace{(Y_i - \theta_0 - \theta_1 x_i)^2}_{\text{誤差の自乗}} \quad \text{を最小にするパラメータ } (\widehat{\theta}_0, \widehat{\theta}_1) \tag{1}$$

式 (1) の解を求める．二乗誤差を最小にするパラメータを求めるために極値条件を考える．

$$\begin{cases} \frac{\partial}{\partial \theta_0} \sum_{i=1}^{n} (Y_i - \theta_0 - \theta_1 x_i)^2 = 0 \\ \frac{\partial}{\partial \theta_1} \sum_{i=1}^{n} (Y_i - \theta_0 - \theta_1 x_i)^2 = 0 \end{cases} \iff \begin{cases} \sum_{i=1}^{n} (Y_i - \theta_0 - \theta_1 x_i) = 0 \\ \sum_{i=1}^{n} (Y_i - \theta_0 - \theta_1 x_i) x_i = 0 \end{cases} \tag{2}$$

式 (2) を行列とベクトルを使って表現すると，最小二乗推定量を計算することができる．n 個のデータ $\{(x_1, Y_1), \ldots, (x_n, Y_n)\}$ に対して n 次元ベクトル Y, $n \times 2$ 行列 X, 2 次元ベクトル θ をそれぞれ

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \theta = \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix}$$

とおく．このとき式 (2) をまとめて

$$X^\top (Y - X\theta) = 0$$

と書くことができる ($X^\top$ は X の転置行列)．したがって最小二乗推定量は

$$\widehat{\theta} = \begin{pmatrix} \widehat{\theta}_0 \\ \widehat{\theta}_1 \end{pmatrix} = (X^\top X)^{-1} X^\top Y = \begin{pmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} x_i Y_i \end{pmatrix}$$

となる．ここでは行列 $X^\top X$ が正則であることを仮定している[1]．独立変数が多次元ベクトル

$$x_i = (x_{i1}, \ldots, x_{id})$$

であっても，行列を使って式を書けば同様に推定量 $\widehat{\theta} = (\widehat{\theta}_0, \widehat{\theta}_1, \ldots, \widehat{\theta}_d)^\top$ を行列とベクトルで表現できる．

多次元の最小二乗法の解法

データ $\{(x_1, Y_1), \ldots, (x_n, Y_n)\}$ が得られているとする．ここで $x_i = (x_{i1}, \ldots, x_{id}) \in \boldsymbol{R}^d$ とする．このときパラメータを $\theta = (\theta_1, \ldots, \theta_d)$ として (多次元) 線形回帰モデル

$$Y_i = x_{i1}\theta_1 + x_{i2}\theta_2 + \cdots + x_{id}\theta_d + \varepsilon_i, \quad i = 1 \ldots, n$$

を仮定する．これをベクトルと行列を用いて表現すると

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} x_{11} & \cdots & x_{1d} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{nd} \end{pmatrix}, \quad \theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_d \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

[1] 正則でないときには一般化逆行列が必要になるが，講義では扱わない．

として

$$Y = X\theta + \varepsilon$$

と書ける．

最小二乗推定量は二乗誤差

$$\sum_{i=1}^{n}\left(Y_i - \sum_{k=1}^{d} x_{ik}\theta_k\right)^2$$

を最小にするパラメータは，極値条件から求めることができる．二乗誤差をパラメータ θ_t で微分して極値条件を計算すると

$$\begin{aligned}
&\sum_{i=1}^{n}\left(Y_i - \sum_{k=1}^{d} x_{ik}\widehat{\theta}_k\right)x_{it} = 0, \quad t = 1, \ldots, d \\
\iff \quad &\sum_{k=1}^{d}\Bigg(\underbrace{\sum_{i=1}^{n} x_{it}x_{ik}}_{(X^\top X)_{tk}}\Bigg)\widehat{\theta}_k = \underbrace{\sum_{i=1}^{n} y_i x_{it}}_{(X^\top Y)_t}, \quad t = 1, \ldots, d
\end{aligned}$$

となる．この式を行列を用いて表現すると

$$X^\top X\widehat{\theta} = X^\top Y$$

となる．したがって最小二乗法の解は

$$\widehat{\theta} = \begin{pmatrix} \widehat{\theta}_1 \\ \vdots \\ \widehat{\theta}_d \end{pmatrix} = (X^\top X)^{-1}X^\top Y$$

となる．$\widehat{\theta} = (X^\top X)^{-1}X^\top Y$ は θ の不偏推定量であり，分散共分散行列は $\sigma^2(X^\top X)^{-1}$ である．これは以下のように確かめられる：

$$\begin{aligned}
E[\widehat{\theta}] &= (X^\top X)^{-1}X^\top E[Y] = (X^\top X)^{-1}X^\top(X\theta + E[\vartheta]) = (X^\top X)^{-1}X^\top X\theta = \theta, \\
V[\widehat{\theta}] &= (X^\top X)^{-1}X^\top V[Y]X((X^\top X)^{-1})^\top = (X^\top X)^{-1}X^\top(\sigma^2 I)X(X^\top X)^{-1} = \sigma^2(X^\top X)^{-1}X^\top X(X^\top X)^{-1} \\
&= \sigma^2(X^\top X)^{-1}.
\end{aligned}$$

上の公式を使えば，4 節の $\widehat{\theta}_1$ の分散は $\sigma^2(X^\top X)^{-1}$ の $(2,2)$-成分であることが分かる．

3 最小二乗推定量の性質

線形回帰モデル

$$Y_i = \theta_0 + \theta_1 x_i + \varepsilon_i, \quad E[\varepsilon_i] = 0, \quad V[\varepsilon_i] = \sigma^2, \quad i = 1, \ldots, n$$

に対する最小二乗推定量

$$\widehat{\theta} = \begin{pmatrix} \widehat{\theta}_0 \\ \widehat{\theta}_1 \end{pmatrix} = (X^\top X)^{-1}X^\top Y$$

はパラメータ $\theta = (\theta_0, \theta_1)^\top$ の不偏推定量である．これを以下に示す．誤差項をまとめて $Z = (\varepsilon_1, \ldots, \varepsilon_n)^\top$ とする．すると

$$Y = X\theta + Z$$

となる．したがって最小２乗推定量は

$$\widehat{\theta} = \begin{pmatrix} \widehat{\theta}_0 \\ \widehat{\theta}_1 \end{pmatrix} = (X^\top X)^{-1}X^\top Y = (X^\top X)^{-1}X^\top(X\theta + Z) = \theta + (X^\top X)^{-1}X^\top Z$$

となる．以上から

$$E[Z]=0 \text{ より } E[\widehat{\theta}]=\theta$$

を得る．

次に推定量の分散と共分散を求める．多変数の確率変数 $Z=(Z_1,\ldots,Z_k)^\top$ に対して，Z の分散共分散行列を

$$V[Z]=E[(Z-E[Z])(Z-E[Z])^\top], \qquad (k\times k \text{ 行列})$$

と定義する．すると $V[Z]$ の各要素は

$$\begin{aligned}(V[Z])_{ii} &= Z_i \text{ の分散},\\ (V[Z])_{ij} &= Z_i \text{ と } Z_j \text{ の共分散 } \mathrm{Cov}(Z_i,Z_j) \quad (i\neq j)\end{aligned}$$

となる．行列 C と ベクトル $\boldsymbol{b}$ に対して $CZ+\boldsymbol{b}$ の分散共分散行列は $V[CZ+\boldsymbol{b}]=CV[Z]C^\top$ となる：

$$\because \quad V[CZ]=E[(CZ+\boldsymbol{b}-E[CZ+\boldsymbol{b}])(CZ+\boldsymbol{b}-E[CZ+\boldsymbol{b}])^\top]=CE[(Z-E[Z])(Z-E[Z])^\top]\top C=CV[Z]C^\top.$$

以上をふまえると，最小２乗推定量 $\widehat{\theta}$ の分散共分散行列は

$$\begin{aligned}V[\widehat{\theta}] &= V[(X^\top X)^{-1}X^\top Y] \;=\; (X^\top X)^{-1}X^\top V[Y]X(X^\top X)^{-1} \;=\; (X^\top X)^{-1}X^\top(\sigma^2 I)X(X^\top X)^{-1}\\ &= \sigma^2(X^\top X)^{-1}\end{aligned}$$

となる．上の計算で $V[Y]=V[X\theta+\varepsilon]=V[\varepsilon]=\sigma^2 I$ を用いた．

まとめると以下のようになる．

最小２乗推定量の統計的性質

Y の 分布：$Y=\theta_0+\theta_1 x+\varepsilon$，$\varepsilon_1,\ldots,\varepsilon_n$ は独立に同一の分布にしたがい，分散は σ^2．
推定量：$\widehat{\theta}=(X^\top X)^{-1}X^\top Y$．

不偏性： $E[\widehat{\theta}]=\theta$．

分散共分散行列： $V[\widehat{\theta}] = \sigma^2(X^\top X)^{-1}$． したがって $\widehat{\theta}_0$ の分散は $\sigma^2((X^\top X)^{-1})_{11}$，$\widehat{\theta}_1$ の分散は $\sigma^2((X^\top X)^{-1})_{22}$，$\widehat{\theta}_0$ と $\widehat{\theta}_1$ の共分散は $\sigma^2((X^\top X)^{-1})_{12}$ (または 21 成分)．

4　線形回帰モデルに対する仮説検定

実際のデータ解析では，線形回帰モデルのパラメータについて仮説検定を行うことが多い．まず以下の線形モデルを考えよう．

$$Y_i=\theta_0+\theta_1 x_i+\varepsilon_i, \quad i=1,\ldots,n.$$

ここで $\varepsilon_i, i=1,\ldots,n$ は独立に正規分布 $N(0,\sigma^2)$ にしたがうとする．ここで分散 σ^2 は既知とする．観測された n 個のデータ $(x_1,y_1),\ldots,(x_n,y_n)$ から傾き θ_1 に関する検定を行う．

このような検定をおこなう動機は，「変数 x が変化したとき，それに伴って y も変化するかどうか」を検証することである．このような問題は応用上重要である．たとえば，薬の投与量とその効果の関連を調べるのに用いられる．

検定を構成するために，推定量の分布を計算する．行列 X を

$$X=\begin{pmatrix}1 & x_1\\ 1 & x_2\\ \vdots & \vdots\\ 1 & x_n\end{pmatrix}$$

として，$(X^\top X)^{-1}$ が存在すると仮定する．傾き θ_1 に関する検定

$$\text{帰無仮説 } H_0 : \theta_1 = 0 \qquad \text{対立仮説 } H_1 : \theta_1 \neq 0$$

を考える．帰無仮説のもとで，まず最小二乗法で得られる $\hat{\theta}_1$ の分布を計算する．$\widehat{\theta} = (X^\top X)^{-1} X^\top Y$, $Y = (Y_1, \ldots, Y_n)^\top$ より

$$\widehat{\theta}_1 \;=\; \frac{1}{n\sum_i x_i^2 - (\sum_i x_i)^2}\left(-\sum_i x_i \sum_j Y_j + n\sum_j x_j Y_j\right) \;=\; \frac{\sum_j (nx_j - \sum_i x_i) Y_j}{n\sum_i x_i^2 - (\sum_i x_i)^2}$$

となる．Y_j は独立に $N(\theta_0 + \theta_1 x_i, \sigma^2)$ の正規分布にしたがうのでその線形和である $\widehat{\theta}_1$ も正規分布にしたがう．さらに $\widehat{\theta}_1$ は θ_1 の不偏推定量なので $\theta_1 = 0$ のときには $E[\widehat{\theta}_1] = 0$ となる．つぎに $\widehat{\theta}_1$ の分散を求める．

$$V[\widehat{\theta}_1] \;=\; \frac{1}{(n\sum_i x_i^2 - (\sum_i x_i)^2)^2}\sum_j (nx_j - \sum_i x_i)^2 V[Y_j] \;=\; \frac{\sigma^2 \sum_j (nx_j - \sum_i x_i)^2}{(n\sum_i x_i^2 - (\sum_i x_i)^2)^2} \;=\; \frac{\sigma^2}{\sum_i x_i^2 - \frac{1}{n}(\sum_i x_i)^2}$$

となるので，帰無仮説のもとでは

$$\widehat{\theta}_1 \sim N\left(0, \frac{\sigma^2}{\sum_i x_i^2 - \frac{1}{n}(\sum_i x_i)^2}\right)$$

となることが分かる．標準正規分布の両側 α 点を z_α^* とすると，有意水準が α となる棄却域は

$$W = \left\{\widehat{\theta}_1 \;\middle|\; |\widehat{\theta}_1| \;>\; z_\alpha^* \sqrt{\frac{\sigma^2}{\sum_i x_i^2 - \frac{1}{n}(\sum_i x_i)^2}}\right\}$$

となる．

5 予測値の信頼区間

データ $(x_1, Y_1), \ldots, (x_n, Y_n)$ が与えられているとき，独立変数が x のときの Y の期待値 $E[Y] = \theta_0 + \theta_1 x$ の値を予測し，その信頼区間を作ることを考える．ここではノイズに正規分布の仮定を加える：

$$Y(x) = \theta_0 + \theta_1 x + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2).$$

ここで σ^2 は既知とする．

まず $Y(x)$ の期待値 $E[Y(x)] = \theta_0 + \theta_1 x$ を推定することを考える．データから θ を最小二乗法で推定した結果を $\widehat{\theta} = (\widehat{\theta}_0, \widehat{\theta}_1)^\top$ とすると $\theta_0 + \theta_1 x$ を

$$\widehat{Y}(x) \;=\; \widehat{\theta}_0 + \widehat{\theta}_1 x$$

で推定することができる．$\widehat{\theta}$ が θ の不偏推定量なので，$\widehat{Y}(x)$ も $\theta_0 + \theta_1 x$ の不偏推定量になることが分かる：

$$E[\widehat{Y}(x)] = E[\widehat{\theta}_0 + \widehat{\theta}_1 x] = E[\widehat{\theta}_0] + E[\widehat{\theta}_1]x = \theta_0 + \theta_1 x.$$

次に信頼区間を構成する．

$$\widehat{Y}(x) \;=\; \widehat{\theta}_0 + \widehat{\theta}_1 x \;=\; (1\ x)\begin{pmatrix}\widehat{\theta}_0 \\ \widehat{\theta}_1\end{pmatrix} \;=\; (1\ x)(X^\top X)^{-1} X^\top Y$$

となる．ここで $(1\ x)(X^\top X)^{-1}X^\top$ は $1 \times n$ 行列 (横ベクトル) であり $Y_1, \ldots, Y_n$ は独立に正規分布 $N(\theta_0 + \theta_1 x_1, \sigma^2)$ にしがたうので，$\widehat{Y}(x)$ は独立な正規分布の和として書ける．したがって $\widehat{Y}(x)$ も正規分布にしたがう．$\widehat{Y}(x)$ の分散を計算すると

$$\begin{aligned} V[\widehat{Y}(x)] &= V[(1\ x)(X^\top X)^{-1}X^\top Y] = (1\ x)(X^\top X)^{-1}X^\top V[Y] X (X^\top X)^{-1}\begin{pmatrix}1 \\ x\end{pmatrix} \\ &= \sigma^2 (1\ x)(X^\top X)^{-1}\begin{pmatrix}1 \\ x\end{pmatrix} \end{aligned}$$

となる．この分散を $v(x)$ とおくと

$$\widehat{Y}(x) \ \sim \ N(E[Y(x)],\ v(x))$$

となる．したがって $E[Y(x)] = \theta_0 + \theta_1 x$ の $100(1-\alpha)\%$ 信頼区間は

$$\frac{\widehat{Y}(x) - E[Y(x)]}{\sqrt{v(x)}} \ \sim \ N(0,1) \quad \Longrightarrow \quad P\left(\left|\frac{\widehat{Y}(x) - E[Y(x)]}{\sqrt{v(x)}}\right| \le z_\alpha^*\right) = 1-\alpha$$

より

$$\left[\ \widehat{Y}(x) - z_\alpha^* \sqrt{v(x)},\ \widehat{Y}(x) + z_\alpha^* \sqrt{v(x)}\ \right]$$

となる．

6 線形回帰の幾何学的解釈

行列 X，ベクトル Y, θ, ε を

$$X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad \theta = \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

とすると，線形回帰モデルは

$$Y = X\theta + \varepsilon$$

と簡潔に表現できる．さらに列ベクトル a_0, a_1 を

$$a_0 = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \quad a_1 = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

と定義すると

$$X = (a_0\ a_1)$$

と書ける．したがって

$$X\theta = \theta_0 a_0 + \theta_1 a_1$$

となって，これは a_0 と a_1 の線形和で表現される n 次元ベクトルである．パラメータ θ の最小二乗推定量は

$$\text{二乗誤差} = \|Y - \theta_0 a_0 - \theta_1 a_1\|^2$$

を最小にするように $\theta = (\theta_0, \theta_1)$ を決めるという推定方法である．つまり $\{a_0, a_1\}$ で張られる $\boldsymbol{R}^n$ の部分空間へ Y を射影することでパラメータ $\widehat{\theta}$ を求めることができる．

最小二乗推定量 $\widehat{\theta}$ は

$$\widehat{\theta} = (X^\top X)^{-1} X^\top Y$$

となる．したがって

$$X\widehat{\theta} = X(X^\top X)^{-1} X^\top Y$$

が成り立つ．以上の考察から

$$\{a_0, a_1\}\text{ で張られる部分空間への }Y\text{ の射影} \quad : \quad Y \longrightarrow X(X^\top X)^{-1} X^\top Y$$

という関係になっている．つまり Y を $\{a_0, a_1\}$ で張られる部分空間へ射影したベクトルを得るには，行列 $X(X^\top X)^{-1}X$ を Y に掛ければよい．射影を表す行列 Π を

$$\Pi = X(X^\top X)^{-1}X^\top$$

とすると

$$X\widehat{\theta} = \Pi Y$$

となる (図 2).

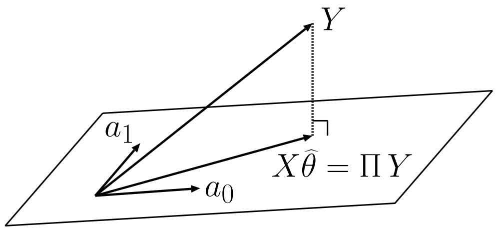


図 2: a_0 と a_1 で張られる部分空間に Y を射影すると $X\widehat{\theta}$ が得られる．

一般に $X = (a_1 \ \cdots \ a_d)$ のように $n \times d$ 行列になっているときも，同じように計算できる．このときの最小二乗推定量を $\widehat{\theta}$ とする．$\Pi = X(X^\top X)^{-1}X^\top$ によって $\{a_1, \ldots, a_d\}$ で張られる部分空間にベクトル Y を射影すると

$$X\widehat{\theta} = \Pi Y$$

が成り立つ．

7 誤差項の分散の推定

誤差項 ε_i の分散を $V[\varepsilon_i] = \sigma^2 \ (i = 1, \ldots, n)$ とする．データ $(x_1, Y_1), \ldots, (x_n, Y_n)$ から分散 σ^2 を推定する．ここで，分散はデータ点 x_i に依存せずに一定の値 σ^2 をとると仮定している．線形モデル $Y_i = \theta_0 + \theta_1 x_i + \varepsilon_i$ より

$$\sigma^2 \ = \ \frac{1}{n}\sum_{i=1}^{n} V[\varepsilon_i] \ = \ \frac{1}{n}\sum_{i=1}^{n} E[\varepsilon_i^2], \quad \varepsilon_i \ = \ Y_i - (\theta_0 + \theta_1 x_i)$$

となるので

$$\sigma^2 \ = \ \frac{1}{n}\sum_{i=1}^{n} E[(Y_i - (\theta_0 + \theta_1 x_i))^2]$$

と書ける．したがって，σ^2 の推定量として，θ を 推定量 $\widehat{\theta}$ で置き換えて

$$\widehat{\sigma}^2 \ = \ \frac{1}{n}\sum_{i=1}^{n}(Y_i - (\widehat{\theta}_0 + \widehat{\theta}_1 x_i))^2$$

とすれば良さそうである．実際，上の $\widehat{\sigma}^2$ は ε_i が独立に同一の正規分布にしたがうと仮定したときの最尤推定量になっている．

次に，$\widehat{\sigma}^2$ を修正して σ^2 の不偏推定量 S^2 を導出する．

$$\sum_{i=1}^{n}(Y_i - (\widehat{\theta}_0 + \widehat{\theta}_1 x_i))^2 \ = \ (Y - X\widehat{\theta})^\top (Y - X\widehat{\theta}) \ = \ Z^\top (I - \Pi) Z$$

となるので，

$$E\left[\sum_{i=1}^{n}(Y_i-(\widehat{\theta}_0+\widehat{\theta}_1 x_i))^2\right]=E[\mathtt{Tr}(I-\Pi)ZZ^\top]=\sigma^2(n-\mathtt{Tr}(X^\top X)^{-1}X^\top X)=\sigma^2(n-2).$$

したがって

$$S^2 \;=\; \frac{1}{n-2}\sum_{i=1}^{n}(Y_i-(\widehat{\theta}_0+\widehat{\theta}_1 x_i))^2$$

は σ^2 の不偏推定量である．

演習問題

1. 線形回帰モデル

$$Y_i=\theta_0+\theta_1 x_i+\varepsilon_i,\quad i=1,\ldots,n$$

を考える．ここで θ_0,θ_1 はそれぞれ1次元パラメータであり，$x_i\in\mathbb{R}$ とする．データ $(x_1,Y_1),\ldots,(x_n,Y_n)$ が得られたとき，パラメータ θ_0,θ_1 の最小二乗推定量を $\widehat{\theta}_0,\widehat{\theta}_1$ とする．このとき回帰直線

$$y=\widehat{\theta}_0+\widehat{\theta}_1\, x$$

はデータの標本平均点 $\left(\frac{1}{n}\sum_{i=1}^{n}x_i,\ \frac{1}{n}\sum_{i=1}^{n}Y_i\right)$ を通ることを示せ．

2. n 個のデータ $Y_1,\ldots,Y_n$ が独立に得られているとする．各データ Y_i がしたがう分布は

$$Y_i=\theta_0+\theta_1 x_i+\varepsilon_i,\ \ \varepsilon_i\sim N(0,\sigma^2),\ \ i=1,\ldots,n$$

のような線形回帰モデルで記述できるとする．独立変数 $x_i\in\mathbb{R}, i=1,\ldots,n$ の値はすべて既知の定数とし，分散 σ^2 は既知とする．行列 X を

$$X=\begin{pmatrix}1 & x_1\\ 1 & x_2\\ \vdots & \vdots\\ 1 & x_n\end{pmatrix}$$

としたとき $(X^\top X)^{-1}$ が存在すると仮定する．パラメータ θ_0,θ_1 に関する検定

$$\text{帰無仮説 } H_0:(\theta_0,\theta_1)=(1,1),\qquad \text{対立仮説 } H_1:(\theta_0,\theta_1)=(0,2)$$

をおこなう．ネイマン・ピアソンの補題を用いて，有意水準 α の検定の棄却域を構成せよ．ただし標準正規分布の片側パーセント点を z_α とする．

3. データ $(x_1,Y_1),\ldots,(x_n,Y_n)$ は，線形回帰モデル

$$Y_i=\theta x_i+\varepsilon_i,\ \ \varepsilon_i\sim N(0,\sigma^2),\ \ i=1,\ldots,n$$

にしたがって生成されたとする．$x_1,\ldots,x_n\in\mathbb{R}$ は定数であり，少なくとも 1 つは 0 ではないと仮定する．また $\varepsilon_i, i=1,\ldots,n$ は独立とする．データから傾き θ と分散 σ^2 を推定する．事前の知識から，傾き θ は $0\le\theta$ の範囲にあることが分かっているとする．

   (a) θ と σ^2 の最尤推定量 $\widehat{\theta},\widehat{\sigma}^2$ をそれぞれ計算せよ．推定量が (対数) 尤度を最大化していることも証明すること．

   (b) 標準正規分布の分布関数を Φ とする．$\widehat{\theta}$ の分布関数を Φ を用いて記述せよ．

(a) 最小二乗法によって θ_0, θ_1 を推定する．このとき θ_0 の推定量 $\widehat{\theta}_0$ の分布は $N(\theta_0, \sigma^2 h_{11})$ となることを示せ．

注：正規分布にしたがう確率変数の線形和も正規分布にしたがうことは，証明なしに用いてよい．

(b) パラメータに関する検定

$$\text{帰無仮説 } H_0 : (\theta_0, \theta_1) = (0, 0), \qquad \text{対立仮説 } H_1 : (\theta_0, \theta_1) = (0, 1)$$

をおこなう．ネイマン・ピアソンの補題を用いて，有意水準 α の検定の棄却域を構成せよ．ここで標準正規分布の片側パーセント点を z_α とする．

4. n 次元確率変数 Y は，期待値が $0 \in \boldsymbol{R}^n$ で分散共分散行列が I_n ($n \times n$ 単位行列) であるような n 次元正規分布 $N_n(0, I_n)$ にしたがうとする．Y は列ベクトルとして記述する．また Π は $n \times n$ 行列であり

$$\Pi^2 = \Pi, \quad \Pi^\top = \Pi$$

を満たすとする (上の性質を満たす行列を射影行列という)．このとき，２つの確率変数，

$$Y^\top \Pi Y, \quad Y^\top (I_n - \Pi) Y$$

は独立であることを示せ．

ヒント： Π は対称行列なので，直交行列 Q と対角行列 Λ が存在して $\Pi = Q^\top \Lambda Q$ のように書ける．さらに $Z = QY$ とおくと, $Z \sim N(0, I_n)$ となる．

(a) Π のランクは d とする．$Y^\top \Pi Y$ は自由度 d のカイ二乗分布にしたがうことを示せ．

(b) $Y^\top \Pi Y$ と $Y^\top (I_n - \Pi) Y$ は確率変数として独立であることを示せ．

注：カイ二乗分布については，テキストなどを参照のこと．

5. データ $Y_i, i = 1, \ldots, n$ は独立に指数分布 $\mathrm{Ex}(\lambda_i)$ にしたがうとする．ここでパラメータ $\lambda_1, \ldots, \lambda_n$ は

$$\lambda_i = x_i \theta, \quad i = 1, \ldots, n$$

と書けるとする．$x_1, \ldots, x_n$ は全て正の定数で既知とする．θ の最尤推定量 $\widehat{\theta}$ を計算せよ．

注：尤度方程式を解いて $\widehat{\theta}$ を求めればよい．最尤推定量が (対数) 尤度を最大化していることまで証明する必要はない．

6. X を $n \times d$ の行列とする．ここで $n \geq d$ として X のランク (階数) は d に等しいとする．行列 Π を

$$\Pi = X(X^\top X)^{-1} X^\top$$

と定義する．次の問に答えよ．

(a) Π の固有値は 0 または 1 であることを示せ．

(b) $X = (a_1, \ldots, a_d)$ と書く．ここで $a_i, i = 1, \ldots, d$ は n 次元列ベクトルである．$\{a_i, i = 1, \ldots, d\}$ が張る部分空間に列ベクトル b が含まれるとき

$$\Pi b = b$$

が成り立つことを示せ．

(c) 独立変数が d 次元の線形回帰モデル

$$Y_i = \alpha + \beta^\top x_i + \varepsilon_i$$

を考える．ここで α は 1 次元パラメータ，β は d 次元パラメータ (列ベクトル)，また $x_i \in \mathbb{R}^d$ (列ベクトル) とする．データ $(x_1, Y_1), \ldots, (x_n, Y_n)$ が得られたとき，パラメータ α, β の最小二乗推定量をそれぞれ $\widehat{\alpha}, \widehat{\beta}$ とする (最小二乗推定量は一意に定まるとする)．このとき回帰直線

$$y = \widehat{\alpha} + \widehat{\beta}^\top x$$

はデータの標本平均点 $\left(\dfrac{1}{n}\displaystyle\sum_{i=1}^n x_i,\ \dfrac{1}{n}\sum_{i=1}^n Y_i\right)$ を通ることを示せ．

7. n 次元確率変数 Y は，期待値が $0 \in \boldsymbol{R}^n$ で分散共分散行列が I_n ($n \times n$ 単位行列) であるような正規分布 $N_n(0, I_n)$ にしたがうとして，Y を列ベクトルで記述する．また Π を 問題 7-4 で定義される行列とする．

   (a) $Y^\top \Pi Y$ は自由度 d のカイ二乗分布にしたがうことを示せ．

   (b) $Y^\top \Pi Y$ と $Y^\top (I_n - \Pi) Y$ は確率変数として独立であることを示せ．

   注：カイ二乗分布については適当なテキストを参照のこと．

8. n 個のデータ $Y_1, \ldots, Y_n$ が独立に得られているとする．各データ Y_i がしたがう分布は

$$Y_i = \theta_0 + \theta_1 x_i + \varepsilon_i$$

   のような線形回帰モデルで記述できるとする．ここで ε_i は正規分布 $N(0, \sigma^2)$ にしたがう確率変数であり，分散 σ^2 は既知とする．また独立変数 $x_i, i = 1, \ldots, n$ の値はすべて既知である．行列 X を

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

   としたとき $(X'X)^{-1}$ が存在すると仮定して

$$(X'X)^{-1} = H = \begin{pmatrix} h_{11} & h_{12} \\ h_{12} & h_{22} \end{pmatrix}$$

   とおく．

   (a) 分散 σ^2 は既知とする．傾き θ_1 に関する検定

$$\text{帰無仮説 } H_0 : \theta_1 = 0 \qquad \text{対立仮説 } H_1 : \theta_1 > 0$$

   をおこなう．有意水準が α となるような一様最強力検定を構成せよ．

   (b) 2 つのデータ $(x_1, Y_1), (x_2, Y_2)$ を得たとき，前問で求めた傾きに関する一様最強力検定をおこなう．このとき第 II 種の誤り確率は一般に x_1, x_2 の選び方に依存する．独立変数は $[-1, 1]$ の範囲で選ぶことができるとき，第 II 種の誤り確率を一様に最小にするような $x_1, x_2 \in [-1, 1]$ を求めよ．

   Note : 実際の実験では，実験者がいくつかのパラメータ (独立変数) を設定する必要がある場合が多い．独立変数をうまく選ぶことで統計的推論の精度を上げることができる．