

線型推測論

第07回 確率分布の仮定と線形仮説の検定(2)

2022/5/20

慶應義塾大学病院

長島 健悟

メモ

- $\text{tr}\{\mathbf{Y}'\mathbf{A}\mathbf{Y}\} = \text{tr}\{\mathbf{A}\mathbf{Y}\mathbf{Y}'\}$ (定理1-6)
- 二次形式の分布： $\left(\frac{\mathbf{Y}-\boldsymbol{\mu}}{\sigma}\right)' \mathbf{A} \left(\frac{\mathbf{Y}-\boldsymbol{\mu}}{\sigma}\right) \sim \chi^2(r)$
- 冪等行列 $\mathbf{A} = \mathbf{A}\mathbf{A}$ ： $\text{rank}(\mathbf{A}) = \text{trace}(\mathbf{A})$

複数仮説の検定

- 単一仮説の検定： $\beta_1 = 0, \beta_1 - \beta_2 = 0$
 - t 統計量のような検定統計量を用いる
- 複数仮説の検定： $\beta_1 = \beta_2 = 0, \boldsymbol{\beta} = \mathbf{0}$
 - t 統計量では原理的に無理
 - β_1 と β_2 が同時に0になるときだけ0になる統計量を構成したい！

複数仮説の検定

- 和 $\beta_1 + \beta_2$ を評価する統計量はどうか？
 - ダメ： $\beta_1 = -\beta_2$ のときにも $\beta_1 + \beta_2 = 0$ になってしまう
- 候補は色々あるが...
 - 平方和： $\beta_1^2 + \beta_2^2$
 - 絶対値の和： $|\beta_1| + |\beta_2|$
 - etc...
- 平方和はかなり扱いやすい

平方和に基づく複数仮説の検定

- $\beta_1 = \beta_2 = 0$ だけでなく、一般化すると

$$H_0: \mathbf{C}\boldsymbol{\beta} = \begin{pmatrix} \mathbf{c}'_1 \boldsymbol{\beta} \\ \vdots \\ \mathbf{c}'_m \boldsymbol{\beta} \end{pmatrix} = \mathbf{0}$$

$$\Leftrightarrow H_0: (\mathbf{c}'_1 \boldsymbol{\beta})' \mathbf{c}'_1 \boldsymbol{\beta} + \cdots + (\mathbf{c}'_m \boldsymbol{\beta})' \mathbf{c}'_m \boldsymbol{\beta} = 0$$

$$\Leftrightarrow H_0: \left(\sum_{j=1}^p c_{1j} \beta_j \right)^2 + \cdots + \left(\sum_{j=1}^p c_{mj} \beta_j \right)^2 = 0$$

- 平方和を使うと複数仮説の検定をこのように考えることができる

平方和の統計量

- 平方和の統計量はどうしたらよい？
- 二次形式の分布は知っている (Aは冪等)

$$\left(\frac{\mathbf{Y} - \boldsymbol{\mu}}{\sigma}\right)' \mathbf{A} \left(\frac{\mathbf{Y} - \boldsymbol{\mu}}{\sigma}\right) \sim \chi^2(r)$$

- 二次形式は平方和の統計量を含む

$$\sum_i \sum_j a_{ij} \left(\frac{y_i - \mu}{\sigma}\right) \left(\frac{y_j - \mu}{\sigma}\right)$$

- 二次形式の分布をうまく使って複数仮説の検定統計量を構成してみよう

F統計量

- U_1, U_2 が独立に自由度 d_1, d_2 のカイ二乗分布に従うとき

$$\frac{U_1/d_1}{U_2/d_2} \sim F(d_1, d_2)$$

- カイ二乗分布に従う確率変数を自由度で割った統計量の比をF統計量と呼ぶのであった

一つの平方和の仮説検定

- 一番簡単な仮説の平方和版を考える

- $m = 1, \mathbf{c}'_1 = (c_{11}, c_{12}, \dots, c_{1p})$

- $H_0: (\mathbf{c}'_1 \boldsymbol{\beta})' \mathbf{c}'_1 \boldsymbol{\beta} = \left(\sum_{j=1}^p c_{1j} \beta_j \right)^2 = 0$

- これは当然

$$\Leftrightarrow H_0: \mathbf{c}'_1 \boldsymbol{\beta} = 0$$

F統計量：平方和に基づく統計量

- 前回導いた t 検定統計量を, 二乗和を含む様に拡張してみよう

- $t = \frac{Z}{\sqrt{V/(n-p)}}, Z \sim N(0,1), V \sim \chi^2(n-p)$
であった

- 二乗和が含まれれば良いので...

- $t^2 = \frac{Z^2}{V/(n-p)} = \frac{Z^2/1}{V/(n-p)} \sim F(1, n-p)$

F統計量：平方和に基づく統計量

- 同様に, $\mathbf{c}'_1\boldsymbol{\beta}$ の場合は
- 分子 Z_1 の二乗を考えれば良い

- 帰無仮説のもとで：
$$Z_1 = \frac{\mathbf{c}'_1\hat{\boldsymbol{\beta}}}{\sqrt{\sigma^2\mathbf{c}'_1(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}_1}}$$

- $Z_1^2 = ?$

$$Z_1^2 = (\mathbf{c}'_1\hat{\boldsymbol{\beta}})' [\mathbf{c}'_1(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}_1]^{-1} (\mathbf{c}'_1\hat{\boldsymbol{\beta}}) / \sigma^2$$

F統計量：平方和に基づく統計量

- $(\mathbf{c}'_1 \hat{\boldsymbol{\beta}})' \mathbf{V}^{-1} (\mathbf{c}'_1 \hat{\boldsymbol{\beta}}), \mathbf{V} = \mathbf{c}'_1 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{c}_1$
 $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$
- 代入すると
- $(\mathbf{c}'_1 \hat{\boldsymbol{\beta}})' \mathbf{V}^{-1} (\mathbf{c}'_1 \hat{\boldsymbol{\beta}}) = ?$
 $\mathbf{Y}' \{ \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{c}_1 \mathbf{V}^{-1} \mathbf{c}'_1 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \} \mathbf{Y}$
- これは二次形式になっている

F統計量：平方和に基づく統計量

- $\mathbf{A} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}_1\mathbf{V}^{-1}\mathbf{c}_1'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$
 - これは冪等行列である
- $\text{rank}(\mathbf{A}) = \text{trace}(\mathbf{A}) = ?$
 - (定理1-6) $\text{tr}\{\mathbf{Y}'\mathbf{A}\mathbf{Y}\} = \text{tr}\{\mathbf{A}\mathbf{Y}\mathbf{Y}'\}$
 - $\text{tr}\{\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}_1\mathbf{V}^{-1}\mathbf{c}_1'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\} =$
 $\text{tr}\{\mathbf{V}^{-1}\mathbf{c}_1'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}_1\} =$
 $\text{tr}\{\mathbf{V}^{-1}\mathbf{V}\} = 1$
- $\mathbf{c}_1 : (p \times 1)$

F統計量：平方和に基づく統計量

- 分子の二乗がカイ二乗分布に従うことから, 結局
 - $Z_1^2 \sim \chi^2(1)$
 - $V = (n - p) \frac{s^2}{\sigma^2} \sim \chi^2(n - p)$
 - $t^2 = \frac{Z_1^2/1}{V/(n-p)} \sim F(1, n - p)$
- F統計量を使ってもt検定と同じ仮説を評価できる

複数仮説版のF統計量

- $\mathbf{C}\boldsymbol{\beta}$ の場合の一般化を考える
- $\mathbf{C} : m \times p$ である
- 仮定より $\mathbf{C}\hat{\boldsymbol{\beta}} \sim N(\mathbf{C}\boldsymbol{\beta}, \sigma^2 \mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}')$

- $\mathbf{C}\hat{\boldsymbol{\beta}} = \begin{pmatrix} \mathbf{c}'_1 \hat{\boldsymbol{\beta}} \\ \vdots \\ \mathbf{c}'_m \hat{\boldsymbol{\beta}} \end{pmatrix}$

複数仮説版のF統計量

- m 個の仮説について
- 帰無仮説のもとで

$$\mathbf{C}\hat{\boldsymbol{\beta}} \sim N(\mathbf{0}, \sigma^2 \mathbf{C}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{C}')$$

- うまく二乗和が出てくる統計量求める...

$$(\mathbf{C}\hat{\boldsymbol{\beta}})' \{ \mathbf{C}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{C} \}^{-1} \mathbf{C}\hat{\boldsymbol{\beta}} / \sigma^2$$

複数仮説版のF統計量

- $(\mathbf{C}\hat{\boldsymbol{\beta}})' \mathbf{W}^{-1} (\mathbf{C}\hat{\boldsymbol{\beta}}), \mathbf{W} = \mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'$
 $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$
- 代入すると
- $(\mathbf{C}\hat{\boldsymbol{\beta}})' \mathbf{W}^{-1} (\mathbf{C}\hat{\boldsymbol{\beta}}) = ?$
 $\mathbf{Y}' \{ \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\mathbf{W}^{-1}\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \} \mathbf{Y}$
- これも当然二次形式になっている

F統計量：平方和に基づく統計量

- $\mathbf{B} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\mathbf{W}^{-1}\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$
 - これは冪等行列である
- $\text{rank}(\mathbf{B}) = \text{trace}(\mathbf{B}) = ?$
 - (定理1-6) $\text{tr}\{\mathbf{Y}'\mathbf{A}\mathbf{Y}\} = \text{tr}\{\mathbf{A}\mathbf{Y}\mathbf{Y}'\}$
 - $\text{tr}\{\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\mathbf{W}^{-1}\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\} =$
 $\text{tr}\{\mathbf{W}^{-1}\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\} =$
 $\text{tr}\{\mathbf{W}^{-1}\mathbf{W}\} = m$
- $\mathbf{C} : (m \times p)$

F統計量：平方和に基づく統計量

- 結局

- $(\mathbf{C}\hat{\boldsymbol{\beta}})' \{ \mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}' \}^{-1} \mathbf{C}\hat{\boldsymbol{\beta}} / \sigma^2 \sim \chi^2(m)$

- $V = (n - p) \frac{s^2}{\sigma^2} \sim \chi^2(n - p)$

- $$\frac{(\mathbf{C}\hat{\boldsymbol{\beta}})' \{ \mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}' \}^{-1} \mathbf{C}\hat{\boldsymbol{\beta}} / m \sigma^2}{V / (n - p)} =$$
$$\frac{(\mathbf{C}\hat{\boldsymbol{\beta}})' \{ \mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}' \}^{-1} \mathbf{C}\hat{\boldsymbol{\beta}} / m}{s^2} \sim F(m, n - p)$$

具体的な例1

- $C\beta = 0$ の具体例を見て, どのような仮説を検定できるのか検討しよう

- $C = I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$

- $H_0: ?$

- $H_0: \beta_1 = \beta_2 = \cdots = \beta_p = 0$

具体的な例2

- $\mathbf{C} = \begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}$

- $H_0: \beta_1 - \beta_2 = 0, \beta_2 - \beta_3 = 0, \dots, \beta_{p-1} - \beta_p = 0$

- $H_0: \beta_1 = \beta_2 = \cdots = \beta_p$

おさらい

- 単一の線型仮説の t 統計量を導出した
- 複数仮説を同時に検定する F 統計量を導出した
- 二次形式の分布の応用と冪等行列が重要となる
- 一般線型モデルにおける F 検定は t 検定の自然な拡張であった