線型推測論

第07回 確率分布の仮定と線形仮説の検定(2)

2017/5/19

千葉大学大学院医学研究院
長島 健悟

メモ

- $\mathrm{tr}\{\mathbf{Y}'\mathbf{A}\mathbf{Y}\} = \mathrm{tr}\{\mathbf{A}\mathbf{Y}\mathbf{Y}'\}$ (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$\mathbf{A} = \mathbf{A}\mathbf{A}$：$\mathrm{rank}(\mathbf{A}) = \mathrm{trace}(\mathbf{A})$

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
  - 例えば・・・
  - 平方和：$\beta_1^2 + \beta_2^2$
  - 絶対値の和：$|\beta_1| + |\beta_2|$
  - etc…

F統計量：平方和に基づく統計量

- 候補が色々あるとはいったものの・・・
  - 我々は平方和を含む二次形式の分布を知っていた！
    - $\left(\frac{\mathbf{Y}-\boldsymbol{\mu}}{\sigma}\right)'\mathbf{A}\left(\frac{\mathbf{Y}-\boldsymbol{\mu}}{\sigma}\right) \sim \chi^2(r)$
- また, カイ二乗分布に従う確率変数を自由度で割ったものの比を取った統計量の事をF統計量と呼ぶのであった

F分布

- 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

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

F統計量：平方和に基づく統計量

- 二乗和を含む様に前回導いたt検定統計量を拡張することを考える
  - $t = \frac{Z}{\sqrt{V/(n-p)}}, Z \sim N(0,1), V \sim \chi^2(n-p)$ 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

- $\mathbf{c}_1'\boldsymbol{\beta}$の場合からはじめよう
- 分子Z_1の二乗について考える
- 帰無仮説のもとで：$Z_1 = \dfrac{\mathbf{c}_1'\widehat{\boldsymbol{\beta}}}{\sqrt{\sigma^2\mathbf{c}_1'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}_1}}$
- $Z_1^2 =?$

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

F統計量：平方和に基づく統計量

- $\left(\mathbf{c}_1'\hat{\boldsymbol{\beta}}\right)'\mathbf{V}^{-1}\left(\mathbf{c}_1'\hat{\boldsymbol{\beta}}\right)$, $\mathbf{V} = \mathbf{c}_1'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}_1$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

- 代入すると
  - $\left(\mathbf{c}_1'\hat{\boldsymbol{\beta}}\right)'\mathbf{V}^{-1}\left(\mathbf{c}_1'\hat{\boldsymbol{\beta}}\right) =?$

    $$\mathbf{Y}'\{\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}_1\mathbf{V}^{-1}\mathbf{c}_1'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\}\mathbf{Y}$$
  - 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
- $\mathrm{rank}(\mathbf{A}) = \mathrm{trace}(\mathbf{A}) = ?$
  - (定理1-6) $\mathrm{tr}\{\mathbf{Y}'\mathbf{A}\mathbf{Y}\} = \mathrm{tr}\{\mathbf{A}\mathbf{Y}\mathbf{Y}'\}$
  - $\mathrm{tr}\{\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}_1\mathbf{V}^{-1}\mathbf{c}_1'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\} = \mathrm{tr}\{\mathbf{V}^{-1}\mathbf{c}_1'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}_1\} = \mathrm{tr}\{\mathbf{V}^{-1}\mathbf{V}\} = 1$
  - $\mathbf{c}_1$：$(p \times 1)$

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
  - $Z_1^2 \sim \chi^2(1)$
  - $V = (n-p)\dfrac{s^2}{\sigma^2} \sim \chi^2(n-p)$
  - $t^2 = \dfrac{Z_1^2/1}{V/(n-p)} \sim F(1, n-p)$
- 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$\mathbf{C}\widehat{\boldsymbol{\beta}} \sim N(\mathbf{C}\boldsymbol{\beta}, \sigma^2\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}')$
  - $\mathbf{C}\widehat{\boldsymbol{\beta}} = \begin{pmatrix} \mathbf{c}_1'\widehat{\boldsymbol{\beta}} \\ \vdots \\ \mathbf{c}_m'\widehat{\boldsymbol{\beta}} \end{pmatrix}$

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

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

- これの二乗和を求めれば…

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

複数仮説版のF統計量

- $\left(\mathbf{C}\hat{\boldsymbol{\beta}}\right)'\mathbf{W}^{-1}\left(\mathbf{C}\hat{\boldsymbol{\beta}}\right)$, $\mathbf{W}=\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'$

$$\hat{\boldsymbol{\beta}}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

- 代入すると
  - $\left(\mathbf{C}\hat{\boldsymbol{\beta}}\right)'\mathbf{W}^{-1}\left(\mathbf{C}\hat{\boldsymbol{\beta}}\right)=?$

    $\mathbf{Y}'\{\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\mathbf{W}^{-1}\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\}\mathbf{Y}$
  - 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
- $\mathrm{rank}(\mathbf{B}) = \mathrm{trace}(\mathbf{B}) = ?$
  - (定理1-6) $\mathrm{tr}\{\mathbf{Y}'\mathbf{A}\mathbf{Y}\} = \mathrm{tr}\{\mathbf{A}\mathbf{Y}\mathbf{Y}'\}$
  - $\mathrm{tr}\{\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\mathbf{W}^{-1}\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\} = \mathrm{tr}\{\mathbf{W}^{-1}\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\} = \mathrm{tr}\{\mathbf{W}^{-1}\mathbf{W}\} = m$
  - $\mathbf{C}$：$(m \times p)$

F統計量：平方和に基づく統計量

- 結局
  - $\left(\mathbf{C}\hat{\boldsymbol{\beta}}\right)'\{\mathbf{C}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}\}^{-1}\mathbf{C}\hat{\boldsymbol{\beta}}/\sigma^2 \sim \chi^2(m)$
  - $V = (n-p)\dfrac{s^2}{\sigma^2} \sim \chi^2(n-p)$
  - $\dfrac{(\mathbf{C}\hat{\boldsymbol{\beta}})'\left\{\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}\right\}^{-1}\mathbf{C}\hat{\boldsymbol{\beta}}/m\sigma^2}{V/(n-p)} =$

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

具体的な例1

- $\mathbf{C\beta} = \mathbf{0}$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
- $\mathbf{C} = \mathbf{I} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$
    - $H_0: ?$
    - $H_0: \beta_1 = \beta_2 = \cdots = \beta_p = 0$

具体的な例2

- $\mathbf{C} = \begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}$
  - $H_0: \beta_1 - \beta_2 = 0, \beta_2 - \beta_3 = 0, \ldots, \beta_{p-1} - \beta_p = 0$
  - $H_0: \beta_1 = \beta_2 = \cdots = \beta_p$

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