線型推測論

第12回 総論

2021/6/4

慶應義塾大学病院

長島 健悟

最終回

- 線形仮説の検定の補足
  - 定理6-1の証明
- 発展的な話題の紹介
  - 一般化線型モデル
  - 線型混合モデル
  - 一般化推定方程式
- 残りの時間は質問の時間に充てます

線形仮説の検定の補足

- 定理6-1
  - $\mathbf{A}$が冪等行列であり, $r = \mathrm{rank}(\mathbf{A})$であるとする
  - $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{I})$のとき$\mathbf{Z}'\mathbf{A}\mathbf{Z} \sim \chi^2(r)$

定理12-1

- $\mathbf{A}$を定数対称行列とし, $\mathbf{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$とする
- $\mathbf{Y}'\mathbf{A}\mathbf{Y}$のモーメント母関数は
  $M_{\mathbf{Y}'\mathbf{A}\mathbf{Y}}(t)$
  $= |\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma}|^{-1/2} \exp\{-\boldsymbol{\mu}'[\mathbf{I} - (\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma})^{-1}]\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}/2\}$
  である

定理12-1の証明

- モーメント母関数 (積率母関数) の定義は

$$M_X(t) = \mathrm{E}[\exp(tX)]$$

- 定義から

$$M_{\mathbf{Y}'\mathbf{A}\mathbf{Y}}(t) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp(t\mathbf{Y}'\mathbf{A}\mathbf{Y})\, c_1 \exp\left\{-\frac{(\mathbf{Y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\boldsymbol{\mu})}{2}\right\} \mathrm{d}\mathbf{Y}$$

$$= c_1 \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left(\frac{2t\mathbf{Y}'\mathbf{A}\mathbf{Y} - (\mathbf{Y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\boldsymbol{\mu})}{2}\right) \mathrm{d}\mathbf{Y}$$

$$= c_1 \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left(\frac{-\mathbf{Y}'[\mathbf{I} - 2t\mathbf{A}\boldsymbol{\Sigma}]\boldsymbol{\Sigma}^{-1}\mathbf{Y} - 2\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{Y} + \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{2}\right) \mathrm{d}\mathbf{Y}$$

- ただし$c_1 = [(2\pi)^p |\boldsymbol{\Sigma}|]^{-1/2}$

定理12-1の証明

$$c_1 \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left(\frac{-\mathbf{Y}'[\mathbf{I}-2t\mathbf{A\Sigma}]\mathbf{\Sigma}^{-1}\mathbf{Y}-2\boldsymbol{\mu}'\mathbf{\Sigma}^{-1}\mathbf{Y}+\boldsymbol{\mu}'\mathbf{\Sigma}^{-1}\boldsymbol{\mu}}{2}\right) \mathrm{d}\mathbf{Y}$$

$$= c_1 c_2 \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{1}{[(2\pi)^p|\mathbf{V}|]^{1/2}} \exp\left(-\frac{(\mathbf{Y}-\boldsymbol{\theta})'\mathbf{V}^{-1}(\mathbf{Y}-\boldsymbol{\theta})}{2}\right) \mathrm{d}\mathbf{Y}$$

$$= c_1 c_2$$

- ただし, $t \in (-\delta, \delta)$の範囲で$\mathbf{I}-2t\mathbf{A\Sigma}$が正則になるような$\delta > 0$が存在することを仮定
  - $\boldsymbol{\theta}' = \boldsymbol{\mu}'[\mathbf{I}-2t\mathbf{A\Sigma}]^{-1}$
  - $\mathbf{V}^{-1} = [\mathbf{I}-2t\mathbf{A\Sigma}]\mathbf{\Sigma}^{-1}$
  - $c_2 = [(2\pi)^p|\mathbf{V}|]^{1/2} \exp\left(-\frac{\boldsymbol{\mu}'\mathbf{\Sigma}^{-1}\boldsymbol{\mu}-\boldsymbol{\theta}'\mathbf{V}^{-1}\boldsymbol{\theta}}{2}\right)$
- よって

$$M_{\mathbf{Y}'\mathbf{AY}}(t) = |\mathbf{I}-2t\mathbf{A\Sigma}|^{-1/2} \exp\{-\boldsymbol{\mu}'[\mathbf{I}-(\mathbf{I}-2t\mathbf{A\Sigma})^{-1}]\mathbf{\Sigma}^{-1}\boldsymbol{\mu}/2\}$$

定理6-1の証明

- $\mathbf{Z}'\mathbf{A}\mathbf{Z}$のモーメント母関数は
$$M_{\mathbf{Z}'\mathbf{A}\mathbf{Z}}(t) = |\mathbf{I} - 2t\mathbf{A}|^{-1/2}$$
- $\mathbf{I} - 2t\mathbf{A}$の固有値は$1 - 2t\lambda_i$であり, λ_iは$\mathbf{A}$の固有値であるから, $|\mathbf{I} - 2t\mathbf{A}| = \prod_{i=1}^{p} 1 - 2t\lambda_i$
$$M_{\mathbf{Z}'\mathbf{A}\mathbf{Z}}(t) = \left(\prod_{i=1}^{p} 1 - 2t\lambda_i\right)^{-1/2}$$

定理6-1の証明

- $\mathbf{A}$が冪等行列のとき, $\lambda_i = 0 \text{ or } 1$であり, $r = \mathrm{rank}(\mathbf{A}) = \mathrm{tr}(\mathbf{A})$であるから ($r$個の1の固有値と$p-r$個の0の固有値が出てくる)
- よって

$$M_{\mathbf{Z}'\mathbf{A}\mathbf{Z}}(t) = \left(\prod_{i=1}^{r} 1-2t\right)^{-1/2} = \left(\frac{1}{1-2t}\right)^{r/2}$$

となり, 自由度rのカイ二乗分布のモーメント母関数に等しい

- $\mathbf{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$の場合も証明でき, 非心カイ二乗分布が得られる

一般線型モデル

- 今でも研究はされているが一般線型モデルの結構な部分はかなり整理されている
- 統計モデルを扱う上で非常に重要な概念が沢山あるため, 教科書を一冊読み次の学習に進むとよい？
  - 一般線型モデルとモデルの構築, BLUEなどの推定量, 線型仮説の検定, 一般化逆行列

一般線型モデル

- 一般線型モデルだけではあまり扱われない重要な概念はものすごく沢山ある
  - 漸近理論, 最尤法, その他のモデル, etc...
  - これらは他の講義と自習で補って欲しい

一般化線型モデル

- 尤度理論をベースとし, モデルは線型でない場合も含む
  - 例えば, 線型回帰, ロジスティック回帰, ポアソン回帰などが含まれる
- 確率変数$Y_i\ (i = 1, \dots, n)$が互いに独立に正準形の指数型分布族に従うことを仮定
- Y_iの期待値$\mathrm{E}[Y_i] = \mu_i$に対して既知のリンク関数gを用いて$g(\mu_i) = \mathbf{x}_i'\boldsymbol{\beta}$を仮定

一般化線型モデル

- 指数型分布属
  - 確率密度関数が以下の形をした分布

$$f(y;\theta) = \exp\{\eta(\theta)T(y) - A(\theta) + B(y)\}$$

  - $\eta(\theta) = \theta$の場合を正準形という (正準形: シンプルな数式)

一般化線型モデル

- 特に一般化線型モデルの場合は以下の指数型分布属を想定する

$$f(y_i; \theta_i, \phi)$$
$$= \exp\{[y_i\theta_i - b(\theta_i)]/\phi^2 - c(y_i, \phi)\}$$

- $b(\theta_i)$やϕは想定した確率分布によって定まる

一般化線型モデル

- よく使う一部の例を紹介

モデル	分布	リンク関数
線型回帰	正規分布	$g(\mu_i) = \mu_i$
ロジスティック回帰	二項分布	$g(\mu_i) = \log \frac{\mu_i}{1 - \mu_i}$
ポアソン回帰	ポアソン分布	$g(\mu_i) = \log \mu_i$

一般化線型モデル

- 一般化線型モデルも多くの回帰モデルを統一的に表現でき, 同様の考え方で推定や検定を構築できる
  - 理論的な基盤は最尤法なので, 最尤法についてきちんと学習しておくとよい
  - McCulloch, Searle & Neuhaus (2008) などがおすすめ (邦訳あり)

線型混合モデル

- 線型混合モデルは固定効果と変量効果の両方を含む線型モデル (最尤法がベース)

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon}$$

$$\boldsymbol{\epsilon} \sim \mathrm{N}(\mathbf{0}, \mathbf{R}), \boldsymbol{\gamma} \sim \mathrm{N}(\mathbf{0}, \mathbf{G})$$

- $\mathbf{Z}$：変量効果の$n \times q$デザイン行列
- $\boldsymbol{\gamma}$：変量効果ベクトル$q \times 1$ ($\boldsymbol{\epsilon}$と独立)
- $\mathbf{R}$：誤差分散共分散行列
- $\mathbf{G}$：変量効果の分散共分散行列

線型混合モデル

- 前述の仮定の下で
  - $\mathbf{Y} \sim N(\mathbf{X\beta}, \mathbf{V}), \mathbf{V} = \mathbf{ZGZ} + \mathbf{R}$
  - これは一般線型モデルの時に検討済であった
  - $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$
  - $\hat{\mathbf{V}}$はプロファイル尤度法により求める

線型混合モデル

- 独立性の仮定より$\mathbf{Y},\boldsymbol{\gamma}$の同時分布は

$$f(\mathbf{Y},\boldsymbol{\gamma}) = g(\mathbf{Y} \mid \boldsymbol{\gamma})h(\boldsymbol{\gamma})$$

$$\log f(\mathbf{Y},\boldsymbol{\gamma}) = -\frac{1}{2}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\boldsymbol{\gamma})'\mathbf{R}^{-1}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\boldsymbol{\gamma}) -\frac{1}{2}\boldsymbol{\gamma}'\mathbf{G}^{-1}\boldsymbol{\gamma} + \text{constant}$$

線型混合モデル

- *Mixed Model Equations* (Henderson et al., 1959)：同時分布から導いた対数尤度関数について$\boldsymbol{\beta}$と$\boldsymbol{\gamma}$について偏微分したもの$= \mathbf{0}$とおいた式

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\widetilde{\boldsymbol{\beta}} + \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\tilde{\boldsymbol{\gamma}} = \mathbf{X}'\mathbf{R}^{-1}\mathbf{Y}$$
$$\mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}\widetilde{\boldsymbol{\beta}} + (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})\tilde{\boldsymbol{\gamma}} = \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Y}$$

- $\therefore \tilde{\boldsymbol{\gamma}} = \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}), \widetilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}$

線型混合モデル

- 経時測定データの解析に比較的よく使われている
- 昨今では欠測データの解析にもよく使われる
  - Fitzmaurice, Laird & Ware (2011) や McCulloch, Searle & Neuhaus (2008) などがおすすめ
- 正規分布以外の場合の一般化線型混合モデルもある

一般化推定方程式

- Liang & Zeger (1986)のGeneralized Estimating Equations (GEE)
  - (一般化)線型混合モデルでは結果変数の確率分布を完全に仮定したが, GEEでは周辺分布だけを仮定する
  - 周辺分布では相関が定まらないため, 作業相関行列というものを与える
  - Wedderburn (1974)のquasi-likelihoodを多変量に一般化した方法といえる

一般化推定方程式

- $\mathbf{Y}_i = \left(Y_{i1}, \dots, Y_{it}, \dots, Y_{in_i}\right)'$：$n_i \times 1$次元結果変数ベクトル, $\mathbf{Y} = (\mathbf{Y}_1', \dots, \mathbf{Y}_K')'$
- $\mathbf{x}_{it} = \left(x_{it1}, \dots, x_{itj}, \dots, x_{itp}\right)'$：$p \times 1$次元のデザインベクトル, $\mathbf{X}_i = \left(\mathbf{x}_{i1}, \dots, \mathbf{x}_{in_i}\right)'$：$n_i \times 1$次元のデザイン行列
- $\boldsymbol{\beta}$：$p \times 1$次元モデルパラメータベクトル
- $i = 1, \dots, K$：個体を表わす添え字
- $t = 1, \dots, n_i$：各個体の時点の添え字

一般化推定方程式

- 仮定：指数型分布族の周辺密度を持つ
  - $f(Y_{it} = y_{it}) = \exp\{[y_{it}\theta_{it} - a(\theta_{it}) + b(y_{it})]/\phi^2\}$
  - $\theta_{it} = h(\eta_{it}), \eta_{it} = \mathbf{x}'_{it}\boldsymbol{\beta}$
- このとき

$$\mathrm{E}[Y_{it}] = a'(\theta_{it}), \mathrm{Var}[Y_{it}] = a''(\theta_{it})/\phi$$

がなりたつ

一般化推定方程式

- 作業相関行列$\mathbf{R}_i(\boldsymbol{\alpha})$：$n_i \times n_i$次元
  - $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_s)$：$s \times 1$次元パラメータベクトル
- $\mathbf{V}_i = \mathbf{A}_i^{1/2}\mathbf{R}_i(\boldsymbol{\alpha})\mathbf{A}_i^{1/2}/\phi$とすると
  - $\mathbf{R}_i(\boldsymbol{\alpha})$が真の相関行列に等しいとき
  - $\mathbf{V}_i = \mathrm{Cov}[\mathbf{Y}_i]$が成り立つ

一般化推定方程式

- これらより以下の推定方程式を定義する

$$\sum_{i=1}^{K} \mathbf{U}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{i=1}^{K} \mathbf{D}_i \mathbf{V}_i^{-1} \mathbf{S}_i = \mathbf{0}$$

- $\mathbf{D}_i = \frac{\partial a'(\boldsymbol{\theta}_i)}{\partial \boldsymbol{\beta}'}, \mathbf{S}_i = \mathbf{Y}_i - a'(\boldsymbol{\theta}_i), a'(\boldsymbol{\theta}_i) = \left(a'(\theta_{i1}), \dots, a'\left(\theta_{in_i}\right)\right)'$

一般化推定方程式

- 実際には$\boldsymbol{\alpha}$とϕが未知のため, 推定量をプラグインした二段階推定が行われる

$$\sum_{i=1}^{K} \mathbf{U}_i\left(\boldsymbol{\beta}, \widehat{\boldsymbol{\alpha}}\left\{\boldsymbol{\beta}, \hat{\phi}(\boldsymbol{\beta})\right\}\right) = \mathbf{0}$$

- 上の推定方程式に基づく方法をGEEとよび, 推定方程式の解が$\widehat{\boldsymbol{\beta}}$の推定量になる

一般化推定方程式

- $K \to \infty$のとき, $\mathbf{R}_i(\boldsymbol{\alpha})$が誤っていたとしても$\hat{\boldsymbol{\beta}}$は一致性を持つ
  - ただし, 誤っていた場合は検出力が低下する
  - そして現実は$K \to \infty$とはいかない

一般化推定方程式

- こちらも経時測定データの解析に比較的よく使われている
  - 理論的なベースは独自のものなので, 原典(Liang & Zeger, 1986)が理解できるように読んでおくとよい
  - Fitzmaurice, Laird & Ware (2011) などもおすすめ

正則化

- 過学習の回避
  - リッジ回帰
$$Q_R = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + \lambda_R \boldsymbol{\beta}'\boldsymbol{\beta}$$
$$\boldsymbol{\beta}_R = (\mathbf{X}'\mathbf{X} + \lambda_R \mathbf{I})^{-1}\mathbf{X}'\mathbf{Y}$$
  - $(\mathbf{X}'\mathbf{X})^{-1}$が存在しなくても推定できる
- 過学習の回避＋スパース推定
  - Lasso
$$Q_L = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + \lambda_L \sum_{j=1}^{p} |\beta_j|$$

まとめ

- 一般化線型モデル, 線型混合モデル, 一般化推定方程式を紹介した
  - これらは別の講義でも詳しく扱われる

参考文献

- McCulloch CE, Searle SR, Neuhaus JM. *Generalized, Linear, and Mixed Models, 2nd Edition*. Wiley, 2008.
- Henderson CR, Kempthorne O, Searle SR, von Krosigk CM. The estimation of environmental and genetic trends from records subject to culling. *Biometrics* 1959;**15**(2):192–218.
- Fitzmaurice GM, Laird NM, Ware JH. *Applied Longitudinal Analysis, 2nd Edition*. Wiley, 2011.
- Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986;**73**(1):13–22.
- Wedderburn RWM. Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method. *Biometrika* 1974;**61**:439–447.
- Friedman J, Hastie T, Hoefling H, Tibshirani R. Pathwise coordinate optimization. *The Annals of Applied Statistics* 2007;**2**(1):302–332.