応用システム工学
第5回　指数型分布族と一般化線形モデル

平成21年7月03日

指数型分布族の対数尤度関数 最尤推定の準備

- 指数型分布族の確率密度関数

$$f(y,\theta)=\exp[a(y)b(\theta)+c(\theta)+d(y)]$$

- 指数型分布族の尤度関数

$$L(\theta,y)=\exp[a(y)b(\theta)+c(\theta)+d(y)]$$

- 指数型分布族の対数尤度関数

$$l(\theta,y)=\log L(\theta,y)=a(y)b(\theta)+c(\theta)+d(y)$$

- スコア統計量
  - 対数尤度関数の θ に関する微分
    - パラメータの推測に利用できる→最尤推定

指数型分布族の対数尤度関数

- スコア統計量U

$$U(\theta, y) = \frac{d}{d\theta} l(\theta, y) = a(y) b'(\theta) + c'(\theta)$$

  – Uはyに依存するので確率変数

$$U(Y) = a(Y) b'(\theta) + c'(\theta)$$

  – Uの期待値はaの期待値で表される
    - aはYに依存する確率変数

$$E[U(Y)] = E[a(Y) b'(\theta) + c'(\theta)]$$

指数型分布族の対数尤度関数

- スコア統計量Uの期待値

$$
\begin{aligned}
E[U(Y)] &= E[a(Y)b'(\theta)+c'(\theta)] \\
&= b'(\theta)E[a(Y)]+c'(\theta) \\
&= b'(\theta)\left\{-\frac{c'(\theta)}{b'(\theta)}\right\}+c'(\theta) \\
&= -c'(\theta)+c'(\theta) \\
&= 0
\end{aligned}
$$

  - スコア統計量の期待値は0となる

指数型分布族の対数尤度関数

- スコア統計量Uの分散→情報量を表す
  - スコア統計量は確率変数a(Y)の線形変換

$$U(Y) = a(Y)b'(\theta) + c'(\theta)$$

$$\mathfrak{I} = \mathrm{var}[U(Y)] = \mathrm{var}[a(Y)b'(\theta) + c'(\theta)] = b'(\theta)^2 \ \mathrm{var}[a(Y)]$$

$$\mathrm{var}[U(Y)] = b'(\theta)^2 \ \mathrm{var}[a(Y)]$$

$$= b'(\theta)^2 \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{b'(\theta)^3}$$

$$= \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{b'(\theta)}$$

指数型分布族の対数尤度関数

- スコア統計量のほかの性質1

$$\mathrm{var}\left[U(Y)\right] = E\left[U(Y)^2\right] - E\left[U(Y)\right]^2$$
$$= E\left[U(Y)^2\right]$$

– $\because \; E\left[U(Y)\right] = 0$

- スコア統計量のほかの性質2

$$\frac{d}{d\theta}U(Y) = U'(Y) = \frac{d}{d\theta}\left[a(Y)b'(\theta) + c'(\theta)\right]$$
$$= a(Y)b''(\theta) + c''(\theta)$$

指数型分布族の対数尤度関数

- スコア統計量のほかの性質

$$
\begin{aligned}
E[U'(Y)] &= E[a(Y)b''(\theta)+c''(\theta)] \\
&= b''(\theta)E[a(Y)]+c''(\theta) \\
&= b''(\theta)\left\{-\frac{c'(\theta)}{b'(\theta)}\right\}+c''(\theta) \\
&= \frac{-b''(\theta)c'(\theta)+b'(\theta)c''(\theta)}{b'(\theta)} \\
&= -\mathrm{var}[U(Y)] = -\mathfrak{I}
\end{aligned}
$$

最尤推定で利用

一般化線形モデル

一般化線形モデル

- 指数型分布族の分布を持つ
- 独立な確率変数Y1,・・・,YNの集合の条件
  - 条件1:各確率変数Yiの分布は正準形(a(y)=y)で一つのパラメータθiに依存

$$f\left(y_i;\theta_i\right)=\exp\left[y_i b\left(\theta_i\right)+c\left(\theta_i\right)+d\left(y_i\right)\right]$$

  - 条件2:全ての確率変数Yiの分布は同じ型

一般化線形モデル

- 確率変数Y1,･･･,YNの同時確率密度関数

$$f(y_1,\cdots,y_N;\theta_1,\cdots,\theta_N)=\prod_{i=1}^{N}\exp\left[y_i b(\theta_i)+c(\theta_i)+d(y_i)\right]$$

$$=\exp\left[\sum_{i=1}^{N}y_i b(\theta_i)+\sum_{i=1}^{N}c(\theta_i)+\sum_{i=1}^{N}d(\theta_i)\right]$$

一般化線形モデル

- 少数のパラメータでモデルを表現
  - 確率変数Yiの期待値μiがパラメータβ1,・・・,βp(p<N)の関数として表される
    - 説明変数ベクトル
      - ダミー変数
    - パラメータベクトル
    - 単調かつ微分可能な連結関数

$$x_i^T = \left[x_{i1}, \cdots, x_{ip}\right]$$

$$\beta^T = \left[\beta_1, \cdots, \beta_p\right]$$

$$g(\mu_i) = x_i^T \beta$$

一般化線形モデル

- モデルの構成
  - 確率変数Y1,・・・,YN(指数型分布族)
  - パラメータベクトル β の集合と説明変数の行列

$$\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \qquad X = \begin{bmatrix} x_1^T \\ \vdots \\ x_N^T \end{bmatrix} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{N1} & \cdots & x_{Np} \end{bmatrix}$$

  - 単調な連結関数g $g(\mu_i) = x_i^T \beta \qquad \mu_i = E[Y_i]$

パラメータの推定

- 推定法
  - 最尤推定法　→　これをする
  - 最小二乗法
  - その他
- 一般化線形モデル
  - 最尤法によるパラメータの点推定・区間推定
    - 数値解析
    - 収束計算(ニュートンラプソン法)

一般化線形モデルの最尤推定

- 一般化線形モデルの確率変数Y1,・・・,YN
  - 期待値 $\mu_i = E[Y_i]$
  - 連結関数 $g(\mu_i) = x_i{}^T \beta = \eta_i$
    - 確率変数に関係するパラメータβの推定
  - 各確率変数に対する対数尤度関数

$$l_i = y_i b(\theta_i) + c(\theta_i) + d(y_i)$$

$$\mathrm{var}[Y_i] = \frac{b''(\theta_i)c'(\theta_i) - c''(\theta_i)b'(\theta_i)}{b'(\theta_i)^3}$$

$$E[Y_i] = \mu_i = -\frac{c'(\theta_i)}{b'(\theta_i)}$$

一般化線形モデルの最尤推定

- 全ての確率変数に対する対数尤度関数

$$l = \sum_{i=1}^{N} l_i = \sum_{i=1}^{N} y_i b(\theta_i) + \sum_{i=1}^{N} c(\theta_i) + \sum_{i=1}^{N} d(\theta_i)$$

  – 対数尤度関数を最大化するパラメータ β
    - スコア関数　$U(\beta) = 0$

$$\frac{\partial l}{\partial \beta_j} = U_j = \sum_{i=1}^{N} \left[\frac{\partial l_i}{\partial \beta_j} \right] = \sum_{i=1}^{N} \left[\frac{\partial l_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \beta_j} \right]$$

    - 非線形関数の解を求める→ニュートンラプソン法

ニュートンラプソン法

- 非線形関数yがx軸と交わる点を求める

$$y(x) = 0$$

  – 値 $x^{(m-1)}$ での関数の傾き

$$\left.\frac{dy}{dx}\right|_{x=x^{(m-1)}} = y'\left(x^{(m-1)}\right) \cong \frac{y\left(x^{(m)}\right) - y\left(x^{(m-1)}\right)}{x^{(m)} - x^{(m-1)}}$$

  – $y\left(x^{(m)}\right) = 0$ なら

$$y'\left(x^{(m-1)}\right) = \frac{0 - y\left(x^{(m-1)}\right)}{x^{(m)} - x^{(m-1)}}$$

    - 収束するまで計算

$$x^{(m)} = x^{(m-1)} - \frac{y\left(x^{(m-1)}\right)}{y'\left(x^{(m-1)}\right)}$$

      – スコア関数の微分を求める必要有

一般化線形モデルの最尤推定

$$\left[\frac{\partial l_i}{\partial \theta_i}\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial \beta_j}\right]$$

- スコア関数内微分第一項

$$\frac{\partial l_i}{\partial \theta_i} = \frac{\partial}{\partial \theta_i}\left[y_i b(\theta_i) + c(\theta_i) + d(y_i)\right] = y_i b'(\theta_i) + c'(\theta_i)$$

  – 期待値の関係 $E[Y_i] = \mu_i = -\dfrac{c'(\theta_i)}{b'(\theta_i)}$

$$c'(\theta_i) = -\mu_i b'(\theta_i)$$

$$\frac{\partial l_i}{\partial \theta_i} = y_i b'(\theta_i) + c'(\theta_i) = y_i b'(\theta_i) - \mu_i b'(\theta_i) = b'(\theta_i)\left[y_i - \mu_i\right]$$

一般化線形モデルの最尤推定

- ∴ $$\frac{\partial l_i}{\partial \theta_i} = y_i b'(\theta_i) + c'(\theta_i) = y_i b'(\theta_i) - \mu_i b'(\theta_i) = b'(\theta_i)[y_i - \mu_i]$$
- スコア関数内微分第二項 $$\left[\frac{\partial l_i}{\partial \theta_i}\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial \beta_j}\right]$$

$$\frac{\partial \mu_i}{\partial \theta_i} = \frac{\partial}{\partial \theta_j}\left[-\frac{c'(\theta_i)}{b'(\theta_i)}\right] = -\frac{b'(\theta_i)c''(\theta_i) - b''(\theta_i)c'(\theta_i)}{b'(\theta_i)^2}$$

– 前出の分散 $$\mathrm{var}[Y_i] = \frac{b''(\theta_i)c'(\theta_i) - c''(\theta_i)b'(\theta_i)}{b'(\theta_i)^3}$$

一般化線形モデルの最尤推定

$$\left[\frac{\partial l_i}{\partial \theta_i}\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial \beta_j}\right]$$

- 分散を代入

$$\frac{\partial \mu_i}{\partial \theta_i} = b'(\theta_i)\mathrm{var}[Y_i]$$

$$\frac{\partial \theta_i}{\partial \mu_i} = \frac{1}{\frac{\partial \mu_i}{\partial \theta_i}} = \frac{1}{b'(\theta_i)\mathrm{var}[Y_i]}$$

一般化線形モデルの最尤推定

- スコア関数内微分第三項 $\left[\frac{\partial l_i}{\partial \theta_i}\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial \beta_j}\right]$

$$\frac{\partial \mu_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_j}\frac{\partial \eta_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_j}\frac{\partial}{\partial \beta_j}\left[x_i{}^T \beta\right] = \frac{\partial \mu_i}{\partial \eta_j} x_{ij}$$

- スコア関数

$$U_j = \sum_{i=1}^{N}\left[\frac{\partial l_i}{\partial \theta_i}\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial \beta_j}\right] = \sum_{i=1}^{N}\left[b'(\theta_i)\left[y_i - \mu_i\right]\frac{1}{b'(\theta_i)\mathrm{var}\left[Y_i\right]}\frac{\partial \mu_i}{\partial \eta_j} x_{ij}\right]$$

$$= \sum_{i=1}^{N}\left[\frac{y_i - \mu_i}{\mathrm{var}\left[Y_i\right]}\frac{\partial \mu_i}{\partial \eta_j} x_{ij}\right]$$