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22年7月02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
  - パラメータθ，確率変数Y

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

- 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

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

指数型分布族の対数尤度関数 最尤推定の準備

$$l(\theta; y) = a(y)b(\theta) + c(\theta) + d(y)$$

- 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
  - 対数尤度関数の微分の期待値・分散を調べる
    - まずa(y)の期待値・分散を調べておく

指数型分布族の性質

- 確率変数a(Y)の期待値を求める

$$E[a(Y)]=\int a(y)f(y;\theta)dy$$

  - 確率密度関数の性質(全体の確率は1)

$$\int f(y;\theta)dy=1$$

    - パラメータに対する微分の性質

$$\frac{d}{d\theta}\int f(y;\theta)dy=\frac{d}{d\theta}1=0$$

$$\Rightarrow \quad \int \frac{d}{d\theta}f(y;\theta)dy=0$$

指数型分布族の性質

- 指数型分布族の確率密度関数f(y;θ)のパラメータθに対する微分

$$
\begin{aligned}
\frac{d}{d\theta} f(y;\theta) &= \frac{d}{d\theta} \exp\left[a(y)b(\theta) + c(\theta) + d(y)\right] \\
&= \left[a(y)b'(\theta) + c'(\theta)\right] \\
&\quad \times \exp\left[a(y)b(\theta) + c(\theta) + d(y)\right] \\
&= \left[a(y)b'(\theta) + c'(\theta)\right] f(y;\theta)
\end{aligned}
$$

指数型分布族の性質

- パラメータθに対する確率密度関数f(y; θ)の微分の性質を考える

$$\int \frac{d}{d\theta} f\left(y;\theta\right)dy = 0$$ より

$$\int \left[a\left(y\right)b'\left(\theta\right) + c'\left(\theta\right)\right] f\left(y;\theta\right)dy$$

$$= b'\left(\theta\right)\int a\left(y\right) f\left(y;\theta\right)dy + c'\left(\theta\right)\int f\left(y;\theta\right)dy$$

$$= 0$$

指数型分布族の性質

- 期待値 $E[a(Y)] = \int a(y) f(y;\theta) dy$
- 確率密度分布 $\int f(y;\theta) dy = 1$

$\left.\begin{matrix} \\ \end{matrix}\right\}$ を代入すると

$$b'(\theta)\int a(y) f(y;\theta) dy + c'(\theta)\int f(y;\theta) dy$$
$$= b'(\theta) E[a(Y)] + c'(\theta)$$
$$= 0$$

- 確率変数a(y)の期待値 $E[a(Y)] = -\dfrac{c'(\theta)}{b'(\theta)}$

と表される

指数型分布族の性質

- 確率変数a(Y)の分散

$$\mathrm{var}[a(Y)] = \int \{a(y) - E[a(y)]\}^2 f(y;\theta)dy$$

- 確率密度関数のパラメータに対する2階微分

$$\frac{d^2}{d\theta^2} f(y;\theta) = \frac{d^2}{d\theta^2} \exp[a(y)b(\theta) + c(\theta) + d(y)]$$

$$= \frac{d}{d\theta}\{[a(y)b'(\theta) + c'(\theta)]f(y;\theta)\}$$

$$= [a(y)b''(\theta) + c''(\theta)]f(y;\theta) + [a(y)b'(\theta) + c'(\theta)]\frac{d}{d\theta} f(y;\theta)$$

$$= [a(y)b''(\theta) + c''(\theta)]f(y;\theta) + [a(y)b'(\theta) + c'(\theta)][a(y)b'(\theta) + c'(\theta)]f(y;\theta)$$

$$= [a(y)b''(\theta) + c''(\theta)]f(y;\theta) + [a(y)b'(\theta) + c'(\theta)]^2 f(y;\theta)$$

指数型分布族の性質

前頁式第二項

$$\boxed{\left[a(y)b'(\theta)+c'(\theta)\right]^2 f(y,\theta)=}\, b'(\theta)^2\left[a(y)+\frac{c'(\theta)}{b'(\theta)}\right]^2 f(y,\theta)$$

$$=b'(\theta)^2\left[a(y)-E[a(Y)]\right]^2 f(y,\theta)$$

- 全体確率のパラメータに対する2階微分の性質

$$\frac{d^2}{d\theta^2}\int f(y;\theta)dy=\int\frac{d^2}{d\theta^2}f(y;\theta)dy=\frac{d^2}{d\theta^2}1=0$$

前頁式第二項を分散で表す

$$\int\left[a(y)b'(\theta)+c'(\theta)\right]^2 f(y,\theta)dy=\int b'(\theta)^2\left[a(y)-E[a(Y)]\right]^2 f(y,\theta)dy$$

$$=b'(\theta)^2\int\left[a(y)-E[a(Y)]\right]^2 f(y,\theta)dy$$

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

指数型分布族の性質

前前頁式第一項を積分

$$\boxed{\int\left[a(y)b''(\theta)+c''(\theta)\right]f(y,\theta)dy}=\int a(y)b''(\theta)f(y,\theta)dy+\int c''(\theta)f(y,\theta)dy$$
$$=b''(\theta)\int a(y)f(y,\theta)dy+c''(\theta)\int f(y,\theta)dy$$
$$=b''(\theta)E\left[a(Y)\right]+c''(\theta)$$

前前頁式の積分

$$0=b''(\theta)E\left[a(Y)\right]+c''(\theta)+b'(\theta)^2\,\mathrm{var}\left[a(Y)\right]$$
$$=b''(\theta)\left\{-\frac{c'(\theta)}{b'(\theta)}\right\}+c''(\theta)+b'(\theta)^2\,\mathrm{var}\left[a(Y)\right]$$

指数型分布族の性質

$$b'(\theta)^2 \operatorname{var}[a(Y)] = \frac{b''(\theta)c'(\theta)}{b'(\theta)} - c''(\theta)$$

- 確率変数a(y)の分散

$$\operatorname{var}[a(Y)] = \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{b'(\theta)^3}$$

  – 分布形式とパラメータが分かれば期待値と分散が求まる

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である

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

  – 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

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

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

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

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

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

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

$$\operatorname{var}[U(Y)] = E\left[U(Y)^2\right] - E[U(Y)]^2$$
$$= E\left[U(Y)^2\right]$$
$$- \because \quad E[U(Y)] = 0$$

- 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2のつづき

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

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

$$f\left(y_i;\theta_i\right)=\exp\left[y_i b_i\left(\theta_i\right)+c_i\left(\theta_i\right)+d_i\left(y_i\right)\right]$$

  - 条件2:全ての確率変数Yiの分布は同じ型(正規分布，二項分布等)

$$\begin{cases} b=b_1,b_2,\cdots b_i,\cdots,b_N \\ c=c_1,c_2,\cdots c_i,\cdots,c_N \\ d=d_1,d_2,\cdots d_i,\cdots,d_N \end{cases}$$

一般化線形モデル

- 確率変数Y1,・・・,YN(各々同じ分布)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　$x_i^T = [x_{i1}, \cdots, x_{ip}]$
      - ダミー変数(デザイン行列Xのi行)
    - パラメータベクトル　$\beta^T = [\beta_1, \cdots, \beta_p]$
    - 単調かつ微分可能な連結関数g

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

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 $g(\mu_i) = x_i{}^T \beta \qquad \mu_i = E[Y_i]$