

第8回 母数の推定

ベイズ解
最小2乗解

- 説明変数が複数ある場合のロジスティック回帰 `logistic6.R`
- Newton-Raphson 法で `gradient` や `hessian` がない場合 `NR4.R`

ベイズ統計学の事後分布

y をデータ、 ξ をモデルのパラメタとする。

モデル分布 = パラメタが所与の時のデータの同時分布 = 尤度

$$f(y|\xi) = L(\xi) \quad (84)$$

$$= \prod_{i=1}^N f(y_i|\xi) \quad (85)$$

パラメタの事前分布は ρ を超母数 (hyperparameter) として

$$h(\xi|\rho) \quad (86)$$

事後分布は、データとパラメタの同時分布に比例する。

$$h(\xi|y, \rho) \propto g(y, \xi|\rho) = f(y|\xi) \times h(\xi|\rho) \quad (87)$$

$$= L(\xi) \times h(\xi|\rho) \quad (88)$$

比例常数は ξ に関する定積分が 1 となるように：

$$h(\xi|y, \rho) = \frac{1}{\int_{\xi} L(\xi) \times h(\xi|\rho) d\xi} \times L(\xi) \times h(\xi|\rho) \quad (89)$$

3

点推定値

事後分布の評価、特に点推定値は

- EAP 事後分布から計算した g の平均や事後分散共分散行列を利用
- MAP 事後分布のモードを利用

事後平均や分散共分散行列は事後分布から発生した ξ の乱数の平均や分散共分散行列を用いる。(MCMC)

MAP は最尤解のプログラムが利用可能。

4

R のコード

- 最尤解 : $L(\xi)$ の最大値を求める。
- MAP : $L(\xi) \times h(\xi|\rho)$ の最大値を求める。

```
# Bayes MAP
# prior: (mu, invSigma) = hyperparameters
logpriorpdf <- function( beta, mu, invSigma ){
  # multivariate normal pdf
  lpdf=-0.5*c(beta-mu)%*%invSigma%*%c(beta-mu) + 0.5*det(invSigma)
  return(lpdf)
} # logpriorpdf

mlogpostpdf <- function( beta, data, mu, invSigma ){
  locx=which(regexpr("x",colnames(data)) > 0)
  X=as.matrix(data[,locx]); f=data$f; r=data$r
  p=P(X,beta)
  # log post pdf of x is log likelihood + log prior pdf
  llh=sum( r*log(p)+(f-r)*log(1-p) ) + logpriorpdf( beta, mu, invSigma )
  return( -llh )
} # end of mlogpostpdf

mu=rep(0,length(beta0)); invSigma=0.01*diag(length(beta0))
resNR=NR( x=beta0, mlogpostpdf, data=data , mu=mu, invSigma=invSigma )
resnlminb=nlminb( beta0, mlogpostpdf, data=data
, mu=mu, invSigma=invSigma, control=list(trace=1) )
```

5

数値例

最尤解
1.464336 0.2127619

MAP rho = N(0, 100 I)
1.462796 0.2124356

MAP rho = N(0, I)
1.334631 0.1857594

6

フルベイズ解 よろしくお願いします。

7

正規分布する加算的誤差

データは、その期待値に正規分布をする誤差が加わって生成されている。

$$y_i = \hat{y}_i + e_i, \quad i = 1, 2, \dots, N \quad (90)$$

$$\hat{y}_i = g(\mathbf{x}_i | \boldsymbol{\xi}) \quad (91)$$

$$e_i \sim N(\mathbf{0}, \sigma^2 \mathbf{I}) \quad (92)$$

したがって

$$y_i \sim N(\hat{y}_i, \sigma^2 \mathbf{I}) \quad (93)$$

例えば

$$y_i = \xi_1 + \exp(\xi_2(x_i - \xi_3)) + e_i$$

$$y_{ij} = \sqrt{\sum_{a=1}^A (\xi_{ia} - \xi_{ja})^2} + e_{ij}$$

8

データの同時分布 = 尤度関数

$$L(\boldsymbol{\xi}) = \prod_{i=1}^N f(y_i|\boldsymbol{\xi}) \quad (94)$$

$$= \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}(y_i - \hat{y}_i)^2/\sigma^2\right) \quad (95)$$

対数尤度

$$\ln L(\boldsymbol{\xi}) = -\frac{1}{2} \left(N \log \sigma^2 + \sum_{i=1}^N (y_i - \hat{y}_i)^2/\sigma^2 \right) \quad (96)$$

$$= -\frac{1}{2} \left(N \log \sigma^2 + \sum_{i=1}^N (y_i - g(\mathbf{x}_i|\boldsymbol{\xi}))^2/\sigma^2 \right) \quad (97)$$

9

$\boldsymbol{\xi}$ が所与の場合の σ^2 の推定値は

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (y_i - g(\mathbf{x}_i|\boldsymbol{\xi}))^2 \quad (98)$$

したがって、最尤解は

$$RSS(\boldsymbol{\xi}) = \sum_{i=1}^N (y_i - g(\mathbf{x}_i|\boldsymbol{\xi}))^2 \quad (99)$$

を最小とする $\boldsymbol{\xi}$ として求めることが出来る。

最小2乗解

10

Gauss-Newton 法

最適化されるべき関数が最小2乗基準の場合 $r(\boldsymbol{\xi}) = \text{RSS}(\boldsymbol{\xi})$

いま、求めるべきパラメタの仮の値を $\boldsymbol{\xi}_{(\ell)}$ とする。モデルの各要素 $g(\boldsymbol{x}_i|\boldsymbol{\xi})$ を $\boldsymbol{\xi}_{(\ell)}$ のまわりで1次の項まで Taylor 展開したものを $N \times 1$ のベクトルとしてまとめれば、

$$g(\boldsymbol{\xi}) \approx g(\boldsymbol{\xi}_{(\ell)}) + \boldsymbol{F}(\boldsymbol{\xi}_{(\ell)})(\boldsymbol{\xi} - \boldsymbol{\xi}_{(\ell)}) \quad (102)$$

となるが、ここに、

$$\boldsymbol{F}(\boldsymbol{\xi}) = \frac{\partial \boldsymbol{g}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} = \left[\frac{\partial g(\boldsymbol{x}_i|\boldsymbol{\xi})}{\partial \boldsymbol{\xi}_k} \right], \quad i = 1, 2, \dots, N, \quad k = 1, 2, \dots, Q \quad (103)$$

は、ヤコビアン行列と呼ばれる $N \times Q$ の1次微分行列、また、 $\boldsymbol{F}(\boldsymbol{\xi}_{(\ell)})$ はそれを $\boldsymbol{\xi}_{(\ell)}$ で評価した行列

$$\boldsymbol{F}(\boldsymbol{\xi}_{(\ell)}) = \left. \frac{\partial \boldsymbol{g}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} \right|_{\boldsymbol{\xi}=\boldsymbol{\xi}_{(\ell)}} \quad (104)$$

である。

11

この1次近似を RSS に代入すれば、

$$\text{RSS}(\boldsymbol{\xi}) = \sum_{i=1}^N (y_i - g(\boldsymbol{x}_i|\boldsymbol{\xi}))^2 = \|\boldsymbol{y} - \boldsymbol{g}(\boldsymbol{\xi})\|^2 \quad (105)$$

$$\approx \|\boldsymbol{y} - (g(\boldsymbol{\xi}_{(\ell)}) + \boldsymbol{F}(\boldsymbol{\xi}_{(\ell)})(\boldsymbol{\xi} - \boldsymbol{\xi}_{(\ell)}))\|^2 \quad (106)$$

$$= \|\boldsymbol{y} - g(\boldsymbol{\xi}_{(\ell)}) - \boldsymbol{F}(\boldsymbol{\xi}_{(\ell)})(\boldsymbol{\xi} - \boldsymbol{\xi}_{(\ell)})\|^2 \quad (107)$$

となるから、

$$\frac{\partial \text{RSS}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} \approx -2\boldsymbol{F}(\boldsymbol{\xi}_{(\ell)})'(\boldsymbol{y} - \boldsymbol{g}(\boldsymbol{\xi}_{(\ell)})) + 2\boldsymbol{F}(\boldsymbol{\xi}_{(\ell)})'\boldsymbol{F}(\boldsymbol{\xi}_{(\ell)})(\boldsymbol{\xi} - \boldsymbol{\xi}_{(\ell)}) = \mathbf{0}$$

の解は

$$\boldsymbol{\xi} = \boldsymbol{\xi}_{(\ell)} + (\boldsymbol{F}(\boldsymbol{\xi}_{(\ell)})'\boldsymbol{F}(\boldsymbol{\xi}_{(\ell)}))^{-1}\boldsymbol{F}(\boldsymbol{\xi}_{(\ell)})'(\boldsymbol{y} - \boldsymbol{g}(\boldsymbol{\xi}_{(\ell)})) \quad (108)$$

となる。

12

また、先に示したグレイディエントベクトルが

$$\mathbf{g} = \frac{\partial \text{RSS}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} = -2\mathbf{F}(\boldsymbol{\xi})'(\mathbf{y} - \mathbf{g}(\boldsymbol{\xi})) \quad (109)$$

となるため、Gauss-Newton 法ではパラメタの更新式において

$$\mathbf{d}_{(\ell)} = -\frac{1}{2}(\mathbf{F}'(\boldsymbol{\xi}_{(\ell)})'\mathbf{F}(\boldsymbol{\xi}_{(\ell)}))^{-1}\mathbf{g}(\boldsymbol{\xi}_{(\ell)}) \quad (110)$$

と置いたものになっていることがわかる。

多変量の場合

基準変数が多変量の場合は、

$$\text{RSS} = \|\mathbf{Y} - \mathbf{G}(\boldsymbol{\xi})\|^2 = \sum_{j=1}^p \|\mathbf{y}_{(j)} - \mathbf{g}_{(j)}(\boldsymbol{\xi})\|^2 \quad (111)$$

と書けるため、変数ごとのヤコビアン行列を

$$\mathbf{F}_j(\boldsymbol{\xi}) = \frac{\partial \mathbf{g}_{(j)}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} = \left[\frac{\partial g_{ij}(\boldsymbol{\xi})}{\partial \xi_k} \right], \quad i = 1, 2, \dots, N, \quad k = 1, 2, \dots, Q$$

と定義すれば、グレイディエントならびに更新の方向は

$$\mathbf{g} = \frac{\partial \text{RSS}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} = -2 \sum_{j=1}^p \mathbf{F}_{(j)}(\boldsymbol{\xi})'(\mathbf{y}_{(j)} - \boldsymbol{\theta}_{(j)}(\boldsymbol{\xi})) \quad (112)$$

$$\mathbf{d}_{(\ell)} = -\frac{1}{2} \left(\sum_{j=1}^p \mathbf{F}_{(j)}(\boldsymbol{\xi}_{(\ell)})' \mathbf{F}_{(j)}(\boldsymbol{\xi}_{(\ell)}) \right)^{-1} \mathbf{g}(\boldsymbol{\xi}_{(\ell)}) \quad (113)$$

で与えられる。

今後の話題

- 交互最小2乗解 Alternating Least Squares
 - ◆ 最尤解にも応用可能
- 応用例
 - ◆ 因子分析、多次元尺度法、クラスター分析、ニューラルネット等
- 周辺最尤解