データ解析

Rによる多変量解析入門

最尤法，モデル選択

lec20031120 下平英寿 shimo@is.titech.ac.jp

概要

パラメタの推定と確率モデルの推定の一般論を「尤度原理」で説明する

- **最尤法（尤度）**
- **モデル選択（修正決定係数，赤池情報量規準）**

回帰分析の説明変数選択を例題に取り上げる

最尤法

尤度（likelhood）

- **正規線形回帰モデル**

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \epsilon_i; \quad i = 1, \ldots, n$$

$$\epsilon_1, \ldots, \epsilon_n \sim N(0, \sigma^2) \quad \text{i.i.d.}$$

$$\boldsymbol{\theta} = (\beta_0, \beta_1, \ldots, \beta_p, \sigma^2)$$

- **確率密度関数**

$$f(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi\sigma^2)^n}} \exp\left[-\sum_{i=1}^n \frac{(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} - y_i)^2}{2\sigma^2}\right]$$

- **尤度**

$$L(\boldsymbol{\theta}; \mathcal{X}) = f(\mathcal{X}; \boldsymbol{\theta})$$

最尤法（maximum likelihood method）

- **一般にパラメタベクトルをθ，データを$\mathcal{X}$と書く**

$$L(\boldsymbol{\theta}; \mathcal{X}) = f(\mathcal{X}; \boldsymbol{\theta})$$

- **対数尤度（log-likelihood）**

$$\ell(\boldsymbol{\theta}; \mathcal{X}) = \log L(\boldsymbol{\theta}; \mathcal{X}) = \log f(\mathcal{X}; \boldsymbol{\theta})$$

- **パラメタの最尤推定量(MLE)を$\hat{\theta}$と書く**

$$\max_{\boldsymbol{\theta} \in \Theta} \ell(\boldsymbol{\theta}; \mathcal{X}) = \ell(\hat{\boldsymbol{\theta}}; \mathcal{X})$$

$$\left.\frac{\partial \ell(\boldsymbol{\theta}; \mathcal{X})}{\partial \boldsymbol{\theta}}\right|_{\hat{\boldsymbol{\theta}}} = 0$$

重回帰モデルの最尤推定

$$\ell(\boldsymbol{\beta}, \sigma^2; \boldsymbol{x}, \boldsymbol{y}) = -\frac{n}{2}\log(2\pi\sigma^2) - \sum_{i=1}^n \frac{(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} - y_i)^2}{2\sigma^2}$$

$$\frac{\partial \ell}{\partial \beta_k} = \sum_{i=1}^n \frac{(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} - y_i)\, x_{ik}}{\sigma^2}$$

$$\frac{\partial \ell}{\partial (\sigma^2)} = -\frac{n}{2\sigma^2} + \sum_{i=1}^n \frac{(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} - y_i)^2}{2\sigma^4}$$

最尤推定量 $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}, \quad \hat{\sigma}^2 = \dfrac{\|\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\|^2}{n}$

c.f. σ^2**の不偏推定量** $\tilde{\sigma}^2 = \dfrac{\|\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\|^2}{n-p-1}, \quad E(\tilde{\sigma}^2) = \sigma^2$

ここでは最尤推定は$\hat{\sigma}^2$，不偏推定は$\tilde{\sigma}^2$ と書いて区別する

モデル選択

モデルの候補

確率モデルの候補が複数ある場合

$$f_1(\mathcal{X}; \boldsymbol{\theta}_1), \ldots, f_m(\mathcal{X}; \boldsymbol{\theta}_m)$$

データ$\mathcal{X}$を最もよく説明するモデルを選ぶ

[例] 重回帰分析の説明変数の選択
f_1**は定数項**β_0**のみ（**$\beta_1 = \cdots = \beta_p = 0$**を仮定）**
f_2**は**x_1**を使う（**$\beta_2 = \cdots = \beta_p = 0$**）**
f_3**は**x_2**を使う（**$\beta_1 = \beta_3 = \cdots = \beta_p = 0$**）**
f_4**は**x_1**と**x_2**を使う（**$\beta_3 = \cdots = \beta_p = 0$**）**
f_m**は**$x_1, \ldots, x_p$**を使う**

[例] 多項式回帰の次数選択
f_1**は**$y = \beta_0 + \epsilon$**，**f_2**は**$y = \beta_0 + \beta_1 x + \epsilon$**，**
f_3**は**$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon$**，**
f_{p+1}**は**$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_p x^p + \epsilon$

モデル選択

- **モデルの良さを表す規準を$C(f_i, \mathcal{X})$と書く**

$$C(f_1, \mathcal{X}), \ldots, C(f_m, \mathcal{X})$$

を最大（または最小）にするモデルf_iを選択する

- **様々な規準$C(f_i, \mathcal{X})$が提案されている**

[例] 修正決定係数，赤池情報量規準

修正決定係数

- 回帰分析の決定係数 R^2
  「観測データへのモデルの当てはまりの良さ」を測る
- 決定係数の問題点
  モデルが大きくなるほど大きな値を取る（多項式回帰では最大次数のモデルが R^2 を最大にする）
- 説明変数の数 p が増えるに従い推定誤差が増える影響を考慮して決定係数を修正する
- 修正決定係数 $\bar{R}^2$ を最大にするモデルを選ぶ

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1}(1-R^2)$$

$$n = \text{サンプルサイズ}, \quad k = \text{変数の数}$$

修正決定係数の導出

- モデル (k) の決定係数

$$R^2 = 1 - \frac{\hat{\sigma}_k^2}{\hat{\sigma}_0^2} = 1 - \frac{RSS^{(k)}/n}{RSS^{(0)}/n}$$

ただしモデル (k) における σ^2 の最尤推定を $\hat{\sigma}_k^2$ と書いた．

$$RSS^{(k)} = \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad RSS^{(0)} = \sum_{i=1}^n (y_i - \bar{y}_i)^2$$

- 最尤推定を不偏推定で置き換えて決定係数を修正する．

$$\bar{R}^2 = 1 - \frac{\tilde{\sigma}_k^2}{\tilde{\sigma}_0^2} = 1 - \frac{RSS^{(k)}/(n-k-1)}{RSS^{(0)}/(n-1)}$$

ただしモデル (k) における σ^2 の不偏推定を $\tilde{\sigma}_k^2$ と書いた．

- $\bar{R}^2$ 最大化は $\tilde{\sigma}_k^2$ 最小化と同じこと．

赤池情報量規準（Akaike Information Criterion）

- モデル選択の「最尤法」
  （回帰分析に限らずさまざまなモデルに適用可能）
- モデル候補 $1,\ldots,m$ の最大対数尤度

$$\hat{\ell}_i = \log f_i(\mathcal{X};\hat{\boldsymbol{\theta}}_i), \quad i = 1,\ldots,m$$

- パラメタ数 $\dim\theta_i$ の大きいモデルが $\hat{\ell}_i$ を大きくする
- パラメタ数の増加に伴い推定誤差が増える影響を考慮して $\hat{\ell}$ を修正する

$$\mathrm{AIC} = -2 \times (\hat{\ell} - \dim\boldsymbol{\theta})$$

AIC を最小にするモデルを選ぶ

- 回帰分析の場合，最大対数尤度とパラメタ数は

$$\hat{\ell} = -\frac{n}{2}\left\{1 + \log(2\pi\hat{\sigma}_k^2)\right\}, \quad \dim\boldsymbol{\theta} = k+2$$

重回帰分析のモデル選択

セッションファイル
edu:~shimo/class/gakubu200209/note20031010.Rt

```
> ### 重回帰分析の説明変数選択
> source("myfunc20031010.R")
> ax <- c("E09504","A0410302","C01301","B02101"); x <- X2000$x[,ax]
> ay <- "A05203"; y <- X2000$x[,ay]
> X2000$jitem[c(ax,ay)]
                              E09504                              A0410302
"最終学歴が大学・大学院卒の者の割合 " "未  婚  者  割  合［20～24歳・女］ "
                              C01301                                B02101
          "県民１人当たり県民所得  "                "年  平  均  気  温  "
                              A05203
            "合 計 特 殊 出 生 率  "
> ff <- mylsfitv(x,y)  # すべての変数のくみあわせ (16通り)
> ff$v  #  使った変数のリスト
         1     2     3     4
()     FALSE FALSE FALSE FALSE
```



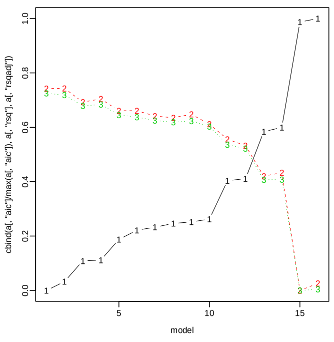
```
(1)     TRUE FALSE FALSE FALSE
(2)    FALSE  TRUE FALSE FALSE
(12)    TRUE  TRUE FALSE FALSE
(3)    FALSE FALSE  TRUE FALSE
(13)    TRUE FALSE  TRUE FALSE
(23)   FALSE  TRUE  TRUE FALSE
(123)   TRUE  TRUE  TRUE FALSE
(4)    FALSE FALSE FALSE  TRUE
(14)    TRUE FALSE FALSE  TRUE
(24)   FALSE  TRUE FALSE  TRUE
(124)   TRUE  TRUE FALSE  TRUE
(34)   FALSE FALSE  TRUE  TRUE
(134)   TRUE FALSE  TRUE  TRUE
(234)  FALSE  TRUE  TRUE  TRUE
(1234)  TRUE  TRUE  TRUE  TRUE
> ff$p # 変数の数
   ()    (1)    (2)   (12)    (3)   (13)   (23)  (123)    (4)   (14)   (24)
    0      1      1      2      1      2      2      3      1      2      2
 (124)   (34)  (134)  (234) (1234)
```

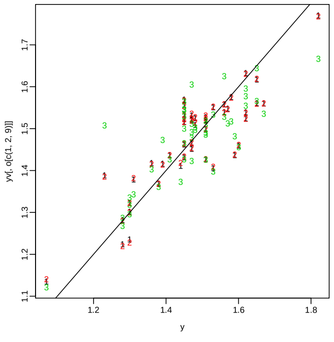
```
     3       2       3       3       4
> round(ff$coef,3) #  推定した回帰係数
       Intercept      1       2 3     4
()         1.473  0.000   0.000 0 0.000
(1)        1.742 -0.028   0.000 0 0.000
(2)        5.099  0.000  -0.042 0 0.000
(12)       4.175 -0.010  -0.030 0 0.000
(3)        2.132  0.000   0.000 0 0.000
(13)       1.914 -0.022   0.000 0 0.000
(23)       4.624  0.000  -0.034 0 0.000
(123)      4.189 -0.006  -0.029 0 0.000
(4)        1.329  0.000   0.000 0 0.009
(14)       1.447 -0.032   0.000 0 0.021
(24)       5.047  0.000  -0.044 0 0.017
(124)      3.653 -0.016  -0.027 0 0.020
(34)       2.021  0.000   0.000 0 0.007
(134)      1.459 -0.031   0.000 0 0.021
(234)      4.745  0.000  -0.039 0 0.015
(1234)     3.636 -0.017  -0.027 0 0.020
```

```
> round(ff$aic,2) # 情報量規準AIC
     ()     (1)     (2)    (12)     (3)    (13)    (23)   (123)     (4)    (1
 -53.17  -86.90  -95.61  -96.55  -76.78  -87.34  -97.35  -96.22  -52.41  -99
   (24)   (124)    (34)   (134)   (234)  (1234)
-104.61 -110.97  -75.86  -98.00 -104.43 -109.05
> o <- order(ff$aic) # AICの小さい順序
> o
 [1] 12 16 11 15 10 14  7  4  8  3  6  2  5 13  1  9
> round(ff$aic[o] - min(ff$aic),2) # AIC最小モデルからの差
 (124) (1234)   (24)  (234)   (14)  (134)   (23)   (12)  (123)    (2)   (13)
  0.00   1.92   6.36   6.54  10.98  12.97  13.62  14.42  14.75  15.36  23.63
   (1)    (3)   (34)     ()    (4)
 24.07  34.19  35.11  57.80  58.56
> round(ff$rsq[o],2) # 決定係数
 (124) (1234)   (24)  (234)   (14)  (134)   (23)   (12)  (123)    (2)   (13)
  0.74   0.74   0.69   0.70   0.66   0.66   0.64   0.64   0.65   0.61   0.56
   (1)    (3)   (34)     ()    (4)
  0.53   0.42   0.43   0.00   0.03
> round(ff$rsqadj[o],2) # 修正決定係数
```

```
 (124) (1234)   (24)  (234)   (14)  (134)   (23)   (12)  (123)    (2)   (13)
  0.72   0.72   0.68   0.68   0.65   0.64   0.62   0.62   0.62   0.60   0.54
   (1)    (3)   (34)     ()    (4)
  0.52   0.41   0.41   0.00   0.00
> a <- mylsfitv.table(ff)   # 上記の結果をまとめた表
> round(a,3)
             aic       rsq     rsqadj
(124)    1  0.000  2 0.743  1  0.725
(1234)   2  1.921  1 0.743  2  0.719
(24)     3  6.357  4 0.693  4  0.679
(234)    4  6.539  3 0.704  3  0.684
(14)     5 10.983  6 0.661  5  0.645
(134)    6 12.974  5 0.661  6  0.637
(23)     7 13.625  8 0.641  7  0.625
(12)     8 14.423  9 0.635  9  0.618
(123)    9 14.753  7 0.648  8  0.623
(2)     10 15.361 10 0.612 10  0.603
(13)    11 23.629 11 0.556 11  0.536
(1)     12 24.073 12 0.532 12  0.522
```

```
(3)     13 34.188 14 0.420 14  0.407
(34)    14 35.110 13 0.433 13  0.408
()      15 57.801 16 0.000 16  0.000
(4)     16 58.557 15 0.026 15  0.004
> a[,"aic"]/max(a[,"aic"])  # ちょうど０～１の範囲に収まるようにする.
      (124)      (1234)        (24)       (234)        (14)       (134)        (23)
0.00000000 0.03281278 0.10856896 0.11166437 0.18755748 0.22156638 0.23267936
       (12)       (123)         (2)        (13)         (1)         (3)        (34)
0.24630189 0.25193588 0.26232587 0.40353000 0.41111183 0.58383451 0.59959282
         ()         (4)
0.98708914 1.00000000
> matplot(cbind(a[,"aic"]/max(a[,"aic"]),a[,"rsq"],a[,"rsqadj"]),type="b",xla
> yv <- apply(ff$coef,1,function(beta) cbind(1,x) %*% beta) # 予測値
> matplot(y,yv[,o[c(1,2,9)]])
> abline(0,1)
```



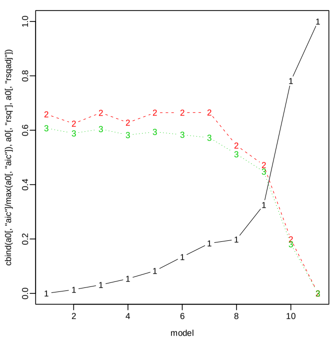


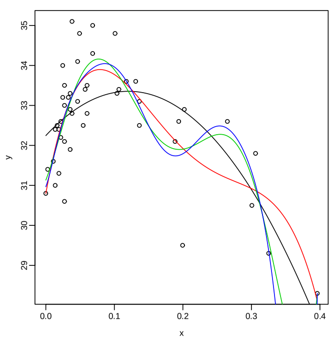
多項式回帰の次数選択

```
> ## 次数1,..,10で回帰分析
> ax <- "B02304"; ay <- "B02102"
> X2000$jitem[c(ax,ay)]
                                        B02304
                   "雪    日    数（年  間）  "
                                        B02102
"最  高  気  温（日最高気温の月平均の最高値）  "
> x <- X2000$x[,ax]/366  # 雪が降る頻度にしておく
> y <- X2000$x[,ay]
> m0 <- 10 # 10次回帰までしらべる
> xx0 <- apply(as.matrix(1:m0),1,function(i) x^i) # データ行列
> dimnames(xx0)[[2]] <- paste("X",1:m0,sep="")
> v0 <- matrix(F,m0+1,m0) # 変数の組み合わせを指定する場所
> for(i in 1:m0) { v0[i+1,1:i] <- T }
> dimnames(v0) <- list(paste("deg",0:m0,sep=""),dimnames(xx0)[[2]])
> v0
         X1    X2    X3    X4    X5    X6    X7    X8    X9   X10
```



```
deg0  FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
deg1   TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
deg2   TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
deg3   TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
deg4   TRUE  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE
deg5   TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE
deg6   TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE
deg7   TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE FALSE
deg8   TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE
deg9   TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE
deg10  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
> ff0 <- mylsfitv(xx0,y,v0)
> ff0$p # 変数の数
 deg0  deg1  deg2  deg3  deg4  deg5  deg6  deg7  deg8  deg9 deg10
    0     1     2     3     4     5     6     7     8     9    10
> round(ff0$coef,1) #  推定した回帰係数
      Intercept    X1      X2        X3        X4         X5          X6
deg0       32.6   0.0     0.0       0.0       0.0        0.0         0.0
deg1       33.2  -6.6     0.0       0.0       0.0        0.0         0.0
deg2       32.2  18.4   -76.3       0.0       0.0        0.0         0.0
deg3       31.6  44.9  -274.9     365.3       0.0        0.0         0.0
deg4       30.8  94.3  -918.4    3092.5   -3579.6        0.0         0.0
deg5       30.6 110.8 -1274.6    5837.6  -12078.9     9046.4         0.0
deg6       31.1  43.6   910.1  -20998.8  135541.7  -358887.2    337864.1
deg7       30.9  84.0  -808.9    7724.7  -94117.8   579773.2  -1551290.9
deg8       30.9  90.1 -1175.9   16609.0 -198963.7  1240524.4  -3819285.7
deg9       30.9  76.7  -190.0  -12454.7  230323.8 -2302948.9  13266130.5
deg10      31.0  44.3  2848.9 -127348.3 2457603.9 -27018280.4 178824766.5
             X7         X8           X9          X10
deg0          0          0            0            0
deg1          0          0            0            0
deg2          0          0            0            0
deg3          0          0            0            0
deg4          0          0            0            0
deg5          0          0            0            0
deg6          0          0            0            0
deg7    1481261          0            0            0
deg8    5473821   -2817197            0            0
deg9   -42275331   68749831   -44439762            0
deg10 -722024954 1739416842 -2297848728 1281301538
> a0 <- mylsfitv.table(ff0)  # ＡＩＣの小さい順序に並べ替え
> a0
                 aic            rsq          rsqadj
deg6    1  0.0000000  5 0.6588093  1 0.6076307
deg4    2  0.5266367  7 0.6243143  4 0.5885347
deg7    3  1.2129178  4 0.6644755  2 0.6042531
deg5    4  2.0765810  6 0.6278945  6 0.5825158
deg8    5  3.2002414  3 0.6645659  3 0.5939482
deg9    6  5.1707424  2 0.6647764  5 0.5832355
deg10   7  7.0907457  1 0.6653465  7 0.5723872
deg3    8  7.6532499  8 0.5437978  8 0.5119697
deg2    9 12.5295707  9 0.4719236  9 0.4479201
deg1   10 30.1167515 10 0.1988971 10 0.1810948
deg0   11 38.5397450 11 0.0000000 11 0.0000000
> matplot(cbind(a0[,"aic"]/max(a0[,"aic"]),a0[,"rsq"],a0[,"rsqadj"]),type="b"
> psinit("20031010k2.eps",pty="s")
> matplot(cbind(a0[,"aic"]/max(a0[,"aic"]),a0[,"rsq"],a0[,"rsqadj"]),type="b"
> dev.off()
X11
  2
> ##
> xs <- seq(min(x),max(x),length=300) # グラフを書くためにxの範囲を３００等
分
> xx1 <- apply(as.matrix(1:m0),1,function(i) xs^i) # データ行列
> xx1 <- cbind(1,xx1) # 定数項の列を加える
> a <- c(2,4,6,10) # この次数の曲線だけ描く
> plot(x,y) # 散布図
> for(i in seq(along=a)) lines(xs,xx1 %*% ff0$coef[a[i]+1,],col=i)
> psinit("20031010p2.eps",pty="s")
> plot(x,y) # 散布図
> for(i in seq(along=a)) lines(xs,xx1 %*% ff0$coef[a[i]+1,],col=i)
> dev.off()
X11
  2
```





演習問題

4-1. `mylsfitv`**は残差平方和**$RSS=\sum_{i=1}^{n} e_i^2$**を各モデル毎に計算したベクトルを**`rss`**として返す．これを利用して，各モデルの不偏分散**$\tilde{\sigma}^2$**と AIC を計算する関数**`modelcrit`**を作る．**

```
ff <- mylsfitv(x,y,v) # 例題を参考にして適当にx,y,vを決める．
ff$n # サンプルサイズ
ff$p # 説明変数の数のベクトル
ff$rss # RSSのベクトル
modelcrit <- function(ff) {
   # ff$rss, ff$n, ff$rssをつかって不偏分散とAICを計算
 sqe <- ....  # 不偏分散の計算
 aic <- ....  # AICの計算
  return(sqe,aic)
}
```

4-2. `note20031010.Rt`**の「重回帰分析の説明変数選択」とおなじ数値例に**`modelcrit`**を適用し，各モデルの**$\tilde{\sigma}^2$**と AIC を計算せよ．**`note20031010.Rt`

で計算したAICと`modelcrit`**で計算したAICが等しいことを確認せよ．また，**$\tilde{\sigma}^2$**の小さい順が**$\bar{R}^2$**の大きい順と等しいことを確認せよ．**

4-3. 情報量規準BICは次式で定義される．

$$\mathrm{BIC} = -2\times\hat{\ell} + (\log n)\times \dim\boldsymbol{\theta}$$

これはAICの定義とは第２項が異なっている．`modelcrit`**を修正して，BICも計算するようにせよ．**

```
modelcrit <- function(ff) {
 sqe <- ....  # 不偏分散の計算
 aic <- ....  # AICの計算
 bic <- ....  # BICの計算
  return(sqe,aic,bic)
}
```

4-4. `note20031010.Rt`**の「重回帰分析」と「多項式回帰」の二つの例について，**`modelcrit`**を用いて**$\bar{R}^2$**，AIC，BICのそれぞれの規準でモデル選択を行い，選択されたモデルとそのときの規準値を示せ．**

4-5．`X2000$x`から適当な項目を選んで重回帰分析をする．ただし説明変数は４個以上とする．4-4と同様に３つの規準でモデル選択を行い，それらの結果を比較せよ．