

クレジット:

UTokyo Online Education 数理手法Ⅶ 2019 北川源四郎

ライセンス:

利用者は、本講義資料を、教育的な目的に限ってページ単位で利用することができます。特に記載のない限り、本講義資料はページ単位でクリエイティブ・コモンズ 表示-非営利-改変禁止 ライセンスの下に提供されています。

<http://creativecommons.org/licenses/by-nc-nd/4.0/>

本講義資料内には、東京大学が第三者より許諾を得て利用している画像等や、各種ライセンスによって提供されている画像等が含まれています。個々の画像等を本講義資料から切り離して利用することはできません。個々の画像等の利用については、それぞれの権利者の定めるところに従ってください。

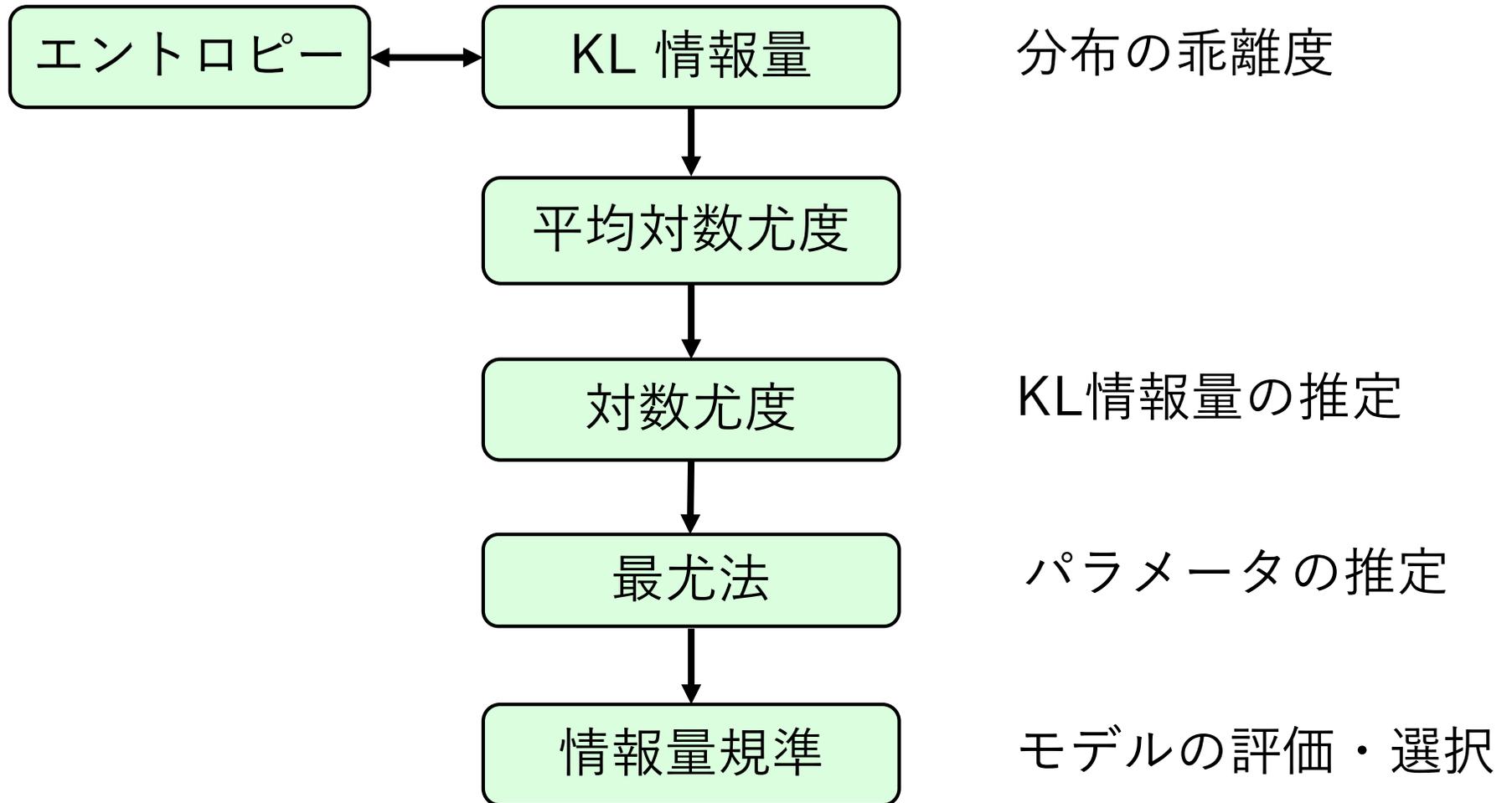


時系列解析（４）

－ モデルの推定・選択 －

東京大学 数理・情報教育研究センター
北川 源四郎

前回の復習



概要

1. AICによるモデル選択例
 - (1) ヒストグラムのBin数の選択
 - (2) 分布の形状の選択
 - (3) Box-Cox変換の選択
 - (4) 分布の同一性
2. 重回帰モデル, 多項式回帰モデル
3. 最小二乗法, Householder法
4. 部分回帰モデル

ヒストグラムのビン数の決定

$$P(\{n_j\} | \{p_j\}) = \frac{n!}{n_1! \cdots n_k!} p_1^{n_1} \cdots p_k^{n_k}$$

$$\ell(p_1, \dots, p_k) = C + \sum_{j=1}^k n_j \log p_j$$

$$\hat{p}_j = \frac{n_j}{n}$$

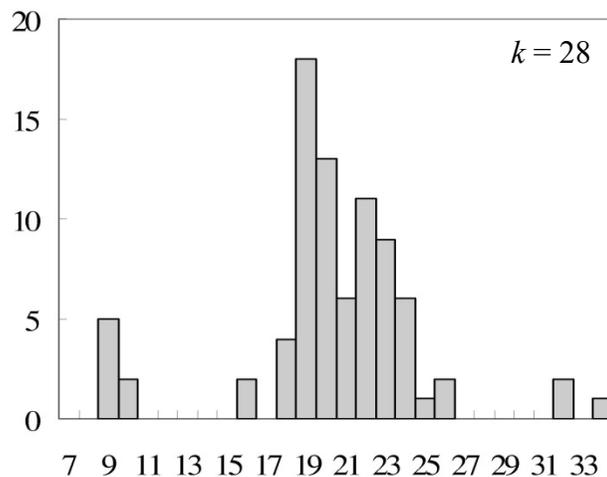
$$\text{AIC}_k = (-2) \left\{ C + \sum_{j=1}^k n_j \log \left(\frac{n_j}{n} \right) \right\} + 2(k-1)$$

銀河データ (Roeder (1990))

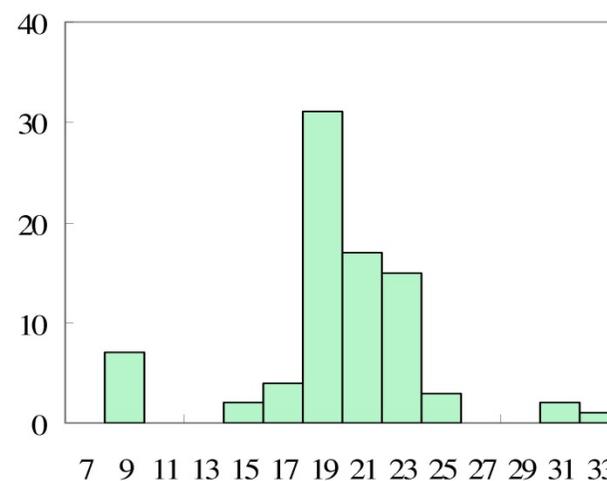
0	5	2	0	0	0	0	0	2	0	4	18	13	6
11	9	6	1	2	0	0	0	0	0	2	0	1	0

Bin Size	log-LK	AIC
28	-189.19	432.38
14	-197.72	421.43
7	-209.52	431.03

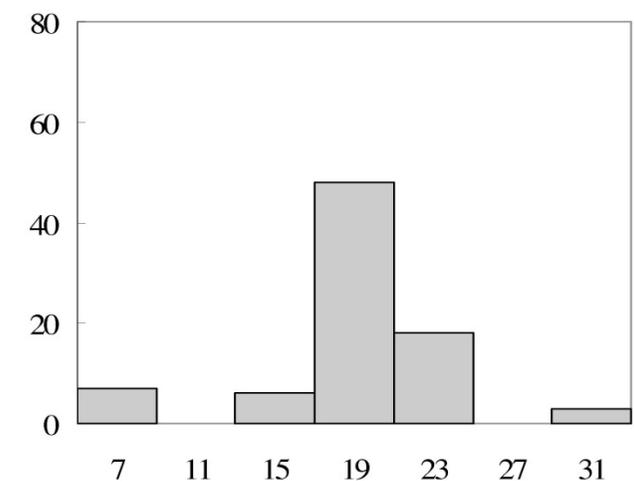
銀河データのヒストグラム



AIC最小

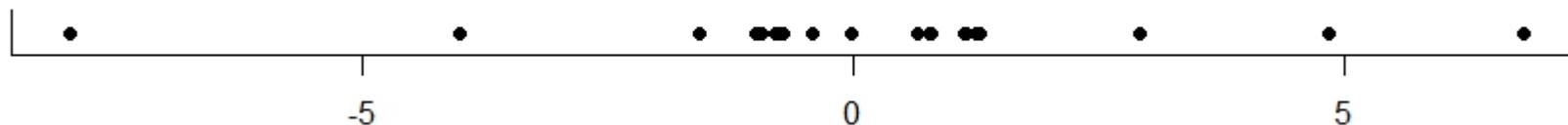


少なすぎ



モデル選択例：分布の形状の選択

-7.99	-4.01	-1.56	-0.99	-0.93	-0.80	-0.77	-0.71	-0.42	-0.02
0.65	0.78	0.80	1.14	1.15	1.24	1.29	2.81	4.84	6.82



Pearson 分布族 (type IV)

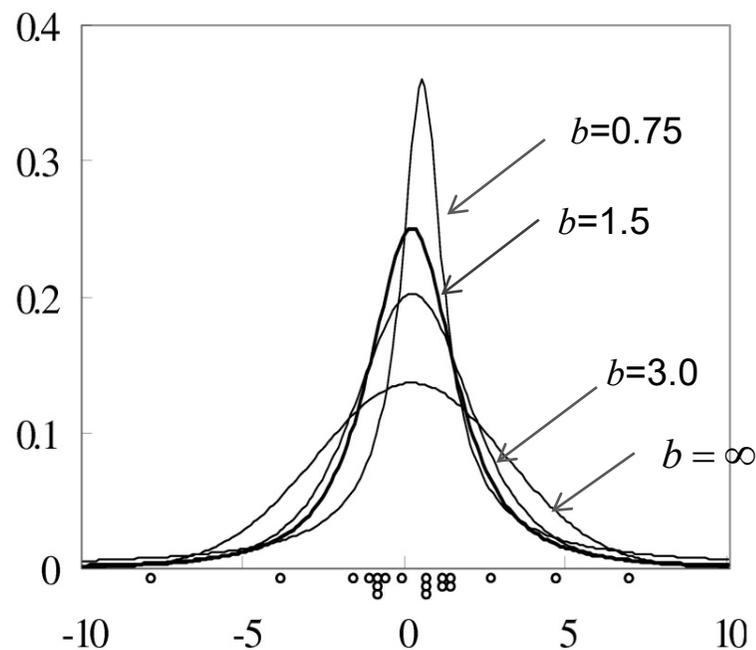
$$f(y | \mu, \tau^2, b) = \frac{C}{((y - \mu)^2 + \tau^2)^b}$$

$$C = \tau^{2b-1} \Gamma(b) / \Gamma(b - \frac{1}{2}) \Gamma(\frac{1}{2})$$

b : 形状パラメータの選択

- 1 : Cauchy分布
- $(k+1)/2$: 自由度 k の t 分布
- ∞ : 正規分布

応用：非ガウス型平滑化



AICによる分布の形状の選択

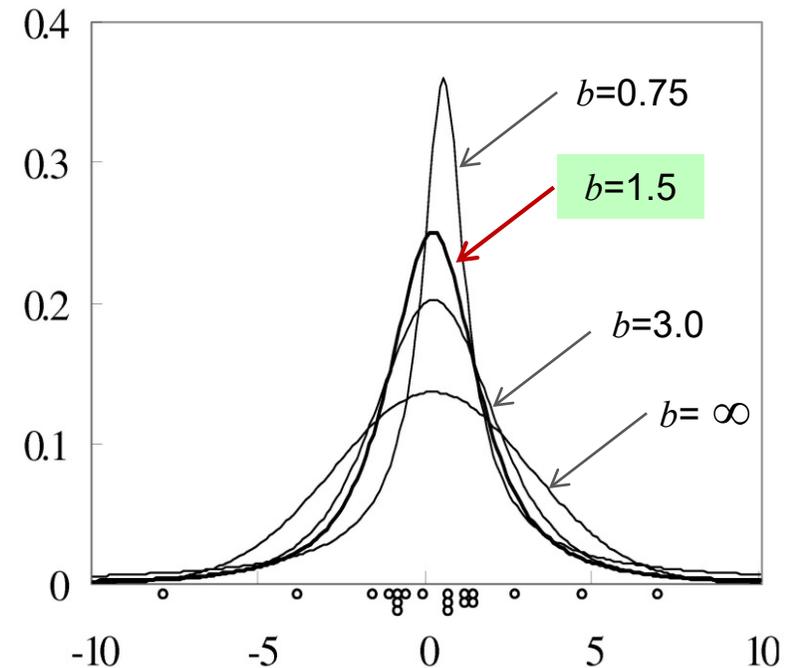
$$f(y | \mu, \tau^2, b) = \frac{C}{((y - \mu)^2 + \tau^2)^b}$$

$$\ell(\mu, \tau^2, b) = \sum_{n=1}^N \log f(y_n | \mu, \tau^2, b)$$

$$= N \left\{ (b - \frac{1}{2}) \log \tau^2 + \log \Gamma(b) - \log(b - \frac{1}{2}) - \log \Gamma(\frac{1}{2}) \right\} - b \sum_{n=1}^N \log \left\{ (y_n - \mu)^2 + \tau^2 \right\}$$

b によるパラメータとAICの変化

b	μ	τ^2	Log-L	AIC
0.60	0.801	0.030	-58.84	121.69
0.75	0.506	0.431	-51.40	106.79
1.00	0.189	1.380	-47.87	99.73
1.50	0.185	4.152	-47.07	98.14
2.00	0.201	8.395	-47.43	98.86
2.50	0.214	13.87	-47.82	99.63
3.00	0.222	20.21	-48.12	100.25
∞	0.166	8.545	-49.83	103.66



モデル選択例：Box-Cox変換

Box-Cox 変換

$$z_n = \begin{cases} \lambda^{-1}(y_n^\lambda - 1) & \text{for } \lambda \neq 0 \\ \log y_n & \text{for } \lambda = 0 \end{cases}$$

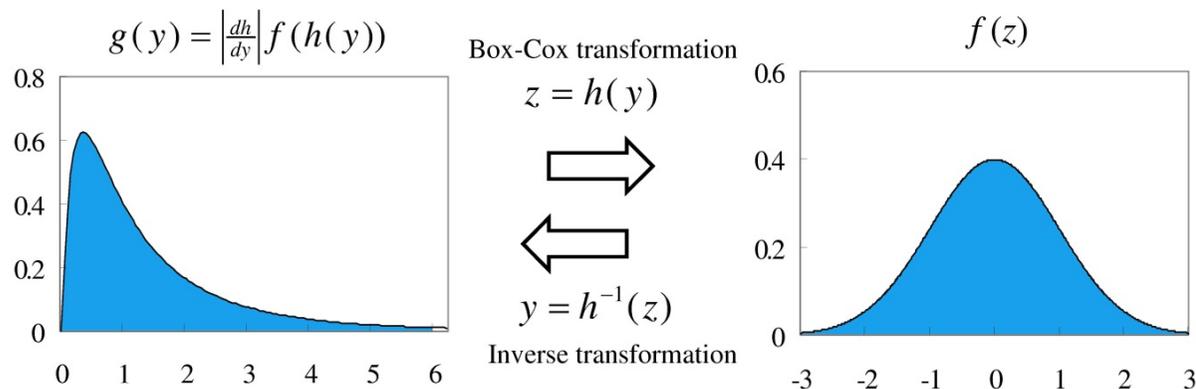
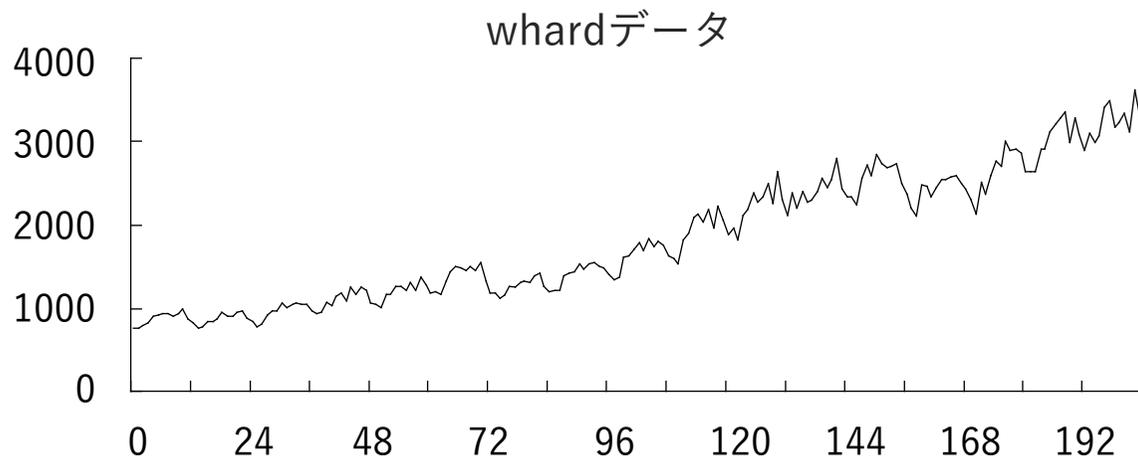
- 変換後のモデルのAICではモデル比較はできない

$$g(y) = \left| \frac{dh_\lambda}{dy} \right| f(h(y))$$

$$\text{AIC}'_z = \text{AIC}_z - 2 \log \left| \frac{dh_\lambda}{dy} \right|$$

- 原データに関するAIC

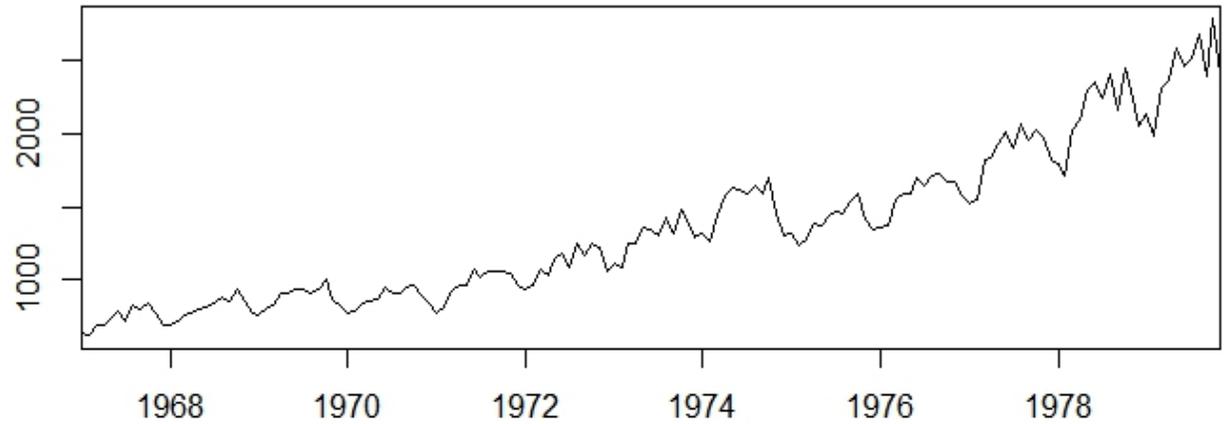
➡ 比較可能



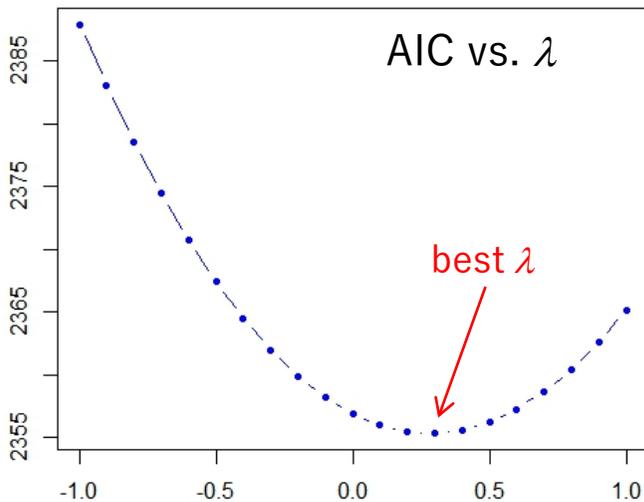
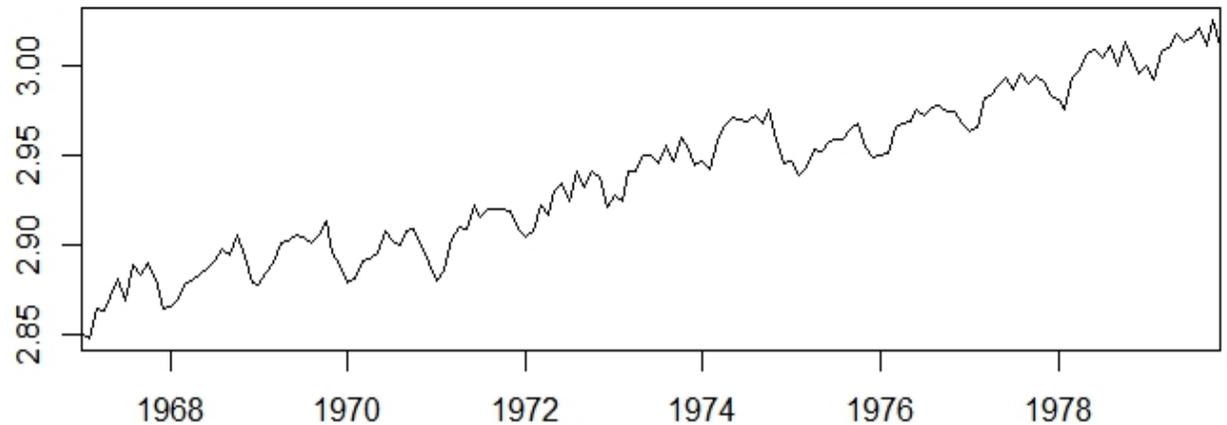
Box-Cox変換の選択

λ	対数尤度	AIC	AIC'
1.0	-1191	2388	2388
0.8	-1187	1936	2379
0.6	-1183	1486	2371
0.4	-1180	1038	2365
0.3	-1179	814	2362
0.2	-1178	591	3273
0.0	-1176	146	2357
-0.2	-1176	-298	2355
-0.4	-1176	-740	2356
-0.6	-1177	-1181	2357
-0.8	-1178	-1620	2360
-1.0	-1181	-2058	2365

WHARD 原データ($\lambda=1$)

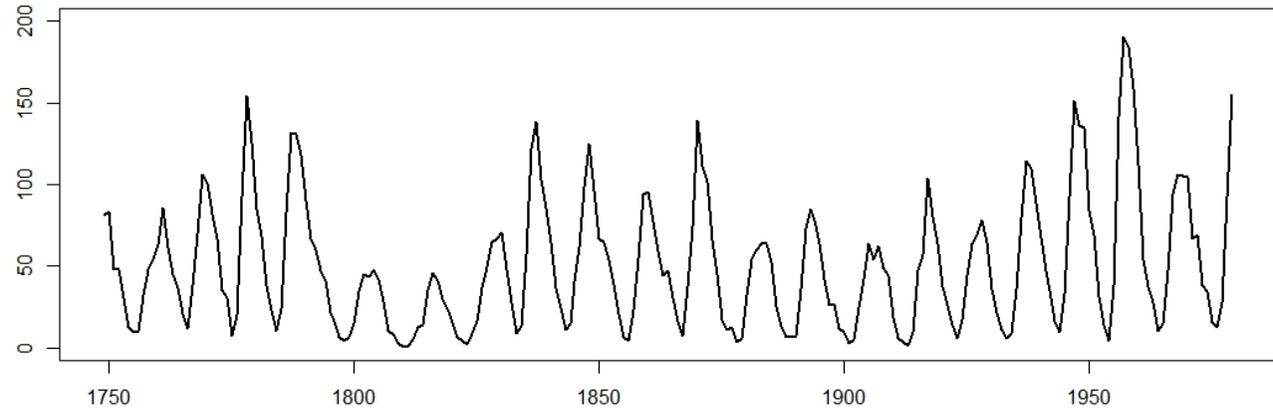


transformed data AIC最小($\lambda=0.3$)

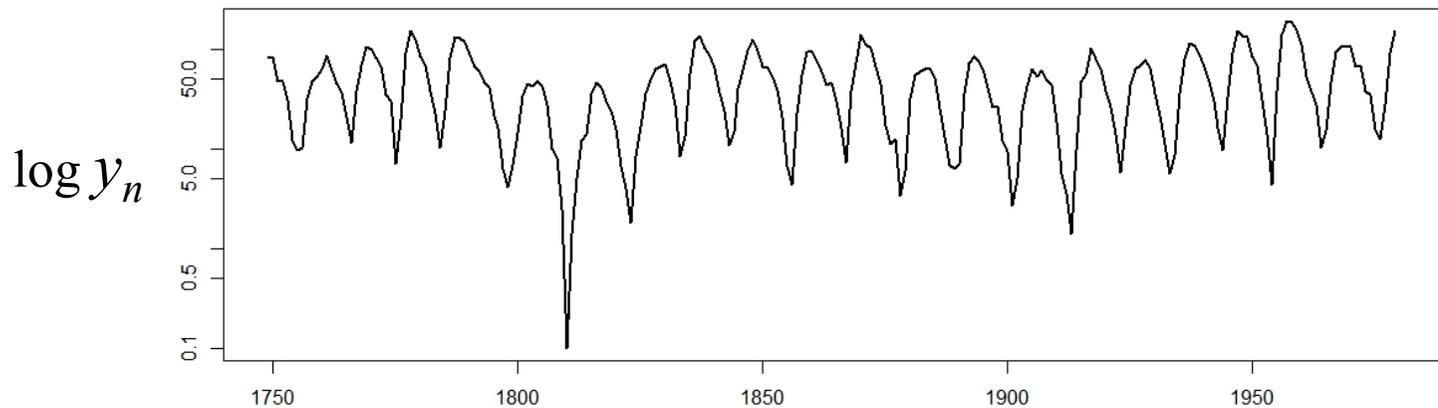


太陽黒点数データ

```
data(Sunspot)
plot(Sunspot,ylim=c(0,200))
plot(Sunspot,log="y")
```



- 太陽黒点数の原データは上に尖っている



- 対数変換すると下に尖る

AIC'による変換パラメータの選択

data(Sunspot) # Sun spot number data

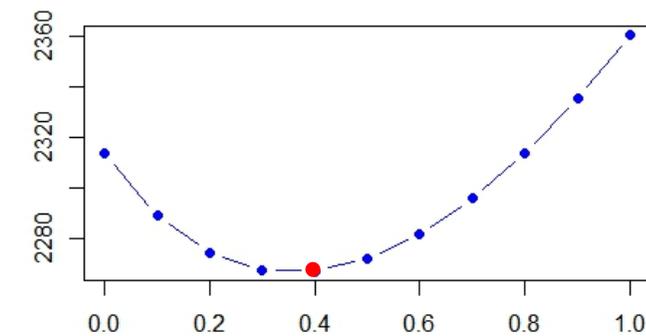
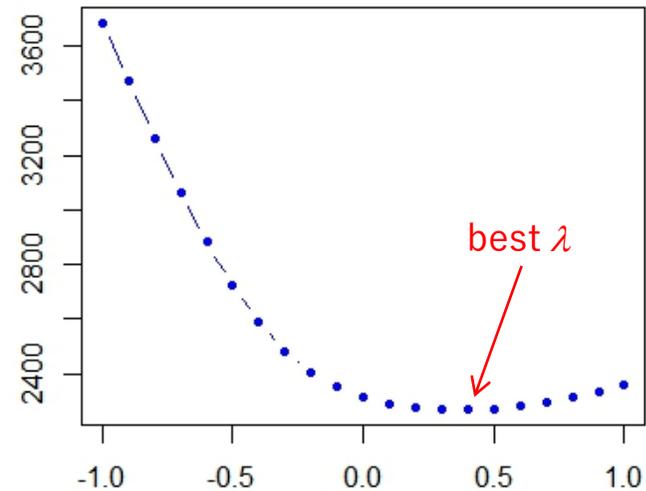
boxcox(Sunspot)

lambda	aic'	LL'	aic	LL	mean	variance
1.00	2360.26	-1178.13	2360.26	-1178.13	4.909502e+01	1.575552e+03
0.90	2335.22	-1165.61	2174.47	-1085.24	3.545844e+01	7.049401e+02
0.80	2313.48	-1154.74	1991.98	-993.99	2.591126e+01	3.199262e+02
0.70	2295.33	-1145.66	1813.07	-904.54	1.917397e+01	1.474669e+02
0.60	2281.11	-1138.56	1638.11	-817.05	1.437922e+01	6.914276e+01
0.50	2271.26	-1133.63	1467.50	-731.75	1.093610e+01	3.303737e+01
0.40	2266.32	-1131.16	1301.81	-648.91	8.439901e+00	1.612487e+01
0.30	2267.05	-1131.52	1141.79	-568.90	6.611858e+00	8.065706e+00
0.20	2274.59	-1135.29	988.58	-492.29	5.258840e+00	4.155209e+00
0.10	2290.79	-1143.40	844.03	-420.01	4.246205e+00	2.222464e+00
0.00	2318.78	-1157.39	711.27	-353.63	3.479466e+00	1.250918e+00
-0.10	2363.66	-1179.83	595.39	-295.70	2.891856e+00	7.574966e-01
-0.20	2432.86	-1214.43	503.84	-249.92	2.435839e+00	5.096385e-01
-0.30	2534.61	-1265.31	444.85	-220.42	2.077302e+00	3.947690e-01
-0.40	2673.75	-1334.88	423.23	-209.62	1.791544e+00	3.595107e-01
-0.50	2848.16	-1422.08	436.89	-216.45	1.560501e+00	3.814048e-01
-0.60	3050.32	-1523.16	478.30	-237.15	1.370809e+00	4.562814e-01
-0.70	3271.90	-1633.95	539.12	-267.56	1.212437e+00	5.937308e-01
-0.80	3506.54	-1751.27	613.01	-304.51	1.077716e+00	8.175441e-01
-0.90	3750.16	-1873.08	695.88	-345.94	9.606427e-01	1.170321e+00
-1.00	4000.25	-1998.13	785.23	-390.61	8.563591e-01	1.722986e+00

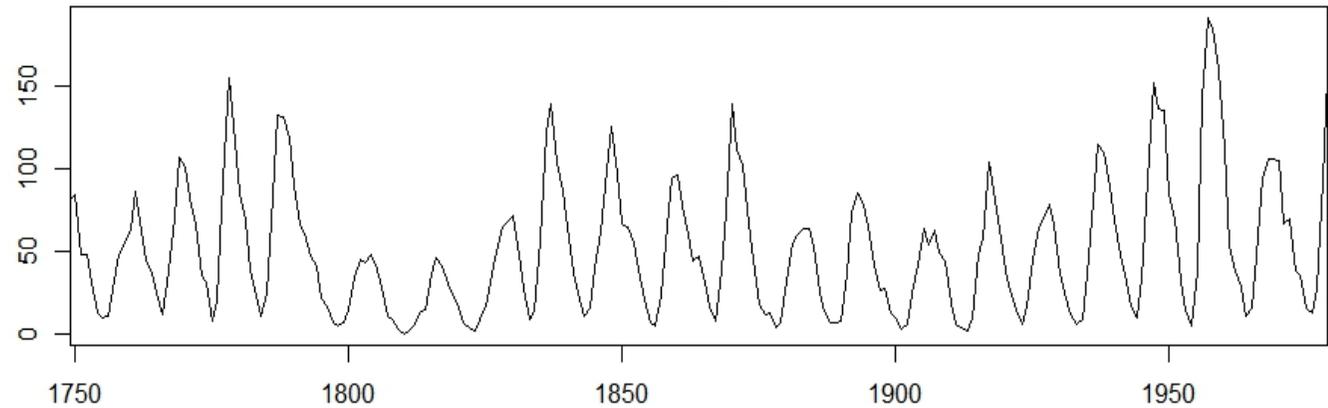
lambda = 0.40 AIC' minimum = 2266.32

太陽黒点数データのBox-Cox変換

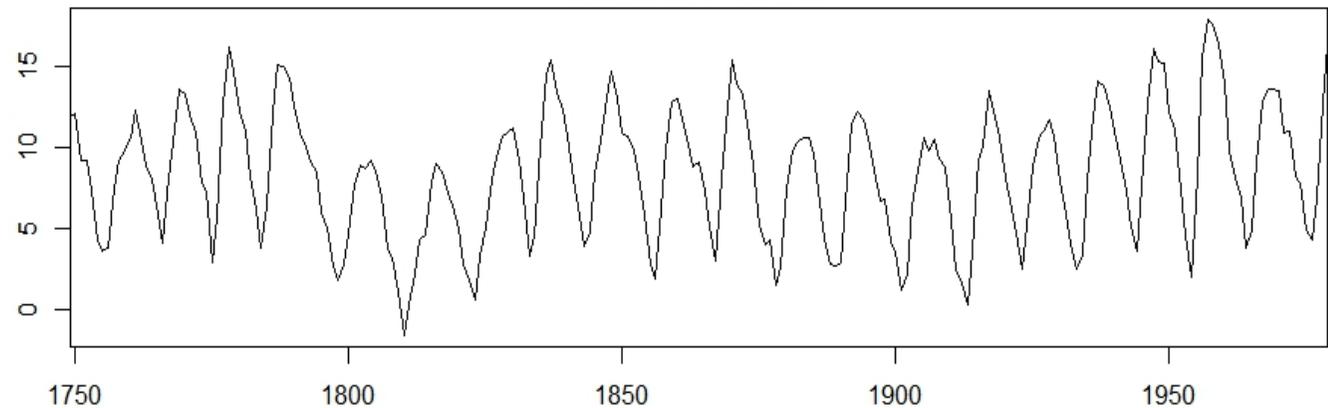
AIC'



Sunspot



transformed data $\lambda = 0.4$



```
data(Sunspot) # Sun spot number data
boxcox(Sunspot)
# AIC'のプロット
z <- boxcox(Sunspot)
x <- seq(1,-1,length=21)
plot(x,z$aic.z,col="blue",type="b",pch=19)
```

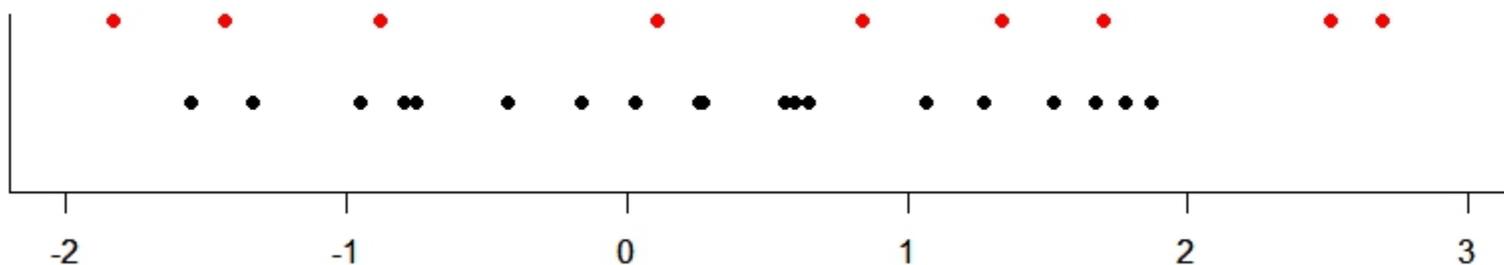
分布の同一性

Data Set 1

0.26 -1.33 1.07 1.78 -0.16 0.03 -0.79 -1.55 1.27 0.56
-0.95 0.60 0.27 1.67 0.60 -0.42 1.87 0.65 -0.75 1.52

Data Set 2

1.70 0.84 1.34 0.11 -0.88 -1.43 3.52 2.69 2.51 -1.83



$$f(y | \mu_1, \tau_1^2), \quad y \in \text{Data Set 1}$$

$$f(y | \mu_2, \tau_2^2), \quad y \in \text{Data Set 2}$$

(1) 制約なし

$$(2) \sigma_1^2 = \sigma_2^2$$

$$(3) \mu_1 = \mu_2$$

$$(4) \mu_1 = \mu_2, \sigma_1^2 = \sigma_2^2$$

分布の同一性

制約	ℓ	AIC	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$
なし	-48.411	104.823	0.310	0.857	1.033	3.015
$\hat{\sigma}_1^2 = \hat{\sigma}_2^2$	-50.473	106.946	0.310	0.857	1.694	
$\hat{\mu}_1 = \hat{\mu}_2$	-48.852	103.703	0.438		1.049	3.191
$\hat{\mu}_1 = \hat{\mu}_2, \hat{\sigma}_1^2 = \hat{\sigma}_2^2$	-51.050	106.101	0.492		1.760	

応用（モデルの同一性）

1. 変化点検出（例：景気の転換点）
2. 局所定常時系列モデル（例：地震波到着時刻）

回帰モデル

連続的な値をとる変数間の確率的な構造を記述するモデル

- 単回帰モデル . . . ひとつの変数で説明 (多項式など)
- 重回帰モデル . . . いくつかの変数で説明
- 自己回帰モデル . . . 過去の値で説明

- | | |
|------------|-----------|
| 1. パラメータ推定 | 最尤法、最小二乗法 |
| 2. 次数選択 | AIC |
| 3. 変数選択 | AIC |
| 4. モデル選択 | AIC |

- 単回帰 次数選択, モデル選択 (多項式, 三角関数など)
- 重回帰 変数選択
- 自己回帰 変数選択, 次数選択
- 非線形回帰 モデル選択, 変数選択, 次数選択

重回帰モデル

$$y_n = a_1 x_{n1} + \cdots + a_m x_{nm} + \varepsilon_n \quad \varepsilon_n \sim N(0, \sigma^2)$$

y_n : 目的変数 m : 次数 ε_n : 残差

$x_{n1} \cdots x_{nm}$: 説明変数 a_i : 回帰係数 σ^2 : 残差分散

パラメータ $\theta = (a_1, \cdots, a_m, \sigma^2)^T$

重回帰モデルが定める確率分布

$$\begin{aligned} p(y_n | \theta, x_{n1}, \dots, x_{nm}) &\sim N\left(\sum_{j=1}^m a_j x_{nj}, \sigma^2\right) \\ &= (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2} \left(y_n - \sum_{j=1}^m a_j x_{nj}\right)^2\right\} \end{aligned}$$

重回帰モデルの対数尤度

パラメータ $\theta = (a_1, \dots, a_m, \sigma^2)^T$

データ $\{y_n, x_{n1}, \dots, x_{nm}\} \quad (n = 1, \dots, N)$

$$L(\theta) = \prod_{n=1}^N p(y_n | \theta, x_{n1}, \dots, x_{nm}) \quad \text{尤度}$$

$$\ell(\theta) = \log L(\theta) = \sum_{n=1}^N \log p(y_n | \theta, x_{n1}, \dots, x_{nm}) \quad \text{対数尤度}$$

$$p(y_n | \theta, x_{n1}, \dots, x_{nm}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} \left(y_n - \sum_{j=1}^m a_j x_{nj} \right)^2 \right\}$$

$$\log p(y_n | \theta, x_{n1}, \dots, x_{nm}) = -\frac{1}{2} \log 2\pi\sigma^2 - \frac{1}{2\sigma^2} \left(y_n - \sum_{j=1}^m a_j x_{nj} \right)^2$$

$$\ell(\theta) = -\frac{N}{2} \log 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum_{n=1}^N \left(y_n - \sum_{j=1}^m a_j x_{nj} \right)^2$$

重回帰モデルの最尤推定

$$\max \ell(\theta) \Rightarrow \hat{\theta} = (\hat{a}_1, \dots, \hat{a}_m, \hat{\sigma}^2)$$

$$\frac{\partial \ell(\theta)}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{n=1}^N (y_n - \sum_{j=1}^m a_j x_{nj})^2 = 0$$

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{n=1}^N (y_n - \sum_{j=1}^m a_j x_{nj})^2 \quad a_j \text{ の値に関係なく } \hat{\sigma}^2 \text{ の最尤推定値は決まる}$$

$$\ell(a_1, \dots, a_m) = -\frac{N}{2} \log 2\pi\hat{\sigma}^2 - \frac{N}{2}$$

$$\max \ell(a_1, \dots, a_m) \Leftrightarrow \min \hat{\sigma}^2$$

線形正規型回帰モデルの場合：最尤法 \longleftrightarrow 最小二乗法

重回帰モデルの最小二乗推定

$$y_n = a_1 x_{n1} + \cdots + a_m x_{nm} + \varepsilon_n \quad \varepsilon_n \sim N(0, \sigma^2)$$

y_n : 目的変数 $x_{n1} \cdots x_{nm}$: 説明変数

m : 次数 a_i : 回帰係数 ε_n : 残差 σ^2 : 残差分散

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \quad Z = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ x_{21} & \cdots & x_{2m} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{Nm} \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}, \quad a = \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix}$$

$$y = Za + \varepsilon$$

y 目的変数ベクトル
 Z 説明変数行列
 a 回帰係数ベクトル
 ε 残差ベクトル

重回帰モデルの最小二乗法

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{n=1}^N (y_n - \sum_{j=1}^m a_j x_{nj})^2 = \|y - Za\|_N^2 = \|\varepsilon\|_N^2$$

$$\frac{\partial \hat{\sigma}^2}{\partial a} = -Z^T (y - Za) = 0 \quad \Rightarrow \quad Z^T y = Z^T Za \quad \Rightarrow \quad \hat{a} = (Z^T Z)^{-1} Z^T y$$

$$\begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} \sum x_{i1}^2 & \cdots & \sum x_{i1} x_{im} \\ \vdots & \ddots & \vdots \\ \sum x_{im} x_{i1} & \cdots & \sum x_{im} x_{im} \end{bmatrix}^{-1} \begin{bmatrix} \sum x_{i1} y_i \\ \vdots \\ \sum x_{im} y_i \end{bmatrix}$$

$$\Rightarrow \hat{a} = (\hat{a}_1, \dots, \hat{a}_m)^T$$

$$y = (y_1, \dots, y_N)^T \text{ のとき } \|y\|_N^2 = y^T y = \sum_{n=1}^N y_n^2$$

重回帰モデル（最大対数尤度とAIC）

$$\text{AIC}_m = -2\ell_m(\hat{\theta}) + 2k$$

$$\ell_m(\hat{\theta}) = -\frac{N}{2} \log 2\pi\hat{\sigma}_m^2 - \frac{N}{2}$$

$$k = m + 1$$

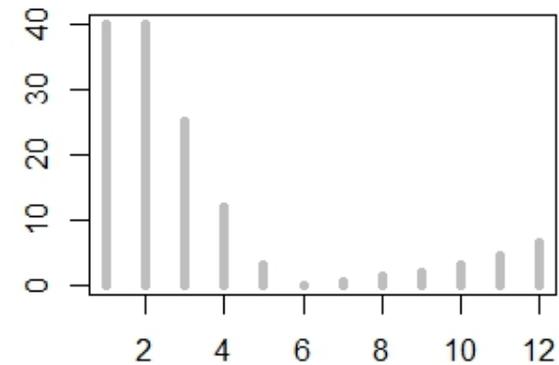
m : モデルの次数

k : パラメータ数

$$\text{AIC}_m = N \log 2\pi\hat{\sigma}_m^2 + N + 2(m + 1)$$

$$\text{AIC}_m^* = N \log \hat{\sigma}_m^2 + 2m$$

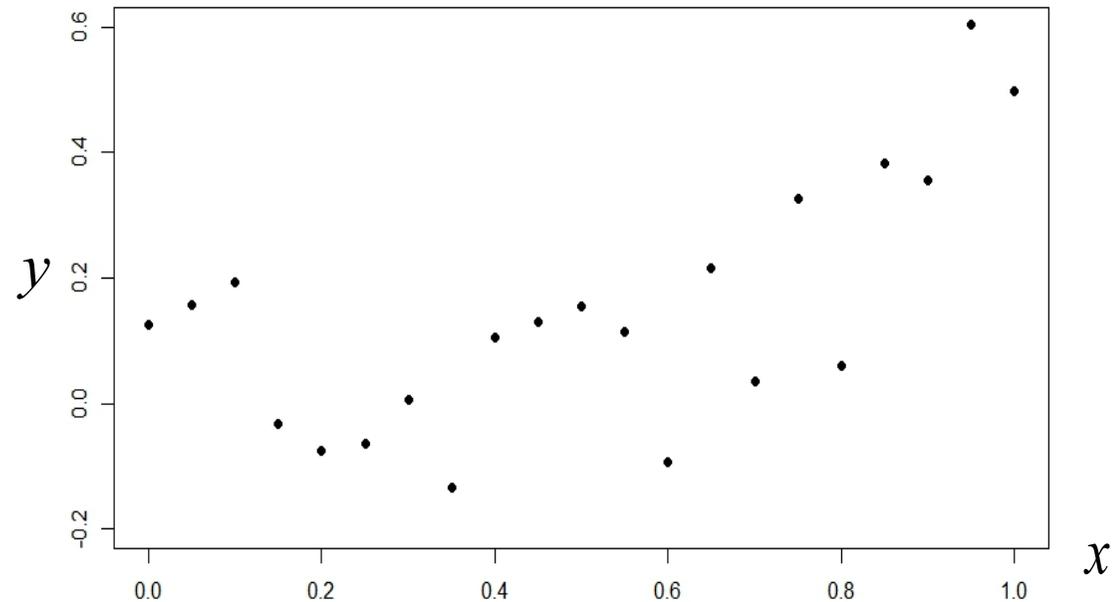
- ・常に定数項まで計算しておいた方がよい
- ・別のモデル（族）との比較が可能になる



例：多項式回帰

i	x_i	y_i
1	0.00	0.125
2	0.05	0.156
3	0.10	0.193
4	0.15	-0.032
5	0.20	-0.075
6	0.25	-0.064
7	0.30	0.006
8	0.35	-0.135
9	0.40	0.105
10	0.45	0.131
11	0.50	0.154
12	0.55	0.114
13	0.60	-0.094
14	0.65	0.215
15	0.70	0.035
16	0.75	0.327
17	0.80	0.061
18	0.85	0.383
19	0.90	0.357
20	0.95	0.605
21	1.00	0.499

y : 目的変数 x : 説明変数



$$y = a_0 + a_1x + \cdots + a_px^p + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2)$$

$$\theta = (a_0, a_1, \cdots, a_p, \sigma^2)$$

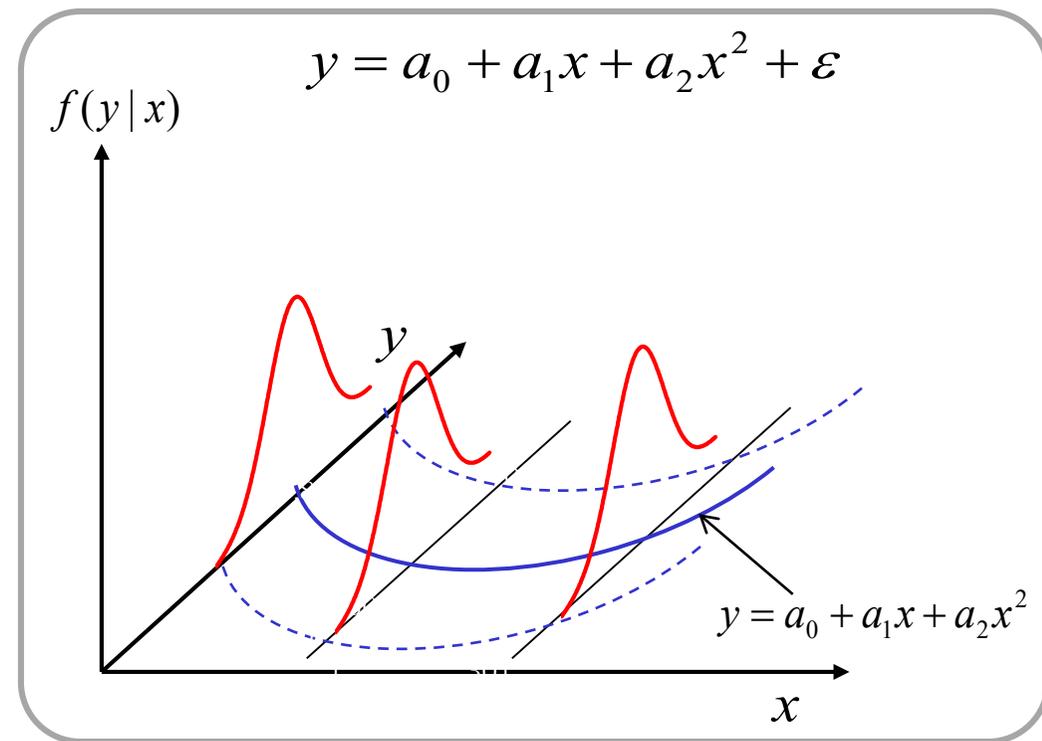
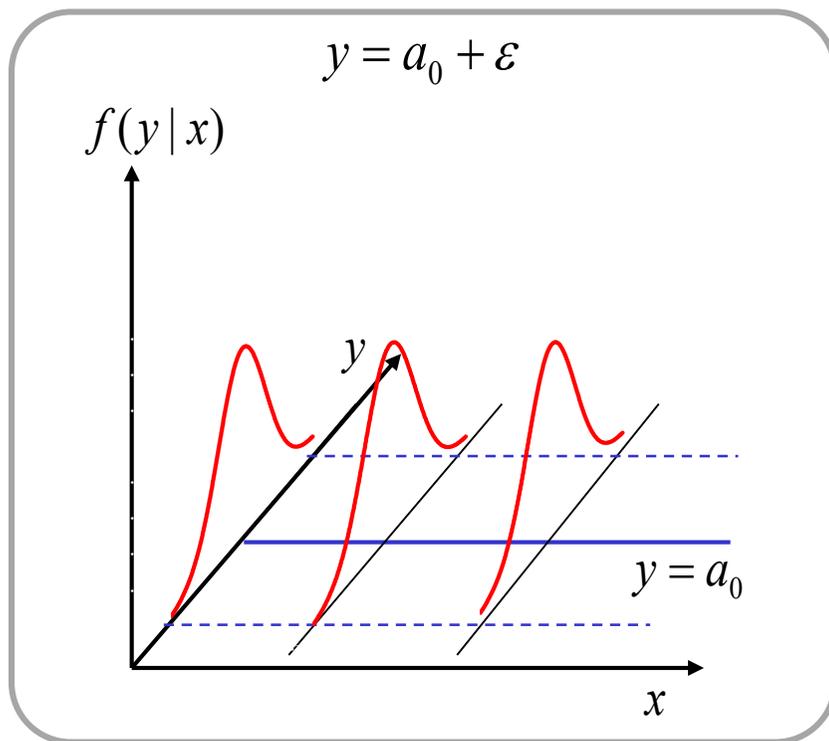
```
x <- seq(0,1,length=21)
y <- c(0.125, 0.156, 0.193, -0.032, -0.075, -0.064, 0.006, -0.135, 0.105, 0.131, 0.154, 0.114, -
0.094, 0.215, 0.035, 0.327, 0.061, 0.383, 0.357, 0.605, 0.499)
plot(x,y,pch=19,ylim=c(-0.2,0.6))
```

多項式回帰モデル

$$y_i = f(x_i) + \varepsilon_i$$

$$f(x) = a_0 + a_1x + \cdots + a_mx^m, \quad \varepsilon_i \sim N(0, \sigma^2)$$

$$y_i \sim N(\mu_i, \sigma^2), \quad \mu_i = a_0 + a_1x_i + \cdots + a_mx_i^m$$



多項式回帰モデル

$$x_{nj} = x^{j-1}, \quad j = 1, \dots, m+1$$

$$a = (a_0, a_1, \dots, a_m)^T, \quad \theta = (a_0, a_1, \dots, a_m, \sigma^2)^T$$

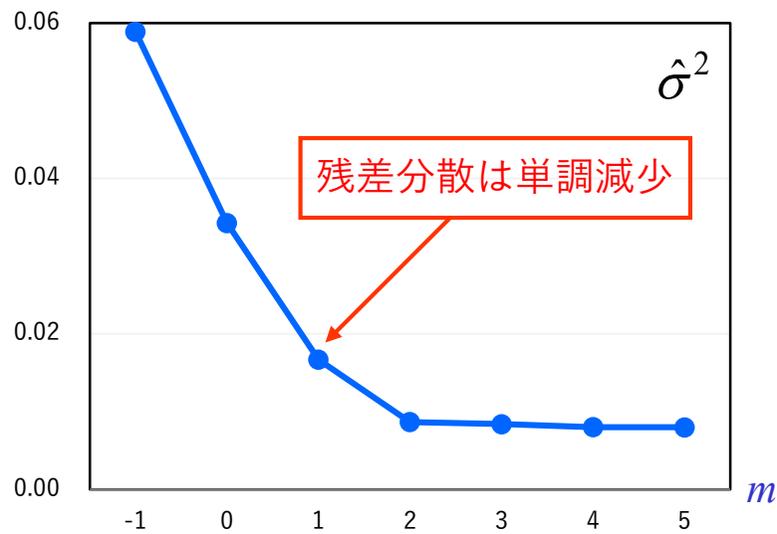
対数尤度 $\ell_m(\hat{\theta}) = -\frac{N}{2} \log 2\pi\hat{\sigma}_m^2 - \frac{N}{2}$

パラメータ数 $k = m + 2$

$$\text{AIC}_m = N \log 2\pi\hat{\sigma}_m^2 + N + 2(m + 2)$$

$$\text{AIC}_m^* = N \log \hat{\sigma}_m^2 + 2m$$

多項式回帰の次数選択

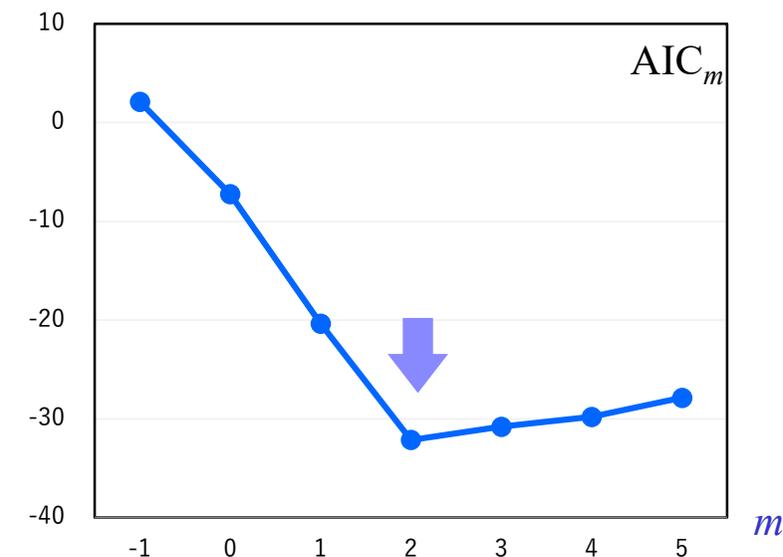


$$\theta = (a_0, a_1, \dots, a_m, \sigma^2)$$

$$\ell(\theta) = -\frac{N}{2} \log(2\pi\sigma^2) - \frac{1}{2\pi\sigma^2} \sum_{i=1}^N \left(y_i - \sum_{j=0}^m a_j y_{i-j} \right)^2$$

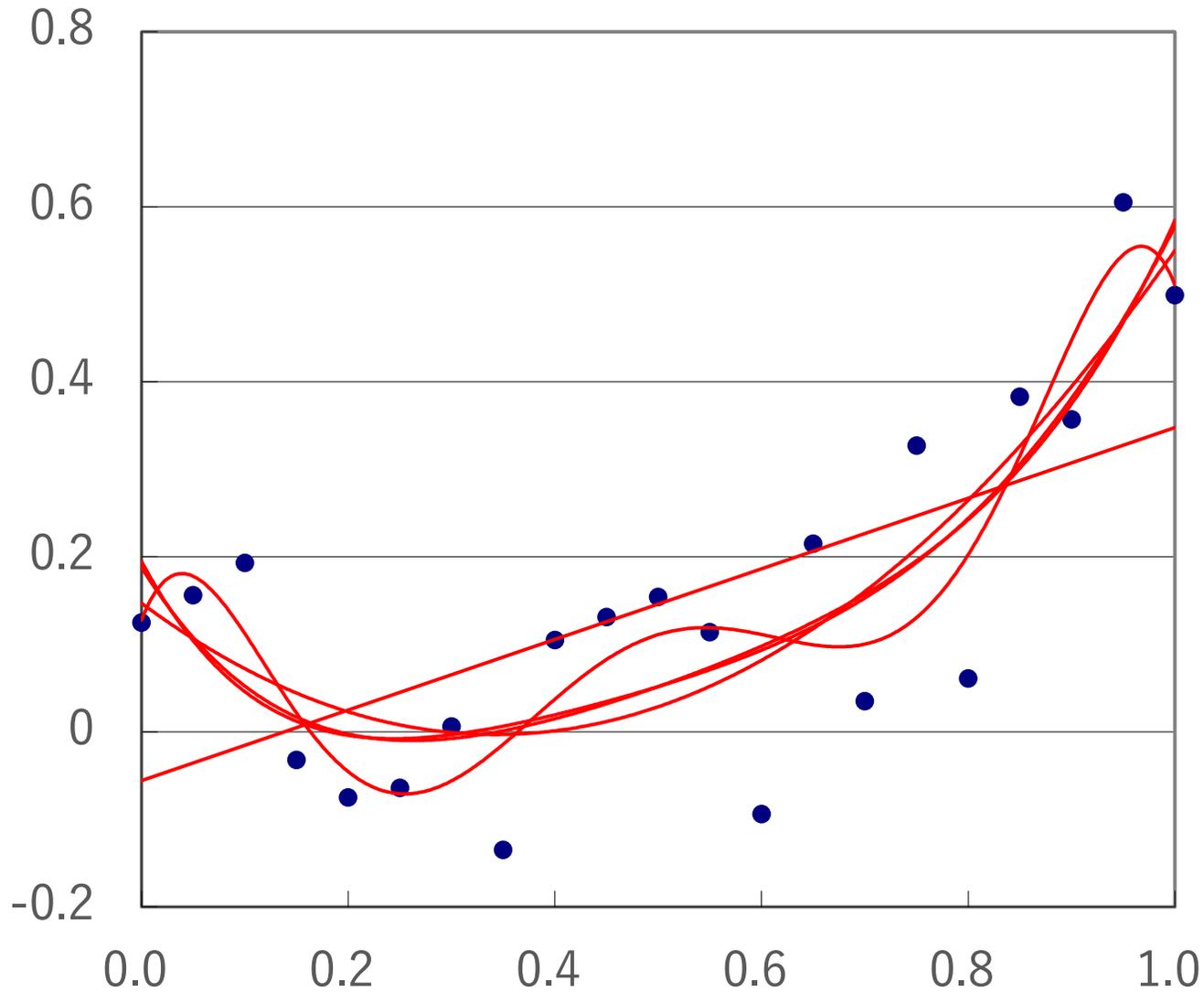
$$\ell(\hat{\theta}) = -\frac{N}{2} \log(2\pi\hat{\sigma}^2) - \frac{N}{2}$$

$$\text{AIC}_m = N(\log 2\pi + 1) + N \log \hat{\sigma}^2 + 2(m+2)$$



次数	パラメータ数	残差分散	AIC
-1	1	0.05889	2.12
0	2	0.03427	-7.25
1	3	0.01669	-20.35
2	4	0.00866	-32.13
3	5	0.00839	-30.80
4	6	0.00800	-29.79
5	7	0.00798	-27.86

推定された多項式回帰曲線 ($m=1, \dots, 5$)



シミュレーション

True Model :

$$y_i = \exp\{(x_i - 0.3)^2\} - 1 + 0.1\varepsilon_i$$

$$\varepsilon_i \sim N(0,1)$$

$$x_i = 0.05(i - 1), \quad i = 1, \dots, 21$$

注意：

多項式回帰モデルはTrue Modelを含んでいない。

データ生成

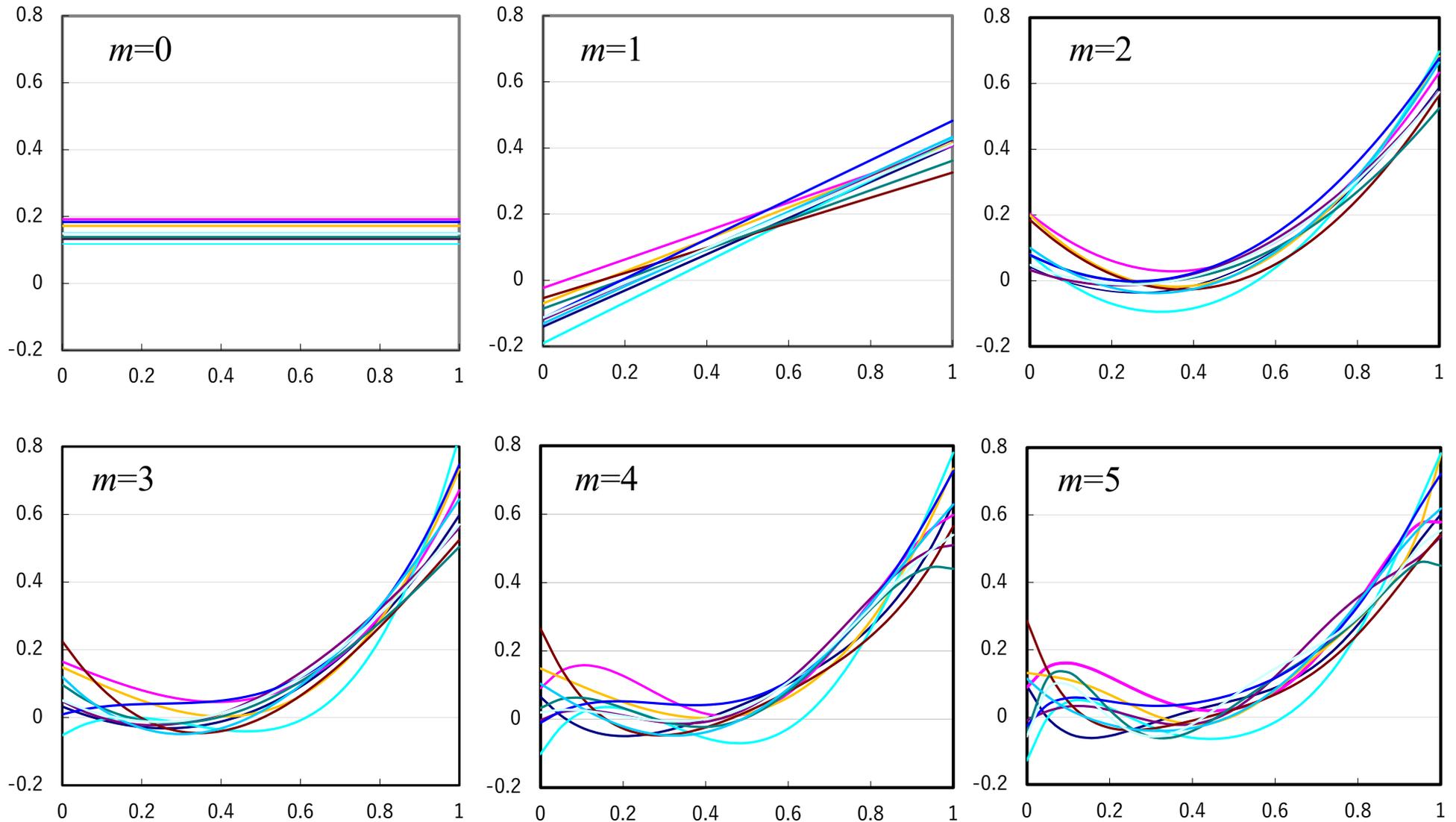


モデル推定



L 回繰り返す, 重ね書き

シミュレーション結果 ($L=10$)



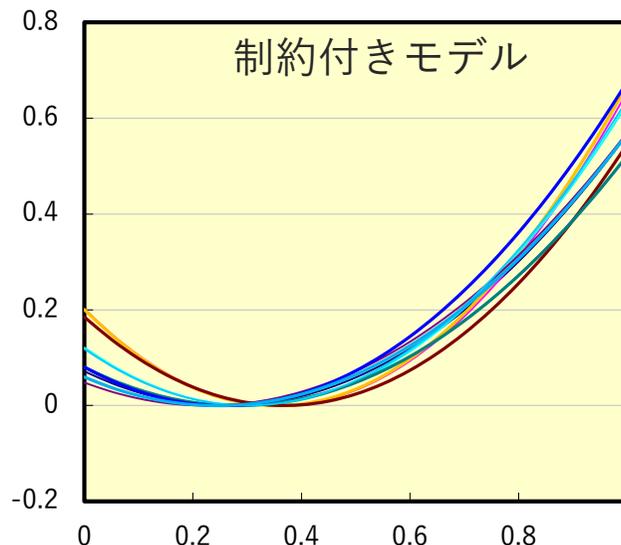
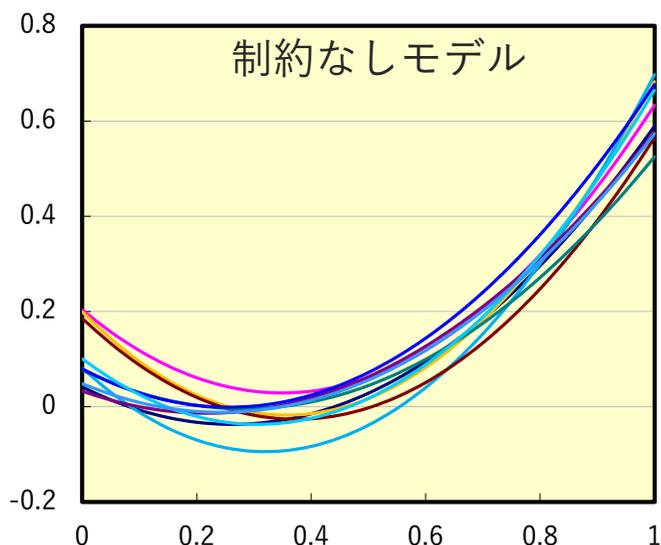
制約つきモデル

例えば, 回帰式の最小値が 0 となることを知っていたら

$$y_i = \alpha(x_i + \beta)^2 + \varepsilon_i \quad \text{パラメータが1個少ないモデル}$$

	\hat{a}_0	\hat{a}_1	\hat{a}_2	$\hat{\sigma}^2$	AIC
制約なし	0.1116	-0.6590	1.0969	0.00866	-32.13
制約付き	0.1061	-0.6990	1.1515	0.00875	-33.92

$$\begin{aligned}\hat{\alpha} &= 1.1515 \\ \hat{\beta} &= -0.3035 \\ \hat{a}_0 &= \hat{\alpha} \\ \hat{a}_1 &= 2\hat{\alpha}\hat{\beta} \\ \hat{a}_2 &= \hat{\alpha}\hat{\beta}^2\end{aligned}$$



時系列への応用

単回帰モデル（多項式，三角関数）によるトレンド推定

$$y_i = f(x_i) + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2)$$

$$f(x) = a_0 + a_1x + \cdots + a_mx^m$$

$$f(x) = a + \sum_{j=1}^m b_j \sin(j\omega n) + \sum_{j=1}^{\ell} c_j \cos(j\omega n)$$

最高気温データ（三角関数回帰モデル）

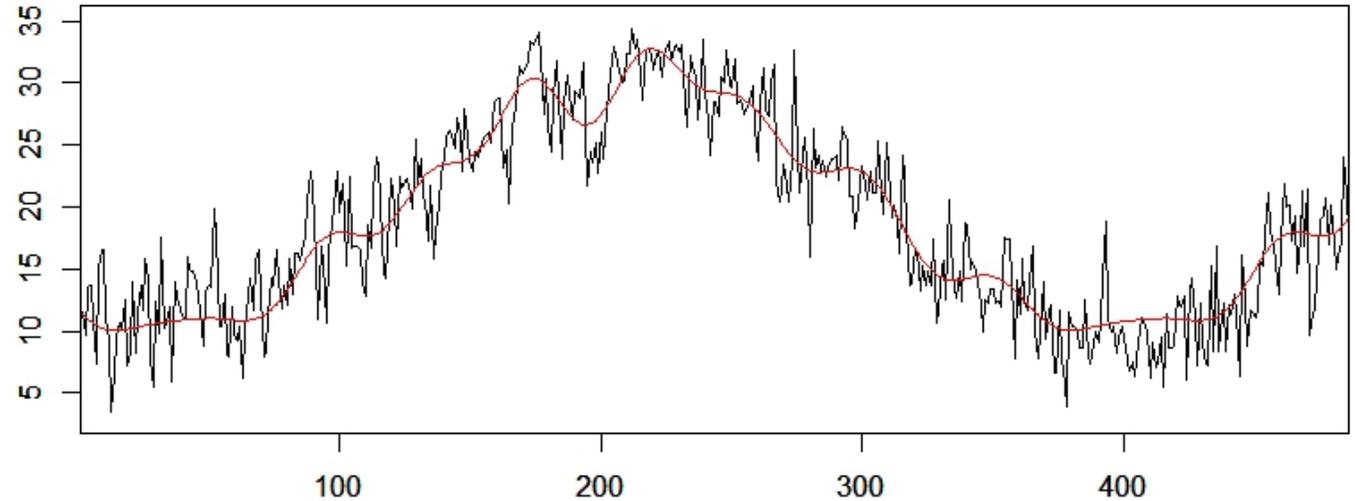
```
data(Temperature) # Highest Temperature Data of Tokyo
lsqr(Temperature)
```

$$y_i = f(x_i) + \varepsilon_i$$

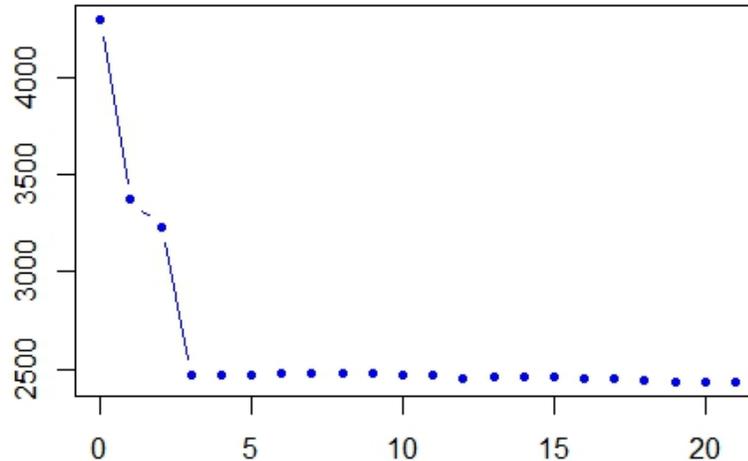
$$\varepsilon_i \sim N(0, \sigma^2)$$

$$f(x) = a + \sum_{j=1}^m b_j \sin(jwn) + \sum_{j=1}^{\ell} c_j \cos(jwn)$$

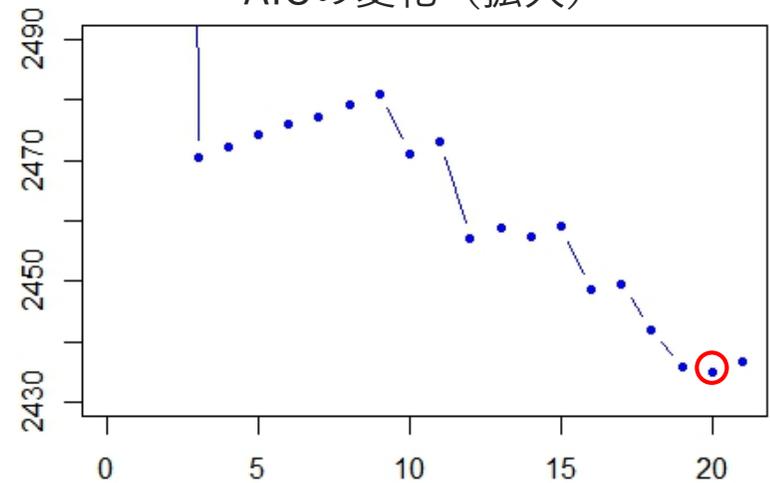
Temperature and regression curve of the model with order 20



AICの変化



AICの変化（拡大）



$\ell = m$ または $m-1$

説明変数 20
 ・ 定数項
 ・ sin 10
 ・ cos 9

最高気温データ（多項式回帰モデル）

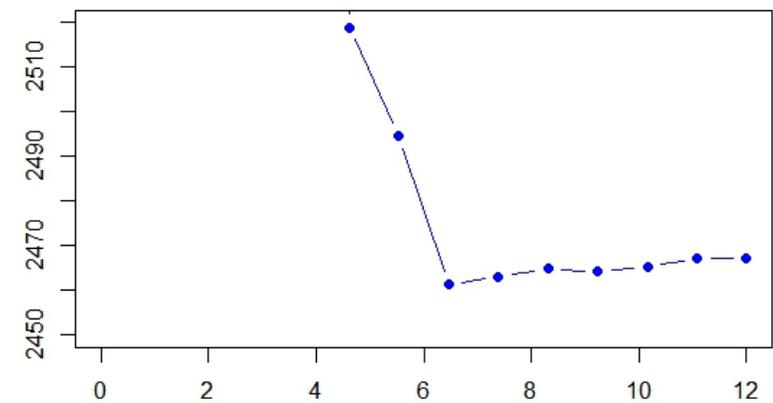
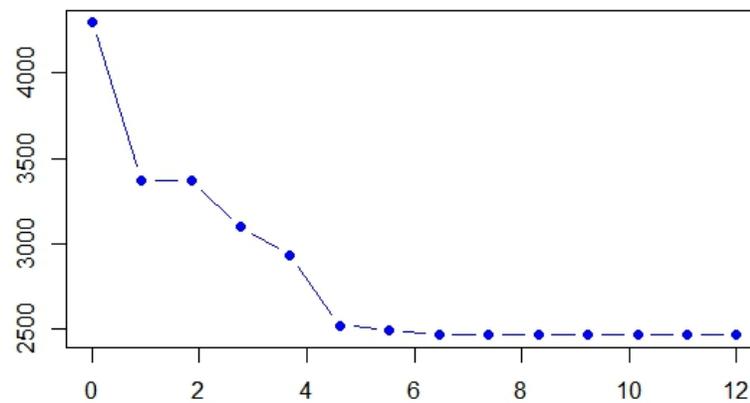
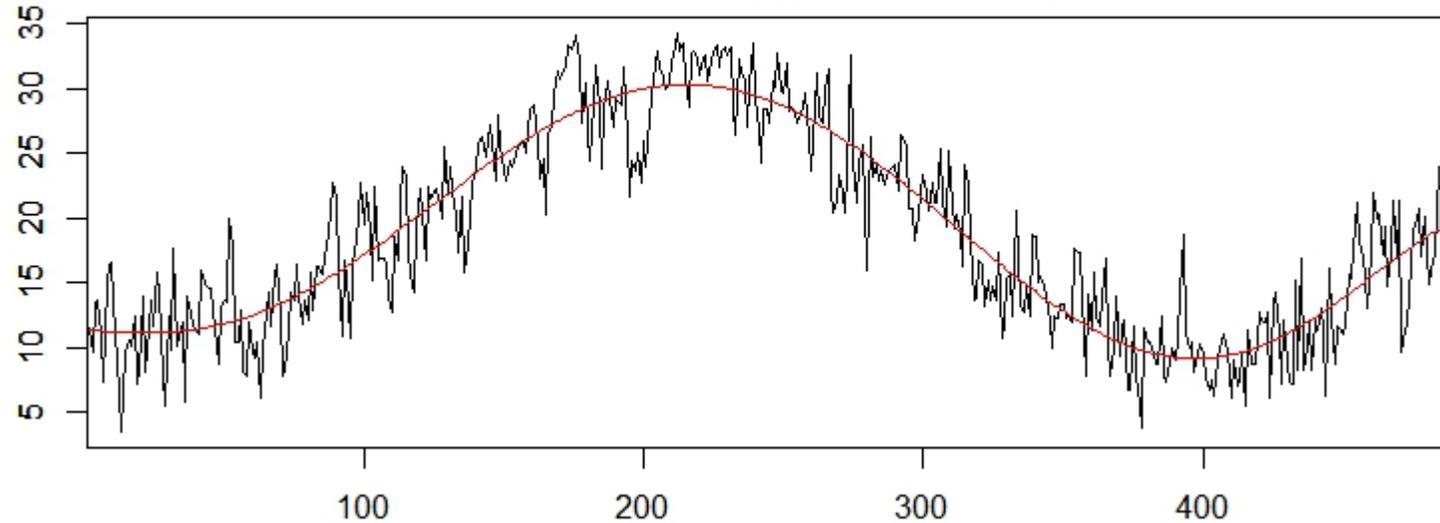
```
polreg(Temperature,13)
```

$$y_i = f(x_i) + \varepsilon_i$$

$$\varepsilon_i \sim N(0, \sigma^2)$$

$$f(x) = a_0 + a_1x + \dots + a_mx^m$$

Temperature and trend component
minimum aic = 2461.42 at order 7



```
z <- polreg(Temperature,13)
```

```
x <- seq(0,12,length=14)
```

```
plot(x,z$aic,col="blue",type="b",pch=19)
```

```
plot(x,z$aic,col="blue",type="b",pch=19,ylim=c(2450,2520))
```

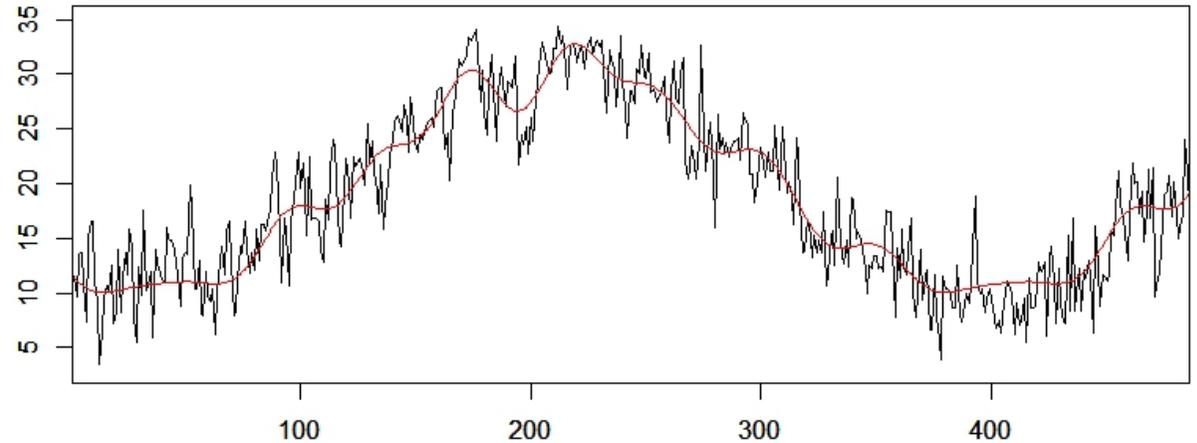
最高気温データ（モデル比較）

```
data(Temperature)  
lsqr(Temperature)
```

- 三角関数回帰モデル

AIC = 2435.06

Temperature
and regression curve of the model with order 20

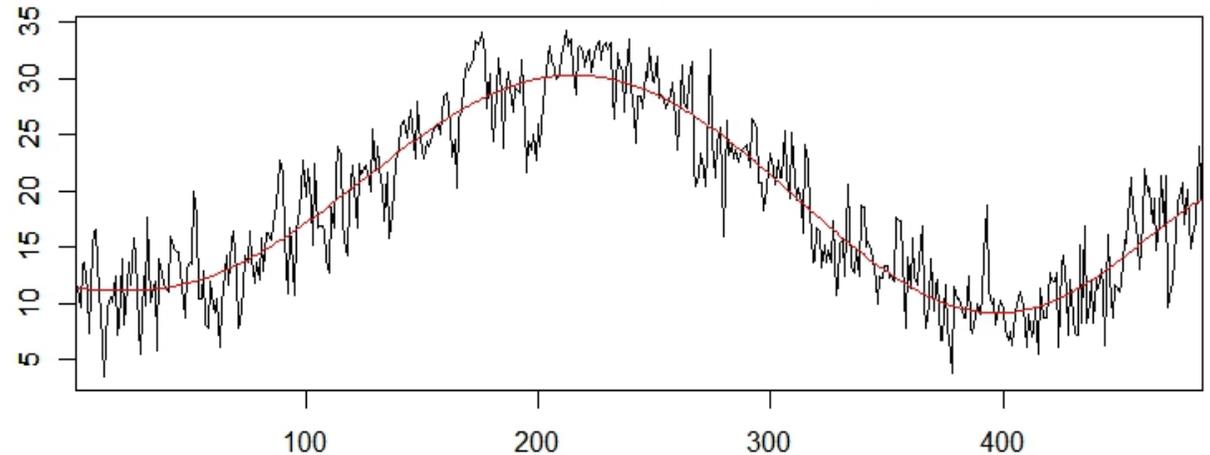


```
polreg(Temperature,7)
```

- 多項式回帰モデル

AIC=2461.42

Temperature and trend component
minimum aic = 2461.42 at order 7

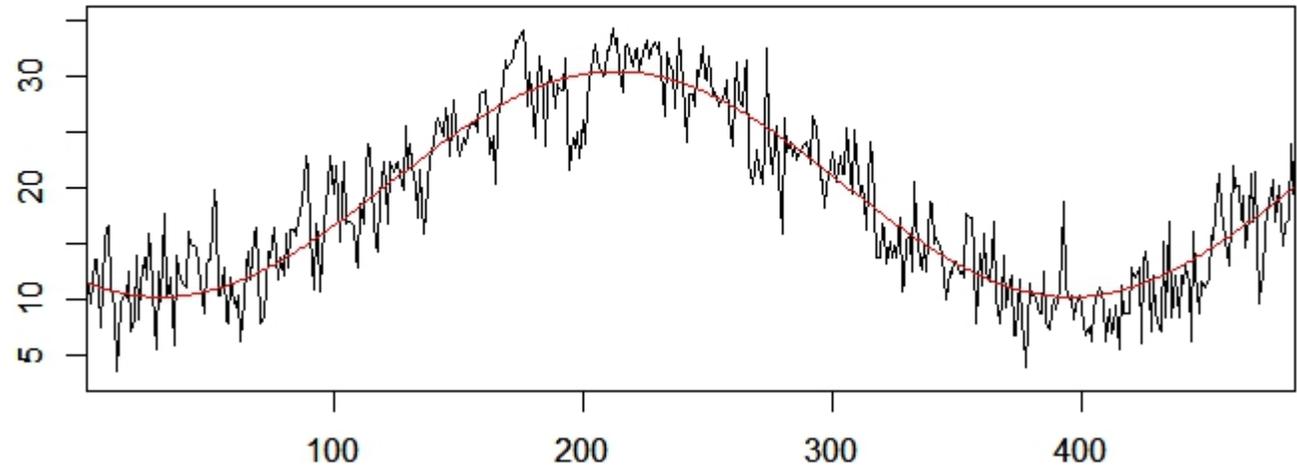
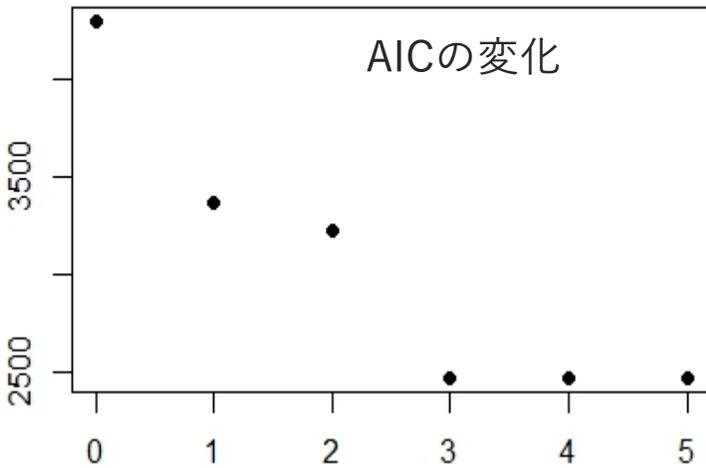


次数選択における注意

```
data(Temperature) # Highest Temperature Data of Tokyo  
lsqr(Temperature, lag=2)
```

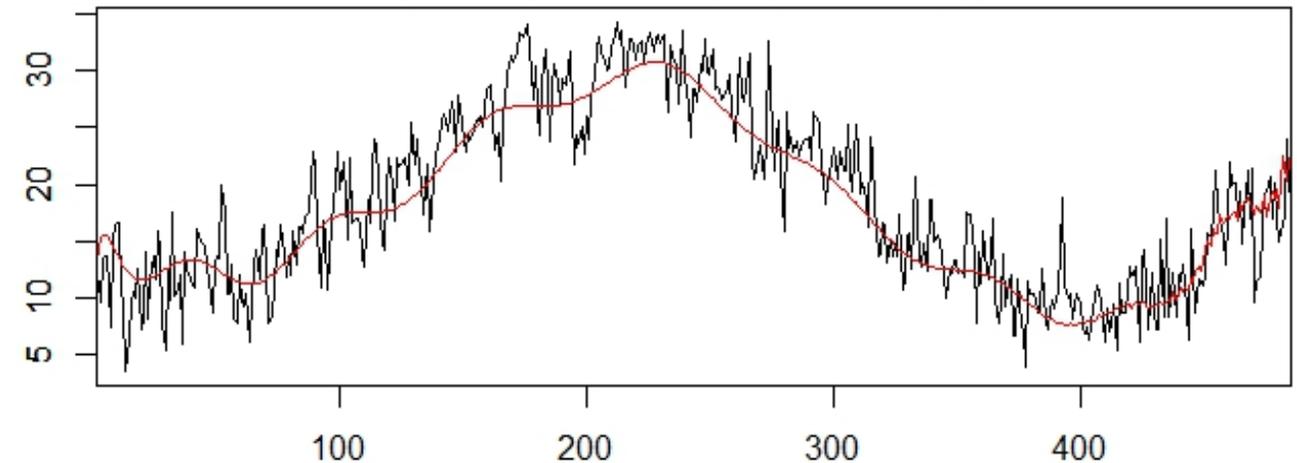
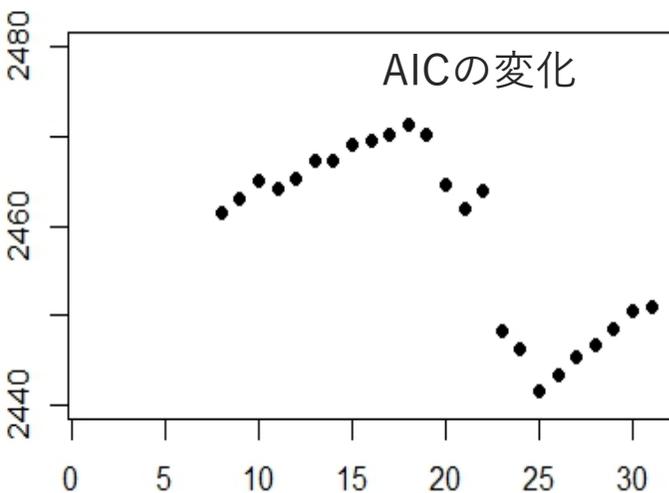
三角関数回帰

Temperature and regression curve of the model with order 3



多項式回帰

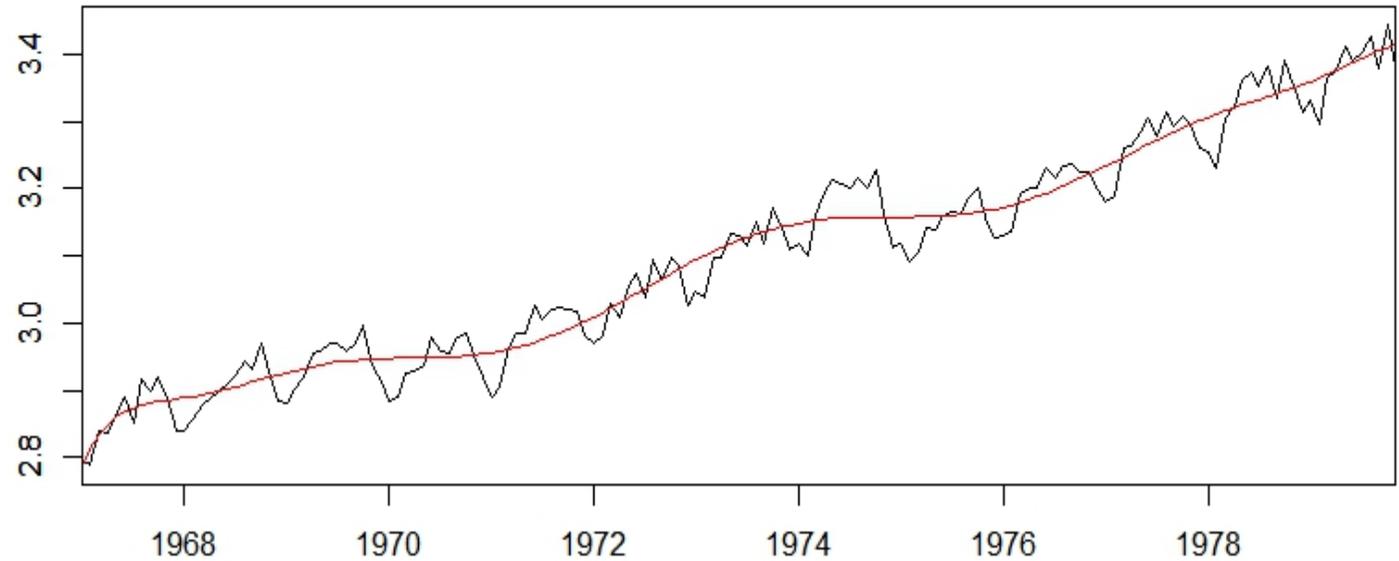
Temperature and trend component minimum aic = 2441.47 at order 24



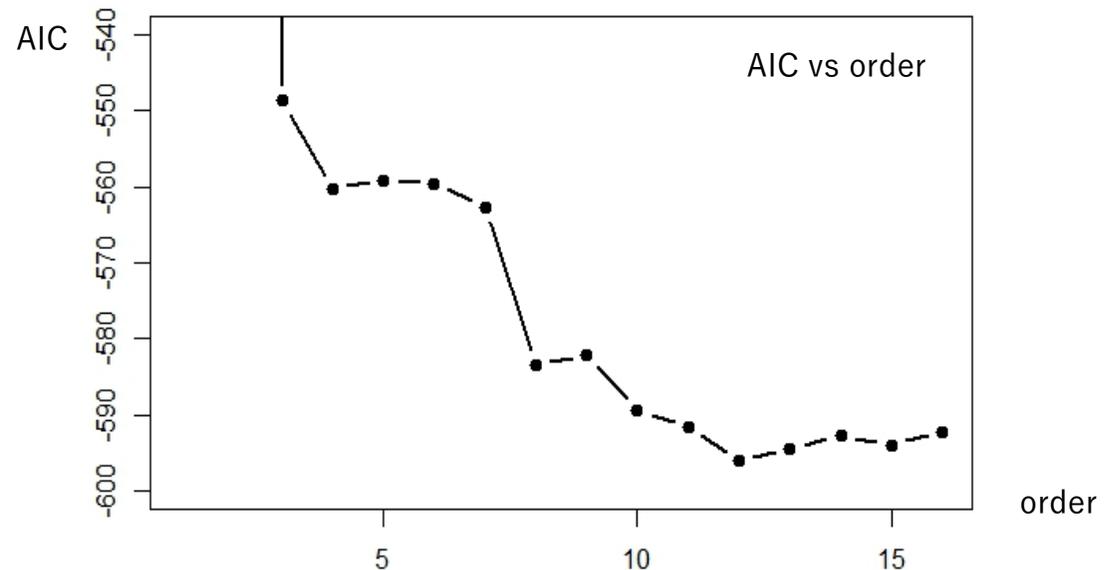
Whardデータ (多項式回帰)

```
data(WHARD)# Wholesale hardware data  
y <- log10(WHARD)  
polreg(y, 14)
```

y and trend component
minimum aic = -595.98 at order 11



```
# AIC's of polinomial regression models  
z <- polreg(y,14)  
par(mar=c(2,2,2,1)+0.1)  
plot(z$aic,type="b",pch=19,lwd=2,ylim=  
c(-600,-540))
```



多変量データ

気温：1月

都市名	気温 y	緯度 x_1	経度 x_2	標高 x_3
稚内	-8.9	45.42	141.68	2.8
旭川	-13.6	43.77	142.37	111.9
札幌	-9.5	43.05	141.33	17.2
青森	-5.4	40.82	140.78	3.0
盛岡	-6.7	39.70	141.17	155.2
仙台	-3.2	38.27	140.90	38.9
金沢	-0.1	36.55	136.65	26.1
長野	-5.5	36.67	138.20	418.2
高山	-7.6	36.15	137.25	560.2
軽井沢	-10.0	36.33	138.55	999.1
名古屋	-0.9	35.17	136.97	51.1
飯田	-4.7	35.52	137.83	481.8
東京	-0.4	35.68	139.77	5.3
鳥取	0.5	35.48	134.23	7.1
京都	-0.6	35.02	135.73	41.4
広島	0.2	34.37	132.43	29.3
福岡	1.5	33.58	130.38	2.5
鹿児島	2.0	31.57	130.55	4.3
高知	0.1	33.55	133.53	1.9
那覇	13.5	26.23	127.68	34.9

回帰モデルの変数選択

y_n : 気温, x_{1n} : 緯度, x_{2n} : 経度, x_{3n} : 標高

$$y_n = a_0 + a_1 x_{1n} + a_2 x_{2n} + a_3 x_{3n} + \varepsilon_n, \quad \varepsilon_n \sim N(0, \sigma^2)$$

説明変数 x_1, x_2, x_3 のなかで y_n の予測に有用な変数を選択

$$\ell(a_0, a_1, a_2, a_3, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n \left(y_n - a_0 - \sum_{j=1}^3 a_j x_{jn} x_{jn} \right)^2$$

$$\text{AIC}(x_{1n}, x_{2n}, x_{3n}) = n(\log 2\pi + 1) + n \log \hat{\sigma}^2 + 2(k + 2)$$

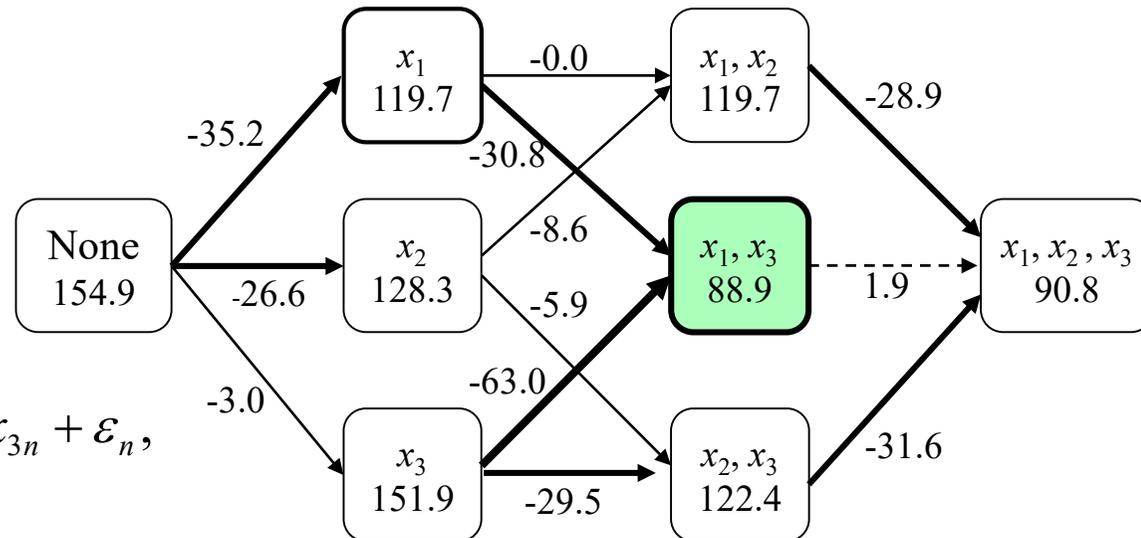
選択された変数

x_{1n}, x_{3n}

選択されたモデル

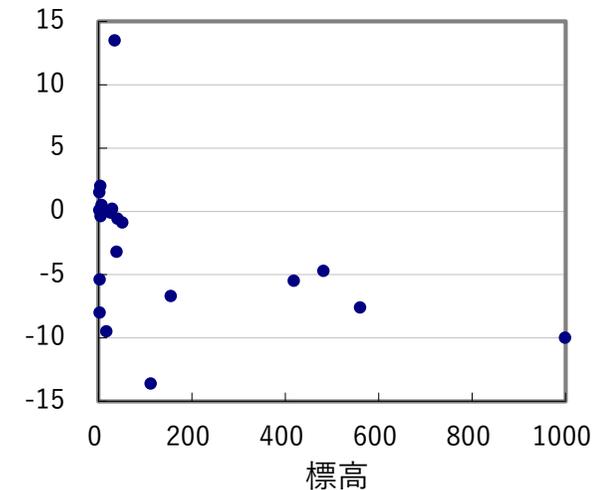
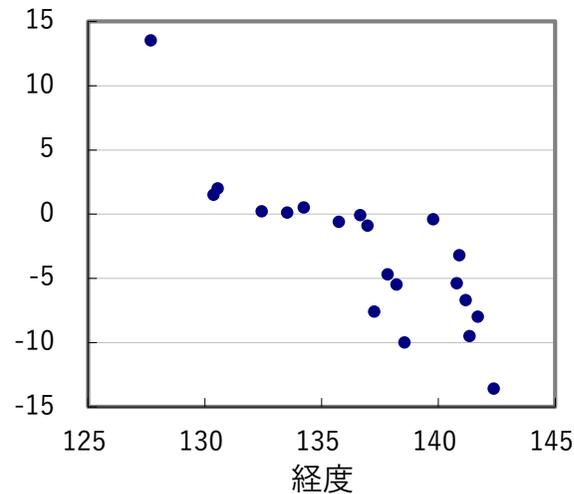
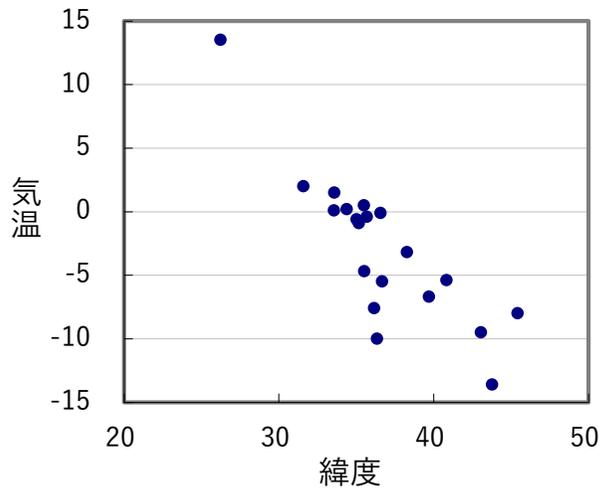
$$y_n = 40.490 - 1.208x_{1n} - 0.010x_{3n} + \varepsilon_n,$$

$$\varepsilon_n \sim N(0, 1.490)$$



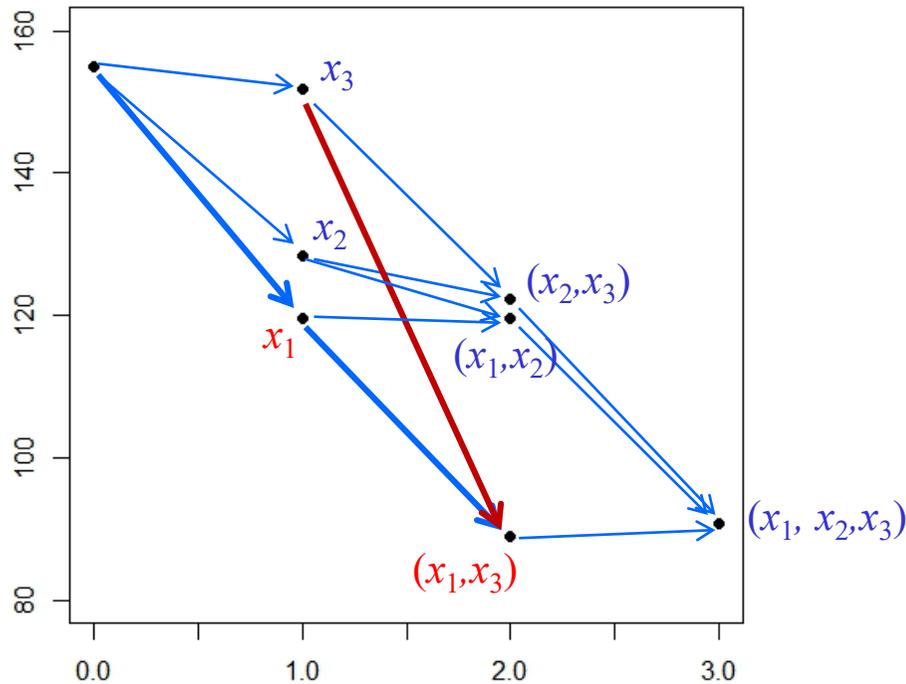
モデル比較

モデル番号	説明変数	次数	残差分散	AIC	回帰係数			
					a_0	a_1	a_2	a_3
1	1, 緯度, 標高	3	2.46	82.8	40.7	-1.15		-0.0098
2	1, 緯度, 経度, 標高	4	2.46	84.8	38.3	-1.17	0.02	-0.0098
3	1, 経度, 標高	3	7.01	103.7	147.1		-1.09	-0.0064
4	1, 緯度, 経度	3	7.74	105.7	94.3	-0.69	-0.53	
5	1, 緯度	2	8.73	106.1	39.0	-1.14		
6	1, 経度	2	9.60	108.0	155.8		-1.16	
7	1, 標高	2	26.68	128.4	-1.5			
8	1	1	32.66	130.5	-2.9			
9	なし	0	41.19	133.1				



代理変数

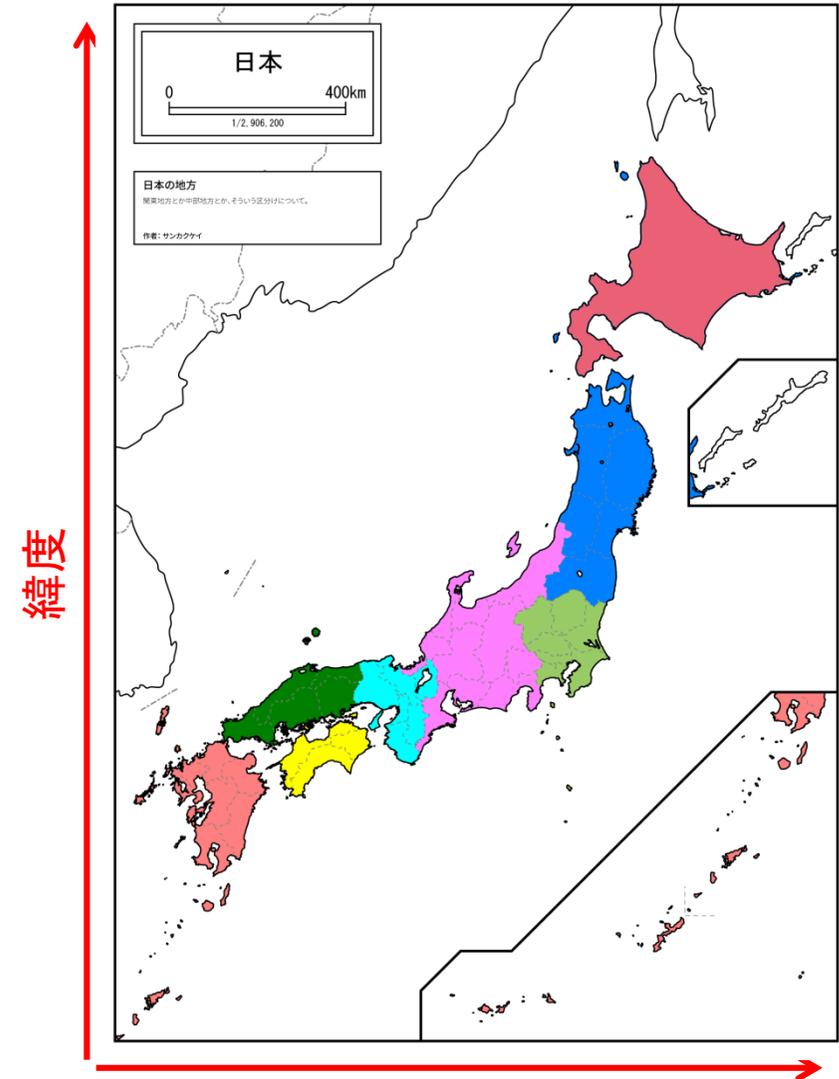
変数の採用によるAICの変化



$$AIC(x_1) < AIC(x_2) \square AIC(x_3)$$

$$AIC(x_1, x_2) > AIC(x_1, x_3)$$

- モデルに x_1 (緯度) が含まれない場合, x_2 (経度) が代理変数となる.
- 日本列島では, 緯度と経度の相関が高い (相関係数: 0.84)



予測誤差

$$y = 40.7 - 1.15x_1 - 0.0098x_3 + \varepsilon, \quad \varepsilon \sim N(0, 2.46)$$

$$\text{福島} : y = -3.1, x_1 = 37.75, x_2 = 140.45, x_3 = 67.4$$

$$\text{神戸} : y = 1.2, x_1 = 34.68, x_2 = 135.19, x_3 = 59.3$$

$$\hat{y} = 40.7 - (1.15)(37.75) - (0.0098)(67.4) = -3.37$$

$$\hat{y} = 40.7 - (1.15)(34.68) - (0.0098)(59.3) = 0.24$$

	実測値	予測値	予測値(2)
福島	-3.1	-3.37	-3.43
神戸	1.2	0.24	-0.15

注：予測値(2)はAICが2位のフルモデルによる予測

最小二乗法の解法

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, Z = \begin{bmatrix} z_{11} & \cdots & z_{1m} \\ \vdots & & \vdots \\ z_{n1} & \cdots & z_{nm} \end{bmatrix}, a = \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$y = Za + \varepsilon \quad \|\varepsilon\|^2 = \|y - Za\|^2$$

- 従来の方法

$$\hat{a} = (Z^T Z)^{-1} Z^T y$$

- 直交変換に基づく方法

- Householder法
- 修正Gram-Schmidt法
- Givens法

Householder法

U : 任意の直交変換 (ベクトルの長さを変えない)

$$\|\varepsilon\|_N^2 = \|y - Za\|_N^2 = \|U(y - Za)\|_N^2 = \|Uy - UZa\|_N^2$$

$$\min_a \|\varepsilon\|^2 \Leftrightarrow \min_a \|Uy - UZa\|^2$$

$$X = \left[\begin{array}{c|c} \overbrace{Z}^{m+1} & y \end{array} \right] \Bigg|_N \Rightarrow UX = S = \begin{bmatrix} s_{11} & \cdots & s_{1m} & s_{1,m+1} \\ & \ddots & \vdots & \vdots \\ & & s_{mm} & s_{m,m+1} \\ & & & s_{m+1,m+1} \\ & 0 & & \end{bmatrix}$$

最小二乘法 (Householder法)

$$\begin{aligned}
 \|Uy - UZa\|_N^2 &= \left\| \begin{bmatrix} s_{1,m+1} \\ \vdots \\ s_{1,m+1} \\ s_{1,m+1} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} s_{11} & \cdots & s_{11} \\ & \ddots & \vdots \\ & & s_{11} \\ & & \\ \mathbf{0} & & \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix} \right\|_N^2 \\
 &= \left\| \begin{bmatrix} s_{1,m+1} \\ \vdots \\ s_{m,m+1} \end{bmatrix} - \begin{bmatrix} s_{11} & \cdots & s_{1m} \\ & \ddots & \vdots \\ & & s_{mm} \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix} \right\|_m^2 + s_{m+1,m+1}^2
 \end{aligned}$$

$$m \leq N$$

最小二乗解 (Householder法)

$$\begin{bmatrix} s_{11} & \cdots & s_{1m} \\ & \ddots & \vdots \\ & & s_{mm} \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} s_{1,m+1} \\ \vdots \\ s_{m,m+1} \end{bmatrix}$$

$$\hat{a}_m = \frac{s_{m,m+1}}{s_{mm}}$$

$$\hat{a}_i = \frac{s_{i,m+1} - s_{i,i+1}\hat{a}_{i+1} - \cdots - s_{i,m}\hat{a}_m}{s_{ii}} \quad i = m-1, \dots, 1$$

$$\hat{\sigma}_m^2 = \frac{s_{m+1,m+1}^2}{n}$$

後退代入だけで求まるので逆行列不要

AICによる次数選択

$$\ell(\hat{\theta}) = -\frac{N}{2} \log 2\pi\hat{\sigma}_m^2 - \frac{N}{2}$$

$$\begin{aligned} \text{AIC}_m &= -2\ell(\hat{\theta}) + 2(\text{パラメータ数}) \\ &= N(\log 2\pi\hat{\sigma}_m^2 + 1) + 2(m+1) \end{aligned}$$

for $k = 1, \dots, m$

$$\begin{bmatrix} s_{11} & \cdots & s_{1k} \\ & \ddots & \vdots \\ & & s_{kk} \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_k \end{bmatrix} = \begin{bmatrix} s_{1,m+1} \\ \vdots \\ s_{k,m+1} \end{bmatrix}$$

$$\begin{bmatrix} s_{11} & \cdots & s_{1k} & \cdots & s_{1m} & s_{1,m+1} \\ & \ddots & \vdots & & \vdots & \vdots \\ & & s_{kk} & \cdots & s_{km} & s_{k,m+1} \\ & & & \ddots & \vdots & \vdots \\ & & & & s_{mm} & s_{m,m+1} \\ & & & & & s_{m+1,m+1} \\ & & & & & 0 \end{bmatrix}$$

$$\hat{\sigma}_k^2 = \frac{1}{n} (s_{k+1,m+1}^2 + \cdots + s_{m+1,m+1}^2)$$

$$\text{AIC}_k = N(\log 2\pi\hat{\sigma}_k^2 + 1) + 2(k+1)$$

$\hat{\sigma}_k^2$ および AIC_k の計算に
回帰係数 $\hat{a}_1, \dots, \hat{a}_k$ は不要

Householder変換

行列を三角化する直交変換の求め方

$$b = Ua = a - 2w \cdot (w^T a)$$

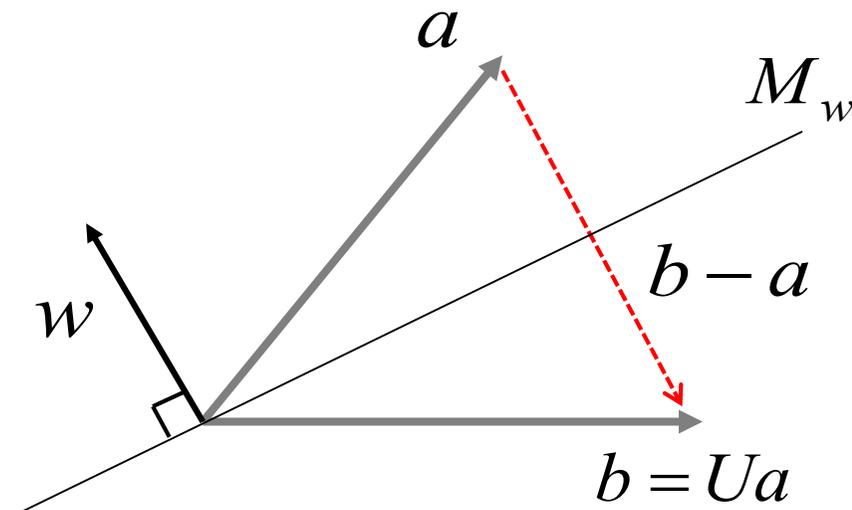
単位ベクトル w に直交する平面 M_w に関するベクトル a の鏡像 b

$$U = I - 2ww^T$$

鏡映変換

$$\begin{aligned} UU^T &= (I - 2ww^T)(I - 2ww^T)^T \\ &= I - 4ww^T + 4ww^T ww^T = I \end{aligned}$$

U は直交変換



Householder変換

$$\|a\|^2 = \|b\|^2 \Rightarrow \exists U \quad \text{such that } Ua = b$$

$$\begin{aligned} w = \frac{a-b}{\|a-b\|} \quad \text{とすると} \quad & Ua = (I - 2ww^T)a \\ &= a - \frac{(a-b)(a-b)^T}{\|a-b\|^2} 2a \\ &= a - \frac{(a-b)(a-b)^T}{\|a-b\|^2} \{(a-b) + (a+b)\} \\ &= a - (a-b) - \frac{(a-b)(\|a\|^2 - \|b\|^2)}{\|a-b\|^2} \\ &= b \end{aligned}$$

Householder変換

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}, \quad a_1 = \begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{bmatrix}, \quad b_1 = \begin{bmatrix} x_{11}^{(1)} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad x_{11}^{(1)} = \mp \|a_1\|$$

$$U_1 X = \begin{bmatrix} x_{11}^{(1)} & x_{12}^{(1)} & \cdots & x_{1m}^{(1)} \\ 0 & x_{22}^{(1)} & \cdots & x_{2m}^{(1)} \\ \vdots & \vdots & & \vdots \\ 0 & x_{n2}^{(1)} & \cdots & x_{nm}^{(1)} \end{bmatrix} \quad U_2 U_1 X = \begin{bmatrix} x_{11}^{(1)} & x_{12}^{(1)} & x_{13}^{(1)} & \cdots & x_{1m}^{(1)} \\ 0 & x_{22}^{(2)} & x_{23}^{(2)} & \cdots & x_{2m}^{(2)} \\ 0 & 0 & x_{33}^{(2)} & \cdots & x_{3m}^{(2)} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & x_{n3}^{(2)} & \cdots & x_{nm}^{(2)} \end{bmatrix}$$

Householder法のメリット

1. 精度が良い（倍精度に相当）
 2. 計算効率的
 3. 上三角行列がモデル推定に必要な全情報を持つ
 4. 時系列モデリングに便利
 5. データの併合・モデルの併合等が容易
- 時系列モデルの最小二乗法に利用される

部分回帰モデル

y_n : 目的変数

x_{nj} : 説明変数 ($j = 1, \dots, m$)

$\{x_{n,j_1}, \dots, x_{n,j_k}\}$: 部分回帰変数

$$y_n = \sum_{i=1}^k a_{j_i} x_{n,j_i} + \varepsilon_n, \quad \varepsilon_n \sim N(0, \sigma^2)$$

次数 k

パラメータ $\theta = (a_{j_1}, \dots, a_{j_k}, \sigma^2)$

次数 k には ${}_m C_k$ 個の部分回帰モデルがある

部分回帰モデルの推定

(ℓ_1, \dots, ℓ_m) : 説明変数の優先順位

(j_1, \dots, j_m) : (ℓ_1, \dots, ℓ_m) の逆関数 ($\ell_{j_i} = i$)

左から順に非零要素の数を $(j_1, \dots, j_m, m+1)$ にする.

例

$$S \xrightarrow{H} T = \begin{bmatrix} t_{11} & t_{12} & t_{13} & t_{14} & t_{15} \\ t_{21} & 0 & t_{23} & t_{24} & t_{25} \\ t_{31} & 0 & t_{33} & 0 & t_{35} \\ t_{41} & 0 & 0 & 0 & t_{45} \\ 0 & 0 & 0 & 0 & t_{55} \end{bmatrix}$$

部分回帰モデルの推定

例： x_2, x_4, x_3, x_1 の順に採用する場合

$$T = \begin{bmatrix} t_{11} & t_{12} & t_{13} & t_{14} & t_{15} \\ t_{21} & 0 & t_{23} & t_{24} & t_{25} \\ t_{31} & 0 & t_{33} & 0 & t_{35} \\ t_{41} & 0 & 0 & 0 & t_{45} \\ 0 & 0 & 0 & 0 & t_{55} \end{bmatrix}$$

$$\hat{\sigma}_j^2(\ell_1, \dots, \ell_k) = \frac{1}{N} \sum_{i=k+1}^{m+1} t_{j,m+1}^2$$

$$\text{AIC}(\ell_1, \dots, \ell_k) = N(\log 2\pi \hat{\sigma}_j^2(\ell_1, \dots, \ell_k) + 1) + 2(k+1)$$

$$\begin{bmatrix} t_{1,\ell_1} & \cdots & t_{1,\ell_k} \\ & \ddots & \vdots \\ 0 & & t_{k,j_k} \end{bmatrix} \begin{bmatrix} b_{\ell_1} \\ \vdots \\ b_{\ell_k} \end{bmatrix} = \begin{bmatrix} t_{1,m+1} \\ \vdots \\ t_{k,m+1} \end{bmatrix}$$

説明変数(m)と部分回帰モデルの数(2^m)

m	2^m
5	32
10	1,024
15	32,768
20	1,048,576
25	33,554,432
30	1,073,741,824
35	34,359,738,368
40	1,099,511,627,776

- $m > 35$ 以上ではあまり実用的でない
- 多変量時系列モデルはもう少し構造を入れて探索する

.

Hocking-Leslie アルゴリズム

$$\sigma_1^2 \leq \sigma_2^2 \leq \dots \leq \sigma_m^2$$

k 個以下の変数の部分集合を削除したモデルの残差分散が σ_{k+1}^2 以下ならば, $k+1$ 個以上の変数をひとつでも削除したモデルの残差分散がそれより小さくなることはない.

実際に推定したモデル数

$m=10$ の場合

k	Number of models		
	AIC	ALL	H-L
0	521.0	1	1
1	127.3	10	10
2	110.1	45	9
3	95.1	120	36
4	94.7	210	84
5	95.3	252	21
6	95.8	210	35
7	97.8	120	10
8	99.7	45	6
9	101.7	10	10
10	103.7	1	1
		1024	223

$m=20$ の場合

k	Number of Models					
	ALL	H-L	H-L(2)	H-L(3)	H-L(4)	
0	1	1	1	1	1	
1	20	20	20	20	20	
2	190	19	19	19	19	
3	1140	171	171	171	171	
4	4845	969	969	514	514	
5	15504	3876	3876	1601	1535	
6	38760	11628	7260	3620	3256	
7	77520	27132	9023	4315	1021	
8	125970	6188	1832	555	140	
9	167960	4368	2366	585	85	
10	184756	3003	1056	205	73	
11	167960	2002	46	19	19	
12	125970	495	53	29	21	
13	77520	330	71	13	12	
14	38760	84	13	7	7	
15	15504	21	6	6	6	
16	4845	15	6	6	4845	
17	1140	10	5	1140	1140	
18	190	6	190	190	190	
19	20	20	20	20	20	
20	1	1	1	1	1	
		1048576	60359	27004	13037	13096
			5.7%	2.6%	1.2%	1.2%

実際に推定したモデル数 ($m=30$ の場合)

k	Number of Models				
	ALL	H-L	H-L(2)	H-L(3)	H-L(4)
0	1	1	1	1	1
1	30	30	30	30	30
2	435	29	29	29	29
3	4060	406	406	406	406
4	27405	3654	3654	3654	3654
5	142506	2371	23751	23751	11101
6	593775	118755	118755	65625	52975
7	2035800	475020	475020	244690	179516
8	5852926	1560780	1560780	672751	416592
9	14307150	4292145	2072070	1456774	474832
10	30045015	10015005	5328180	1100320	220510
11	54627300	20030010	4263050	1049505	272794
12	86493225	34597290	2771131	1127403	224052
13	119759850	21474180	3860990	811995	114793
14	145422675	30421755	5327239	411535	76331
15	155117520	17383860	2163454	233228	43745
16	145422674	9657700	1360306	125643	18592
17	119759850	5200300	613225	30823	2308
18	86493225	2704156	108418	3482	171
19	54627300	2496144	25235	557	239
20	30045015	646646	2121	353	85
21	14307150	167960	372	159	46
22	5852925	43758	407	91	31
23	2035800	1716	205	31	14
24	593775	210	19	13	9
25	142506	126	7	6	6
26	27405	15	9	8	27405
27	4060	10	7	4060	4060
28	435	10	435	435	435
29	30	30	30	30	30
30	1	1	1	1	1
	1073741824	161315453	30079337	7367489	2144793
		15.02%	2.80%	0.69%	0.20%

AICによるモデル選択上の注意

1. AICの絶対値には意味がないが、AICの差に意味がある。スケール変換してもAICの差は変化しない。

2. AICは小数点以下2桁程度の表示で十分

$$\exp\left\{-\frac{1}{2} \text{AIC}_k\right\} \approx \text{相対事後確率}$$

3. AICが最小点以上でほぼ2ずつ増加する場合

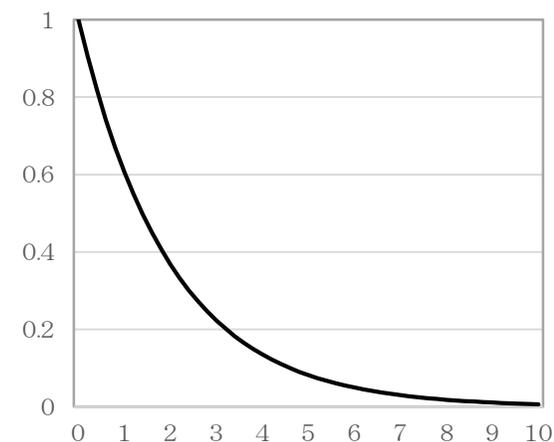
✓ モデル族が真のモデルを含む

4. 次数の増加に伴ってAICが減少し続ける場合

✓ モデル族が真のモデルを含まない

例：ARMA時系列にARモデルを適用

モデルの相対事後確率



4.の場合

● $g(y)$

● $g(y)$

$\{f(y|\theta), \theta \in \Theta\}$

今後の予定

1	4/10	時系列の前処理と可視化	イントロダクション, 時系列の前処理
2	4/17		共分散関数, スペクトルとピリオドグラム
3	4/24	モデリング	統計的モデリング・情報量規準AIC
4	5/8		最尤法, 最小二乗法, ベイズモデル
5	5/15	定常時系列モデル	ARMAモデルによる時系列の解析
6	5/22		ARモデルの推定・応用
7	5/29		局所定常ARモデル, 統計的制御
8	6/5	状態空間モデル	状態空間モデルによる時系列の解析
9	6/12		ARMAモデルの推定, トレンドの推定
10	6/19		季節調整モデル
11	6/26		ボラティリティ, 時変係数ARモデル
12	7/3	非線形・非ガウス型モデル	一般状態空間モデル, ガウス和フィルタ
13	7/10		粒子フィルタ, シミュレーション