時系列解析（４）

1．情報量規準AIC（続）
2．AICによるモデル選択（例）
3．回帰モデルと最小二乗法

－ 配布用 －

東京大学 数理・情報教育研究センター
北川源四郎

記号と準備（復習）

$g(x)$　真のモデル
$f(x\mid\theta)$　パラメトリックモデル

θ^*　「真」の値

$$E_Y \log f(Y\mid\theta^*) = \max E_Y \log f(Y\mid\theta)$$

$$\Rightarrow \tfrac{\partial}{\partial\theta} E_Y \log f(Y\mid\theta^*) = 0$$

$\hat{\theta}$　最尤推定値

$$\sum_{i=1}^{n} \log f(x_i\mid\hat{\theta}) = \max \sum_{i=1}^{n} \log f(x_i\mid\theta)$$

$$\Rightarrow \sum_{i=1}^{n} \tfrac{\partial}{\partial\theta} \log f(Y\mid\hat{\theta}) = 0$$

$I(\theta)$：Fisher情報行列，$J(\theta)$：Expected Hessian

$$I(\theta) \equiv E\left\{\left(\tfrac{\partial}{\partial\theta}\log f(Y\mid\theta)\right)\left(\tfrac{\partial}{\partial\theta}\log f(Y\mid\theta)\right)^T\right\}$$

$$J(\theta) = -E\left\{\tfrac{\partial^2}{\partial\theta\partial\theta^T} f(Y\mid\theta)\right\}$$

$g(x)=f(x\mid\theta_0)$，$\theta_0\in\Theta$　が存在しない場合でも

$$\hat{\theta}_n \to \theta_0$$

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \to N(0, J(\theta_0)^{-1} I(\theta_0) J(\theta_0)^{-1})$$

$$E\left[(\hat{\theta}-\theta_0)^T J(\theta_0)(\hat{\theta}-\theta_0)\right]$$

$$= \mathrm{tr}\left\{J(\theta_0) E\left[(\hat{\theta}-\theta_0)(\hat{\theta}-\theta_0)^T\right]\right\}$$

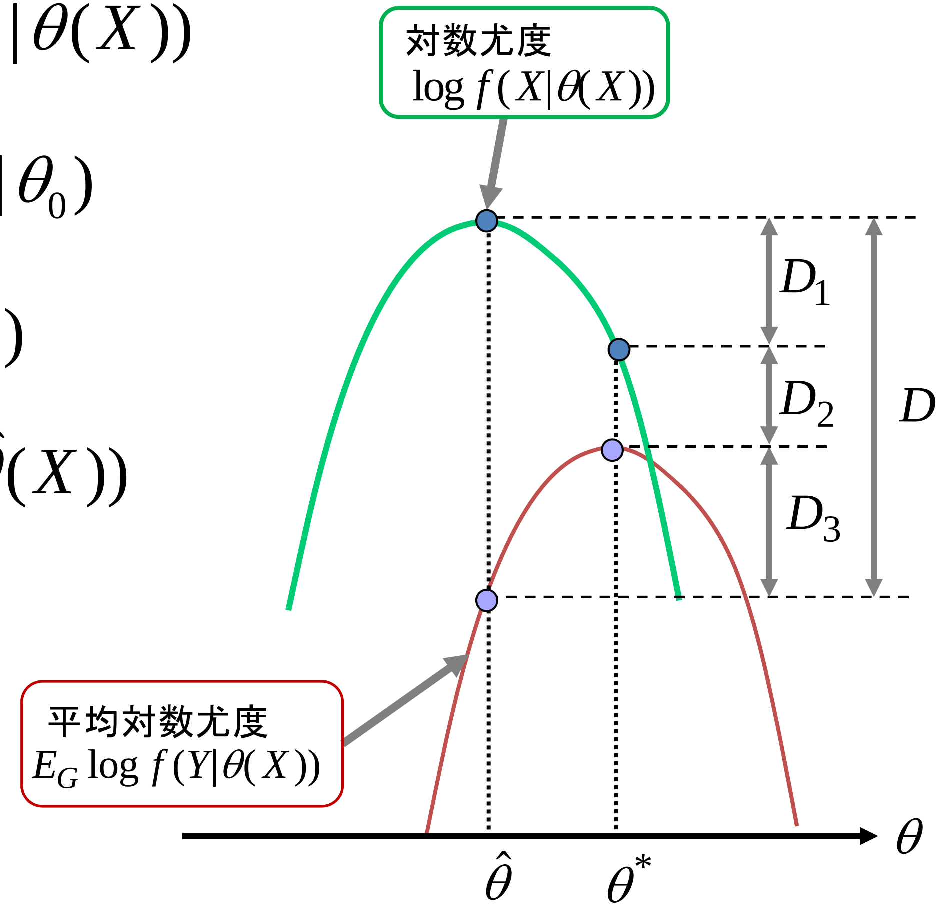
$$= \mathrm{tr}\left\{J(\theta_0)\tfrac{1}{n}J(\theta_0)^{-1} I(\theta_0) J(\theta_0)^{-1}\right\}$$

$$= \tfrac{1}{n}\mathrm{tr}\left\{I(\theta_0) J(\theta_0)^{-1}\right\}$$

バイアスの評価

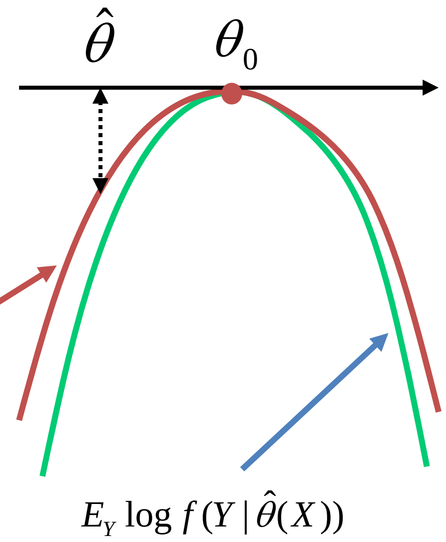
対数尤度　　平均対数尤度

$$
\begin{aligned}
D &= \frac{1}{n}\log f(X\,|\,\hat{\theta}(X)) - E_Y \log f(Y\,|\,\hat{\theta}(X)) \\
&= \frac{1}{n}\log f(X\,|\,\hat{\theta}(X)) - \frac{1}{n}\log f(X\,|\,\theta_0) \\
&+ \frac{1}{n}\log f(X\,|\,\theta_0) - E_Y \log f(Y\,|\,\theta_0) \\
&+ E_Y \log f(Y\,|\,\theta_0) - E_Y \log f(Y\,|\,\hat{\theta}(X)) \\
&= D_1 + D_2 + D_3
\end{aligned}
$$



D_1, D_2, D_3 の評価

$$
\begin{aligned}
&E_Y \log f(Y \mid \hat{\theta}(X)) \\
&\approx E_Y \log f(Y \mid \theta_0) + \underbrace{\tfrac{\partial}{\partial \theta} E_Y \log f(Y \mid \theta_0)}_{=0}(\hat{\theta}-\theta_0) \\
&= \qquad + \tfrac{1}{2}(\hat{\theta}-\theta_0)^T \boxed{E_Y \tfrac{\partial^2}{\partial\theta\partial\theta} \log f(Y \mid \theta_0)}(\hat{\theta}-\theta_0) \\
&= E_Y \log f(Y \mid \theta_0) - \tfrac{1}{2}(\hat{\theta}-\theta_0)^T \boxed{J(\theta_0)}(\hat{\theta}-\theta_0)
\end{aligned}
$$



$$E_X\{D_3\} = E_X\left\{E_Y \log f(Y \mid \theta_0) - E_Y \log f(Y \mid \hat{\theta}(X))\right\} \approx \frac{1}{2n} tr\left\{IJ^{-1}\right\}$$

$$E_X\{D_1\} = E_X\left\{\log f(X \mid \hat{\theta}(X)) - \log f(X \mid \theta_0)\right\} \approx \frac{1}{2n} \mathrm{tr}\left\{IJ^{-1}\right\}$$

$$E_X\{D_2\} = E_X\left\{\frac{1}{n}\log f(X \mid \theta_0) - E_Y \log f(Y \mid \theta_0)\right\} = 0$$

情報量規準

$$\text{IC} = -2\log f(x\,|\,\hat{\theta}) + 2b(G)$$

竹内(1976)

$$b(G) = E[D_1] + E[D_2] + E[D_3] = \text{tr}\left\{I(G)J(G)^{-1}\right\}$$

$$\text{TIC} = -2\log f(x\,|\,\hat{\theta}) + 2\text{tr}\{I(G)J(G)^{-1}\}$$

$\exists\theta\in\Theta$ such that $g(x) = f(x\,|\,\theta_0) \quad \Rightarrow \quad I(\theta_0) = J(\theta_0)$

$$b(G) = \text{tr}\{I(\theta_0)J(\theta_0)^{-1}\} = \text{tr}\{I_k\} = k$$

$$\text{AIC} = -2\log f(X\,|\,\hat{\theta}(X)) + 2k$$

k:　パラメータ数（θの次元）

その他の情報量規準

AIC$_c$　有限修正

$$b(G) = \frac{n(p+1)}{n-p-2}$$

GIC　統計的汎函数で定義される任意の推定量

$$b(G) = \mathrm{tr}\left\{ \int T^{(1)}(x;G) \frac{\partial \log f(x|\theta)}{\partial \theta} dG(x) \right\}$$

EIC　Bootstrap法によるバイアス推定

$$b^*(G) = \frac{1}{n} E_{X^*}\left\{ \log f(X^* \mid \hat{\theta}(X^*)) - \log f(X \mid \hat{\theta}(X^*)) \right\}$$

ABIC　ベイズ型情報量規準

$$\mathrm{ABIC} == -2 \max \log \int f(x \mid \theta) \pi(\theta \mid \lambda) d\theta + 2q$$

参考書

- 坂元慶行, 石黒真木夫, 北川源四郎(1983). 「情報量統計学」, 共立出版, 情報科学講座　A.5.4
- Y.Sakamoto, M.Ishiguro and G.Kitagawa (1986) *Akaike Information Criterion Statistics,* D.Reidel, Dordrecht.
- Burnham, K. P., & Anderson, D. R. (2003). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer.
- 小西貞則，北川源四郎(2004)「情報量規準」，朝倉書店，予測と発見の科学２
- 竹内・下平・伊藤・久保川(2004)：モデル選択，統計科学のフロンティア，岩波書店
- 赤池弘次・甘利俊一・北川源四郎・樺島祥介・下平英寿，編者 室田一雄・土谷隆(2007)「赤池情報量規準AIC－モデリング・予測・知識発見」共立出版
- S. Konishi and G. Kitagawa (2008). *Information Criteria and Statistical Modeling*, Springer Verlag

関連論文リスト

- **Akaike, H. (1973), “**Information theory and an extension of the maximum likelihood principle.” *Proc. 2nd International Symposium on Information Theory* , B. N. Petrov and F. Csaki eds., Akademiai Kiado, Budapest, 267-281.
- **Akaike, H. (1974),** “A new look at the statistical model identification.” *IEEE Trans. Automat. Contrl*., AC-19, No. 6, 716-723.
- 竹内啓, (1976). 情報統計量の分布とモデルの適切さの規準, < 特集> 情報量規準. 数理科学, 14(3), 12-18.
- **Konishi and Kitagawa (1996)**, “Generalized Information Criteria in Model Selection”, *Biometrika*, Vol. 83, No.4, 875-890.
- **Ishiguro, Sakamoto and Kitagawa (1997)**, “Bootstrapping Log Likelihood and EIC, an Extension of AIC”, *Annals of the Institute of Statistical Mathematics*, Vol. 49, No. 3, 411-434.

ヒストグラムのBin Size の決定

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

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

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

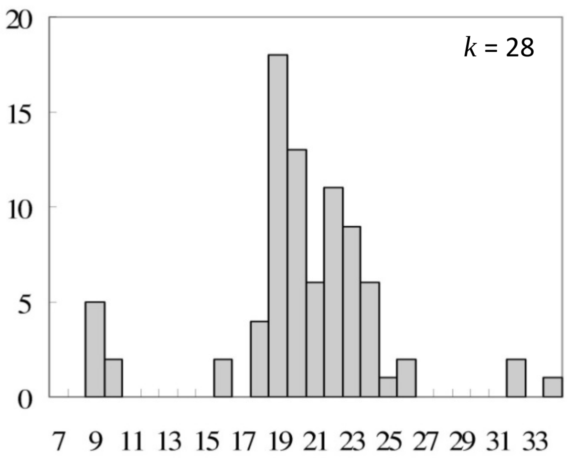
$$\mathrm{AIC}_k = (-2)\left\{C + \sum_{j=1}^{k} n_j \log\left(\frac{n_j}{j}\right)\right\} + 2(k-1)$$

Galaxy data (Roeder (1990))

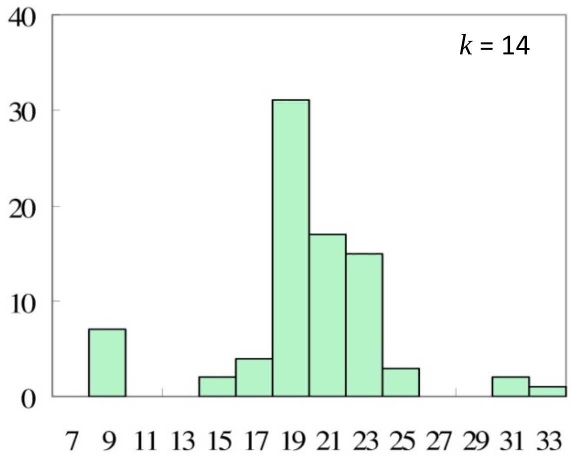
0 5 2 0 0 0 0 0 2 0 4 18 13 6
11 9 6 1 2 0 0 0 0 0 2 0 1 0

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

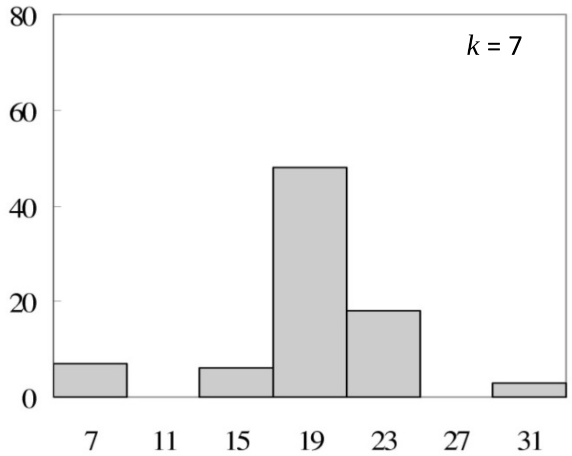
Histogram of galaxy data



Best



Too small



モデル選択例：分布の形状の選択

$$f(y \mid \mu, \tau^2, b) = \frac{C}{\left(y^2 + \tau^2\right)^b}$$

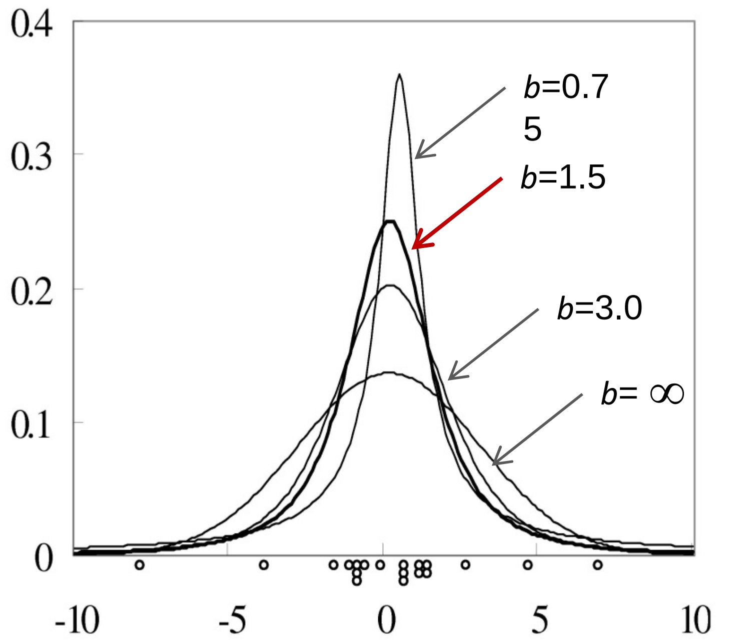
Pearson's family of distributions
Select the shape parameter *b*

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

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

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

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

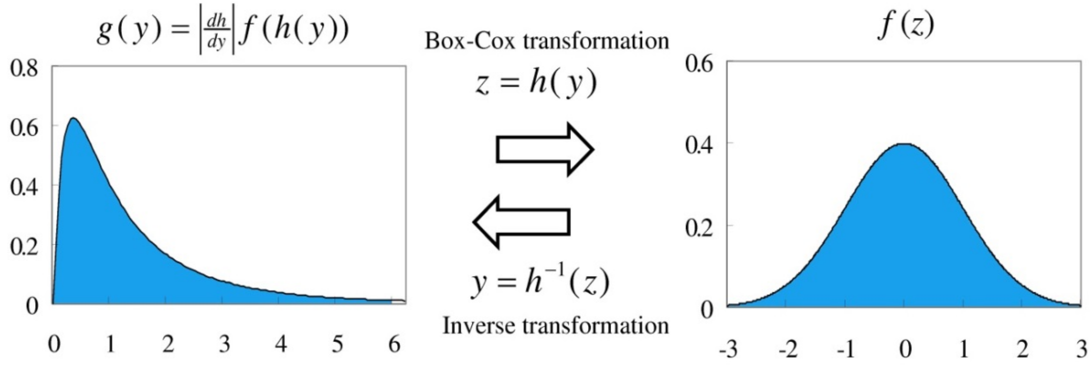


Box-Cox 変換のパラメータ決定

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

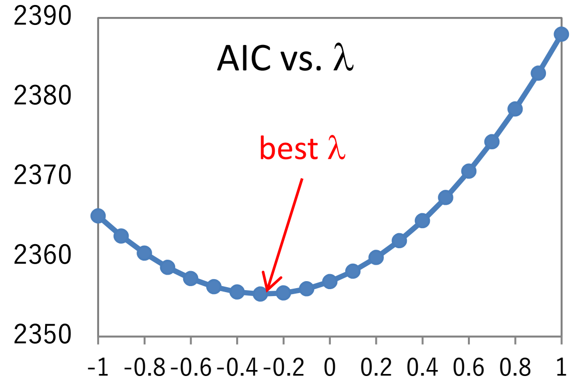
$$g(y) = \left| \frac{dh_\lambda}{dy} \right| f(h(y))$$

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

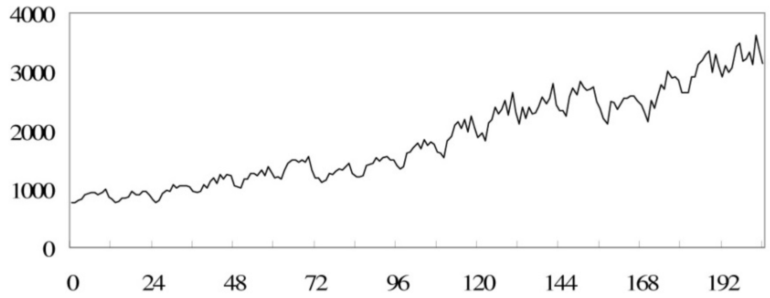


Jacobian of the transformation

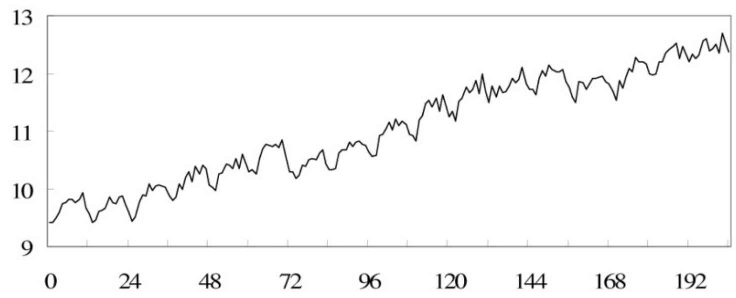
λ	-1.0	-0.5	-0.3	0.0	0.3	0.5	1.0
log-L	1030.8	482.4	-261.7	-70.8	-405.0	-628.9	-1191.9
AIC	-2057.6	-960.8	-519.5	145.5	814.1	1261.7	2387.9
AIC'	2365.1	2356.2	2355.3	2356.9	2362.0	2367.4	2387.9



Original WHARD data (US BLS)

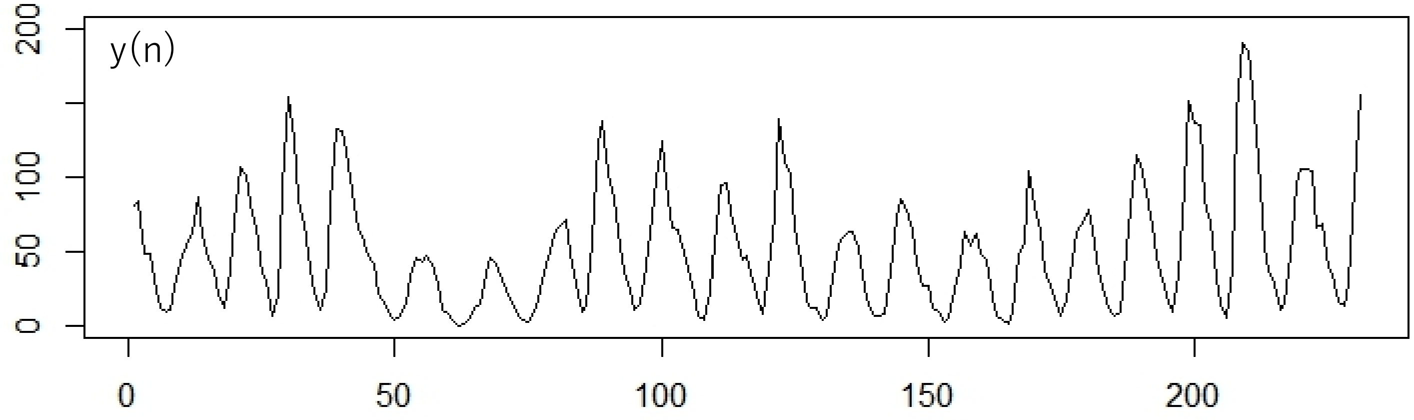


Best Box-Cox transformation (λ=0.1)

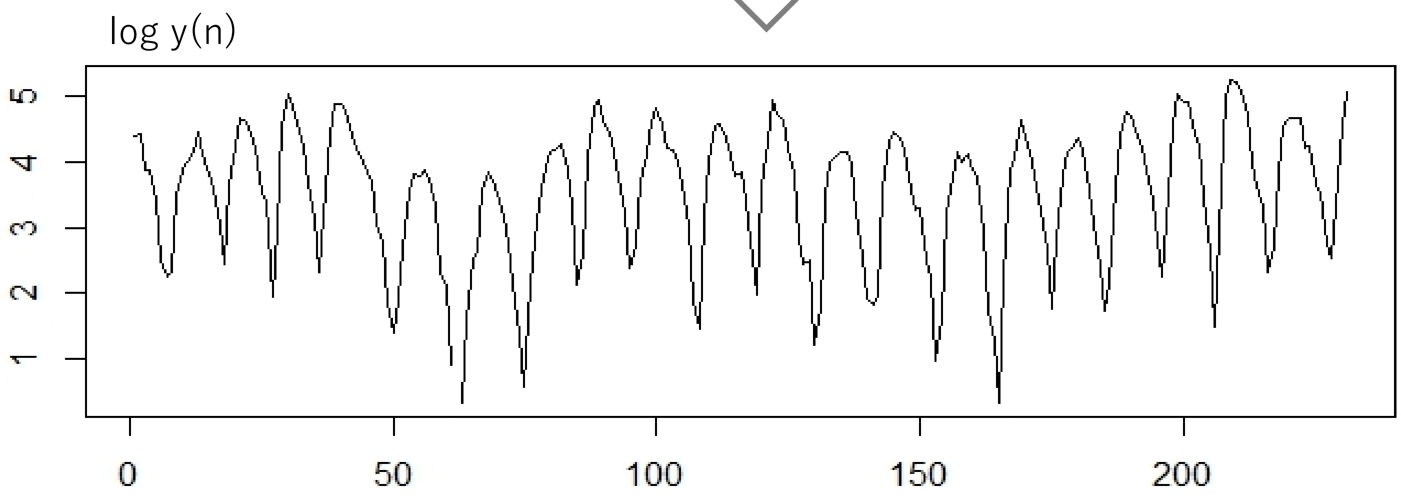


太陽黒点数データ

```
plot(sunspot,ylim=c(0,200))
y <- log( sunspot )
plot(y)
```



$$y_n = \log x_n$$



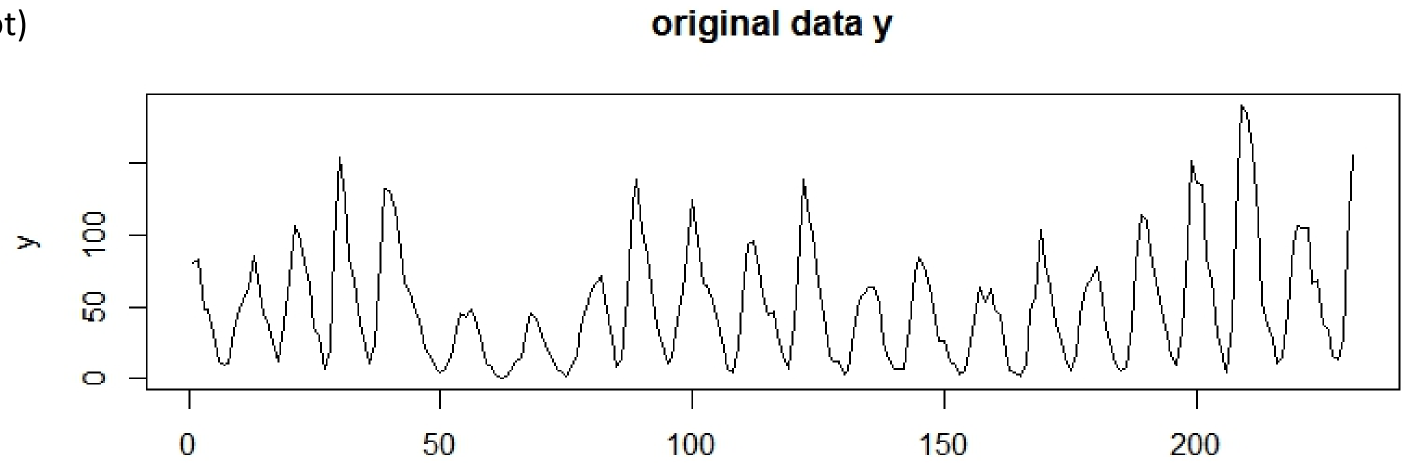
AIC'による変換パラメータの選択

data(Sunspot) # Sun spot number data
boxcox(Sunspot)

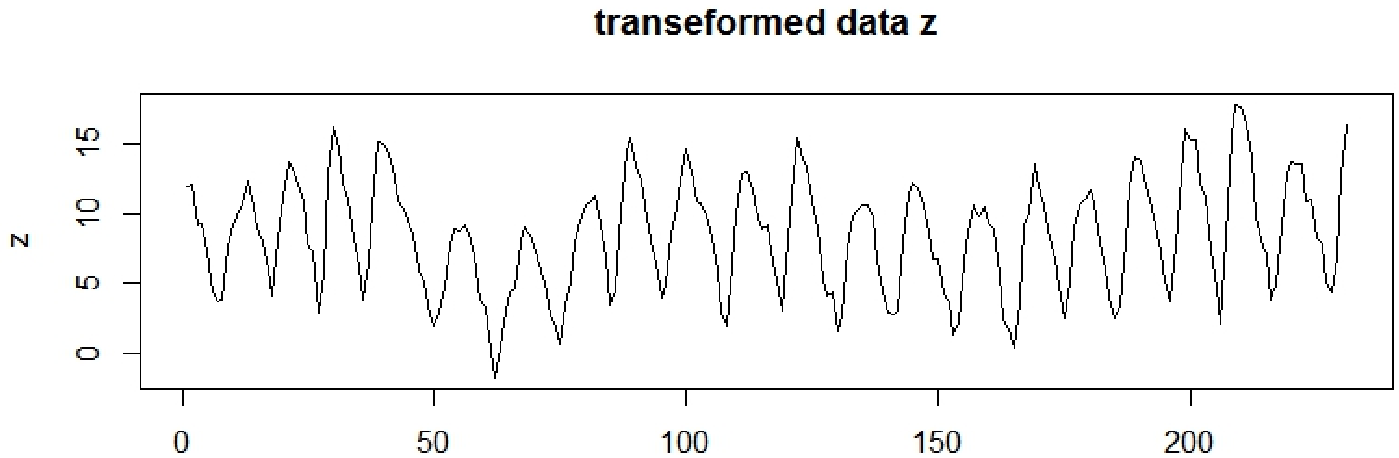
lambda	aic'	LL'	aic	LL	mean	variance
1.00	2360.26	-1178.13	2360.26	-1178.13	4.909502e+01	1.575552e+03
0.90	2335.22	-1165.61	2174.47	-1085.24	3.545844e+01	7.049401e+02
0.80	2313.48	-1154.74	1991.98	-993.99	2.591126e+01	3.199262e+02
0.70	2295.33	-1145.66	1813.07	-904.54	1.917397e+01	1.474669e+02
0.60	2281.11	-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

data(Sunspot) # Sun spot number data
boxcox(Sunspot)



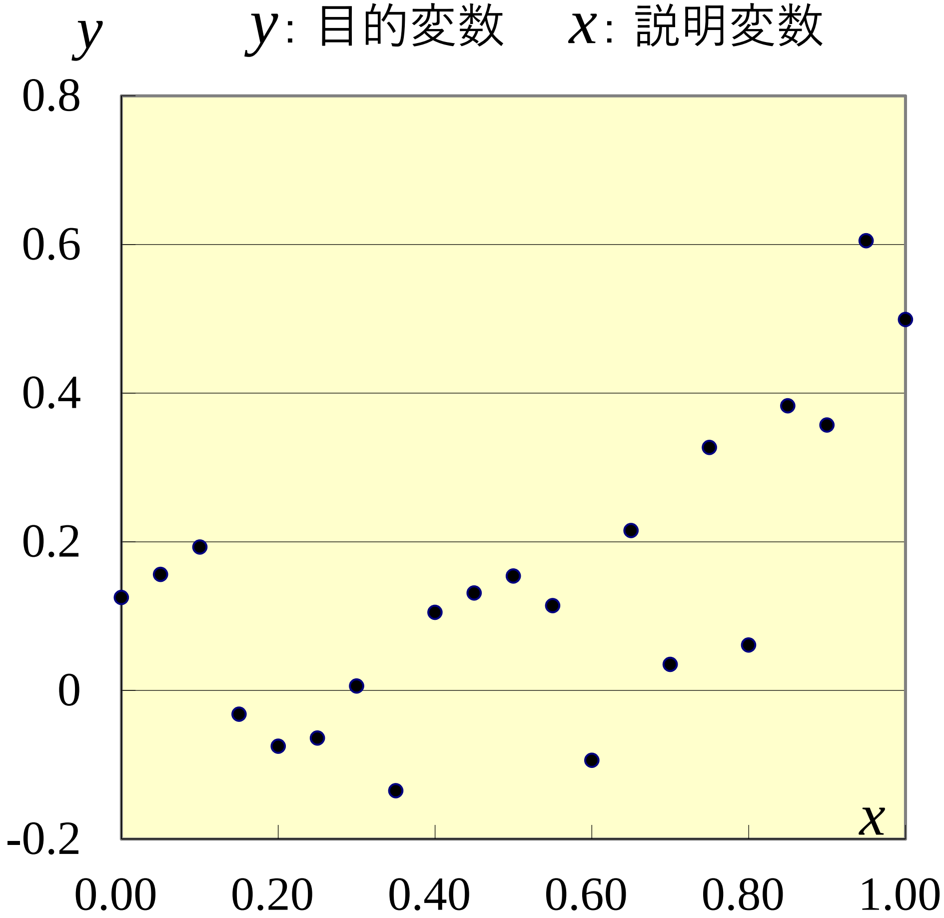
$$\lambda = 0.3$$



多項式回帰（例）

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

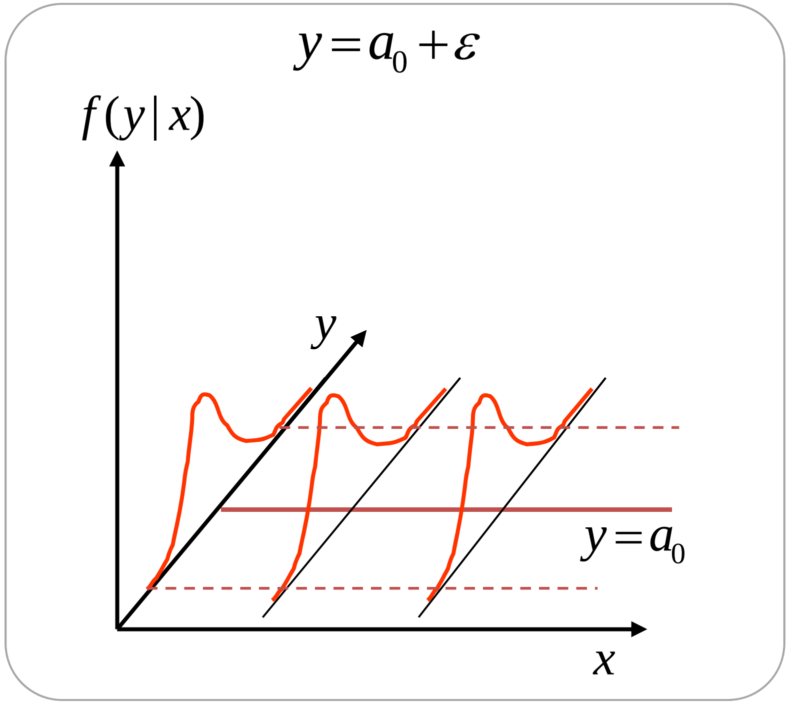


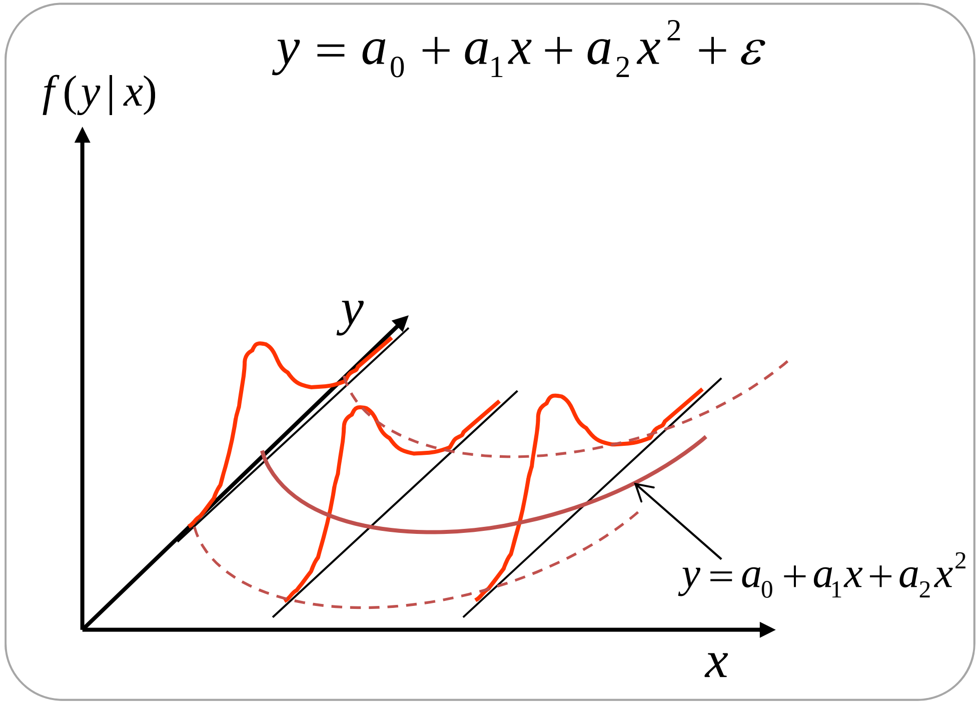
多項式回帰モデル

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

$$f(x) = a_0 + a_1 x + \cdots + a_m x^m, \qquad \varepsilon_i \sim N(0, \sigma^2)$$

$$y_i \sim N(\mu_i, \sigma^2), \ \mu_i = a_0 + a_1 x_i + \cdots + a_m x_i^m$$

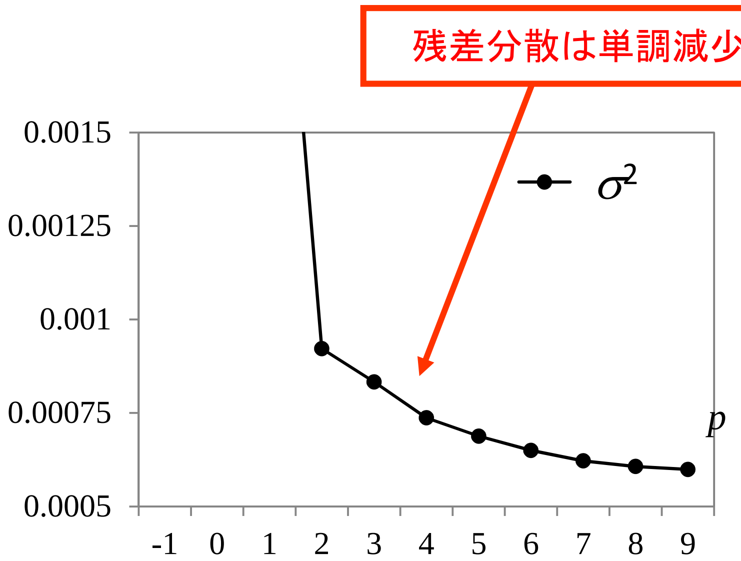


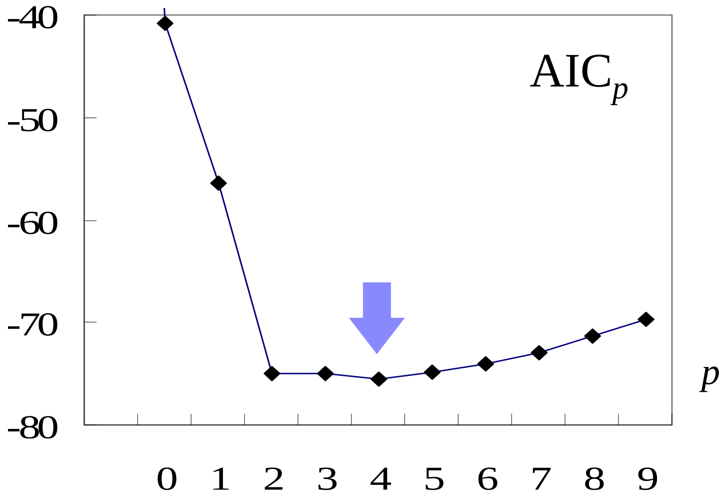


モデル選択

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

モデル選択例：多項式回帰の次数





$$y = \beta_0 + \beta_1 x + \cdots + \beta_p x^p + \varepsilon, \qquad \varepsilon \sim N(0, \sigma^2)$$

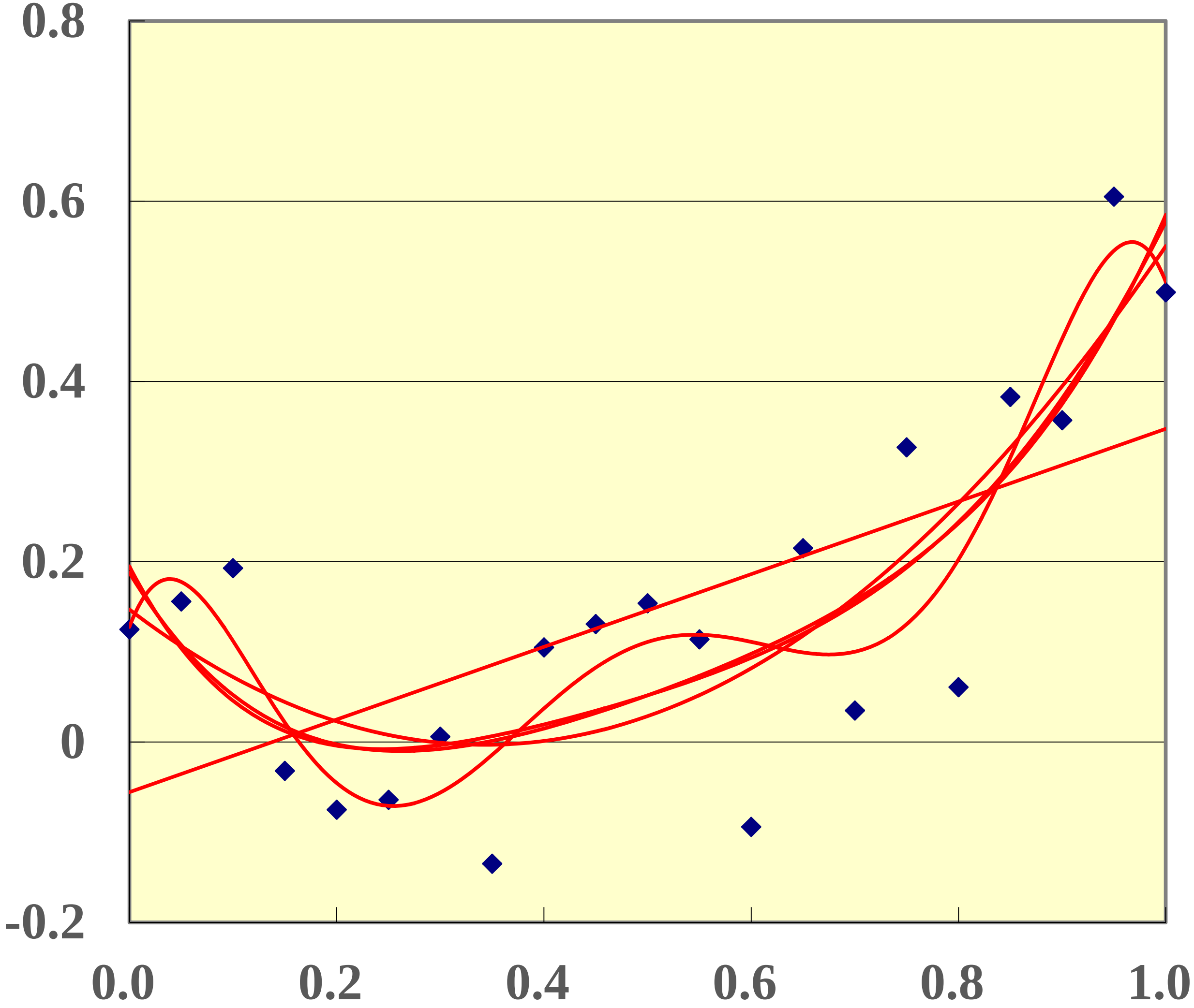
$$\theta = (\beta_0, \beta_1, \cdots, \beta_p, \sigma^2)$$

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

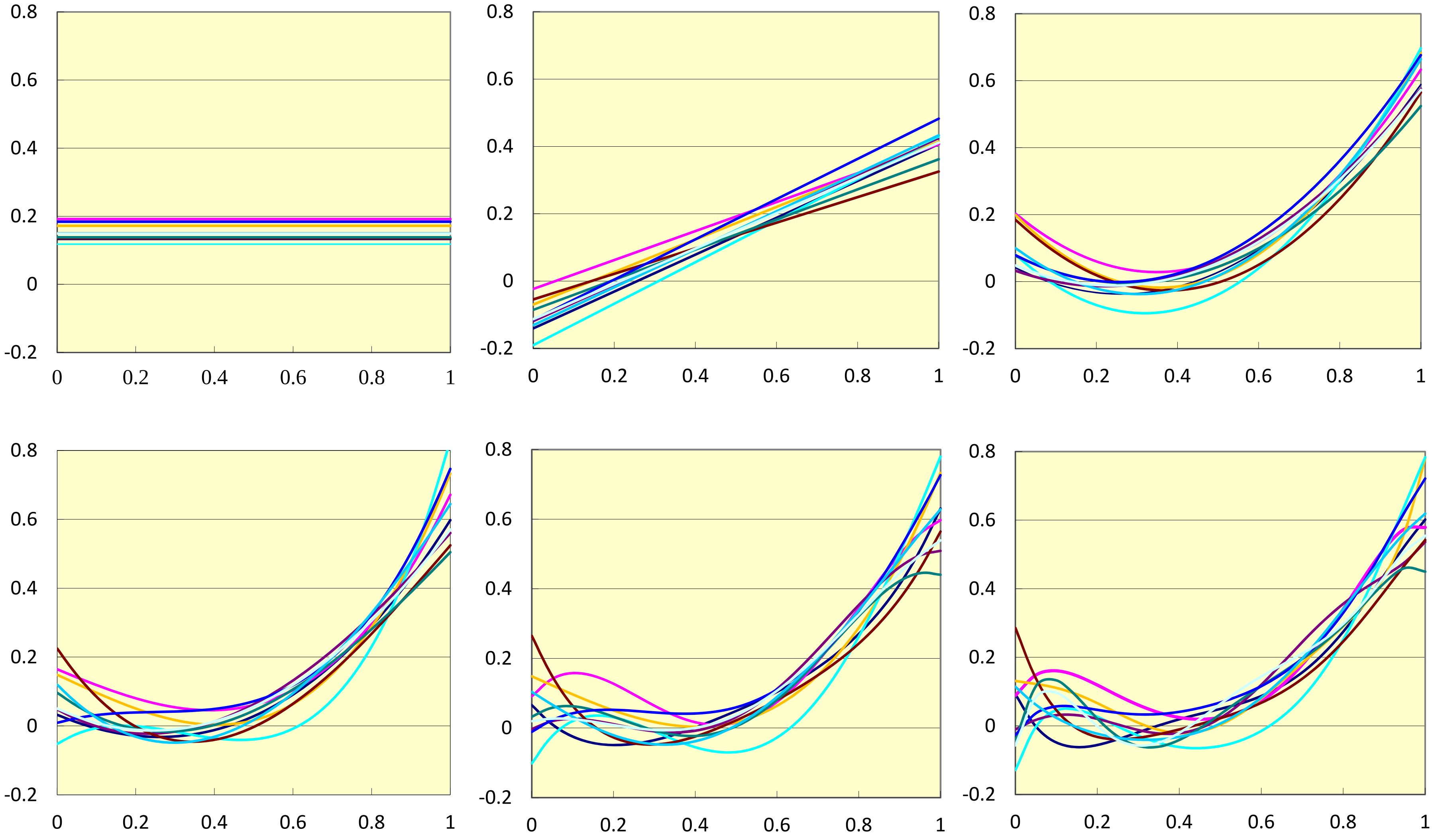
$$\ell(\hat{\theta}) = -\frac{n}{2}\log(2\pi\hat{\sigma}^2) - \frac{n}{2}$$

$$\mathrm{AIC}_p = n(\log 2\pi + 1) + n\log\hat{\sigma}^2 + 2(p+2)$$

Estimated Regression Curves



シミュレーション



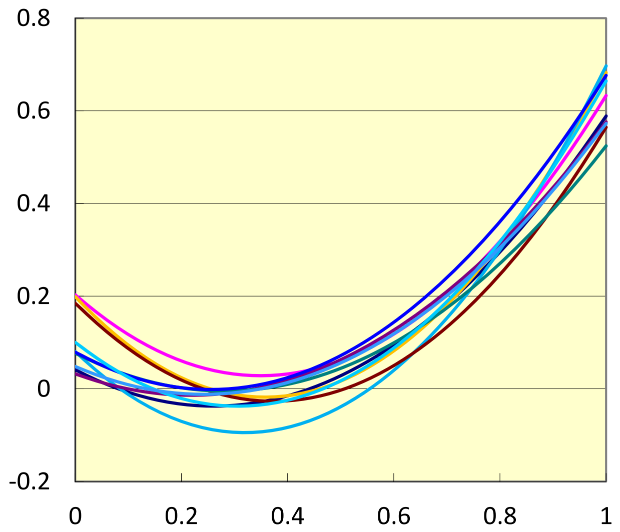
制約モデル

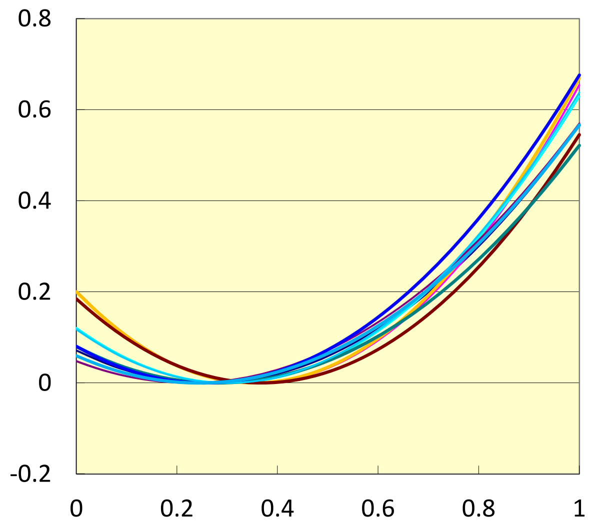
例えば回帰式の最小値が 0 となることを知っていたら

$$y_i = a(x_i + b)^2 + \varepsilon_i$$

$$\hat{a} = 1.1515, \quad \hat{b} = -0.3035, \quad \hat{\sigma}^2 = 0.00875$$

$$\text{AIC} = 21(\log 2\pi\hat{\sigma}^2 + 1 + \log 0.00875) + 2\times 3 = -33.92 \ < \ -32.13$$



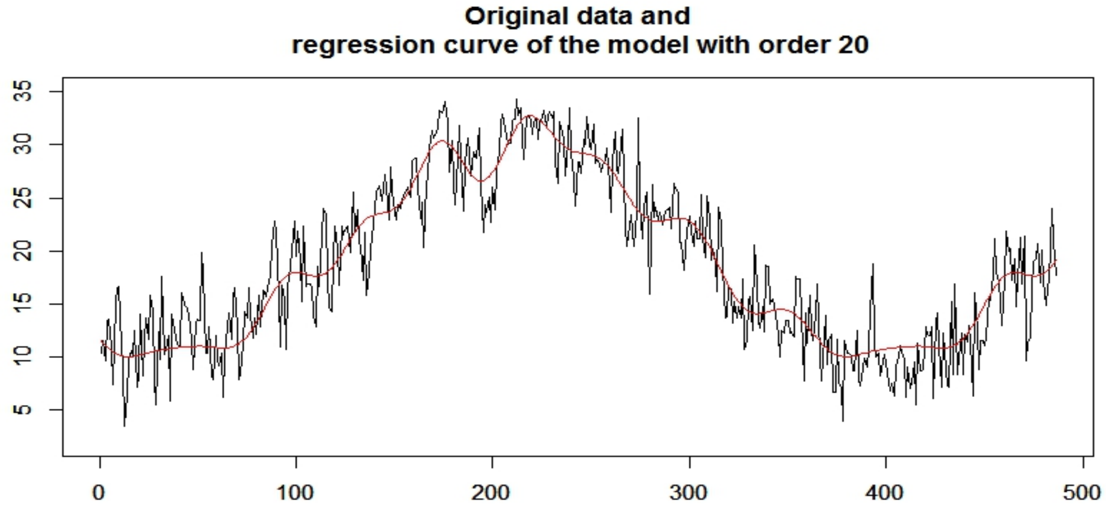


$$y_i = f(x_i) + \varepsilon_i \qquad \varepsilon_i \sim N(0, \sigma^2)$$

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

三角関数回帰モデル

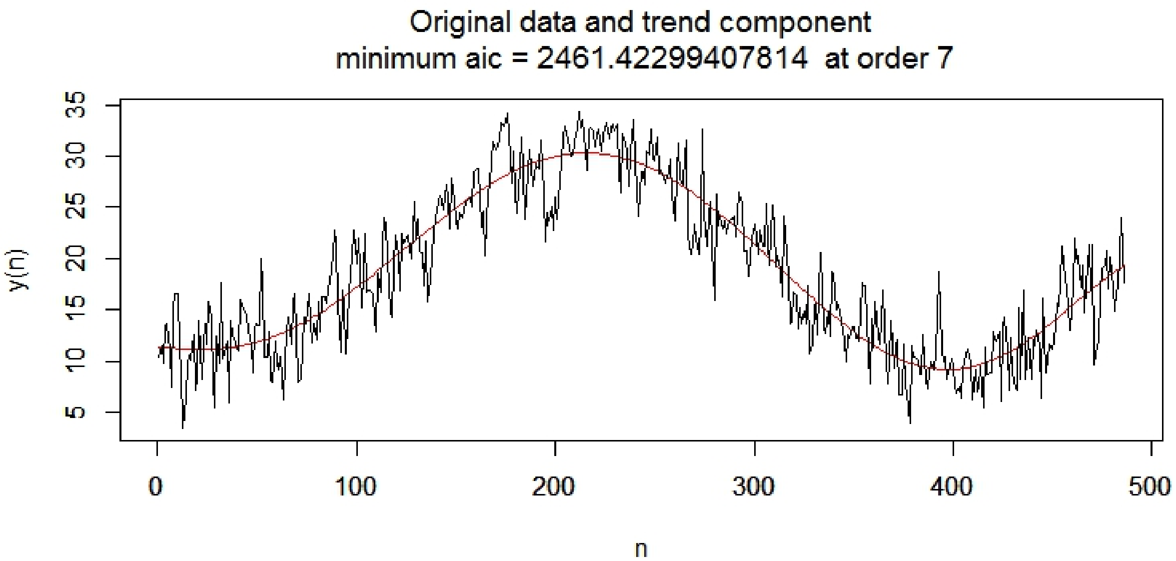
$$f(x) = a + \sum_{j=1}^{m} b_j \sin(jwn) + \sum_{j=1}^{\ell} c_j \cos(jwn)$$



polreg(TemperData,7)

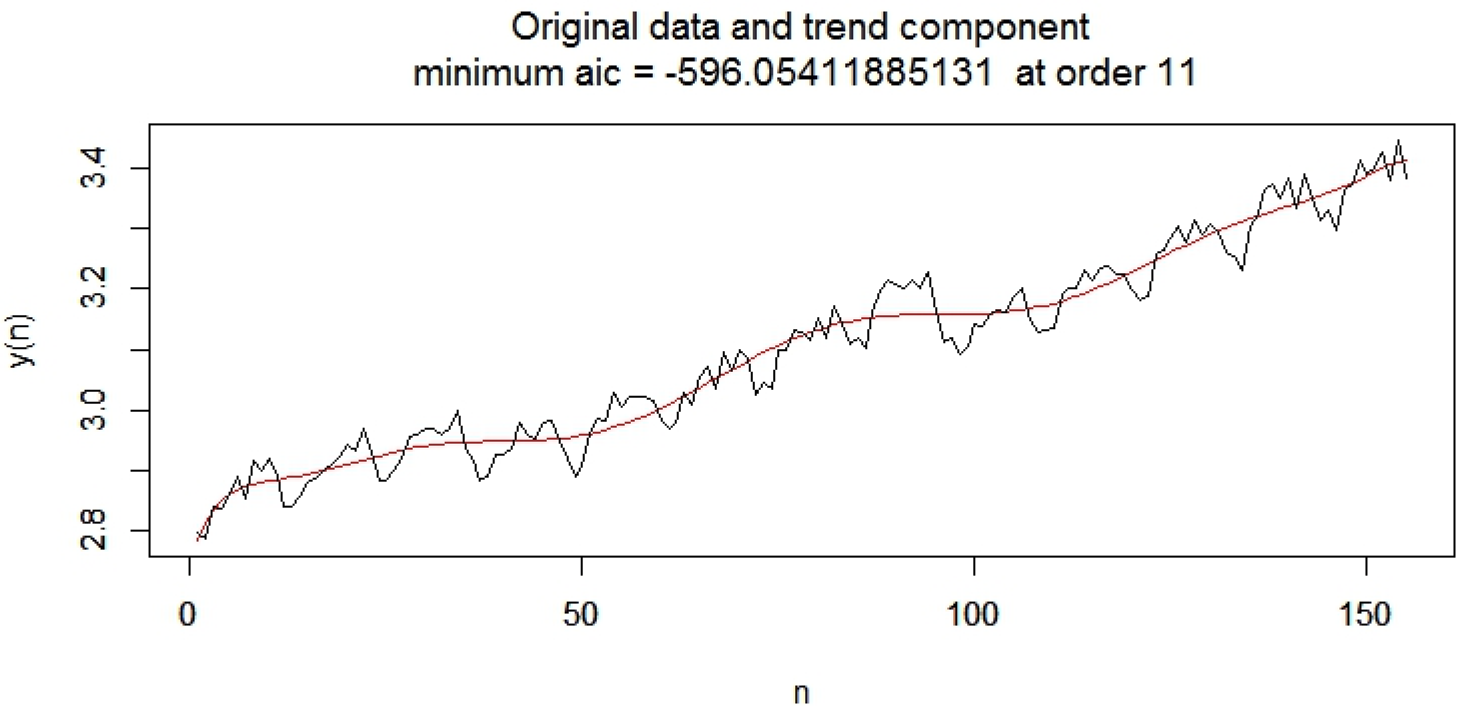
多項式回帰モデル

$$f(x) = a_0 + a_1 x + \cdots + a_m x^m$$



Whardデータ（多項式回帰）

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



重回帰モデル

都市名	気温 y	緯度 x1	経度 x2	標高 x3
稚内	−8.0	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	25.68	139.77	5.3
鳥取	0.5	35.48	134.23	7.1
京都	−0.6	25.02	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

Variable Selection for a Regressi

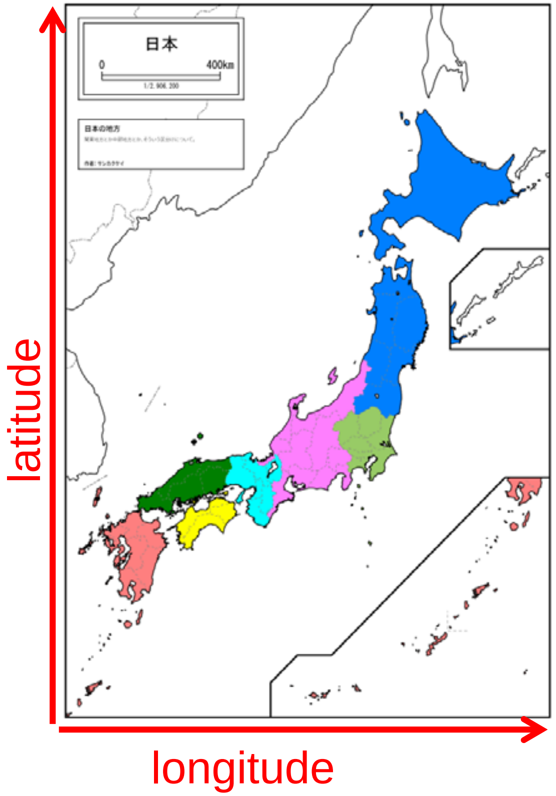
y_n: Temperature, x_{1n}: Latitude, x_{2n}: Longitude, x_{3n}: Altitude

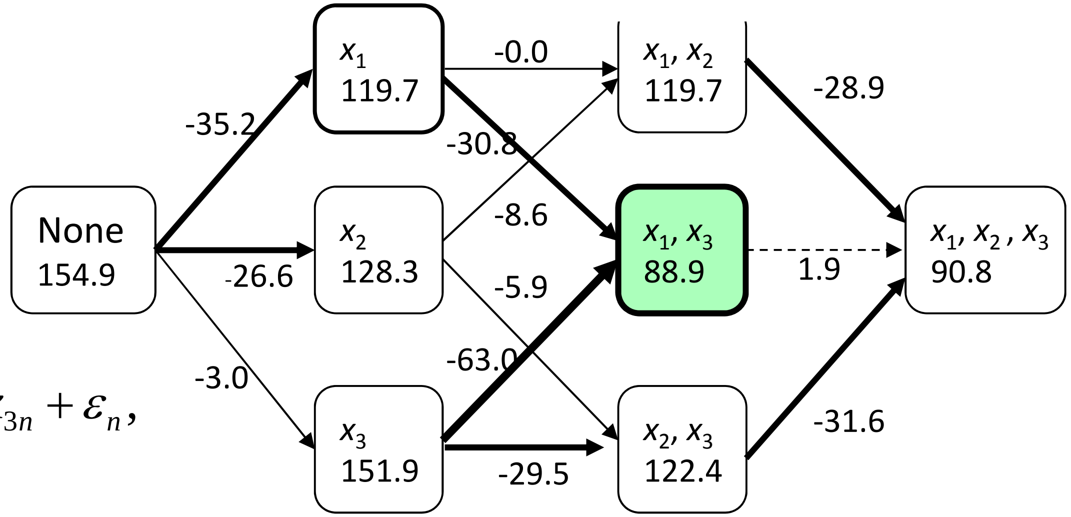
$$y_n = a_0 + a_1 x_{1n} + a_2 x_{2n} + a_3 x_{3n} + \varepsilon_n, \qquad \varepsilon_n \sim N(0, \sigma^2)$$

Select variables among x_1, x_2, x_3 appropriate to predict y_n

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

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





Selected model

$$y_n = 40.490 - 1.208 x_{1n} - 0.010 x_{3n} + \varepsilon_n,$$
$$\varepsilon_n \sim N(0, 1.490)$$

Householder法

U :　任意の直交変換　（ベクトルの長さを変えない）

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

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

$$X = \underbrace{[\, z \mid y \,]}_{m+1} \updownarrow N \quad \Rightarrow \quad UX = S = \begin{bmatrix} s_{11} & \cdots & s_{1m} & s_{1,m+1} \\ & \ddots & \vdots & \vdots \\ & & s_{mm} & s_{m,m+1} \\ & & & s_{m+1,m+1} \\ & 0 & & \end{bmatrix}$$

最小二乗法（Householder法）

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

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

最小二乗解

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

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

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

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

AICによる次数選択

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

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

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

for $k = 1,\ldots,m$

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

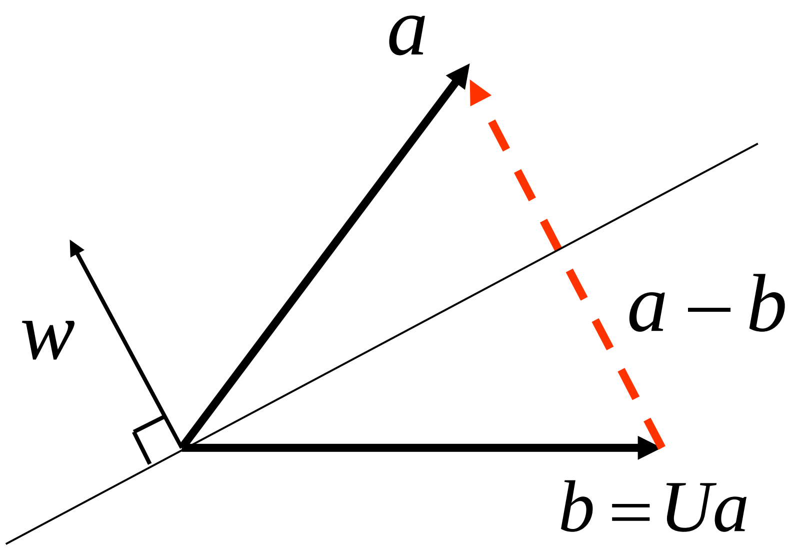
$$\hat{\sigma}_k^2 = \frac{1}{n}\left(s_{k+1,m+1}^2 + \cdots + s_{m+1,m+1}^2\right)$$

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

Householder変換

$$a - 2w \cdot (w^T a)$$

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



$$UU^T = (I - 2ww^T)(I - 2ww^T)^T$$
$$= I - 4ww^T + 4ww^T ww^T = I$$

Householder変換

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

$$w = \frac{a-b}{\|a-b\|}$$

$$\begin{aligned} Ua &= (I - 2ww^T)a \\ &= a - \frac{(a-b)(a-b)^T}{\|a-b\|^2} 2a \\ &= a - \frac{(a-b)(a-b)^T}{\|a-b\|^2} \{(a-b)+(a+b)\} \\ &= a - (a-b) - \frac{(a-b)(\|a\|^2 - \|b\|^2)}{\|a-b\|^2} \\ &= b \end{aligned}$$

Householder変換

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

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