時系列解析（３）

配布資料（2018/4/25）

東京大学 数理・情報教育研究センター
北川 源四郎

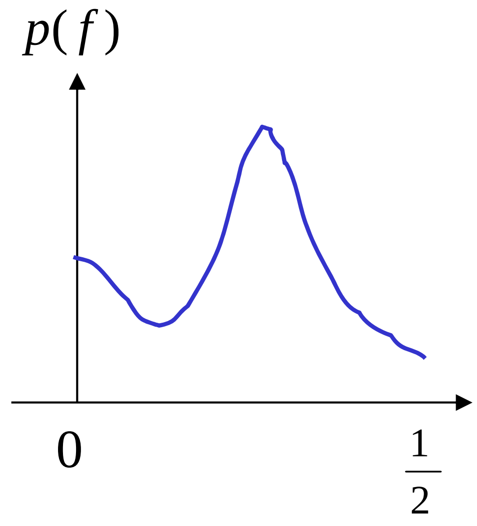
（パワー）スペクトル

C_k　**自己共分散関数**　$\sum_{k=-\infty}^{\infty} |C_k| < \infty$

$$p(f) = \sum_{k=-\infty}^{\infty} C_k e^{-2\pi ikf} \qquad -\frac{1}{2} \le f \le \frac{1}{2}$$

$$= C_0 + 2\sum_{k=1}^{\infty} C_k \cos 2\pi kf$$

$$p(-f) = p(f)$$



$$C_k = \int_{-\frac{1}{2}}^{\frac{1}{2}} p(f) e^{2\pi ikf} df = \int_{-\frac{1}{2}}^{\frac{1}{2}} p(f) \cos(2\pi kf) df$$

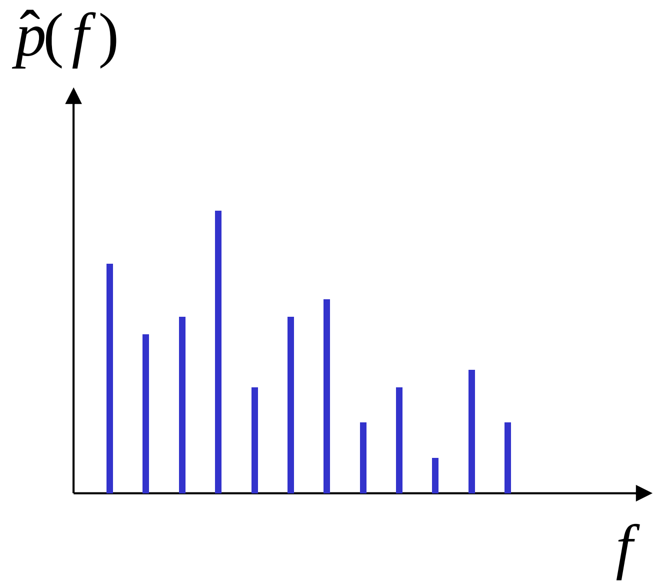
ピリオドグラム

$$y_1, \cdots, y_N \Rightarrow \hat{C}_0, \cdots, \hat{C}_{N-1}$$

$$\hat{p}(f) = \sum_{k=-N+1}^{N-1} \hat{C}_k e^{-2\pi ikf}$$

$$= \hat{C}_0 + 2\sum_{k=1}^{N-1} \hat{C}_k \cos(2\pi kf)$$

$$f = 0, \frac{1}{N}, \frac{2}{N}, \cdots, \frac{1}{2}$$



$$\hat{C}_k = \int_{-\frac{1}{2}}^{\frac{1}{2}} \hat{p}(f) e^{2\pi ikf} df \qquad k = 0, 1, \ldots, N-1$$

$$\hat{p}(f) = \sum_{k=-N+1}^{N-1} \hat{C}_k e^{-2\pi ikf} \qquad -\frac{1}{2} \le f \le \frac{1}{2}$$

ピリオドグラムの性質

（１）漸近的に不偏

$$\lim_{n\to\infty} E\left[\hat{p}(f)\right] = p(f)$$

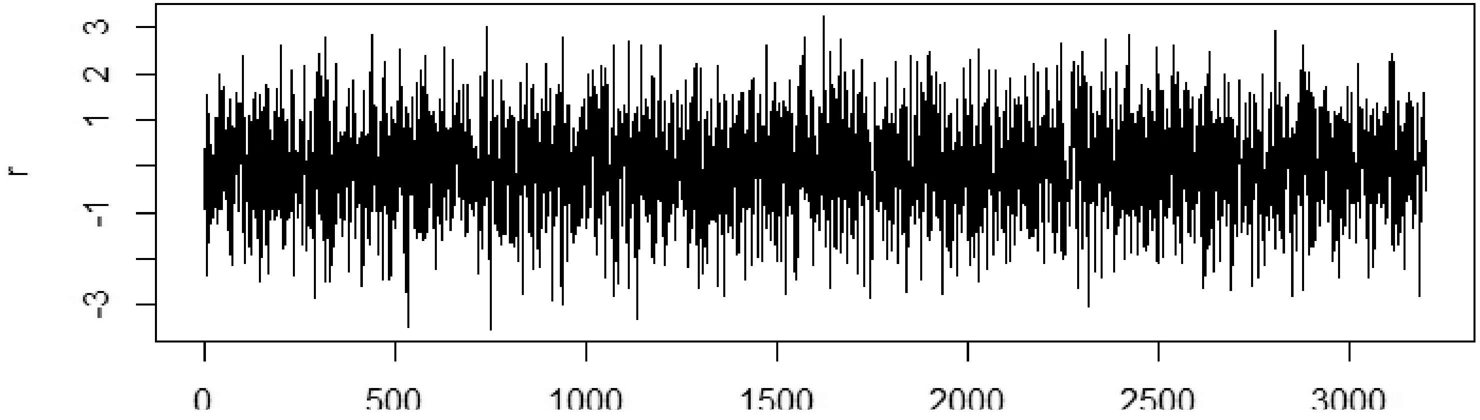
（２）一致推定量でない

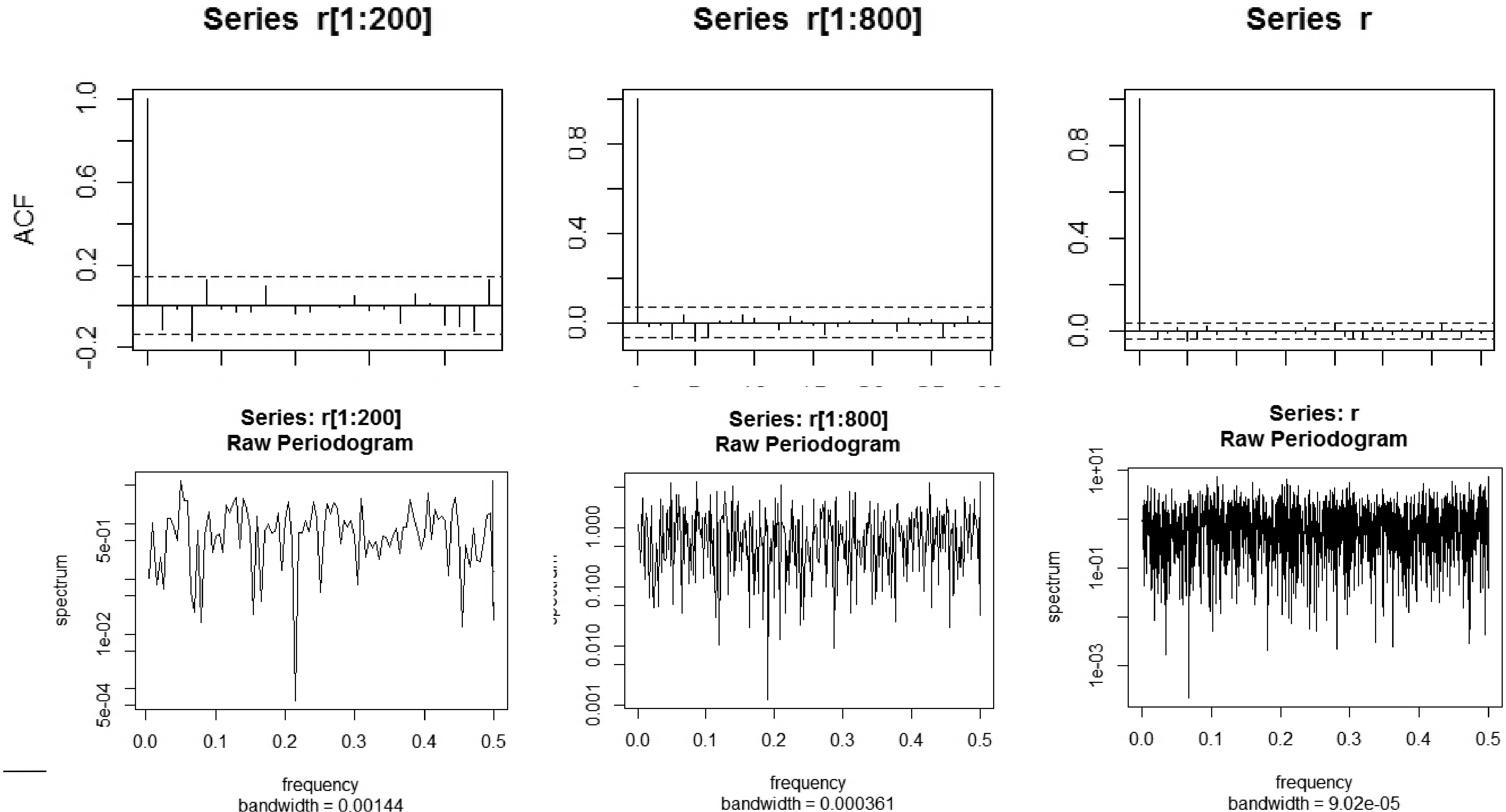
$$\lim_{n\to\infty} \hat{p}(f) \neq p(f)$$

$$\frac{2\hat{p}(f_1)}{p(f_1)}, \cdots, \frac{2\hat{p}(f_m)}{p(f_m)} \sim \chi_2^2 \qquad m = \left[\tfrac{N}{2}\right] - 1$$

$$\frac{\hat{p}(0)}{p(0)}, \frac{\hat{p}(\frac{1}{2})}{p(\frac{1}{2})} \sim \chi_1^2$$

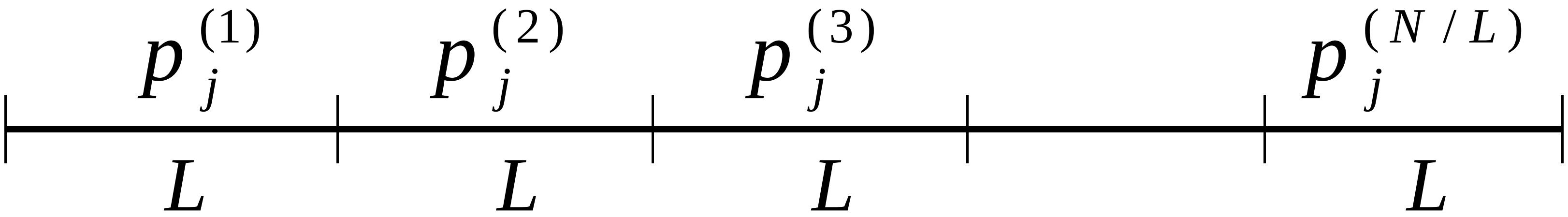
```
r <- rnorm(3200)
```


r
-3
-1
1
2
3
0
500
1000
1500
2000
2500
3000


Series r[1:200]
Series r[1:800]
Series r
ACF
Series: r[1:200]
Raw Periodogram
Series: r[1:800]
Raw Periodogram
Series: r
Raw Periodogram
spectrum
frequency
bandwidth = 0.00144
bandwidth = 0.000361
bandwidth = 9.02e-05

ピリオドグラムの平均

$$\ell = \frac{N}{L}$$



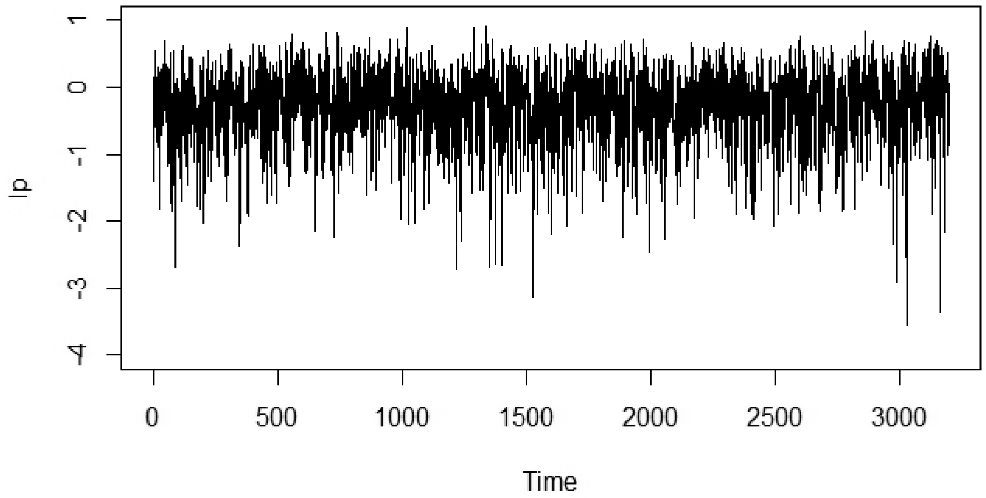
$$\hat{p}_j = \sum_{k=1}^{N/L} p_j^{(k)}$$

分散は $1/\ell$ に減少

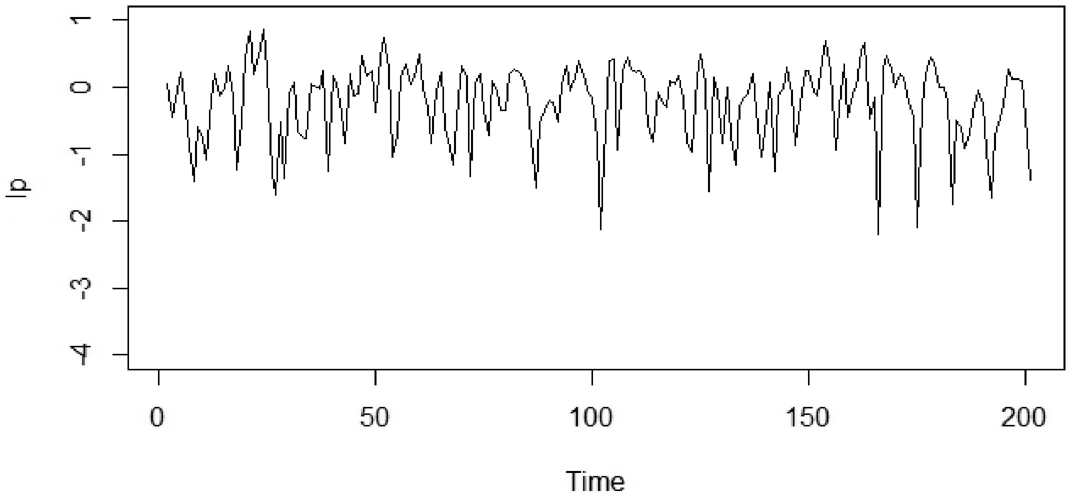
実はラグL-1までのフーリエ変換でよい

$$p_j = \hat{C}_0 + 2\sum_{k=1}^{L-1} \hat{C}_k \cos(2\pi k f_j)$$

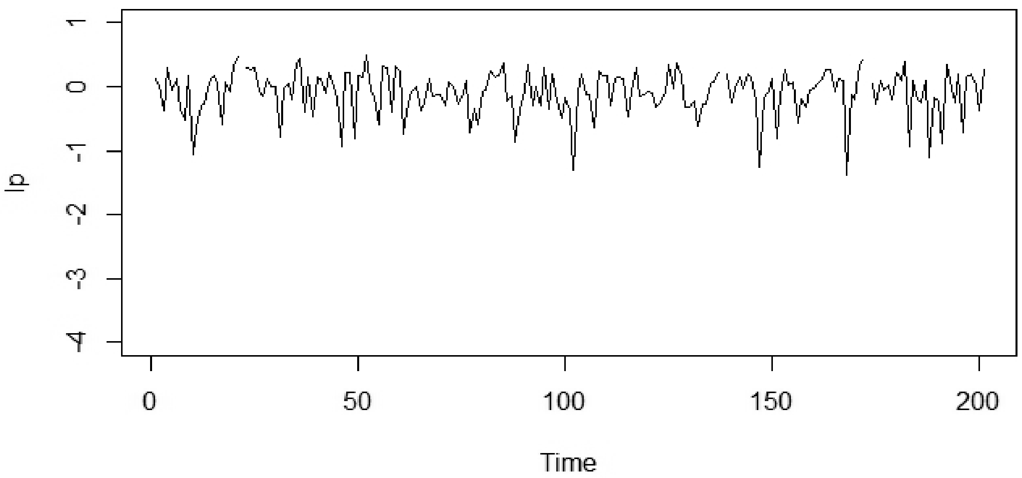
N=3200, lag=3200



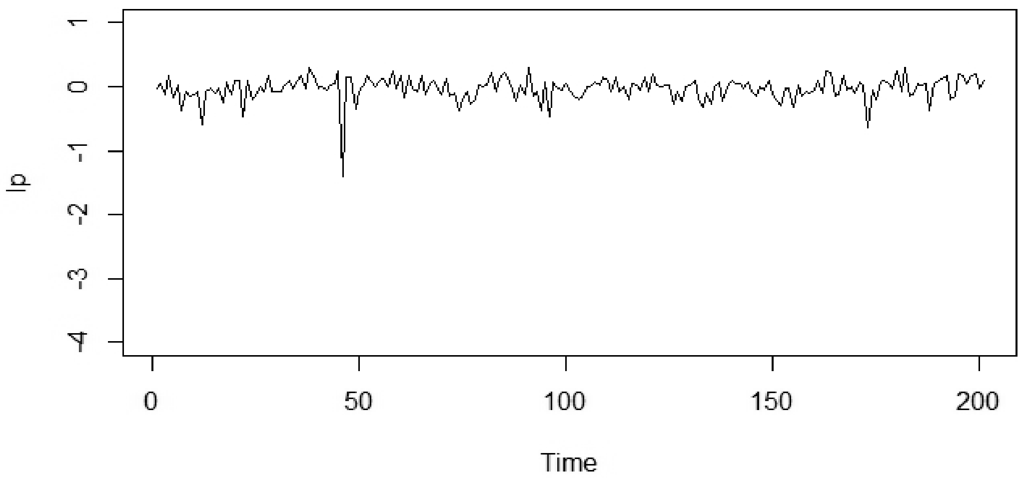
N=200, lag=200



N=800, lag=200



N=3200, lag=200



ピリオドグラムの平滑化

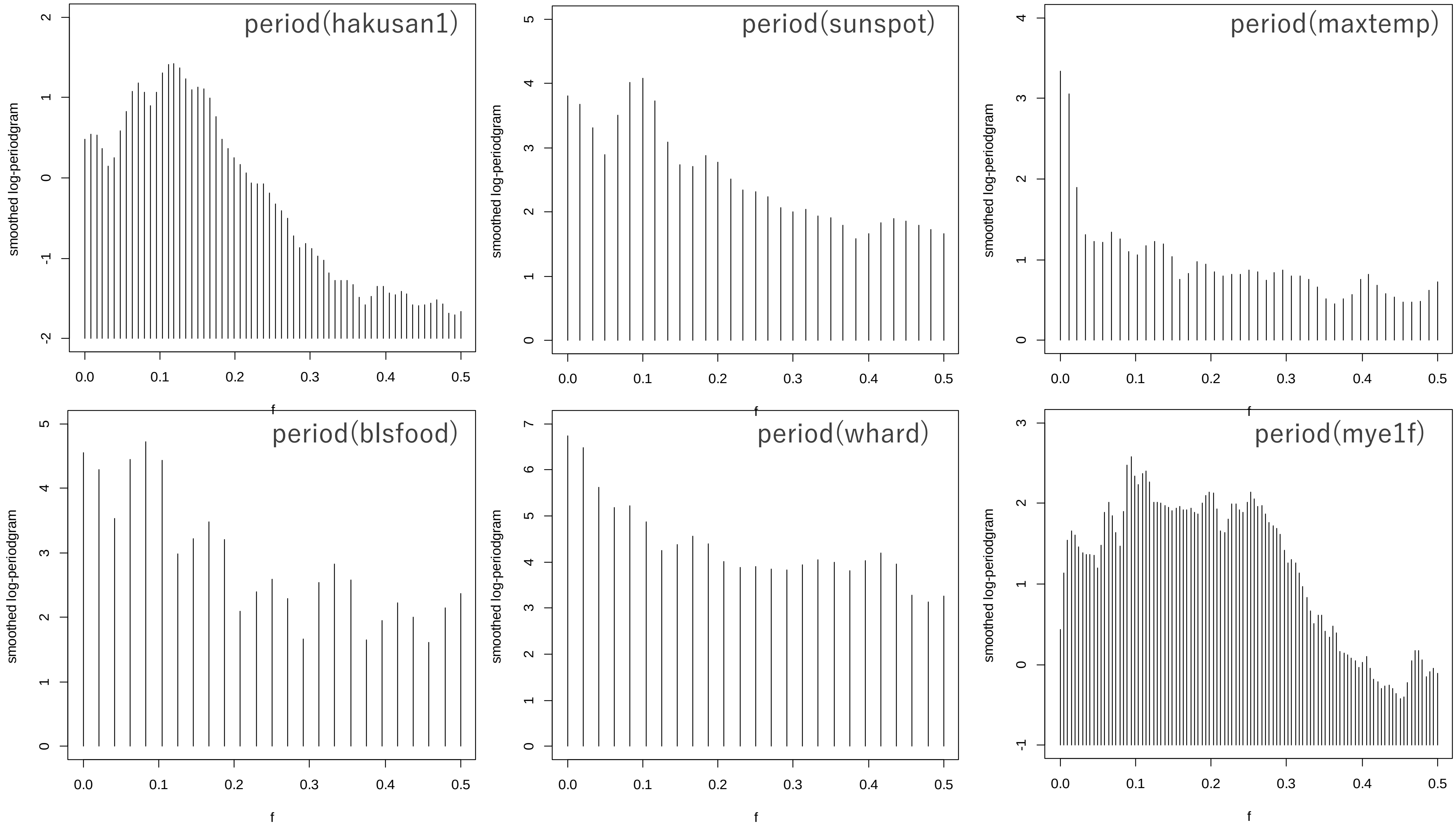
$$\tilde{p}_j = \sum_{i=-m}^{m} W_i \hat{p}_{j-i}$$

W_i： スペクトル　ウィンドウ

Window	m	W_0	W_1
Hanning	1	0.50	0.25
Hamming	1	0.54	-.23

Blackman-Tukey 法

Smoothed Periodogram



FFT （Fast Fourier Transform）

$N = p^{\ell}$ のとき，計算量　$N^2 \to Np\ell$

$$\frac{Np\ell}{N^2} = \frac{p\ell}{N}$$

$N = 1024 = 2^{10}$　$\dfrac{2\times 10}{1024} \sim \dfrac{1}{50}$

$N = 4096 = 2^{12}$　$\dfrac{2\times 12}{4096} \sim \dfrac{1}{170}$

ただし $\hat{C}_k$ の計算も $O(N^2)$

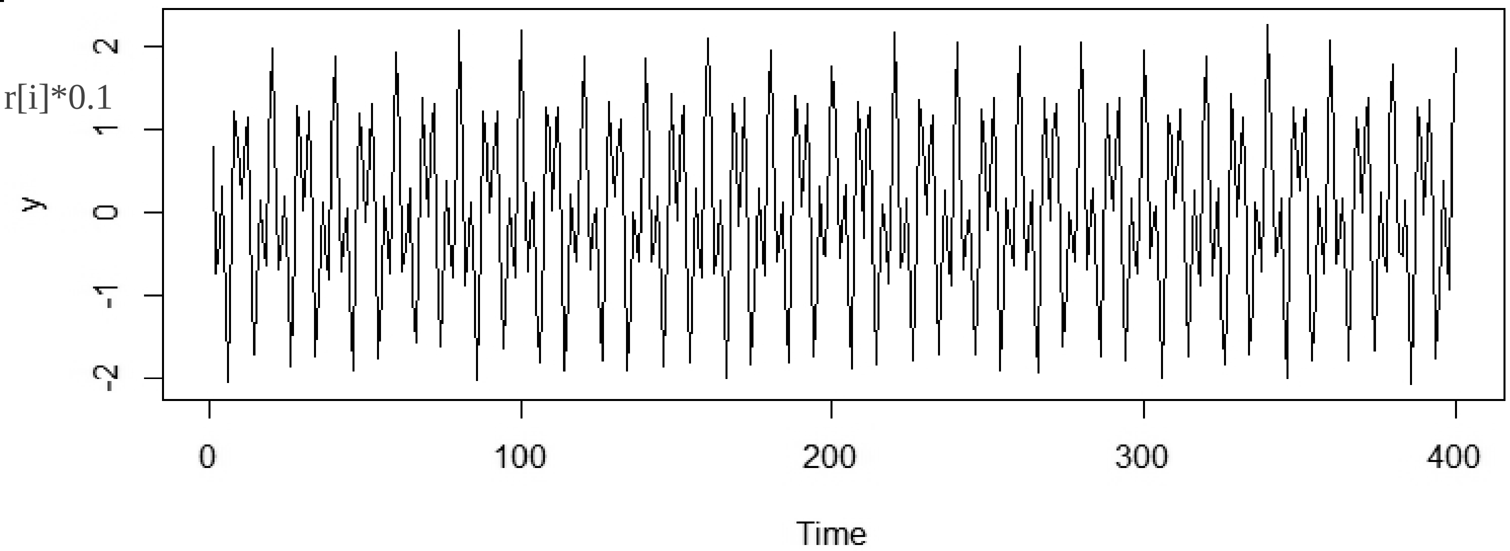
FFTによるピリオドグラムの計算

$$
\begin{aligned}
Z_j &= \sum_{n=1}^{N} y_n e^{-\frac{2\pi i(n-1)j}{N}} \\
&= \sum_{n=1}^{N} y_n \cos\left(\frac{2\pi(n-1)j}{N}\right) - i\sum_{n=1}^{N} y_n \sin\left(\frac{2\pi(n-1)j}{N}\right) \\
&= FC_j - iFS_j
\end{aligned}
$$

$$
\begin{aligned}
p_j &= \frac{|Z_j|^2}{N} = \frac{1}{N}\left|\sum_{n=1}^{N} y_n e^{-2\pi i(n-1)j/N}\right|^2 \\
&= \frac{1}{N}\left(FC_j^2 + FS_j^2\right)
\end{aligned}
$$

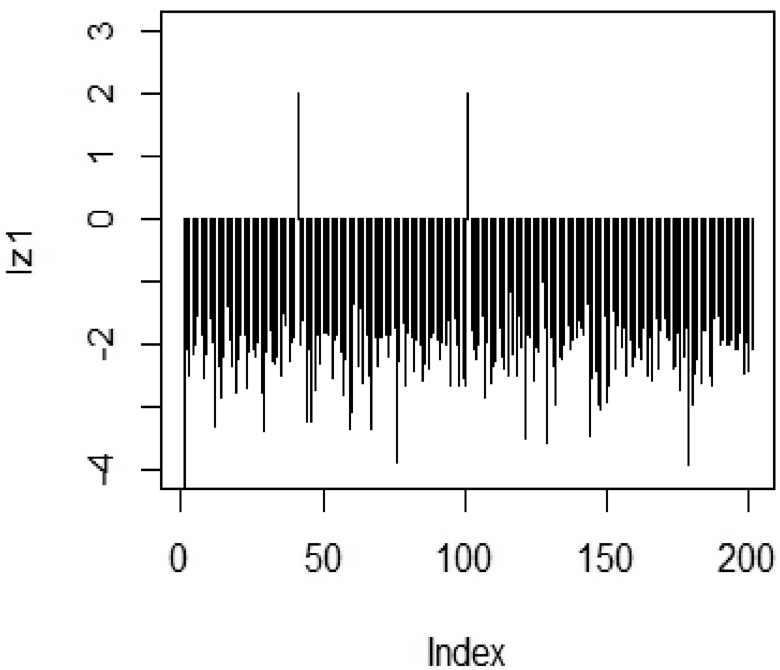
$$
\begin{tikzcd}
y_n \arrow[r, "y*y"] \arrow[d, "F"'] & C_k \arrow[d, "F"] \\
Z(f) \arrow[r, "|Z|^2"'] & p(f)
\end{tikzcd}
$$

```
# simulation by 2 cosine function
t <- 1:400
r <- rnorm(400)
for (i in t) {
y[i] <- cos(2*pi*i/10) + cos(2*pi*i/4) + r[i]*0.1
}
y <- as.ts(y)
plot(y)
```

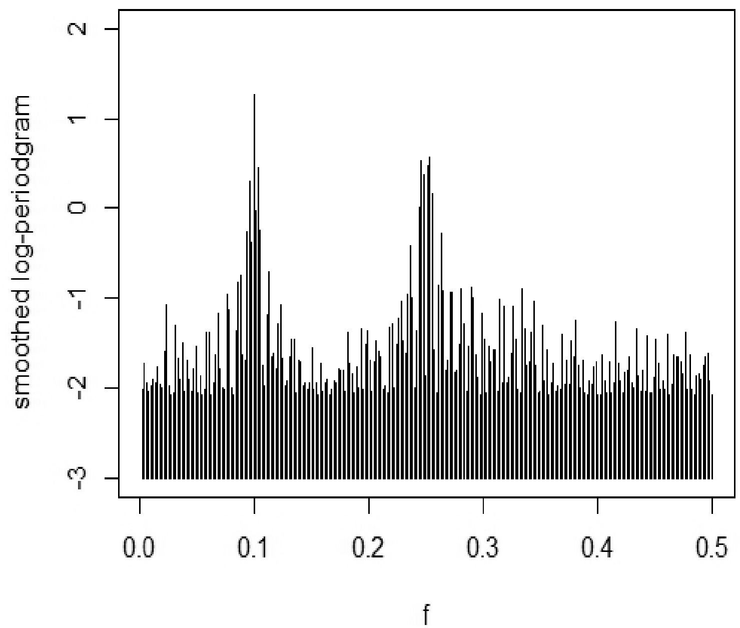



```
z <- period(y,window=0)
z1 <- z$period
lz1 <- log10(z1)
plot(lz1,type="h")
```

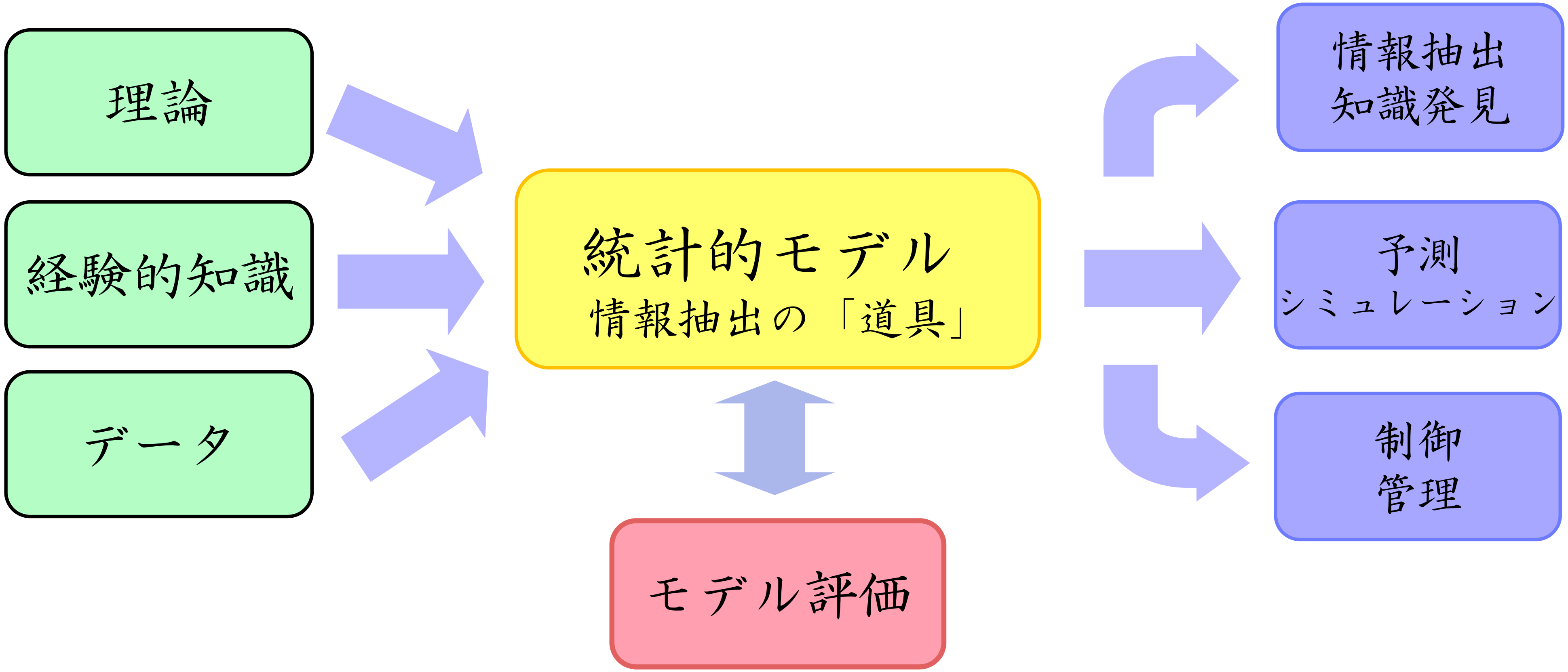
ピリオドグラム



FFTピリオドグラム



統計的モデリングと評価



- 統計的推論の結果はモデルに依存する
- モデル評価・選択が重要

H. Akaike Became Google's Top Logo 11/6/2017

If you've ever conducted a statistical analysis, you might've spent hours thinking about which variables to include and the impact each would have on the outcome. But to ensure the model itself is accurate, shouldn't someone measure the measurers?

In the early 1950s, a young Japanese scientist named Hirotugu Akaike asked this simple but crucial question. More than two decades of research later, he presented the answer as a simple equation known as the Akaike Information Criterion. With AIC, analysts select a model from a set of options by measuring how close the results are to the (hypothetical) truth.

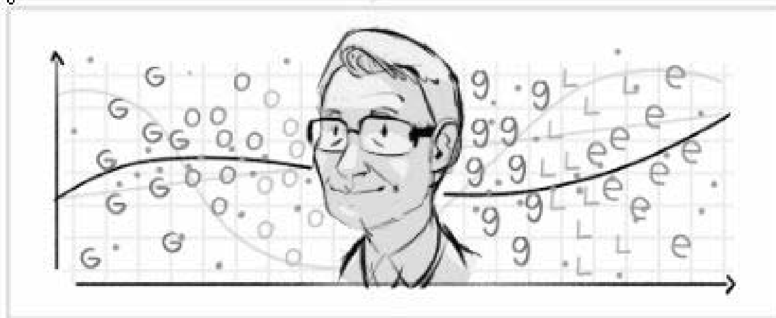
For Dr. Akaike, experience was core to creativity. To get 'a direct feel of random vibrations,' for example, he bought a scooter and rode it around Mount Fuji. This first-hand experience helped him differentiate between the vibrations of riding on normal and heavily-trucked roads.

Today's Doodle portrays Dr. Akaike against a Google-inspired approximation of functions, parameters, and their respective curves.

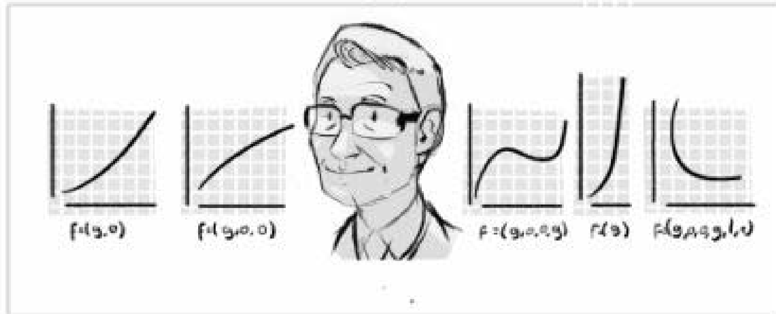
Below are a few initial conceptualizations of the Doodle.

①

②



③



K-L情報量とエントロピー

$$B(g;f) = -I(g;f) = \sum g_i \log \frac{g_i}{f_i}$$

Boltzmannのエントロピー

モデル　$f = (f_1, \ldots, f_k)$

n個の独立な観測値　$(n_1, \ldots, n_k)$　$n_1 + \cdot\cdot\cdot + n_k = n$

相対度数　$(g_1, \ldots, g_k)$

$g_i = n_i / n$　$(n_1, \ldots, n_k)$が得られる確率

$$B(g;f) \sim \frac{1}{n} \log W$$

W:　想定したモデルから得られたサンプルの相対度数が真の分布と一致する確率

$$W = \frac{n!}{n_1! \cdots n_k!} f_1^{n_1} \cdots f_k^{n_k}$$

$$\log W! = \log n! - \sum_{i=1}^k \log n_i! + \sum_{i=1}^k n_i \log f_i$$

$$\sim n \log n - n - \sum_{i=1}^k n_i \log n_i + \sum_{i=1}^k n_i + \sum_{i=1}^k n_i \log f_i$$

$$= -\sum_{i=1}^k n_i \log \frac{n_i}{n} + \sum_{i=1}^k n_i \log f_i$$

$$= \sum_{i=1}^k n_i \log \frac{f_i}{g_i}$$

$$= n \sum_{i=1}^k g_i \log \frac{f_i}{g_i} = nB(g;f)$$

数値積分によるKL情報量の計算

```
# g:gauss, f:gauss
klinfo(1, c(0, 1), 1, c(0.1, 1.5), 8)
```

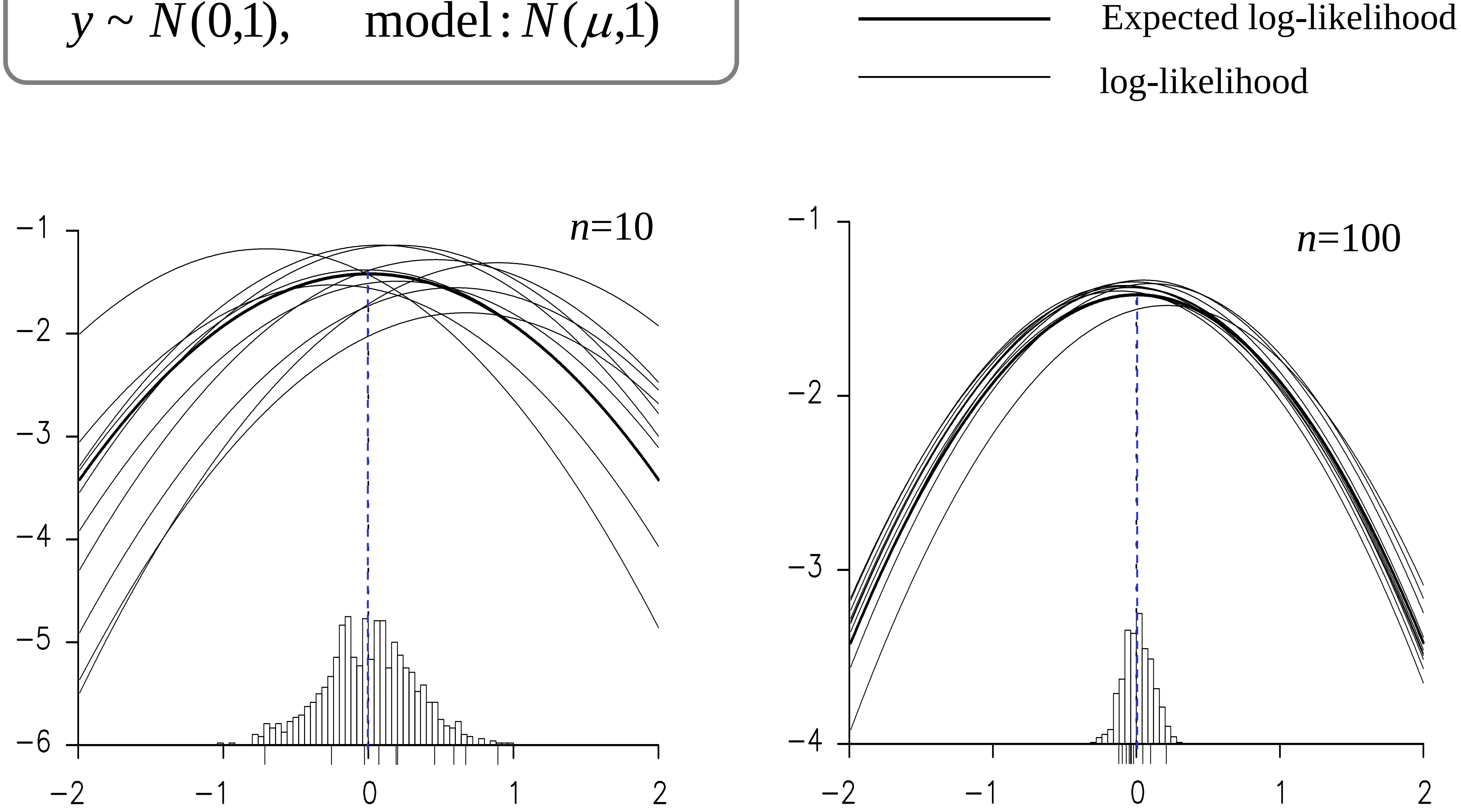
x_k	k	dx	KLI	gint
8.00	16	1.0000	0.03939926	1.00000001
8.00	32	0.5000	0.03939922	1.00000000
8.00	64	0.2500	0.03939922	1.00000000
8.00	128	0.1250	0.03939922	1.00000000

```
# g:gauss, f:cauchy
klinfo(1, c(0, 1), 2, c(0, 1), 8)
```

x_k	k	dx	KLI	gint
8.00	16	1.0000	0.25620181	1.00000001
8.00	32	0.5000	0.25924202	1.00000000
8.00	64	0.2500	0.25924453	1.00000000
8.00	128	0.1250	0.25924453	1.00000000

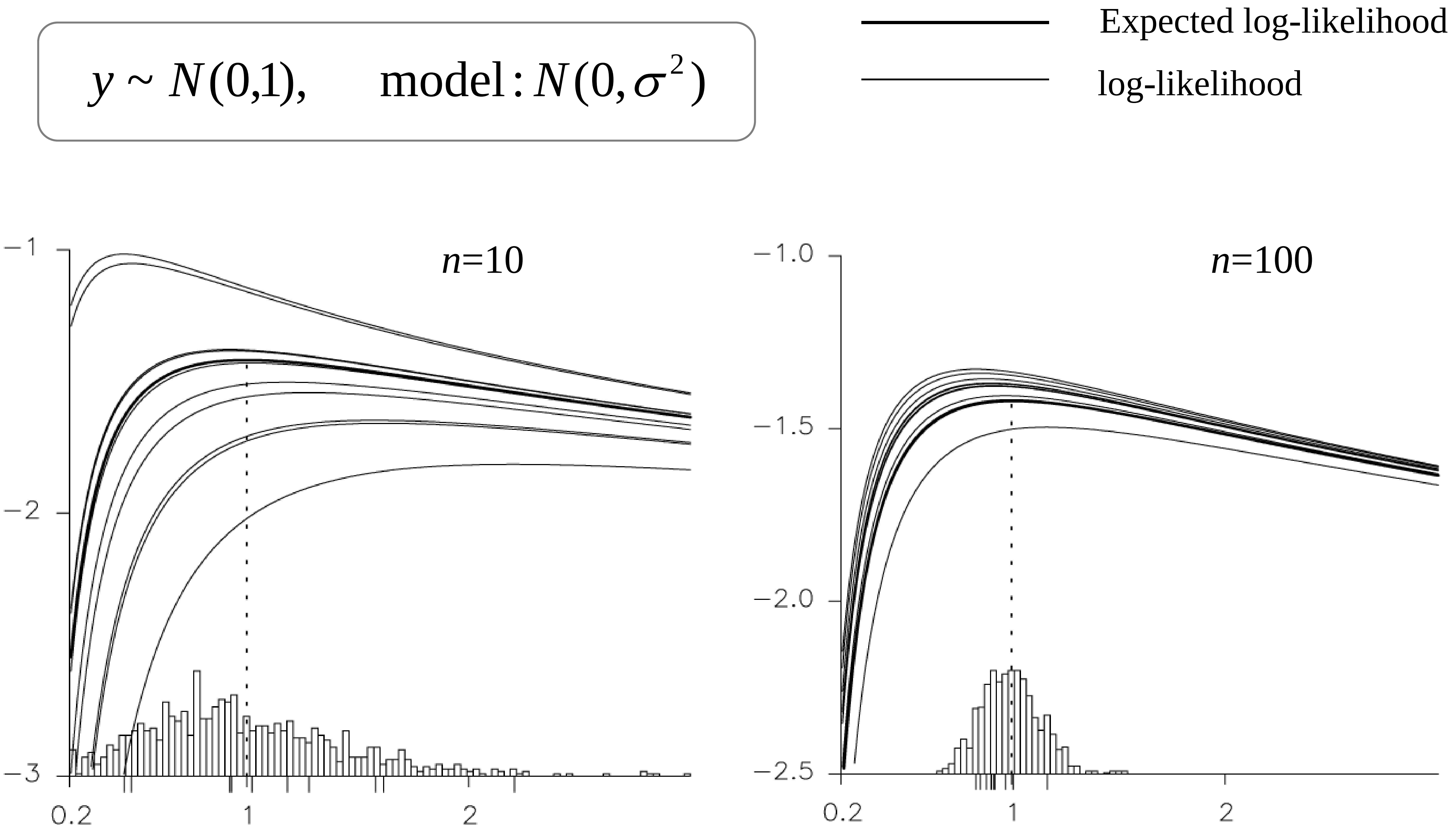
最尤推定値の例(正規分布の平均)

$$y \sim N(0,1), \quad \text{model}: N(\mu,1)$$



最尤推定値の例（正規分布の分散）

$$y \sim N(0,1), \quad \text{model}: N(0,\sigma^2)$$

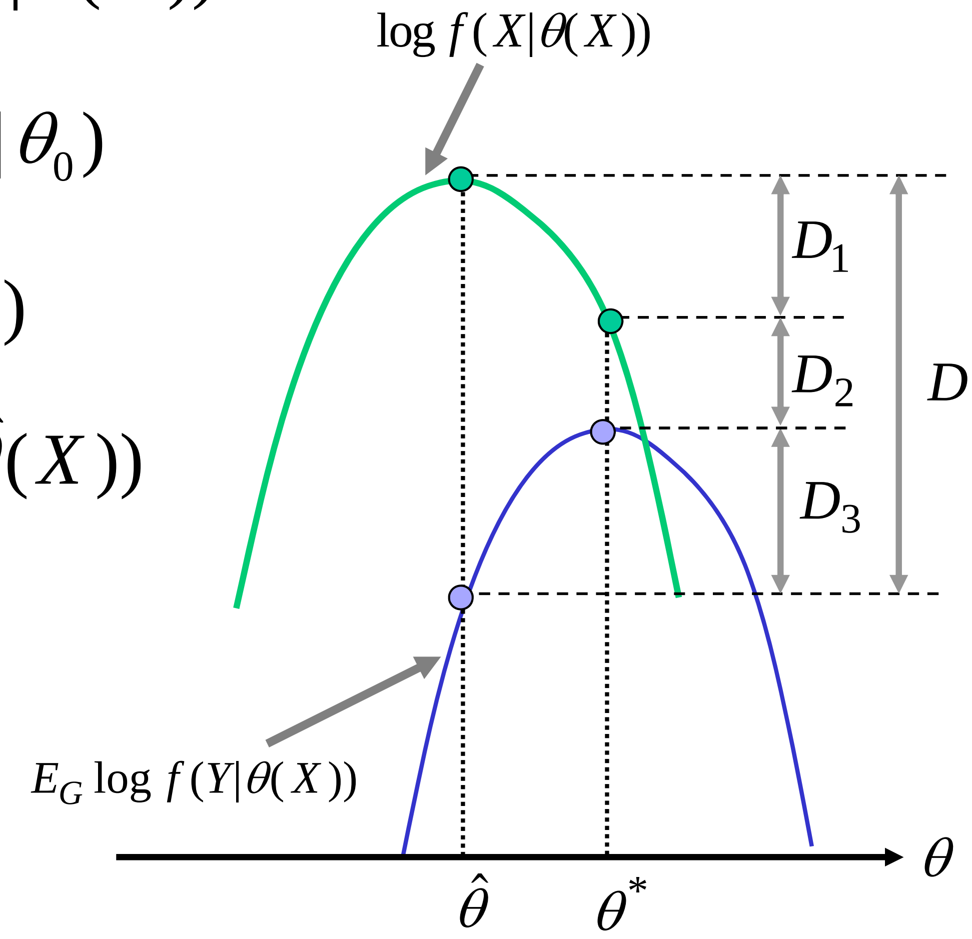


バイアスの評価

対数尤度

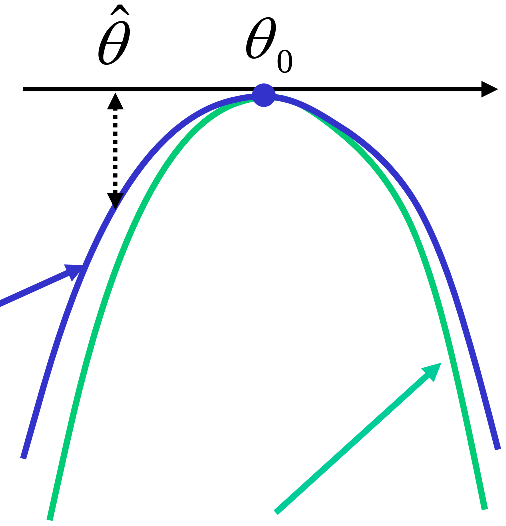
平均対数尤度

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



D_3 , D_1, D_2 の評価

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



$E_Y \log f(Y \mid \hat\theta(X))$

$$E_X\left\{(\hat\theta - \theta_0)^T J(\theta_0)(\hat\theta - \theta_0)\right\} = \frac{1}{n} tr\left\{I(\theta_0) J^{-1}(\theta_0)\right\}$$

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

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

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

情報量規準（Generic）

$$b(G) = E_X\{D\} = E_X\{D_1 + D_2 + D_3\}$$

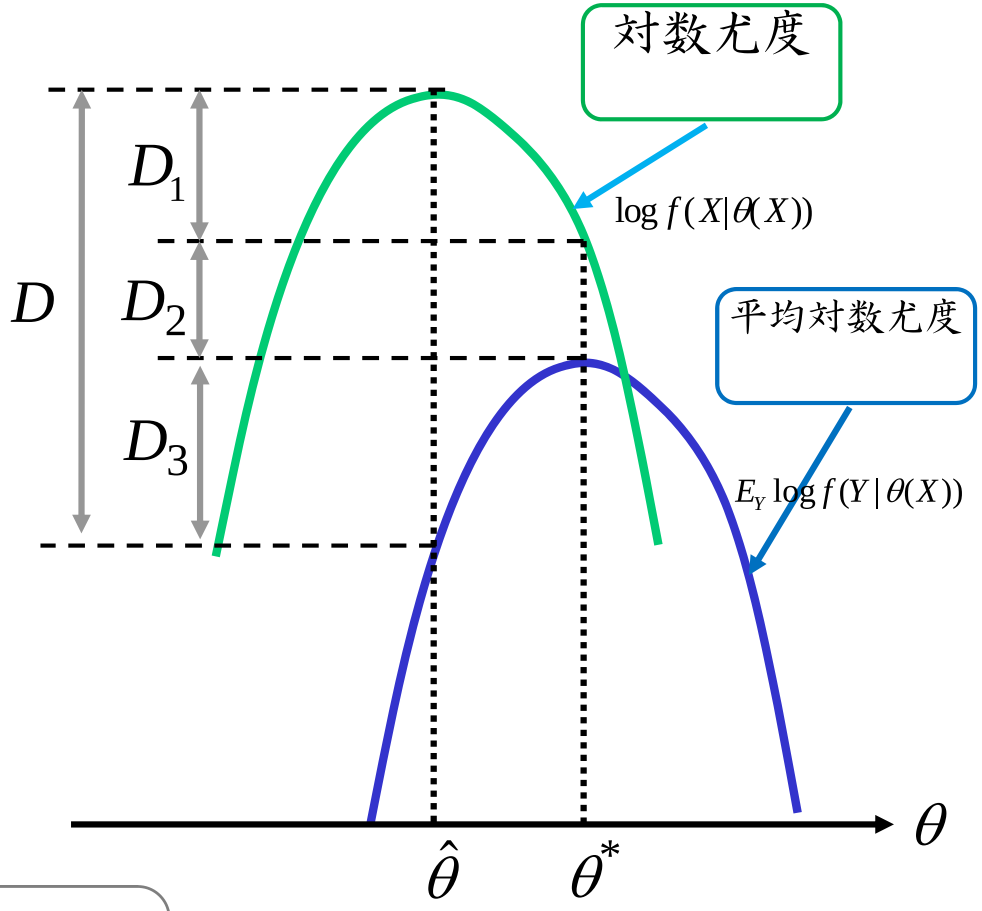
$$b_{\mathrm{TIC}}(G) = \mathrm{tr}\left\{I(G)J(G)^{-1}\right\}$$

$$I(G) = E_X\left[\frac{\partial \log f(X\mid\theta)}{\partial\theta}\frac{\partial \log f(X\mid\theta)}{\partial\theta'}\right]$$

Fisher情報量

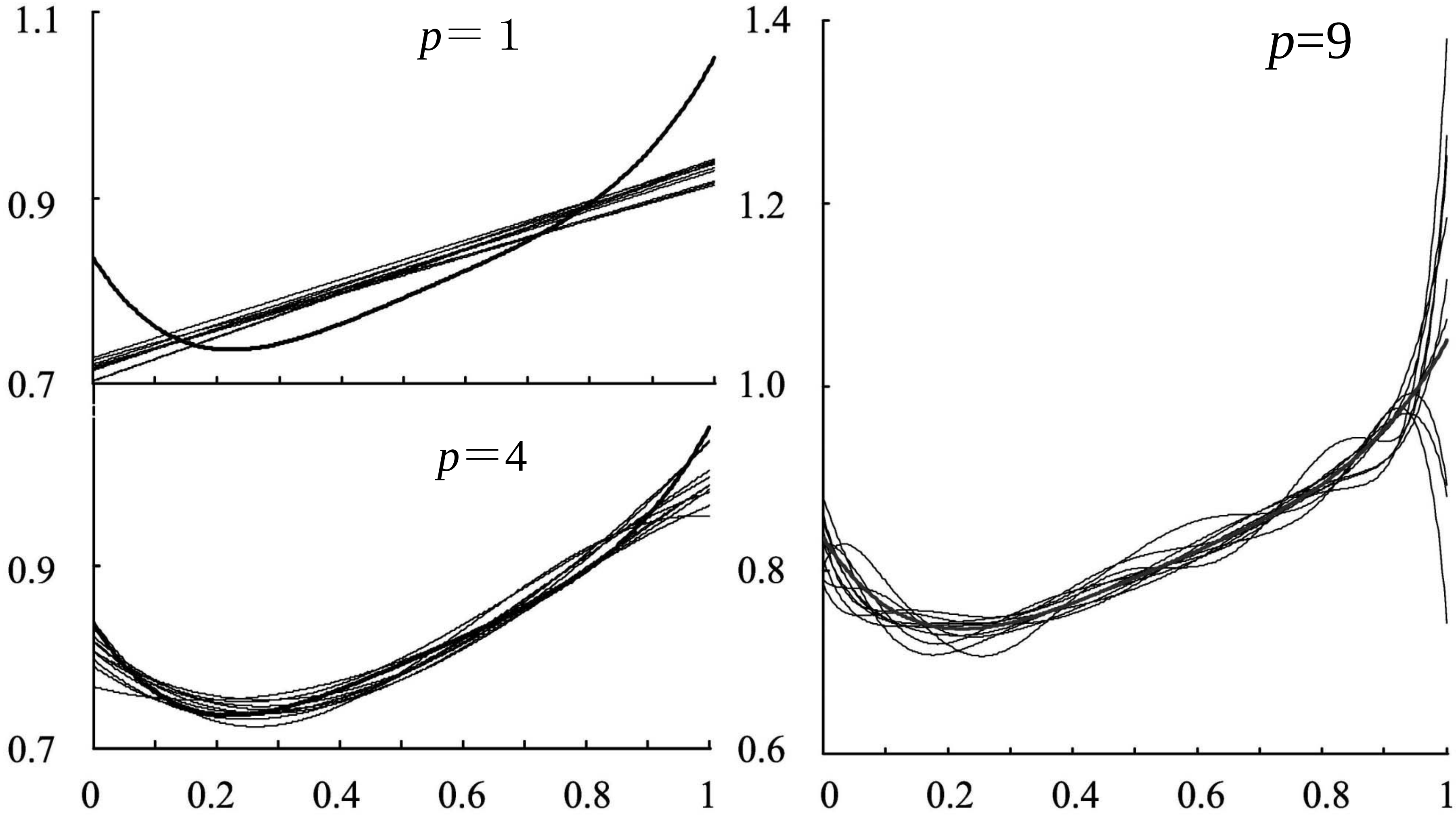
$$J(G) = -E_X\left[\frac{\partial^2 \log f(X\mid\theta)}{\partial^2\theta}\right]$$

ヘッセ行列の期待値



$$\mathrm{IC} = -2\log f(x\mid\hat{\theta}) + 2\hat{b}(G)$$

モデル選択例：多項式回帰の次数



Selection of the Bin Size of a Histogram

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

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

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

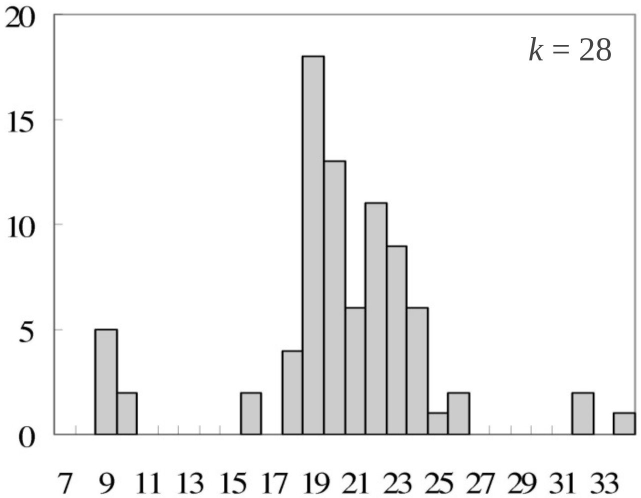
$$\mathrm{AIC}_k = (-2)\left\{C + \sum_{j=1}^{k} n_j \log\left(\frac{n_j}{j}\right)\right\} + 2(k-1)$$

Galaxy data (Roeder (1990))

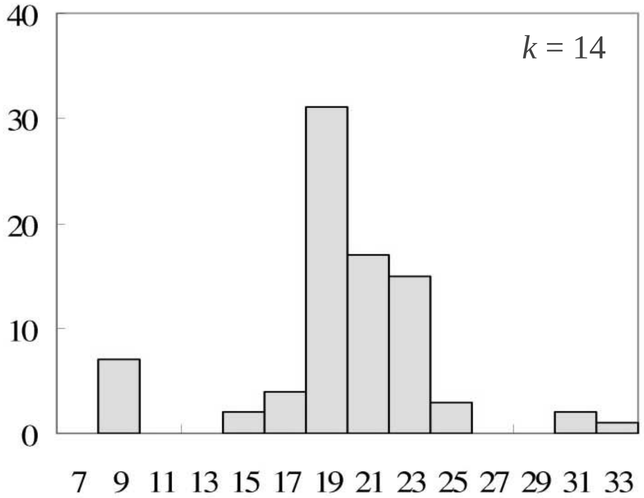
0 5 2 0 0 0 0 0 2 0 4 18 13 6
11 9 6 1 2 0 0 0 0 0 2 0 1 0

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

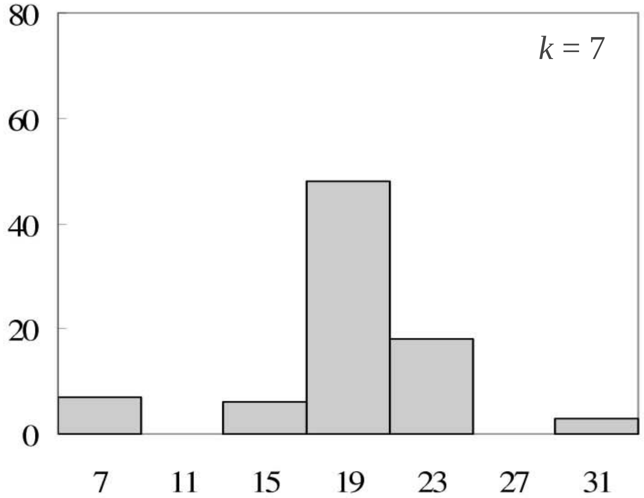
Histogram of galaxy data



Best



Too small



Variable Selection for a Regression M

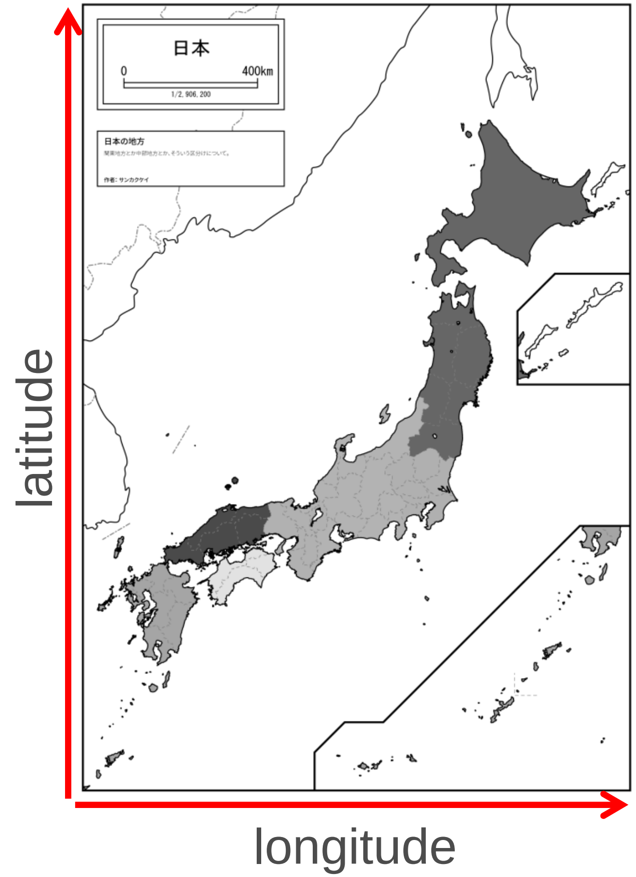
y_n: Temperature, x_{1n}: Latitude, x_{2n}: Longitude, x_{3n}: Alti

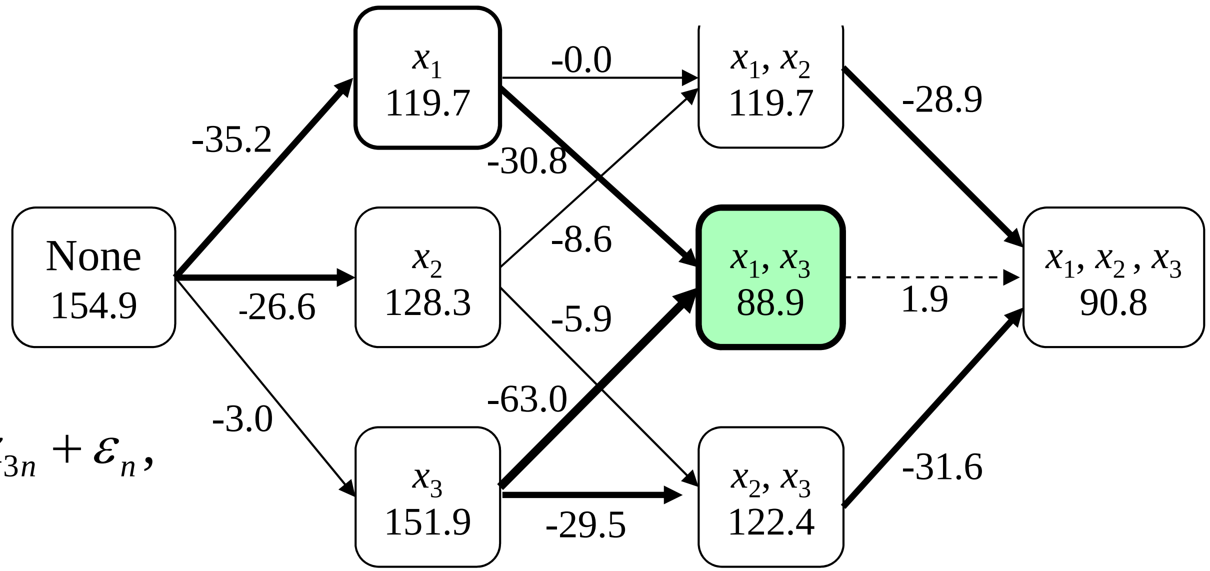
$$y_n = a_0 + a_1 x_{1n} + a_2 x_{2n} + a_3 x_{3n} + \varepsilon_n, \qquad \varepsilon_n \sim N(0, \sigma^2)$$

Select variables among x_1, x_2, x_3 appropriate to predic

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

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





Selected model

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

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

モデル選択例：分布の形状の選択

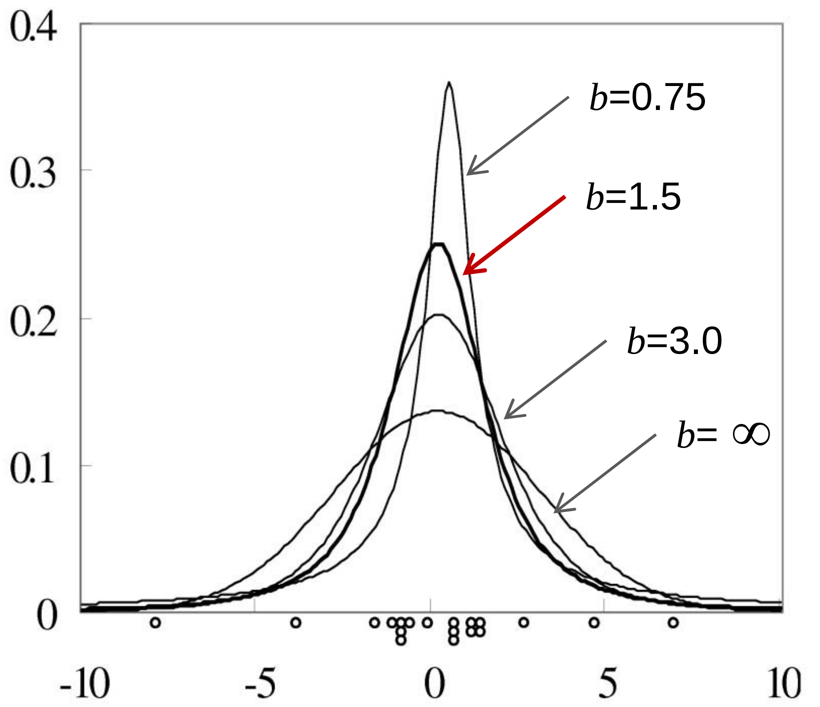
$$f(y \mid \mu, \tau^2, b) = \frac{C}{\left(y^2 + \tau^2\right)^b}$$

Pearson's family of distributions
Select the shape parameter b

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

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

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

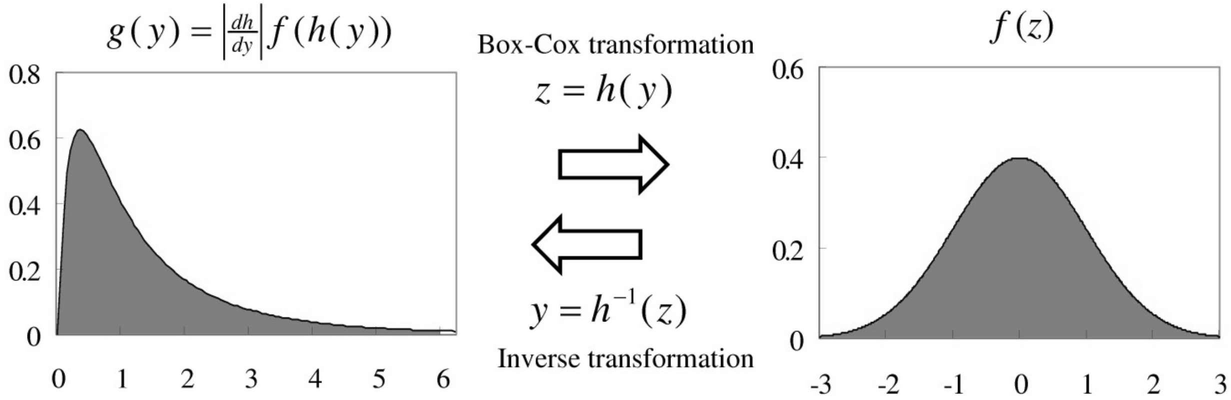


Selection of the Box-Cox transformation

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

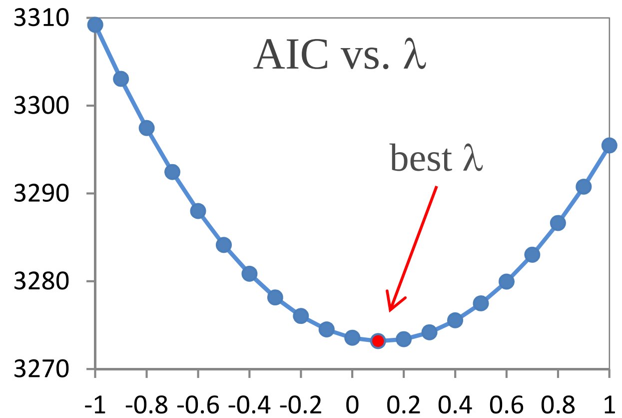
$$g(y) = \left| \frac{dh_{\lambda}}{dy} \right| f(h(y))$$

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



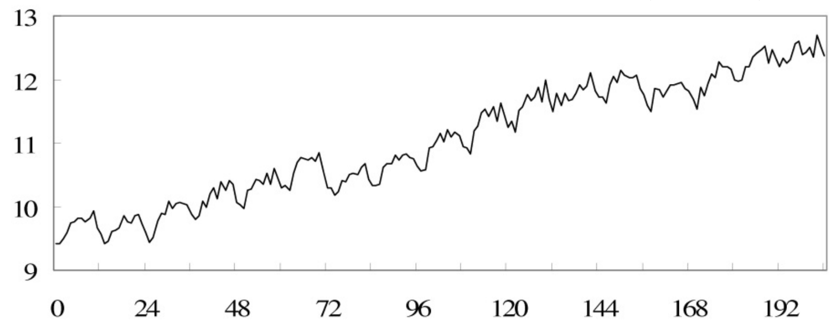
Jacobian of the transformation

λ	-1.0	-0.5	0.0	0.1	0.2	0.5	1.0
log-L	1374.9	630.5	-121.1	-272.2	-423.7	-879.9	-1645.3
AIC	-2745.7	-1257.0	246.1	548.5	854.4	1763.8	3295.5
AIC’	3309.2	3284.1	3273.6	3273.2	3273.4	3277.5	3295.5



Original WHARD data (US BLS)

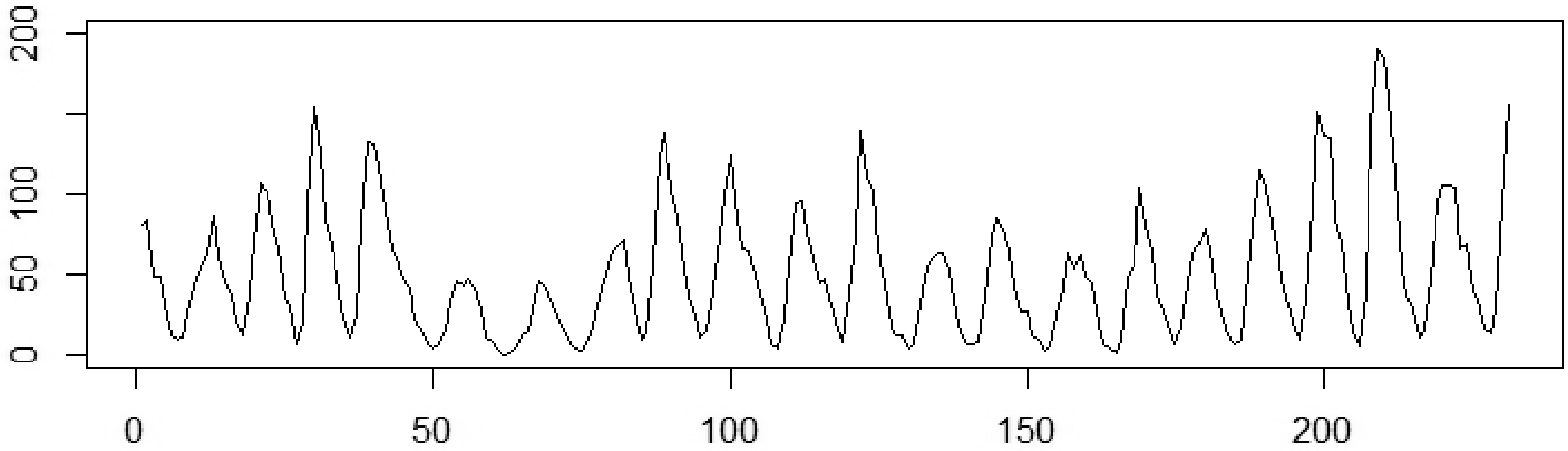
Best Box-Cox transformation (λ=0.1)



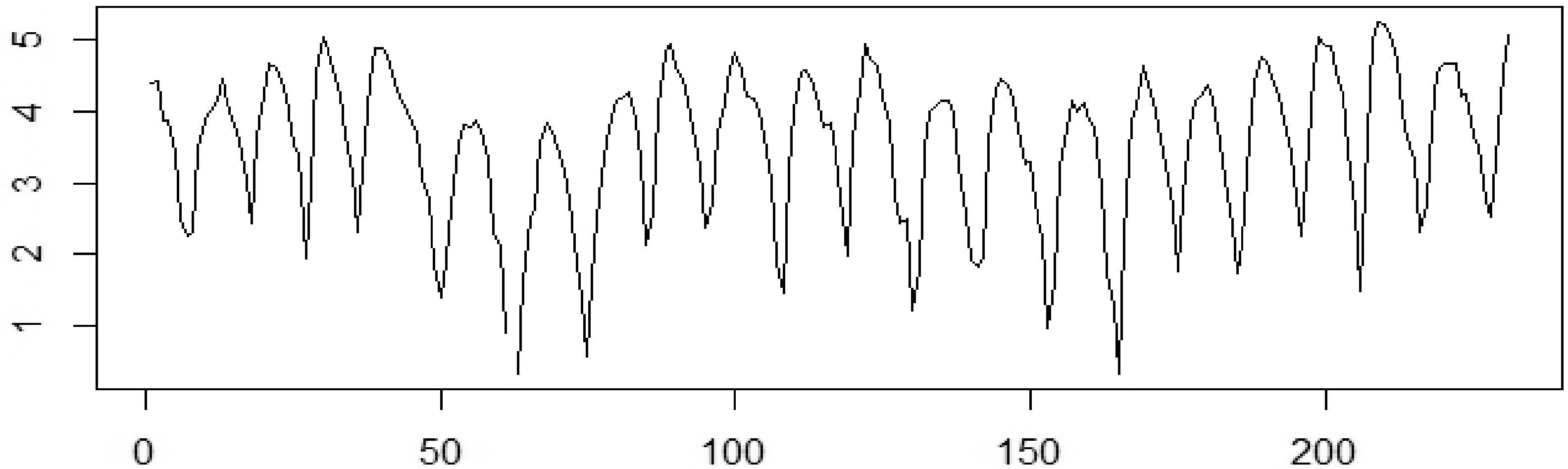
太陽黒点数データ

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

y(n)



log y(n)

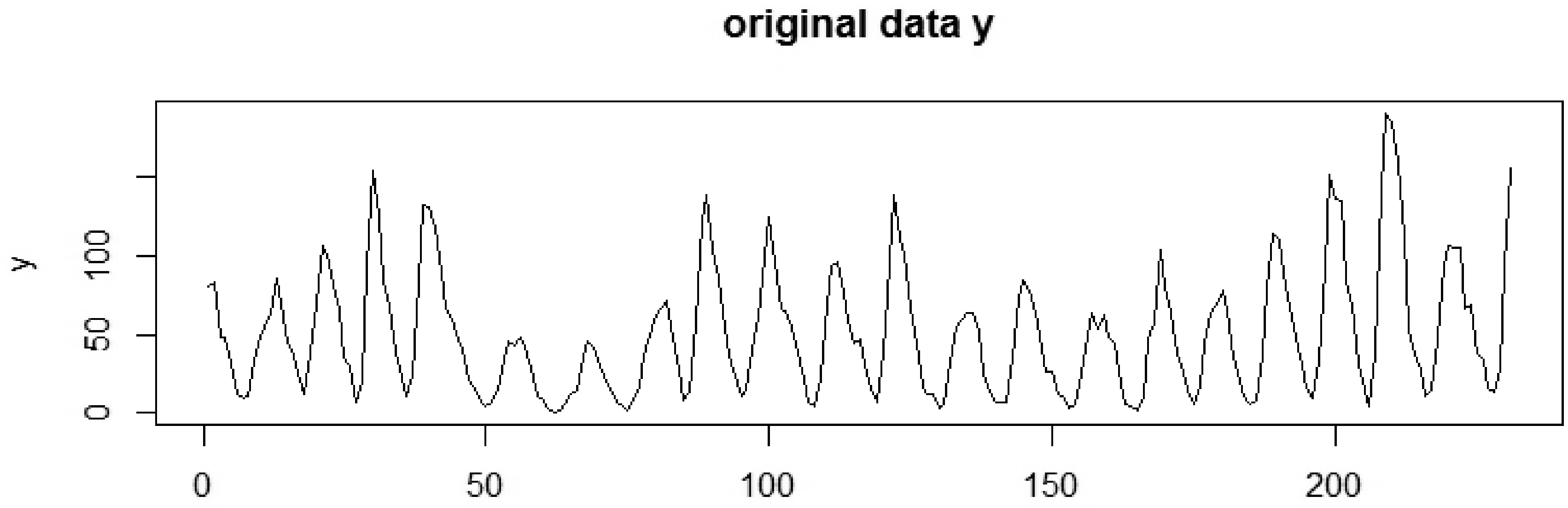


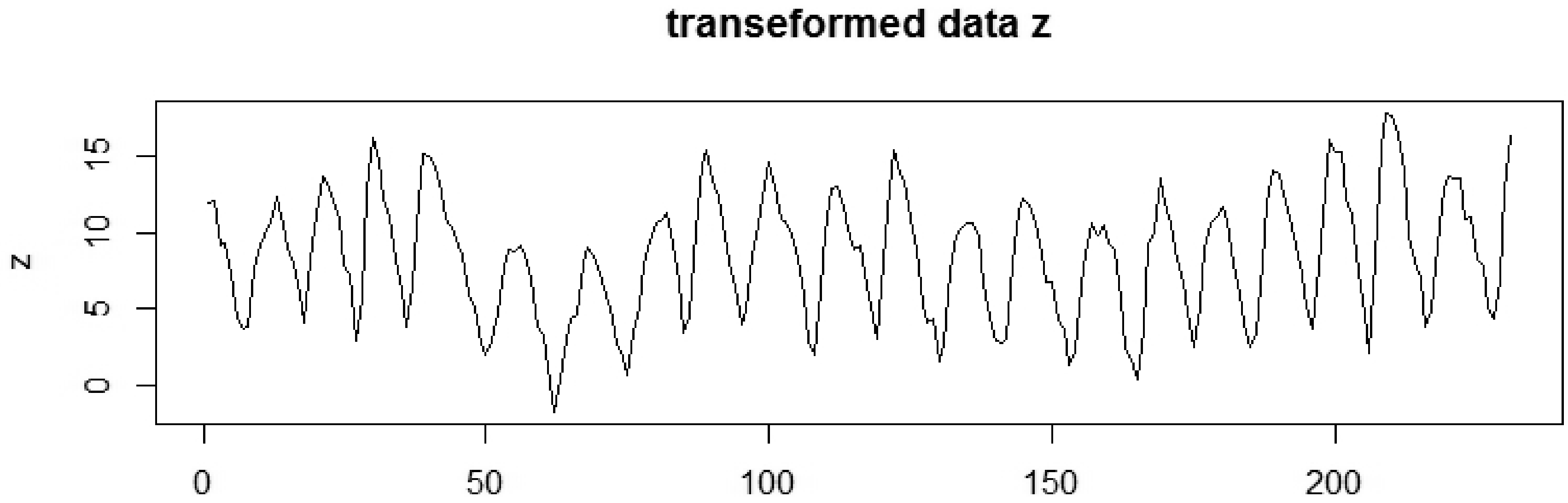
```
data(Sunspot) # Sun spot number data
boxcox(Sunspot)
```

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

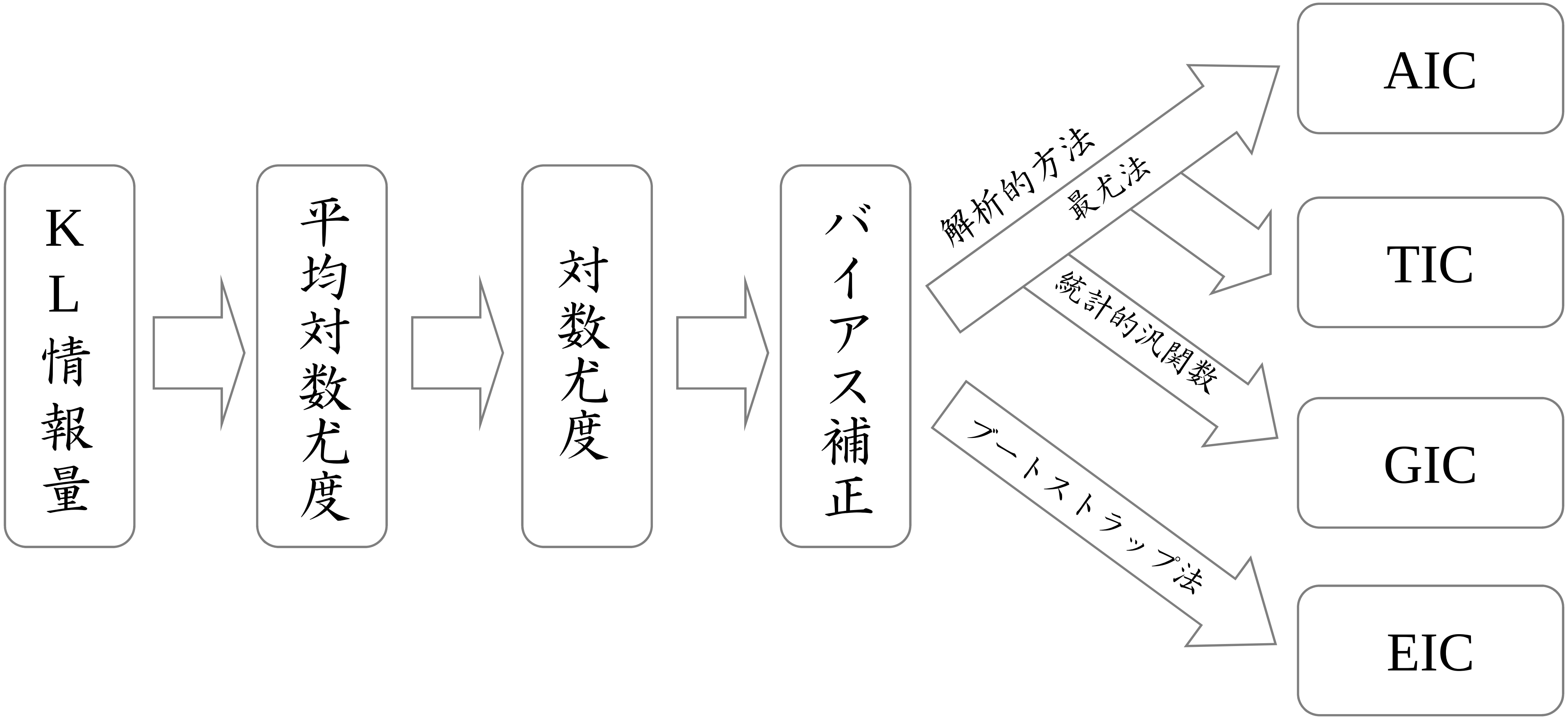
lambda = 0.40 AIC' minimum = 2266.32

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





情報量規準の系譜


KL情報量
平均対数尤度
対数尤度
バイアス補正
解析的方法
最尤法
統計的汎関数
ブートストラップ法
AIC
TIC
GIC
EIC

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.