

AIC のゆらぎについて

統計数理研究所 石 黒 真 木 夫

(1994 年 1 月 受付)

1. はじめに

AIC によるモデル評価は, AIC が期待平均対数尤度の推定量であることに依拠している. この評価の信頼度が評価できれば AIC が安心して使えるようになる. AIC のゆらぎに関する知見をまとめておくことをこの稿の目的とする.

2 つのモデルの AIC の差のゆらぎの理論の本質的な部分はすでに Kishino and Hasegawa (1989) によって明らかにされている. 最近下平 (1993a, 1993b) によって精密に論じられていることでもある. そのかならずしも新しくないことがらを, 筆者もその一翼をになった『情報量統計学』(坂元 他 (1983)) の論理展開につづけるかたちで整理した上で, 簡単な数値例で AIC の同時分布の明確なイメージを与え, 最後に「AIC の真値の bootstrap 推定」を提案する.

2. 理 論

2.1 定義と仮定と記号

$x=(x_1, x_2, \dots, x_n)$ で確率変数 $X=(X_1, X_2, \dots, X_n)$ の実現値を表す. X の確率密度関数を

$$(2.1) \quad g_n(x) = f_n(x | \theta^*) = \prod_{i=1}^n f(x_i | \theta^*)$$

とする. ただし f はパラメータベクトル $\theta \in \Theta$ を持つモデルとする. このパラメータの最尤推定値 $\hat{\theta}$ は対数尤度

$$(2.2) \quad l_n(\theta) = \sum_{i=1}^n \log f(x_i | \theta)$$

の最大化で求められる.

モデル $f_n(\cdot | \theta)$ の平均対数尤度は

$$(2.3) \quad l_n^*(\theta) = E\{l_n(\theta)\} = E\left\{\sum_i \log f(X_i | \theta)\right\} = n \int f(z | \theta^*) \log f(z | \theta) dz$$

で与えられる. $[\partial^2 h / \partial \theta^2]$ で (j, l) 成分が $\partial^2 h / \partial \theta_j \partial \theta_l$ である行列を表すことにしよう. 行列 J_* を

$$(2.4) \quad J_* = -E_z \left[\frac{\partial^2}{\partial \theta^2} \log f(Z | \theta) \right]_{\theta^*}$$

で定義する. ここで $[]_{\theta^*}$ は角カッコの中の式を $\theta = \theta^*$ で評価する意味である. 記号

$$\| \theta \|^2 = \theta J_* \theta^T$$

を使って、平均対数尤度の θ^* のまわりでの Taylor 展開が近似的に

$$(2.5) \quad l_n^*(\theta) = l_n^*(\theta^*) - \frac{n}{2} \|\theta - \theta^*\|^2$$

と表される。適当な正則条件のもとで、最尤推定値 $\hat{\theta}$ は漸近正規性

$$(2.6) \quad \sqrt{n}(\hat{\theta} - \theta^*) \rightarrow N(0, J_*^{-1}) \quad (n \rightarrow \infty)$$

を持ち、大数の法則から

$$(2.7) \quad \frac{1}{n} \left[\frac{\partial^2 l}{\partial \theta^2} \right]_{\theta^*} = \frac{1}{n} \sum_{i=1}^n \left[\frac{\partial^2}{\partial \theta^2} \log f(x_i | \theta) \right]_{\theta^*}$$

$$(2.8) \quad \rightarrow E_Z \left[\frac{\partial^2}{\partial \theta^2} \log f(Z | \theta) \right]_{\theta^*} = -J_*$$

が $n \rightarrow \infty$ の極限で成立する。 $n \rightarrow \infty$ で $\hat{\theta} \rightarrow \theta^*$ となることから、

$$(2.9) \quad \frac{1}{n} \left[\frac{\partial^2 l}{\partial \theta^2} \right]_{\hat{\theta}} \rightarrow -J_* \quad (n \rightarrow \infty)$$

も成り立ち、このことを使って最尤推定値 $\hat{\theta}$ のまわりで対数尤度を Taylor 展開して近似式

$$(2.10) \quad l_n(\theta) = l_n(\hat{\theta}) - \frac{n}{2} \|\theta - \hat{\theta}\|^2$$

を得る。

パラメータ空間を

$$(2.11) \quad \Theta_k \subset \Theta$$

なる Θ_k に制約したモデルを MODEL(k) で表す。このモデルのパラメータの最尤推定値 $\hat{\theta}_k$ は

$$(2.12) \quad l_n(\hat{\theta}_k) = \max_{\theta \in \Theta_k} l_n(\theta)$$

を満たす。 $\theta \in \Theta_k$ のもとで平均対数尤度を最大化するパラメータ θ_k^* は

$$(2.13) \quad l_n^*(\theta_k^*) = \max_{\theta \in \Theta_k} l_n^*(\theta)$$

$$(2.14) \quad = l_n^*(\theta^*) - \frac{n}{2} \|\theta_k^* - \theta^*\|^2$$

を満足する。 θ_k^* は MODEL(k) の中で最もよいパラメータの値である。任意の $\theta \in \Theta_k$ に対して、

$$(2.15) \quad \|\theta - \theta^*\|^2 = \|\theta - \theta_k^*\|^2 + \|\theta_k^* - \theta^*\|^2$$

と

$$(2.16) \quad l_n^*(\theta) = l_n^*(\theta_k^*) - \frac{n}{2} \|\theta - \theta_k^*\|^2$$

が成立する。同様に $\hat{\theta}$ と $\hat{\theta}_k$ と任意の $\theta \in \Theta_k$ に対して、

$$(2.17) \quad \|\theta - \hat{\theta}\|^2 = \|\theta - \hat{\theta}_k\|^2 + \|\hat{\theta}_k - \hat{\theta}\|^2$$

が成り立つ。パラメータ空間 Θ_k の次元を m_k で表すことにしよう。漸近正規性から、 $n\|\hat{\theta}_k - \theta_k^*\|^2$ が自由度 m_k の χ^2 分布に従う確率変数になり、誤差評価

$$(2.18) \quad nE_X\{\|\hat{\theta}_k - \theta_k^*\|^2\} = m_k$$

が得られる。

2.2 AIC の分布

最尤モデルの平均対数尤度

$$(2.19) \quad l_n^*(\hat{\theta}_k) = l_n^*(\theta_k^*) - \frac{n}{2} \|\hat{\theta}_k - \theta_k^*\|^2$$

は“定数 - χ^2 分布に従う確率変数”の形をしていて

$$(2.20) \quad l_n^*(\hat{\theta}_k) = l_n^*(\theta_k^*) - \frac{m_k}{2} - \frac{1}{2} \{n\|\hat{\theta}_k - \theta_k^*\|^2 - m_k\}$$

と書き直せる。 χ^2 分布の性質から $1/2 \cdot \{n\|\hat{\theta}_k - \theta_k^*\|^2 - m_k\}$ の平均は 0, 分散は $m_k/2$ であり, 期待平均対数尤度は

$$E\{l_n^*(\hat{\theta}_k)\} = l_n^*(\theta_k^*) - m_k/2$$

で与えられることが分かる。 $\Theta_1 \subset \Theta_2 \subset \dots$ であれば $l_n^*(\theta_1^*) \leq l_n^*(\theta_2^*) \leq \dots$ がなりたつが, $m_1 \leq m_2 \leq \dots$ でもあるから $E\{l_n^*(\hat{\theta}_k)\}$ を最大化する k ($\hat{k}(n)$ で表す) は $\theta_k^* = \theta^*$ を満足するとは限らない。 $l_n^*(\theta_1^*), l_n^*(\theta_2^*), \dots$ が n に比例するから $\hat{k}(n)$ は n の単調増加関数である。

MODEL(k) の最大対数尤度を以下のように変形することができる:

$$\begin{aligned} l_n(\hat{\theta}_k) &= l_n(\hat{\theta}) - \frac{n}{2} \|\hat{\theta}_k - \hat{\theta}\|^2 \\ &= l_n(\hat{\theta}) - \frac{n}{2} \|\theta_k^* - \hat{\theta}\|^2 + \frac{n}{2} \|\theta_k^* - \hat{\theta}_k\|^2 \\ &= l_n(\hat{\theta}) - \frac{n}{2} \|\theta_k^* - \theta^* + \theta^* - \hat{\theta}\|^2 + \frac{n}{2} \|\theta_k^* - \hat{\theta}_k\|^2 \\ &= l_n(\hat{\theta}) - \frac{n}{2} \|\theta^* - \hat{\theta}\|^2 - n(\theta_k^* - \theta^*)J_*(\theta^* - \hat{\theta})^T \\ &\quad - \frac{n}{2} \|\theta_k^* - \theta^*\|^2 + \frac{n}{2} \|\theta_k^* - \hat{\theta}_k\|^2 \\ &= l_n(\theta^*) - n(\theta_k^* - \theta^*)J_*(\theta^* - \hat{\theta})^T - \frac{n}{2} \|\theta_k^* - \theta^*\|^2 + \frac{n}{2} \|\theta_k^* - \hat{\theta}_k\|^2 \\ &= l_n^*(\theta^*) - \frac{n}{2} \|\theta_k^* - \theta^*\|^2 + l_n(\theta^*) - l_n^*(\theta^*) \\ &\quad - n(\theta_k^* - \theta^*)J_*(\theta^* - \hat{\theta})^T + \frac{n}{2} \|\theta_k^* - \hat{\theta}_k\|^2 \\ (2.21) \quad &= l_n^*(\theta_k^*) + l_n(\theta^*) - l_n^*(\theta^*) - n(\theta_k^* - \theta^*)J_*(\theta^* - \hat{\theta}) + \frac{n}{2} \|\theta_k^* - \hat{\theta}_k\|^2. \end{aligned}$$

従って, AIC_k は

$$(2.22) \quad \begin{aligned} AIC_k &= -2l_n(\hat{\theta}_k) + 2m_k \\ &= -2l_n^*(\theta_k^*) + m_k - 2\{l_n(\theta^*) - l_n^*(\theta^*)\} + 2n(\theta_k^* - \theta^*)J_*(\theta^* - \hat{\theta})^T \\ &\quad - \{n\|\theta_k^* - \hat{\theta}_k\|^2 - m_k\} \end{aligned}$$

と表される。明らかに, 等式

$$(2.23) \quad E\{AIC_k\} = -2l_n^*(\theta_k^*) + m_k = -2E\{l_n^*(\hat{\theta}_k)\}$$

が成り立つ。この式の値を ELL_k^* で表すことにする。

2.3 AIC の差の分布

n が大きい時、漸近的に

$$(2.24) \quad AIC_j - AIC_k = -2l_n^*(\theta_j^*) + 2l_n^*(\theta_k^*) + 2n(\theta_j^* - \theta_k^*)J_*(\theta^* - \hat{\theta}).$$

が成立する。この量の平均と分散は、それぞれ、

$$(2.25) \quad n \|\theta_j^* - \theta_k^*\|^2 - n \|\theta_k^* - \theta^*\|^2$$

と

$$(2.26) \quad 4n\|\theta_j^* - \theta_k^*\|^2$$

で与えられる。この見積りは $\hat{\theta}$ の漸近正規性から導かれる。 AIC_j と AIC_k の差の標準偏差は

$$(2.27) \quad 2\sqrt{n} \|\theta_j^* - \theta_k^*\|$$

で与えられる事がわかる。 AIC のばらつきが θ_k^* と θ_j^* に依存するという事は2つのモデルと真の分布の三者の関係で決まるということである。

2.4 AIC の差の高次のふるまい

最尤モデルの良さを情報量規準で評価する量 ELL_k を

$$(2.28) \quad ELL_k = -2l_n^*(\hat{\theta}_k) = -2l_n^*(\theta_k^*) + m_k + \{n \|\hat{\theta}_k - \theta_k^*\|^2 - m_k\}$$

で定義しよう。 ELL_k が小さいほどパラメータ $\hat{\theta}_k$ で決まる分布は真の分布に近いことになる。 $\theta_k^* = \theta_j^*$ の場合には、

$$ELL_j - ELL_k = (m_j - m_k) + \{n \|\theta_j^* - \hat{\theta}_j\|^2 - m_j\} - \{n \|\theta_k^* - \hat{\theta}_k\|^2 - m_k\}$$

と

$$(2.29) \quad AIC_j - AIC_k = (m_j - m_k) + \{n \|\theta_k^* - \hat{\theta}_k\|^2 - m_k\} - \{n \|\theta_j^* - \hat{\theta}_j\|^2 - m_j\}$$

から、

$$AIC_j - AIC_k = 2(m_j - m_k) - (ELL_j - ELL_k)$$

が導かれる。この関係式は $|m_j - m_k|$ に比べて $l_n^*(\hat{\theta}_j) - l_n^*(\hat{\theta}_k)$ のあばれかたが大きい場合に ELL と AIC の値の間に逆相関が現れることを意味している。Shimizu (1978) が自己回帰モデルの次数選択に関して指摘したことである。このことによるダメージは $m_k = m_j \gg 1$ のときに大きい。しかし、この場合、 $\Theta_c = \Theta_j \cap \Theta_k$ の次元がかなり小さければ AIC_c が最小値をとることが期待できる。

3. 数値例 —— AIC の分布のイメージ ——

AIC の揺らぎを簡単な重回帰モデルにおける変数選択の例で見よう。データ $\{(y_i, x_{i1}, \dots, x_{ik}) | i=1, 2, \dots, n\}$ をそれぞれ k 次元と1次元の標準正規分布に従う互いに独立な確率変数 (x_{i1}, \dots, x_{ik}) と ε_i と、

$$(3.1) \quad y_i = \sum_{j=1}^k x_{ij} a_j^* + \varepsilon_i$$

で生成した。 $n=100$, a^* は、7番目と8番目の要素が0.3, 他はすべて0のベクトルとした。
このデータに確率密度関数

$$\prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}\left(y_i - \sum_{j=1}^k x_{ij} a_j\right)^2\right\} \prod_{j=1}^k \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x_{ij}^2}{2}\right\}$$

のパラメータに“ $j \in (j_1, \dots, j_r)$ なる j に対して $a_j=0$ ” という形の制約を加えたモデル MODEL(j_1, j_2, \dots, j_r) をあてはめた。たとえば MODEL(1, 2, 3, 4, 5, 6, 7, 8) は真の構造 a^* を含んでいるが, MODEL(9, 10, 11, 12) は含まない。AIC の同時分布の様子を見るために, 表1に示した典型的な6通りのモデルの組合せにおける AIC と ELL の値の同時分布を調べて見た。モデルの対と“真の構造”の関係を図式的に示したのが図1である。

まず, 100組のデータを発生して, そのたびに ELL を計算した結果を図2に示す。シミュレーションデータの解析であるから最尤法であてはめたモデルの (2.28) の値が計算できる。ELL(j_1, \dots, j_r) が MODEL(j_1, \dots, j_r) の ELL を示す。ELL はその周辺分布が「定数」+ χ^2 分布になるようにばらつく。この図で点群の左下で密度が高く, 右上でまばらなのはそのためである。図の中の○印の中心は点群の重心, つまり点の数が無限大であれば (2.20) 式で定義される期待平均対数尤度 ELL* に一致する点である。

表1. モデルの組合せ.

I	MODEL (5, 6, 7, 8, 9, 10, 11, 12)-MODEL (7, 8, 9, 10)
II	MODEL (5, 6, 7, 8, 9, 10, 11, 12)-MODEL (8, 9, 10, 11)
III	MODEL (5, 6, 7, 8, 9, 10, 11, 12)-MODEL (9, 10, 11, 12)
IV	MODEL (5, 6, 7, 8, 9, 10, 11, 12)-MODEL (1, 2, 3, 4, 5, 6, 7, 8)
V	MODEL (5, 6, 7, 8, 9, 10, 11, 12)-MODEL (1, 2, 3, 4, 5, 6, 7, 9)
VI	MODEL (5, 6, 7, 8, 9, 10, 11, 12)-MODEL (1, 2, 3, 4, 5, 6, 9, 10)

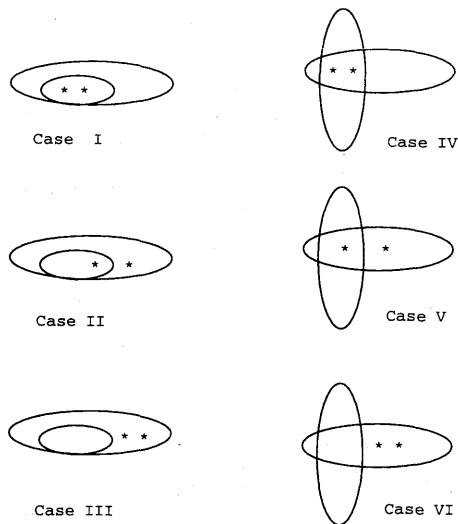


図1. 2つのモデルとパラメータの非零要素の関係.

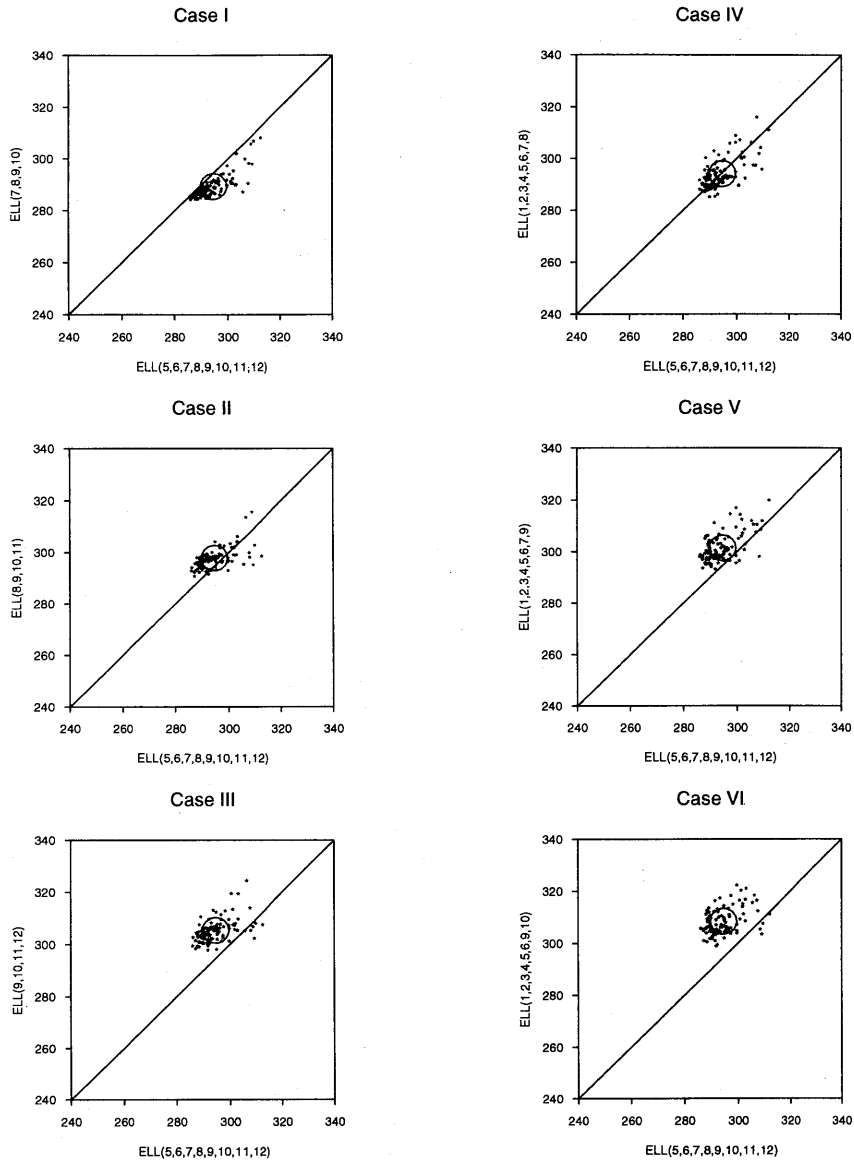


図2. ELL の値の同時分布.

図2から、期待平均対数尤度の意味で、MODEL(7, 8, 9, 10) が MODEL(5, 6, 7, 8, 9, 10, 11, 12) よりよいこと、MODEL(8, 9, 10, 11) より MODEL(5, 6, 7, 8, 9, 10, 11, 12) の方がよいけれど、この場合にはデータの出かたによっては MODEL(8, 9, 10, 11) による結果の方が平均対数尤度の意味ではよくなる場合もあることなどが読みとれる。

ELL が推定できるか否かを見るために6つのモデルの ELL と AIC の関係を見たのが図3である。AIC(j_1, \dots, j_r) は MODEL(j_1, \dots, j_r) の AIC である。この点群の重心の位置に □ がプロットしてある。□ が対角線に重なっていることがすなわち、AIC の不偏性 (2.23) を示している。点群が上下にのびていること、つまり AIC のばらつきがかなり大きいことはモデルの平

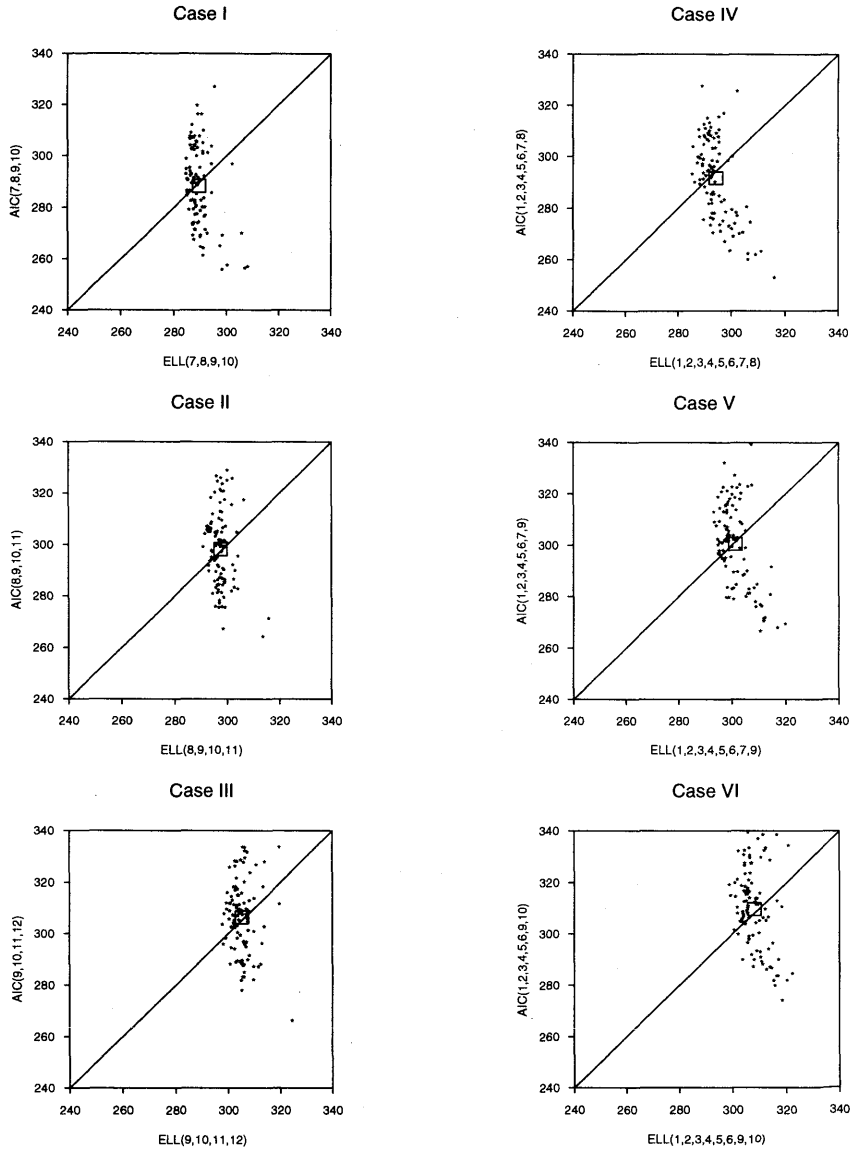


図3. ELL と AIC の同時分布.

均対数尤度,あるいは期待平均対数尤度の推定量として AIC がそれほどシャープでないことを示している. 2つのモデルの対の ELL の差と, AIC の差をプロットしたのが図4である. 図中の AIC0 と ELL0 はそれぞれ AIC(5,6,7,8,9,10,11,12) と ELL(5,6,7,8,9,10,11,12) の意味である. この各図の右上第 I 象元あるいは, 左下の第 III 象元に点が集中していれば, そのモデル対に関する ELL によるモデル選択と AIC によるモデル選択が一致することを意味する.

MODEL(1,2,3,4,5,6,7,8) と MODEL(5,6,7,8,9,10,11,12) を比較する Case IV は 2.4 節で指摘した破滅的な場合である. もちろん, 左上の Case I を見れば分かるように, この場合に

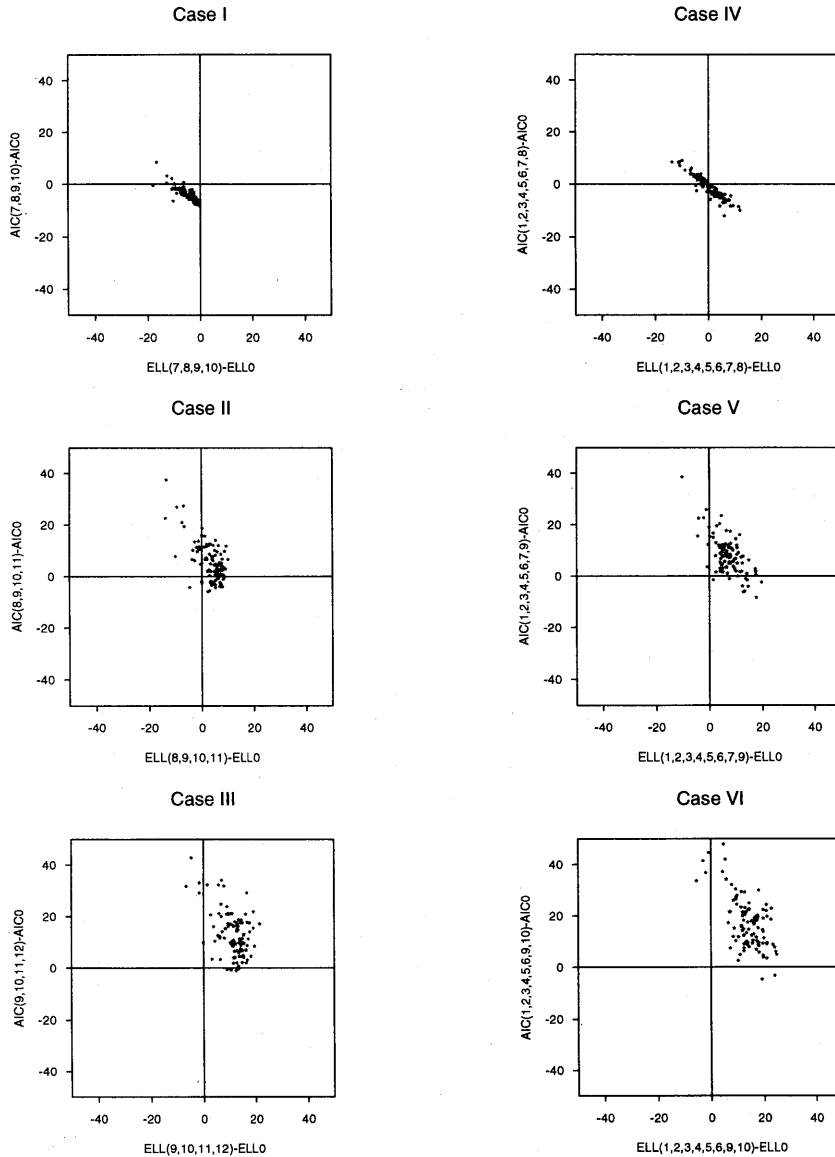


図4. AIC の差と ELL の差の同時分布.

これら2つのモデルと $MODEL(7, 8, 9, 10)$ を比べれば, $MODEL(7, 8, 9, 10)$ を選択するということが ELL と AIC の判定はほとんどの場合に一致する.

以上, ELL と AIC の関係を見たが実データの解析では AIC の値しか求めることはできない. モデル対の AIC の値の同時分布を見てみよう (図5). 図2と同じように各モデルの期待平均対数尤度の位置を中心に○が置いてある. 対角線に関して○と同じ側に点がある場合に, AIC によるモデルの比較が期待平均対数尤度によるものと一致することを考えると, この例題ではモデル比較がうまくいく事が分かる. 1組のデータに対応する AIC の値の組に+印がつけてある. このデータが次章の bootstrap 法で使われるデータである.

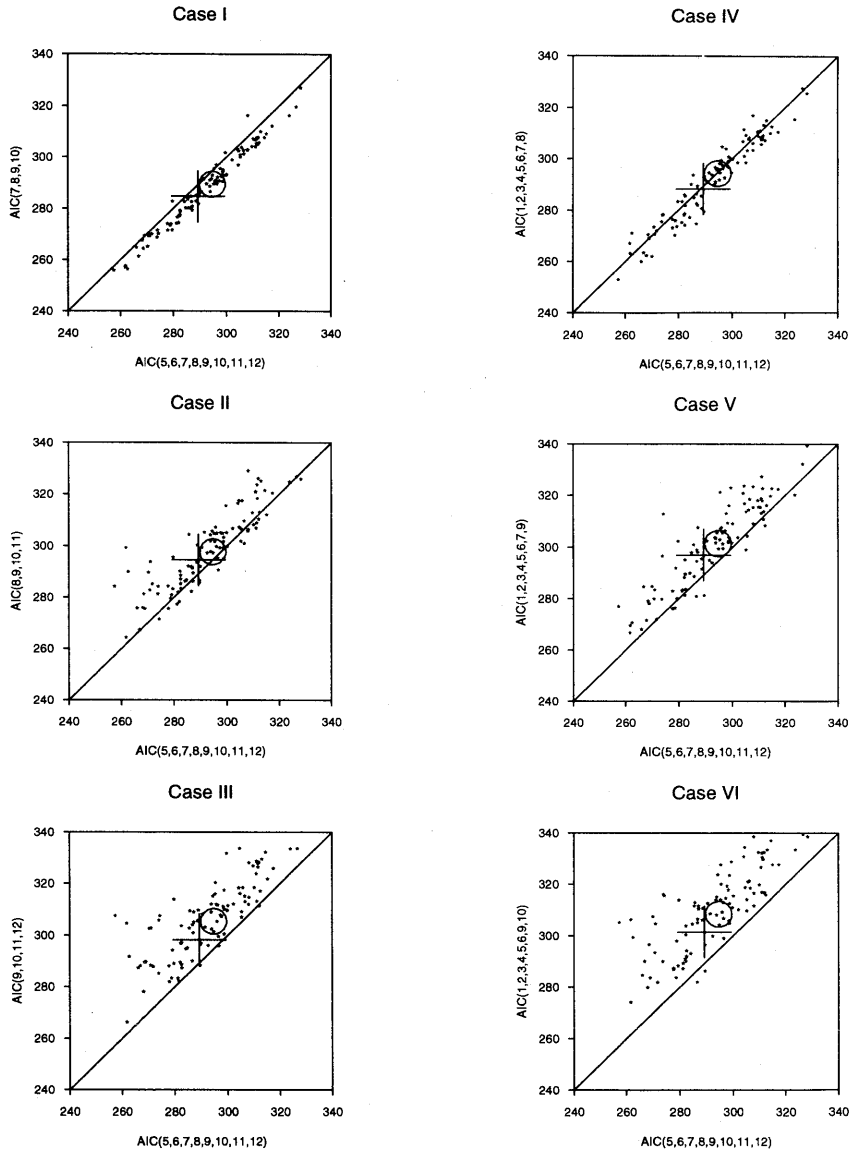


図5. AIC の値の同時分布.

4. AIC の bootstrap 推定

データ $x = \{x_i; i=1, \dots, n\}$ を考える。この時、 $1/n$ の確率で x_i ($i=1, \dots, n$) という値をとる互いに独立な確率変数を X_1^*, X_2^*, \dots で表し、 X^* を $X^* = (X_1^*, X_2^*, \dots, X_n^*)$ で定義する。 X^* の実現値を bootstrap 標本という。 X^* と同じ分布に従う、 X^* とは独立な確率変数を Y^* で表す事にする。

以下、最尤推定量 $\hat{\theta}_k$ が確率変数 X の実現値に依存して決まるときに $\hat{\theta}_k(X)$ と書くことにする。

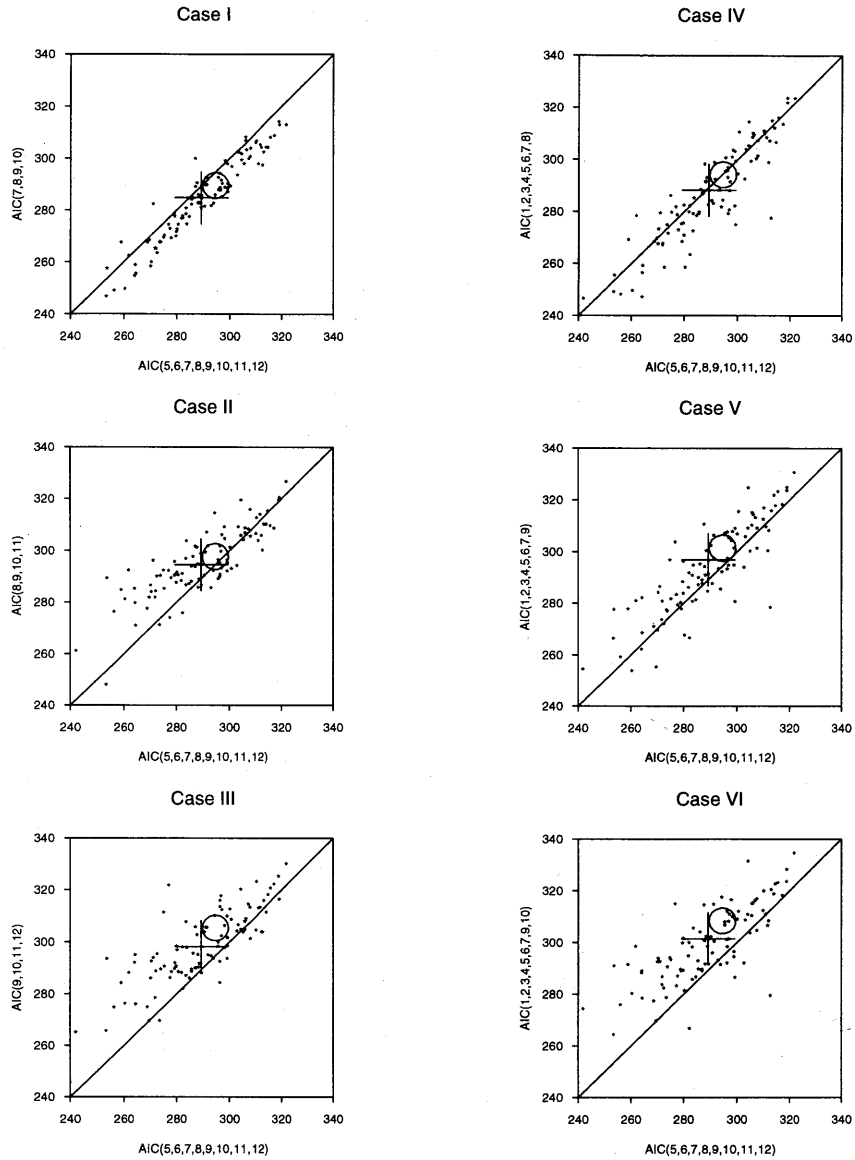


図6. 「AICの真値」のbootstrap推定.

(2.19) から

$$(4.1) \quad E_Y \log f_n(Y | \theta_k^*) = E_{X,Y} \log f_n(Y | \hat{\theta}_k(X)) + \frac{m_k}{2}$$

が得られる。Bootstrap 標本 Y^* についても同様に

$$(4.2) \quad E_{Y^*} \log f_n(Y^* | \theta_k^*) = E_{X^*,Y^*} \log f_n(Y^* | \hat{\theta}_k(X^*)) + \frac{m_k}{2}$$

が成り立つ。ここで θ_k^* は (2.13) 式に対応する

$$(4.3) \quad E_{Y^*} \log f_n(Y^* | \theta_k^*) = \max_{\theta \in \theta_k} E_{Y^*} \log f_n(Y^* | \theta)$$

で定義されるが、bootstrap 標本に関して

$$(4.4) \quad E_{Y^*} \log f_n(Y^* | \theta) = \log f_n(x | \theta)$$

が成立するから $\theta_k^* = \hat{\theta}_k(x)$ となり、(4.2) 式は、結局

$$(4.5) \quad \log f_n(x | \hat{\theta}_k(x)) = E_{X^*, Y^*} \log f_n(Y^* | \hat{\theta}_k(X^*)) + \frac{m_k}{2}$$

に帰着する。(2.21) から得られる

$$(4.6) \quad E_X \log f_n(X | \hat{\theta}_k(X)) = E_Y \log f_n(Y | \theta_k^*) + \frac{m_k}{2}$$

と (4.1) 式から $AIC_k = -2 \log f_n(x | \hat{\theta}_k(x)) + 2m_k$ が

$$(4.7) \quad -2E_{X, Y} \log f_n(Y | \hat{\theta}_k(X))$$

のまわりに分布することが導かれる。同様に bootstrap 標本から計算される

$$AIC_k^* = -2 \log f_n(X^* | \hat{\theta}_k(X^*)) + 2m_k$$

は $-2E_{X^*, Y^*} \log f_n(Y^* | \hat{\theta}_k(X^*))$ のまわりに分布する。これは (4.5) 式から

$$-2 \log f_n(x | \hat{\theta}_k(x)) + m_k = AIC_k - m_k$$

に等しい。結局

$$AIC_k^{*+} = -2 \log f_n(X^* | \hat{\theta}_k(X^*)) + 3m_k$$

と AIC_k の間に、 AIC_k とその「真値」の間の関係が再現されることになる。 AIC_k の同時分布が「真値」に関して対称であれば、 AIC_k の真値、 $-2E_{X, Y} \log f_n(Y | \hat{\theta}_k(X))$ は AIC_k^{*+} が現れる範囲にあるはずである。

点推定、区間推定ということばに対するものとして、 AIC_k^{*+} をプロットする方法をここでは「AIC の bootstrap 推定」と呼ぶ。bootstrap 法で統計量の分散を見ることはしばしば行なわれるが、AIC の場合不偏性がかなりゆるい条件で保障されているので、「推定」というのは正当な表現である。

図5の+印に対応するデータ $x = (x_1, \dots, x_n)$ に基づく AIC の bootstrap 推定を図6に示す。図5の点群が○を中心に分布していて、+印がその中心からはずれていたのに対して、この図では+印のまわりに点がちらばっているのが分かる。図5の○の位置をこの図にも入れてある。点の群の中に○印がとり込まれる形になっている事が分かる。

5. 最後 に

(2.27) 式は下平 (1993a, 1993b) によるものと本質的に同じものであり、Kishino and Hasegawa (1989) で推定量が提案されている。(2.27) 式にただちにつながる (2.21) 式は筆者も共著者の一人である坂元 他 (1983) で求めていたが、差をとることに気がつかなかったのは不覚である。

6. 謝 辞

査読者の方々の適切なご指摘に感謝する。

参 考 文 献

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *2nd International Symposium on Information Theory* (eds. B.N. Petrov and F. Csaki), 267-281, Akademiai Kiado, Budapest (Reproduced in 1992, *Breakthroughs in Statistics*, Vol. 1 (eds. S. Kotz and N.L. Johnson), Springer, New York).
- Efron, B. (1979). Bootstrap methods: another look at the jackknife, *Ann. Statist.*, **7**(1), 1-26 (Reproduced in 1992, *Breakthroughs in Statistics*, Vol. 2 (eds. S. Kotz and N.L. Johnson), Springer, New York).
- Ishiguro, M. (1993). A note on the variance of AIC, Research Memo., No. 488, The Institute of Statistical Mathematics, Tokyo.
- 石黒真木夫 (1994). AIC はなぜ役にたつのか?, *応用数理*, **4**(2), 125-138.
- Kishino, H. and Hasegawa, M. (1989). Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea, *Journal of Molecular Evolution*, **29**, 170-179.
- 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, 東京.
- Shimizu, R. (1978). Entropy maximization principle and selection of the order of an autoregressive Gaussian process, *Ann. Inst. Statist. Math.*, **30**(2), 263-270.
- 下平英寿 (1993a). (不完全データに基づく)情報量規準とその分散を考慮したモデル選択, 第61回 日本統計学会講演報告集, 9-11.
- 下平英寿 (1993b). モデルの信頼集合と地図によるモデル探索, *統計数理*, **41**(2), 131-147.

On Variance/Covariance of AIC 's

Makio Ishiguro

(The Institute of Statistical Mathematics)

Multivariate distribution of AIC 's of models is discussed. Especially the attention is focused on the theoretical variance of the difference of AIC 's. A higher order behavior of AIC is also discussed. A "Bootstrap estimation procedure of AIC " is proposed to provide a visual expression of the variance of AIC 's. A numerical example is given.