統計数理 (1999)
第 47 巻 第 2 号 375-394

一般化情報量規準 GIC とブートストラップ

統計数理研究所　北川源四郎
九州大学*　小西貞則

(受付 1999 年 4 月 7 日；改訂 1999 年 6 月 3 日)

要　旨

情報量規準 AIC は最尤法の利用を前提として導出された．しかし，対数尤度を平均対数尤度の推定量とみなしてバイアス補正を行うという考え方は，より一般の推定法に対しても適用可能である．竹内 (1976) はモデル族が真の分布を含まない場合を想定し TIC を導出している．また Konishi and Kitagawa (1996) はこのバイアス補正の方法が統計的汎関数で定義される一般の推定量に対しても適用できることを示し一般化情報量規準 GIC を導出した．一方，Ishiguro et al. (1997) はブートストラップ法にもとづき，広範なモデルや推定法に適用できる拡張情報量規準 EIC を提案している．

最近，Konishi and Kitagawa (1998) と Kitagawa and Konishi (1998) は GIC の導出法を拡張し，バイアスの二次補正や対数尤度のブートストラップ補正の分散減少法に関する理論を確立した．本稿では，簡単な例を用いながら，GIC とその精密化の解説をする．第 2 節では赤池のバイアス補正の方法を説明し，情報量規準 AIC, TIC, GIC と EIC を示す．第 3 節では GIC の導出法を簡単に説明する．さらにこの方法の拡張として，第 4 節では二次補正された情報量規準について，また第 5 節では EIC の分散減少法について考える．第 6 節では簡単な数値例を示す．

キーワード：モデリング，モデル評価，対数尤度，Kullback-Leibler 情報量，統計的汎関数，バイアス補正，二次バイアス補正．

1. はじめに

統計的モデルの評価規準として提案された情報量規準 AIC (Akaike (1973, 1974)) は，モデルの評価・比較のための客観的な基準を与えたものであり，従来の推測論の枠組みを超える新しいパラダイムとしての統計的モデリングの方向の確立に大きく寄与した．AIC の実質科学へのインパクトの大きさは，5,000 近くにも達し今なお指数的に増加し続ける上記 2 つの論文の引用件数や Bozdogan (ed.) (1994)，赤池・北川 編 (1994, 1995) などにみられる諸科学の現象解明に果たした役割の大きさから窺い知ることができる．

情報量規準 AIC の特徴は，統計的モデルのよさをその予測能力で評価することにある．すなわち，モデルのよさを「真の構造」への近さではなく，観測データとは別にそこから生成されるデータに関する予測精度を見る二標本論の立場をとることにある．しかも，その予測精度を

* 大学院数理学研究科：〒812-8581 福岡市東区箱崎 6-10-1.

二乗誤差のような点推定としてのよさを評価しているのではなく，確率分布で表現した統計的モデルに対して分布としての近さをKullback-Leibler情報量によって評価している．すなわち，AICの背景にある基本的な考え方は

(i) モデルのよさをその予測能力でみる
(ii) 予測は分布の形で行う
(iii) 分布の近さをKullback-Leibler情報量で測る

の3点にあるといえる．

実際のAICはパラメータの推定に最尤法を用いた場合を想定し，モデルの対数尤度の漸近バイアスの近似値による補正を行っているが，上記の考え方はより広く一般の場合に適用できる．竹内(1976)は，想定したモデルが真の分布を含まない場合の最尤推定量の性質を用いて漸近的バイアスを評価した．Konishi and Kitagawa (1996)は一般に，統計的汎関数で定義されるパラメータの推定量を用いた場合にも，同様に漸近的バイアスが評価できることを示し，一般化情報量規準GIC (Generalized Information Criterion)を提案した．一方，Ishiguro et al. (1997)は，ブートストラップ法によってバイアス補正を行うことにより，さまざまな推定量やベイズモデルなどの広汎な場面に適用できる拡張情報量規準EIC (Extended Information Criterion)を提案した．また，Kitagawa (1997)はベイズモデルの予測分布の評価のための規準PICを提案している．

最近，Konishi and Kitagawa (1998)，Kitagawa and Konishi (1998)は統計的汎関数の枠組みでGICの方法を拡張し，対数尤度の2次のバイアス補正量やブートストラップバイアス推定に関わる変動減少法の理論を確立した．これによって，小標本の場合やモデルと真の分布がかなり離れている場合にも，より正確な補正が可能となった．さらに，上記の結果の利用によって，TICやGICのようにバイアス補正量をデータから推定した場合に生じる偏りに関してもいろいろな知見を得ることができる．

本稿は，GICとその拡張の方法を解説し，簡単な例を通して様々な情報量規準を比較検討する．第2節では，赤池によって提案された情報量規準構成のための統計的モデル評価の考え方を述べ，これまでに提案されている情報量規準AIC, TIC, GIC, EICなどを紹介する．第3節ではGICの導出過程を概説し，第4節ではその方法を拡張することによってGICの精密化も可能であることを簡単なモデルを使って説明する．第5節ではブートストラップ情報量規準EICの性質や分散減少の方法について考察する．

2. 統計的モデリングと情報量規準

2.1 統計的モデル

データ $\boldsymbol{X}=\{X_1,\dots,X_n\}$ は，密度関数 $g(x)$ をもつ未知の分布関数 $G(x)$ から生成されるものとする．このとき，$g(x)$ あるいは $G(x)$ を真の分布と呼ぶ．これに対して，この真の分布を近似する密度関数 $f(x)$ を一般にモデルと呼ぶ．特に，データに基づいて構成されたモデルが統計的モデルである．モデルが m 次元のパラメータ $\boldsymbol{\theta}=(\theta_1,\dots,\theta_m)'$ によって規定される場合には，モデルは $f(x|\boldsymbol{\theta})$ と表される．パラメータが適当な空間の集合 Θ の一点として与えられるとき，$\{f(x|\boldsymbol{\theta});\ \boldsymbol{\theta}\in\Theta\}$ はモデル族と呼ばれる．

2.2 Kullback-Leibler 情報量によるモデル評価

Akaike (1973)は真の分布 $g(x)$ に対するモデル $f(x)$ の近さを測る規準として，Kullback-Leibler情報量(1951)

$$I(g\,;f) = E_Y\left[\log \frac{g(Y)}{f(Y)}\right] \tag{2.1}$$

を用いることを提案した. $I(g\,;f)$ の値が小さく 0 に近いほどモデル $f(x)$ は真の分布 $g(x)$ に近いとみなされる. ただし, E_Y は真の分布 $G(y)$ に関する期待値を表す. ここで, $I(g\,;f)$ は

$$I(g\,;f) = E_Y[\log g(Y)] - E_Y[\log f(Y)] \tag{2.2}$$

と分解でき, 右辺の第 1 項はモデル $f(x)$ に依存しない定数であることから, 右辺第 2 項の平均対数尤度が大きいほどモデルは真の分布に近く, 良い近似を与えるモデルであることになる. すなわち, モデルの比較•選択においては Kullback-Leibler 情報量の最小化と平均対数尤度の最大化は同等であることがわかる. 以後, 平均対数尤度が大きいほどモデル $f(x)$ は真の分布 $g(x)$ の良い近似であるとみなすことにする. ただし, 以下の議論からわかるように, 情報量規準の導出には真の分布から生成されたデータ $\boldsymbol{X}$ だけが必要であり, 真の分布 $g(x)$ を具体的な形で表現することは必要ないことに注意する.

実際の統計的な問題では, 真の分布 $G(x)$ は未知であり, $G(x)$ から得られた標本 $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ だけが利用できる. そこで, 平均対数尤度 $E_Y[\log f(Y)] = \int \log f(x)\, dG(x)$ の $G(x)$ を, 各観測値に $1/n$ の等確率を与えた経験分布関数 $\hat{G}_n$ で置き換えると対数尤度関数

$$\ell = n \int \log f(x)\, d\hat{G}_n(x) = \sum_{i=1}^{n} \log f(X_i)$$

が求まる. 対数尤度は, 平均対数尤度 (の n 倍) の自然な推定値であることがわかる.

また, モデルがパラメータ $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)' \in \Theta$ をもつ密度関数 $f(x|\boldsymbol{\theta})$ の形で与えられる場合には, 対数尤度関数

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(X_i|\boldsymbol{\theta}) \equiv \log f(\boldsymbol{X}|\boldsymbol{\theta}) \tag{2.3}$$

の最大化によって最尤推定量 $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\boldsymbol{X})$ が自然に得られる.

2.3 情報量規準

(2.3) 式の最大化によって与えられる最大対数尤度 (の n^{-1} 倍), $n^{-1}\ell(\hat{\boldsymbol{\theta}}(\boldsymbol{X}))$, は平均対数尤度 $E_Y \log f(Y|\hat{\boldsymbol{\theta}}(\boldsymbol{X}))$ の自然な推定量ではあるが, 正のバイアスをもっており, この値を直接モデル選択や比較に用いることはできない. バイアスは, 同じデータ $\boldsymbol{X}$ をパラメータ推定と推定されたモデルの平均対数尤度の推定に二度用いたことによって生じる. この対数尤度のバイアス

$$b(G) = nE_X\left\{\frac{1}{n}\log f(\boldsymbol{X}|\hat{\boldsymbol{\theta}}(\boldsymbol{X})) - E_Y \log f(Y|\hat{\boldsymbol{\theta}}(\boldsymbol{X}))\right\} \tag{2.4}$$

を補正することによって, 平均対数尤度の不偏推定量 $n^{-1}\{\log f(Y|\hat{\boldsymbol{\theta}}(\boldsymbol{X})) - b(G)\}$ が求まる. ただし, 期待値 E_X は標本 $\boldsymbol{X}$ の同時分布に関してとるものとする. したがって, AIC の定義にならって, 一般に情報量規準を

$$\mathrm{IC} = -2\log f(\boldsymbol{X}|\hat{\boldsymbol{\theta}}(\boldsymbol{X})) + 2b(G) \tag{2.5}$$

と定義することにする.

ただし, 一般的枠組みではバイアス $b(G)$ を明示的に求めることは難しいので, 通常は $b(G)$ の近似値として (2.4) 式において $n \to +\infty$ としたときの漸近バイアスを利用する. 真の分布を必ずしも含まないモデルに対する最尤推定量 $\hat{\boldsymbol{\theta}}_{ML}$ の性質 (Huber (1976)) より, この漸近バ

イアスは

$$b_T(G) = \mathrm{tr}\{I(G)J(G)^{-1}\}, \tag{2.6}$$

で与えられる (竹内 (1976))．ただし，$I(G)$ と $J(G)$ はそれぞれ

$$\begin{aligned} I(G) &= E_Y\left[\frac{\partial \log f(Y|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \log f(Y|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right] \\ J(G) &= -E_Y\left[\frac{\partial^2 \log f(Y|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right] \end{aligned} \tag{2.7}$$

で定義される Fisher 情報量行列および Hessian 行列の期待値である．この対数尤度の漸近バイアスを補正して情報量規準

$$\mathrm{TIC} = -2\log f(\boldsymbol{X}|\hat{\boldsymbol{\theta}}_{ML}) + 2\,\mathrm{tr}\{\hat{J}(G)^{-1}\hat{I}(G)\}, \tag{2.8}$$

が得られる．ただし，$\hat{J}(G)$ と $\hat{I}(G)$ はそれぞれ $J(G)$ と $I(G)$ の一致推定量とする．

モデルが真の分布を含む場合には $I(G) = J(G)$ が成り立ち，漸近バイアスは $b_{\mathrm{AIC}}(G) = m$ となる．ただし，m はパラメータベクトル $\boldsymbol{\theta}$ の次元である．したがって，この場合には赤池情報量規準 (Akaike (1973)，坂元 他 (1983))

$$\mathrm{AIC} = -2\log f(\boldsymbol{X}|\hat{\boldsymbol{\theta}}_{ML}) + 2m \tag{2.9}$$

が得られる．

AIC および TIC は最尤法で推定されたモデルの利用を前提としているが，赤池によって提案された対数尤度のバイアス補正の方法は，統計的汎関数 $\boldsymbol{T}(G)$ で定義される一般の推定量 $\hat{\boldsymbol{\theta}} = \boldsymbol{T}(\hat{G}_n)$ に基づくモデル $f(x|\hat{\boldsymbol{\theta}})$ の評価規準の構成に対しても適用できる．ただし，$\boldsymbol{T}(G) = (T_1(G),\ldots, T_m(G))'$ は一点分布と真の分布を含むすべての分布関数上で定義された実数値関数 $T_i(\cdot)$ をその要素とする m 次元汎関数ベクトルとする．実際の推定ではこのような一般に統計的汎関数で定義される推定量に対しては，次節で示すようにその漸近バイアスは

$$b_1(\hat{G}_n) = \frac{1}{n}\sum_{\alpha=1}^{n} \mathrm{tr}\left\{\boldsymbol{T}^{(1)}(X_\alpha;\hat{G}_n)\frac{\partial \log f(X_\alpha|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'}\right\} \tag{2.10}$$

で与えられる (Konishi and Kitagawa (1996))．ただし，$\boldsymbol{T}^{(1)}(X_\alpha;G) = (T_1^{(1)}(X_\alpha;G),\ldots, T_m^{(1)}(X_\alpha;G))'$ は，m 次元影響関数ベクトルでその第 i 要素 $T_i^{(1)}(X_\alpha;G)$ は δ_α を X_α 上のデルタ関数とするとき

$$T_i^{(1)}(X_\alpha;G) = \lim_{\varepsilon\to 0}\frac{T_i((1-\varepsilon)G+\varepsilon\delta_\alpha) - T_i(G)}{\varepsilon} \tag{2.11}$$

で定義される．この漸近バイアスを対数尤度から引いて，一般化情報量規準

$$\mathrm{GIC} = -2\log f(\boldsymbol{X}|\boldsymbol{\theta}) + 2b_1(\hat{G}_n) \tag{2.12}$$

が得られる．最尤推定量に対しては GIC は TIC に一致する．

ブートストラップ法によっても対数尤度のバイアスを推定することができる (Ishiguro et al. (1997), Konishi and Kitagawa (1996), Cavanaugh and Shumway (1997), Shibata (1997))．この方法の特徴はバイアスの存在さえ証明できればその表現を明示的に与える必要がないことから，複雑なモデルや推定量に対して自由に適用できることにある．

ブートストラップ法においては，真の分布 $G(x)$ はデータによって定まる経験分布関数 $\hat{G}_n(x)$ で置き換えられる．ここで，E_{Y^*} は経験分布関数 $\hat{G}_n(y)$ に関する期待値を表すとすると，

$\boldsymbol{\theta}$ の推定量 $\tilde{\boldsymbol{\theta}}(\cdot)$ に対して，(2.4) 式のバイアス $b(G)$ のブートストラップ・バイアス推定量は

$$b_B(\hat{G}_n) = nE_{X^*}\left\{\frac{1}{n}\log f(\boldsymbol{X}^*|\tilde{\boldsymbol{\theta}}(\boldsymbol{X}^*)) - E_{Y^*}\log f(Y^*|\tilde{\boldsymbol{\theta}}(\boldsymbol{X}^*))\right\} \tag{2.13}$$

で与えられる．したがって，$G(x)$ から得られた標本 $\boldsymbol{X}$ と確率変数 Y は，$\hat{G}_n(x)$ から得られた標本，すなわちブートストラップ標本 $\boldsymbol{X}^*$ と $\hat{G}_n(x)$ に従う確率変数 Y^* に置き換えられる．これにともなって，期待値 $E_Y \log f(Y|\cdot)$ は $E_{Y^*}\log f(Y^*|\cdot)$ となることがわかる．ここで

$$E_{Y^*}\log f(Y^*|\tilde{\boldsymbol{\theta}}(\boldsymbol{X}^*)) = \int \log f(y^*|\tilde{\boldsymbol{\theta}}(\boldsymbol{X}^*))\,d\hat{G}_n(y^*) = \frac{1}{n}\log f(\boldsymbol{X}|\tilde{\boldsymbol{\theta}}(\boldsymbol{X}^*)). \tag{2.14}$$

したがって，Y^* に関する期待値は不要となるので，バイアスのブートストラップ推定値は

$$b_B(\hat{G}_n) = E_{X^*}\{\log f(\boldsymbol{X}^*|\tilde{\boldsymbol{\theta}}(\boldsymbol{X}^*)) - \log f(\boldsymbol{X}|\tilde{\boldsymbol{\theta}}(\boldsymbol{X}^*))\} \tag{2.15}$$

で求められる．

実際の計算においては，ブートストラップ•バイアス補正項 $b_B(\hat{G}_n)$ は，$\hat{G}_n(\boldsymbol{X})$ から生成された独立な M 組の標本数 n のブートストラップ標本 $\boldsymbol{X}^*_{(1)},\ldots,\boldsymbol{X}^*_{(M)}$ を用いて，

$$b^*_B(\hat{G}_n) = \frac{1}{M}\sum_{i=1}^{M}\{\log f(\boldsymbol{X}^*_{(i)}|\tilde{\boldsymbol{\theta}}(\boldsymbol{X}^*_{(i)})) - \log f(\boldsymbol{X}|\tilde{\boldsymbol{\theta}}(\boldsymbol{X}^*_{(i)}))\} \tag{2.16}$$

で推定される．

このとき，対数尤度のブートストラップ•バイアス補正は $\log f(\boldsymbol{X}|\tilde{\boldsymbol{\theta}}(\boldsymbol{X})) - b^*_B(\hat{G}_n)$ となり，AIC の定義に従って，ブートストラップ情報量規準が

$$\mathrm{EIC} = -2\log f(\boldsymbol{X}|\tilde{\boldsymbol{\theta}}(\boldsymbol{X})) + 2b^*_B(\hat{G}_n) \tag{2.17}$$

で与えられる (Ishiguro et al. (1997))．このブートストラップ・バイアス修正の方法は，容易に $p(y|\boldsymbol{X}) = \int p(y|\boldsymbol{\theta})\,\pi(\boldsymbol{\theta}|\boldsymbol{X})\,d\boldsymbol{\theta}$ で定義されるベイズモデルの予測分布の場合に拡張することができる．ただし，$\pi(\boldsymbol{\theta}|\boldsymbol{X})$ は $\boldsymbol{X}$ が与えられた場合の $\boldsymbol{\theta}$ の事後分布である．

例. (正規分布モデル)　平均 μ，分散 σ^2 の正規分布を最尤法で推定したモデル

$$f(y|\hat{\mu},\,\hat{\sigma}^2) = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}}\exp\left\{-\frac{(y-\hat{\mu})^2}{2\hat{\sigma}^2}\right\} \tag{2.18}$$

に対する情報量規準，特に対数尤度の漸近バイアスを求めてみる．

(i)　モデルのパラメータは平均と分散の二つであるから，AIC の漸近バイアスは $b_{\mathrm{AIC}} = 2$ となる．

(ii)　TIC を求めるために (2.7) 式の行列 $I(G)$ および $J(G)$ を計算すると以下のようになる．

(2.19)
$$I(G)=E_Y\left[\frac{\partial\log f(Y|\boldsymbol{\theta})}{\partial\theta}\cdot\frac{\partial\log f(Y|\boldsymbol{\theta})}{\partial\theta'}\right]=\begin{bmatrix}\dfrac{1}{\sigma^2} & \dfrac{\mu_3}{2\sigma^6}\\ \dfrac{\mu_3}{2\sigma^6} & -\dfrac{1}{4\sigma^4}\left(1-\dfrac{\mu_4}{\sigma^4}\right)\end{bmatrix}$$
$$J(G)=-E_Y\left[\frac{\partial^2\log f(Y|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]=\begin{bmatrix}\dfrac{1}{\sigma^2} & 0\\ 0 & \dfrac{1}{2\sigma^4}\end{bmatrix}$$

ただし，$\sigma^2(G)$，$\mu_3(G)$，$\mu_4(G)$はそれぞれ真の分布Gの分散および3次，4次の中心化モーメントである．したがって対数尤度の漸近バイアスは

(2.20)
$$b_T(G)=\mathrm{tr}\{J^{-1}(G)\,I(G)\}=\frac{1}{2}+\frac{\mu_4}{2\sigma^4}$$

となる．

(iii) 平均μと分散σ^2の最尤推定量は統計的汎関数を用いて

(2.21)
$$T_\mu(G)=\int xdG(x),\qquad T_{\sigma^2}(G)=\int(x-T_\mu(G))^2\,dG(x)$$

と表すことができる．したがって，汎関数ベクトル$\boldsymbol{T}_{ML}(G)=(T_\mu(G),T_{\sigma^2}(G))'$に対して，その影響関数は

(2.22)
$$T_\mu^{(1)}(x\,;G)=x-\mu,\qquad T_{\sigma^2}^{(1)}(x\,;G)=(x-\mu)^2-\sigma^2$$

となる．一方

(2.23)
$$\left.\frac{\partial\log f(y|\boldsymbol{\theta})}{\partial\mu}\right|_{T_{ML}(G)}=\frac{y-T_\mu(G)}{T_{\sigma^2}(G)}$$
$$\left.\frac{\partial\log f(y|\boldsymbol{\theta})}{\partial\sigma^2}\right|_{T_{ML}(G)}=-\frac{1}{2T_\sigma(G)}+\frac{(y-T_\mu(G))^2}{2(T_{\sigma^2}(G))^2}$$

を用いると，GICのバイアス修正項(2.10)は分散と4次のモーメントを用いて

$$\begin{aligned}b_1(G)&=\int T_\mu^{(1)}(x\,;G)\left.\frac{\partial\log f(x|\boldsymbol{\theta})}{\partial\mu}\right|_{T_{ML}(G)}dG(x)\\&\quad+\int T_{\sigma^2}^{(1)}(x\,;G)\left.\frac{\partial\log f(x|\boldsymbol{\theta})}{\partial\sigma^2}\right|_{T_{ML}(G)}dG(x)\\&=\frac{1}{2}\left(1+\frac{\mu_4}{\sigma^4}\right)\end{aligned}\tag{2.24}$$

となる．当然のことながら，これはTICと一致する．ただし，TICとGICではGは実際には未知なので，Gの代りに経験分布関数$\hat{G}_n$を用いる．

表1.

	正規分布	Laplace 分布
b_{AIC}	2	2
$b_T(G)$	2	3.5
$b_T(\hat{G}_{25})$	1.88	2.60
$b_B^*(\hat{G}_{25})$	2.20	3.09

表 1 は真の分布を正規分布 ($\sigma^2=1,\ \mu_4=3$) あるいは Laplace 分布 ($\sigma^2=1,\ \mu_4=6$) とし，そのパラメータを最尤法で推定した場合に，これまでに議論した情報量規準で用いるバイアス評価量がどうなるかをまとめたものである．TIC，GIC の漸近バイアス $b_T(G)\ (=b_1(G))$ はそれぞれの分布に対して 2 と 3.5 となるが，データ数が $n=25$ の場合にはデータから推定した $b_T(\hat{G}_{25})$ はそれより小さな値となっている．とくに正規分布の場合には 1.88 と AIC の 2 よりも小さくなっている．一方，EIC の補正値 $b_B^*(\hat{G}_{25})$ は正規分布の場合には $b_T(G)$ よりも大きく，また Laplace 分布の場合には $b_T(G)$ と $b_T(\hat{G}_{25})$ の中間の値となっている．これらの理由は 4 節で明らかにする．

2.4 ロバスト推定に基づく統計的モデル

一般に標本空間とパラメータ空間 ($\Theta\subset R^m$) の直積空間上で定義された実数値関数 $\psi_i(X,\boldsymbol{\theta})$ に対して

$$\sum_{\alpha=1}^{n}\psi_i(X_\alpha,\ \hat{\boldsymbol{\theta}}_M)=0,\qquad i=1,\ldots,m \tag{2.25}$$

の解として定義されるのが，ロバスト推定の一つである M-推定量である (Huber (1964))．この場合 M-推定量 $\hat{\boldsymbol{\theta}}_M$ は

$$\int\psi_i(z,\ \boldsymbol{T}_M(G))\,dG(z)=0,\qquad i=1,\ldots,m \tag{2.26}$$

の解として非明示的に定義されるm次元汎関数ベクトル $\boldsymbol{T}_M(G)$ を用いて $\hat{\boldsymbol{\theta}}_M=\boldsymbol{T}_M(\hat{G}_n)$ と表される．ここで，$\boldsymbol{\psi}=(\psi_1,\ldots,\psi_m)'$ とし

$$M(\boldsymbol{\psi},\ G)=-\int\frac{\partial\boldsymbol{\psi}(z,\ \boldsymbol{\theta})'}{\partial\boldsymbol{\theta}}\bigg|_{\boldsymbol{T}_M(G)}dG(z) \tag{2.27}$$

とおくとき，M-推定量の影響関数は

$$\boldsymbol{T}_M^{(1)}(z\,;\,G)=M(\boldsymbol{\psi},\ G)^{-1}\,\boldsymbol{\psi}(z,\ \boldsymbol{T}_M(G)) \tag{2.28}$$

で与えられる (Hampel et al. (1986), p. 230)．したがって (2.12) 式の GIC を適用すると，M-推定量を用いて推定されたモデル $f(z|\hat{\boldsymbol{\theta}}_M)$ の評価規準

$$\begin{aligned}\mathrm{GIC}_M=&-2\sum_{\alpha=1}^{n}\log f(X_\alpha|\hat{\boldsymbol{\theta}}_M)\\&+\frac{2}{n}\sum_{\alpha=1}^{n}\mathrm{tr}\left\{M(\boldsymbol{\psi},\ \hat{G}_n)^{-1}\,\boldsymbol{\psi}(X_\alpha,\ \hat{\boldsymbol{\theta}}_M)\frac{\partial\log f(X_\alpha|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\bigg|_{\hat{\boldsymbol{\theta}}_M}\right\}\end{aligned} \tag{2.29}$$

が得られる．とくに，モデルの族が真の分布を含み Fisher 一致性，すなわちパラメータ空間の任意の $\boldsymbol{\theta}$ に対して $\boldsymbol{T}_M(F_{\boldsymbol{\theta}})=\boldsymbol{\theta}$ が成り立つ場合には，GIC_M の漸近バイアスは $2m$ となり，M-推定量に基づくモデルに対しても AIC と同様な式

$$\mathrm{GIC}_M=-2\sum_{\alpha=1}^{n}\log f(X_\alpha|\hat{\boldsymbol{\theta}}_M)+2m \tag{2.30}$$

が求まる (Konishi and Kitagawa (1996), Konishi (1999))．

2.5 罰則付き最尤法

罰則付き対数尤度関数を一般に

$$\ell(\boldsymbol{\theta}) = \sum_{\alpha=1}^{n} \log f(X_\alpha|\boldsymbol{\theta}) - \frac{n}{2}\lambda\boldsymbol{\theta}' A\boldsymbol{\theta} \tag{2.31}$$

と表す．ただし，Aは$m\times m$非負値定符号行列，$\lambda(>0)$は平滑化パラメータとする．この罰則付き対数尤度関数の最大化に基づく推定量は，Good and Gaskins (1980) によって密度推定の枠組みで提唱され，その後，縮小推定量やベイズモデルとの関係が明らかにされた (Akaike (1980), Kitagawa and Gersch (1984, 1996), Shibata (1989), 田辺 (1989, p. 375))．

罰則付き対数尤度関数 (2.31) 式の最大化によって与えられる推定量を $\hat{\boldsymbol{\theta}}_{PL}$ とすると，モデル $f(X|\hat{\boldsymbol{\theta}}_{PL})$ の評価規準は，(2.29) 式の $\boldsymbol{\psi}$-関数を

$$\boldsymbol{\psi}(X_\alpha, \boldsymbol{\theta}) = \frac{\partial\left\{\log f(X_\alpha|\boldsymbol{\theta}) - \dfrac{\lambda}{2}\boldsymbol{\theta}' A\boldsymbol{\theta}\right\}}{\partial\boldsymbol{\theta}} \tag{2.32}$$

と置いた式によって与えられる (Konishi and Kitagawa (1996), p. 883)．ここでは，例としてロジスティックモデルの推定を取り上げる (Konishi (1999))．

例．(ロジスティックモデル) 目的変数Yとp個の説明変数$\boldsymbol{X}$に対して，n個のデータ$\{(y_\alpha, \boldsymbol{x}_\alpha);\ \alpha=1,\dots, n\}$を観測したとする．ここで，$y_\alpha$はレベル$\boldsymbol{x}_\alpha$の刺激に対して個体が反応すれば$y_\alpha=1$，そうでなければ$y_\alpha=0$を対応させる．このとき，$\boldsymbol{x}_\alpha$の刺激に対する個体の反応確率を$\pi(\boldsymbol{x}_\alpha)$とすると

$$\Pr(Y_\alpha=1|\boldsymbol{x}_\alpha) = \pi(\boldsymbol{x}_\alpha), \qquad \Pr(Y_\alpha=0|\boldsymbol{x}_\alpha) = 1-\pi(\boldsymbol{x}_\alpha) \tag{2.33}$$

であり，Y_αの確率分布はベルヌーイ分布

$$f(y_\alpha|\boldsymbol{x}_\alpha;\boldsymbol{\beta}) = \pi(\boldsymbol{x}_\alpha)^{y_\alpha}(1-\pi(\boldsymbol{x}_\alpha))^{(1-y_\alpha)} \tag{2.34}$$

となる．さらに，反応確率$\pi(\boldsymbol{x}_\alpha)$に対して

$$\pi(\boldsymbol{x}_\alpha) = \frac{\exp(\boldsymbol{x}_\alpha'\boldsymbol{\beta})}{1+\exp(\boldsymbol{x}_\alpha'\boldsymbol{\beta})} \qquad \text{または} \qquad \log\frac{\pi(\boldsymbol{x}_\alpha)}{1-\pi(\boldsymbol{x}_\alpha)} = \boldsymbol{x}_\alpha'\boldsymbol{\beta} \tag{2.35}$$

を仮定したモデルがロジスティックモデルと呼ばれる．ここで$\boldsymbol{\beta}$はm次元パラメータベクトルとする．

パラメータ$\boldsymbol{\beta}$を対数尤度関数にペナルティ項を付与した次の罰則付き対数尤度関数の最大化によって推定する．

$$\ell_\lambda(\boldsymbol{\beta}) = \sum_{\alpha=1}^{n}[y_\alpha\log\pi(\boldsymbol{x}_\alpha) + (1-y_\alpha)\log\{1-\pi(\boldsymbol{x}_\alpha)\}] - \frac{n}{2}\lambda\boldsymbol{\beta}' A\boldsymbol{\beta} \tag{2.36}$$

ここで，$p\times p$行列Aは既知の非負値定符号行列とする．(2.36) 式の解 $\hat{\boldsymbol{\beta}}_\lambda$ はリッジ推定量，λはリッジパラメータとも呼ばれ，λの値を無限に大きくすると$\boldsymbol{\beta}_\lambda$は0へと近づいていく．罰則付き対数尤度関数 (2.36) 式の$\boldsymbol{\beta}$に関する一次微分は

$$\frac{\partial\ell_\lambda(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} = \sum_{\alpha=1}^{n}[\{y_\alpha-\pi(\boldsymbol{x}_\alpha)\}\boldsymbol{x}_\alpha - \lambda A\boldsymbol{\beta}]$$

となる．したがって$\boldsymbol{\psi}$-関数を

$$\boldsymbol{\psi}(y_\alpha, \boldsymbol{\beta}_\lambda) = \{y_\alpha-\pi(\boldsymbol{x}_\alpha)\}\boldsymbol{x}_\alpha - \lambda A\boldsymbol{\beta}_\lambda$$

と置いて，(2.29) 式の情報量規準GIC_Mへ代入すると，罰則付き対数尤度法によって推定され

たモデル $f(y_\alpha|\boldsymbol{x}_\alpha;\ \hat{\boldsymbol{\beta}}_\lambda)$ の評価規準

$$\text{(2.37)} \qquad \mathrm{IC}(\lambda) = -2\sum_{\alpha=1}^{n}[y_\alpha \log \hat{\pi}(\boldsymbol{x}_\alpha) + (1-y_\alpha)\log\{1-\hat{\pi}(\boldsymbol{x}_\alpha)\}] + 2b_\lambda$$

を得る．ただし

$$b_\lambda = \frac{1}{n}\sum_{\alpha=1}^{n}\mathrm{tr}\,[M(\boldsymbol{\psi},\ \hat{G}_n)^{-1}\,\boldsymbol{\psi}(y_\alpha,\ \hat{\boldsymbol{\beta}}_\lambda)\{y_\alpha - \hat{\pi}(\boldsymbol{x}_\alpha)\}\,\boldsymbol{x}_\alpha'],$$

$$M(\boldsymbol{\psi},\ \hat{G}_n) = \frac{1}{n}\sum_{\alpha=1}^{n}\hat{\pi}(\boldsymbol{x}_\alpha)\{1-\hat{\pi}(\boldsymbol{x}_\alpha)\}\,\boldsymbol{x}_\alpha\boldsymbol{x}_\alpha' + \lambda A,$$

$$\hat{\pi}(\boldsymbol{x}_\alpha) = \frac{\exp(\boldsymbol{x}_\alpha'\hat{\boldsymbol{\beta}}_\lambda)}{1+\exp(\boldsymbol{x}_\alpha'\hat{\boldsymbol{\beta}}_\lambda)}$$

とする．このとき，$\mathrm{IC}(\lambda)$ を最小にするパラメータ λ の値を選択する．

3. GIC の導出

本節では (2.10) 式のバイアス補正項の導出法を簡単に示す．m次元パラメータベクトル $\boldsymbol{\theta} = (\theta_1,\ldots,\ \theta_m)'$ の各成分の推定量 $\hat{\theta}_p$ が統計的汎関数により $\hat{\theta}_p = T_p(\hat{G}_n)$ と定義されているものとする．ここで $\hat{\theta}_p = T_p(\hat{G}_n)$ の $T_p(G)$ の周りでの n^{-1} 次までの確率展開は

$$\text{(3.1)} \qquad \hat{\theta}_p = T_p(G) + \frac{1}{n}\sum_{\alpha=1}^{n}T_p^{(1)}(X_\alpha;\ G) + \frac{1}{2n^2}\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}T_p^{(2)}(X_\alpha,\ X_\beta;\ G) + o(n^{-1}),$$

で与えられる．ただし，$T_p^{(j)}$ は対称な汎関数で，任意の分布Hに対して $\varepsilon = 0$ のとき

$$\text{(3.2)} \qquad \frac{d^j}{d\varepsilon^j}T_p\{(1-\varepsilon)G + \varepsilon H\} = \int\cdots\int T_p^{(j)}(x_1,\ldots,\ x_j;\ G)\prod_{i=1}^{j}d\{H(x_i) - G(x_i)\}$$

$$\text{(3.3)} \qquad \int T_p^{(j)}(x_1,\ldots,\ x_j;\ G)\,dG(x_i) = 0 \qquad (1 \leq i \leq j)$$

の二つの式を満たすものとする．

平均対数尤度 $\int \log f(y|\hat{\boldsymbol{\theta}})\,dG(y)$ を $\hat{\boldsymbol{\theta}} = \boldsymbol{T}(G) = (T_1(G),\ldots,\ T_m(G))'$ のまわりで Taylor 展開して，(3.1) 式を代入すると次の確率展開式が得られる．

$$\begin{aligned}\text{(3.4)} \qquad & \int \log f(z|\hat{\boldsymbol{\theta}})\,dG(z) \\ &= \int \log f(z|\boldsymbol{T}(G))\,dG(z) + \frac{1}{n}\sum_{p=1}^{m}\sum_{\alpha=1}^{n}T_p^{(1)}(X_\alpha;\ G)\int\frac{\partial \log f(z|\boldsymbol{\theta})}{\partial\theta_p}dG(z) \\ &\quad + \frac{1}{2n^2}\sum_{p=1}^{m}\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}T_p^{(2)}(X_\alpha,\ X_\beta;\ G)\int\frac{\partial \log f(z|\boldsymbol{\theta})}{\partial\theta_p}dG(z) \\ &\quad + \frac{1}{2n^2}\sum_{p=1}^{m}\sum_{q=1}^{m}\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}T_p^{(1)}(X_\alpha;\ G)\,T_q^{(1)}(X_\beta;\ G)\int\frac{\partial^2 \log f(z|\boldsymbol{\theta})}{\partial\theta_p\partial\theta_q}dG(z) + o(n^{-1})\end{aligned}$$

以下では簡単のために汎関数の微分を

$$\text{(3.5)} \qquad T_{p\alpha}^{(1)} = T_p^{(1)}(X_\alpha;\ G), \qquad T_{p\alpha\beta}^{(2)} = T_p^{(2)}(X_\alpha,\ X_\beta;\ G)$$

と略記すると，(3.4) 式の各項の真の分布Gに関する期待値は

$$
(3.6)\qquad
\begin{aligned}
&E_X\left[\sum_{\alpha=1}^{n} T_{p\alpha}^{(1)}\right]=0\\
&E_X\left[\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n} T_{p\alpha\beta}^{(2)}\right]=E_X\left[\sum_{\alpha=1}^{n} T_{p\alpha\alpha}^{(2)}\right]=nS_p^{(2)}\\
&E_X\left[\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n} T_{p\alpha}^{(1)}T_{q\beta}^{(1)}\right]=\sum_{p=1}^{m}\sum_{q=1}^{m} S_{pq}^{(11)}
\end{aligned}
$$

となる．ただし

$$
(3.7)\qquad
\begin{aligned}
S_p^{(2)} &\equiv \int T_p^{(2)}(x,\,x\,;\,G)\,dG(x),\\
S_{pq}^{(11)} &\equiv \int T_p^{(1)}(x\,;\,G)\,T_q^{(1)}(x\,;\,G)\,dG(x)
\end{aligned}
$$

とする．これから (3.4) 式の平均対数尤度の確率展開式の期待値

$$
(3.8)\qquad
\begin{aligned}
E_X[D_3(\boldsymbol{X}\,;\,G)] &:= E_X\left[\int \log f(z|\boldsymbol{T}(G))\,dG(z)-\int \log f(z|\hat{\boldsymbol{\theta}})\,dG(z)\right]\\
&= -\frac{1}{2}\sum_{p=1}^{m}\left(S_p^{(2)}F_p+\sum_{q=1}^{m} S_{pq}^{(11)}F_{pq}\right)+O(n^{-1})
\end{aligned}
$$

が求まる．ただし，

$$
(3.9)\qquad F_p=\int \frac{\partial \log f(y|\boldsymbol{\theta})}{\partial\theta_p}\,dG(y),\qquad F_{pq}=\int \frac{\partial^2 \log f(y|\boldsymbol{\theta})}{\partial\theta_p\partial\theta_q}\,dG(y)
$$

とする．

同様に，対数尤度の確率展開は以下のようになる．

$$
\begin{aligned}
(3.10)\quad &\frac{1}{n}\sum_{\alpha=1}^{n}\log f(X_\alpha|\hat{\theta})\\
&=\frac{1}{n}\sum_{\alpha=1}^{n}\log f(X_\alpha|\boldsymbol{T}(G))+\frac{1}{n^2}\sum_{p=1}^{m}\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n} T_p^{(1)}(X_\alpha\,;\,G)\frac{\partial \log f(X_\beta|\boldsymbol{\theta})}{\partial\theta_p}\\
&\quad+\frac{1}{2n^3}\sum_{p=1}^{m}\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}\sum_{\gamma=1}^{n} T_p^{(2)}(X_\alpha,\,X_\beta\,;\,G)\frac{\partial \log f(X_\gamma|\boldsymbol{\theta})}{\partial\theta_p}\\
&\quad+\frac{1}{2n^3}\sum_{p=1}^{m}\sum_{q=1}^{m}\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}\sum_{\gamma=1}^{n} T_p^{(1)}(X_\alpha\,;\,G)\,T_q^{(1)}(X_\beta\,;\,G)\frac{\partial^2 \log f(X_\gamma|\boldsymbol{\theta})}{\partial\theta_p\partial\theta_q}+o(n^{-1})
\end{aligned}
$$

ここで

$$
(3.11)\qquad f_p^\alpha=\frac{\partial \log f(X_\alpha|\boldsymbol{\theta})}{\partial\theta_p},\qquad f_{pq}^\alpha=\frac{\partial^2 \log f(X_\alpha|\boldsymbol{\theta})}{\partial\theta_p\partial\theta_q}
$$

と表すことにすると (3.10) 式の各項の期待値は

$$
\begin{aligned}
E_X\left[\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n} T_{p\alpha}^{(1)}f_p^\beta\right] &= \sum_{\alpha=1}^{n} E[T_{p\alpha}^{(1)}f_p^\alpha]=n\int T_{p\alpha}^{(1)}f_p^\alpha dG\\
E_X\left[\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}\sum_{\gamma=1}^{n} T_{p\alpha\beta}^{(2)}f_p^\gamma\right] &= \sum_{\alpha=1}^{n}\sum_{\gamma=1}^{n} E[T_{p\alpha\alpha}^{(2)}f_p^\gamma]=n^2S_p^{(2)}F_p+O(n)\\
E_X\left[\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}\sum_{\gamma=1}^{n} T_{p\alpha}^{(1)}T_{q\beta}^{(1)}f_{pq}^\gamma\right] &= \sum_{\alpha=1}^{n}\sum_{\gamma=1}^{n} E[T_{p\alpha}^{(1)}T_{q\alpha}^{(1)}f_{pq}^\gamma]=n^2S_{pq}^{(11)}F_{pq}+O(n)
\end{aligned}
$$

となる．したがって

$$E_X[D_1(\boldsymbol{X};G)] := E_X\left[\frac{1}{n}\sum_{i=1}^{n}\log f(X_i|\hat{\theta}) - \frac{1}{n}\sum_{i=1}^{n}\log f(X_i|\boldsymbol{T}(G))\right]$$

$$= \frac{1}{n}\int T_{p\alpha}^{(1)} f_p^\alpha dG + \frac{1}{2n}(S_p^{(2)}F_p + S_{pq}^{(11)}F_{pq}) + O(n^{-2}) \tag{3.12}$$

が得られる．

以上の結果を用いると対数尤度のバイアス補正項を評価することができる．まず，第一に (2.4) のバイアス補正項 $b(G) = E_X[D(\boldsymbol{X};G)]$ の中の $D(\boldsymbol{X};G)$ は以下のように 3 つの項に分解することができる．

$$D(\boldsymbol{X};G) = \sum_{i=1}^{n}\log f(x_i|\boldsymbol{T}(\hat{G}_n)) - n\int\log f(y|\boldsymbol{T}(\hat{G}_n))\,dG(y)$$

$$= D_1(\boldsymbol{X};G) + D_2(\boldsymbol{X};G) + D_3(\boldsymbol{X};G) \tag{3.13}$$

ただし，

$$D_1(\boldsymbol{X};G) = \sum_{i=1}^{n}\log f(X_i|\boldsymbol{T}(\hat{G}_n)) - \sum_{i=1}^{n}\log f(X_i|\boldsymbol{T}(G))$$

$$D_2(\boldsymbol{X};G) = \sum_{i=1}^{n}\log f(X_i|\boldsymbol{T}(G)) - n\int\log f(y|\boldsymbol{T}(G))\,dG(y) \tag{3.14}$$

$$D_3(\boldsymbol{X};G) = n\int\log f(y|\boldsymbol{T}(G))\,dG(y) - n\int\log f(y|\boldsymbol{T}(\hat{G}_n))\,dG(y)$$

である．

ここで，$E[D_2(\boldsymbol{X};G)] = 0$ であることから

$$b(G) = E_X[D(\boldsymbol{X};G)] = E_X[D_1(\boldsymbol{X};G)] + E_X[D_3(\boldsymbol{X};G)]$$

$$= \frac{1}{n}\int T_{p\alpha}^{(1)} f_p^\alpha dG + O(n^{-2}) \tag{3.15}$$

$$= \frac{1}{n}\left(\sum_{p=1}^{m}\int T_{p\alpha}^{(1)}(x;G)\frac{\partial\log f(x|\boldsymbol{\theta})}{\partial\theta_p}dG(x)\right) + O(n^{-2}) \tag{3.16}$$

と表すことができる．

4. 対数尤度のバイアスの二次補正

以下では簡単のために次の記号を使うことにする．

$$T_{p\alpha\beta\gamma}^{(3)} = T_p^{(3)}(X_\alpha, X_\beta, X_\gamma;G), \qquad T_{p\alpha\beta\gamma\delta}^{(4)} = T_p^{(4)}(X_\alpha, X_\beta, X_\gamma, X_\delta;G) \tag{4.1}$$

さらに，簡単のために和の記号は省略する．例えば，$T_{p\alpha}^{(1)}T_{q\beta}^{(1)}f_{pq}^\gamma$ は

$$\sum_{p=1}^{m}\sum_{q=1}^{m}\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}\sum_{\gamma=1}^{n} T_{p\alpha}^{(1)}T_{q\beta}^{(1)}f_{pq}^\gamma \tag{4.2}$$

を意味するものとする．

2 節で紹介した情報量規準 AIC，TIC，GIC，EIC は，モデルの平均対数尤度 $E_Y\log f(Y|\hat{\boldsymbol{\theta}})$ の推定量として，対数尤度の漸近バイアスを補正することによって導かれた．一般に漸近バイアスを補正した情報量規準を

$$\log f(\boldsymbol{X}|\hat{\boldsymbol{\theta}}) - b_1(\hat{G}_n) \tag{4.3}$$

と表す．ただし，$b_1(\hat{G}_n)$ は (2.10) で与えられる一次のバイアス補正項とする．本節では，この

情報量規準で平均対数尤度を推定したときの(漸近)バイアスをさらに補正することによって，情報量規準の精密化を行う方法について述べる (Konishi and Kitagawa (1998))．

(4.3) 式の情報量規準で平均対数尤度を推定したときのバイアスは

$$E_X[\log f(\boldsymbol{X}|\hat{\boldsymbol{\theta}}) - b_1(\hat{G}_n) - nE_Y \log f(Y|\hat{\boldsymbol{\theta}})] \\ = E_X[\log f(\boldsymbol{X}|\hat{\boldsymbol{\theta}}) - nE_Y \log f(Y|\hat{\boldsymbol{\theta}})] - E_X[b_1(\hat{G}_n)] \tag{4.4}$$

である．上式右辺の最初の $\boldsymbol{X}$ に関する期待値は対数尤度のバイアスであり，これが次のように展開されたとする．

$$b(G) = E_X[\log f(\boldsymbol{X}|\hat{\boldsymbol{\theta}}) - nE_Y \log f(Y|\hat{\boldsymbol{\theta}})] \\ = b_1(G) + \frac{1}{n} b_2(G) + O(n^{-2}) \tag{4.5}$$

ここで，$b_1(G)$ は (2.10) 式で与えられる対数尤度の一次のバイアス (漸近バイアス) で，二次のバイアス $b_2(G)$ は

$$b_2(G) = \frac{1}{2}\Big(-S_p^{(2)}F_p - S_{qp}^{(11)}F_{pq} + \int T_{pa}^{(1)}T_{qa}^{(1)}f_{pq}^{a}dG + S_q^{(2)}\int T_{pa}^{(1)}f_{pq}^{a}dG + \int T_{paa}^{(2)}f_p^{a}dG \\ + \int T_{paa\beta}^{(3)}f_p^{\beta}dG + 2\int T_{pa}^{(1)}T_{qa\beta}^{(2)}f_{pq}^{\beta}dG + S_{pq}^{(11)}\int T_{ra}^{(1)}f_{pqr}^{a}dG\Big) \tag{4.6}$$

で与えられる．

また，(4.4) 式右辺の一次のバイアスの推定量 $b_1(\hat{G}_n)$ の期待値は

$$E_X[b_1(\hat{G})] = b_1(G) + \frac{1}{n}\varDelta b_1(G) + O(n^{-2}) \tag{4.7}$$

と展開され，$\varDelta b_1(G)$ は次で与えられる．

$$\varDelta b_1(G) = -S_p^{(2)}F_p - S_{pq}^{(11)}F_{pq} + \int T_{paa}^{(2)}f_p^{a}dG - \int T_{pa}^{(1)}f_p^{a}dG \\ + \frac{1}{2}S_{qr}^{(11)}\int T_{pa}^{(1)}f_{pqr}^{(1)}dG + \frac{1}{2}S_q^{(2)}\int T_{pa}^{(1)}f_{pq}^{a}dG + \int T_{pa\beta}^{(2)}T_{q\beta}^{(1)}f_{pq}^{a}dG \\ + \int T_{pa}^{(1)}T_{qa}^{(1)}f_{pq}^{a}dG + \frac{1}{2}\int T_{pa\beta\beta}^{(3)}f_p^{a}dG \tag{4.8}$$

以上より，対数尤度に一次のバイアス補正を施した情報量規準で平均対数尤度を推定したときのバイアスは

$$E_X[\log f(\boldsymbol{X}|\hat{G}_n) - b_1(\hat{G}_n) - nE_Y \log f(Y|\hat{\boldsymbol{\theta}})] \\ = \frac{1}{n}\{b_2(G) - \varDelta b_1(G)\} + O(n^{-2}) \tag{4.9}$$

となる．ただし，$b_2(G)$，$\varDelta b_1(G)$ は各々 (4.6)，(4.8) 式で与えられる．

したがって，対数尤度に二次のバイアス補正を施した情報量規準

$$\mathrm{GIC}_S = -2\log f(\boldsymbol{X}|\hat{\boldsymbol{\theta}}) + 2\left\{b_1(\hat{G}_n) + \frac{1}{n}(b_2(\hat{G}_n) - \varDelta b_1(\hat{G}_n))\right\} \tag{4.10}$$

が求まる．

前節で求めた対数尤度の二次のバイアス $b_2(G)$ や漸近バイアスの推定量 $b_1(\hat{G}_n)$ のもつバイアス $\varDelta b_1(G)$ は，一般の枠組みで求めるとかなり複雑な形をしている．そこで，(2.18) の正規分

表 2.

真の分布	$\frac{1}{n}b_2(G)$	$\frac{1}{n}\Delta b_1(G)$
正規分布	$\frac{6}{n}$	$-\frac{3}{n}$
Laplace 分布	$\frac{6}{n}$	$-\frac{42}{n}$

布モデルの場合，これらの補正項が具体的にどのようになるかを考えてみる．まず統計的汎関数の微分は以下のように与えられる．

$$
\begin{aligned}
&T_\mu^{(1)}(x\,;G)=x-\mu, \qquad T_\mu^{(j)}(x_1,\ldots,x_j\,;G)=0 \qquad (j\geq 2),\\
&T_{\sigma^2}^{(1)}(x\,;G)=(x-\mu)^2-\sigma^2, \qquad T_{\sigma^2}^{(2)}(x,y\,;G)=-2(x-\mu)(y-\mu),\\
&T_{\sigma^2}^{(j)}(x_1,\ldots,x_j\,;G)=0 \qquad (j\geq 3).
\end{aligned} \tag{4.11}
$$

この結果を利用して (4.6)，(4.8) から，次の結果を得る．

$$
b_2(G)=3-\frac{\mu_4}{\sigma^4}-\frac{1}{2}\frac{\mu_6}{\sigma^6}+4\frac{\mu_3^2}{\sigma^6}+\frac{3}{2}\frac{\mu_4^2}{\sigma^8}, \tag{4.12}
$$

$$
\varDelta b_1(G)=3-\frac{3}{2}\frac{\mu_4}{\sigma^4}-\frac{\mu_6}{\sigma^6}+4\frac{\mu_3^2}{\sigma^6}+\frac{3}{2}\frac{\mu_4^2}{\sigma^8}, \tag{4.13}
$$

$$
b_2(G)-\varDelta b_1(G)=\frac{1}{2}\left(\frac{\mu_4}{\sigma^4}+\frac{\mu_6}{\sigma^6}\right), \tag{4.14}
$$

$$
b_1(G)-\frac{1}{n}\varDelta b_1(G)+\frac{1}{n}b_2(G)=\frac{1}{2}\left(1+\frac{\mu_4}{\sigma^4}\right)+\frac{1}{2n}\left(\frac{\mu_4}{\sigma^4}+\frac{\mu_6}{\sigma^6}\right), \tag{4.15}
$$

ただし，μ_j は真の分布 G の j-次のモーメントである．

特に，真の分布 G が正規分布 ($\sigma^2=1$, $\mu_3=0$, $\mu_4=3$, $\mu_6=15$) と Laplace 分布 ($\sigma^2=1$, $\mu_3=0$, $\mu_4=6$, $\mu_6=90$) の場合を表 2 に示す．

以上から，バイアスの推定値 $b_1(\hat{G}_n)$ は $b_1(G)$ の推定量としてバイアスを持っており，その差は n が小さいときには無視できないことを示唆している．AIC の長所のひとつは，バイアス補正項が分布 G に依存せず，したがって $\varDelta b_A(\hat{G}_n)=0$ が成り立つことである．

5. ブートストラップ情報量規準 EIC について

5.1 ブートストラップによる二次補正

(2.4) の対数尤度のバイアスは

$$
\frac{1}{n}b(G)=\frac{1}{n}b_1(G)+\frac{1}{n^2}b_2(G)+\frac{1}{n^3}b_3(G)+\cdots \tag{5.1}
$$

と展開できる．したがって，バイアスのブートストラップ推定値の期待値は

$$
\begin{aligned}
E_X[b_B(\hat{G}_n)]&=E_X\left[b_1(\hat{G}_n)+\frac{1}{n}b_2(\hat{G}_n)\right]+o(n^{-1})\\
&=b_1(G)+\frac{1}{n}\varDelta b_1(G)+\frac{1}{n}b_2(G)+o(n^{-1})
\end{aligned} \tag{5.2}
$$

となる．ただし，$\Delta b_1(G)$ は (4.8) で与えられる 1 次のバイアスの推定量 $b_1(\hat{G}_n)$ のバイアスである．これから，$\Delta b_1(G)=0$ の場合には，ブートストラップ推定量は自動的に 2 次の補正を行っていることが示される．

一方，これとは対照的に (2.8) の GIC (最尤推定量の場合には (2.4) の TIC も同様) の場合は

$$E_X[b_1(\hat{G}_n)]=b_1(G)+\frac{1}{n}\Delta b_1(G)+o(n^{-1}) \tag{5.3}$$

となり，たとえ $\Delta b_1(G)=0$ の場合でも二次のバイアス補正は行っていない．

前節で 2 次のバイアス補正項を解析的に求めたが，モデルが複雑な場合には，実際上 2 次補正項はブートストラップ法によって求める方が有効である．$b_1(G)$ が解析的に求められる場合には，対数尤度のバイアス補正量はブートストラップ推定値

$$\frac{1}{n}b_2^*(\hat{G}_n)=E_{X^*}[\log f(\boldsymbol{X}^*|\hat{\theta}(\boldsymbol{X}^*))-b_1(\hat{G}_n)-nE_{Y^*}\log f(Y^*|\hat{\theta}(\boldsymbol{X}^*))] \tag{5.4}$$

で求められる．一方，1 次補正項を解析的に求めることが困難な場合にも，以下のように二段階ブートストラップによって求めることができる．

$$\frac{1}{n}b_2^{**}(\hat{G}_n)=E_{X^*}[\log f(\boldsymbol{X}^*|\hat{\theta}(\boldsymbol{X}^*))-b_B^*(\hat{G}_n)-nE_{Y^*}\log f(Y^*|\hat{\theta}(\boldsymbol{X}^*))] \tag{5.5}$$

ただし，$b_B^*(G)$ は (2.18) 式で求められる 1 次の補正項のブートストラップ推定値である．二段階ブートストラップ法については，例えば小西 ((1990)，p. 149) を参照されたい．

例．(ベイズ型予測分布モデル) ベイズアプローチに基づく予測分布モデルは，パラメトリックモデル $\{f(x|\boldsymbol{\theta});\,\boldsymbol{\theta}\in\Theta\subset R^p\}$ とパラメータ $\boldsymbol{\theta}$ に対する事前分布 $\pi(\boldsymbol{\theta}|\omega)$ に基づいて

$$h(z|\boldsymbol{X}_n)=\int f(z|\boldsymbol{\theta})\,\pi(\boldsymbol{\theta}|\boldsymbol{X}_n;\,\omega)\,d\boldsymbol{\theta} \tag{5.6}$$

で与えられる．ただし，$\pi(\boldsymbol{\theta}|\boldsymbol{X}_n;\,\omega)$ は，次式で定義される $\boldsymbol{\theta}$ の事後分布とする．

$$\pi(\boldsymbol{\theta}|\boldsymbol{X}_n;\,\omega)=\frac{\prod_{\alpha=1}^{n}f(X_\alpha|\boldsymbol{\theta})\,\pi(\boldsymbol{\theta}|\omega)}{\int\prod_{\alpha=1}^{n}f(X_\alpha|\boldsymbol{\theta})\,\pi(\boldsymbol{\theta}|\omega)\,d\boldsymbol{\theta}} \tag{5.7}$$

ここでは，ハイパーパラメータ ω は与えられたものとする．

予測分布は，(5.7) 式の事後分布を代入することによって

$$h(z|\boldsymbol{X}_n)=\frac{\int\exp\left[n\left\{q(\boldsymbol{\theta}|\boldsymbol{X}_n)+\frac{1}{n}\log f(z|\boldsymbol{\theta})\right\}\right]d\boldsymbol{\theta}}{\int\exp\{nq(\boldsymbol{\theta}|\boldsymbol{X}_n)\}\,d\boldsymbol{\theta}} \tag{5.8}$$

と書き直すことができる．ただし，

$$q(\boldsymbol{\theta}|\boldsymbol{X}_n)=\frac{1}{n}\sum_{\alpha=1}^{n}\log f(X_\alpha|\boldsymbol{\theta})+\frac{1}{n}\log\pi(\boldsymbol{\theta}|\omega) \tag{5.9}$$

とする．

いま，$\hat{\boldsymbol{\theta}}_q$, $\hat{\boldsymbol{\theta}}_q(z)$ を各々 $q(\boldsymbol{\theta}|\boldsymbol{X}_n)$, $q(\boldsymbol{\theta}|\boldsymbol{X}_n)+n^{-1}\log f(z|\boldsymbol{\theta})$ のモードとする．このモードの周りで予測分布の分母，分子の被積分関数をそれぞれ Taylor 展開し，積分の近似式を導く方法が Laplace's methods (Tierney and Kadane (1986), Davison (1986)) である．このラプラス近似を適用すると，次のような形で予測分布の漸近展開式が導かれる．

$$h(z|\boldsymbol{X}_n)=f(z|\hat{\boldsymbol{\theta}})\left\{1+\frac{1}{n}a(z|\hat{\boldsymbol{\theta}})+o_p(n^{-1})\right\} \tag{5.10}$$

ここで，$a(z|\hat{\boldsymbol{\theta}})$ は予測分布の対数尤度の漸近バイアスには影響しない項である．

推定量 $\hat{\boldsymbol{\theta}}$ は，事前分布 $\pi(\boldsymbol{\theta}|\omega)$ が標本数 n に依存するか否かによって異なる．もし，事前分布が $\log\pi(\boldsymbol{\theta}|\omega)=O(1)$ であれば $\hat{\boldsymbol{\theta}}$ は最尤推定量 $\hat{\boldsymbol{\theta}}_{ML}$ となり，$\log\pi(\boldsymbol{\theta}|\omega)=O(n)$ であれば，$\hat{\boldsymbol{\theta}}$ は (5.7) 式の事後分布のモード $\hat{\boldsymbol{\theta}}_B$ となる．

したがって，ベイズ型予測分布モデル $h(z|\boldsymbol{X}_n)$ の評価規準は，(2.29) 式において $\boldsymbol{\psi}$-関数を $\boldsymbol{\psi}(x,\ \hat{\boldsymbol{\theta}})=\partial\log f(x|\hat{\boldsymbol{\theta}}_{ML})/\partial\boldsymbol{\theta}$ および $\boldsymbol{\psi}(x,\ \hat{\boldsymbol{\theta}})=\partial\{\log f(x|\hat{\boldsymbol{\theta}}_B)+\log\pi(\hat{\boldsymbol{\theta}}_B)\}/\partial\boldsymbol{\theta}$ とおくと，一般に

$$\mathrm{GIC}_B=-2\sum_{\alpha=1}^{n}\log h(X_\alpha|\boldsymbol{X}_n)+2\,\mathrm{tr}\{J(\boldsymbol{\psi},\ \hat{G}_n)^{-1}I(\boldsymbol{\psi},\ \hat{G}_n)\} \tag{5.11}$$

で与えられる．

しかしながら，予測分布モデルに対しては，漸近バイアスの補正だけでは十分でない場合が多く，このような場合には対数尤度に対してバイアスの二次補正を行う必要がある．予測分布モデルの二次の漸近バイアス補正項は

$$\begin{aligned}E_X\left[\sum_{\alpha=1}^{n}\log h(X_\alpha|\boldsymbol{X}_n)-\mathrm{tr}\{J(\boldsymbol{\psi},\ \hat{G}_n)^{-1}I(\boldsymbol{\psi},\ \hat{G}_n)\}-n\mathrm{E}_G[h(Z|\boldsymbol{X}_n)]\right]\\=\frac{1}{n}b_{(2)}(G)+O(n^{-2})\end{aligned}$$

で与えられる $b_{(2)}(G)$ の推定量として定義される．予測分布モデルに対しては，一般にこの項を精確に求めることは難しく，このような問題に対しては (5.4) 式のブートストラップ法の適用が実用的である．このとき，二次のバイアス補正を施したベイズ型予測分布モデルの評価規準は，次式で与えられる．

$$\mathrm{GIC}_B=-2\sum_{\alpha=1}^{n}\log h(X_\alpha|\boldsymbol{X}_n)+2\,\mathrm{tr}\{J(\boldsymbol{\psi},\ \hat{G}_n)^{-1}I(\boldsymbol{\psi},\ \hat{G}_n)\}+\frac{2}{n}b_{(2)}^{*}(\hat{G}_n) \tag{5.12}$$

5.2 バイアス項の分解とブートストラップ分散の減少

ブートストラップ法の適用上の問題点は，バイアス推定値の分散の大きさにある．ブートストラップ・バイアス推定量の分散は，標本数 n とブートストラップ標本の反復回数Mに依存する．したがって，(2.16) 式において精度の良い $b_B^*(\hat{G}_n)$ を求めるためには，多数の反復回数Mを必要とし，とくに複雑なモデルに対しては膨大な計算量となる．(2.16) のバイアスのブートストラップ推定量の分散は以下のようにバイアス項を分解することによって著しく減少させることができる (Konishi and Kitagawa (1996), Ishiguro et al. (1997))．この分散減少の方法は Konishi and Kitagawa (1996) でも示されているが，汎関数で定義される推定量に対する証明はなされていない．ここでは，ブートストラップ・バイアス推定の分散の減少法を理論的・実際的な観点から検討することにする．

まず，第一に (2.4) のバイアス補正項 $b(G)=E_X[D(\boldsymbol{X}\,;\,G)]$ の中の $D(\boldsymbol{X}\,;\,G)$ を (3.13) 式

と同様に，以下のように3つの項に分解する.

$$D(\boldsymbol{X}\,;\,G)=D_1(\boldsymbol{X}\,;\,G)+D_2(\boldsymbol{X}\,;\,G)+D_3(\boldsymbol{X}\,;\,G) \tag{5.13}$$

ただし，$D_i(i=1,2,3)$ は (3.14) 式で定義されるものとする．ここで，$\hat{\boldsymbol{\theta}}$ が最尤推定量の場合には，$\boldsymbol{T}(G)$ と $\boldsymbol{T}(\hat{G}_n)$ はそれぞれ $\int\log f(y|\boldsymbol{T}(G))\,dG(y)$ と $\sum_{i=1}^{n}\log f(X_i|\boldsymbol{T}(\hat{G}_n))$ を最大にする値である.

統計的汎関数 $\hat{\boldsymbol{\theta}}=\boldsymbol{T}(\hat{G}_n)$ で定義される推定量に対しては,

$$\begin{aligned}
&b_{11}(\boldsymbol{X})\equiv E_X[D_1]=\int T_{pa}^{(1)}f_p^a dG+\frac{1}{2}(S_p^{(2)}F_p+S_{pq}^{(11)}F_{pq})+O(n^{-1})\\
&b_{13}(\boldsymbol{X})\equiv E_X[D_3]=-\frac{1}{2}(S_p^{(2)}F_p+S_{pq}^{(11)}F_{pq})+O(n^{-1})\\
&\mathrm{Var}\,(D_1)\equiv\mathrm{Var}\,(D_3)=nF_pS_{pq}^{(11)}F_q+O(1)\\
&b_{12}(\boldsymbol{X})\equiv E_X[D_2]=0,\qquad \mathrm{Var}\,(D_2)=an+O(1),
\end{aligned} \tag{5.14}$$

が成り立つことが示される．ただし，$a=n^2E_X[\{\log f(\boldsymbol{X}|\hat{\boldsymbol{\theta}}_{ML})-E_Y\log f(Y|\hat{\boldsymbol{\theta}}_{ML})\}^2]$ である (Kitagawa and Konishi (1998)).

さらに，$\mathrm{Var}\,\{D\}=O(n)$ であるのに対して，$\mathrm{Var}\,\{D_1+D_3\}=O(1)$ が成り立つ．したがって，(5.6) 式の対数尤度のブートストラップ・バイアス推定値は $E_X(D_2)=0$ であることから，バイアスを

$$b^*(\hat{G}_n)=E_{X^*}[D_1+D_3], \tag{5.15}$$

によって推定することにより，n が大きな場合には，分散を著しく減少させることができる．特に，最尤推定値の場合には (5.14) 式で $F_p=0$ が成り立つので $\mathrm{Var}\,(D_1)=O(1)$ および $\mathrm{Var}\,(D_3)=O(1)$ が成り立つことに注意する必要がある.

Cavanaugh and Shumway (1997) は最尤推定量の場合，ブートストラップ推定値を $2D_3$ で求めることを提案している．最尤推定量に対しては漸近的に $E_X[2D_3]=E_X[D_1+D_3]$ が成り立ち，さらに D_2 を除去することにより分散減少を実現できる．しかしながら，実際には $E_X[D_1]$ と $E_X[D_3]$ の高次の項は異なっており，特に，最尤推定量ではない推定量に対しては表6に示すようにまったく異なった値となる．一方，Shibata (1997) では，我々の記号では，D, $2D_1$, $2D_3$, D_1+D_2, D_2+D_3, D_2 に対応する6つの尤度比型のブートストラップ推定量を比較している．ただし，最も自然な推定量と考えられる D_1+D_3 は取り上げられていない.

6. 数 値 例

本節では，二次補正，分散減少などを2.4節で用いた正規分布モデルを用いて例示する．ただし真の分布としては以下の2つの分布を考える.

$$\begin{aligned}
&\text{Normal:}\quad g(x)=\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2}\right\}\\
&\text{Laplace:}\quad g(x)=\frac{1}{\sqrt{2}}\exp\{-\sqrt{2}|x|\}
\end{aligned} \tag{6.1}$$

これらの場合，モーメントは $\mu_3=0$, $\mu_4=3$, $\mu_6=15$ (正規分布の場合) $\mu_3=0$, $\mu_4=6$, $\mu_6=90$ (Laplace 分布の場合) となり，バイアス補正量は表2に示したように標本数だけの関数となる.

表 3．正規分布モデル，最尤推定量の場合：真のモデル正規分布．

n	True	B_1	B_2	$\hat{B}_1$	$\hat{B}_2$	B_1^*	B_2^*	B_2^{**}
25	2.27	2.00	2.24	1.89	2.18	2.20	2.24	2.33
100	2.06	2.00	2.06	1.97	2.06	2.04	2.06	2.06
400	2.02	2.00	2.02	1.99	2.02	2.01	2.02	2.02

表 4．分散減少法の効果．

n	25		100		400	
D	0.023	0.237	0.057	0.113	0.206	0.223
D_1+D_3	0.008	0.231	0.005	0.061	0.004	0.019

表 5．正規分布モデル，最尤推定量の場合：真のモデル Laplace 分布．

n	True	B_1	B_2	$\hat{B}_1$	$\hat{B}_2$	B_1^*	B_2^*	B_2^{**}
25	3.87	3.50	3.74	2.60	3.28	3.09	3.30	3.52
100	3.57	3.50	3.56	3.16	3.49	3.33	3.50	3.50
400	3.56	3.50	3.52	3.40	3.51	3.43	3.51	3.50

表 3 は，真の分布を正規分布と仮定し，標本数を $n=25$，100，400 とした 3 つの場合について 10,000 回のモンテカルロ実験で求めたバイアス補正量を示す．True は真の値 $b(G)$ で表 3 の場合には解析的に評価でき $2n/(n-3)$ で与えられる．B_1，B_2 はそれぞれ，1 次と 2 次の補正量を表す．また，$\hat{}$ は真の分布 G の代わりに経験分布関数 $\hat{G}_n$ を代入した場合，* と ** はそれぞれ 1000 回のブートストラップおよび二段階ブートストラップ法による推定値を表す．

この場合，モデルが真の分布を含むので B_1 は AIC の補正項と一致する．$n=400$ の場合，漸近バイアス B_1 や他のすべてのバイアス推定値は真の値に近く，よい近似値となっている．一方 $n=25$ の場合には，B_1 は真の値とかなり異なっているが，B_2 は良い近似値を与えている．実際の場合には，真の分布 G は未知であり常に $\hat{B}_1$ や $\hat{B}_2$ を用いることに注意する必要がある．この場合，$\hat{B}_1=1.89$ は B_1 よりかなり小さな値をとることに注意する必要があるが，その差 0.11 は 1 次のバイアス補正項のバイアス値 $\Delta B_1=-3/25=-0.12$ とよく対応している．2 次の補正項 $\hat{B}_2$ は $n=100$ や 400 の場合，ほとんど厳密な値を与えている．1 次のブートストラップ推定値 B_1^* は 2 次の解析的補正項 B_2 に近い値を与えている．これは，この例の場合にはモデルが真の分布を含むので，5.1 節で議論したように B_1 は定数となり，したがって $\Delta B_1=0$ となることからブートストラップ推定値は自動的に 2 次の補正を行っていることになる．

表 4 は $n=25$，100，400 の場合のブートストラップ推定値の分散を示す．各 n に対して左側の数値は (2.13) 式のブートストラップ補正項の $\hat{G}_n$ の変化による分散を示す．一方，右側の数値は，(2.16) 式によるブートストラップ推定値の分散でこの項は M に反比例するものと考えられる．この表から，D_1 と D_3 に分解する方法は特に n が大きいときに劇的に効果があることがわかる．また，右側の数値が示すように標本自身の変動があるので，むやみにブートストラップの反復数を増やしても意味がないことがわかる．

表 5 は Laplace 分布（両側指数分布）を真の分布とした場合である．この場合，B_1 と B_2 は AIC の 2 と比較して，真の値のかなりよい推定値を与える．しかしながら，$\hat{G}_n$ を用いて推定した $\hat{B}_1$ と $\hat{B}_2$ はかなり大きなバイアスを持っている．これは 1 次のバイアス補正項のバイアス $\Delta b_1=-42/n$ から理解できる．この場合にも，ブートストラップ推定値 B_1^* は $\hat{B}_2$ と同程度に D_1+D_3 に近い値を与える．

表6. 正規分布モデル，Median 推定量の場合：真のモデル正規分布.

n	25			100			400		
	D_1+D_3	D_1	D_3	D_1+D_3	D_1	D_3	D_1+D_3	D_1	D_3
True	2.58	−0.47	3.04	2.12	−0.18	2.30	2.02	−0.16	2.18
B_1	1.89	0.94	0.94	1.97	0.98	0.98	1.99	0.99	0.99
B_1^*	2.57	−0.56	3.14	2.25	−0.37	2.61	2.06	−0.19	2.25
B_1^{**}	2.63	−0.54	3.16	2.27	−0.35	2.62	2.06	−0.19	2.26

最尤推定量以外の方法でパラメータを推定した場合の例として，表6はメディアン $\hat{\mu}_m=\mathrm{med}_i\{X_i\}$ および平均偏差のメディアン $\hat{\sigma}_m=c^{-1}\mathrm{med}_i\{|X_i-\mathrm{med}_j\{X_j\}|\}$，ただし $c=\Phi^{-1}(0.75)$ をそれぞれ μ および σ の推定量として用いた場合の結果を示す．ブートストラップ法はこのような推定値に対しても適用できる．この場合，D_1 と D_3 の平均は全く異なった値をとるが，ブートストラップ法はこの場合にも適当な推定値を与える．一方，漸近バイアス $b_1(G)$ は最尤推定量の場合と同じであるが，$n=100$ や 400 の場合には，AIC が適当な近似値を与えることは興味深い (Konishi and Kitagawa (1996))．

Cavanaugh and Shumway (1997) によって提案されたブートストラップ推定値は $2D_3$ と同等である．彼らの方法は，もともと最尤推定量に対して提案されたものであるが，表6からそれ以外の推定量に対してはまったく利用できないことがわかる．また，表5のように，$\hat{\theta}$ が最尤推定量の場合にも，D_3 は一般に D_1 より大きいので，$2D_3$ は D_1+D_3 の過大評価値を与える．

参考文献

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *2nd International Symposium in Information Theory* (eds. B. N. Petrov and F. Csaki), 267-281, Akademiai Kiado, Budapest. (Reproduced in *Breakthroughs in Statistics, Vol. I, Foundations and Basic Theory* (eds. S. Kots and N. L. Johnson), 610-624, Springer, New York, (1992).)

Akaike, H. (1974). A new look at the statistical model identification, *IEEE Trans. Automat. Control*, **AC-19**, 716-723.

Akaike, H. (1980). Likelihood and the Bayes procedure, *Bayesian Statistics* (eds. J. M. Bernardo, M. H. De Groot, D. V. Lindley and A. F. M. Smith), 143-166, University Press, Valencia, Spain.

赤池弘次，北川源四郎 編 (1994).『時系列解析の実際 I』，統計科学選書 3，朝倉書店，東京.

赤池弘次，北川源四郎 編 (1995).『時系列解析の実際 II』，統計科学選書 4，朝倉書店，東京.

Bozdogan, H. (ed.) (1994). *Proceedings of the First US/Japan Conference on the Frontiers of Statistical Modeling: An Informational Approach*, Kluwer, Dordrecht.

Cavanaugh, J. E. and Shumway, R. H. (1997). A bootstrap variant of AIC for state space model selection, *Statistica Sinica*, **7**, 473-496.

Davison, A. C. (1986). Approximate predictive likelihood, *Biometrika*, **73**, 323-332.

Good, I. J. and Gaskins, J. R. (1980). Density estimation and bump hunting by the penalized likelihood method exemplified by scattering and meteorite data, *J. Amer. Statist. Assoc.*, **75**, 42-73.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust Statistics. The Approach Based on Influence Functions*, Wiley, New York.

Huber, P. J. (1964). Robust estimation of a location parameter, *Ann. Math. Statist.*, **35**, 73-101.

Huber, P. J. (1976). The behavior of maximum likelihood estimates under nonstandard conditions, *Proc. Fifth Berkeley Symp. on Statist.*, 221-233, University of California Press, Berkeley.

Ishiguro, M., Sakamoto, Y. and Kitagawa, G. (1997). Bootstrapping log likelihood and EIC, an extention of AIC, *Ann. Inst. Statist. Math.*, **49**(3), 411-434.

Kitagawa, G. (1997). Information criteria for the predictive evaluation of Bayesian models, *Comm. Statist. Theory Methods*, **26**(9), 2223-2246.

Kitagawa, G. and Gersch, W. (1984). A smoothness priors-state space modeling of time series with trend and seasonality, *J. Amer. Statist. Assoc.*, **79**(386), 378-389.

Kitagawa, G. and Gersch, W. (1996). *Smoothness Priors Analysis of Time Series*, Lecture Notes in Statist., No. 116, Springer, New York.

Kitagawa, G. and Konishi, S. (1998). Bias and variance reduction techniques for log-likelihood based information criterion, Research Memo., No. 713, The Institute of Statistical Mathematics, Tokyo.

小西貞則 (1990). ブートストラップ法と信頼区間の構成, 応用統計学, **19**(3), 137-162.

Konishi, S. (1999). Statistical model evaluation and information criteria, *Multivariate, Design of Experimemts and Sampling* (ed. S. Ghosh), 369-399, Marcel Dekker, New York.

Konishi, S. and Kitagawa, G. (1996). Generalised information criteria in model selection, *Biometrika*, **83**(4), 875-890.

Konishi, S. and Kitagawa, G. (1998). Second order bias correction for generalized information criterion, Research Memo., No. 663, The Institute of Statistical Mathematics, Tokyo.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency, *Ann. Statist.*, **22**, 79-86.

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.

Shibata, R. (1989). Statistical aspects of model selection, *From Data to Model* (ed. J. C. Willems), 215-240, Springer, New York.

Shibata, R. (1997). Bootstrap estimate of Kullback-Leibler information for model selection, *Statistica Sinica*, **7**, 375-394.

竹内 啓 (1976). 情報統計量の分布とモデルの適切さの規準, 数理科学, **3**, 12-18.

田辺國士 (1989). 平滑化, 『統計学辞典』 (竹内 啓 編), 375-380, 東洋経済新報社, 東京.

Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities, *J. Amer. Statist. Assoc.*, **81**, 82-86.

Generalized Information Criterion GIC and the Bootstrap

Genshiro Kitagawa
(The Institute of Statistical Mathematics)

Sadanori Konishi
(Graduate School of Mathematics, Kyushu University)

The information criterion AIC is obtained by assuming the use of the maximum likelihood estimators. However, the basic idea of the bias correction for the log-likelihood as an estimator of the expected log-likelihood can be applied to a wider class of models and estimation procedures. Actually, Takeuchi (1976) proposed the criterion TIC for the situation where the model class does not contain the true model. Konishi and Kitagawa (1996) showed that this method of bias correction of the log-likelihood can be generalized to estimators defined by statistical functionals and derived the GIC (Generalized Information Criterion). On the other hand, Ishiguro et al. (1997) proposed a bootstrap based information criterion EIC (Extended Information Criterion) which can be applied to very broad class of models and estimation methods.

Recently, Konishi and Kitagawa (1998) and Kitagawa and Konishi (1998) extend the method used for the derivation of GIC, and established a theory for the second order bias correction and the variance reduction for the bootstraping log-likelihood. The amount of the bias terms in estimating the information criteria such as TIC and GIC are also explained by this method.

In this article, we explain the GIC criterion and its refinement and exemplify by using a simple model. In Section 2, we review Akaike's method of statistical model evaluation and show information criteria AIC, TIC, GIC and EIC. Section 3 is devoted to a brief derivation of GIC. The second order bias corrected information criterion is shown in Section 4 and a method of reducing the bootstrap variance in computing EIC is discussed in Section 5. Numerical examples are shown in Section 6.

Key words: Modeling, model evaluation, log-likelihood, Kullback-Leibler information, statistical functional, bias correction, second order correction.