遠洋水研報　第38号　21-28頁　平成13年3月
Bull. Nat. Res. Inst. Far Seas Fish., No. 38, 21-28, March 2001

情報量規準ＴＩＣとｃ-ＡＩＣによるモデル選択の有効性

庄野　宏*

Efficiency of Model Selection by Information Criteria, TIC and c-AIC

Hiroshi SHONO*

In a model selection, AIC tends to overestimate the number of unknown parameters in several cases. However, c-AIC and TIC correct the bias of AIC in small samples and where the true model does not include the candidate one in nested model, respectively. I calculated the values of AIC, c-AIC, TIC and BIC using a virtual example dealing with CPUE analysis. A finally selected model by c-AIC is simpler than that by other criteria (AIC, TIC and BIC). I carried out the computer simulation by ANOVA type model corresponding to CPUE standardization. Especially, I compared the selection performance of TIC to that of AIC c-AIC, and BIC using the nested model that the true model does not include candidate one. As a result of simulation, the efficiency of TIC is slightly better than that of AIC and c-AIC. In addition, the value of TIC is not so greatly different from that of AIC in the statistical model with normal error (such as ANOVA or linear regression), theoretically. Therefore, there seems no need to use TIC instead of AIC in such cases because of complexity of TIC formula.

Key words: TIC, AIC, c-AIC, BIC, nested model, CPUE standardization

1．はじめに

ＣＰＵＥ標準化，網目選択曲線の推定，標識放流モデルの推定など水産資源分野の多くの問題において，情報量規準ＡＩＣ（Akaike, 1973）がモデル選択の指標として広く用いられている．ＡＩＣは真の分布との距離を測る一つの基準であるKulbuck-Leibler情報量の期待値を計算したものであるが，その導出時に標本数が無限大であるといった漸近性な近似を用いているため，有限の標本（特に小標本）の場合には偏りを生じ，その結果パラメーター数を過大に推定する傾向がある（i.e. 複雑なモデルを選択する）ことが知られている（Shono, 2000）．このため，ＡＩＣの結果を盲目的に信じるとモデル選択を誤る危険性がある．

そこで，本報告ではこのＡＩＣが持つ過大推定の問題点について系統的に整理し，これらの問題点を解消する情報量規準として提示されているＴＩＣ（竹内，1976），ｃ-ＡＩＣ（Sugiura, 1978）について検討した．ここでは例として分散分析モデルに対応するＣＰＵＥ標準化を取り上げ，仮想例によるモデル選択やシミュレーション実験を通じてこれら情報量規準のパフォーマンス評価を行った．

2．情報量規準ＡＩＣにおける偏り

まず，本稿で扱う正規モデルの式と記号を次のように定義する．

$$Y = X\Theta + \varepsilon = \mu + \varepsilon,\ \varepsilon \sim N(0, \sigma^2 I_n) \rightarrow Y \sim N(\mu, \sigma^2 I_n) \tag{2.1}$$

$l(\Theta|Y)$：対数尤度関数

$\hat{\Theta}$：未知母数Θの最尤推定量

n：標本数

p：（局外母数σ^2を除く）パラメーター数

RSS：残差平方和（residual sum of squares）

このとき，候補となるモデルMのＡＩＣは

$$AIC(M) = -2\ l(\hat{\Theta}|Y) + 2\ (p+1) \tag{2.2}$$

と表され，ＡＩＣが最小のモデルが最適なモデルとして選択される．しかし，

1）小標本の場合

2）p/n(パラメーター数/標本数)が大きい場合

3）真のモデルが候補のモデルを含まない場合

2000年12月14日受理　遠洋水産研究所業績　第423号
Accepted on December 14, 2000. Contribution No. 423 from the National Research Institute of Far Seas Fisheries.
*遠洋水産研究所（National Research Institute of Far Seas Fisheries; 5-7-1 Orido, Shimizu, Shizuoka 424-8633, Japan）.

には，ＡＩＣの値に偏りが生じ，パラメーター数を過大推定してしまう傾向があることが知られている（坂元ほか，1983；竹内，1976）．

そこで，これらの場合に情報量規準ＡＩＣを用いることによって生じる問題点について整理し，偏りを修正するような情報量規準について述べる．

２－１．小標本の場合

ＡＩＣはその導出過程で漸近理論（標本数が無限にあると仮定した上での理論）を用いているため，標本数が20ないし50以下であるような小標本の場合には偏りが生じてしまう．ＡＩＣの導出については坂元（1985），庄野（2000）などに詳しく書かれているが，端的に言えば，標本Yと独立な変量をZとしたときにZの真の分布と予測分布のKulbuck-Leibler情報量の期待値を漸近的に評価していることになる．

すなわち，標本の確率密度関数$f(\cdot|\Theta)$に対して，

$$E^{y}\left[\int f(z|\Theta)\log\frac{f(z|\Theta)}{f(z|\hat{\Theta}(y))}dz\right] = \int f(z|\Theta)\log f(z|\Theta)dz - E^{y}\left[\int f(z|\Theta)\log f(z|\hat{\Theta}(y))dz\right] \tag{2.3}$$

の右辺第2項の漸近的な不偏推定量を計算することによってＡＩＣが求められたのだが，Sugiura（1978）は誤差項が正規分布に従う回帰分析・分散分析モデルに対して漸近的でない精確（exact）な不偏推定量を導出した．この有限修正を行った情報量規準がｃ-ＡＩＣ（correction of AIC）と呼ばれているものであり，次のような単純な形で表される．

$$\begin{aligned} \text{c-AIC}(M) &= -2^{*}l(\hat{\Theta}|Y) + 2n\ (p+1)/(n-p-2) \\ &= \text{AIC}(M) + 2(p+1)(p+2)/(n-p-2) \end{aligned} \tag{2.4}$$

そして，この導出過程により，小標本の場合にＡＩＣが持つパラメーターを過大推定してしまうような偏りを修正していることが分かる．なお，ｃ-ＡＩＣは正規誤差を持つような一般化線形モデル（線形重回帰，分散分析，自己回帰モデル等を含む）の場合にのみ適用可能であり，Poisson分布，２項分布，多項分布などには適用出来ない．

２－２．ｐ／ｎが大きい場合

この場合にはＡＩＣの一致性が成り立たないことから偏りが生じてしまうが，前述のｃ-ＡＩＣを用いることによって修正可能である．実際，M*：真のモデル，Mj*：候補となるモデルの中でＡＩＣの値が最小となるモデルとしたときに，下の(2.5)式によって定義されるＡＩＣの一致性

$$\Pr[\text{AIC}(Mj^{*})=\text{AIC}(M^{*})]\to 1 \quad (\text{as } n\to\infty) \tag{2.5}$$

は$n\to\infty$の場合などには成立し得ないため，p/nが大きいときには偏りを持ってしまう．しかし，ｃ-ＡＩＣにおいては導出過程で漸近的な近似によらない精確（exact）な統計量を用いているため，このような場合での一致性については考慮する必要がない（Sugiura, 1978）．

２－３．真のモデルが候補のモデルを含まない場合

ＡＩＣでは，真のモデルが候補となるモデルを含んでいない場合にも偏りが生じることが知られている．すなわち，モデルがネスト構造になっている必要があり，例えば下のような回帰モデル

$$\begin{aligned} &(1).\ Z = a + bX + e, \\ &(2).\ Z = a + bX + cY + e, \end{aligned} \tag{2.6}$$

（但し，X, Y, Z：確率変数，$e\sim N(0,\sigma^2)$とする）

を考えたとき，(1)⊂(2)となっている．そのため，(1)が真のモデルで(2)が候補となるモデルであれば問題は生じないが，逆の場合，すなわち(2)が真のモデルで(1)が候補となるモデルである場合には，包含関係に関する条件が満たされていないことになる．しかし，このように候補となるモデルが真のモデルを含んでいない場合にも，竹内（1976）が提案した情報量規準ＴＩＣを用いることによって偏りを修正出来る．ＴＩＣはＡＩＣのペナルティ項を精密に評価したものであり，下の形になる．

$$\text{TIC}(M) = -2^{*}l(\hat{\Theta}|Y) + 2\hat{t} \tag{2.7}$$

但し，$\hat{t}$ は trace$\{J(\Theta)^{-1}I(\Theta)\}$ の部分をその推定量で置き換えたものであり，

$$I(\Theta) = E_{g}\left[\frac{\partial}{\partial\Theta}l(\Theta|Y)\cdot\frac{\partial}{\partial\Theta'}l(\Theta|Y)\right] = \int\left\{\frac{\partial}{\partial\Theta}l(\Theta|y)\cdot\frac{\partial}{\partial\Theta'}l(\Theta|y)\right\}g(y)dy,$$

$$J(\Theta) = -E_{g}\left[\frac{\partial}{\partial\Theta\partial\Theta'}l(\Theta|Y)\right] = -\int\left\{\frac{\partial}{\partial\Theta\partial\Theta'}l(\Theta|y)\right\}g(y)dy$$

である．また，g(y)は標本Yの真の確率密度関数を表す．ＴＩＣのペナルティ項$\hat{t}$は複雑な形をしており分布(誤差構造)によって変化するため，実際の計算は厄介なことも多いが，基本的には trace$\{J(\Theta)^{-1}I(\Theta)\}$ をΘの最尤推定量$\hat{\Theta}$と標本モーメントを用いて具体的に表現することが可能である．

なお，（候補となるモデルが真のモデルを含むという）モデルに関する包含条件を満たしている場合の情報量規準ＴＩＣは，理論的にはＡＩＣと完全に一致することが

知られている．

3．情報量規準ＢＩＣ

本論文では，上で紹介した情報量規準

$$
\begin{aligned}
\mathrm{AIC(M)} &= -2 * l(\hat{\Theta}|Y) + 2\,(p+1)\,, \\
\mathrm{c\text{-}AIC(M)} &= -2 * l(\hat{\Theta}|Y) + 2\,n\,(p+1) / (\,n-p-2\,)\,, \\
\mathrm{TIC(M)} &= -2 * l(\hat{\Theta}|Y) + 2\hat{t}
\end{aligned} \tag{3.1}
$$

の他に，下の規準

$$\mathrm{BIC(M)} = -2 * l(\hat{\Theta}\,|\,Y) + (p+1) * \log(n)$$

を用いてモデル選択を行った．

ＢＩＣはSchwarz(1978)が考案したBayes流の事後確率最大化の考えに基づく情報量規準であり，下の形で表される．

$$\mathrm{BIC(M)} = -2 * l(\hat{\Theta}\,|\,Y) + p * \mathrm{Log}(n) \tag{3.2}$$

この規準は式の形がＡＩＣと良く似ているが，導出に関してはBayes流の考え方を用いている．そのため，平均二乗予測誤差最小の考えに基づくＡＩＣとは全く考え方が異なる．実際，候補となるｋ個のモデルMj (j=1,---, k)に対して

$p(\Theta_j, M_j)$: モデルMjに対する母数Θ_jの事前確率密度

$L(\Theta_j, M_j\,|\,Y)$:モデルMjに対する母数Θ_jの尤度関数

とおいたとき，標本Yに対してモデルMjが正しい事後確率密度

$$\frac{\int L(\Theta_j, M_j\,|\,Y)\, p(\Theta_j, M_j) d\Theta}{\sum_{j=1}^{k} \int L(\Theta_j, M_j\,|\,Y)\, p(\Theta_j, M_j) d\Theta} \tag{3.3}$$

を最大にするモデルを選ぼうとする規準である．

4．ＣＰＵＥ標準化におけるモデル選択

4－1．ＣＰＵＥ標準化とは？

資源量指数としてＣＰＵＥが良く用いられているが，季節変化や海区分け，船の性能など資源密度以外の要因が影響していることも多い．そのため，正確な指数を求めるためにはこれらの要因の影響を取り除いて資源の年変動の効果のみを取り出す必要がある．

生のＣＰＵＥから年変動の効果を取り出す作業をＣＰＵＥ標準化といい，その手法としてＧＬＭ（generalized linear model，一般化線形モデル）が使われることが多い．

例えば，下のようなモデル

$$\mathrm{Log}(\mathrm{CPUE}_{ijk}) = I + Y_i + M_j + A_k + (MA)_{jk} + \varepsilon_{ijk} \tag{4.1}$$

I：切片（平均）

Y：年（Year）効果

M：月（Month）効果

A：エリア（Area）効果

(MA)：月（Month）とエリア（Area）の交互作用効果

ε：誤差（Error）効果

を考えることも多く，誤差項が正規分布に従うと仮定した場合には多元配置（上のケースでは3元配置）分散分析と同一である．実際には，仮定した主効果及び交互作用効果をモデルに含めるか否かをＡＩＣなどの情報量規準，もしくはカイ二乗検定やＦ検定に代表されるようなステップワイズ検定を用いて判断するのが一般的である（庄野，2000）．

本稿では，3節で述べた5つの情報量規準について，正しいモデルが候補となるモデルの中に存在するという仮定の下で真のモデルを選択する確率（i.e., セレクションパフォーマンス）を仮想例や計算機シミュレーションを用いてチェックし， 主にモデルに関する包含関係が満たされない場合にどのような情報量規準用いれば良いのかを検討した．

なお，ＣＰＵＥ標準化の考え方・手法については平松(1995）に詳しく述べられている．また，ＧＬＭの標準的な教科書としては，McCullagh and Nelder (1989), Dobson (1990）などが挙げられる．

4－2．仮想例によるモデル選択

Hilborn and Walters (1992）による仮想例（Table1）を取り上げて，下の１－４のモデルにおける種々の情報量の値を計算した．但し，Error $\sim N(0, \sigma^2)$ とする．

1. log(CPUE) = Intercept + Error
2. log(CPUE) = Intercept + Class + Error

Table 1. Virtual data for CPUE standardization (Hilborn and Walters, 1992). The value show catch rate (tons per hour) for three classes of vessel in four different years.

Year	Class-1	Class-2	Class-3
1	0.63	0.85	1.28
2	0.46	0.65	1.09
3	0.35	0.66	1.01
4	0.43	0.48	0.84

Table 2. Results of model selection by four information criteria (AIC, BIC, TIC, and c-AIC) using the data for CPUE standardization from Table 1.

Model	n	p	RSS	u(4)	n*u(4)/(RSS)^2	AIC	c-AIC	BIC	TIC
1	12	1	1.80717	0.51452	1.89054	15.3368	16.6701	13.8217	14.2273
2	12	3	0.43375	0.03314	2.11385	2.2122	**7.9265**	1.6669	1.3261
3	12	4	1.45701	0.29818	1.68553	18.7522	28.7522	18.6919	17.4378
4	12	6	0.08360	0.00154	2.63773	**-11.5449**	16.4551	**-10.6355**	**-11.9076**

Remarks u(4) and RSS show sample fourth moment about the mean, and resudual sum of squares, respectively. p(number of unknow parameters) does not include the nuisance parameter, σ^2.

Table 3. Dataset on the model of CPUE standardization for computer simulation, which corresponds to the analysis of variance model two way layout and eight replicates.

Year	Area1	Area2	Area3
Dataset-1			
1	1	0.8	0.9
2	0.9	0.72	1.08
3	1.1	0.88	1.32
4	0.8	0.64	0.96
Dataset-2			
1	1	0.9	1.1
2	0.95	0.855	1.045
3	1.1	0.99	1.21
4	0.9	0.81	0.99

3. log(CPUE) = Intercept + Year + Error
4. log(CPUE) = Intercept + Year + Class + Error (4.2)

結果はTable 2のようになるが，有限修正を行った情報量規準c-ＡＩＣについてはモデル２を，そうでないものに対しては一番複雑なモデル４を選ぶ傾向が認められた．このことから，Sugiura (1978) による有限修正を行った場合には，そうでない場合に比べてパラメーター数が少ないモデルが選択されると考えられる．

４－３．シミュレーションによるモデル選択と理論面からの考察

４－２のような仮想例による計算を行うことにより，モデル選択における情報量規準の持つ性質は調べられるが，どの選択が正しいかについての情報は与えられない．

そこで，これらの情報量規準を用いて真のモデルを選択する確率（セレクションパフォーマンス）を調べるため，計算機によるシミュレーション実験を行った． ただ，c-ＡＩＣのセレクションパフォーマンスについてはShono (2000) で既に調べているため，今回は主にＴＩＣによって偏りが補正されるような場合（２－３節で説明），すなわちモデルの包含構造に関する条件が成り立っていない場合について取り上げる． ４－２の仮想例と同様にＣＰＵＥ標準化を想定し，前回 (Shono, 2000) の計算機実験を参考にして，繰り返し数6の2元配置分散分析モデルを仮定した（Table 3)．下のⅢのモデルを真と考え，4つの情報量規準によってモデル選択を行ったときに次のⅠ－Ⅳのどのモデルが選択されるかについての計算機シミュレーションを1000回行った．

Ⅰ. log(CPUE) = Intercept + Error
Ⅱ. log(CPUE) = Intercept + Year + Error
Ⅲ. log(CPUE) = Intercept + Year + Area + Error (4.3)
Ⅳ. log(CPUE) = Intercept + Year + Area + Year*Area + Error

Case-1: Dataset-1を使用し，正規誤差の分散を0.5（i.e. Error～ N(0, 0.5)）と設定．
Case-2: Dataset-2を使用し，正規誤差の分散を0.3（i.e. Error～ N(0, 0.3)）と設定．

ここでは分散分析の特性を生かして，最適化計算ではなく精確（exact）なパラメーター推定を行い，計算には数式処理ソフトのMathematica (Ver.4.0) を使用した．

2種類のシミュレーション結果はTable 4のようになるが，今回の場合にはＢＩＣ以外の３つの情報量規準（ＡＩＣ，c-ＡＩＣ，ＴＩＣ）にはほとんど違いが見られないことから，偏り修正の効果が顕著に表れているとは言えない．

実際，c-ＡＩＣのセレクションパフォーマンスがＡＩＣと比較してさほど良くなかった原因としては，標本数（=96）がある程度大きかったことが挙げられる．しかし，実験の主目的であるＴＩＣのセレクションパフォーマンスが（多少は良くなっているとは言え）ＡＩＣのそれとほとんど変わらなかったことから，モデルの包含構造に関する仮定が満たされていない場合にもＡＩＣの代わりにＴＩＣを使用すべきであるとは断言出来ない．なお，分散分析モデルにおけるＴＩＣの計算に関しては4次の平均まわりのモーメント推定量が必要になってくることから（付録参照），統計パッケージのＳＡＳなどを利用した実際の計算においてはマクロによってこの統計量を求めない限りＴＩＣの計算は困難であり，重回帰分析の場合にも同様のことが言える．

Table 4. Results of computer experiments (frequency of selecting the true model of an ANOVA in 1000 replications of the simulation).

Model	AIC	BIC	c-AIC	TIC
Case-1				
I	79	643	125	78
II	51	50	62	51
III (true)	**776**	**307**	**778**	**781**
IV	94	0	35	90
Case-2				
I	114	724	159	111
II	110	80	123	108
III (true)	**691**	**196**	**693**	**695**
IV	85	0	25	86

また，付録におけるＴＩＣの導出過程に関する検討結果から，分散分析モデルや重回帰分析モデルなど正規誤差を持つ場合には，ＴＩＣとＡＩＣの値が理論的には完全に一致することが示された．そのため，理論と実用の両方の観点から考え，分散分析や重回帰分析を含む正規誤差モデルにおいては，（真のモデルが候補となるモデルを含むという）モデルの包含構造に関する仮定を満たさない場合でもＴＩＣの代わりにＡＩＣを用いることが妥当であると考えられる．

謝　辞

原稿を精査し多くの有益なコメントを下さった遠洋水産研究所の平松一彦博士と所内の査読者の方々に厚く御礼申し上げる．

この報告は，1999年4月に行われた平成11年度日本水産学会春季大会（東京水産大学）で講演した内容に対して，大幅に加筆・修正を行ったものである．

文　献

Akaike, H. 1973: Information theory and an extention of the maximum likelihood principle, *2nd International Symposium on Information Theory*. Petrov, B. N., and Csaki, F. (eds.), Akadimiai Kiado, Budapest: 267-281.

Dobson, A. J. 1990: *An introduction to generalized linear models*. Chapman and Hall. 174 pp.（田中　豊・森川敏彦・栗原孝次共訳．統計モデル入門．共立出版，176 pp.）

Hilborn, R., and C. L. Walters. 1992: *Quantitative fisheries stock assessment*. Chapman and Hall. 570 pp.

平松一彦．1995: 統計モデルによるCPUE標準化．漁業資源研究会議北日本底魚部会報，28: 87-97.

McCullagh, P. and J. A. Nelder. 1989: *Generalized linear models*. 2nd ed. Chapman and Hall. 511 pp.

坂元慶行．1985: カテゴリカルデータのモデル分析．共立出版，221 pp.

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，236 pp.

Schwarz, G. 1978: Estimating the dimension of a model. *Ann. Statist.*, **6**: 461-464.

Shono, H. 2000: Efficiency of finite correction of AIC. *Fisheries Science*, **66**: 608-610.

庄野　宏．2000: 情報量規準とステップワイズ検定の比較と水産資源解析への応用．遠洋水産研究所研究報告，**37**: 1-8.

Sugiura, N. 1978: Further analysis of the data by Akaike's information criterion and the finite corrections. *Commun. Statist.-Theor. Meth.*, **7**(1): 13-26.

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，**153**: 12-18.

付録　分散分析モデルにおけるＴＩＣの導出ならびにＡＩＣとの比較

ＡＩＣの有限修正の場合と比較するとＴＩＣのペナルティ項は複雑な形をしており，理論上は計算可能であるが実際には厄介なことが多い．そこで，付録では論文中の仮想例やシミュレーション実験で仮定したような，正規誤差を持つ2元配置分散分析モデル（分散分析モデルはＣＰＵＥ標準化に良く用いられる）におけるＴＩＣの導出及び理論的な側面からＡＩＣとの比較を行った．

まず，Table 5のような2元配置分散分析モデルを考える．標本数n，パラメーター数p (n>p)，及びn/p の全てが整数であるとし，下のような要因を含むモデルを仮定する．

$$\text{Log(CPUE)} = \text{Intercept} + \text{Year} + \text{Error}, \ \text{Error} \sim N(0, \sigma^2) \tag{A.1}$$

実際には，Log(CPUE) = Intercept + Year + Area + Error の形のモデルでAreaの主効果が認められないものと考えることにする（Table 5参照）．

このとき，未知パラメーターベクトル $\Theta := (\mu, \sigma^2), \mu = (\mu_1, \cdots, \mu_p)$ に対する対数尤度関数 l はその確率密度関数 f を書き下すことによって次のように表される．

$$
\begin{aligned}
\mathrm{l}(\Theta \mid \mathrm{Y}) &= \sum_{i=1}^{n} \log \mathrm{f}(\mathrm{y}_i \mid \mu_i, \sigma^2) \\
&= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 \\
&\quad - \frac{1}{2\sigma^2}\{(\mathrm{y}_1-\mu_1)^2+\cdots+(\mathrm{y}_{\frac{n}{p}}-\mu_1)^2 \\
&\qquad + (\mathrm{y}_{\frac{n}{p}+1}-\mu_2)^2+\cdots+(\mathrm{y}_{2\frac{n}{p}}-\mu_2)^2 \\
&\qquad + \cdots \\
&\qquad + (\mathrm{y}_{(p-1)\frac{n}{p}+1}-\mu_p)^2+\cdots+(\mathrm{y}_n-\mu_p)^2\}
\end{aligned} \tag{A.2}
$$

そして，以後は計算の簡略化のために下記の{　}の中身を☆とおく．すなわち，

$$
\begin{aligned}
☆ &= (\mathrm{y}_1-\mu_1)^2+\cdots+(\mathrm{y}_{\frac{n}{p}}-\mu_1)^2 \\
&\quad + (\mathrm{y}_{\frac{n}{p}+1}-\mu_2)^2+\cdots+(\mathrm{y}_{2\frac{n}{p}}-\mu_2)^2 \\
&\quad + \cdots \\
&\quad + (\mathrm{y}_{(p-1)\frac{n}{p}+1}-\mu_p)^2+\cdots+(\mathrm{y}_n-\mu_p)^2
\end{aligned} \tag{A.3}
$$

である．

　このとき，j=1,…,p に対して

$$
\begin{aligned}
&\frac{\partial \mathrm{l}}{\partial \mu_j} = \frac{1}{\sigma^2}\sum_{k=(j-1)n/p+1}^{j*n/p}(\mathrm{y}_k-\mu_j), \\
&\frac{\partial^2 \mathrm{l}}{\partial \mu_j^2} = -\frac{1}{\sigma^2}\frac{n}{p}, \\
&\frac{\partial^2 \mathrm{l}}{\partial \mu_i \partial \mu_j} = 0 \quad (i \neq j), \\
&\frac{\partial^2 \mathrm{l}}{\partial \mu_j \partial \sigma^2} = -\frac{1}{\sigma^4}\sum_{k=(j-1)n/p+1}^{j*n/p}(\mathrm{y}_k-\mu_j),
\end{aligned} \tag{A.4}
$$

となることと

$$
\begin{aligned}
&\frac{\partial \mathrm{l}}{\partial \sigma^2} = \frac{1}{2\sigma^4}\{☆\} - \frac{n}{2\sigma^2}, \\
&\frac{\partial^2 \mathrm{l}}{\partial (\sigma^2)^2} = -\frac{1}{\sigma^6}\{☆\} + \frac{n}{2\sigma^4}
\end{aligned} \tag{A.5}
$$

より，また下の仮定

$$
\mathrm{E}\left[(\mathrm{y}_k-\mu_j)\right] = 0, \quad \mathrm{E}\left[(\mathrm{y}_k-\mu_j)^2\right] = \sigma^2 \quad (k=(j-1)*n/p+1,\cdots,j*n/p) \tag{A.6}
$$

を用いると

$$
\begin{aligned}
&\mathrm{E}\left[\frac{\partial^2 \mathrm{l}}{\partial \mu_j^2}\right] = -\frac{1}{\sigma^2}\frac{n}{p}, \\
&\mathrm{E}\left[\frac{\partial^2 \mathrm{l}}{\partial \mu_i \partial \mu_j}\right] = 0 \quad (i \neq j), \\
&\mathrm{E}\left[\frac{\partial^2 \mathrm{l}}{\partial (\sigma^2)^2}\right] = -\frac{1}{\sigma^6}\mathrm{E}\left[\{☆\}\right] + \frac{n}{2\sigma^4} = -\frac{n}{2\sigma^4}
\end{aligned} \tag{A.7}
$$

となる．よって，

$$
\mathrm{J}(\Theta) = \mathrm{J}(\mu, \sigma^2) = -\mathrm{E}\left[\frac{\partial}{\partial\Theta\partial\Theta'}\mathrm{l}(\Theta \mid \mathrm{Y})\right]
= \begin{bmatrix}
\frac{1}{\sigma^2}\frac{n}{p} & & & \\
 & \ddots & & \\
 & & \frac{1}{\sigma^2}\frac{n}{p} & \\
 & & & \frac{n}{2\sigma^4}
\end{bmatrix} \tag{A.8}
$$

$$
\mathrm{J}(\Theta)^{-1} = \mathrm{J}(\mu, \sigma^2)^{-1} = \begin{bmatrix}
\sigma^2\frac{p}{n} & & & \\
 & \ddots & & \\
 & & \sigma^2\frac{p}{n} & \\
 & & & \frac{2\sigma^4}{n}
\end{bmatrix} \tag{A.9}
$$

と表される．

　また，

$$
\mathrm{E}\left[\frac{\partial \mathrm{l}}{\partial \mu_j}\cdot\frac{\partial \mathrm{l}}{\partial \mu_j}\right] = \frac{1}{\sigma^2}\frac{n}{p} \quad (\because \text{標本の独立性の仮定より}), \tag{A.10}
$$

$$
\begin{aligned}
\mathrm{E}\left[\frac{\partial \mathrm{l}}{\partial \sigma^2}\cdot\frac{\partial \mathrm{l}}{\partial \sigma^2}\right] &= \mathrm{E}\left[\left(\frac{1}{2\sigma^4}\{☆\} - \frac{n}{2\sigma^2}\right)\left(\frac{1}{2\sigma^4}\{☆\} - \frac{n}{2\sigma^2}\right)\right] \\
&= \frac{1}{4\sigma^8}\mathrm{E}\left[\{☆\}^2\right] - \frac{n}{2\sigma^6}\mathrm{E}\left[\{☆\}\right] + \frac{n^2}{4\sigma^4} \\
&= \frac{1}{4\sigma^8}\mathrm{E}\left[\{☆\}^2\right] - \frac{n^2}{4\sigma^4} = \frac{1}{4\sigma^8}\left(\frac{\kappa}{\sigma^4} - n^2\right)
\end{aligned} \tag{A.11}
$$

とおくと（上のように κ を定める），

$$
\{☆\}^2 = \sum_{j=1}^{p}\sum_{k=(j-1)n/p+1}^{j*n/p}(\mathrm{y}_k-\mu_j)^4 + \sum_{i,j,k\neq l}(\mathrm{y}_k-\mu_i)^2(\mathrm{y}_k-\mu_j)^2 \tag{A.12}
$$

$$
\mathrm{E}\left[\sum_{j=1}^{p}\sum_{k=(j-1)n/p+1}^{j*n/p}(\mathrm{y}_k-\mu_j)^4\right] = \mathrm{E}\left[\sum_{i=1}^{n}(\mathrm{y}_i-\mu_i)^4\right] = n\mathrm{E}\left[(\mathrm{y}_i-\mu_i)^4\right]
$$

$$
(\because \text{添字の付替え}) \tag{A.13}
$$

Table 5. ANOVA model dealing with CPUE Standardization for TIC derivation.

Area	1	2	•	•	•	n/p	Parameter
Year-1	Y(1)	Y(2)	•	•	•	Y(n/p)	u(1)
Year-2	Y(n/p+1)	Y(n/p+2)	•	•	•	Y(2n/p)	u(2)
•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•
Year-p	Y((p-1)n/p+1)	Y((p-1)n/p+2)	•	•	•	Y(n)	u(p)

Remarks　Y(•) and u(•) show the corresponding samples and unknown parameters, respectively.

$$E\left[\sum_{i,j,k\neq 1}(y_k-\mu_i)^2(y_k-\mu_j)^2\right]=n(n-1)\sigma^4 \tag{A.14}$$

となることから

$$\begin{aligned}\frac{\kappa}{\sigma^4}-n^2&=\frac{n}{(\sigma^2)^2}E\left[(y_i-\mu_i)^4\right]+\frac{n(n-1)\sigma^4}{\sigma^4}-n^2\\&=\frac{n}{(\sigma^2)^2}E\left[(y_i-\mu_i)^4\right]-n\\&=n\left\{\frac{1}{(\sigma^2)^2}E\left[(y_i-\mu_i)^4\right]-1\right\}=(*)\end{aligned} \tag{A.15}$$

と変形出来る．

ここで，μ_i,σ^2をそれぞれ最尤推定量$\hat{\mu}_i$，$\hat{\sigma}^2=\mathrm{RSS}/n$で置き換える．また，平均まわりの4次の積率を，標本における平均まわりの4次の積率$\hat{\mu}(4)$を用いて表現すると，

$$\begin{aligned}(*)&=n\left\{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{\mu}_i)^4\Big/(\hat{\sigma}^2)^2-1\right\}\\&=n\left[\frac{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{\mu}_i)^4}{\left\{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{\mu}_i)^2\right\}^2}-1\right]=n\left\{n^2\cdot\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{\mu}_i)^4\Big/(\mathrm{RSS})^2-1\right\}\\&=n\left\{\frac{n^2\cdot\hat{\mu}(4)}{(\mathrm{RSS})^2}-1\right\}\end{aligned} \tag{A.16}$$

となる．但し，

$$\begin{aligned}&\mathrm{RSS}:=\sum_{i=1}^{n}(y_i-\hat{\mu}_i)^2 \qquad \text{(residual sum of squares)},\\&\hat{\mu}(4):=\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{\mu}_i)^4\end{aligned} \tag{A.17}$$

とおく．

以上により

$$\begin{aligned}I(\hat{\Theta})=I(\hat{\mu},\hat{\sigma}^2)&=E_g\left[\frac{\partial}{\partial\Theta}l(\hat{\Theta}|Y)\cdot\frac{\partial}{\partial\Theta'}l(\hat{\Theta}|Y)\right]\\&=\begin{bmatrix}\frac{1}{\hat{\sigma}^2}\frac{n}{p} & & & *\\ & \ddots & & \\ & & \frac{1}{\hat{\sigma}^2}\frac{n}{p} & \\ * & & & \frac{n}{4\hat{\sigma}^4}\left\{\frac{n^2\cdot\hat{\mu}(4)}{(\mathrm{RSS})^2}-1\right\}\end{bmatrix}\end{aligned} \tag{A.18}$$

$$J(\hat{\Theta})^{-1}I(\hat{\Theta})=\begin{bmatrix}1 & & & *\\ & \ddots & & \\ & & 1 & \\ * & & & \frac{1}{2}\left\{\frac{n^2\cdot\hat{\mu}(4)}{(\mathrm{RSS})^2}-1\right\}\end{bmatrix} \tag{A.19}$$

となり（上の行列の対角成分以外（*印の部分）については最終的な結果に影響しないなめ，計算を省略する），

$$\begin{aligned}\mathrm{trace}\left\{J(\hat{\Theta})^{-1}I(\hat{\Theta})\right\}&=p+\frac{1}{2}\left\{\frac{n^2\cdot\hat{\mu}(4)}{(\mathrm{RSS})^2}-1\right\}\\&=\frac{1}{2}\left\{2p-1+\frac{n^2\cdot\hat{\mu}(4)}{(\mathrm{RSS})^2}\right\}\end{aligned} \tag{A.20}$$

と表せることから，この場合のＴＩＣは

$$\begin{aligned}\mathrm{TIC}(M)&=-2*l(\hat{\Theta}|Y)+2\hat{t}\\&=-2*l(\hat{\Theta}|Y)+2p-1+\frac{n^2\cdot\hat{\mu}(4)}{(\mathrm{RSS})^2}\end{aligned} \tag{A.21}$$

となり，ペナルティ項の評価が出来たことになる．

なお，上式は(A.1)のような分散分析型のモデルに適用可能なことに注意する必要がある．

また，これとＡＩＣの形

$$\mathrm{AIC}(M)=-2*l(\hat{\Theta}|Y)+2(p+1) \tag{A.22}$$

を比較することにより，

$$\begin{aligned}&\mathrm{TIC}(M)>\mathrm{AIC}(M)\Rightarrow\frac{n^2\cdot\hat{\mu}(4)}{(\mathrm{RSS})^2}>3,\\&\mathrm{TIC}(M)=\mathrm{AIC}(M)\Rightarrow\frac{n^2\cdot\hat{\mu}(4)}{(\mathrm{RSS})^2}=3,\\&\mathrm{TIC}(M)<\mathrm{AIC}(M)\Rightarrow\frac{n^2\cdot\hat{\mu}(4)}{(\mathrm{RSS})^2}<3\end{aligned} \tag{A.23}$$

という関係式が成立していることが分かる．

この分散分析モデルでは，両者（ＡＩＣとＴＩＣ）のペナルティ項の値を比較することにより，精密評価の度合いを実際の標本を用いて数値的に検討することが可能である．

次に，理論的な側面からＴＩＣとＡＩＣを比較することを考える．

すなわち，TICのペナルティ項の（真の分布gに対する）期待値を考えると，

$$\begin{aligned}\mathrm{trace}\left\{J(\Theta)^{-1}I(\Theta)\right\}&=p+\frac{1}{2}\left\{\frac{1}{(\sigma^2)^2}E\left[(y_i-\mu_i)^4\right]-1\right\}\\&=\frac{1}{2}\left\{2p-1+\frac{1}{(\sigma^2)^2}E\left[(y_i-\mu_i)^4\right]\right\}\end{aligned} \tag{A.24}$$

より

$$-2*l(\hat{\Theta}|Y)+2*E_g[\hat{t}]=-2*l(\hat{\Theta}|Y)+2p-1+\frac{E\left[(y_i-\mu_i)^4\right]}{(\sigma^2)^2} \tag{A.25}$$

と表現出来る．これとＡＩＣのペナルティ項の期待値を取ったもの

$$-2*l(\hat{\Theta}|Y)+2(p+1)\ (=\mathrm{AIC}(M)) \tag{A.26}$$

との差を考えると，

$$E_g[\mathrm{TIC}(M)-\mathrm{AIC}(M)]=\frac{E\left[(y_i-\mu_i)^4\right]}{(\sigma^2)^2}-3=\beta-3 \tag{A.27}$$

となる．このとき

$$\beta=\frac{E\left[(y_i-\mu_i)^4\right]}{(\sigma^2)^2} \tag{A.28}$$

は確率分布の中央での尖り具合や裾の重さ・長さを測る尖度（kurtosis）という指標になっている．よく知られているように，正規分布の尖度は3であることから，今回の分散分析モデルにおけるＴＩＣとＡＩＣの差の理論値は0になることが証明される．よって，この場合にはＴＩＣとＡＩＣは実用上もほとんど差がないと考えられ，本論文の計算機実験においてもそのような結果が示された．

また，p個の回帰係数を持つ下のような重回帰分析モ

デル

$$Y = X\Theta + \varepsilon,\quad \varepsilon \sim N(0, \sigma^2 I_n) \tag{A.29}$$
$$(\text{但し}, \Theta = (\theta_1, \cdots, \theta_p)\text{とする})$$

におけるＴＩＣも今回の分散分析の場合とほぼ同じ手順にて計算することが可能であり，

$$TIC(M) = -2 * l(\hat{\Theta}|Y) + 2p - 1 + \frac{n^2 \cdot \hat{\mu}(4)}{(RSS)^2} \tag{A.30}$$

と表される．従って，上と同様の尖度βを用いた理論的な検討により，

$$E_g[TIC(M) - AIC(M)] = \beta - 3 = 0 \tag{A.31}$$

となり，ＡＩＣとＴＩＣの理論値に完全に一致することが示された．

今回の付録で取り上げた統計モデルは分散分析と重回帰分析のみであるが，ＴＩＣのペナルティ項は上の意味での統計モデルには依存しないため，この論理を正規分布に従う他の統計モデル（e.g. 時系列モデルetc.）に拡張することは比較的容易である．ゆえに，正規誤差モデルにおけるＴＩＣの理論値はＡＩＣのそれと同じであり，標本に基づく実測値についてもほとんど差がないと考えられる．なお，Poisson分布，二項分布など他の確率分布に従う場合には

$$E_g[TIC(M) - AIC(M)] = \frac{E\left[(y_i - \mu_i)^4\right]}{(\sigma^2)^2} - 3 = \beta - 3 \tag{A.32}$$

となる保証がないため（実際にはＴＩＣの具体的な形を求めること自体困難な場合が多いと思われる），ここでの推論は適用出来ないことに注意する必要がある．