日本統計学会誌
第27巻，第3号，1997年
205頁〜220頁

会長就任講演

統計データと数理統計学のはざまに35年

杉　浦　成　昭*

あるテレビ対談の席上ですが，有名なゴルフの尾崎将司選手が“自分は猪年生まれであるが，これまで過去を振り返ることなく前のみを目指して生きてきた．恐れるものは何もなかったが，ただ恐いのは前方の落し穴に気付かず落ちてしまい再び這上がれないことだけだ．”と云っています．私は一廻り上の猪に大変近いのですが彼の言葉に共感したのを覚えています．今日はその猪もどきが異例にも過去を振り返る話をしたいと思います．講演報告集には35年を振り返り同じ題の下で項目とKey Wordsだけですが，本日の演題として（1）順序統計量の平均値の上界，（2）多変量Wilcoxon検定，（3）正規多変量解析（大標本論），（4）分割表とLogit model，（5）AICの有限修正，（6）生存曲線の解析，（7）正規多変量解析（小標本論），（8）変化点模型，（9）順序の付いた平均に対するBayes検定，（10）統計学と統計学会の将来，を掲げました．（2）は終生nonparametric推測の研究に打ち込まれた故田村亮二先生（熊本大学理学部教授）と多変量Wilcoxon検定に関わる話しで，当時島根大学におられ大阪統計談話会で初めてお会いしたこと，Tamura (1966) や私の学位論文Sugiura (1965) 等の経緯が高田，橋本，坂田，柳川 (1981) の序文に書いてあります．1次元のWilcoxon検定は丘本正先生（現大阪大学名誉教授）の紹介で歯学データに適用する機会に恵まれ（Sugiura 1963, 1964），このお陰で順位検定の必要性を実感し研究したのですが，多次元Wilcoxon検定は実際の場面に適用する機会が遂になくて残念に思っています．（3）は分布の漸近展開や検定の不偏性，局所最強力性に関するもの（Sugiura and Nagao 1968, Sugiura and Fujikoshi 1969, Sugiura 1969, 1972, 1973a, b, 1976a, b, Nagao 1973）．（4），（6）が医学データに関係した話しです（Sugiura and Otake 1973, 1974等）．実際の大規模な臨床実験の一部にも参加したのですが何も統計学上の文献に残るような貢献が出来なくて残念に思っています．（7）は縮小推定量やrisk，検定力等の精密評価，グラフ表示に関するもの（Sugiura 1988, 1989, 1990, 1992, 1995, 1997, Sugiura and Konno 1987, Sugiura and Kubokawa 1988等）．（8）は線形変化点模型の尤度比検定や順位検定に関するもの（Sugiura and Ogden 1994）．（9）は平均に順序があるときの一般化Bayes検定の近似に関するもの（Sugiura 1994）等です．しかしその後昔やった仕事に少し進展がありましたので，今日は予定を変更し（1）と（5）について話し最後に（10）について触れたいと思います．このような機会が与えられたことに感謝します．

1. 順序統計量の平均の上界，下界

一般性を失うこと無く平均0，分散1とし，連続型分布関数$F(x)$をもつ分布から大きさnの無作為標本を大きさの順に並べてr番目に小さい標本を$X_{r|n}$とする．平均，分散が制約されているため順序統計量の期待値$E(X_{r|n})$も当然制約され無限に大きくはなれない．事実平均が0，分散が1である限りどのような連続型分布であろうとも必ず

* 筑波大学　数学系

$$|E(X_{r|n})| \le \sqrt{\frac{B(2r-1, 2n-2r+1)}{B(r, n-r+1)^2} - 1} \quad \text{Hartley and David (1954)} \tag{1.1}$$

が成り立つ．ここで等号は $F(x)$ の逆関数を $x(u)=\inf\{x|F(x)>u\}$ とするとき

$$x(u) = \pm \frac{1}{\sqrt{\frac{B(2r-1, 2n-2r+1)}{B(r, n-r+1)^2} - 1}} \left\{1 - \frac{1}{B(r, n-r+1)} u^{r-1}(1-u)^{n-r}\right\} \tag{1.2}$$

に限るが，$x(u)$ は単調増加でなければならないので右辺がそうなるのは $r=1$ または $r=n$ のときに限ることがわかる．これより $r=n$ のとき上界を達成する分布は有限区間上の Pareto 型の分布で分布関数は

$$F_n(x) = \frac{1}{n^{\frac{1}{n-1}}} \left(1 + \frac{n-1}{\sqrt{2n-1}} x\right)^{\frac{1}{n-1}}, \quad -\frac{\sqrt{2n-1}}{n-1} \le x \le \sqrt{2n-1} \tag{1.3}$$

その確率密度関数は

$$f_n(x) = \frac{1}{n^{\frac{1}{n-1}} \sqrt{2n-1} \left(1 + \frac{n-1}{\sqrt{2n-1}} x\right)^{\frac{n-2}{n-1}}}, \quad -\frac{\sqrt{2n-1}}{n-1} \le x \le \sqrt{2n-1} \tag{1.4}$$

となる．この密度関数は x の左端で ∞ となり，右端では $1/n\sqrt{2n-1}$ となっている．$n=2, 3, 4, 5, 10$ のとき次の図 1.1 のグラフを得る．直観的には右のほうに高い確率を持つ分布かと思われるが事実は逆である．

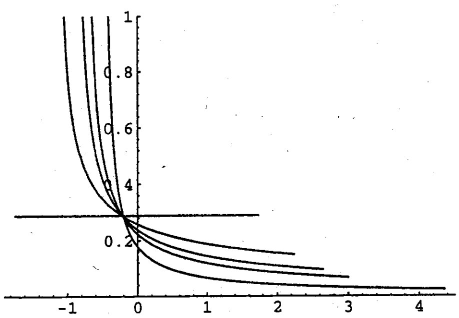

図 1.1　$f_n(x)$ グラフ
(左から $n=2, 3, 4, 5, 10$)
Gumbel (1954), David (1981)

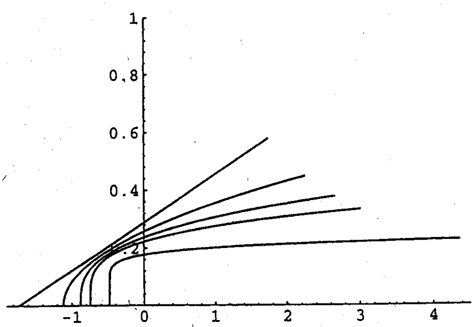

図 1.2　$F_n(x)$ のもとで $X_{n|n}$ の確率密度
(左から $n=2, 3, 4, 5, 10$)

このとき最大値の分布の確率密度関数 $(d/dx)F(x)^n$ は図 1.2 のグラフとなりこれは直観に合っている．

更に母分布は原点に関して対称 $F(x)+F(-x)=1$ であるものに限れば

$$|E(X_{n|n})| \le \sqrt{1 - \frac{1}{\binom{2n-2}{n-1}}} \frac{n}{\sqrt{2(2n-1)}} \tag{1.5}$$

であり等号は

$$x(u)=\sqrt{n-\frac{1}{2}}\,\frac{1}{\sqrt{1-\frac{1}{\binom{2n-2}{n-1}}}}\{u^{n-1}-(1-u)^{n-1}\} \tag{1.6}$$

に限る．これらはやはり有限区間上の分布であり，$n=2$ と $n=3$ は同じ一様分布となるが，それ以外は（1）の様に確率密度関数を陽に求めることは出来ない．数値計算によりそのグラフを求めれば，図1.1，図1.2に対応して次の図1.3，図1.4を得る．

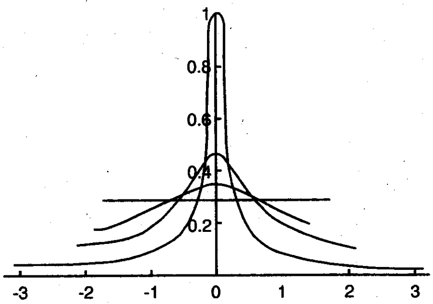

図1.3　等号を与える確率密度関数
(左端上から $n=3, 4, 5, 10$)
Moriguti (1951)，森口 (1995)

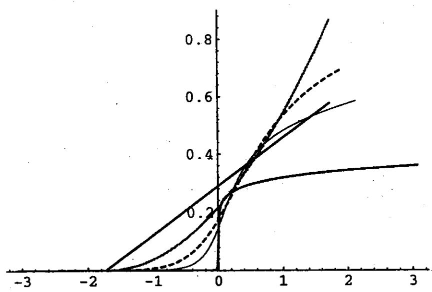

図1.4　$X_{n|n}$ の確率密度関数
(左端上から $n=2, 3, 4, 5, 10$)

$\varphi_v(u)$ を区間 $[0, 1]$ 上の正規化された v 次 Legendre 多項式とすれば，

$$\varphi_v(u)=\frac{\sqrt{2v+1}}{v!}\frac{d^v}{du^v}u^v(u-1)^v \quad (v=0, 1, 2, \cdots) \tag{1.7}$$
$$\varphi_0(u)=1,\ \varphi_1(u)=\sqrt{3}(2u-1),\ \varphi_2(u)=\sqrt{5}(6u^2-6u+1), \cdots$$

であるが，$\{\varphi_{2v}(u)\}$ は $L^2(0, 1)$ の部分空間

$$L_e^2(0, 1)=\left\{f(u)\left|\int_0^1 f(u)^2du<\infty,\ f(u)=f(1-u)\right.\right\}$$

の完備正規直交基底であり $\{\varphi_{2v-1}(u)\}$ は $L^2(0, 1)$ の部分空間

$$L_o^2(0, 1)=\left\{f(u)\left|\int_0^1 f(u)^2du<\infty,\ f(u)=-f(1-u)\right.\right\}$$

の完備正規直交基底である．$x(u), g(u)=u^{r-1}(1-u)^{n-r}/B(r, n-r+1)$ の $\varphi_v(u)$ による展開係数を

$$a_v=\int_0^1 x(u)\varphi_v(u)du, \quad b_v=\int_0^1 g(u)\varphi_v(u)du \tag{1.8}$$

とすれば $a_0=0, b_0=1$ であるが，さらに $x(u)=-x(1-u)$ であるから $a_{2v}=0$ が成り立つ．

定理 1. (Sugiura 1962)　母分布が絶対連続で有限な分散 σ^2 をもち，原点に関して対称であるとき

$$(1.9)\quad \left|E(X_{r|n})-\sum_{v=1}^{k}a_{2v-1}b_{2v-1}\right| \leq\sqrt{\sigma^2-\sum_{v=1}^{k}a_{2v-1}^2}\sqrt{\frac{B(2r-1,2n-2r+1-B(n,n)}{2B(r,n-r+1)^2}-1-\sum_{v=1}^{k}b_{2v-1}^2}$$

Sugiura (1962) ではこの定理を用いて正規分布の場合に順序統計量の平均値の近似とその誤差について計算したが，誤差の評価が最良のものでないこと，特に $r\neq 1, n$ のときに等号を与える分布がどうなっているのか気になったままであった．Chebyshev の不等式の等号を達成する分布が離散型分布であることからこの場合も離散成分をもつ分布になるのではないかと想像される．最近の著書で森口 (1995) は御自身の論文 Moriguti (1951, 1953) の本質的な部分を明快に解説しそのなかに既に私の疑問が解ける基本的な定理が完成されているのを知った．定理があることは David (1981) 等により知っていたが，今回初めて具体的に解決することが出来た．Moriguti (1953) の結果を我々に適した形で述べれば，次の通りである．

定理 2. (Moriguti 1953, 森口 1995) $x(u), \varphi(u)$ は $L^2(0,1)$ に属し $x(u)$ は単調増加とする．$\Phi(u)=\int_0^u\varphi(u)du$ に対して $\Phi(u)$ の最大凸劣関数を $\overline{\Phi}(u)$，最小凹優関数を $\underline{\Phi}(u)$ とするとき

$$(1.10)\quad -\sqrt{\int_0^1 x(u)^2du\int_0^1\left\{\frac{d}{du}\underline{\Phi}(u)\right\}^2du}\leq\int_0^1 x(u)\varphi(u)du\leq\sqrt{\int_0^1 x(u)^2du\int_0^1\left\{\frac{d}{du}\overline{\Phi}(u)\right\}^2du}$$

左側の不等式で等号は $x(u)=-C\cdot(d/du)\underline{\Phi}(u)$ のとき，右側の不等式で等号は $x(u)=C\cdot(d/du)\overline{\Phi}(u)$ に限る．ただし $C>0$ は定数とする．

これは $x(u)$ が単調であるときに Schwarz の不等式を改良したもので等号条件も与えられているからこれ以上は改良できない．Moriguti (1953)，森口 (1995) には右側の不等式のみ述べてあるが左側の不等式も同様に証明できる．これを用いて定理 1 を次のように改良できる．

定理 3. 定理 1 の仮定の下で，$x(u), \{g(u)-g(1-u)\}/2$ を正規直交基底 $\{\varphi_v(u)\}$ で近似したときの誤差を

$$(1.11)\quad x_{2k-1}(u)=x(u)-\sum_{v=1}^{k}a_{2v-1}\varphi_{2v-1}(u),$$
$$g_{2k-1}(u)=\frac{1}{2}\{g(u)-g(1-u)\}-\sum_{v=1}^{k}b_{2v-1}\varphi_{2v-1}(u),$$

としその積分を $G_{2k-1}(u)=\int_0^u g_{2k-1}(v)dv$ とおく．$x_{2k-1}(u)$ が単調増加または単調減少の領域は有限個の互いに交わらない区間 $I_i(i=1,\cdots,p)$ または $D_j(j=1,\cdots,q)$ の和集合で表わされるとする．区間 I_i 上で $G_{2k-1}(u)$ の最大凸劣関数を $\overline{G}_{2k-1,I_i}(u)$，最小凹優関数を $\underline{G}_{2k-1,I_i}(u)$ としその導関数をそれぞれ $\bar{g}_{2k-1,I_i}(u), \underline{g}_{2k-1,I_i}(u)$ とおく．同様に区間 D_j 上で $\overline{G}_{2k-1,D_j}(u), \underline{G}_{2k-1,D_j}(u)$ およびその導関数 $\bar{g}_{2k-1,D_j}(u), \underline{g}_{2k-1,D_j}(u)$ を定義する．このとき次の不等式が成り立つ．

$$(1.12)\quad -\sum_{i=1}^{p}\sqrt{\int_{I_i}x_{2k-1}(u)^2du\int_{I_i}\{\underline{g}_{2k-1,I_i}(u)\}^2du}-\sum_{j=1}^{q}\sqrt{\int_{D_j}x_{2k-1}(u)^2du\int_{D_j}\{\bar{g}_{2k-1,D_j}(u)\}^2du}$$
$$\leq E(X_{r|n})-\sum_{v=1}^{k}a_{2v-1}b_{2v-1}\leq\sum_{i=1}^{p}\sqrt{\int_{I_i}x_{2k-1}(u)^2du\int_{I_i}\{\bar{g}_{2k-1,I_i}(u)\}^2du}$$
$$+\sum_{j=1}^{q}\sqrt{\int_{D_j}x_{2k-1}(u)^2du\int_{D_j}\{\underline{g}_{2k-1,D_j}(u)\}^2du}$$

左辺の不等式で等号は正の定数 C_j に対して区間 I_j 上で $x_{2k-1}(u)=-C_j\underline{g}_{2k-1,I_j}(u)$ が，区間 D_j 上

で $x_{2k-1}(u)=-C_j\bar{g}_{2k-1,D_j}(u)$ が成り立つときに限る．右辺の不等式で等号は区間 I_j 上で $x_{2k-1}(u)=C_j\bar{g}_{2k-1,I_j}(u)$ が，区間 D_j 上で $x_{2k-1}(u)=C_j\underline{g}_{2k-1,D_j}(u)$ が成り立つときに限る．

系 3.1. 定理3の下で

$$(1.13)\qquad -\sqrt{\sigma^2-\sum_{v=1}^{k}a_{2v-1}^2}\left[\sum_{i=1}^{p}\sqrt{\int_{I_i}\{\underline{g}_{2k-1,I_i}(u)\}^2du}+\sum_{j=1}^{q}\sqrt{\int_{D_j}\{\bar{g}_{2k-1,D_j}(u)\}^2du}\right]$$
$$\le E(X_{r|n})-\sum_{v=1}^{k}a_{2v-1}b_{2v-1}\le\sqrt{\sigma^2-\sum_{v=1}^{k}a_{2v-1}^2}\left[\sum_{i=1}^{p}\sqrt{\int_{I_i}\{\bar{g}_{2k-1,I_i}(u)\}^2du}\right.$$
$$\left.+\sum_{j=1}^{q}\sqrt{\int_{D_j}\{\underline{g}_{2k-1,D_j}(u)\}^2du}\right]$$

系3.1は定理3よりも評価としては甘いが係数 a_v と区間 I_i, D_j のみで誤差の計算が出来るという利点がある．

系 3.2. (Moriguti 1953) 定理3.1において $k=0$ のときは $x_{-1}(u)=x(u), g_{-1}(u)=(g(u)-g(1-u))/2$ となり

$$(1.14)\qquad -\sigma\sqrt{\int_0^1\underline{g}_{-1}(u)^2du}\le E(X_{r|n})\le\sigma\sqrt{\int_0^1\bar{g}_{-1}(u)^2du}$$

が成り立つ．左辺の等号は正の定数 C が存在して $x(u)=-C\underline{g}_{-1}(u)$ のとき，また右辺の等号は $x(u)=C\bar{g}_{-1}(u)$ のときに限る．

例 1. $k=0, n=10, r=9$ のとき系3.2より $E(X_{9/10})$ の上限と下限を求めると $g_{-1}(u)=45\{u^8(1-u)-u(1-u)^8\}$ であるから $G_{-1}(u)=\int_0^u g_{-1}(v)dv$ のグラフは次の図1.5で与えられ，$G_{-1}(u)$ の最小凹優関数（明らかに0）と最大凸劣関数を微分して

$$\underline{g}_{-1}(u)=0,\quad 0<u<1$$
$$\bar{g}_{-1}(u)=\begin{cases}-1.56135, & 0<u\le 0.193116\\ -45(1-u)^8u+45(1-u)u^8, & 0.193116<u<0.806884\\ 1.56135, & 0.806884<u<1\end{cases}$$

を得る．点 $u=0.193116$ は図1.5で原点からこの曲線に引いた接線の接点，$\underline{g}_{-1}(u)$ の値 -1.56135 はその勾配である．$\bar{g}_{-1}(u)$ のグラフが図1.6である．

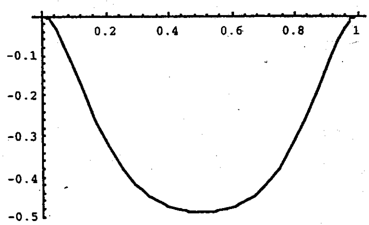

図1.5 $G_{-1}(u)$ のグラフ

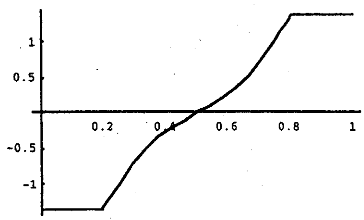

図1.6 $\bar{g}_{-1}(u)$ のグラフ

これより

$$0\le E(X_{9|10})\le\sqrt{\int_0^1\bar{g}_{-1}(u)^2du}=1.14004$$

となり右側の不等式で等号を達成するのは混合型の分布で

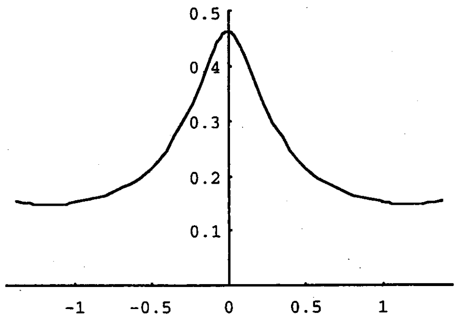

図1.7 連続成分の確率密度関数（$-1.36956<x<1.36956$）

$$x(u)=\frac{1}{1.14004}\bar{g}_{-1}(u)$$

を解けばよい．その結果離散成分は

$$P(X=-1.36956)=P(X=1.36956)=0.193116$$

連続成分の確率密度関数のグラフは上の図1.7で与えられる．

例 2. (Sugiura 1962) $F(x)$ として標準正規分布関数 $\Phi(x)$ をとりその逆関数を $\Phi^{-1}(u)$ とする．定理3において $n=10, r=9, k=1$ のとき $x_1(u)$ のグラフは次の図1,8で与えられる．
これより $x_1(u)$ が単調増加の区間は $I_1=(0, 0.219115), I_2=(0.780885, 1)$，単調減少の区間は $D_1=(0.219115, 0.780885)$ となり $\int_{I_1}x_1(u)^2du=\int_{I_2}x_1(u)^2du=0.018504, \int_{D_1}x_1(u)^2du=0.008063$ を得る．

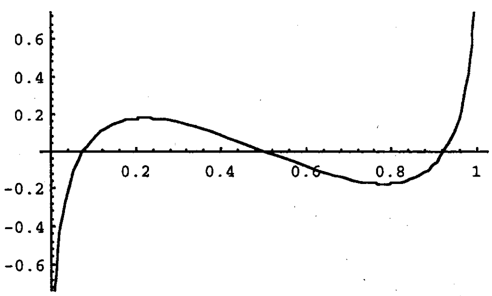

図1.8 $x_1(u)=\Phi^{-1}(u)-\frac{3}{\sqrt{\pi}}(2u-1)$ のグラフ

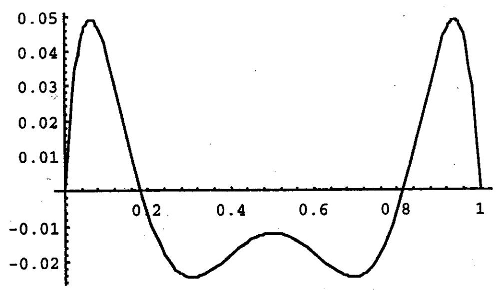

図1.9 $G_1(u)=\int_0^u g_1(u)du$ のグラフ

$G_1(u)$ のグラフは $g_1(u)=45u^8(1-u)-45u(1-u)^8-(21/11)(2u-1)$ を用いて次の図1.9で与えられる．

これより誤差の上限を計算するには

$$\bar{g}_{1,I_1}(u)=-0.0588952, \quad 0<u\leq 0.219115$$

$$\underline{g}_{1,D_1}(u)=\begin{cases} 0.00328858 & 0.219115<u\leq 0.497577 \\ \frac{21}{11}(1-2u)-45(1-u)^8u+45(1-u)u^8 & 0.497577<u\leq 0.5 \\ -\underline{g}_{1,D_1}(1-u) & 0.5<u\leq 0.780885 \end{cases}$$

$$\bar{g}_{1,I_2}(u)=0.0588952, \quad 0.780885<u<1$$

より $\int_{I_1}\bar{g}_{1,I_1}(u)^2du=\int_{I_2}\bar{g}_{1,I_2}(u)^2du=0.000760$, $\int_{D_1}\bar{g}_{1,D_1}(u)^2du=6.04\times10^{-6}$ を用いて次の表 1 の上界 0.0077 が計算される.

下限を計算するには

$$\underline{g}_{1,I_1}(u)=\begin{cases}\frac{21}{11}(1-2u)-45(1-u)^8u+45(1-u)u^8 & 0<u\leq0.106411\\ -0.443916 & 0.106411<u\leq0.219115\end{cases}$$

$$\bar{g}_{1,D_1}(u)=\begin{cases}\frac{21}{11}(1-2u)-45(1-u)^8u+45(1-u)u^8 & 0.219115<u\leq0.305098\\ 0 & 0.305098<u\leq0.5\\ -\bar{g}_{1,D_1}(1-u) & 0.5<u<0.780885\end{cases}$$

$$\underline{g}_{1,I_2}(u)=-\underline{g}_{1,I_2}(1-u),\quad 0.780885<u<1$$

より $\int_{I_1}\underline{g}_{1,I_1}(u)^2du=\int_{I_2}\underline{g}_{1,I_2}(u)^2du=0.0706723$, $\int_{D_1}g_{1,D_1}(u)^2du=0.004318$ を用いて次の表 1.1 の下界 -0.078 を得る.

表 1.1　$E(X_{9|10})-a_1b_1$ の上界と下界

a_1b_1	定理 1 の上下界	定理 3 の上界	定理 3 の下界	系 3.1 の上界	系 3.1 の下界	$E(X_{9\|10})$
1.077	±0.089	0.0077	−0.078	0.0083	−0.0878	1.00136

この例についてみれば計算は面倒であるが定理 3 を用いれば定理 1 より上界で 1/10 以下に, 下界で約 1 割強改良されている. 下界についてはこれが限界であることも正確な値をみればわかる. 系 3.1 による評価も悪くない.

ここでは分布 $F(x)$ が原点について対称である場合に定理 3 を述べたが, 対称でないときも Sugiura (1962) を拡張した形で同様な定理が成り立つ.

順序統計量の正確な積率を計算することは昔は難しい問題であったが最近の計算機を使えば数値計算で簡単に求められるので, 今やこのような近似を問題にする必要性は少ないかも知れない. しかしここで述べた様な限界を明らかにしておくことは意味があるといえよう.

1962 年に定理 1 等を発表したところ, しばらくして R. L. Plackett 先生から全く面識が無いにもかかわらず, お手紙を頂き Mathematical Review のために論文を読んだが, たいへん elegant な結果であるとお褒めの言葉をいただき励みになったことがあります. 私も他人の論文を読み感激することはあるのですが若い方に対してこのような労を取っていなかったではないかと最近反省しています.

2.　情報量基準の有限修正

データ Y の確率密度関数を $f(y|\theta)$ とし母数 θ に関するモデル $M_j(j=1,\cdots,J)$ を考える. M_j のもとでの最尤推定量を $\hat{\theta}_j=\hat{\theta}_j(Y)$($p_j$ 次元) とする. このとき赤池情報量基準は $\mathrm{AIC}=-2\log f(y|\hat{\theta}_j)+2p_j$ となりこれを用いてモデル選択を行うと良いことは, Akaike (1973) により導入されて以来良く知られている. Z を Y と独立な Y の未来の観測変量とするとき Z の真の分布と Z の予測分布との Kullback-Leibler 情報量の期待値

$$(2.1)\quad E^Y\left[\int f(z|\theta)\log\frac{f(z|\theta)}{f(z|\hat{\theta}_j)}dz\right]=\int f(z|\theta)\log f(z|\theta)dz-E\left[\int f(z|\theta)\log f(z|\hat{\theta}_j)dz\right]$$

の右辺第2項の漸近的不偏推定量を計算することによりAICが求められたが，モデルの分布が$f(y|\theta)$と与えられているなら最尤推定量の漸近正規性を使って漸近的不偏推定量を求めなくても直接正確な不偏推定量を計算した方が原理に忠実ではないかと考えいくつかの標準的な場合について実際それが可能であることを示したのがSugiura (1978)である．n行1列のデータ行列$Y(n\times 1)$に対してn行k列の説明変数行列$X(n\times k)$とk行1列の回帰係数ベクトル$\beta(k\times 1)$による多重線形回帰模型

(2.2)
$$Y(n\times 1)=X(n\times k)\beta(k\times 1)+e(n\times 1)$$

を考える．ここで説明変数行列Xは既知でその階数はkとし，回帰係数ベクトルβは未知母数，eはn行1列の誤差ベクトルで各成分が互いに独立に正規分布$N(0,\sigma^2)$に従う．このとき回帰係数βと誤差分散σ^2の最尤推定量は

(2.3)
$$\hat{\beta}=(X'X)^{-1}X'Y,\quad \hat{\sigma}^2=\frac{1}{n}Y'(I-X(X'X)^{-1}X')Y$$

である．Xの列ベクトルに関する選択問題をM_j(k_j個の列ベクトルに対する回帰係数のみが0でない)$j=1,\cdots,J$としM_Jを最大のモデル$k_J=k$とすればM_jに対して

$$\mathrm{AIC}=n\log(\hat{\sigma}_j^2)+2(k_j+1)$$

となる．しかしM_jが真のモデルを含むとき正確な不偏推定量を用いればAICの有限修正として

(2.4)
$$\text{c-AIC}=n\log(\hat{\sigma}_j^2)+\frac{2(k_j+1)n}{n-k_j-2}$$

が得られる．これは少なくともAICの改良になってなっているはずで$n\to\infty$とすればAICに一致するがnが小さくk_jに近いときは違いが大きくなる．この有限修正はその後Hurvich and Tsai (1989, 1991) により再発見され詳しく調べられた結果nがk_jに近いとき有効であることが示されている．そこでは又Sugiura (1978) でなぜMonte Carlo simulationによりc-AICのperformanceを調べなかったのか奇妙であると批判されている．しかし当時の我が研究室にある計算機といえば機械式か電卓位で計算センターまで出向いてプログラムでもしない限り，せいぜい10回位の繰り返しが限度であった．やっとの思いで当時最新のIBM personal computerを買って貰いSugiura (1979) で簡単な場合に1000回繰り返しのsiumulationを行い真のモデルを良く選択するのはどれか20種類の情報量基準について比較した．その結果はAICとc-AICは余り変わらないことAICよりもむしろSchwarz (1978) のBICの系統の基準の方が良いことがわかった．AICが予測分布の近似を問題にしていてどのモデルが真であるかを判定するための基準ではないからこれは当然ではあるが当時必ずしも明らかではなかったように思う．回帰分析を行うとき現実には予測のために行うのか，影響のある因子（説明変数）の特定のために行うのか渾然となって必ずしもはっきりとは区別できない場合があると思い，そのときでもAICがよい基準であることを期待したのである．当時，経済系の方からは割に柔軟に考えていただいたと思うがAIC関係の専門家からは回帰問題では予測こそが大事であって真のモデルを当てることは余り意味がない，不可能なことをやろうとしているのではないかという批判が強かった．結果が余り面白くないと考えたこともありSugiura (1979)を論文にする機会を失った．しかしどこから聞きつけるのかつい最近までたまに外国からこの資料の請求が届き，論文にしておかなかった事が悔やまれる．

実は Sugiura (1978) も最初に Communications に投稿したときは却下されたという経緯がある。理由は F-統計量を用いればよいではないか，既に Kullback による minimum discrimination information procedure 等あるのに新たに AIC で解析する必要はない，AIC でモデル選択するときはその信頼性が評価できない，すでにある方法と比べて優れていることが示されていない，等である．2 名の審査員のうちどちらも論文の後半で議論している有限修正については一言も触れていない．そこで 5 頁に及ぶ反論の手紙を編集長の Owen 先生に書いたところ，彼は 2 人の審査員から批判は方式に対するもので論文の technical flow に対するものではないとの同意をとった上でこのようなとき自分は論文を出版してどちらが正しいか時の判断に待ちたいと云ってくれ編集長の支持だけでやっと出版された次第です．

この経験以来私は編集に携わるときは Owen 先生の様でなければいけないと思い一方的な判断はしないよう心掛けています．一度だけお会いしたことがあるのですが，Owen 先生は 1991 年 5 月 5 日に亡くなられもはやこのようなお話しが出来ない事が残念です．

なおその後より詳しい比較が Hashimoto, Honda, Inoue and Taguri (1981), Nishii (1984) によりなされ，更に最近，Fujikoshi and Satoh (1995), Satoh, Kobayashi and Fujikoshi (1997) はモデル M_J は真のモデルを含むが考えているモデル M_j は真のモデルを含まないときの偏りも漸近的に修正した基準を多次元の場合に求めている．定数項を除いて 1 次元のときに述べれば

(2.5) $$\mathrm{AIC}_{\mathrm{FS}}=n\log\ \hat{\sigma}_j^2+\frac{2n(k_j+1)}{n-k_j-2}+2k_j(\lambda_j-1)-2(\lambda_j-1)^2$$

ただし $$\lambda_j=\frac{n-k_j\hat{\sigma}_J^2}{n-k_J\hat{\sigma}_j^2}$$

となる．右辺第 2 項は c-AIC の修正項と同じである．Noda, Miyaoka and Itoh (1996) も同様な立場からやや異なる基準を導いている．最尤推定量を使わない場合や真のモデルが含まれていないときの AIC の一般化について Konishi and Kitagawa (1996) の結果は注目に値する．

母数空間の次元が標本の大きさ n と共に大きくなるときの AIC の最適性は Sibata (1981) により示されたが，次元が固定されているときは AIC は一致性を持たず Bayes 流の事後分布を最大にする基準 Schwarz (1978) の BIC

(2.6) $$\mathrm{BIC}=-2\log L(\hat{\theta}_j)+(\log n)p_j$$

がよい．これはモデル M_j が正しい先験確率を π_j とし，M_j が正しいとき母数 θ(p_j 次元) の先験確率密度関数を $p_j(\theta)$ とおくとき，得られたデータ Y(n 次元) に対して，モデル M_j が正しい事後確率

(2.7) $$\frac{\int f(Y|\theta, M_j)\pi_j\rho_j(\theta)d\theta}{\sum_{j=1}^{J}\int f(Y|\theta, M_j)\pi_j\rho_j(\theta)d\theta}$$

を最大とするモデルを選ぼうとする基準で，指数型分布族と自然母数を仮定すると漸近的に BIC が得られる．Neath and Cavanaugh (1997) は指数型分布族と自然母数の仮定をはずしても，分布が充分に滑らかで $\rho_j(\theta)=\mathrm{const.}$ であれば BIC 基準が漸近的に得られること更に $\log n$ の次の項まで評価して

(2.8) $$\mathrm{BIC}_{\mathrm{NC}}=-2\log L(\hat{\theta}_j)+p_j\log\frac{n}{2\pi}+\log|I(\hat{\theta}_j)|-2\log\pi_j$$

を示している．ここで $I(\theta)$ は Fisher 情報量を表わす．いずれにせよ BIC の有限修正を考えるには適当な先験分布に対して正確な事後分布を最大にする基準ということになる．母数に何ら情報がないときは無情報先験分布 (noninformative prior) を仮定して事後分布を評価するのが自然であろう．不幸にして固有の確率分布にはならないが位置母数であれば $d\theta$，尺度母数であれば $d\theta/\theta$ が無情報先験分布である．では初めに述べた正規多重線形回帰模型で $\theta=(\beta,\sigma^2)$ のとき何が無情報先験分布であろうか？一つの解答は Jeffrey prior $|I(\theta)|^{1/2}d\theta$ である．これは θ の座標変換に対して不変となりその意味で一様分布の性質を持っている．又 1 次元正規分布のとき確かに位置母数と尺度母数について無情報先験分布を与えている．正規多重線形回帰模型 (2.2) のとき計算すれば

$$I(\theta)=\begin{pmatrix}\dfrac{1}{\sigma^2}X'X & 0\\ 0 & \dfrac{n}{2\sigma^4}\end{pmatrix} \tag{2.9}$$

であるから

$$\text{Jeffreys prior}\propto\frac{1}{\sigma^{k+2}}|X'X|^{\frac{1}{2}}d\beta d\sigma^2 \tag{2.10}$$

となり

$$\mathrm{BIC}_{\mathrm{Jef}}=n\log(\hat{\sigma}_j^2)-k_j\log(2\pi) \tag{2.11}$$

を得るが，これは k_j が大きくなるほど小さくなるので情報量基準として不適切である．Jeffrey prior は 1 次元のときは良いが多次元のときは良く機能しないということは Bayes 統計の分野では良く知られていて，情報量基準においても不変性だけでは不十分であることが再確認されたといえる．これを補うものとして Berger and Bernardo (1992a, b) 等は基準先験分布 (reference prior) を導入している．この reference と云う用語も uniform や noninformative その他に対して慎重に選ばれたようである．

n が充分大きければ事後分布 $f(\theta|Y)$ は真の先験分布に近いとし，先験分布 $\rho(\theta)$ の期待情報量損失を

$$\int f(Y)\int f(\theta|Y)\log\left\{\frac{f(\theta|Y)}{\rho(\theta)}\right\}d\theta dY \tag{2.12}$$

で定めこれを最大にする $\rho(\theta)$ を考える．$f(Y)$ は Y の周辺確率密度関数である．このとき

$$\rho(\theta)\propto\exp\left(\int f(Y|\theta)\right)\log f(\theta|Y)dY) \tag{2.13}$$

であるが，これは最も好ましくない先験分布あるいは Y の与える情報量が最も大きな先験分布といってもよい．最尤推定量 $\hat{\theta}$ が漸近正規性 $N(\theta,I(\theta)^{-1})$ を持つ正則な場合に $\hat{\theta}$ が与えられたときの θ の事後分布も $N(\hat{\theta},I(\hat{\theta})^{-1})$ で近似出来るとし，母数が $\theta=(\theta_1,\theta_2)$ と分割され θ_1 が θ_2 より重要な母数とする．母数空間が有界閉集合であるとき

（1）$\rho(\theta_2|\theta_1)$ の期待情報量損失を最大にする条件付先験分布は (θ_1,θ_2) の大きさに応じて分割した小行列で $I(\theta)^{-1}=(I^{ij})_{i,j=1,2}$ とおいたとき

$$\rho(\theta_2|\theta_1)\propto|I^{22}-I^{21}(I^{11})^{-1}I^{12}|^{-1/2} \tag{2.14}$$

で与えられる．右辺は $|I_{22}|^{1/2}$ と書いてもよいが母数が3つ以上のグループに分割されているときこの表現が便利である．

（2） $\rho(\theta_1)$ の情報量損失を最大にする先験分布は（1）で得られた $\rho(\theta_2|\theta_1)$ を用いて

(2.15) $$\rho(\theta_1)\propto \exp\left(\int \rho(\theta_2|\theta_1)\log|I^{11}(\theta)|^{-\frac{1}{2}}d\theta_2\right)$$

で与えられる．

（1）と（2）で得た $\rho(\theta_2|\theta_1)$ と $\rho(\theta_1)$ の積として基準先験分布 $\rho(\theta_1, \theta_2)$ が与えれる．特に多重線形回帰模型で $\theta_1=\beta, \theta_2=\sigma^2$ のときは

(2.16) $$\text{reference prior}\propto\frac{c^{-\frac{k}{4}}}{\sigma^2}|X'X|^{\frac{1}{2}}d\beta d\sigma^2$$

となるが，ここで正数 c は母数空間を有界閉集合から全空間へ拡張するときに積分が収束するようにとるために現われる任意の定数である．Jeffrey prior (2.10) との主な違いは σ^2 の指数部に回帰係数の次元 k がないことである．これより (2.7) の分子を評価してモデル M_j のもとで

(2.17) $$\mathrm{BIC}_{\mathrm{Ref}}=(n-k_j)\log(n\hat{\sigma}_j^2)-2\log\Gamma\left(\frac{n-k_j}{2}\right)+k_j\log\frac{c^{1/2}}{\pi}$$

を得る．n が充分に大きいとき $\mathrm{BIC}_{\mathrm{ref}}$ は漸近的に

(2.18) $$\sim(n-k_j)\log\hat{\sigma}_j^2+k_j\log\frac{c^{1/2}}{2\pi}$$

となるので，(2.17) において $c=1$ と置いたものを $\mathrm{BIC}_{\mathrm{ref1}}$, $c=(2\pi n)^2$ と置いたものを $\mathrm{BIC}_{\mathrm{ref2}}$ とする．

Sugiura (1979) では Mallow の C_p や Allen (1974) の PSS (prediction Sum of Squares) 等20種類の基準を比較したがそのなかの幾つかは恣意的に先験分布をとり Schwarz の BIC の有限修正を試みている．すなわちモデル M_j でとられた X の k_j 個の列ベクトルの作る小行列を $X(j)$ として，M_j のもとで

$\rho_j(\theta)d\theta\propto d\beta d(1/\sigma^2)$ ととれば

(2.19) $$\mathrm{UBIC}=n\log\hat{\sigma}_j^2-(k_j+1)\log(2\pi\hat{\sigma}_j^2)+\log|X(j)X(j)'|$$

$\rho_j(\theta)d\theta\propto e^{-(\beta'\beta/2\sigma^2)}d\beta d(1/\sigma^2)$ ととれば

(2.20) $$\begin{aligned}\mathrm{NBIC1}=&(n-1)\log[Y'\{I-X(j)(X(j)'X(j)+I)^{-1}X(j)'\}Y]\\&+\log|I+X(j)'X(j)|\end{aligned}$$

$\rho_j(\theta)d\theta\propto e^{-(\beta'\beta/2\sigma^2)}d\beta e^{-(1/2\sigma^2)}d(1/\sigma^2)$ ととれば

(2.21) $$\begin{aligned}\mathrm{NBIC2}=&(n+2)\log[1+Y'\{I-X(j)(X(j)'X(j)+I)^{-1}X(j)'\}Y]\\&+\log|I+X(j)'X(j)|\end{aligned}$$

が得られるが，20種類の基準のうちで良い結果を与えたのは NBIC1, NBIC2 であった．ここでも Sugiura (1979) と同じ簡単な2つの例についてその後に新たに得られた情報量基準を適用してみる．説明変数が定数項を含めて3つある場合で真のモデルが $Y=1+X_1+X_2+e$ のとき

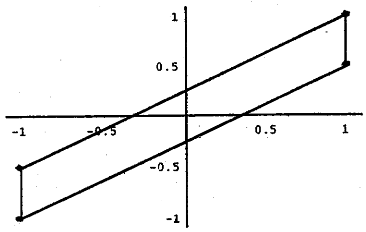

図 2.1 (X_1, X_2) のグラフ $\rho(X_1, X_2)=0.95$

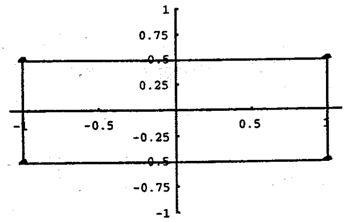

図 2.2 (X_1, X_2) グラフ $\rho(X_1, X_2)=0$

を $X(012)$; $Y=1+X_1+e$ のときを $X(01)$; $Y=1+X_2+e$ のときを $X(02)$; $Y=1+e$ のときを $X(0)$ で表わす．第一の例は説明変数の間の相関が高い場合で図 2.1 のように 4 点 $(X_1, X_2)=(-1,-1), (-1,-0.5), (1, 0.5), (1, 1)$ を同じ回数繰り返す（Gorman and Toman 1966）．このとき X_1 と X_2 の相関係数は $\rho(X_1, X_2)=0.95$ となりモデル選択はやりにくい例である．第 2 の例は図 2.2 のように 4 点 $(X_1, X_2)=(-1,-0.5), (-1, 0.5), (1,-0.5), (1, 0.5)$ を同じ回数繰り返す場合で $\rho(X_1, X_2)=0$ となる．

誤差 e は平均 0 分散 σ^2 の正規分布で互いに独立とする．標本の大きさを $n=16$ または 100 とし，母分散を $\sigma^2=1$ または 1/4 ととり各モデルの下で正規乱数により 1000 回の繰り返しで真のモデルが選択された回数を次の表 2.1 と表 2.2 に示す．この例については c-AIC は AIC と NBIC1 は NBIC2 とほとんど同じ結果であったので省略してある．ごくわずかに NBIC2 が NBIC1 より良いかという程度であった．また UBIC も結果が不安定で良くなかったのでここではとりあげていない（Sugiura 1979）．

表 2.1，表 2.2 より次のことがわかる．

（1） AIC, $\mathrm{AIC_{FS}}$, BIC 共に $X(012)$ が真のモデルのとき極端に悪いことがある．

表 2.1 1000回の simulation により真のモデルが選択された回数

$\rho(X_1, X_2)=0.95$

	(i) $n=16, \sigma^2=1$				(ii) $n=16, \sigma^2=1/4$			
	$X(012)$	$X(01)$	$X(02)$	$X(0)$	$X(012)$	$X(01)$	$X(02)$	$X(0)$
AIC	58	617	552	699	648	782	722	704
$\mathrm{AIC_{FS}}$	4	685	608	872	244	881	808	875
BIC	20	640	574	801	513	836	762	805
NBIC2	678	660	636	850	991	614	637	911
$\mathrm{BIC_{NC}}$	520	279	500	780	635	784	783	921
$\mathrm{BIC_{ref1}}$	1000	0	0	0	991	40	30	11
$\mathrm{BIC_{ref2}}$	17	636	573	759	290	878	806	894

	(iii) $n=100, \sigma^2=1$			
	$X(012)$	$X(01)$	$X(02)$	$X(0)$
AIC	818	835	824	767
$\mathrm{AIC_{FS}}$	792	853	837	792
BIC	479	920	891	952
NBIC2	901	860	890	944
$\mathrm{BIC_{NC}}$	812	834	867	933
$\mathrm{BIC_{ref1}}$	1000	0	0	0
$\mathrm{BIC_{ref2}}$	564	913	885	931

表2.2 1000回の simulation により真のモデルが選択された回数

$\rho(X_1, X_2)=0$

	(i) $n=16, \sigma^2=1$				(ii) $n=16, \sigma^2=1/4$			
	$X(012)$	$X(01)$	$X(02)$	$X(0)$	$X(012)$	$X(01)$	$X(02)$	$X(0)$
AIC	736	789	585	638	998	788	787	631
$\mathrm{AIC_{FS}}$	486	920	549	831	974	934	927	808
BIC	659	858	568	755	994	856	852	751
NBIC2	713	812	661	720	991	902	957	816
$\mathrm{BIC_{NC}}$	744	782	705	696	987	915	956	879
$\mathrm{BIC_{ref1}}$	1000	0	1	0	1000	32	29	14
$\mathrm{BIC_{ref2}}$	715	835	606	679	987	924	923	851

	(iii) $n=100, \sigma^2=1$			
	$X(012)$	$X(01)$	$X(02)$	$X(0)$
AIC	999	837	820	767
$\mathrm{AIC_{FS}}$	999	861	842	792
BIC	998	962	956	952
NBIC 2	999	924	961	944
$\mathrm{BIC_{NC}}$	999	923	961	933
$\mathrm{BIC_{ref1}}$	1000	0	0	0
$\mathrm{BIC_{ref2}}$	998	949	947	931

(2) $\mathrm{AIC_{FS}}$ は大体において AIC よりよいが，$X(012)$ が真のとき悪くなることがある．c-AIC も全く同じ傾向が認められた．

(3) $\mathrm{BIC_{NC}}$ は大体において BIC より良い．

(4) NBIC2 は多くの場合 $\mathrm{BIC_{NC}}$ よりよい．

(5) $\mathrm{BIC_{ref1}}$ は常に $X(0, 12)$ を選択する傾向が強く良くない．

(6) $\mathrm{BIC_{ref2}}$ は BIC とくらべて特によいわけではない．

(7) $\mathrm{BIC_{ref1}}$ は不安定で使えない．これは Jeffrey prior をとった場合に近い．

(8) NBIC2 は先験分布に理論的根拠はないがこの例に関する限りいつも割に安定してよい結果を示す．

以上によりモデル選択においても Jeffreys prior より Reference prior の方が良いことがわかったが，無情報先験分布としては未だ充分満足出来るまでには至ってないと思われる．依然として昔の NBIC2 等が良いからである．基準先験分布は現在いろいろな問題に対して試されていて困難に出会う度に改良されているので今後の発展が期待される．

3. 統計学と統計学会の将来

21世紀は計算機と宗教の時代であるとかいわれています．確かに計算機の発達と社会への普及その影響の大きさは計り知れないものがあります．では統計学もそれにつれて一般社会でより多く使われるようになりより盛んになるでしょうか？日本の現実では結論が先にあり説明に困るので統計学の用語を利用してなんとか説明できないかという使われ方もあるのが残念ながら事実です．母集団の概念さえなかなか受け入れてもらえません．米国ではもう少し専門外の方でも信頼係数はとか有意水準は……，とかに理解があったように思います．しかしその米国で現に SUNY (State University of New York) in Buffalo の統計学科が社会予防医学科に吸収されようとしているとか，Cornell 大学では学科間の統計活動が一つの新しい学科内に整理

統合されようとしているとか暗い話しが一方であります(Kettenring 1997). 私は今まで統計学が過剰に期待されていた部分がありそれが段々に整理されようとしているのではないかと云う気がしています. 統計学が適用される場面は社会の限られた部分に過ぎませんが, 確かに実社会に貢献できる面がありその限りにおいて将来も相応の立場を保持できると思います. そのためには出来る限り多くの窓口を外の世界に保持すべきです. Kettenring (1997)は必ずしも統計学の研究者とはならない修士課程の学生教育をより重視し修士号が博士号に達成出来なかった"残念賞"として授与されるような統計学科であってはならないと述べています.

私の所属する筑波大学でもいくつかの学科が自ら外部評価を受けて報告書をだしています. 教員の任期制とか大学の自由化とか新しい問題も起きています. 抽象的な言い方ですが, 当学会も外部評価に耐えられる学会でありたいと望んでいます. そのためには若い人々を惹き付けられる魅力がないといけません. 積極的に勧誘するのではなく自然に集まるだけの魅力が必要です. 最近の新入会員の加入の様子を見ておりますと人数はさておき割にうまく行っているのではないでしょうか. 更に大事なことは会員の研究活動の一つの指針となる学会誌があります. 10数年前ですが日本人ではなく, 外国人同士がSouth Africa Statist. J. はCurrent Index にちゃんと引用されているのにJ. Japan Statist Soc. は引用されていないからCurrent Index が必ずしも公平ではないと話していたのを側で聞いたことがあります. 今では日本統計学会誌も引用されていると思いますが, 私はこのような会話も一つの学会に対する外部評価と思っています. 最近会員の方から日本統計学会誌の審査が遅い, 審査内容が貧弱である, 問い合わせに対して迅速な返事がなかった, 等の声を耳にし学会誌に対する関心の深いのを知り喜んでいます. 今回の大会や理事会の際に欧文誌, 和文誌の編集担当理事の方と話し合う機会が持て, 活発に動いて下さっていますので, やがて事態は改善されると思います. 学会誌が今年度の学会報告集位の厚さになると我々の存在感がもっと強力になるのではないかと望んでいます.

連続変量のデータのヒストグラムを描くのに柱状に描くのでは密度関数の過小評価になるので折れ線でちょうど不偏推定となるよう新しい度数多角形の表示法を議論した論文 Casteren (1997)や, 神経回路網模型で回帰直線を推定すると線形回帰理論を用いたものとほとんど同じ答えが得られ, 更に統計理論では扱いが難しい様な曲線回帰でも真の曲線に近いものが推定できることを示した解説論文 Warner and Misra (1996) 等を見ますと未だ統計学は研究する事が入口のところに色々あって若い学問であり未来は明るいという気が致しますが, 楽観的過ぎるでしょうか?

最後にここでは触れませんでしたが, 振り返って見ますと私が今日あるのは全く予期せぬ時に, 予期せぬ所で, 立派な先輩の先生方に出会えたお陰であると思います. 改めて深く感謝致します.

なおここでの数値計算には Mac 9500/200 に Mathematica V. 2.2 を使用しています.

参考文献

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *2nd Int. Symp. Inf. Theory* (Eds. B. N. Petrov and F. Csaki), Budapest : Akademiai Kiado, 267-281.

Allen, D. M. (1974). The relationship between variable selection and data augmentation and a method for prediction. *Technometrics* **16**, 125-127.

Berger, J. O. and Bernardo, J. M. (1992a). On the development of reference priors. *Bayesian Statistics 4* (Eds. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), Oxford : University Press, 35-60.

Berger, J. O. and Bernardo, J. M. (1992b). Reference priors in a variance components problem. *Bayesian Analysis in Statist. Econometrics.* (eds. P. Goel & N. S. Iyengar) Springer, New York, 177-194.

Casteren, P. H. F. M. (1997). The frequency polygon reconsidered. *Comp. Statist. Data Anal.* **24**, 1-8.

David, H. A. (1981). *Order Statistics*. 2nd Ed. Wiley, New York.

Fujikoshi, Y. and Satoh, K. (1995). Modified AIC and C_p criterion in multivariate linear regression. TR95-21 Dept. Math. Hiroshima University.

Gorman, J. W. and Toman, R. J. (1966). Selection of variables for fitting equations to data. *Technometrics* 8, 27-51.

Gumbel, E. J. (1954). The maxima of the mean largest value and of the range. *Ann. Math. Statist.* **25**, 76-84.

Hartley, H. O. and David, H. A. (1954). Universal bounds for mean range and extreme observation. *Ann. Math. Statist.* **25**, 85-99.

Hashimoto, A. Honnda, M. Inoue, T. and Taguri, M. (1981). Selection of regression models by several information criterions. *Rep. Stat. Appl. Res. JUSE* **28**, 57-72.

Hurvich, C. M. and Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika* **76**, 297-307.

Hurvich, C. M. and Tsai, C.-L. (1991). Bias of the corrected AIC criterion for underfitted regression and time series models. *Biometrika* **78**, 499-509.

Kettenring, J. (1997). The birth, life, and death of statistical department. President's Corner, *AMSTAT NEWS* **245**, 9-10. Amer. Statist. Assoc.

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

Moriguti, S. (1951). Extremal properties of extreme value distributions. *Ann Math. Statist.* **22**, 423-536.

Moriguti, S. (1953). A modification of Schwarz's inequality with applications to distributions. *Ann. Math. Statist.* **24**, 107-113.

森口繁一 (1995). 確率表現関数 東京大学出版会.

Nagao, H. (1973). On some test criteria for covariance matrix. *Ann. Statist.* **1**, 700-709.

Neath, A. A. and Cavanaugh, J. E. (1997). Regression and time series model selection using variants of the Schwarz information criterion. *Commun. Statist.-Theory Meth.* **26** (3) ,559-580.

Nishii, R. (1984). Asymptotic properties of criteria for selection of variables in multiple regression. *Ann. Statist.* **12**, 758-765.

Noda, K., Miyaoka, E. and Itoh, M. (1996). On bias correction of the Akaike's information criterion in linear models. *Commun. Staist.-Theory Meth.*, **25** (8), 1845-1857.

Satoh, K., Kobayashi, M. and Fujikoshi, Y. (1997). Variable selection for the growth curve model. *J. Mult. Analysis* **60**, 277-292.

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

Shibata, R. (1981). An optimal selection of regression variables. *Biometrika* **68**, 45-54.

Sugiura, N. (1962). On the orthogonal inverse expansion with an application to the moments of order statistics. *Osaka Math. J.* **14**, 253-263.

Sugiura, N. (1963). On a generalization of the wilcoxon test for censored data. *Osaka Math. J.* **15**, 257-268. Correction : *Osaka J. Math.* **2** (1965), 283.

Sugiura, N. (1964). A generalization of the wilcoxon test for censored data II, —several sample problem—. *Osaka J. Math.* **1**, 165-174.

Sugiura, N. (1965). Multisample and multivariate nonparametric tests based on U statistics and their asymptotic efficiencies. *Osaka J. Math.* **2**, 385-426.

Sugiura, N. (1969). Asymptotic expansions of the distributions of the likelihood ratio criteria for covariance matrix. *Ann. Math. Statist.* **40**, 2051-2063.

Sugiura, N. (1972). Locally best invariant test for sphericity and the limiting distributions. *Ann. Math. Statist.* **43**, 1312-1316.

Sugiura, N. (1973a). Derivatives of the characteristic root of a symmetric or Hermitian matrix with two applications in multivariate analysis. *Commun. Statist.* **1**, 393-417.

Sugiura, N. (1973b). Asymptotic non-null distributions of the likelihood ratio criteria for covariance matrix under local alternatives. *Ann. Statist.* **1**, 718-728.

Sugiura, N. (1976a). Asymptotic expansions of the distributions of the canonical correlations. Commemoration Vol. of Founding of Fac. of Integrated Arts and Sci. Hiroshima Univ. 141-163.

Sugiura, N. (1976b). Asymptotic expansions of the distributions of the latent roots or latent vector of the wishart and multivariate F matrices. *J. Mult. Analysis* **6**, 500-525.

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.

Sugiura, N. (1979). Comparison of information criteria in the selection of regressor variables. *International Conference in Statistics, Tokyo* 178-181.

Sugiura, N. (1988). A class of improved estimators of powers of the generalized variance and precision under squared loss. *Statistical Theory and Data Analysis II* (ed. K. Matusita) 421-428, North-Holland.

Sugiura, N. (1989). Entropy loss and a class of improved estimators for powers of the generalized variance. *Sankhyā* Ser. A, **51**, 328-333.

Sugiura, N. (1990). Graphs of the distributions of bivariate Wishart roots and their cumulants. *J. Japan Statist. Soc.* **20**, 117-136.

Sugiura, N. (1992). Lower bounds for the risks of equivariant estimators of generalized variance. *Nonparametric Statistics and Related Topics* (ed. A. K. Md. E. Saleh) 253-267. Elsevier Sci. Pub. Amsterdam.

Sugiura, N. (1994). Approximating Bayes critical region for testing simple and tree ordered normal means. *J. Statist. Research* **28**, 1-20.

Sugiura, N. (1995). Exact nonnull distributions of sphericity tests for trivariate normal population with power comparison. Amer. *J. Math. & Manag. Sci.* **15**, 355-374.

Sugiura, N. (1997). Exact powers of classical tests for bivariate linear hypotheses. *Statistical Inference and Related Topics,* (eds. S. E. Ahmed, M. Ahsanullah and B. K. Sinha), Nova Science Publishers, Inc., New York, in press.

Sugiura, N. and Fujikoshi, Y. (1969). Asymptotic expansions of the non-null distributions of the likelihood ratio criteria for multivariate linear hypotheses and independence. *Ann. Math. Statist.* **40**, 942-952.

Sugiura, N. and Ishibayashi, H. (1997). Referece prior Bayes estimator for bivariate normal covariance matrix with risk comparison. *Commun. Statist.-Theory Meth.* **26** (9), 2203-2221.

Sugiura, N. and Konno, Y. (1987). Risk of improved estimators for generalized variance and precision. *Advances in Multivariate Statistical Analysis* (ed. A. K. Gupta) 353-371 D. Reidel, Holland.

Sugiura, N. and Kubokawa, T. (1988). Estimating common parameters of growth curve models. *Ann. Inst. Statist. Math.* **40**, 119-135.

Sugiura, N. and Nagao, H. (1968). Unbiasedness of some test criteria for the equality of one or two covariance matrices. *Ann. Math. Statist.* **39**, 1686-1692.

Sugiura, N. and Ogden R. T. (1994). Testing change points with linear trend. *Commun. Statist.-Simul.* **23** (2), 287-322.

Sugiura, N. and Otake, M. (1973). Approximate distribution of the maximum of $c-1$ χ^2 statistics derived from $2\times c$ contingency table. *Commun. Statist.* **1**, 9-16.

Sugiura, N. and Otake, M. (1974). An extension of the Mantel-Haensel procedures to k $2\times c$ contingency tables and the relation to the logit model. *Commun. Statist.* **3**, 829-842.

Sugiura, N. and Takagi, Y. (1996). Dominating James-Stein positive part estimator for normal mean with unknown covariance matrix. *Commun. Statist.-Theory Meth.* **25**, 2875-2900.

高田佳和, 橋本 勲, 坂田年男, 柳川 堯 (1981). 田村亮二教授論文集.

Tamura, R. (1966). Multivariate nonparametric several sample tests. *Ann. Math. Statist.* **37**, 611-618.

Warner, B. and Misra, M. (1996). Understanding neural networks as statistical tools. *Amer. Statistician* **50** (4), 284-293.