

②

赤池のベイズ型情報量規準を用いた  
種畜評価モデルの選択法に関する  
研 究

和田 康彦

1991

## 目 次

緒 論	1
第1章 赤池のベイズ型情報量規準を用いたモデル比較法	5
緒 言	5
第1節 赤池のベイズ型情報量規準	7
第2節 混合モデルにおける赤池のベイズ型情報量規準	9
第3節 より複雑なモデルにおける赤池のベイズ型情報量規準	12
第4節 計算例	15
摘 要	18
第2章 シミュレーションによる有効性の検討	20
緒 言	20
方 法	21
結 果	25
考 察	29
摘 要	29
第3章 大規模データのための計算手法の開発	30
緒 言	30
材料および方法	30
結 果	32
考 察	32
摘 要	36
第4章 乳牛のフィールドデータを用いた有効性の検討	37
緒 言	37
材料および方法	38
結 果	41
考 察	43
摘 要	44
総 括	46
謝 辞	48

## 緒 論

昔から牛を飼育していた人々は、良い性質を持った親の子牛はやはり優れていることが多いということを知っていた。特に、雄牛は多くの子牛の親となるため若雄牛の間から毛色や角質、体型といった外貌形質を手がかりとして選抜が行われてきた。20世紀にはいと、遺伝学の進歩にともなって後代の記録から種雄牛の遺伝的な能力を評価しようという動きがでてきた。まず、乳牛において Hansson(1913) と Yapp(1925) は、娘牛の記録の2倍から母牛の記録をひいた値を種雄牛の評価値とすることを提案した。この値は Hansson と Yapp の指数と呼ばれ、種々の検討ののちに世界各国の検定場で広く使われていった。

Wright(1931) と Lush(1931) は遺伝率の概念をはじめて種雄牛の評価手法に取り入れて、娘牛の頭数に従って評価値を補正する方法を提案した。また、Lush(1935) は能力検定と後代検定を比較し、娘牛の頭数を増加させることによって能力検定よりも後代検定の方がより正確に育種価を推定できることを示した。しかし、彼らの方法はその先進性と必要な遺伝パラメータを推定するためのデータが得られなかったことから長い間にわたって積極的に使われることはなかった。

第2次世界大戦後、人工授精の普及にともなって、1頭の種雄牛が生産する後代牛数が飛躍的に増加した。これにより、種雄牛の遺伝的な能力を正確に評価することが牛の育種改良にとっての最重要課題となるとともに、能力の保証のある後代検定済みの種雄牛の人気が高まった。たとえば世界で最も早く人工授精が普及したデンマークでははじめは母娘比較によって乳用種雄牛を評価して後代検定済みの種雄牛を供給していた。しかし、戦時中の低エネルギー飼料で検定を受けた母牛と戦後の高エネルギー飼料で検定された娘牛を比較することには大いに問題があったので、大規模な中央検定場を設けて、同一の飼養管理条件のもとでの後代検定を行うようになっていった。

その後、Robertson と Rendel(1950) はフィールドデータを用いた同期比較によって種雄牛を評価する方法を提案した。彼らの方法は同期比較法と呼ばれ、世界中でその国の条件に合うように修正を施されながら、広く使われていった。彼らの方法は、(1) 各種雄牛の後代が各牛群にランダムに分布して

いること、(2) 牛群間に遺伝的な能力水準の差がないことを前提条件としていたが、これらの条件は当時のイギリスではほぼ満たされていた (Robertson and Rendel(1954))。アメリカでは1962年から同群比較法 (Henderson et al.(1954)) が乳用種雄牛の評価に用いられた。この方法ではデータ数を増すことをねらいとして、初産のみならずすべての産次、すべての年次・季節のデータを取り入れた。同群比較法は、同期比較法での前提条件に追加して、(3) 遺伝的趨勢があまり大きくなく、(4) すべての牛群において同一の改良目標に沿って選抜が行われていることを前提条件とすることによって、検定に使用するデータ数を増してより多くの種雄牛をより正確に評価することをめざしたのであった。

さらに、アメリカでは1974年から乳用種雄牛の評価法が変更になり、修正同期比較法が採用された (Dickinson et al.(1976), Norman(1976))。この方法の長所は、(1) 同じ牛群に属する母牛の父親の遺伝的レベルを考慮することにより、年次を越えた比較が可能なこと、(2) 当該雄牛の父親と母方祖父の遺伝的レベルを考慮している点である。

一方、Henderson(1949)は混合モデルを用いた統一的な種畜の遺伝能力の評価手法を開発し、1973年のHendersonの論文 (Henderson(1973)) をきっかけとして注目され、急速に各国に普及していった。彼の方法は混合線形モデルを基本として、現代数理統計学で定義されているところの最良線形不偏予測量 (best linear unbiased predictor; BLUP) を導くものである。このBLUP法によって牛群間の遺伝的なレベルの差や年次の経過による遺伝的趨勢が認められたり、各牛群における種雄牛の分布に相関が認められるようなデータに対しても、モデルを選択することによって、より正確な評価が可能となった。

BLUP法の最も大きな長所は、モデルを工夫することによって、同期比較法や同群比較法における前提条件を取り去ることが可能となったことである。そのためデータに合わせたいろいろなモデルが提案されて、検討されてきた。まず、種雄牛間の血縁関係を考慮するモデルが提案され (Henderson(1975b))、実際のデータについて検討された (Kennedy and Moxley(1975), 佐々木と祝前(1980))。このモデルにより、評価する種雄牛間に濃い血縁関係が存在する場合でも、分子血縁係数行列 (A行列) を考慮することによって正確な評価ができるようになった。また、このモデルを使えば後代のデータのない種雄牛でも、血縁関係からある程度の正確度で評価できるようになる (Henderson(1977))。この考え方を拡張してモデルに種雄牛だけではなくて雌牛も含めて評価する種雄牛-雌牛モデル (Sire and Dam Model) や、自分自身のデータを評価に用いる個体モデル (Animal Model) が提案され、乳用種においてはアメリカ、カナダをはじめ今までの種雄牛モデルから個体モデルに移行する国が増えつつある。種雄牛の遺伝的なグループをモデルに加えるこ

とは簡単にできるが、実際の評価にあたってモデルに加えるべきかどうかは、いまだに議論の対象となっている (Keown(1974), Kennedy(1981), 佐々木と佐々江(1988))。各雌牛の遺伝的な能力の違いを考慮するために、母方祖父の効果をモデルに取り入れることも検討された (Everett et al.(1979), Quaas et al.(1979), 伊藤と佐々木(1985))。また、HendersonとQuaas(1976)は、BLUP法を遺伝相関のある複数形質の場合に拡張した。その後、Cornell大学では複数形質のBLUP法について種々の検討が行われてきた (Quaas and Pollak(1980), Pollak and Quaas(1981a,b), Pollak et al. (1984))。そのほか、複数の産次のデータを扱うリピータビリティモデル (Henderson(1974), Ufford et al.(1979))、種雄牛と雌牛を同時に評価する際に計算の便を考えた縮約化個体モデル (Quaas and Pollak(1980), Henderson(1985c))、直接検定モデル (Sasaki and Henderson(1986))、非相加的な遺伝効果を考慮したモデル (Henderson(1985a,b))、カテゴリカルな形質に対するモデル (Gianola(1980), Quaas and Van Vleck(1980), Van Vleck and Karner(1980)) など枚挙に暇がない。

BLUP法の欠点は、実際のフィールドデータを前にしたときに、これだけたくさんあるモデルのなかのどのモデルを種雄牛評価のモデルとして選ぶかを決定するための手法がないことである。さらに、ひとつのモデルを選んだとしても、どのように各効果の水準を分けるかも大きな問題である。何人かの研究者がBLUP法におけるモデル選択の方法に取り組んだが (Henderson(1975c), Kennedy and Moxley(1975), Pollak et al.(1977), Kennedy(1981))、真の母数効果の値が必要であったり、一部のモデルにのみ適用可能であったりして、実際の種雄牛評価の場面で使用されるにはいたっていない。

またBLUP法においては評価の前に種雄牛分散と誤差分散の比が既知でなければならない。複数形質モデルの場合には、種雄牛効果の分散共分散行列と誤差の分散共分散行列が必要となる。BLUP法が普及するまでは、混合モデルの分散成分の推定法としてはHendersonの方法3 (Henderson(1953)) が広く使われていた。しかし、この方法を実際のデータに適用してみると遺伝率が1を越えたり遺伝相関の絶対値が1を越えるなど、データによってはかなり不正確な値が推定されることもあった。さらに、この方法は係数行列の逆行列とその逆行列の各効果に相当する部分の小行列の逆行列が必要となる。実験計画法にもとづくようなデータの場合はこれでもよいが、フィールドデータによる種雄牛評価の場合には、係数行列の次元が数十万と言うオーダーになり、小行列ですら数万のオーダーで、計算は非常に困難である。

そこで、最尤法やREML法 (Restricted Maximum Likelihood Method) といったおもに反復によって分散成分の推定値を求める手法が提案されてきた (Hartley and Rao(1967), Henderson(1973), Schaeffer(1976); Patterson and Thompson(1971), Harville(1977), Schaeffer et al.(1978), Hender-

son(1984),Henderson(1985b))。これらの方法は、各反復でBLUP法の計算と分散成分の推定の計算を合わせておこない、収束した時点で、分散比が既知であるBLUP法として考えて、その最良線形不偏予測量を評価値とするもので、種畜評価の立場から見れば「反復BLUP法」と呼ぶべき方法である(和田と建部(1986)、和田ら(1987))。しかし、反復法は、場合によってはなかなか収束せずに大量の計算コストが必要になったり、初期値の与え方によっては真の値とは全く異なる局所的な最適値に収束してしまう可能性がある。

本研究では、以上に述べた種雄牛評価の歴史を背景として、BLUP法の問題点を解決すべく次の諸点について検討する。第1章では日本から始まった新しい数理統計学の流れである情報量統計学(Akaike(1973), Akaike(1980))の視点からBLUP法を見直して、赤池のベイズ型情報量規準(Akaike's Bayesian information criterion; ABIC)を用いたモデル選択の手法と分散成分の推定を含めた、統一的でかつ実用的な種雄牛評価手法を提案する。第2章では我が国の乳牛集団のフィールドデータを前提として作成したデータもちいて、新手法の有効性をシミュレーションによって検討する。第3章では混合モデルにおけるABC算出にあたっての計算上の問題点を検討し、現状で利用できるABC算出のためのアルゴリズムについて比較検討する。第4章では日本におけるホルスタイン種のフィールドデータを材料としてABCを用いて選択したモデルが、どの程度正確に種雄牛の遺伝的能力を評価できるのかについて検討する。

## 第1章

# 赤池のベイズ型情報量規準を用いたモデル比較法

### 緒言

BLUP法(Henderson(1949), Henderson(1963), Henderson(1973))は、現在のところ世界で最も広く使用されている種雄牛評価手法であろう。この方法は種雄牛効果を変量効果と考え、牛群や年次、季節などの効果を母数効果と考えて混合線形モデルをたてる。そして、変量効果の最良線形不偏予測量(best linear unbiased predictor; BLUP)を導くものである。Hendersonの方法は、牛群間の遺伝的なレベルの差や、年次の経過による遺伝的趨勢が認められたり、牛群と種雄牛の分布の間に相関が認められるようなデータに対して、モデルにもとづいてそれらの補正をして、より正確な評価をすることができる手法である。BLUP法の出現によって、遺伝的なレベルに差のある広い地域から集められた長期間にわたるフィールドデータを種雄牛評価に利用することが可能となった。

しかし、BLUP法にも問題点がないわけではない。初期に問題となったのは計算コストであったが、これはコンピュータの飛躍的な発達によって、それほど問題とはならなくなってきた。次に問題となったのは、BLUP計算をするにあたって必要となる分散比あるいは種雄牛および誤差の分散共分散成分をどのように決めるかであった。最初は文献値で代用されていたが、次第にフィールドデータを分析して分散共分散成分を推定すべきであるという意見が強まってきた。そこで、より正確な推定が可能な、最尤法(maximum likelihood method)やREML法(restricted maximum likelihood method)といった反復によって分散共分散成分の推定値を求める手法が提案されてきた(Hartley and Rao(1967),Henderson(1973); Patterson and Thompson(1971), Schaeffer et al.(1978), Henderson(1984), Henderson(1985b))。これらの方法は、各反復でBLUP法の計算と分散成分の推定の計算を合わせて行い、

収束した時点で分散成分の推定値が得られたと考えて、その時点でのBLUPの計算の結果を評価値とするもので、実質的には「反復BLUP法」と呼ぶべき方法である(和田と建部(1986)、和田ら(1987))。純粋に数理統計学的に考えればBLUP法における評価値は、分析するデータに対して真のモデルと真の分散比が与えられた時に最良線形不偏予測量となるのである。従って、「反復BLUP法」による評価値は、真の分散比ではなく推定値を用いているのであるから厳密な意味での最良線形不偏予測量ではない。また、一般に反復法は、場合によってはなかなか収束せず多額の計算コストが必要になったり、初期値の与え方によっては真の値とは全く異なる局所的な最適解に収束してしまう可能性もある(Jacobyら(1972))。

しかし、BLUP法の最大の問題点は、実際のフィールドデータを用いて種雄牛評価をしようとするときに、考えられる数多くのモデルの中からどのモデルを選ぶかを科学的に決定する方法がないことである。さらに、ひとつのモデルを選んだとしても、どのように各効果の水準を分けるかも大きな問題になる場合もある。今までは、個々のデータに対して分散分析などにより分析した結果から考察したり(Kennedy and Moxley(1975)、Everettら(1979)、Uffordら(1979)、Van Vleck(1980)、佐々木と祝前(1980)、伊藤と佐々木(1985)、佐々木と佐々江(1988))、シミュレーションによってそのモデルの有効性が検討されてきた(Pollak and Quaas(1981a,b,1983)、Pollak et al.(1984))。しかし、分析結果からの考察では経験と勘に頼る面が多い。また、シミュレーションでは実施されたシミュレーションモデルが、実際のデータを正確にシミュレートしているのかという疑問が消えない。そこで、Henderson(1975)はモデル選択のための基準として、不偏性と予測誤差分散を使うことを提案した。しかし、彼の提案は必ずしも全てのモデルに適用できるものではなく、特に、不偏性の指標となる平均平方誤差(mean square error)を計算するには、モデルに組み入れるかどうかを検討する効果において各水準の真の母数効果の値が必要であり、実際の種雄牛評価の場面で使用するには問題が多い。また、予測誤差分散の計算には混合モデル方程式の係数行列の一般化逆行列を求める必要があり、この計算は評価値を求めるだけのBLUP計算とは比較にならないほど計算時間を必要とする。

分散分析などの旧来の統計手法では、必ずモデルの善し悪しをチェックするために検定を行う。F検定とかt検定と言われているものがそれである。しかし、BLUP法には有効なモデル評価手法が存在しない。これは、BLUP法が立脚しているところのRao(1973)流の統計的推測の理論では、BLUP法で実施している種雄牛と誤差の分散共分散成分を事前の情報として与えて分析するという手法を完全には扱いきれないところからきている。

そこで、本章では事前の情報をモデルのなかに明確な形で取り入れるために、BLUP法の基本モデルをベイズ統計学の観点から見直して、情報量統計

学におけるベイズ型情報量規準(Akaike's Bayesian Information Criterion; ABIC; Akaike(1980))を用いたモデルの選択法を提案する。まず第1節ではABICを用いたモデル選択法を概説し、第2節ではHenderson流の基本的な混合モデルをベイズ型のモデルに書き直し、情報量統計学の理論に従って評価値とABICを求めてモデルを選択する方法を示す。つぎにこの手法を第3節では複数の変量効果を含むモデルや分散比が未知の場合のようなより複雑なモデルに拡張し、第4節では理解を助けるために簡単な計算例を提示し、摘要では新手法の有効性について考察する。

## 1.1 赤池のベイズ型情報量規準を用いたモデルの比較

Fisher(1959)などの古典的な統計手法においては、

1. モデルの設定(specification)、
2. パラメータの推定(estimation)、
3. モデルの検定(test)

を繰り返して、より良いモデルを選び、より正確な推定を行なう。しかし、BLUP法ではモデルの検定のための手法が提示されていない。種雄牛評価のように、完全に計画された実験のデータではない、フィールドのデータを扱わねばならない場合には、このことは非常に深刻な問題となる。現状では、BLUP法のモデルを決める前に通常の分散分析を実施して、意味のある母数効果をモデルに含めるようにしている。しかし、全国規模の種雄牛評価の場合には、通常の分散分析は係数行列の逆行列などが必要となるため、非常に多くの計算費用がかかる。また、種雄牛間の血縁関係のような変量効果の事前分布については、通常の分散分析で検定することは全く不可能である。

そこで、種雄牛評価においてもAkaike(1980)によって提唱された赤池のベイズ型情報量規準(Akaike's Bayesian information criterion; ABIC)によってモデルの良し悪しを計ることを次節で提案したい。この節ではABICとそれを生み出した情報量統計学について簡単に説明し、次節への導入とする。

さて、ボルツマンによって導入された熱力学的エントロピーを確率論的に解釈すれば、ある確率分布が他のひとつの確率分布からどの程度隔たっているかを測るための自然な尺度を与える(Akaike(1973)、日本数学会(1985))。確率密度関数 $f(y)$ をもつ分布の、確率密度関数 $g(y)$ で定められる分布に関するエントロピー(entropy)は次の式によって定義される。

$$B(f;g) = -\int \frac{f(y)}{g(y)} \log \frac{f(y)}{g(y)} g(y) dy$$

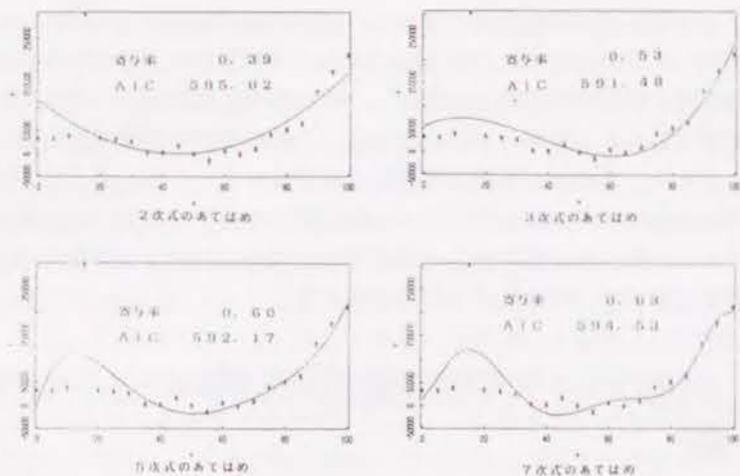


図 1.1: 異常値を含むデータへの直交多項式の当てはめと AIC

ここで  $f$  を任意の統計モデル、 $g$  を真のモデルとするとエントロピー  $B(f; g)$  は統計モデル  $f$  と真のモデル  $g$  との隔たりの程度を表わすことになる。Akaike(1973) はここで述べた統計モデルのエントロピーの考え方と最尤推定値の種々の性質から、統計モデルのエントロピーの逆数の推定値として赤池の情報量規準 (Akaike's information criterion; AIC) を提案し、この値が最小となるモデルを選択する手法 (最小 AIC 法) を提唱した。彼の提案した AIC は次の式で定義されている。

$$AIC = -2(\text{最大対数尤度}) + 2(\text{自由パラメータ数}) \quad (1.1.1)$$

AIC の考え方を示すために直交多項式の当てはめの例を図 1.1 に取り上げた。グラフ中で \* で示したのが当てはめのためのデータで、次の式で作成したデータに異常値として 4 番目のデータの値を 300000 に変更したものである。

$$y = 30000 + 1900x - 100x^2 + x^3 + e$$

$$e \sim N(0, 18000)$$

グラフの中の曲線は、このデータに対してそれぞれ 2 次、3 次、5 次、7 次の直交多項式を当てはめて得た推定曲線である。次数が高くなるにつれて寄与率も高くなり、7 次の時に最大の値となっている。しかし、図からもわかるように異常値があるために本来の式とはかなり形が違ってしまっている。一方、本来の次数である 3 次式の当てはめにおいて AIC は最小値を示しているが 7 次式では AIC はかなり大きな値となっている。このように AIC によるモデルの選択法は、寄与率などの従来の統計量に基づく方法よりも、はるかにデータ本来の構造に即したモデルを選び出すことが可能なので

ある。現在では AIC は経済分析などの分野においてモデル選択の一般的な手法として広く使用されている。また単純なモデルにおいては種雄牛評価においても AIC を用いたモデル選択を行なうことは可能である (Wada and Kashiwagi(1990))。

さて、フィールドデータを用いて種雄牛や種雌牛の遺伝的能力を評価する場合には、種雄牛効果などの変量効果に血縁関係などの種々の事前分布を仮定するのが普通である。そこでそれらのモデルの変量効果に事前分布を仮定したベイズモデルとして捉えておいた方が、理論的にもすっきりとして今後の発展には有利であると考えられる。赤池 (Akaike(1978,1979,1980)) は AIC を提案した後に AIC とベイズモデルとの関係を考察し赤池のベイズ型情報量規準 (Akaike's Bayesian Information Criterion; ABIC) を導入した。この ABIC は AIC をベイズモデルにすなおに拡張したもので AIC と同様にモデル選択のための指標として利用することができる。そこで次節ではベイズモデルにおける AIC、すなわち ABIC を用いたモデル選択法について考えることにする。

## 1.2 混合モデルにおける赤池のベイズ型情報量規準

まず、次式で示す基本的な線形混合モデルを考える。

$$y = X\beta + Zu + e \quad (1.2.1)$$

ここで、 $y$  は  $n \times 1$  のデータベクトル、 $\beta$  は  $p \times 1$  の母数効果ベクトル、 $u$  は  $q \times 1$  の変量効果ベクトル、 $e$  は  $n \times 1$  の誤差ベクトルである。また、 $X$  は  $n \times p$  の計画行列 (design matrix)、 $Z$  は  $n \times q$  の計画行列である。一般に、 $n$  はデータ数、 $p$  は母数効果ベクトルの大きさ、 $q$  は変量効果ベクトルの大きさである。さらに、

$$E(u) = E(e) = 0,$$

$$\text{Var}(u) = \sigma_u^2 A, \text{Var}(e) = \sigma_e^2 R, \text{Cov}(u, e') = 0$$

を仮定する。

このモデルを基本的な種雄牛評価の場面に当てはめると、 $\beta$  は牛群や年次、季節などの母数効果、 $u$  は種雄牛の効果、 $A$  は種雄牛間の分子血縁係数行列 (numerator relationship matrix, Henderson(1975b))、 $\sigma_u^2$  は種雄牛の分散成分、 $\sigma_e^2$  は残差の分散成分と考えることができる。また、 $R$  は通常に単位行列と考えることが多い。そして、1.2.1 式を  $u$  について解いて得られた値が種雄牛の評価値となり、この値は種雄牛の育種価の推定値の 1/2 である。Henderson(1949,1963) は 1.2.1 式を  $u$  について解くには、次に示す連立 1 次

方程式を解けばよいことを示した。

$$\begin{pmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + \lambda^2 A^{-1} \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{u} \end{pmatrix} = \begin{pmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{pmatrix} \quad (1.2.2)$$

ここで  $\lambda^2 = \sigma_e^2 / \sigma_u^2$  である。

この方法がいわゆる BLUP 法で、数理統計学的には  $\hat{u}$  が最良線形不偏予測量 (best linear unbiased predictor) となることから BLUP 法と名付けられた。また、1.2.2 式は Henderson の混合モデル方程式 (Henderson's mixed model equation) と呼ばれる。

次に BLUP 法をベイズ統計学の立場から見直してみる。BLUP 法とベイズ統計学の関係を最初に指摘したのは Dempfle (1977) であった。その後、Gianola と彼の共同研究者らによってベイズ統計学の立場から BLUP 法をとらえた総説が書かれている (Gianola and Fernando (1986), Gianola et al. (1986))。これら従来のベイズ統計学の基本的な推定手法は以下のようになる。

1. 分析するデータにあったベイズモデル、すなわちデータ分布の確率密度関数とパラメータの事前分布の確率密度関数を作成する。
2. ベイズの定理よりパラメータの事後分布の確率密度関数を求める。
3. パラメータの事後分布の確率密度関数を偏微分して、パラメータの事後分布のモードを求める。
4. パラメータの事後分布のモードの式にデータを代入して、パラメータの推定値を求める。

しかし、いつも問題となるのはデータ分布や事前分布をどのように決めるかである。特に事前分布は今までの種々の経験に基づくものであるだけに主観的に決めることになりやすい。ABC は前節でも述べたように統計モデルのエントロピーという尺度を用いて客観的に事前分布を含めたベイズモデルの善し悪しを評価するものである。すなわち変量効果としての種雄牛効果間の血縁関係の有無や分散比といった事前情報の善し悪しも含めて ABC で比較検討できるわけである。

混合モデルでの ABC を導くために、まず 1.2.1 式をベイズモデルで表現することを考える。データの分布を正規分布と仮定するとそのデータ分布モデルは次のようになる。

$$\begin{aligned} & f(y | \beta, u, \sigma_e^2, R) \\ &= (2\pi)^{-n/2} |R|^{-1/2} \sigma_e^{-n} \exp \left\{ -(y - X\beta - Zu)' R^{-1} (y - X\beta - Zu) / 2\sigma_e^2 \right\} \end{aligned} \quad (1.2.3)$$

次に  $u$  の事前分布は平均 0 ベクトル、分散共分散行列が  $\sigma_u^2 A$  である多変量正規分布であるから  $u$  の事前分布モデルは以下のようになる。

$$f(u | \sigma_u^2, A) = (2\pi)^{-q/2} \sigma_u^{-q} |A|^{-1/2} \exp \frac{-u' A^{-1} u}{2\sigma_u^2} \quad (1.2.4)$$

1.2.3 式と 1.2.4 式が種雄牛評価における基本的なベイズモデルである。次にベイズの定理によって、パラメータ  $u$  の事後分布の確率密度関数を求める。ベイズの定理は次の式で表わすことができる。

$$p(\theta | y) = \frac{f(y | \theta) \cdot g(\theta)}{\int f(y | \theta) \cdot g(\theta) d\theta}$$

ここで、 $f(y | \theta)$  はデータ分布を表わす確率密度関数、 $g(\theta)$  は事前分布を表わす確率密度関数、 $p(\theta | y)$  は事後分布を表わす確率密度関数である。1.2.3 式と 1.2.4 式をこの式に代入すればパラメータ  $u$  の事後分布の確率密度関数は次のように求めることができる。

$$f(u | y, \beta, \sigma_e^2, \lambda^2, R, A) \propto f(y | \beta, u, \sigma_e^2, R) \cdot f(u | \sigma_u^2, A) \quad (1.2.5)$$

ここで、 $\lambda^2 = \frac{\sigma_e^2}{\sigma_u^2}$  である。次に 1.2.5 式を偏微分した式をゼロとおくことによって、事後尤度を最大にする  $u$ 、すなわち  $\hat{u}$  を求める式を導くことができる。

$$\hat{u} = (Z'R^{-1}Z + \lambda^2 A^{-1})^{-1} Z'R^{-1}(y - X\beta) \quad (1.2.6)$$

母数効果の推定量  $\hat{\beta}$  を求めるためにはベイズの定理は使えない。ここでは、Akaike (1980) に従ってベイズ型尤度関数を求めて、それを偏微分した式をゼロとおくことによって、ベイズ型尤度を最大にする  $\beta$ 、すなわち  $\hat{\beta}$  を求める式を導くことができる。ベイズ型尤度関数は次の式で定義されている。

$$f(y | \theta) = f(y | \theta, S_1) \cdot f(\theta | S_2)$$

ここで、 $f(y | \theta, S_1)$  はデータ分布、 $f(\theta | S_2)$  が事前分布で、 $S_1$  と  $S_2$  は未知パラメータである。BLUP 法の場合のベイズ型尤度関数は

$$\begin{aligned} & L(y | \beta, \sigma_e^2, \lambda^2) \\ &= \int f(y | \beta, u, \sigma_e^2, R) \cdot f(u | \sigma_u^2, A) du \\ &= (2\pi)^{-\frac{n+q}{2}} \sigma_e^{-n} \lambda^q |R|^{-\frac{1}{2}} |A|^{-\frac{1}{2}} |Z'R^{-1}Z + \lambda^2 A^{-1}|^{-\frac{1}{2}} \\ & \quad \exp \left[ -\{(y - X\beta - Z\hat{u})' R^{-1} (y - X\beta - Z\hat{u}) + \lambda^2 \hat{u}' A^{-1} \hat{u}\} / 2\sigma_e^2 \right] \end{aligned} \quad (1.2.7)$$

となる。従って、母数効果の最尤推定量  $\hat{\beta}$  は

$$\hat{\beta} = (X'R^{-1}Z)^{-1} [X'R^{-1}(y - Z\hat{u})] \quad (1.2.8)$$

となる。1.2.6式と1.2.8式より、 $\beta$ と $u$ の推定値は、ある与えられた分散比( $\lambda^2 = \sigma_e^2/\sigma_u^2$ )の下で次式を解いて得られることがわかる。

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + \lambda^2 A^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix} \quad (1.2.9)$$

これは1.2.2式と同じものであり、これによってHendersonのBLUP法を完全にベイズ統計学の手法で見直すことができたわけである。さてABICはAICと同様に

$$ABIC = -2(\text{最大対数ベイズ型尤度}) + 2(\text{自由パラメータ数})$$

で与えられる。最大対数ベイズ型尤度は1.2.7式の $\beta$ に最尤推定値 $\hat{\beta}$ を、 $\sigma_e^2$ に最尤推定値 $\hat{\sigma}_e^2$ 、そして $\lambda^2$ に所与の値を代入して求めることができる。自由パラメータ数は1.2.7式におけるパラメータ数であるから $\text{rank}(X) + 2$ である。このようにして各モデルから算出したABICを比較してABICが最小のモデルを選択すればよい。

### 1.3 より複雑なモデルにおける赤池のベイズ型情報量規準

第2節の結果は基本的な線形混合モデルについてのものであり、現在の種雄牛評価で用いられる多くのモデルは1.2.1式のモデルの変形であると見なすことができる。本節では互いの間に共分散を持たない複数の変量効果を含むモデルについてABICの算出方法を示すとともに分散比が未知の場合についても考察する。

#### 1.3.1 2つの変量効果を含むモデル

まず、2つの変量効果 $u_s$ と $u_d$ を含むモデルを次のように決める。

$$y = X\beta + Z_s u_s + Z_d u_d + e \quad (1.3.1)$$

ここでは $u_s$ を種雄牛の変量効果、 $u_d$ を雌牛の変量効果としておく。さらに、

$$\text{Var} \begin{bmatrix} u_s \\ u_d \\ e \end{bmatrix} = \begin{bmatrix} \sigma_{u_s}^2 A_s & 0 & 0 \\ 0 & \sigma_{u_d}^2 A_d & 0 \\ 0 & 0 & \sigma_e^2 R \end{bmatrix} \quad (1.3.2)$$

を仮定する。ここで、 $\sigma_{u_s}^2$ は種雄牛の分散成分、 $\sigma_{u_d}^2$ は雌牛の分散成分、 $A_s$ は種雄牛間の分子血縁係数行列、 $A_d$ は雌牛間の分子血縁係数行列であり、 $\text{Cov}(u_s, u_d) = 0$ を仮定した。第2節と同様に、事後分布の確率密度関数は

$$\begin{aligned} & f(u_s, u_d | y, \beta, \sigma_e^2, \sigma_{u_s}^2, \sigma_{u_d}^2, R, A_s, A_d) \\ & \propto f(y | \beta, u_s, u_d, \sigma_e^2, R) \cdot f(u_s, u_d | \sigma_{u_s}^2, \sigma_{u_d}^2, A_s, A_d) \\ & = (2\pi)^{-\frac{n+q_s+q_d}{2}} \sigma_e^{-n} \sigma_{u_s}^{-q_s} \sigma_{u_d}^{-q_d} |R|^{-1/2} |A_s|^{-1/2} |A_d|^{-1/2} \cdot \\ & \quad \exp \left[ -\{(y - X\beta - Z_s u_s - Z_d u_d)' R^{-1} (y - X\beta - Z_s u_s - Z_d u_d) + \right. \\ & \quad \left. \lambda_s^2 u_s' A_s u_s + \lambda_d^2 u_d' A_d u_d\} / 2\sigma_e^2 \right] \end{aligned} \quad (1.3.3)$$

となる。ここで、 $\lambda_s^2 = \frac{\sigma_e^2}{\sigma_{u_s}^2}$ 、 $\lambda_d^2 = \frac{\sigma_e^2}{\sigma_{u_d}^2}$ である。また、ベイズ型尤度関数は

$$\begin{aligned} & L(y | \beta, \sigma_e^2, \lambda^2) \\ & = \int f(y | \beta, u, \sigma_{u_s}, \sigma_{u_d}, R) \cdot f(u_s | \sigma_{u_s}^2, A_s) \cdot f(u_d | \sigma_{u_d}^2, A_d) du_s du_d \\ & = (2\pi)^{-\frac{n}{2}} \sigma_e^{-n} \lambda_s^{q_s} \lambda_d^{q_d} |A_s|^{-\frac{1}{2}} |A_d|^{-\frac{1}{2}} |b' \Omega^{-1} b|^{-\frac{1}{2}} \cdot \\ & \quad \exp \left[ -\frac{1}{2\sigma_e^2} (a - bu)' \Omega^{-1} (a - bu) \right] \end{aligned} \quad (1.3.4)$$

ここで

$$a = \begin{bmatrix} y - X\beta \\ 0 \\ 0 \end{bmatrix}, b = \begin{bmatrix} Z_s & Z_d \\ I & 0 \\ 0 & I \end{bmatrix}, u = \begin{bmatrix} u_s \\ u_d \end{bmatrix}, \Omega^{-1} = \begin{bmatrix} I & 0 & 0 \\ 0 & \lambda_s^2 A_s^{-1} & 0 \\ 0 & 0 & \lambda_d^2 A_d^{-1} \end{bmatrix}$$

である。

1.3.3式と1.3.4式より $u_s$ と $u_d$ の推定値は次に示す連立方程式を解いて得られることがわかる。

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z_s & X'R^{-1}Z_d \\ Z_s'R^{-1}X & Z_s'R^{-1}Z_s + \lambda_s^2 A_s^{-1} & Z_s'R^{-1}Z_d \\ Z_d'R^{-1}X & Z_d'R^{-1}Z_s & Z_d'R^{-1}Z_d + \lambda_d^2 A_d^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u}_s \\ \hat{u}_d \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z_s'R^{-1}y \\ Z_d'R^{-1}y \end{bmatrix} \quad (1.3.5)$$

ABICは1.3.4式に $\hat{u}_s$ と $\hat{u}_d$ を代入して得られたベイズ型尤度と自由パラメータ数より計算することができる。そして、ABICが最も小さいモデルが、最もデータにマッチしたモデルだということになる。

#### 1.3.2 分散比が未知の場合

BLUP法では分散比( $\lambda^2$ )は既知であるという仮定に基づいていて通常、文献値などが用いられている。しかし、フィールドデータを用いた種雄牛評価では、本質的に分散比はデータより推定すべきものである。現在、大量のフィール

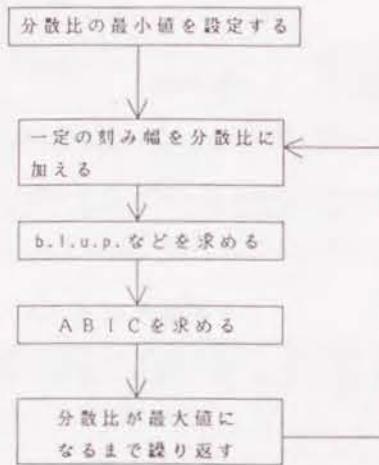


図 1.2: 線形探索法の概略

ドデータから分散成分を推定する手法としては最尤法(Hartley and Rao(1967), Henderson(1973))やREML法(Patterson and Thompson(1971), Schaeffer et al.(1978), Henderson(1984,1985b))といった反復法が使用されている。しかし、反復法は、初期値の値によっては収束しなかったり、真の値とは異なった値に収束することがあるのは良く知られた事実である。また、データによっては意外と収束までに手間取って多大の計算コストを必要とすることがある。そこで、分散成分そのものも事前分布モデルのパラメータであることを考慮して、A B I Cを用いることによって3章で示した方法を分散成分の値が未知の場合にも使用できることを示す。

まず、誤差の分散成分の最尤推定値は $\beta$ を求めた時と同様に1.2.7式を偏微分することによって求めることができる。

$$\sigma_e^2 = \frac{1}{n} \{(y - X\beta - Zu)'R^{-1}(y - X\beta - Zu) + \lambda^2 u' A^{-1} u\} \quad (1.3.6)$$

しかし、種雄牛の分散成分の推定値を直接求めることはできない。そこで、分散比( $\lambda^2$ )について線形探索法(line search method)を実施する。図1.2に線形探索法の概要を示した。すなわち、 $\lambda^2$ を当該形質の遺伝率予想値の周辺の範囲である刻みをもって動かして、そのそれぞれの $\lambda^2$ でのA B I Cを求める。そして、A B I Cが最小の値を示した $\lambda^2$ がデータに最も適した分散比ということになる。刻み幅としては遺伝率換算で0.1ないしは0.05程度が妥当であろう。線形探索法では、ある刻み幅で動かすわけだから正確な推定はできないのではないかという批判があるかもしれない。しかし、この批判は間違っている。なぜなら、われわれの目的はあくまで種畜の遺伝的な能力の評価であり、そのためには遺伝率換算で小数第3位以下まで推定する必要はないからである。

一方、反復法で収束判定するときは、収束の確認のために判定規準を非常に小さくするが、これは分散成分の推定値の有効数字を増やすためではなくて、ほんとうに収束したかどうかを念入りに調べるためである。そのために多くの計算時間を費やすことになる。さらに、収束したからといって局所的な最適値であるかもしれないという疑念をぬぐい去ることは完全にはできない。その点、A B I Cを用いて線形探索する場合は調べた範囲内で最適な分散比であることが保証されるわけである。また実用上、最尤法やREML法に比べて、

1. 係数行列の一般化逆行列を必要とせず、繰り返し数も少なく済むので計算コストが小さい
2. A B I Cを用いたモデル選択法と同じプログラムを使用できる
3. 反復法の欠点である、収束判定の困難さを回避できる

という長所がある。

## 1.4 計算例

理解を助けるために第2節および第3節の方法で簡単な例題を解いてみる。データは表1.1に示した33レコードのものを用いる。

分析には次のモデルを用いることにする。

$$y_{ijkl} = \mu + g_i + u_{ij} + h_k + e_{ijkl} \quad (1.4.1)$$

ここで、 $\mu$ は全平均、 $g_i$ はi番目の種雄牛グループの効果、 $u_{ij}$ はi番目の種雄牛グループ内のj番目の種雄牛の効果、 $h_k$ はk番目の牛群の効果、 $e_{ijkl}$ は誤差である。また、 $Var(u) = \sigma_u^2 I$ 、 $Var(e) = \sigma_e^2 I$ と仮定する。

最初に、 $\sigma_u^2 = 0.01429$ 、 $\sigma_e^2 = 0.1$ (遺伝率で0.5相当)を初期値として選ぶことにする。

この場合の線形混合モデル方程式は次式のようになる。

$$\begin{bmatrix} 330 & 110 & 200 & 50 & 60 & 50 & 110 & 60 \\ 110 & 110 & 90 & 50 & 60 & 0 & 0 & 0 \\ 200 & 90 & 200 & 50 & 40 & 0 & 100 & 10 \\ 50 & 50 & 50 & 119.98 & 0 & 0 & 0 & 0 \\ 60 & 60 & 40 & 0 & 129.98 & 0 & 0 & 0 \\ 50 & 0 & 0 & 0 & 0 & 119.98 & 0 & 0 \\ 110 & 0 & 100 & 0 & 0 & 0 & 179.98 & 0 \\ 60 & 0 & 10 & 0 & 0 & 0 & 0 & 129.98 \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{h}_1 \\ \hat{g}_1 \\ \hat{s}_{11} \\ \hat{s}_{12} \\ \hat{s}_{23} \\ \hat{s}_{24} \\ \hat{s}_{25} \end{bmatrix} = \begin{bmatrix} 322 \\ 121 \\ 209 \\ 67 \\ 54 \\ 42 \\ 106 \\ 53 \end{bmatrix} \quad (1.4.2)$$

表 1.1: 新しい種雄牛評価法のためのテストデータ

種雄牛番号	種雄牛グループ番号	牛群番号	データ
1	1	1	1.6
1	1	1	1.2
1	1	1	1.3
1	1	1	1.5
1	1	1	1.1
2	1	1	0.8
2	1	1	1.0
2	1	1	1.2
2	1	1	1.0
2	1	2	0.8
2	1	2	0.6
3	2	2	0.8
3	2	2	0.8
3	2	2	0.9
3	2	2	0.7
3	2	2	1.0
4	2	1	0.5
4	2	1	1.2
4	2	1	0.6
4	2	1	1.1
4	2	1	0.6
4	2	1	1.3
4	2	1	1.1
4	2	1	0.9
4	2	1	0.7
4	2	1	1.2
4	2	2	1.4
5	2	1	1.0
5	2	2	0.8
5	2	2	1.1
5	2	2	0.6
5	2	2	1.0
5	2	2	0.8

表 1.2: モデル別の A B I C

モデル	A B I C
$\mu + g + u + h$	12.50
$\mu + u + h$	12.05
$\mu + g + u$	11.45
$\mu + u$	12.09

表 1.3: 遺伝率を変化させた場合の A B I C

遺伝率	A B I C
0.1	12.01
0.2	11.76
0.3	11.59
0.4	11.49
0.5	11.45
0.6	11.45
0.7	11.48
0.8	11.55
0.9	11.64

1.4.2 式を解くと、

$$\hat{\mu} = 0.8595, \hat{h}_1 = 0.1625, \hat{g}_1 = 1.0522$$

が得られる。種雄牛評価値 ( $\hat{g} + \hat{u}$ ) は、

$$\hat{g}_1 + \hat{u}_{11} = 0.2512, \hat{g}_1 + \hat{u}_{12} = 0.07381,$$

$$\hat{g}_2 + \hat{u}_{23} = 0.008116, \hat{g}_2 + \hat{u}_{24} = 0.005198, \hat{g}_2 + \hat{u}_{25} = 0.002918$$

となる。

1.3.6 式より求めた誤差の分散成分の推定値は  $\sigma_e^2 = 0.05721$ 、ベイズモデルの対数尤度は  $L = -1.2474$ 、A B I C は  $ABIC = 12.50$  と求められる。

さてモデル選択の実例として 1.4.1 式のモデルから 1 ないし 2 つの母数効果を除いたモデルで A B I C を求めてみると表 1.2 のような結果となる。

種雄牛グループの効果のみを取り入れたモデル ( $\mu + g + u$ ) における A B I C が最小であるので、このモデルが最もこのデータにマッチしているということになる。このように、要因をたくさん加えた方がいいモデルであるとは限らないことが A B I C を計算してみるとよくわかる。

次に、 $\mu + g + u$  のモデルで推定した  $\sigma_e^2$  をもとに、遺伝率 0.1 から 0.9 に対応した  $\lambda^2$  を求め、各々の分散比ごとに BLUP 計算を実施して ABIC を求めるとその結果は表 1.3 のようになる。この結果によると、この例題データでは遺伝率 0.5 と 0.6 で ABIC が最小となっており、先の計算で設定した初期値も分散比についてはおおむね妥当と言える。また、遺伝率 0.5 付近で ABIC の値がほとんど近い値になっていることは、このような少数のデータで分散比を推定することの困難さを示唆している。

## 摘 要

現在、一般的に種雄牛評価に用いられている BLUP 法にも 2-3 の問題がある。BLUP 法はデータに合わせて種々のモデルを選ぶことができるのだが、モデルを選ぶための科学的な方法論が存在しないことが、最も大きな問題点である。そこで、BLUP 法を情報量統計学によって見直すことによって BLUP 法の欠点を克服する新たな手法を開発した。

本章で述べた方法を整理すると次のようになる。

1. モデルをデータ分布と変量効果の事前分布からなるベイズモデルに書き直す。
2. ベイズの定理に基づいて、変量効果の事後分布の確率密度関数を求める。
3. 事後分布の確率密度関数を偏微分して変量効果の推定値を求める式を導く。
4. ベイズモデルのベイズ型尤度関数を求める。
5. ベイズ型尤度関数を偏微分して母数効果と誤差分散の推定値を求める式を導く。
6. 実際のデータを用いて、ベイズ型尤度関数とパラメータの推定値から ABIC を求める。

1 から 6 までの作業を可能性のあるモデルについて実施する。ABIC が最小であったモデルが最良のモデルであり、そのときのパラメータ推定値より種雄牛の評価値を求める。

上記の方法は、ほとんどすべての種雄牛評価モデルで実施することができる。例として基本的な混合線形モデルと複数の変量効果を持つモデルについて、事後分布の確率密度関数やベイズ型尤度関数、パラメータの推定のための方程式などを示した。分散成分の値が未知の場合には、今までは最尤法や REML 法といった反復法によって分散成分を推定し、その値をもとに BLUP 計算をするのが普通であった。しかし、反復法は収束判定の困難さや計

算コストの面で大きな問題がある。そこで、本研究では分散比を与えられた範囲で動かして、各々の分散比で ABIC を計算し、ABIC の値によって最良の分散比を求める方法を提案した。この方法の精度は種雄牛評価には十分であり、計算コストは少なく済む。また、収束判定に悩むこともない。

さらに、簡単なテストデータを用いて、本研究で開発した手法の種雄牛モデルでの計算例を示した。

## 第2章

# シミュレーションによる有効性の検討

### 緒言

第1章でも述べたように、フィールドデータを用いた種雄牛評価に広く用いられているBLUP法(Henderson(1949,1963,1973))では、モデルを科学的に決定するための手法が用意されていない。筆者らは、第1章で詳しく述べたように、BLUP法をベイズモデルとして見直して、情報量統計学の理論(Akaike(1980))に基づくABIC(Akaike's Bayesian Information Criterion)を用いたモデル決定のための方法と、そのいくつかの応用例を明らかにした。

情報量統計学の理論は、統計モデルのエントロピーに基づくもので、数理統計学における有用性は明らかになっており、経済学(Akaike(1980), Yanagimoto and Yanagimoto(1987))、天文学(Ishiguro and Ishiguro(1979))、地球物理学(田辺と田中(1983))、実験計画(柏木(1982))などの種々の分野の問題に適用されつつある。しかし、種雄牛評価の問題はこれらの問題とはいくつかの面で違う側面がある。最も異なる点は、これらの手法は従来はパラメータの事前分布を考慮せずに分析していたのであるが、種雄牛評価では古くから積極的に、遺伝率などの形で種雄牛効果のパラメータの事前分布を、事前分布として意識していなかったにせよ、扱ってきた点である。そのため、情報量統計学の理論が家畜育種の理論と整合性を持つかどうか問題となる。

そこで、本章では、日本におけるホルスタイン種のフィールドデータを前提としたコンピュータシミュレーションを用いて、第1章で筆者らが示した方法、特にABICを用いたモデル選択の方法について、その有効性を検討する。

表 2.1: シミュレーションデータ発生時に用いた母数効果の値

効果	水準	母数効果の値
全平均	$\mu$	8000
種雄牛グループ	$g$	1 1000
		2 -1000
		3 0
牛群	$h$	1 1000
		2 800
		3 600
		4 400
		5 200
		6 0
		7 -200
		8 -400
	9 -600	
	10 -800	
	11 -1000	

## 2.1 方法

### 2.1.1 シミュレーション用データの発生方法

シミュレーションに用いるデータは、日本におけるホルスタイン種の初産、305日補正乳量のフィールドデータを前提として、乱数をもちいてコンピュータにより発生させた。データを発生させるために用いたモデルは、次のようなものである。

$$y_{ijkl} = \mu + h_i + g_j + u_{jk} + e_{ijkl} \quad (2.1.1)$$

ここで、 $\mu$ は全平均、 $h_i$ は*i*番目の牛群の効果、 $g_j$ は*j*番目の種雄牛グループの効果、 $u_{jk}$ は*j*番目の種雄牛グループ内の*k*番目の種雄牛の効果、 $e_{ijkl}$ は残差で、 $y_{ijkl}$ はシミュレーションによって発生させたデータである。 $\mu$ と $h$ と $g$ は母数効果として取り扱い、 $h$ は11水準、 $g$ は3水準とした。また $\mu$ および $h$ と $g$ の各水準の効果の値は表2.1に示した値を使用した。これらの値は、日本におけるホルスタイン種のフィールドデータを念頭において作成したものであるが、計算量を減らすために水準数は実際のデータよりも減らした。

また、種雄牛グループ内の種雄牛頭数は、それぞれ60頭、30頭、120頭で、全体の種雄牛頭数は210頭である。種雄牛の効果と残差は変量効果とし

表 2.2: 分析に用いたモデル

狭義のモデル	分散比
真のモデル	真の分散比
真のモデル	遺伝率で 0.1 から 0.9 までの 9 種
$\mu + u$	真の分散比
$\mu + g + u$	真の分散比
$\mu + h + u$	真の分散比

で扱った。u および e の分散共分散行列は次の式のように仮定した。

$$\text{Var}(u) = \sigma_u^2 A, \quad \text{Var}(e) = \sigma_e^2 I$$

ここで、 $\sigma_u^2$  は種雄牛効果の分散成分、 $\sigma_e^2$  は残差の分散成分、A は種雄牛間の分子血縁係数行列 (numerator relationship matrix; A 行列)、I は単位行列である。本研究では  $\sigma_u^2 = 500000$ 、 $\sigma_e^2 = 6500000$  を仮定した。これは、遺伝率に換算すると  $h^2 = 0.286$  となる。A 行列は次の仮定のもとでシミュレーションを実施する前に乱数を発生させて算出した。

1. 種雄牛の近交係数は平均 0.05 の指数分布に従う。
2. 異なる種雄牛グループに属する種雄牛間の血縁係数は平均 0.005 の指数分布に従う。
3. 同一の種雄牛グループに属する種雄牛間の血縁係数は平均 0.02 の指数分布に従う。

さらに、牛群と種雄牛のクロス分布は、ある程度の差重供用 (differential use) が実施されていると仮定した。そこで、各レコードについて与えられた種雄牛番号 ISIRE をもとに次の式により正規乱数を発生させて、牛群番号 KHYS を決定した。

$$KHYS \propto N(EX2, SD2)$$

ここで ISIRE: グループを込みにした種雄牛の順序番号、KHYS: 決定した牛群番号、 $EX2 = ISIRE/3 + 1$ 、 $SD2 = 3.5$  である。

1 頭の種雄牛あたりのレコード数は、 $\alpha = 0$ 、 $k = 25$  のガンマ乱数に従うとした。ただし、乱数を発生させた結果、20 未満になった場合はもう一度、乱数を発生して、1 種雄牛あたり 20 以上のデータを発生させるようにした。シミュレーションデータの発生手順の概要は図 2.1 にも示した。



図 2.1: シミュレーションデータの発生手順の概略

## 2.1.2 分析モデル

シミュレーションの各回において表 2.2 に示す 4 種の異なる母数効果を取り上げたモデルと真のモデルで分散比を変更した 10 種のモデルを用いて分析した。

## 2.1.3 算出した統計量

各シミュレーションデータに対して、表 2.2 に示した 14 種のモデルによって種雄牛効果の最良線形不偏予測量 (BLUP;  $\hat{u}$ ) と各母数効果の最良線形不偏推定量 (BLUE;  $\hat{\beta}$ ) を算出した。各種雄牛の予測誤差分散 ( $\text{Var}(u - \hat{u})$ ) と BLUE の分散 ( $\text{Var}(\hat{u})$ ) とリピータビリティ、各母数効果の BLUE の分散 ( $\text{Var}(\hat{\beta})$ ) も同時に計算した。さらに、モデル別に誤差分散の最尤推定量と ABC を算出した。これらの計算には ABBLUP (和田 (1989)) を用いた。また、比較のために真のモデルで Henderson の方法 3 (Henderson (1953)) によって分析した種雄牛および誤差の分散成分の推定値も算出した。これらの計算には LSML76 (Harvey (1976)) を用いた。

一部、第 1 章の繰り返しになるが、算出した統計量の計算式は次のようになる。

### a) 基本モデル

$$y = X\beta + Zu + e \quad (2.1.4)$$

ここで、y は  $n \times 1$  のデータベクトル、 $\beta$  は  $p \times 1$  の母数効果ベクトル、u は  $q \times 1$  の変量効果ベクトル、e は  $n \times 1$  の誤差ベクトルである。また、X は  $n \times p$  の計画行列 (design matrix)、Z は  $n \times q$  の計画行列である。一般に、n はデー

タ数、 $p$  は母数効果ベクトルの大きさ、 $q$  は変数効果ベクトルの大きさである。さらに、ここでは

$$E(u) = E(e) = 0, \text{Var}(u) = \sigma_u^2 A, \text{Var}(e) = \sigma_e^2 I, \text{Cov}(u, e') = 0 \quad (2.1.5)$$

を仮定している。

b) 種雄牛評価値

次の連立1次方程式を解いて得られる $\hat{u}$ と $\hat{\beta}$ から $\hat{g} + \hat{u}$ (種雄牛グループを含むモデル)、 $\hat{u}$ (種雄牛グループを含まないモデル)として求めた。

$$\begin{pmatrix} X'X & X'Z \\ Z'X & Z'Z + \lambda^2 A^{-1} \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{u} \end{pmatrix} = \begin{pmatrix} X'y \\ Z'y \end{pmatrix} \quad (2.1.6)$$

ここで、 $\lambda^2 = \frac{\sigma_e^2}{\sigma_u^2}$ である。

c) 誤差分散の最尤推定値は次の式で求めた。

$$\sigma_e^2 = \{(y - X\hat{\beta} - Z\hat{u})'(y - X\hat{\beta} - Z\hat{u}) + \lambda^2 \hat{u}'A^{-1}\hat{u}\} / n \quad (2.1.7)$$

d) ABIC

最大対数尤度をLとすると

$$L = (2\pi)^{-n/2} \sigma_e^{-n} \lambda^q \exp \left[ -\{(y - X\hat{\beta} - Z\hat{u})'(y - X\hat{\beta} - Z\hat{u}) + \lambda^2 \hat{u}'A^{-1}\hat{u}\} / 2\sigma_e^2 \right] \quad (2.1.8)$$

そしてABICは

$$ABIC = -2L + 2(\text{自由パラメータ数}) \quad (2.1.9)$$

で求めた。

e) 予測誤差分散  $\text{Var}(u - \hat{u})$

2.1.6式の係数行列の逆行列をCとして、

$$C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \quad (2.1.10)$$

とすると、予測誤差分散は  $\text{Var}(u - \hat{u}) = \sigma_e^2 C_{22}$ となる。

f) 変数効果のb.l.u.p.の分散は  $\text{Var}(\hat{u}) = \sigma_u^2 A - \sigma_e^2 C_{22}$ として求めた。

g) 母数効果のb.l.u.e.の分散は  $\text{Var}(\hat{\beta}) = \sigma_e^2 C_{11}$ として求めた。

h) リピータビリティ  $r^2$

本来のリピータビリティとは異なるものであるが、種雄牛評価では歴史的に用いられている次の式(Ufford et al.(1979))で計算した。

$$r^2 = \left( \frac{\text{Cov}(\hat{u}', u)}{\text{Var}(\hat{u}) \cdot \text{Var}(u)} \right)^2 = (Z'Z + \lambda^2 A^{-1})^{-1} Z'Z \quad (2.1.12)$$

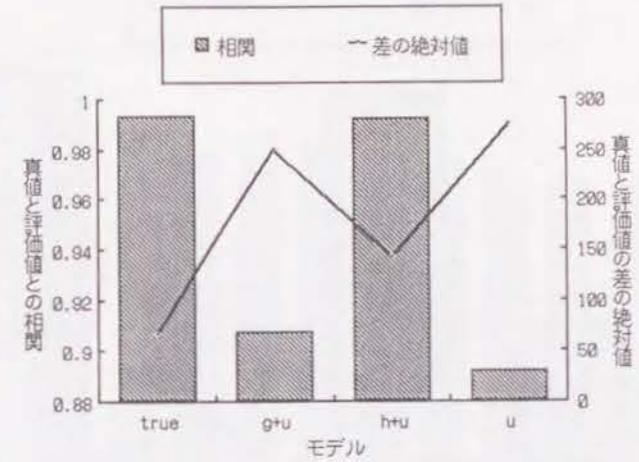


図 2.2: 真の分散比における真値と評価値との相関

## 2.2 シミュレーションの結果

図 2.2 から図 2.7 に分散比は真の値で狭義のモデルのみを変えた場合の結果を示した。

図 2.2 にシミュレーションデータ発生時に設定した真の種雄牛の遺伝的能力と、BLUP法で算出したその評価値との相関と、両者の差の絶対値を示した。これらの値は、評価の正確さを示す指標と考えられる。どちらの結果も最もよい値だったのは真のモデル(true)で、次に牛群効果を加えたモデル、種雄牛グループを加えたモデル、どちらも加えなかったモデルの順に評価の正確さが低下することが示された。これは、表 2.1 に示したシミュレーションデータの発生に用いた牛群効果の値が種雄牛グループの値よりもばらついているためであろう。

図 2.3 に ABIC の値を示した。図 2.2 と比較すればわかるように、よりよく評価できたモデルほど ABIC の値が小さくなっている。ABIC が評価の正確さのモデル別の指標となりうることを示されている。

図 2.4 には予測誤差分散の平均値を示した。一見してわかるように、種雄牛グループを含めたモデルで大きな予測誤差分散を示し、含まないモデルでは小さな予測誤差分散が求められた。

次に狭義のモデルはすべての効果を含んでいて、分散比のみ変化させた結果を示す。図 2.5 にシミュレーションデータ発生時に設定した真の種雄牛の遺伝的能力と、BLUP法で算出したその評価値との相関と、両者の差の絶対値を示した。真の分散比に最も近い遺伝率 0.3 のモデルで、両者とも最もよい値を示した。そして、遺伝率が 0.3 から離れるほど評価の正確さは悪くなったが、左右対称ではなく、遺伝率が小さくなる方では急激に評価の正確

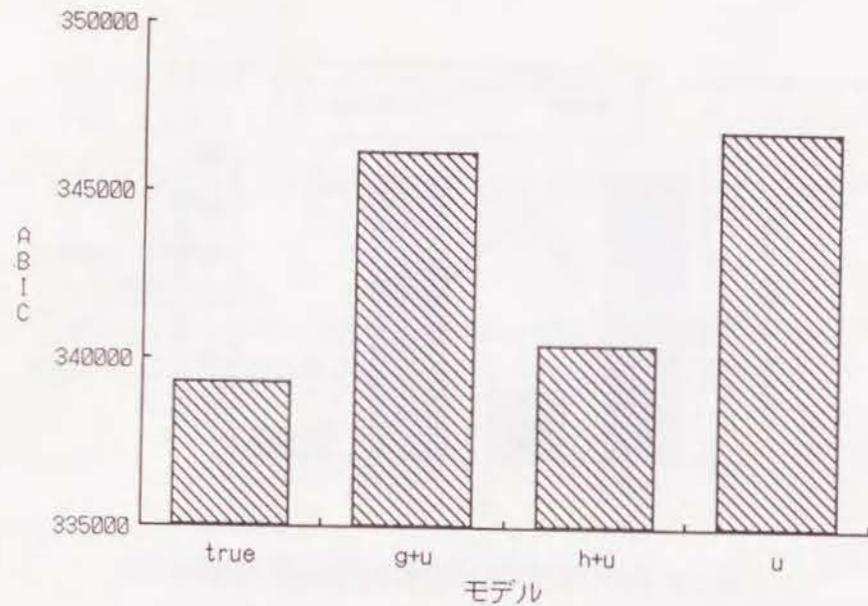


図 2.3: 真の分散比におけるモデル別の A B I C

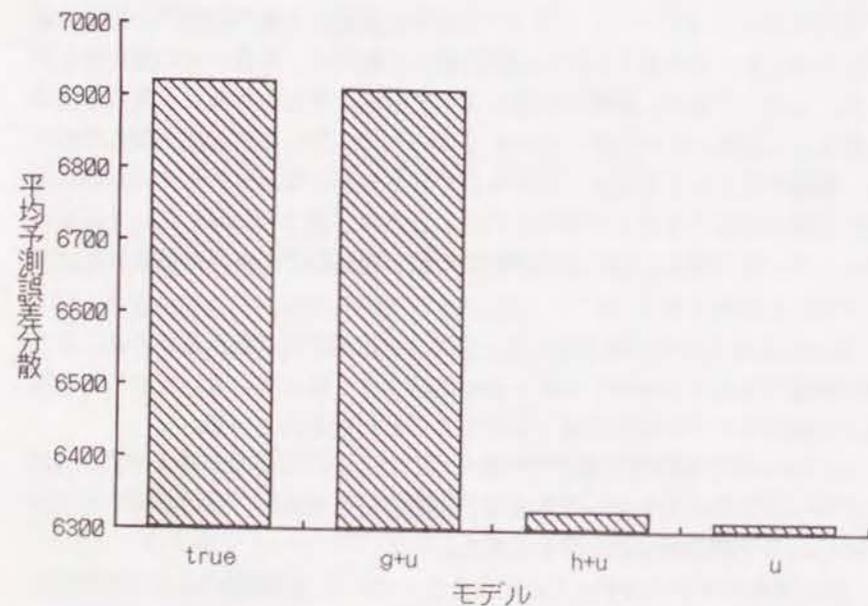


図 2.4: 真の分散比におけるモデル別の平均予測誤差分散

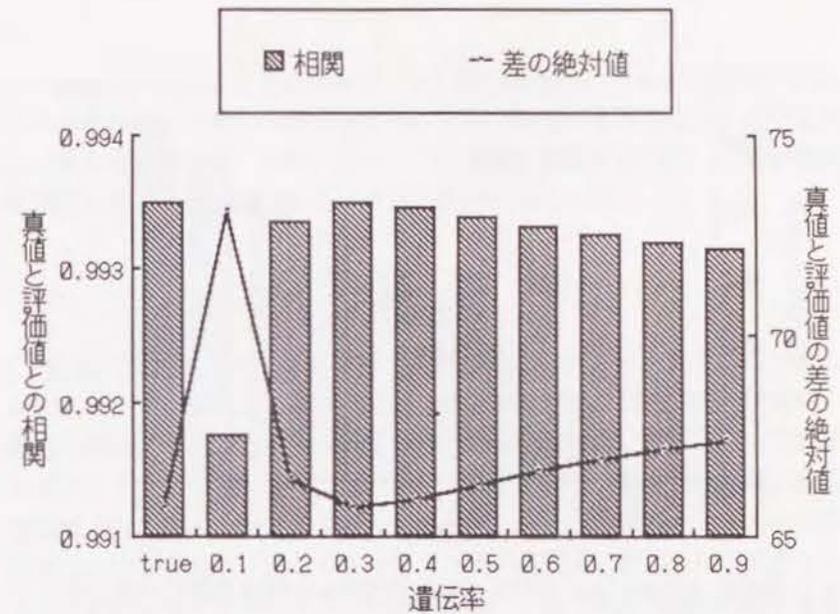


図 2.5: 狭義の真のモデルにおける分散比の変化が真値と評価値との相関および真値と評価値の差の絶対値におよぼす影響

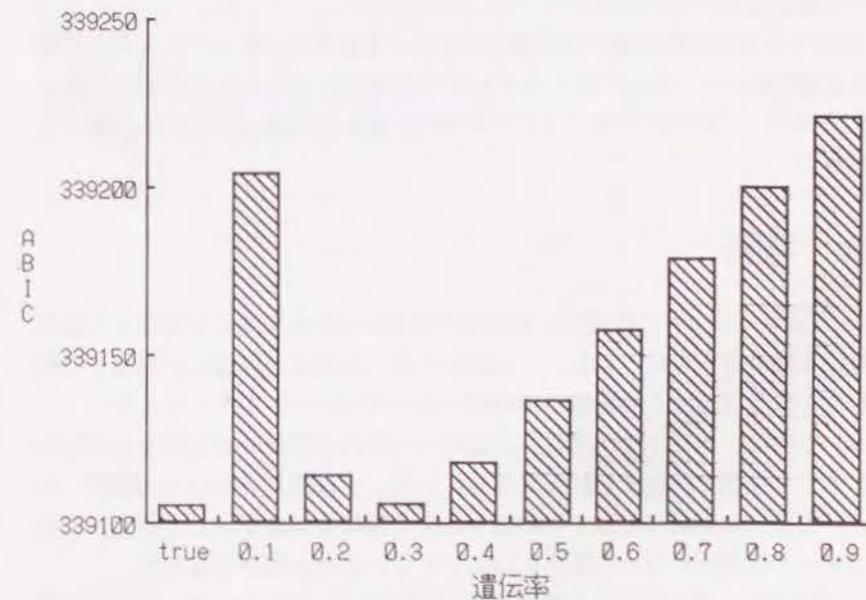


図 2.6: 狭義の真のモデルにおける分散比の変化が A B I C におよぼす影響

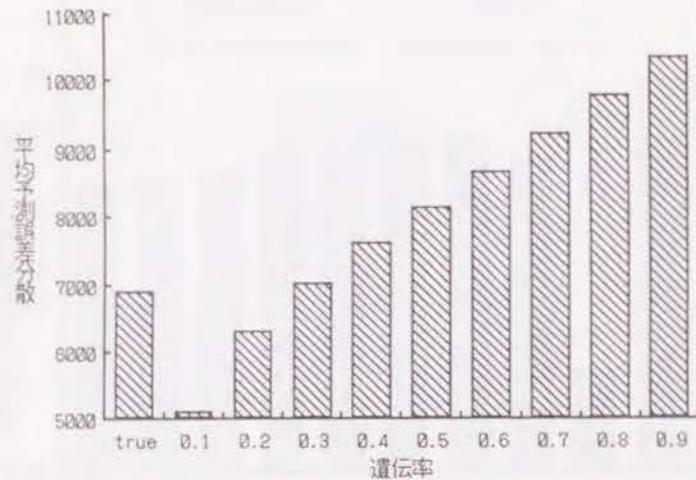


図 2.7: 狭義の真のモデルにおける分散比の変化が平均予測誤差分散におよぼす影響

さは悪くなった。

図 2.6 には A B I C の値を示した。この場合も遺伝率 0.3 で A B I C は最も小さい値を示した。そして遺伝率が 0.3 に近づくほど A B I C が小さくなる傾向が認められた。

図 2.7 に予測誤差分散の平均値を示した。予測誤差分散は A B I C とは異なり、遺伝率 0.1 で最小の値を示し、遺伝率が大きくなるほど単調に増加した。これは、予測誤差分散では分散比の良し悪しが決められないことを示している。

## 2.3 考察

シミュレーションの結果は、母数効果の組み合わせを変えた場合も分散比を変えた場合もともに、A B I C の値が小さくなるほど真値と評価値との相関が高くなり、真値と評価値の差の絶対値の平均が小さくなることを示している。これは、A B I C が種雄牛評価モデルの決定のための指標として有効であることを強く示唆する結果である。一方、予測誤差分散の平均値は、真値と評価値との相関や A B I C の値とは全く異なる動きを示しており、この値はモデルの選択のための基準にはなりえないことが明かとなった。

本論文の第 1 章で述べた方法は最尤法をもとにしている。一方、緒論でも述べたように家畜育種においては分散成分の推定法として REML 法 (restricted maximum likelihood method) がよく用いられている。Lin ら (1989) は数種の手法をシミュレーションデータに適用して反復 REML 法が最もすぐれた分散成分の推定法であると結論づけている。しかし、今回のシミュレーショ

ンの結果をみると A B I C をもちいて分散比を線形探索する方法でも実用上、かなりの精度で分散比を推定できることがわかる。もちろんベイズ型尤度関数の代わりに REML で用いられている制限尤度関数を使用して情報量基準を構築することは可能であり、今後、検討してみたい。

## 摘要

日本におけるホルスタイン種の乳量のフィールドデータを想定したコンピュータシミュレーションを実施し、第 1 章で述べた情報量統計学を用いた種雄牛評価のためのモデル選択法の有効性を検討した。乱数を用いて発生させたデータを 14 種のモデルで分析し、種雄牛評価を実施した結果、分散比を真値として狭義のモデルを変えた場合も、ひとつの狭義のモデルで分散比を変化させた場合も、真のモデルおよび真の分散比に近づくほど算出された A B I C の値は小さくなった。一方、従来、一部でモデル選択に用いられてきた予測誤差分散の平均値は種雄牛グループの効果を含んだモデルでむしろ大きな値を示し、分散比から換算した遺伝率が高くなるほど単調に増加した。これらの結果より、Henderson の示した混合モデルを用いた種雄牛評価モデルの決定に当たっては、予測誤差分散ではなく A B I C を用いるべきことが明らかとなった。また、分散成分を反復法によって求める「反復 B L U P 法」においても、A B I C は収束状態をチェックする指標として利用できることが示唆された。さらに、A B I C を用いて分散比を線形探索することにより、分散比が未知の場合の種雄牛評価が「反復 B L U P 法」より少ない計算コストで実施できる可能性が示された。

# 第3章 大規模な種雄牛評価のための 計算手法の開発

## 緒言

本論文の第1章では情報量統計学にもとづく種畜評価モデルの選択法について詳細に検討し、第2章ではコンピュータシミュレーション手法を用いて第1章で提案した手法の有効性を示した。しかしながら、家畜育種のような応用科学においては実際のデータで使用できてはじめて意味のある手法とすることができる。日本全国を対象とするような大規模な種畜評価の場面においては、通常の種畜評価ですら膨大な計算量になる。情報量統計学にもとづくABCを算出すると、アルゴリズムを工夫しなければその計算量は幾何級数的に増大し、スーパーコンピュータですら計算不能ということになりかねない。

そこで、本章では全国規模の種畜評価モデルに適したABCの計算の中で最大のネックとなる係数行列の変量効果部分の小行列の行列式の算出について、種々のアルゴリズムに基づいて作成したプログラムで実際のデータを分析して計算に費やした時間やメモリ量を測定し、それらのアルゴリズムの優劣を比較検討した。

### 3.1 材料および方法

テストデータとしては第2章で実施したシミュレーションにおける混合モデル方程式の係数行列(A行列; 210x210)と第2章よりも種雄牛数を多くした時の係数行列(B行列; 480x480)を用いた。この2つの行列をLL'分解またはLDL'分解して行列式の値を求めるために要したCPU時間とターンアラウンド時間(計算結果が実際に返ってくるまでに要する時間)を計測した。比較したアルゴリズムは、以前から使用されているガウスの消去法と

表 3.1: ACOS930/20 と EWS4800/50 の性能緒元

	ACOS930/20	EWS4800/50
CPU性能(MIPS)	23	3.2
主記憶容量(MB)	64	16
ディスクアクセス速度(KB/秒)	3000	1250
ディスク容量(MB)	50000	300

表 3.2: 使用したコンパイラとコンパイルオプション

	ACOS930/20	EWS4800/50
コンパイラ	FRT77	optf77
コンパイルオプション	OPT=3,IAP, WARN, V,LSTIN,NOMAP,MASTER,JIPS	-e0 -ffpax -O2 -f3

コレスキー法、スーパーコンピュータ用のアルゴリズムである縦ブロックガウス法、疎行列用のアルゴリズムであるスカイライン法、ウェーブフロント法、疎行列用内積形式改訂コレスキー法の6種類である。計算プログラムには磯田・大野(1971)と村田ら(1985)に掲載されていたサブルーチンを移植したものをを用いた。なお縦ブロックガウス法についてはスーパーコンピュータ以外の機種でも動作可能にするために行列の各要素をランダムファイル上に配置し、1ブロック分のみを主記憶上に格納するように変更した。

使用したマシンは農林水産研究計算センターのACOS-930/20と畜産試験場に配備されているEWS4800/50である。ACOS-930/20は汎用大型機、EWS4800/50はUNIXをOSとするワークステーションである。各々の性能諸元(メーカー公表値)を表3.1に示した。ACOS-930/20はEWS4800/50に対してCPU性能で7倍、主記憶容量で4倍、ディスクアクセス速度で2.4倍の性能である。両マシンともにテストの前に予めテストプログラムをコンパイルし、テスト時には実行形式のプログラムを直接実行して計測した。コンパイル時においてはできる限りの最適化をするように指定した(表3.2)。

ACOS-930/20でのテスト時はほぼ常時20台のTSS端末、15台のリモートバッチ端末が接続されており、5本のバッチジョブがバックグラウンドで実行されている状態であった。EWS4800/50でのテスト時は実行時のユーザー数0-2人で、ネットワーク関係の常駐プログラムが数本実行されている状態であった。

### 3.2 結果

各アルゴリズムにおける主記憶使用量を表 3.3 に示した。行列 A ではウェーブフロント法が最も主記憶使用量が少なかったが、行列 B では縦ブロックガウス法の主記憶使用量が最も少なかった。一般式から明らかなように縦ブロックガウス法では NBL の値を小さくすれば主記憶使用量をもっと減らすことも可能であり、行列の次数が大きくなればなるほど主記憶使用量における縦ブロックガウス法の有利性が增大する。一方、ガウスの消去法とコレスキー法は係数行列の全要素をメモリ上に保存しているために主記憶使用量は他のアルゴリズムに較べてかなり大きかった。

表 3.4 に行列 A (210x210) の行列式の値を ACOS-930/20 で求めたときの結果を示した。縦ブロックガウス法は CPU 時間、ターンアラウンド時間ともに他のアルゴリズムと比較してかなり大きな値を示した。一方、スカイライン法は CPU 時間、ターンアラウンド時間ともに最小の値であった。

表 3.5 に行列 B (480x480) の行列式の値を ACOS-930/20 で求めたときの結果を示した。スカイライン法とともにコレスキー法も CPU 時間、ターンアラウンド時間ともに優れた結果であった。一方、縦ブロックガウス法は CPU 時間ではウェーブフロント法よりも速かったがターンアラウンド時間では平均でも 2 時間半以上かかった。

表 3.6 に行列 A の行列式の値を EWS4800/50 で求めた結果を示した。ACOS-930/20 での結果と同様にスカイライン法が最も速く、縦ブロックガウス法が最も遅いアルゴリズムであった。

表 3.7 に行列 B の行列式の値を EWS4800/50 で求めた結果を示した。この場合も ACOS-930/20 での結果と同様にスカイライン法とコレスキー法が速く、縦ブロックガウス法とウェーブフロント法が遅いという結果であった。ただし縦ブロックガウス法のターンアラウンド時間は ACOS-930/20 での結果よりも EWS4800/50 での結果の方が 3.5 倍以上小さい値であった。

### 3.3 考察

本章で測定した範囲においてはスカイライン法が主記憶使用量、CPU 時間、ターンアラウンド時間のすべてにおいて最も優れたアルゴリズムであることが明らかとなった。しかし今回テストした行列は 210x210 および 480x480 という比較的次数の小さな行列であった。わが国の乳牛における種雄牛評価に当てはめると種雄牛モデルや母方祖父モデルで 3000 ないし 5000、個体モデル (Animal Model) においては数十万という大きな次数を持つ行列の行列式を求めなければならない。これらの場合には縦ブロックガウス法以外のアルゴリズムでは主記憶容量が不足して計算不可能となることは明らかであろう。

表 3.3: 主記憶使用量 (語)

アルゴリズム	一般式	行列 A	行列 B
ガウスの消去法	$2N^2 + N$	88410	461280
コレスキー法	$2N^2 + 4N$	89040	462720
縦ブロックガウス法	$2NBL \times N + 7N$	11970	51360
スカイライン法	$2NSL + 6N$	20170	233698
疎行列用改訂コレスキー法	$8NET + 6N$	74652	116776
ウェーブフロント法	$NWV + 4N$	8100	117360

ここで N は行列の大きさ、NBL はブロック長、NSL はスカイライン法用の配列の大きさ、NET は非零要素の数、NWV はウェーブフロントの表の大きさである。

行列 A、B ともに NBL を 50 として計算した。N、NSL、NET、NWV は行列によって一意に決まる値である。

表 3.4: 行列 A の行列式の値を求める時に要した計算時間 (ACOS-930/20)

アルゴリズム	CPU 時間 (秒)		ターンアラウンド (秒)	
	平均	S.D.	平均	S.D.
ガウスの消去法	3.61	± 0.12	27.8	± 18.5
コレスキー法	0.22	± 0.00	1.4	± 0.7
縦ブロックガウス法	17.28	± 0.76	2013.9	± 227.2
スカイライン法	0.06	± 0.00	0.3	± 0.5
疎行列用改訂コレスキー法	0.57	± 0.00	1.0	± 0.0
ウェーブフロント法	5.84	± 0.04	14.9	± 0.8

表 3.5: 行列 B の行列式の値を求める時に要した計算時間 (ACOS-930/20)

アルゴリズム	CPU 時間 (秒)		ターンアラウンド (秒)	
	平均	S.D.	平均	S.D.
ガウスの消去法	22.69	± 0.17	49.7	± 50.3
コレスキー法	2.92	± 0.05	6.4	± 6.3
縦ブロックガウス法	113.84	± 5.81	9648.6	± 589.3
スカイライン法	2.92	± 0.03	6.1	± 3.4
疎行列用改訂コレスキー法	28.10	± 0.08	48.3	± 16.9
ウェーブフロント法	250.56	± 0.56	500.3	± 221.8

表 3.6: 行列Aの行列式の値を求める時に要した計算時間 (EWS4800/50)

アルゴリズム	CPU時間 (秒)		ターンアラウンド (秒)	
	平均 ± S.D.		平均 ± S.D.	
ガウスの消去法	88.3 ± 0.5		95.7 ± 11.0	
コレスキー法	6.9 ± 1.3		7.3 ± 1.3	
縦ブロックガウス法	104.9 ± 1.2		360.5 ± 30.7	
スカイライン法	1.7 ± 0.02		2.0 ± 0.00	
疎行列用改訂コレスキー法	4.7 ± 0.1		5.9 ± 1.3	
ウェーブフロント法	30.8 ± 0.5		66.9 ± 5.7	

表 3.7: 行列Bの行列式の値を求める時に要した計算時間 (EWS4800/50)

アルゴリズム	CPU時間 (秒)		ターンアラウンド (秒)	
	平均 ± S.D.		平均 ± S.D.	
ガウスの消去法	693.7 ± 32.8		718.6 ± 34.3	
コレスキー法	74.9 ± 8.7		78.1 ± 10.8	
縦ブロックガウス法	1009.0 ± 10.9		2641.2 ± 229.0	
スカイライン法	78.5 ± 0.6		94.0 ± 23.7	
疎行列用改訂コレスキー法	211.6 ± 0.6		225.1 ± 20.8	
ウェーブフロント法	1421.3 ± 2.2		1871.3 ± 28.1	

表 3.8: 牛群検定データを母方祖父モデルで分析した時の行列式の計算に要した計算時間

機種	CPU時間 (秒)	ターンアラウンド (秒)
HITAC S820/80	66.36	411.64
NEC ACOS-930/20	...	...
NEC EWS4800/50	12411.3	38160.0

HITAC は半導体ディスク装置を用いて計算した。HITAC は 2869x2869 の行列を計算した時の値。

ACOS と EWS4800 は 2386x2386 の行列を計算した時の値。ACOS では 10 日間で計算が終了しなかったためアボートした。

用いたモデルは  $y_{ijklmn} = s_i + \frac{1}{2}mgs_j + HY_k + A_l + M_m + e_{ijklmn}$  であった。ここで、 $y_{ijklmn}$ : 観測値、 $s_i$ : i 番目の種雄牛の効果、 $mgs_j$ : j 番目の母方祖父の効果、 $HY_k$ : k 番目の牛群一年次の効果、 $A_l$ : l 番目の雌牛月齢の効果、 $M_m$ : m 番目の分娩月の効果、 $e_{ijklmn}$ : 残差である。

しかしランダムファイルを用いた縦ブロックガウス法の場合には莫大な計算時間が必要となる可能性が高い。やはり半導体ディスク装置を装備しているスーパーコンピュータを使用する必要があると思われる。また、ACOS-930/20 よりも EWS4800/50 の方がターンアラウンド時間が短かったことから、共同利用の汎用機を使用するよりも 1 台のワークステーションを占有して計算した方が縦ブロックガウス法では有利であることが示唆された。

表 3.8 にわが国における乳牛の牛群検定データ (165,703 件) を母方祖父モデルで分析した時の行列式の計算に要した計算時間を示した。アルゴリズムには縦ブロックガウス法を用いた。それぞれ 1 回の計算の結果であるので、あくまでも計算例ということであるがスーパーコンピュータでは問題なく計算できる行列も汎用機では時間がかかりすぎて実用にならなかった。ワークステーションは 1 晩連続運転することによって計算可能となった。

以上の結果より、実際のデータにおいても種雄牛モデルや母方祖父モデルでは縦ブロックガウス法によって当該行列の行列式の値を算出できることが示唆された。行列式の計算以外の部分は全部あわせても通常の BLUP 計算時の計算量の 2 倍程度であることから、これらのモデルについては ABC を用いたモデル選択法は十分に実用的であると思われる。なお個体モデルにおいては実際のデータを用いた場合、通常の BLUP 計算さえも莫大な計算が必要となるため、ABC の算出にあたっては多くの困難が予想される。このような場合、データを 1 万頭くらいずつにわけて各々 ABC を算出してモデル選択をし、それぞれのデータを用いた場合の最適モデルを突き合わ

せて全データ用のモデルを決定するという方法をとるべきかもしれない。

## 摘要

種雄牛評価におけるモデル選択にA B I Cを用いる場合にネックとなる係数行列の変量効果部分の小行列の行列式の値を算出する部分について、種々のアルゴリズムで分析して主記憶使用量と計算時間を計測した。その結果、210x210 や 480x480 程度の行列においてはスカイライン法が最も優れたアルゴリズムであることが明らかとなった。一方、縦ブロックガウス法とウェーブフロント法は計算時間の面で他のアルゴリズムより劣る計測結果であった。しかし、スカイライン法は主記憶を多く消費することから実際の種雄牛評価においてはスーパーコンピュータやワークステーション上で縦ブロックガウス法を用いて行列式の値を算出する方法が現実的であると考察された。

## 第4章 乳牛のフィールドデータを用 いての有効性の検討

### 緒言

わが国の乳牛の育種改良においては、いままでステーション方式の搾乳検定を用いて検定済み種雄牛を作出し、それらの種雄牛を民間に普及することに主眼が置かれてきた。一方、海外では牛群検定制度の普及によるフィールドデータの蓄積と、コンピュータ技術の急速な発達、さらにはHenderson(1973)によるBLUP法の開発によって、フィールドデータを用いた国単位での種雄牛評価が実施されるようになってきた。フィールドデータを用いた種雄牛評価をするメリットは、

1. しっかりした牛群検定組織とそれなりのコンピュータがあれば、中央検定場などの施設や予算の制約を受けずに種雄牛評価が実施できること
2. フィールドにおける飼養環境、飼養技術の現状に基づく評価ができること
3. 過去の種雄牛との比較や他の地域で供用されている種雄牛との比較がある程度可能なこと

などがあげられる。しかしながらBLUP法も万能ではなく、一般に次にあげるような前提条件のもとで正確な種雄牛の評価値を算出する手法である。その前提条件とは

1. モデルが正しいこと。これは第1章で述べたベイズモデルのレベルで正しいこと、すなわち、分散比や分子血縁係数行列も正しいことが必要である。
2. 形質が多変量正規分布に従うこと。ポリジーンに支配されている量的形質であれば通常この条件は満たされる。

3. 選抜および交配計画はデータに対する一定の関数に基づいて実施されること。これは実際に何かの関数を用いて選抜しなければならないということではなくて、結果として統計的に首尾一貫した選抜や交配が行なわれている必要があるという意味である。

この中で2の条件は乳量や乳脂率などの現在評価されている形質では問題なく満たされていると考えられる。3の条件についても若干の時代による変化はあるものの種雄牛評価に使用する5-10年間のデータを考えれば、選抜規準や交配計画の立て方が大きく変化しているとは考えにくい。

もっとも厳しい条件が1の条件である。まず分子血縁係数行列についてはわが国ではおおむね正確な情報が入手可能だと考えられる。分散比についても評価に使うデータをREML法で分析してよりよい推定値を求めることは可能である。しかしながら、どのような要因をモデルに取り込むのかとか、水準をどのようにわけるとか、種雄牛モデルにするのか母方祖父モデルにするのか個体モデルにするのかを決めることはなかなか困難な問題である。通常の分散分析で有意になった要因だけを取り込むというもひとつの方法ではあるが、

1. 分散分析のモデルによって、また有意水準によって有意になったり有意にならなかったりする要因がある
2. 分散分析では分散比などの変量効果の事前情報の良し悪しは判断できない

という欠点がある。

本研究では第1章から第3章まででABICを用いた最適な種雄牛評価モデルを選択する方法について詳細に検討した。そこで、本章ではわが国の乳牛の牛群検定のデータを用いてこれまでに述べてきた方法の有効性を実証的に検討する。

#### 4.1 材料および方法

材料としては、1975年から1987年までの日本のホルスタイン種の牛群検定の初産記録を用いた。実際の分析においては、

1. 娘牛の数が100頭以上ある種雄牛のデータであること、
2. 6頭以上の種雄牛が存在する牛群-年次であること、
3. 乳量が3000kg以上でかつ乳脂量が100kg以上でかつSNF量が200kg以上であること、

表 4.1: 前半データにおける年次別、分娩月別および娘牛年齢区分別のデータ件数

年次	件数	分娩月	件数	娘牛分娩月齢	件数
1975	143	1	6957	22	553
1976	291	2	6906	23	2108
1977	632	3	8727	24	5496
1978	874	4	9090	25	10425
1979	1935	5	8017	26	15207
1980	3672	6	8094	27	16226
1981	10931	7	7238	28	13705
1982	21533	8	7972	29	10899
1983	27318	9	9648	30	7832
1984	30725	10	8493	31	5552
		11	8792	32	3988
		12	8120	33	2793
				34	1926
				35	1344

という3条件を満たした165,703件のデータを用いた。分析データにこのような制限を課したのは、娘牛の数の極めて少ない種雄牛には種雄牛コードが誤記入された可能性があり、また事業開始当初のデータには測定方法などに問題のある異常値があり、これらのデータを分析に含めるとABICを誤って算出することになりかねないためである。

このデータを1975年から1984年までの98,054件のデータ（前半データ；表4.1）と1985年から1987年までの67,649件のデータ（後半データ；表4.2）に分割し、前半データに対して8種のモデルで種雄牛評価を実施して種雄牛評価値やABICを算出した。一方、後半データについては種雄牛と母方祖父牛のペアの平均値を算出した。そして1ペアあたり50件以上のデータを持つペアについてモデルごとに前半データから求めた種雄牛の評価値と母方祖父牛の評価値の1/2の和を求めて、後半データでのペア平均との順位相関を算出した。

種雄牛評価を実施するときのモデルにはすべて農家-年次の母数効果を含めた。そして、他の母数効果について

1. 分娩月と分娩月齢を追加したモデル、
2. 分娩月のみを追加したモデル、

表 4.2: 後半データにおける年次別、分娩月別および娘牛年齢区分別のデータ件数

年次	件数	分娩月	件数	娘牛分娩月齢	件数
1985	30810	1	5448	22	555
1986	34245	2	5053	23	2381
1987	2594	3	6333	24	5893
		4	6265	25	9424
		5	5129	26	11047
		6	5578	27	10218
		7	4795	28	8319
		8	5549	29	6273
		9	6575	30	4513
		10	5506	31	3315
		11	5971	32	2216
		12	5447	33	1632
				34	1073
				35	790

表 4.3: 計算にあたって仮定した遺伝率

形質	遺伝率
乳量	0.35
乳脂率	0.71
SNF 率	0.75

### 3. 分娩月齢のみを追加したモデル

### 4. 農家-年次のみモデル

の4種の数学モデルを分析に用いた。さらに各々のモデルで種雄牛モデルと母方祖父モデルの2通りを検討した。種雄牛モデルの1を式で表わすと(4.1.1)式のようになる。

$$y_{ijklm} = \mu + s_i + h_j + m_k + a_l + e_{ijklm} \quad (4.1.1)$$

ここで

- y : 観測値
- μ : 全平均
- h : 農家-年次の母数効果
- s : 種雄牛の変量効果
- m : 分娩月の母数効果
- a : 分娩月齢の母数効果
- e : 誤差

結局、乳量、乳脂率、SNF率の3形質について8モデルについての計算を行い、種雄牛の評価値、評価値のリピータビリティやABICなどを算出した。これらの値を求める式は第1章で詳しく述べたので本章では割愛する。なお計算にあたっては各形質の遺伝率を表4.3に示したように仮定した。これらの値はABICの線形探索法にもとづく予備分析によって求めたものである。

## 4.2 結果

種雄牛モデルの各モデルで算出したABICの値を表4.4に示した。乳量においてはABICが最小のモデルはモデル1でモデル2、3、4の順にABICの値は大きくなった。モデル1とモデル2の間のABICの差は約2700、ABIC最小のモデル1とABIC最大のモデル4との間のABICの差は約6150であった。

乳脂率においてもABIC最小のモデルはモデル1でモデル2、3、4の順にABICの値は大きくなった。モデル1とモデル2の間のABICの差は66、ABIC最小のモデル1とABIC最大のモデル4との間のABICの差は613であった。

SNF率においてもABIC最小のモデルはモデル1でモデル2、3、4の順にABICの値は大きくなった。ただし、乳量、乳脂率のように直線的に大きくなるのではなくてモデル1とモデル2が-1800程度、モデル3とモデル4が-1240程度と2つのグループに分かれた。

母方祖父モデルの各モデルで算出したABICの値を表4.5に示した。乳量においては種雄牛モデルと同様にABICが最小のモデルはモデル1でモデ

表 4.4: 種雄牛モデルにおける数学モデル別の ABIC

形質	数学モデル			
	1	2	3	4
乳量	1612040	1614744	1615463	1618195
乳脂率	70592	70658	71132	71205
SNF率	-1800	-1794	-1242	-1243

表 4.5: 母方祖父モデルにおける数学モデル別の ABIC

形質	数学モデル			
	1	2	3	4
乳量	1640746	1643121	1643787	1646605
乳脂率	57010	57082	57558	57638
SNF率	-17209	-17277	-16645	-16717

ル 2、3、4 の順に ABIC の値は大きくなった。モデル 1 とモデル 2 の間の ABIC の差は約 2700、ABIC 最小のモデル 1 と ABIC 最大のモデル 4 との間の ABIC の差は約 6150 で、これも種雄牛モデルと同様の傾向であった。また、どのモデルも種雄牛モデルでの ABIC の値にくらべて 28500 程度大きな値を示した。

乳脂率においても種雄牛モデルと同様に ABIC が最小のモデルはモデル 1 でモデル 2、3、4 の順に ABIC の値は大きくなった。モデル 1 とモデル 2 の間の ABIC の差は 72、ABIC 最小のモデル 1 と ABIC 最大のモデル 4 との間の ABIC の差は 628 で、これも種雄牛モデルと同様の傾向であった。また、どのモデルも種雄牛モデルでの ABIC の値にくらべて 13500 程度大きな値を示した。

SNF 率においては種雄牛モデルとは異なりモデル 2 が最小で、以下、モデル 1、モデル 4、モデル 3 の順に ABIC は大きな値を示した。ABIC の最小値と最大値の差は 632 であった。

前半データから算出した種雄牛-母方祖父牛のペアの評価値(予測値)と後半データから算出した種雄牛-母方祖父牛ペアの平均値(実現値)との順位相関を表 4.6 に示した。まず乳量の種雄牛モデルについて見てみると、順位相関はモデル 1、モデル 2 とモデル 3、モデル 4 の順に小さくなっていった。母方祖父モデルについても順位相関はモデル 1、2、3、4 の順に小さくなっ

表 4.6: モデル別の予測値と実現値の順位相関

形質	種雄牛モデル				母方祖父モデル			
	1	2	3	4	1	2	3	4
乳量	.7576	.7465	.7465	.7378	.7572	.7438	.7390	.7320
乳脂率	.8714	.8682	.8778	.8761	.8757	.8742	.8825	.8816
SNF率	.9131	.9127	.9135	.9135	.9155	.9167	.9153	.9163

ていた。また、母方祖父モデルのすべてのモデルは種雄牛モデルよりも小さな順位相関を示した。

乳脂率の種雄牛モデルではモデル 3 が最大の順位相関を示し、以下、モデル 4、モデル 1、モデル 2 の順であったが、最大のモデル 3 と最小のモデル 2 との差は.0960 と小さなものであった。母方祖父モデルにおいても種雄牛モデルと同様の傾向が認められたが、順位相関の値は種雄牛モデルにくらべてわずかに大きな値を示した。

SNF 率の種雄牛モデルにおいては順位相関の値はモデル 3 とモデル 4 が最大で以下、モデル 1、モデル 2 であったが、最大と最小の間の差は.0008 と極めて小さなものであった。母方祖父モデルにおいては順位相関の大きい順にモデル 2、モデル 4、モデル 1、モデル 3 であったが種雄牛モデルと同様に最大と最小の差は.0014 と極めて小さなものであった。種雄牛モデルと母方祖父モデルを比較するとわずかではあるが母方祖父モデルの方が大きな順位相関を示した。

### 4.3 考察

第 1 章でも述べたように ABIC の値が小さいほど良いモデルということが言える。このことを念頭において今回の結果を考察すると乳量においては ABIC で示された良いモデルの順序と予測値と実現値の順位相関の大きさの順序はまったく同じであり ABIC はモデル選択において有効な指標であることが示唆された。乳脂率と SNF 率においても種雄牛モデルよりも母方祖父モデルの方が優れているという点については ABIC と順位相関の両者は同様の結果であった。ただし、種雄牛モデル内の順序や母方祖父モデル内の順序については ABIC と予測値と実現値の順位相関は必ずしも同じ結果を示さなかった。しかし、表にも示したように乳脂率と SNF 率においてはモデル間の差が非常に小さく実質上、差がないと考えてもいいと思われる。

このように差が認められなかったのは乳脂率と SNF 率は遺伝率が高く環

境効果の影響が少ないためにモデルを変更しても結果が変わらないということの意味していると考えられる。しかしまた、今回は実現値として単純な算術平均を用いたので実現値の値が不正確で本来あるはずのモデルの優劣の差が順位相関として現われなかったという可能性も捨てきれない。ただ、実現値を求めるにあたり母数効果を補正するとなると、その補正係数をどのようにして求めるのかということが問題になる。補正係数を求めたときに用いたモデルに近いモデルほど予測値と実現値の順位相関が高くなる可能性が大きいからである。

従って上記の結果は、ABICにもとづくモデルの選択では差がないものを差があるように判断してしまう可能性があることを示唆している。ABICの基本的な考え方は同一のデータ(y)において種々のモデルのABICを求めたときに、ABIC最小のモデルをもっとも当てはまりの良いモデルと見なすというものである。しかし、乳脂率の種雄牛モデルの場合のようにモデル1と3の間にABICで540の差がありながら順位相関ではわずかに逆転している。実現値を正確にはかりにくいということがあるもののABICでどの程度の差があればモデルに明らかな優劣があると言えるのかを今後、検討していく必要がある。

近年、乳牛においては個体モデル(Quaas and Pollak(1980))による雄雌同時評価が実施されることが多くなってきた。この場合、種雄牛モデルの時のように初産データのみを分析するのではなく、複数の産次のデータを同時に分析するのが普通である。従って同一のデータを対象とする本論文でしめたABICを用いた手法では種雄牛モデルと個体モデルとの比較をすることができない。このため今後、相対的な情報量規準を開発する必要がある。

また、わが国の乳牛集団においては地域による遺伝と環境の交互作用が種畜評価におよぼす影響も憂慮されている(和田ら(1991))。この問題についても交互作用を含めたモデルと含めないモデルの双方においてABICを算出してその大小を評価することによって、交互作用を無視することができるかどうかを決定することができるであろう。

このように、ABICを用いたモデル選択手法は今までもすると主観に流れがちであったモデル選択というプロセスを客観的に検討する視点を与えるもので今後、家畜育種の多くの場面で使われていく手法であると思われる。

## 摘 要

わが国における乳牛の牛群検定のデータを用いて、第1章で提案したABICを用いた手法の有効性を検討した。牛群検定データを1975年から1984年までと1985年から1987年までの2つに分割し、前半データを8種のデータで分析して種雄牛評価値やABICを算出した。後半データについては種雄牛と

母方祖父牛のペアの平均値を算出した。そして、1ペアあたり50件以上のデータを持つペアについて前半データから求めた種雄牛の評価値と母方祖父牛の評価値の1/2の和を求めて、後半データでのペア平均との順位相関を算出した。その結果、乳量においてはABICが小さいモデルほど順位相関は大きくなりABICがモデル選択において有効な指標であることが示唆された。乳脂率とSNF率においても種雄牛モデルよりも母方祖父モデルの方が優れているという点についてはABICと順位相関は同様の結果を示した。ただし、種雄牛モデル内や母方祖父モデル内ではモデル間の順位相関に大きな差が認められず、ABICと順位相関は必ずしも同じ傾向を示さなかった。

## 総 括

本論文においては情報量統計学を用いた種畜評価モデルの選択法を提案し(第1章)、シミュレーションによってその有効性を検討した(第2章)。また、この手法を実用データに適用するにあたって必要となる大規模データにおける計算アルゴリズムについて検討した(第3章)。さらに、実際の乳牛のフィールドデータにこの手法を適用して、その有効性を検討した。

第1章ではいままでに提案されてきた尤度比検定や Henderson の方法などの制限のあるモデル選択法にかわって、赤池が提案している情報量統計学を用いて混合線形モデルにおける A B I C を用いた新たなモデル選択手法を提案した。そして、種畜評価において用いられている混合線形ベイズモデルにおける事後尤度関数、変量効果のベイズ推定量、ベイズ型尤度関数、母数効果の最尤推定量と誤差分散の最尤推定量を求め、A B I C を導出した。さらに、複数の変量効果を含む混合線形モデルにおける A B I C を導出し、分散比が未知の場合における A B I C を用いた線形探索法についても紹介した。

第2章では、日本におけるホルスタイン種のフィールドデータを前提としたコンピュータシミュレーションを実施して、本論文において提案したモデル選択の方法について、その有効性を検討した。その結果、すべての反復においてデータ生成時に用いた真のモデルの A B I C の値がすべてのモデルの中で最小であり、モデルが真のモデルに近づくほど A B I C の値は小さくなっていくことがあきらかとなった。また、分散比を変化させた場合にも、データ生成時に用いた真の分散比の場合に A B I C が最小となり、真の分散比に近づくほど A B I C の値は小さくなっていき、A B I C がモデル選択のための指標として有効であることが示唆された。

第3章では A B I C を算出する場合にもっともネックとなる混合モデル方程式の係数行列の変量効果部分の小行列の行列式の値を求めるアルゴリズムについて比較検討した。行列のサイズが 210 の場合と 480 の場合のどちらでもスカイライン法、疎行列用改訂コレスキー法、コレスキー法が他のアルゴリズムよりも計算時間が短かった。ウェーブフロント法、ガウスの消去法、縦ブロックガウス法は桁違いに計算時間がかかった。一方、使用メモリ量においてはウェーブフロント法と縦ブロックガウス法が他のアルゴリズ

ムよりも圧倒的に優れていた。従って、メモリ上で計算可能な問題の場合はスカイライン法を用いると有利であると結論された。また、大規模な問題でスカイライン法ではメモリオーバーとなる場合はウェーブフロント法か縦ブロックガウスを用いることが有利であろうと推察された。

第4章では乳牛のフィールドデータを前半部分と後半部分に分割し、前半部分に対して本論文で提案した手法を適用して、後半部分から得られた結果と比較して本手法の有効性を検討した。その結果、乳量においては A B I C の値が小さいモデルほど予測値と実現値の間の順位相関が高く、A B I C をモデル選択の指標とすることの有効性が示された。乳脂率と SNF 率においては種雄牛モデルよりも母方祖父モデルの方がすぐれている点については A B I C と順位相関は同じ結果を示した。しかし、種雄牛モデル内、および母方祖父モデル内においてはモデル間の順位相関の差がきわめて小さく、A B I C と順位相関は同様の結果を示さなかった。

以上のように種畜評価における A B I C を用いたモデル選択法が有効であることはシミュレーションとフィールドデータの分析からあきらかにすることができた。今後の問題点としては、

1. A B I C の差が小さい場合のモデル選択の判断規準の開発
2. 異なるデータに対する相対的な情報量規準の開発
3. 個体モデルのような次元の大きな行列を扱うモデルにおける A B I C の計算手法の開発

について精力的に研究していく必要があると思われる。

## 謝 辞

本研究を実施するにあたって、終始、懇切なるご指導を賜りました京都大学農学部畜産学科家畜育種研究室の佐々木義之教授に深甚なる感謝の念を捧げます。また、常に有益なご指導と温情あるご激励をいただきました農林水産省畜産試験場の建部 晃 育種第1研究室長、西田 朗 育種第2研究室長、同農業生物資源研究所の小畑 太郎 動物探索評価研究チーム長および同中国農業試験場の松川 正 企画連絡室長に感謝の意を表します。さらに、種々のご助言をいただきました並河 澄 全国和牛登録協会専務理事、阿部猛夫 家畜改良事業団顧問、横内圀生 農林水産省東北農業試験場企画科長、向井文雄 神戸大学農学部畜産学科家畜育種研究室助手の方々にあつくお礼申し上げます。

本論文の第4章で使用したデータの借用を快く許可いただきました家畜改良事業団の皆様にも深く感謝いたします。なお、本論文の計算には農林水産研究計算センターおよび東京大学大型計算機センターを使用させていただきました。

## 引用文献

- Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle., 2nd International Symposium of Information Theory, B.N. Petrov and F. Csaki, eds., Akademiai Kiado, Budapest, 267-281.
- Akaike, H., 1978. A new look at the Bayes procedure., *Biometrika* 65, 53-59.
- Akaike, H., 1978. A Bayesian analysis of the minimum AIC procedure., *Ann. Inst. Statist. Math.* 30, Part A, 9-14.
- Akaike, H., 1979. A Bayesian extension of the minimum AIC procedure of autoregressive model fitting., *Biometrika* 66, 237-242.
- Akaike, H., 1980. Likelihood and bayes procedure, In *Bayesian Statistics* Bernardo et al. eds., University Press, Valencia, Spain, 143-166.
- Dempfle, L., 1977. Relation entre BLUP (Best Linear Unbiased Prediction) et estimateurs bayésiens., *Ann. Génét. Sél. Anim.* 9, 27-32.
- Dinkinson, F. N., R.L. Powell and H. D. Norman, 1976. An introduction to the USDA-DHIA modified contemporary comparison., *NRS, USDA Prod. Res. Report No.165*, 1-7.
- Everett, R.W., R.L. Quaas and A.E. McClintock, 1979. Daughters' maternal grandsires in sire evaluation., *J. Dairy Sci.* 62, 1304-1313.
- Fisher, Sir R.A., 1959. *Statistical methods and scientific inference.*, Oliver and Boyd Limited, Edinburgh, U.K.
- Gianola, D., 1980. Genetic evaluation of animals for traits with categorical response., *J. Anim. Sci.* 51, 1272-1276.
- Gianola, D. and R.L. Fernando, 1986. Bayesian methods in animal breeding theory., *J. Anim. Sci.* 63, 217-244.
- Gianola, D. J.L. Foulley and R.L. Fernando, 1986. Prediction of breeding values when variances are not known., *Génét. Sél. Evol.* 18, 485-498.
- Hansson, N., 1913. Kan man med ekonomisk fördel höja meddelfetthalten i den av vara nötk reatursstammar lömnade mjölken. *Centralanst försökvesendet p. Jordbruksomradef. Meddelande* 75, 85.

- Hartley, H.O. and J.N.K. Rao, 1967. Maximum-likelihood estimation for the mixed analysis of variance model., *Biometrika* 54, 93-108.
- Harvey, W.R., 1976. User's Guide for LSML76. Ohio State University
- Harville, D.A., 1977. Maximum likelihood approaches to variance component estimation and to related problems., *J. Amer. Stat. Assoc.* 72, 320-338.
- Henderson, C.R., 1949. Estimation of changes in herd environment., *J. Dairy Sci.* 32, 706 (Abstr).
- Henderson, C.R., 1953. Estimation of variance and covariance components., *Biometrics* 9, 226-252.
- Henderson, C.R., H.W. Carter and J.T. Godfrey, 1954. Use of the contemporary herd average in appraising progeny tests of dairy bulls., *J. Anim. Sci.* 13, 959 (Abstr).
- Henderson, C.R., O. Kempthorne, S.R. Searle and C.M. von Krosigk, 1959. The estimation of environmental and genetic trends from records subject to culling., *Biometrics* 15, 192-218.
- Henderson, C.R., 1963. Selection index and expected genetics advance., In *Statistical Genetics and Plant Breeding* (Hanson, W.G. and H.F. Robinson eds.). 141-163. Pub.982. NAS-NRC. Washington, D.C..
- Henderson, C.R., 1973. Sire evaluation and genetic trends., *Proc. Anim. Breed. Genet. Symp. in honor of J.L. Lush*, 10-41, ASAS and ADSA, Champaign, Illinois.
- Henderson, C.R., 1974. General flexibility of linear model techniques for sire evaluation., *J. Dairy Sci.* 57, 963-976.
- Henderson, C.R., 1975a. Best linear unbiased estimation and prediction under a selection model., *Biometrics* 31, 423-447.
- Henderson, C.R., 1975b. Use of relationships among sires to increase accuracy of sire evaluation., *J. Dairy Sci.* 58, 1731-1738.
- Henderson, C.R., 1975c. Comparison of alternative sire evaluation methods, *J. Anim. Sci.* 41, 760-770.
- Henderson, C.R. and R.L. Quaas, 1976. Multiple trait evaluation using relatives' records., *J. Anim. Sci.* 43, 1188-1197.

- Henderson, C.R., 1984. Applications of linear models in animal breeding. Univ. of Guelph, Guelph, Ontario, Canada.
- Henderson, C.R., 1985a. Best linear unbiased prediction of nonadditive genetic merits in noninbred populations., *J. Anim. Sci.* 60, 111-117.
- Henderson, C.R., 1985b. MIVQUE and REML estimation of additive and nonadditive genetic variances., *J. Anim. Sci.* 61, 113-121.
- Henderson, C.R., 1985c. Equivalent linear models to reduce computations., *J. Dairy Sci.* 68, 2267-2277.
- Henderson, C.R., 1977. Best linear unbiased prediction of breeding values not in the model for records., *J. Dairy Sci.* 60, 783-787.
- Ishiguro, M. and M. Ishiguro, 1979. Fitting a gaussian model to aperture synthesis data by Akaike's information criterion(AIC)., In *Image Formation from coherence function in astronomy*, Schooneveld, C. van ed., D. Reidel Pub. C.
- 磯田和男・大野 豊、1971. FORTRAN による数値計算ハンドブック、オーム社、東京。
- 伊藤要二・佐々木義之、1985. 枝肉市場成績を用いた種雄牛評価に対する母方祖父の影響、*日畜会報*、56, 619-623.
- Jacoby, S.L.S., J.S. Kowalik and J.T. Pizzo, 1972. Iterative methods for nonlinear optimization problems., Prentice-Hall, Inc., Englewood Cliffs, New Jersey, U.S.A.
- 柏木宣久、1982. 圃場試験に於ける地力の推定、*統計数理研究所輯報*、30, 1-10.
- Keown, J.F., 1974. Comparison of mixed model methods of sire evaluation., *J. Dairy Sci.* 57, 245-250.
- Kendall, Sir M.G. and W.R. Buckland, 1982. A dictionary of statistical terms, 4th edition, Longman Group Limited, London, U.K.
- Kennedy, B.W. and J.E. Moxley, 1975. Comparison of genetic group and relationship methods for mixed model sire evaluation., *J. Dairy Sci.* 58, 1507-1514.
- Kennedy, B.W., 1981. Bias and mean square error from ignoring genetic groups in mixed model sire evaluation., *J. Dairy Sci.* 64, 689-697.

- Lin, C.Y. and A.J. McAllister, 1984. Monte carlo comparison of four methods for estimation of genetic parameters in the univariate case., *J. Dairy Sci.* 67, 2389-2398.
- Lush, J.L., 1931. The number of daughters necessary to prove a bull., *J. Dairy Sci.* 14, 209-220.
- Lush, J.L., 1935. Progeny test and individual performance as indicators of an animal's breeding values., *J. Dairy Sci.* 18, 1-19.
- 村田健郎・小国 力・唐木幸比古、1985. スーパーコンピュータ、pp304、オーム社、東京。
- 長嶺慶隆・林 孝・佐藤 博、1988. 産子の屠殺日齢が種雄牛の評価に及ぼす影響、*日畜会報*、59, 585-589.
- 日本数学会 編、1985. 岩波数学辞典 第3版、pp1609、岩波書店、東京
- Norman, H.D., 1976. Theoretical background for the USDA-DHIA modified contemporary comparison sire summary procedure., *NRS, USDA Prod. Res. Report No.165*, 8-17.
- Patterson, H.D. and R. Thompson, 1971. Recovery of inter-block information when block sizes are unequal., *Biometrika* 58, 545- 554.
- Pollak, E.J., G.R. Ufford and S.J. Gross, 1977. Comparison of alternative models for within-herd genetic evaluation of beef cattle., *J. Anim. Sci.* 45, 1010-1014.
- Pollak, E.J. and R.L. Quaas, 1981a. Monte carlo study of within-herd multiple trait evaluation of beef cattle growth traits., *J. Anim. Sci.* 52, 248-256.
- Pollak, E.J. and R.L. Quaas, 1981b. Monte carlo study of genetic evaluations using sequentially selected records., *J. Anim. Sci.* 52, 257-263.
- Pollak, E.J. and R.L. Quaas, 1983. Definition of group effects in sire evaluation models., *J. Dairy Sci.* 66, 1503-1509.
- Pollak, E.J., J. van der Werf and R.L. Quaas, 1984. Selection bias and multiple trait evaluation., *J. Dairy Sci.* 67, 1590-1595.
- Quaas, R.L., R.W. Everett and A.C. McClintock, 1979. Maternal grandsire model for dairy sire evaluation., *J. Dairy Sci.* 62, 1648-1654.

- Quaas, R.L. and E.J. Pollak, 1980. Mixed model methodology for farm and ranch beef cattle testing programs., *J. Anim. Sci.* 51, 1277- 1287.
- Quaas, R.L. and L.D. Van Vleck, 1980. Categorical traits sire evaluation by best linear unbiased prediction of future progeny category frequencies., *Biometrics* 36, 117-122.
- Rao, C.R., 1973. *Linear Statistical Inference and Its Applications*. 2nd. edition., John Wiley & Sons, Inc., New York.
- Robertson, A. and J.M. Rendel, 1950. The use of progeny testing with artificial insemination in dairy cattle., *J. Genet.* 50, 21-31.
- Robertson, A. and J.M. Rendel, 1954. The performance of heifers got by artificial insemination., *J. Agric. Sci., Camb.* 44, 184-192.
- 佐々木義之・江藤啓一郎・向井文雄・並河 澄、1976. 生産・肥育地帯での肉牛成績による種雄牛の産肉能力に対する現場後代検定の可能性、*日畜会報*、47, 354-358.
- 佐々木義之・祝前博明、1980. B L U P法による増体率および飼料利用性に関する黒毛和種種雄牛の育種価推定、*日畜会報*、51、93-99.
- Sasaki, Y. and C.R. Henderson, 1986. Best linear unbiased prediction with the reduced animal model: An application to evaluation of performance-tested males., *J. Anim. Sci.* 63, 1384-1388.
- 佐々木義之・佐々江洋太郎、1988. フィールド記録を用いたB L U P法による肉用種雄牛評価のためのモデルの検討、*日畜会報*、59、23-30.
- Schaeffer, L.R., 1976. Maximum likelihood estimation of variance components in dairy cattle breeding research., *J. Dairy Sci.* 59, 2146-2151.
- Schaeffer, L.R., J.W. Wilton and R. Thompson, 1978. Simultaneous estimation of variance and covariance components from multitrait mixed model equations., *Biometrics* 34, 199-208.
- 田辺國士・田中輝雄、1983. ベイズモデルによる曲線・曲面のあてはめ、*月刊地球* 5, 179-186.
- Ufford, G.R., C.R. Henderson and L.D. Van Vleck, 1979. An Approximate Procedure for Determining Prediction Error Variances of Sire Evaluations., *J. Dairy Sci.* 62, 621-626.

- Ufford, G.R., C.R.Henderson, J.F.Keown and L.D. Van Vleck, 1979. Accuracy of first lactation versus all lactation sire evaluations by best linear unbiased prediction., J. Dairy Sci. 62, 603-612.
- Van Vleck, L.D. and P.J. Karner, 1980. Sire evaluation by best linear unbiased prediction for categorically scored type traits., J. Dairy Sci. 63, 1328-1333.
- 和田康彦、1986. B L U P法を用いた種雄牛評価、農林水産研究計算センター報告 B10, 1-49.
- 和田康彦・建部 晃、1986. 牛群データに基づく乳用種雄牛評価の可能性、日畜会報 57, 97-102.
- 和田康彦・塩見雅史・松川 正、1987. 黒毛和種の現場後代検定成績に対する反復B L U P法の適用、日畜会報、58, 374-380.
- 和田康彦、1988. A B B L U P利用者マニュアル、未発表。
- Wada, Y. and N. Kashiwagi、1989. A new model selection method of mixed linear models with Akaike's theory in sire evaluation. Proc. of the 6th Internatl. Congr. of SABRAO, 909-912.
- Wada, Y. and N. Kashiwagi、1990. Selecting statistical models with information statistics. J. Dairy Sci. 73, 3575-3582.
- 和田康彦・建部 晃・松本成生・横内圀生、1991. ホルスタイン種の種雄牛評価に及ぼす地域と種雄牛の交互作用の影響、日畜会報、62, 59-64.
- Wright, S., 1931. On the evaluation of dairy sires., Amer. Soc. Anim. Prod. Proc., 71-78.
- Yanagimoto, T. and M. Yanagimoto, 1987. The use of marginal likelihood for a diagnostic test for the goodness of fit of the simple linear regression model., Technometrics 29, 95-101.
- Yapp, W.W., 1924. Transmitting ability of dairy sires., Amer. Soc. Anim. Prod., Ann. Proc. Dec. 1924.