統計数理 (1999)
第 47 巻 第 1 号 49-61

特集「統計基礎理論」
[原著論文]

擬似尤度の構築と非線型回帰モデルへの応用

統計数理研究所　汪　金　芳

(受付 1998 年 10 月 30 日；改訂 1999 年 3 月 10 日)

要　旨

スコア推定関数は母数空間上の勾配ベクトル場を成している．勾配ベクトル場を取り扱うことは便利である．例えば，勾配場の沈点 (源点)，すなわち，尤度方程式の解は，尤度関数の局所最大値 (最小値) として解釈できる．一般に，擬似スコアのような 2 次のモーメントに基づいて構成される推定関数は勾配場ではない．この場合，推定方程式による推測は尤度推論と本質的な違いがある．例えば，推定方程式の根がある統計的に意味をもつスカラー関数の極値としての解釈ができなくなる．したがって，推定方程式の多重根から一致推定値とそうでないものを見分けるようなことは更に困難である．本論文では，推定関数の不可積分問題を紹介するとともに，Wang (1999, *Ann. Inst. Statist. Math.*, **51** (to appear)) で提案された Helmholtz 型擬似尤度を概観し，その正当性について更に議論した後，ブートストラップ法を用いた擬似尤度比検定に基づく多重根の選択法を提案する．例として，測定誤差がある場合のロジスティック回帰モデルへの応用を示す．

キーワード：ブートストラップ法，推定関数，一般化線型模型，変数誤差モデル，多重根，ベクトル場．

1. はじめに

1.1 指数分布族から一般推定方程式へ

説明変数ベクトルを x，応答変量ベクトルを Y とする回帰モデルにおいてしばしば目的とされるのは，関係式

$$g(\mu)=x'\theta, \qquad \mu=EY$$

で定義される構造 (回帰) パラメータ θ に関する推測である．ただし $g(\cdot)$ は連結関数と呼ばれる既知の単調関数であり，また $\eta=g(\mu)=x'\theta$ は線型予測子である．応答変量 Y の確率分布を指数分布族と仮定すれば，θ に関するスコア関数は

$$U(\theta)=\dot{\mu}'(\theta)\, V^{-1}(\theta)\,[\,Y-\mu(\theta)] \tag{1.1}$$

と書ける．ここで，$\mu(\theta)=EY$, $V(\theta)=\mathrm{var}\,Y$, $\dot{\mu}(\theta)=(\partial/\partial\theta)\,\mu(\theta)$ である．このモデルは一般化線型模型 (GLM ; Nelder and Wedderburn (1972), McCullagh and Nelder (1989)) と呼ばれており，線型モデルをはじめ，対数線型モデル，ロジスティックモデルなどを含んでいる．

指数分布族の仮定は理論的に便利であるが，縦断的データのような実際のデータを解析する

際には，確率分布を特定することは困難な場合が多い．しかし，確率分布の代わりに，2次までのモーメントを仮定し，(1.1) 式を形式的なスコア関数とみなせば，θ に関する推論を同様に行うことができる．Wedderburn (1974) は (1.1) を擬似スコアと呼び，それに関する初等的な性質を調べている．

擬似スコア (Wedderburn (1974), McCullagh (1983)) に基づく推測は，分布の仮定が必要でないため，よりロバストである．また，応答変量が独立でなくても容易に拡張できる．例えば，Yが縦断的データの場合，共分散行列Vをブロック対角行列とすればよい．即ち

$$U(\theta)=\sum_i \dot{\mu}_i'(\theta)\, V_i^{-1}(\theta)\,[\,Y_i-\mu_i(\theta)\,]. \tag{1.2}$$

ここで，和は個体について取る．Liang and Zeger (1986) は更に共分散行列を $V_i = A_i^{1/2}R(\alpha)A_i^{1/2}$ と広く仮定し，(1.2) に基づく θ に関する推論を提案している．推定関数 (1.2) は一般推定方程式 (GEE) として知られているものであり，特に生物統計などの分野で幅広く応用されている (Fahrmeir and Tutz (1994) ; Diggle et al. (1994) など参照)．

一般推定方程式 (1.2) のような推定関数は一般に複雑であり，ある確率分布のスコアになるとは限らない．すなわち，推定関数は勾配場ではないことである．実際，滑らかな推定関数の空間の中で，勾配場は“マレ”(atypical) である (Poston and Stewart (1976), p. 33)．本論文では不偏推定関数 $g(\theta)=g(\theta\,;Y)$ を考え，$\dim g(\theta)=\dim\theta=p>1$ を仮定する．非勾配系，或いは，不可積分推定関数 $g(\theta)$ とは，次のいずれかの条件を満たすものをいう．

1．任意のスカラー関数 $\phi(\theta)$ に対し，$g(\theta)\neq \mathrm{grad}\,\phi(\theta)$ である．
2．ヘッシャン行列 $\hat{I}_g=\dot{g}(\theta)$ が非対称である．
3．線積分 $\int_{t_0}^{t_1} g'(\theta)\,\dot{\theta}(t)\,dt$ が R^p 空間の滑らかなパス $\{\theta(t),\, t\in[t_0,\,t_1]\}$ に依存する．
4．推定関数 $g(\theta)=(g_i(\theta))$ を1次微分形式 $g(\theta)=\sum_i^p g_i(\theta)\,d\theta_i$ とみなすとき，

$$d[g(\theta)]\equiv\sum_{i,j=1}^{p}\frac{\partial g_j(\theta)}{\partial\theta_i}d\theta_i\wedge d\theta_j=\sum_{i<j}\left[\frac{\partial g_j(\theta)}{\partial\theta_i}-\frac{\partial g_i(\theta)}{\partial\theta_j}\right]d\theta_i\wedge d\theta_j\neq 0$$

である (ポアンカレの補題；Lovelock and Rund (1975), p. 145)．ここで，d は外微分演算子，$\wedge$ は wedge 積である．

1.2 非勾配系推定関数による推測

尤度推論では，尤度関数に基づいて推論を行う．しかし，不可積分推定関数による推測を行う場合には，ポテンシャルが存在しないため，本質的な困難が起こる．例えば，尤度方程式の根の内，一致推定量が存在すれば，それは，漸近的に，(1)局所最尤推定量であり，また(2)ユニークである，ことが知られている (Huzurbazar (1948))．まず，一致推定量における負のヘッシャンが，大数の法則より，正の定符号である平均フィッシャー情報量に近づくことから，一致推定量は局所最大値に対応することが分かる．すなわち，(1)が成り立つ．また，2つの一致推定量が存在すれば，漸近的にそれぞれ局所最大値であるため，もしそれらが異なれば，両者の間に最小値をとる一致推定量が存在する．これは(1)と矛盾する．つまり，(2)が成立する．このように，尤度推論では尤度関数の存在が本質的な役割を果たしている．しかし，非勾配系推定関数に対しては，ポテンシャルが存在しないため，上の性質(1)の意味さえ持たない．

本論文では，(1.2) 式を含む非勾配系推定関数における擬似尤度の構成法について議論する．さらにその応用例の1つとして，4節では多重根の選択について，擬似尤度比検定に基づく方法を提案する．

2. 非勾配系推定関数

2.1 例

例 2.1. 遷移確率の推定. 政党 P_i ($i = 1,\ldots, p, p \geq 2$) があり，前回 P_i に投票した人が次回 P_1 に移る確率 $\theta = (\theta_1,\ldots, \theta_p)$ の推定に興味があるとする．すでに n 回の投票が独立に行われ，j 回目の P_i の得票数が m_{ij} ($i = 1,\ldots, p; j = 1,\ldots, n$) とする．更に，$j$ 回目の得票数 m_{ij} が既知のとき，次回 P_i から P_1 に移る（観測できない）投票者数 X_{ij} が独立に二項分布 $B(m_{ij}, \theta_i)$ に従うとする．この場合，観測可能な確率変数は $Y_j = \sum_{i=1}^{p} X_{ij}$ のみである．

確率変数 Y_j の平均 $\mu_j(\theta) = \sum_{i=1}^{p} m_{ij}\theta_i$ と分散 $V_j(\theta) = \sum_{i=1}^{p} m_{ij}\theta_i(1-\theta_i)$ から，擬似スコア

$$u_i(\theta) = \sum_{j=1}^{n} m_{ij}[y_j - \mu_j(\theta)]/V_j(\theta), \qquad (i = 1,\ldots, p) \tag{2.1}$$

を構成することができる．単純計算より，

$$\delta = \frac{1}{2}\left(\frac{\partial u_i}{\partial \theta_k} - \frac{\partial u_k}{\partial \theta_i}\right) = \sum_{j=1}^{n} m_{ij}m_{kj}(y_j - \mu_j)(\theta_i - \theta_k)/V_j^2 \neq 0\ (i \neq k)$$

から，推定関数 (2.1) が非勾配系であることが分かる（1.1 節を参照）．期待値 $E\delta$ がゼロになることに注意する．McCullagh and Nelder ((1989), pp. 336-339) と Li and McCullagh (1994) では $p = 2$ の場合，Wang (1999) では $p \geq 2$ の場合についての議論を行っている．

例 2.2. 安定分布. 指数 α が 1（コーシー分布）と 2（正規分布）を除き，特性関数 $\phi(t;\theta) = \exp(i\theta_1 t - |\theta_2 t|^{\alpha})$ で規定される対称安定分布の尤度関数は陽に書けないため，位置母数 θ_1 とスケール母数 θ_2 の最尤推定は難しい．McLeish and Small (1992) では次の推定関数

$$u(\theta;Y) = \sum_{i=1}^{n}\left[\sin\left(t_1\frac{Y_i-\theta_1}{\theta_2}\right), \cos\left(t_1\frac{Y_i-\theta_1}{\theta_2}\right) - \exp(-t_2^{\alpha})\right]' \tag{2.2}$$

を提案している．ここで，α は既知であり，t_1 と t_2 は適当な定数である．不可積分性の定義から推定関数 (2.2) が勾配場でないことはすぐ分かる．

例 2.3. 母数支配モデル. 確率変数 $Y = (Y_1,\ldots, Y_n)$ の従属性と過大分散をモデリングするための有効な方法として，潜在確率過程 ϵ を導入し，Y の条件付きモーメントをモデリングすれば，周辺モーメントに基づく解析ができる (parameter-driven model; Cox (1981))．例えば，点過程の場合 (Zeger (1988))，条件付き平均と分散をそれぞれ

$$u_t = E(Y_t|\epsilon_t) = \exp(x_t'\theta)\,\epsilon_t, \qquad w_t = \mathrm{var}(Y_t|\epsilon_t) = u_t$$

と仮定すれば，Y の周辺モーメントから擬似スコアを構成することができる．ただし，ϵ_t は $E(\epsilon_t) = 1$ と $\mathrm{cov}(\epsilon_t, \epsilon_{t+\tau}) = \sigma^2\rho(\tau)$ を満たすものとする．この場合の擬似スコアの不可積分性も容易に確かめられる．

例 2.4. 一般推定方程式. 擬似スコア (1.1) の可積分性は分散共分散行列 $V[\mu(\theta)]$ の性質によって決まる．一般推定方程式 (1.2) における分散共分散構造 $V = A^{1/2}R(\alpha)A^{1/2}$ は一般に不可積分推定関数を定義する．

推定関数 (1.2) は縦断的データに対して適用されたものであるが，可積分性が変わらないた

め，ここで1個体の場合について議論する．この場合の回帰係数βに関する推定関数は$D'V^{-1}S$となる．ここで，$S=Y-\mu(\theta)$, $V=A^{1/2}R(\alpha)A^{1/2}$, $D=\dot{\mu}(\theta)$であり，Aはデータが独立と仮定した場合の，1変量の指数分布族における分散共分散行列である．例として，相関corr$(Y_t, Y_s)=\alpha$を仮定する$(t\neq s)$．即ち，$R(\alpha)=(r_{ij})$の対角要素は1，非対角要素はすべてαであると仮定する．この場合，定数$-[1+(n-2)\alpha-(n-1)\alpha^2]^{-1}\alpha$を除けば，擬似スコアは$D'V^{-1}S=XWS$と書ける．ここで，$X$は計画行列で，$W=(w_{ij})$は$w_{ii}=-(n-2+1/\alpha)h'(\eta_i)$と$w_{ij}=h'(\eta_i)\sqrt{a''(\theta_i)/a''(\theta_j)}$ $(i\neq j)$を満たす行列である(記号はLiang and Zeger (1986)を参照)．推定関数XWSは一般に不可積分である．例えば，Yがガンマ分布に従うとし，$h(\cdot)$を自然な連結関数とすると，$w_{ii}=-(n-2+1/\alpha)$, $w_{ij}=\mu_i/\mu_j$ $(i\neq j)$が得られる．行列(w_{ij})のμ_kに対する微分配列の非対称性より，XWSの不可積分性が分かる．

例 2.5.　測定誤差がある場合のロジスティック回帰モデル．　この例については6節で詳しく検討する．

2.2　不可積分の度合い

推定関数$g(\theta)$を1次微分形式と見なすとき，外微分$d[g(\theta)]$がゼロであれば$g(\theta)$は可積分である(1.1節)．したがって，2次形式$d[g(\theta)]$が推定関数$g(\theta)$の勾配場からのズレを表す指標と考えられる．指標$\mathrm{NC}(g)\equiv d[g(\theta)]$を推定関数$g(\theta)$の不可積分の度合いと呼ぶことにする．不可積分の度合い$\mathrm{NC}(g)$は，(i)ユニークであり，(ii)ゼロになるための必要十分条件は推定関数$g(\theta)$が可積分である．不可積分の度合い$\mathrm{NC}(g)$の構造を調べるため，Hodgeのスター作用素，$*$，を導入する(Darling (1994), p. 17)．

命題 2.1.　[Wang (1999)]　任意の滑らかな推定関数$g(\theta)$ $(p\geq 2)$に対し，スカラー関数$\phi(\theta)$と2次微分形式$\mathcal{P}(\theta)$が存在し，次の分解が成立する

(2.3)
$$g(\theta)=d\phi(\theta)+*d*\mathcal{P}(\theta).$$

分解(2.3)の第2項目$R=*d*\mathcal{P}(\theta)$は

$$R=\left(-\frac{\partial}{\partial\theta_2},\frac{\partial}{\partial\theta_1}\right)'\psi(\theta)\quad(p=2\text{の場合}),\qquad R=\nabla\times\mathcal{V}(\theta)\quad(p=3\text{の場合})$$

と書き直すことができる．ただし，∇ $(\nabla f=\mathrm{grad}\ f)$をHamiltonの演算子，$\Delta$ $(\Delta=\nabla^2=\sum_i\partial^2/\partial\theta_i^2)$をラプラシアンとし，$\times$をベクトルの外積とする．また，$\psi(\theta)$はスカラー関数で，$\mathcal{V}(\theta)$は通常のベクトルである．3次元の場合，分解(2.3)はHelmholtzの分解定理として知られている．したがって，(2.3)はHelmholtzの分解定理の拡張である．命題2.1によって，不可積分の度合い$\mathrm{NC}(g)$はそれぞれ，$\mathrm{NC}(g)=\Delta\psi(\theta)$ $(p=2)$,　$\mathrm{NC}(g)=\nabla\times[\nabla\times\mathcal{P}(\theta)]$ $(p=3)$, $\mathrm{NC}(g)=(d*)^2\mathcal{P}(\theta)$ $(p\geq 2)$で与えられ，発散ゼロのベクトル場によって一意的に決まる．

2.3　線型近似

次節で議論される擬似尤度の構成に，線型関数$h(\theta)=A\theta+b$が重要な役割を果たす．ここで，Aとbはそれぞれθに無関係の定数行列と定数ベクトルである．この場合，$\mathrm{NC}(h)=(A'-A)/2$となる．これがゼロになる，すなわちhが可積分になるのは行列Aが対称の場合であり，またそのときに限る．更に，ベクトル場$(A-A')\theta$の発散がゼロであることから

$$h(\theta)=\left[\frac{1}{2}(A+A')\,\theta+b\right]+[\mathrm{NC}\,(h)\,\theta] \tag{2.4}$$

は Helmholtz 型分解となっていることが分かる．この分解の特徴は発散ゼロのベクトル場が不可積分の度合いを用いて表されていることである．

不可積分の度合い $\mathrm{NC}\,(g)$ は確率的に変動し，平均 $E[\mathrm{NC}\,(g)]=0$ となる推定関数 g を平均可積分推定関数と呼ぶことにする．推定関数 g が平均可積分であるための必要十分条件は，擬似ヘッシャンの期待値 $I_g=E\dot{g}\,(\theta)$ が対称になることである．擬似スコア (1.1) は重要な平均可積分推定関数の例である (例 2.1 も参照)．平均可積分関数 $g(\theta)$ の線型近似 h を考えると，分解 (2.4) における発散ゼロのベクトル場 $\mathrm{NC}\,(h)\,\theta$ の期待値はゼロとなる．

3. 擬似尤度

任意の滑らかな推定関数 $g(\theta)$ が与えられているとき，(2.3) によって，$g(\theta)$ は勾配場と発散ゼロのベクトル場によって分解される．すなわち，$g(\theta)=d\phi(\theta)+*d*\mathcal{P}(\theta)$ である．Helmholtz 型ポテンシャル $\phi(\theta)$ は調和関数だけの不定性を持つ．すなわち，任意の調和関数 $h(\theta)$ に対し，$\phi(\theta)+h(\theta)$ も Helmholtz 型ポテンシャルである．例えば，ポテンシャル $\phi_1(\theta)=(\theta_1-\theta_2)^2/2$ と $\phi_2(\theta)=(\theta_1+\theta_2)^2/2$ は共にベクトル場 $(\theta_1,\ \theta_2)$ の Helmholtz 型ポテンシャルであることは，次の分解

$$\begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix}=\begin{pmatrix}\theta_1-\theta_2\\ \theta_2-\theta_1\end{pmatrix}+\begin{pmatrix}\theta_2\\ \theta_1\end{pmatrix}, \tag{3.1}$$

$$\begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix}=\begin{pmatrix}\theta_1+\theta_2\\ \theta_1+\theta_2\end{pmatrix}+\begin{pmatrix}-\theta_2\\ -\theta_1\end{pmatrix}, \tag{3.2}$$

から分かる．ポテンシャルの差 $\phi_2(\theta)-\phi_1(\theta)=2\theta_1\theta_2$ が調和関数である．Helmholtz 型分解を応用するためのもう一つの問題点として，推定関数 $g(\theta)$ が可積分のとき，例えばスコア関数の場合，'残差項' $(*d*)\,\mathcal{P}(\theta)$ がゼロになる保証はない．実際，分解 (3.1) と (3.2) はポテンシャル $\phi(\theta)=(\theta_1^2+\theta_2^2)/2$ の勾配場の分解になっている．Helmholtz 型分解 $g(\theta)=d\phi(\theta)+g_r$ におけるポテンシャル $\phi(\theta)$ の選択は，発散ゼロのベクトル場 g_r を規定することと同値であるため，ここで次の規準を考えるのが自然であろう．

1．擬似ヘッシャン $\dot{g}\,(\theta)$ が対称のとき，g_r がゼロである．
2．擬似ヘッシャンの期待値 $E_\theta\dot{g}\,(\theta)$ が対称のとき，$E_\theta g_r$ がゼロである．

適当なスカラー関数 $\psi(\theta)$ が存在し，推定方程式 $g(\theta)$ を $g(\theta)=-d\phi(\theta)+(*d)\,[\psi(\theta)\,(*d)\,g(\theta)]$ と分解できれば，上の規準は満たされる．一般の不偏推定関数 $g(\theta)$ に対して，このようなスカラー関数の存在は保証されない．しかし，線型関数に対して，$\psi(\theta)$ を

$$\psi(\theta)=\frac{(-1)^p}{4}\,\theta'\theta \tag{3.3}$$

と取ればよい (Wang (1999))．この場合，発散ゼロの“残差”ベクトル場 g_r の軌道 (解曲線) は，原点を円心とする同心円を成している．スカラー関数 (3.3) に関するより詳しい考察は Wang (1999) を参照されたい．

以上のような考察から，ここではまず非線型推定関数の局所線型化を行う．パラメータ $\xi\in$

Θ を既知とし，ヘッシャン行列 $A_\xi=\{(\partial/\partial\theta)\,g(\theta\,;\,Y)\}_{\theta=\xi}$ が非退化のとき，パラメータ ξ における $g(\theta)$ の線型近似

$$\tilde{g}(\theta,\xi\,;\,Y)=A_\xi(Y)(\theta-\xi)+b_\xi(Y) \tag{3.4}$$

を考える．ここで，$b_\xi=b_\xi(Y)=g(\xi,\,Y)$ である．線型近似 (3.4) における行列 A_ξ が対称の場合，$\tilde{g}(\theta,\xi\,;\,Y)$ を積分することによって2次のポテンシャルが得られる．行列 A_ξ が任意の ξ に対して対称であれば，元の推定関数 $g(\theta)$ が可積分になり，直接 $g(\theta)$ を積分することによって，一意的にポテンシャルを決めることができる．ここで，パラメータ ξ が非退化臨界点，すなわち，$g(\xi)=0$ で，行列 A_ξ が非退化とすると，適当なパラメータ変換によって，この場合のポテンシャルも2次形式に書くことができる（モースの補題；Poston and Stewart (1978), Chapter 4, § 2).

非線型不可積分推定関数 $g(\theta)$ に対し，スカラー関数 (3.3) が線型ベクトル場 (3.4) の発散ゼロのベクトル場を定義し，これに対応する Helmholtz 型ポテンシャルは

$$\phi(\theta,\xi\,;\,Y)=\frac{1}{2}(\theta-\xi)'B_\xi(\theta-\xi)+b_\xi'(\theta-\xi) \tag{3.5}$$

と書くことができる．ここで $B_\xi=(A_\xi+A_\xi')/2$ である．この論文では，2次関数 (3.5) を推定関数 $g(\theta)$ から誘導される擬似尤度と呼ぶことにする．

擬似スコア (1.1) の場合，擬似尤度 (3.5) は

$$\phi(\theta,\xi,Y)\approx(\mu_\theta-\mu_\xi)'\,V_\xi^{-1}(Y-\mu_\xi)-\frac{1}{2}(\mu_\theta-\mu_\xi)'\,V_\xi^{-1}(\mu_\theta-\mu_\xi) \tag{3.6}$$

となる．(3.6) から分かるように，擬似尤度はパラメータ変換に対して不変である．推定関数が擬似スコアでない場合にもパラメータ変換に対しての不変性は局所的に成り立つ．

いま，対数尤度 $\ell(\theta)=\log L(\theta)$ のパラメータ ξ における2次近似

$$\ell(\theta)\approx\ell(\xi)+S'(\xi)(\theta-\xi)+\frac{1}{2}(\theta-\xi)'\,\dot{S}(\xi)(\theta-\xi) \tag{3.7}$$

を考える．ここで，$S(\xi)=\dot{\ell}(\xi)$ はスコアである．指数分布族の場合，スコア $S(\xi)$ は $\dot{\mu}_\xi'V_\xi^{-1}(Y-\mu_\xi)$ と書け，(3.7) は

$$\ell(\theta)\approx[\dot{\mu}_\xi'V_\xi^{-1}(Y-\mu_\xi)]'(\theta-\xi)+\frac{1}{2}(\theta-\xi)'\,\dot{S}(\xi)(\theta-\xi)+\text{constant} \tag{3.8}$$

と書くことができる．更に，$\dot{S}(\xi)=\partial\dot{\ell}(\xi)/\partial\xi$ を $-\dot{\mu}_\xi'V_\xi^{-1}\dot{\mu}_\xi$ で近似，$(\mu_\theta-\mu_\xi)\approx\dot{\mu}_\xi(\theta-\xi)$ を注意すると，対数尤度 (3.8) は次のように変形することができる．

$$\ell(\theta)\approx(\mu_\theta-\mu_\xi)'\,V_\xi^{-1}(Y-\mu_\xi)-\frac{1}{2}(\mu_\theta-\mu_\xi)'\,V_\xi^{-1}(\mu_\theta-\mu_\xi)+\text{constant}. \tag{3.9}$$

(3.9) 式と (3.6) 式を比較すれば，スコア推定関数に対する擬似尤度は，局所的に尤度に一致することが分かる．

4. 多重根の選択

4.1 はじめに

本節では，擬似尤度の応用として，スコアを含む一般の推定方程式の多重根から一致推定量

を判別するための，ブートストラップ擬似尤度比検定による方法を議論する．

スコア推定関数の場合においても，尤度方程式の多重根から一致推定量を判別することは容易でない場合が多い．二次元の正規分布における相関係数の推定や，コーシー分布におけるロケーションの推定問題などが例である (Stuart and Ord (1991))．二次元正規分布の相関係数に関する尤度方程式の解は漸近的にユニークになる．一方，コーシー分布の場合には，任意のサンプル・サイズに対して多重根が存在することが Fisher (1925) 以来指摘されている．さらに，尤度を多重根選択の基準として用いることが適切でない反例が多く知られている (LeCam (1979), Ferguson (1982) などを参照)．Lehmann (1983) では，(1)正則条件のチェック，(2)別な方法で一致推定量を構成し，尤度方程式の根と比較する，また(3)尤度方程式の近似解を求める，との提案がなされている．Lehmann (1983) でも述べられているように，正則条件のチェックは一般に難しい．また，尤度方程式とは別に適切な一致推定量を構成することは困難な場合が多い (Barendregt and van Pul (1995))．

ここで提案する方法は，難しい正則条件のチェックの必要がなく，また，推定方程式とは別の方法で一致推定量を用意しておく必要もない．まず，4.2 節では，Helmholtz 型擬似尤度比の性質を概観し，4.3 節ではブートストラップ擬似尤度比統計量の性質を検討する．次の 5 節で応用例を示す．

4.2 擬似尤度比検定

不偏推定関数 $g(\theta, Y_1,\ldots, Y_n)$ に基づき，帰無仮説，$H_0: \theta = \theta_0$，を検定する問題を考える．この場合，擬似尤度は

$$\exp[\phi(\theta, \xi, Y)] \propto \exp\left[\frac{1}{2}(\theta-\xi)'B_\xi(\theta-\xi)+g'(\xi, Y)(\theta-\xi)\right] \tag{4.1}$$

であるので，擬似尤度比検定統計量を次のように定義する

$$\gamma(\theta_0, \xi; Y) = -2[\phi(\theta_0, \xi, Y)-\phi(\tilde{\theta}, \xi, Y)]. \tag{4.2}$$

ただし，$\tilde{\theta} = \mathrm{argmax}_\theta[\phi(\theta, \xi, Y)] = \xi - B_\xi^{-1}g(\xi; Y)$ である．擬似尤度比統計量 (4.2) は母数変換に対して不変である．検定統計量の分布を調べるため，ξ を決めなければならない．自然な選択として，$\xi = \theta_0$，或いは，$\xi = \hat{\theta} = \mathrm{argsolve}_\theta\{g(\theta, Y) = 0\}$ が考えられる．このとき，

$$\gamma(\theta_0, \theta_0; Y) = g'(\theta_0; Y)(-B_{\theta_0}^{-1})\,g(\theta_0; Y) \tag{4.3}$$

$$\gamma(\theta_0, \hat{\theta}; Y) = (\hat{\theta}-\theta_0)'(-B_{\hat{\theta}})(\hat{\theta}-\theta_0) \tag{4.4}$$

となることが分かる．帰無仮説の下では，$g(\theta_0; Y)$ の不偏性と漸近正規性を仮定すれば，$\gamma(\theta_0, \theta_0; Y)$ の分布を $\sum_{i=1}^{p}\lambda_i Z_i^2$ で近似できる (Wang (1999))．ここで，$(\lambda_i) = \mathrm{eigen}(\Gamma_0\Sigma_0^{-1})$, $\Gamma_0 = E_{\theta_0}[g(\theta_0; Y)\,g'(\theta_0; Y)]$, $-2\Sigma_0 = E_{\theta_0}\,\dot{g}(\theta_0; Y) + E_{\theta_0}\,\dot{g}'(\theta_0; Y)$ であり，Z_i は独立で平均ゼロ分散 1 の正規分布に従う確率変数である．擬似スコアのように，推定関数 $g(\theta; Y)$ が平均勾配場になっている場合，$\Sigma_0 = -E_0\,\dot{g}(\theta_0; Y)$ となり，また情報不偏の場合には，$\Gamma_0 = \Sigma_0$ となる．すなわち，情報不偏推定関数に対し，検定統計量 (4.3) の帰無分布は $\chi^2(p)$ で近似できる．

根 $\hat{\theta}$ が一致推定量である場合，同様な議論を (4.4) についても適用できる．これらの議論は，推定関数またはその根の漸近正規性が成り立つことを前提としている．これらの正則条件については，McCullagh (1983)，Heyde (1997) などを参照されたい．

4.3 ブートストラップ擬似尤度比検定

この節では推定関数 $g(\theta; Y)$ が $g(\theta; Y) = \sum_{i=1}^{n} g_i(\theta) = \sum_{i=1}^{n} g(\theta; Y_i)$ のように分解可能

と仮定する．共変量 $x_1,\dots, x_n$ が存在する場合，$W_i=(Y_i, x_i)$ とし，$g_i(\theta)=g(\theta;W_i)$ と定義する．例えば，$Y=(Y_1,\dots, Y_n)$ が独立であれば，擬似スコアはこの条件を満たす．また，観測値 $y=\{y_1,\dots, y_n\}$ からのブートストラップ標本を $Y^*=\{Y_1^*,\dots, Y_n^*\}$ とすると，$Y_1^*,\dots, Y_n^*$ は i.i.d. である．GLM のような場合，観測値 y は一般に同じ分布からの実現値ではないことに注意する．

推定方程式 $g(\theta;y)=0$ の解集合の中で，一致推定量が存在すれば，漸近的にユニークであることが知られている．いま，$g(\theta;y)=0$ に対し，複数の解が存在するものとする．我々の目的は，この一組の標本 $y=\{y_1,\dots, y_n\}$ に基づいて，帰無仮説，H_0: $\hat{\theta}$ が一致推定量であること，を検定したいとする．ただし，$g(\hat{\theta};y)=0$ とする．

いま，ブートストラップ擬似尤度を

$$\exp[\phi(\theta, \xi, Y^*)]\propto\exp\left[\frac{1}{2}(\theta-\xi)' B_\xi'(\theta-\xi)+g'(\xi, Y^*)(\theta-\xi)\right] \tag{4.5}$$

と定義する．ただし，$-2B_\xi^*=\dot{g}(\xi;Y^*)+\dot{g}'(\xi;Y^*)$ である．同様に，ブートストラップ擬似尤度比統計量を次のように定義する

$$\gamma(\hat{\theta}, \xi;Y^*)=-2[\phi(\hat{\theta}, \xi, Y^*)-\phi(\tilde{\theta}, \xi, Y^*)]. \tag{4.6}$$

ただし，$\tilde{\theta}=\mathrm{argmax}_\theta\{\phi(\hat{\theta}, \xi, Y^*)\}=\xi-B_\xi^{*-1}g(\xi;Y^*)$ である．ここで，パラメータ ξ の自然な候補として，$\xi=\hat{\theta}$ を採用する．この場合，検定統計量 (4.6) は

$$\gamma(\hat{\theta}, \hat{\theta};Y^*)=g'(\hat{\theta};Y^*)(-B_{\hat{\theta}}^{*-1})g(\hat{\theta};Y^*) \tag{4.7}$$

と書ける．データ $y=\{y_1,\dots, y_n\}$ が既知のもとで，ブートストラップ推定関数 $g(\hat{\theta};Y^*)$ は不偏であり，分散は $\hat{\Sigma}=\sum_{i=1}^n g(\hat{\theta};y_i)g'(\hat{\theta};y_i)$ となる．また，$(-B_{\hat{\theta}}^*)$ のブートストラップ平均は $\hat{\Gamma}=-[\dot{g}(\hat{\theta};y)+\dot{g}'(\hat{\theta};y)]/2$ である．したがって，推定関数 g に対し，前節と同様な正則条件を仮定すると，ブートストラップ検定統計量 $\gamma(\hat{\theta}, \hat{\theta};Y^*)$, (4.6), は漸近的に $\sum_{i=1}^p\lambda_i Z_i^2$ に収束する．ただし，$(\lambda_i)=\mathrm{eigen}(\hat{\Sigma}\hat{\Gamma}^{-1})$ である．

帰無仮説のもとで，すなわち，$\hat{\theta}$ が一致推定量であるとき，大数の法則より，$\gamma(\hat{\theta}, \hat{\theta};Y^*)$ が $\sum_{i=1}^p\lambda_i Z_i^2$ に収束し，$(\lambda_i)=\mathrm{eigen}(\Sigma\Gamma^{-1})$ で，

$$\Sigma=E_{\theta_0}g(\theta_0, y)g'(\theta_0, y),\qquad \Gamma=-[E_{\theta_0}\dot{g}(\theta_0, y)+E_{\theta_0}\dot{g}'(\theta_0, y)]/2$$

である．したがって，g が情報不偏であれば，$\Sigma=\Gamma$ より，ブートストラップ擬似尤度比統計量の帰無分布は，漸近的に自由度 p のカイ 2 乗分布に従う．

4.4 最適ポテンシャル

前節までは，不偏推定関数 $g(\theta, Y)$ について議論してきた．いま，不偏推定関数 $g=g(\theta)=g(Y;\theta)$ のクラス $G=\{g(\theta)|E_\theta g(\theta)=0\}$ を考える．任意の定数の正則行列 $A=A(\theta)$ に対し，$Ag\in G$, すなわち，Ag も不偏推定関数である．また，g と Ag の根の集合が同じである．しかし，g と Ag は異なるベクトル場を定義し，それぞれの誘導される Helmholtz 型ポテンシャルも一般に異なる．この節では，不偏推定関数 $g(\theta, Y)$ の誘導される Helmholtz 型ポテンシャルの族 Ag の中で，最適なポテンシャルについて議論する．

定数の正則行列 A が存在し，不偏推定関数 $h(\theta)=Ag(\theta)$ が積分可能とする．すなわち，スカラー関数 $\phi(\theta)$ が存在し，$h(\theta)=\dot{\phi}(\theta)$ が成り立つ．ポテンシャル $\phi(\theta)$ が対数尤度であれば，バートレットの恒等式 (Bartlett (1953), McCullagh (1987))，$E\dot{\phi}(\theta)=0$, $E[\ddot{\phi}(\theta)+\dot{\phi}(\theta)\dot{\phi}'(\theta)]=0$, が成り立つ．すなわち，

(4.8) $$E_\theta h(\theta)=0\,,\qquad E_\theta[\dot{h}(\theta)+h(\theta)\,h'(\theta)]=0$$

が成立する．これは，推定関数 $h(\theta)$ が不偏かつ情報不偏の定義である．一般に，$h(\theta)=Ag(\theta)$ が可積分になる定数の正則行列 A が存在するとは限らない．しかし，$h(\theta)=Ag(\theta)$ が近似的に積分可能とすれば，(4.8) を局所的に満たすポテンシャルが存在する．

任意の $g\in G$ に対し，標準化操作

$$\mathcal{S}(g)\equiv-(E\dot{g})'(Egg')^{-1}g=[\mathrm{var}^{-1}(I_g^{-1}g)]\,I_g^{-1}g$$

によって，$h=\mathcal{S}(g)$ は情報不偏推定関数になる．ここで，$I_g=-E_\theta\dot{g}(\theta)$ である．すなわち，$E_\theta h(\theta)\,h'(\theta)=-E_\theta\dot{h}(\theta)$ が成り立つ．したがって，不偏かつ情報不偏推定関数の集合は

$$\mathcal{S}(G)=\{\mathcal{S}(g)|g\in G\}$$

となることが分かる．また，$I^{-1}_{\mathcal{S}(g)}\mathcal{S}(g)=I_g^{-1}g$ が成り立ち，$\mathcal{S}[\mathcal{S}(G)]=\mathcal{S}(G)$ になることも確かめられる．

任意の不偏推定関数 $g\in G$ と定数の正則行列 A に対し，$Ag\in G$ なので分散 $\mathrm{var}(g)$ は意味を持たない．しかし，情報不偏推定関数 $h=\mathcal{S}(g)\in\mathcal{S}(G)$ の分散は $\mathrm{var}(h)=\mathrm{var}^{-1}(I_g^{-1}g)$ となり，$-E[\partial(I_g^{-1}g)/\partial\theta]=U$ (単位行列) である．規格分散

$$\mathrm{GEC}(g)=\mathrm{var}(h)=\mathrm{var}^{-1}(I_g^{-1}g)=(E\dot{g})'(Egg')^{-1}E\dot{g}$$

は，推定関数 h または g の Godambe 効率規準と呼ばれている．ここで，最適推定関数を，$h^*=\mathrm{argsup}_{h\in\mathcal{S}(G)}\mathrm{GEC}(h)$ で定義する．1 次元の場合，不偏推定関数 $h^*\in\mathcal{S}(G)$ が最適であるための必要十分条件は

(4.9) $$E_\theta h(\theta)\,h^*(\theta)=-E_\theta\dot{h}(\theta),\qquad\forall\,h(\theta)\in\mathcal{S}(G)$$

である (Small and McLeish (1994), p. 72)．他の同値条件について，Godambe and Heyde (1987)，Heyde ((1997), Chapter 2) などを参照されたい．不偏推定関数 $g(\theta,\ Y)$ から最適化操作によって，情報不偏推定関数 $h(\theta,\ Y)=-E_\theta\dot{g}'(\theta,Y)[E_\theta g(\theta,Y)\,g'(\theta,Y)]^{-1}g(\theta,Y)$ が得られる．推定関数 $h(\theta)$ に対応する Helmholtz ポテンシャルは，Severini (1998) の意味での 1 次局所対数尤度になっていることが確かめられる．2 つの推定方程式の解集合が一致するため，$h(\theta,\ Y)$ から誘導されるポテンシャルに基づく検定方式が望ましい．推定関数 $h(\theta,\ Y)$ が平均的勾配場を成していることにも注意する．

不偏推定関数 $g(\theta,\ Y)=\sum_{i=1}^n g(\theta,\ Y_i)$ から誘導される情報不偏推定関数を $h(\theta,\ Y)$ とし，h に対応するブートストラップ擬似尤度比統計量を $\gamma_h=\gamma(\hat{\theta},\hat{\theta}\,;\,Y^*)$ とする．前節と同様に，統計量 γ_h は $\sum_{i=1}^p\lambda_iZ_i^2$ に収束する．ただし，$(\lambda_i)=\mathrm{eigen}(\Sigma\Gamma^{-1})$ で，

(4.10) $$\begin{aligned}\Sigma&=\left\{C\left[\sum_{i=1}^n g(\hat{\theta},\ y_i)\,g'(\hat{\theta},\ y_i)\right]C'\right\}_{\theta=\hat{\theta}},\\ 2\Gamma&=\left\{\frac{\partial}{\partial\theta}\sum_{i=1}^n[Cg(\theta,\ y_i)+g'(\theta,\ y_i)\,C']\right\}_{\theta=\hat{\theta}}\end{aligned}$$

である．ただし，$C=E_\theta\dot{g}'(\theta,Y)[E_\theta g(\theta,Y)\,g'(\theta,Y)]^{-1}$ である．標準化操作を適用するためには，係数 C を計算しなければならない．帰無仮説のもとで，すなわち，$\hat{\theta}$ が一致推定量のとき，$E_\theta\dot{g}(\theta,Y)$, $E_\theta g(\theta,Y)\,g'(\theta,Y)$ をそれぞれのブートストラップ平均，$E_y\dot{g}(\hat{\theta},Y^*)=\dot{g}(\hat{\theta},y)$，$E_yg(\hat{\theta},Y^*)\,g'(\hat{\theta},Y^*)=\sum_{i=1}^n g(\hat{\theta},y_i)\,g'(\hat{\theta},y_i)$ で近似すると，(4.10) 式より，$\Sigma=\Gamma$ を得る．ゆえに，γ_h の帰無分布は $\chi^2(p)$ で近似される．推定量 $\hat{\theta}$ の一致性と大数の法則を使えば，(4.10) から直接 $\Sigma=\Gamma$ を得ることもできる．しかし，ブートストラップ近似を通じた議論から

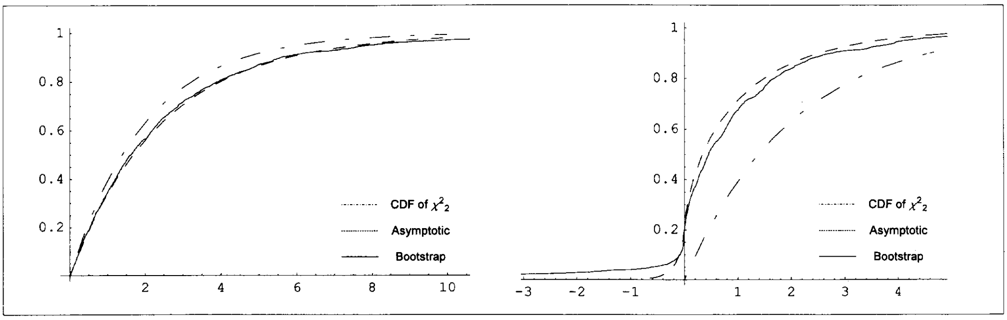

図1. 擬似尤度比統計量の，それぞれの根が一致推定量である，という帰無仮説に対するブートストラップ分布，漸近近似分布 $\sum_{i=1}^{p}\lambda_i Z_i^2$，および帰無分布である $\chi^2(2)$ 分布（5節参照）.

分かるように，カイ2乗分布は良い近似であることが分かる．近似の良さは次節の例においても確かめられている．

5. 応用：測定誤差がある場合のロジスティック回帰モデル

モデル $\mathrm{logit}(\pi)=\alpha+\beta' x$ における $\theta=(\alpha,\beta)'$ の推定を考える．ここで，π は2値反応変数 Y の平均で，x は共変量である．更に，共変量 x の観測ができず，補助変量 $Z=x+\epsilon$ に関する測定しか得られないとする．ただし，誤差 ϵ は Y と独立，$\epsilon\sim N_{p-1}(0,\Psi)$ と仮定し，共分散行列 Ψ を既知とする．撹乱母数 x に対する完備十分統計量 $A=z+y\Psi\beta$ に基づき，Stefanski and Carroll (1987) では条件付きスコア $\sum_{i=1}^{n}(1,x_i)'(y_i-\mu_i^c)$ を導出している．ここで，$\mu_i^c=\{1+\exp\{-[\alpha+(A_i-(1/2)\,\Psi\beta)'\beta]\}\}^{-1}$ は条件付き平均である．条件付きスコアから撹乱母数 x を消去し，Hanfelt and Liang (1995) では推定関数

$$u(\theta)=\sum_{i=1}^{n}(1,d_i)'(y_i-\mu_i^c) \tag{5.1}$$

を提案している．ただし，$d_i=A_i+(\mu_i^c-1)\,\Psi\beta$ である．推定関数 (5.1) は不可積分で，しかも多重根が存在する．Hanfelt and Liang (1995, 1997) はパスに依存する線積分を使って多重根の問題を調べている．Wang (1999) では擬似尤度および擬似尤度比の性質について考察している．

図1では，一組の標本，$(y_i,z_i), i=1,\ldots,100$，に対する擬似尤度比統計量のブートストラップ分布とその漸近分布を示している．通常はこの例のように，標本 y と共変量 z がペアをなしているため，今までのブートストラップ標本も (Y^*,Z^*) と理解しなければならない．真のパラメータの値は $(\alpha,\beta)=(-1.4,1.4)$ とし，また測定誤差を規定するパラメータを $\Phi=.8$ とした．この標本に対して，二つの根 $\hat{\theta}_1=(-1.1,\,2.3)$ と $\hat{\theta}_2=(-1.0,\,8.8)$ が得られた．図1の左と右は，それぞれ，帰無仮説，H_0：$\hat{\theta}_i$ が一致推定量であること，に対してのブートストラップ分布 $(\hat{F}_i)$ と，その漸近分布 $\sum_{j=1}^{2}\lambda_{ij}Z_{ij}^2$ を示してある．ブートストラップ分布は $B=2000$ 回の計算によって得られたものである．また，$(\lambda_{11},\lambda_{12})=(1.57,0.91)$，$(\lambda_{21},\lambda_{22})=(0.97,-0.12)$ であった．上のグラフでは，帰無分布である $\chi^2(2)$ も示してある．分布 $\hat{F}_1$ が $\chi^2(2)$ に近いことより，$\hat{\theta}_1=(-1.1,2.3)$ が $(\alpha,\beta)=(-1.4,1.4)$ のより合理的な推定量であることが分かる．

推定値 $\hat{\theta}_i$ $(i=1,2)$ は推定関数 (5.1) の誘導される力学系の不動点としてみることができる．推定関数の情報不偏性が近似的に成立すれば，一致推定量は，漸近的なアトラクターに対

応する．一致推定量でないものは典型的に鞍点である．この例では，推定値 $\hat{\theta}_1$ がアトラクターで， $\hat{\theta}_2$ は鞍点である．このことから， $\hat{\theta}_2$ の不合理性が再び確証される．

謝　辞

原稿を丁寧に読んでくださった統計数理研究所の栗木哲さん，日本語をチェックしてくれた千葉大学大学院の桜井裕仁君に深謝する．また，二人の査読者から建設的な修正意見を頂いた．ここに記して謝意を表したい．

参 考 文 献

Barendregt, L. G. and van Pul, M. C. (1995). On the estimation of the parameters for the Littlewood model in software reliability, *Statist. Neerlandica*, **49**, 165-184.

Bartlett, M. S. (1953). Approximate confidence intervals, *Biometrika*, **40**, 12-19.

Cox, D. R. (1981). Statistical analysis of time series, some recent developments, *Scand. J. Statist.*, **8**, 93-115.

Darling, R. W. R. (1994). *Differential Forms and Connections*, Cambridge University Press, Cambridge.

Diggle, P. J., Liang, K.-Y. and Zeger, S. L. (1994). *Analysis of Longitudinal Data*, Oxford Statistical Science Series 13, Clarendon Press, Oxford.

Fahrmeir, L. and Tutz, G. (1994). *Multivariate Statistical Modelling Based on Generalized Linear Models*, Springer Series in Statistics, Springer, New York.

Ferguson, T. S. (1982). An inconsistent maximum likelihood estimate, *J. Amer. Statist. Assoc.*, **77**, 831-834.

Fisher, R. A. (1925). Theory of statistical estimation, *Proceedings of the Cambridge Philosophical Society*, **22**, 700-725.

Godambe, V. P. and Heyde, C. C. (1987). Quasi-likelihood and optimal estimation, *International Statistical Review*, **55**, 231-244.

Hanfelt, J. J. and Liang, K.-Y. (1995). Approximate likelihood ratios for general estimating functions, *Biometrika,* **82**, 461-477.

Hanfelt, J. J. and Liang, K.-Y. (1997). Approximate likelihood for generalized linear errors-in-variables models, *J. Roy. Statist. Soc. Ser. B*, **59**, 627-637.

Heyde, C. C. (1997). *Quasi-likelihood and Its Application: A General Approach to Optimal Parameter Estimation*, Springer, New York.

Huzurbazar, V. S. (1948). The likelihood equation, consistency and the maxima of the likelihood function, *Annals of Eugenics*, **14**, 185-200.

LeCam, L. (1979). Maximum likelihood: An introduction, *Lecture Notes in Statist.*, No. 18, University of Maryland, College Park, Maryland.

Lehmann, E. L. (1983). *Theory of Point Estimation*, Wiley, New York.

Li, B. and McCullagh, P. (1994). Potential functions and conservative estimating functions, *Ann. Statist.*, **22**, 340-356.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models, *Biometrika*, **73**, 13-22.

Lovelock, D. and Rund, H. (1975). *Tensors, Differential Forms, and Variational Principles*, Wiley, New York.

McCullagh, P. (1983). Quasi-likelihood function, *Ann. Statist.*, **11**, 59-67.

McCullagh, P. (1987). *Tensor Methods in Statistics*, Chapman and Hall, London.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd ed., Chapman and Hall, London.

McLeish, D. L. and Small, C. G. (1992). A projected likelihood function for semiparametric models, *Biometrika*, **79**, 93-102.

Nelder, J. A. and Wedderburn, W. M. (1972). Generalized linear models, *J. Roy. Statist. Soc. Ser. A*, **135**, 370-384.

Poston, T. and Stewart, I. N. (1976). *Taylor Expansions and Catastrophes*, Pitman, London.

Poston, T. and Stewart, I. N. (1978). *Catastrophe Theory and Its Applications*, Dover, Mineola.

Severini, T. A. (1998). Likelihood functions for inference in the presence of a nuisance parameters, *Biometrika*, **85**, 507-522.

Small, C. G. and McLeish, D. L. (1994). *Hilbert Space Methods in Probability and Statistical Inference*, Wiley, New York.

Stefanski, L. A. and Carroll, R. J. (1987). Conditional scores and optimal scores for generalized linear measurement-error models, *Biometrika*, **74**, 703-716.

Stuart, A. and Ord, J. K. (1991). *Kendall's Advanced Theory of Statistics, Vol. 2, Classical Inference and Relationship*, Edward Arnold, London.

Wang, J. (1999). Nonconservative estimating functions and approximate quasi-likelihoods, *Ann. Inst. Statist. Math.*, **51** (to appear).

Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method, *Biometrika*, **61**, 439-447.

Zeger, S. L. (1988). A regression model for time series of counts, *Biometrika*, **75**, 621-629.

Artificial Likelihoods for General Nonlinear Regressions

Jinfang Wang

(The Institute of Statistical Mathematics)

This paper concerns nonlinear regression models based on estimating functions. A general estimating function, $g(\theta)$, is typically nonconservative, that is, $g(\theta)$ is not the gradient of any scalar function. In such cases, neither quasi-likelihood nor quasi-likelihood ratio can be uniquely defined. In this paper we study the problem of nonconservative estimating functions and the associated difficulties in general linear regression. We propose semi-parametric inference approach based on artificial likelihood functions derived from vector field decomposition associated with estimating functions. Further properties of Helmholtz-type quasi-likelihood proposed by Wang (1999) are studied. In particular, we propose a method for root-selection based on bootstrap quasi-likelihood ratio. The method is applied to logistic regression with measurement error model.

Key words: Bootstrap, estimating function, generalized linear model, logistic regression with measurement error, multiple roots, vector field.