数理解析研究所講究録
第682巻 1989年 1-10

擬似尤度とその一般化について

慶応義塾大学理工学部数理科学科　椿　広計（Hiroe Tsubaki）

1．はじめに

Wedderburn(1974)は一般線形模型をデータYにフィットする際，データの確率分布が曖昧にしか特定できない場合のために”擬似尤度”を提案した．多くの場合擬似尤度は，古典的なＧＬＩＭで想定している”散らばり母数ϕを含む拡張された指数分布族”(1.1)の真尤度に対応する．

$$f(y;\theta,\phi)=\exp[\{y\theta-b(\theta)\}/\phi+c(\theta,\phi)] \tag{1.1}$$

ここにθは，自然母数である．　Wedderburn の擬似尤度を改良しようという試みは，McCullagh(1983),Firth(1987),Nelder and Pregibon(1987)等多い．

Godambe and Heyde(1987)は，擬似尤度に基づく推論は本質的に”最良推定方程式”に基づく推論と等価であると主張した．p次元未知母数ベクトル$\boldsymbol{\theta}$を推定するためにk個の不偏な推定関数$f_j(\boldsymbol{\theta};\mathbf{y})$が利用可能（$j=1,...,k$）としよう．ここで$\mathbf{f}=(f_j)$の期待値は$\mathbf{0}$，分散共分散行列は$\Sigma$とする．

Godambe and Heyde の一般化された擬似尤度表現とは，擬似有効評点関数ベクトルを

$$\mathbf{g}(\boldsymbol{\theta};\mathbf{y})=B'\Sigma^{-1}\mathbf{f}(\boldsymbol{\theta};\mathbf{y}) \tag{1.2}$$

とすることである．但し，Bはヤコビ行列の期待値であり，そのi-j成分は

$$E[\partial f_j/\partial\theta_i]$$

である．

(1.2)がWedderburn,McCullaghの擬似尤度の拡張となっているのは，$\boldsymbol{\theta}$を$\mathbf{y}$の期待値ベクトルとれば明らかである．Godambe and Heyde(1987)は，推定方程式$\mathbf{g}=\mathbf{0}$　がGodambe(1960),Bhapker(1972)の推定方程式の効率の意味で$\mathbf{f}$の最適な線形結合であることを示している．

Godambe and Heyde(1987)の論文は Wedderburnの擬似尤度のある側面を摘出す

るのに成功したといえるが, Wedderburnの思想はそれ以上のものを含んでいると考える. すなわち, 分布(1.1)は分散関数Var[Y]=ϕV(θ)が特定された際の期待値母数μ=E[Y]に関する最小情報量分布なのである. 従って, μの擬似最尤推定量は, 分散関数が同一の形をしている分布族の尤度を用いたM推定量の族の中で漸近的にミニマックス推定量となる. これが, 統計的推論において分布族(1.1)が汎用される最大の原因であると信じる.

以下では擬似尤度の拡張を議論するが, 擬似有効評点関数にGodambe達の意味での最良性のみならず, ”与えられた母数空間の制約”の下で”関心のある母数”に関して”最小情報的”であることを要請する. このような性質を満たす擬似尤度を”最小情報擬似尤度”と呼ぶことにする.

2. 最良推定方程式の正規直交系による表現

Cox and Hinkley(1974)には”ピボット量”の概念が紹介されているが, これは確率変数と母数の関数として定義される量であり, かつその確率分布が母数に依存しない量を意味する. Morton(1981)は, この概念を弱めて単にその期待値と分散共分散行列が母数に依存しないものを pivotlike quantity と呼んだが, ここでは”擬似ピボット”と呼ぶことにする.

仮定1

擬似ピボットの正規直交系 $\{\Psi_i(y,\boldsymbol{\theta})\}$, i=0,...,p, が導入可能であることを仮定しよう. すなわち, 推論の対象となる分布族Fは, 任意のf∈Fに関して,

$$E_f[\Psi_i\Psi_j] = \delta_{ij} \qquad (2.1)$$

を満たすものとする. 但し, $\Psi_0 = 1$ と定義する.

仮定2

更に,

$$g_{ik}(\theta) = E[\partial\Psi_i/\partial\theta_k] \quad i=0,...,p,\ k=1,...,m \qquad (2.2)$$

は, θの既知関数となることを仮定しよう.

これらの仮定の下で, i.i.d.の観測値Y_j, j=1,...,nが与えられれば, Ψから導

かれるθの最良推定方程式は,

$$\sum_{j=1}^{n} Q_k(y_i, \theta) = 0 \qquad k=1,...,m \qquad (2.3\text{-a})$$

と与えられる．ここで，Q_k は，θ_k の”擬似有効評点関数”で次のように定義される．

$$Q_k(y, \theta) = -\sum_{i=0}^{p} g_{ik}(\theta) \ \Psi_i(y, \theta) \qquad (2.3\text{-b})$$

適当な正則条件の下では推定関数Q_k は，次のような有効評点関数と類似した性質を持つことが容易に示される．

$$E[Q_k] = 0$$

$$\mathrm{Cov}[Q_k, Q_{k'}] = -E[\partial Q_k / \partial \theta_{k'}]$$

推定方程式(2.3)を推定方程式の解の周りにテーラー展開すれば，正則条件の下で”推定方程式推定量”の漸近分布は次のようになる．

$$n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \sim Np(\mathbf{0}, I^{-1}(\boldsymbol{\theta}))$$

ここで，$I(\boldsymbol{\theta})$ はQ_kの分散共分散行列であり，擬似情報行列と呼ぶことにする．

Godambe and Heyde(1987)が示唆した，推定方程式(2.3)の最適性は(2.3-b)が正則条件の下で，次のように表現可能なことから明らかである．

$$Q_k(\mathbf{y}, \boldsymbol{\theta}) = \sum_{i=1}^{p} E_f[\Psi_i \ S_k(\mathbf{y}, \boldsymbol{\theta})] \ \Psi_i \qquad (2.3\text{-c})$$

ここに，$S_k(\mathbf{y}, \boldsymbol{\theta}) = \partial \log f(\mathbf{y}, \boldsymbol{\theta}) / \partial \theta_k$，すなわち真の評点関数である．(2.3-c)式の意味するところは，Godambe and Heyde(1987)にも指摘されたように擬似有効評点Q_k が 真の有効評点S_k の最良線形近似となっているとい

うことである.

擬似有効評点の表現(2.3-b) は最小情報擬似尤度の導出の道具として用いられる.

3. Wedderburn の擬似尤度の拡張

Wedderburn の擬似尤度の直接的拡張を考える.

$\mu_i(\mu, \eta)$ は, y の期待値μ周りの i 次モーメント関数とする. ここにμは, 期待値母数であり唯一関心のある母数, ηは攪乱母数ベクトルとする. μ_i, i=1, ..., 2q が特定化されているが, 2q+1次以上のモーメント構造については, 存在するということ以外には何等情報がない状況を考察する. すなわち, 推論の対象となる分布族Fは

$$F = \{ f \mid E_f [(y-\mu)^i] = \mu_i(\mu, \eta), i=1, \ldots, 2q, \mu_i, i>2q \text{は存在} \}$$

となる. μに関する最小情報擬似有効評点は, 以下のような手順で定めることになる.

先ず、分布のモーメント関数が特定されていれば, 次のようなグラムシュミット直交化によって直交系を順次構成できることに注意されたい.

$$\Psi_0 = 1$$

$$\Psi_1 = (Y-\mu)/[\mu_2(\mu, \eta)]^{1/2}$$

$$\Psi_i = [(Y-\mu)^i - \sum_{k=0}^{i-1} c_{ik} \Psi_k]/[\mu_{2i} - (\sum_{k=0}^{i-1} c_{ik}^2)]^{1/2} \qquad (3.1)$$

$$c_{ji} = (\mu_{i+j} - \sum_{k=0}^{i-1} c_{ik} c_{jk})/[\mu_{2i} - (\sum_{k=0}^{i-1} c_{ik}^2)]^{1/2}$$

この直交化を用いれば, Ψ_j, $j \leq q$ までの直交基底は容易に構成できることが分かる. なぜならば, Ψ_j には 2j+1次以上のモーメント関数は含まれないからである.

一方, Ψ_j, $j > q$ の決定は如何にすべきであろうか? 実は, ここに推定関数Q_μが最小情報的となること(Q_μの分散が最大となること)を要請するのが今回の

報告の骨子である．すなわち，

$$E\,[\,\partial\Psi_j/\partial\mu\,]=0,\ \text{for } j>q \qquad (3.2)$$

を要請する．(3.1)式をμで偏微分して期待値をとり，要請(3.2)を課すと2q+1次以上のモーメント関数が順次定まる．すなわちモーメント関数については次の微差分方程式が成立する．

$$i\,\mu_{i-1}-\partial\mu_i/\partial\mu-\sum_{j=1}^{q}c_{ij}\ E[\partial\Psi_j/\partial\mu]=0\quad \text{for } i>q \qquad (3.3)$$

ここでc_{ij}はi+j次のモーメント関数までを含む．また，$E[\partial\Psi_j/\partial\mu]$は，２節の仮定２より既知関数である．従って，微差分方程式(3.3)は$\mu_{i-1},\ldots,\mu_{i+q-1}$を与えた下で$\mu_{i+q}$を順次定める規則となっていることが分かる．特に，i=q+1とすれば，(3.3)は，$\mu_{q-1},\ldots,\mu_{2q}$から$\mu_{2q+1}$を定める式となっていることが分かる．

このようにして，事前には特定しなかった次数のモーメント関数が順次定まる．そして，これらのモーメント関数から正規直交系Ψ_j，j>qも定まる．従って，μ，ηの擬似有効評点も(2.3)式に従って導出される．

(3.3)式で決定される高次モーメント構造を持つ確率分布f_*が実在すれば，その分布はFの中でμに関して最小情報分布であることは，分布族Fに属する任意の分布に対応する直交基底の$\Psi_0,\ldots,\Psi_q$が等しいことから明らかである．

4．応用１：正規分布からの拡張

３節の方法を一例として正規分布型の擬似尤度の拡張に用いよう．Wedderburnの擬似尤度で正規分布に対応するのは，$\mu_2=\phi$の場合であった．

ここでは更に，３次，４次モーメントについても以下のように想定する．

$$\mu_2=\sigma^2$$

$$\mu_3=\tau_1\ \sigma^3$$

$$\mu_4=(\tau_2+3)\,\sigma^4$$

このとき，正規直交基底を４項まで示すと次のようになる．

$$\phi_0 = 1 \tag{4.1-a}$$

$$\phi_1 = (y-\mu)/\sigma \tag{4.1-b}$$

$$\phi_2 = \{(y-\mu)^2-\sigma^2(1+\tau_1\phi_1)\}/(\sigma^2\sqrt{\tau_2-\tau_1^2+2}) \tag{4.1-c}$$

$$\begin{aligned}\phi_3 = {} & [(y-\mu)^3-\sigma^3\{\tau_1+(\tau_2+3)\phi_1+(\tau_2/\tau_1)\sqrt{\tau_2-\tau_1^2+2}\ \ \phi_2\}] \\ & \div [\sigma^3\sqrt{\{(\tau_2-\tau_1^2+2)(2\tau_2-3\tau_1^2)/\tau_1^2\}}]\end{aligned} \tag{4.1-d}$$

$$\begin{aligned}\phi_4 = {} & [(y-\mu)^4-\sigma^4\ \{(\tau_2+3)+\frac{\tau_2(\tau_2+2)+4\tau_1^2}{\tau_1}\ \phi_1 \\ & \qquad + \frac{\tau_2(\tau_2+2)\sqrt{\tau_2-\tau_1^2+2}}{\tau_1^2}\ \phi_2 \\ & \qquad + \frac{2\{\tau_2(\tau_2+2)-2\tau_1^2(\tau_2+3)\}\sqrt{\tau_2-\tau_1^2+2}}{\sqrt{\tau_1^2(2\tau_2-3\tau_1^2)}}\ \phi_3\}\ ] \\ & \div\ [\sigma^4(\tau_2-\tau_1^2+2)\ \{2\ \frac{6(\tau_2-3\tau_1^2)-\tau_2^2}{\tau_1^2(2\tau_2-3\tau_1^2)}\ \}^{1/2}]\end{aligned} \tag{4.1-e}$$

従って μ，σ，τ_1，τ_2 の基底第４項までを用いた最良推定関数は次のようになる．

$$Q_\mu = (1/\sigma)\phi_1 - \{\tau_1/(\sigma\sqrt{\tau_2-\tau_1^2+2})\}\phi_2$$

$$\begin{aligned}Q_\sigma = {} & \{2/(\sigma\sqrt{\tau_2-\tau_1^2+2})\}\phi_2 \\ & - \frac{\mathrm{sgn}(\tau_1)\ \sqrt{(2\tau_2-3\tau_1^2)}}{\sigma\ \sqrt{(\tau_2-\tau_1^2+2)}}\ \phi_3 \\ & + \frac{2\tau_1(2-\tau_1^3)(\tau_2+3)+(\tau_1^2-1)\tau_2(\tau_2+2)}{\sigma\tau_1\ (\tau_2-\tau_1^2+2)\ \{2\ \frac{6(\tau_2-3\tau_1^2)-\tau_2^2}{\tau_1^2(2\tau_2-3\tau_1^2)}\}^{1/2}}\ \phi_4\end{aligned}$$

$$Q_{\tau_1} = \sqrt{[\tau_1^2/\{(\tau_2-\tau_1^2+2)(2\tau_2-3\tau_1^2)\}]}\ \phi_3$$

$$-\ \frac{\{\tau_2(\tau_2+2)-2\tau_1^2(\tau_2+3)\}\sqrt{2\tau_1^2}}{(\tau_2-\tau_1^2+2)\sqrt{[(2\tau_2-3\tau_1^2)\{6(\tau_2-3\tau_1^2)-\tau_2^2\}]}}\ \phi_4$$

$$Q_{\tau_2} = \frac{\sqrt{\{\tau_1^2(2\tau_2-3\tau_1^2)\}}}{(\tau_2-\tau_1^2+2)\sqrt{[2\{6(\tau_2-3\tau_1^2)-\tau_2^2\}]}}\ \phi_4$$

特にμの擬似情報量は次のようになる.

$$[(\tau_2+2)\ /(\tau_2-\tau_1^2\ +2)]/\sigma^2$$

さて、以下では応用例として、i.i.d. の場合の仮説H：$\tau_1 = \tau_2 = 0$ に対する擬似有効評点検定を導こう。

仮説のもとで、基底ϕ_2, ϕ_3, ϕ_4 は次のように表せる。

$$\phi_{H2} = \{(y-\mu)^2-\sigma^2\}/(\sigma^2\sqrt{2}) \tag{4.2-a}$$

$$\phi_{H3} = \{(y-\mu)^3-3\sigma^2(y-\mu)\}/(\sigma^3\sqrt{6}) \tag{4.2-b}$$

$$\phi_{H4} = \{(y-\mu)^4-\sigma^4(3+\ 6\sqrt{2}\ \ \phi_{H2}\}/(2\sigma^4\sqrt{6}) \tag{4.2-c}$$

(4.1)式と(4.2)式を見比べれば、仮説の近傍で次のような関係が成立していることが分かる。

$$\tau_2 = 3\tau_1^2 + 2\tau_1^4 + o(\tau_1^4)$$

従って、仮説のもとでの擬似有効評点は次のように表せる。

$$Q_{H,\mu} = (y-\mu)/\sigma^2$$

$$Q_{H,\sigma} = \{(\sqrt{2})/\sigma\}\ \ \phi_{H2}$$

$$Q_{H,\tau_1} = \phi_{H3}/\sqrt{6}$$

$$Q_{H,\tau_2} = \phi_{H4}/(2\sqrt{6})$$

また、仮説のもとでμ、σの擬似最尤推定量は次のようになる。

$$\hat{\mu} = \overline{y}$$

$$\hat{\sigma}^2 = \hat{\mu}_2$$

ただし、$\hat{\mu}_k = \Sigma(y_i - \overline{y})^k / n$ 。
以上より、擬似有効評点検定統計量は、次のようになる。この統計量は仮説のもとで、漸近的に自由度2のカイ二乗分布に従う。

$$\chi^2{}_D = n\{\hat{\mu}_3{}^2/(6\hat{\sigma}^6) + (\hat{\mu}_4 - 3\hat{\sigma}^4)^2/(24\hat{\sigma}^8)\} \qquad (4.3)$$

(4.3)式で与えられた検定統計量は D'Agostino and Pearson(1973)の提案した正規性の検定統計量の一つと一致する。

5．応用2：寿命分布

寿命データなどでは、データの平均値μよりも，分布のスソに興味がある場合が多い。このようなときには、期待値に関する最小情報擬似尤度を構成しても応用上不合理な評点関数が導かれる場合もあろう。そこで、例えばモーメント間に次のような構造を想定しよう。

$$\mu_1 = \theta - k\sigma s(\theta)$$

$$\mu_2 = \sigma^2 s(\theta)^2$$

ここで、kは解析者が与える定数である。
この構造の下で母数θに関する最小情報擬似尤度を構成する。これは、期待値より、標準偏差のk倍だけ大きい点の情報量を小さくすることになる。このような点における生存関数は、カンテリの不等式より$1/(1+k^2)$以下（$k>0$）である（$k<0$の場合には分布関数が$1/(1+k^2)$以下となる）。研究者がビー・テン・ライフを興味のある母数と考えるならば、kを－3以上の負値に選んでやれば良いであろう。

推論は、σとθ（稀にk）に関して行うので、基底は2項（3項）用意すれば十分である。興味のある母数θに関しては、基底第1項のみに情報量を持つとすれば、モーメント間の漸化式

$$\mu_{m+1} = \sigma^2 s^2 \{ m\mu_{m-1} + \mu_m / \theta / (1 - k\sigma s') \}$$

が得られる。

従って、

$$\mu_3 = 2\sigma^4 s^3 s' / (1 - k\sigma s')$$

$$\mu_4 = 3\sigma^4 s^4 + 2\sigma^6 s^2 (3s^2 s'^2 + s^3 s'' - 3k s^2 s'^3) / (1 - k\sigma s')^3$$

を得る。寿命分布は多くの場合、s（θ）＝θのものがよく使われるので、

$$\mu = (1 - k\sigma)\theta$$

$$\mu_2 = \sigma^2 \theta^2$$

$$\mu_3 = 2\sigma^4 \theta^3 / (1 - k\sigma)$$

$$\mu_4 = 3\sigma^4 \theta^4 \{ 1 + 2(1-k)\sigma^2 / (1 - k\sigma)^3 \}$$

を得る。もちろんk＝0の場合は、ガンマ分布のモーメント構造に一致する。k＞0では、ガンマ分布よりも右に歪み、k＜0では左に歪む。

観測値が、i.i.d. の場合には、上のθとσの擬似最尤推定量は,

$$\hat{\theta} = \bar{y} + kS$$

$$\hat{\sigma}^2 = S^2 / \hat{\theta}^2$$

となる。ただし、$\bar{y}$，Sは、それぞれ観測値の標本平均及び、標本標準偏差である。

参考文献

Bhapkar, V.P. (1972) On a measure of efficiency in an estimating equation. Sankhyā, A, **34**, 467-472.

Cox, D.R. and Hinkley, D.V. (1974) Theoretical Statistics, Chapman & Hall.

D'Agostino and Pearson, E. S. (1973) Tests for departure from normality. Empirical results for the distributions of b_2 and $\sqrt{b_1}$. Biometrika, **60**, 613-622.

Firth, D. (1987) On the efficiency of quasi-likelihood estimation. Biometrika, **74**, 233-245.

Godambe, V. P. (1960) An optimum property of maximum likelihood estimation. Ann. Math. Statist., **31**, 1208-1211.

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

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

Morton R. (1981) Optimal estimating equations with applications to insect development times. Austr. J. Statist., 204-213.

Nelder, J. A. and Pregibon, D. (1983) Quasi-likelihood models and data analysis. Bell Laboratories Techinical Memorandum.

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

椿　広計(1988)一般化線形模型の問題点と擬似尤度の一般化, 応用統計学, 17, 1-12.

椿　広計(1988)一般化擬似尤度とその線形推測論への応用, 東京大学工学部学位論文.