数理解析研究所講究録
第 507 巻 1983 年 60-74

R. A. Fisherにおける非正則推定論

竹内 啓
(Kei Takeuchi)

R. A. Fisherが統計的推定論に関連して，十分性，一致性，漸近有効性等の概念を導入して，正則な場合における最尤推定量の漸近有効性を主張したことはよく知られている。それは数学的には厳密な証明ではなかったにしても，その後多くの人々によって数学的に十分厳密な結果が確立されるに至る過程の出発点となった。また彼は推定量を用いることによる情報量損失の概念を用いて，これまた正則な場合には，一次の漸近有効性を持つ推定量の中で，最尤推定量が漸近的情報量損失を最小にすること，そういう意味で2次の漸近有効性を持つことを示した。この点は後にC. R. Raoによって改めて取り上げられ，そうして最近になってPfanzagl，Ghosh，Takeuchi-Akahiraらによる推定量の高次漸近有効性の理論として，完成された。このようにFisherは特に比較的初期の2論文：

Mathematical Foundations of Theoretical Statistics 1922年，および

Theory of Statistical Estimation 1925年

において，50年以上も後に完成されるような，多くの理論的

成果を予見しているが、その中で彼はまた正則条件がなり立たない場合に何が生ずるであろうかについても、多くの結果と予想を与えている。ここではそれを取り上げて検討することにしよう。

最初に推定理論における正則条件とは何を意味するかについてのべておこう。$X_1, X_2, \cdots X_n, \cdots$ を観測値の系列とするとき、それは通常次のことを意味する。

1） $X_1, X_2, \cdots X_n, \cdots$ は互いに独立に母数 $\theta (\in \Theta \subset R^p)$ をふくむ分布に従う。

2） 分布は密度関数 $f(x, \theta)$ を持つ。

3） 分布の台 support：$S = \{x \mid f(x, \theta) > 0\}$ は θ に無関係に定められる

4） $f(x, \theta)$ は θ に関して連続微分可能。

5） Fisher 情報量行列

$$I(\theta) = E_\theta \left\{ \frac{\partial}{\partial \theta} \log f(X, \theta) \frac{\partial}{\partial \theta'} \log f(X, \theta) \right\}$$

が有限。

このとき $\hat{\theta}_n$ を $(X_1, \cdots X_n)$ にもとづく最尤推定量とすると、$\sqrt{n}(\hat{\theta}_n - \theta)$ は n が大きくなるとき、漸近的に正規分布 $N(0, I^{-1})$ に従い、かつそれが一次の漸近有効性（定義はこ

こではのべない）を持つことが知られている。そうして更に $f(x,\theta)$ の θ に関する高次の連続微分可能性と $\partial^j \log f(x,\theta)/\partial\theta^j$ のモーメントの存在を仮定すると，適当に偏りを補正した最尤推定量が，2次ないし3次の漸近有効性を持つことが証明される（Akahira-Takeuchi［1］参照）。

そこでこれらの條件のうちの1つ，或いは2つ以上がなり立たないとき，推定量の漸近的な性質について，どういうことが起こり得るかをしらべることが，単に病理的なケースに対する興味というよりも，これらの正則條件の意義を明確にするために重要になる。

Fisherは上記の1922年の長大な論文の第8節で，$X_1, \cdots, X_n$ が次のようなガンマ分布に従う場合を取り上げている。

$$f(x,\mu,\nu,\alpha)=\frac{1}{\alpha\Gamma(\nu)}\left(\frac{x-\mu}{\alpha}\right)^{\nu-1}e^{-(x-\mu)/\alpha}:\quad x>\mu$$

Fisherはここで，母数の推定法としてモーメント法と最尤法の比較を行っている。

まず α，ν がともに既知の場合，モーメント法による μ の推定量は，

$$\hat{\mu}=\bar{X}-\alpha\nu$$

で与えられ，その分散は $\alpha^2\nu/n$ となる。

他方 $L = \sum \log f(X_i)$ として

$$\frac{\partial^2 L}{\partial \mu^2} = -(\nu-1)\sum \frac{1}{(X_i-\mu)^2}$$

の期待値を計算すると，$\nu > 2$ ならば

$$nI(\mu) = -E\left(\frac{\partial^2 L}{\partial \mu^2}\right) = \frac{n}{\alpha^2(\nu-2)}$$

となるから，最尤推定量の漸近分散は $\alpha^2(\nu-2)/n$ となり，従ってモーメント法の漸近効率は $(\nu-2)/\nu$ となる。

しかしこの議論にはやや問題がある。というのはここでは上記の条件 3) が破られているから，最尤推定量の漸近分布に関する議論をそのまま適用することはできないのである。しかしこの場合結論自体は正しいので，そのことをもう少しくわしくチェックしよう。まず

$$\frac{\partial L}{\partial \mu} = -(\nu-1)\sum \frac{1}{(X_i-\mu)} + \frac{n}{\alpha}$$

であって，$\mu \to \min X_i$ のとき $\partial L/\partial \mu \to -\infty$，$\mu \to -\infty$ のとき $\partial L/\partial \mu \to n/\alpha$，また上記の式から $\nu > 1$ ならば $\partial^2 L/\partial \mu^2 < 0$ であるから，$\partial L/\partial \mu = 0$ はつねに唯一の解 $\mu^* < \min X_i$ を持ち，それが最尤推定量を与える。そうして $\sqrt{n}(\mu^* - \mu)$ が漸近的に平均 0，分散 $1/I(\mu)$ の正規分布

に従うことは、通常の正則な場合と同様に、$\partial L/\partial\mu$ をテイラー展開することによって証明できる。次にこれが一次の漸近有効性を持つことは、仮説 $\mu=\mu_0$ を対立仮説 $\mu=\mu_0+t/\sqrt{n}$ に対して検定するとき、$\hat{\mu}^*$ にもとづく検定の検出力が、漸近的に最強力検定の検出力に一致することを示せばよい。

$\nu\leqq 2$ の場合については、Fisherは次のようにのべている。「$\nu=2$ のときには、位置母数について標本平均の効率は0になる。それは分布の端で密度関数がx軸と正の角をなしているからである。」とのべている。このことの意味づけについては後に他の例題との関連でのべているが、彼は $\nu<2$ の場合についても「標本平均の効率が0になる」ことには十分気づいていたと思われる。

実際 $\nu=2$ の場合には最尤推定量を $\hat{\mu}_n^*$ とすると $\sqrt{n\log n}\times(\hat{\mu}_n^*-\mu)$ が漸近的に正規分布に従うことが示されるので、その漸近分散のオーダーは $(n\log n)^{-1}$ になり、$\hat{\mu}$ の分散のオーダーより小さくなるのである。また $\nu<2$ の場合には $\hat{\hat{\mu}}=\min X_i-c_n$ という形のものを作ると、その分散のオーダーは $n^{-2/\nu}$ となるので、n^{-1} より小さくなるのである。更にここでもう一つ興味あることは、$\nu=2$ のときには最尤推定量は漸近有効性を持つことが示されるが(Woodroofe [4])

$\nu < 2$の場合には，それを持たないという点である．従って$\nu < 2$の場合について，Fisherがことさらのべることをさけたのは，その場合には最尤推定量の漸近的な性質について明確なことがいえないということを予感したためであるかもしれない．

そこで$\nu < 2$の場合についてもう少し考えてみよう．まず$\nu > 1$ならば，上記の議論により，最尤推定量$\hat{\mu}^*$が，尤度方程式$\partial L/\partial\mu = 0$の一義的な解として求められることがわかる．そこで$\partial L/\partial\mu$を展開すると

$$\frac{\partial L}{\partial\mu}\Big|_{\hat{\mu}^*} = 0 = \frac{\partial L}{\partial\mu}\Big|_{\mu} + \frac{\partial^2 L}{\partial\mu^2}\Big|_{\mu=\mu^0}(\hat{\mu}^* - \mu)$$

となる．ただしここでμ_0はμと$\hat{\mu}^*$の間の値である．

この場合でも，$\hat{\mu}^*$が一致性を持つことは明らかなので，μ^0をμでおきかえて

$$\hat{\mu}^* - \mu \sim -\left(\frac{\partial L}{\partial\mu}\right) \Big/ \left(\frac{\partial^2\mu}{\partial\mu^2}\right)$$

とすることができる．ところが$\nu < 2$のときは

$$\frac{\partial L}{\partial\mu} = \sum\left(-\frac{\nu-1}{(x_i-\mu)} + \frac{1}{\alpha}\right)$$

の各項について，分散が無限大となるから，中心極限定理は$\partial L/\partial\mu$について成立しない．そうして$n^{-\frac{1}{2}}(\partial L/\partial\mu)$の分布が

漸近的に指数 ν の stable law に収束することが証明される。また分母についても大数法則が成立せず，$n^{-\frac{2}{\nu}}(\partial^2 L/\partial\mu^2)$ はある確率分布を持ち，定数に確率収束しない。従って

$$n^{\frac{1}{\nu}}(\hat{\mu}-\mu) \sim \frac{n^{-\frac{1}{\nu}}(\partial L/\partial\mu)}{n^{-\frac{2}{\nu}}(\partial^2 L/\partial\mu^2)} \qquad (*)$$

は，正規分布とは異なる，非常に複雑な分布を持つ。

他方仮説 $\mu=\mu_0$ を対立仮説 $\mu=\mu_0+tn^{-\frac{1}{\nu}}$ に対して検定するときの最強力検定の棄却域は $t>0$ のときは，

$$\sum \log \frac{f(x_i,\mu_0+tn^{-\frac{1}{\nu}})}{f(x_i,\mu_0)} \sim tn^{-\frac{1}{\nu}} \frac{\partial L}{\partial\mu}\Big|_{\mu_0} > c$$

で与えられ，これは最尤推定量から得られる棄却域 $\hat{\mu}<\mu_0+c'n^{-\frac{1}{\nu}}$ とは，漸近的に一致しない。それが $(*)$ 式の右辺の分母が定数に確率収束しないことによるものであることは明らかであろう；$t<0$ のときは最強力検定の棄却域は

$$\min x_i < \mu_0+tn^{-\frac{1}{\nu}} \quad \text{および} \quad \sum \log \frac{f(x_i,\mu_0+tn^{-\frac{1}{\nu}})}{f(x_i,\mu_0)} > c$$

という2つの範囲から構成され，これも最尤推定量から得られる棄却域には一致しない。

このような場合には，一様に（一次の）有効性を持つ推定量は存在しない。

$\nu<1$ のときは $\partial L/\partial\mu$ はつねに単調増加関数となるから，

最尤推定量はμの可能な範囲の上限に一致し，$\hat{\mu}^* = \min X_i$となる．そうして$\nu = 1$の場合には$\min X_i$は十分統計量となるから，$\hat{\mu}^*$は（任意の次数の）漸近有効性を持つ．$\nu < 1$のときには$\hat{\mu}^*$は漸近有効性を持たないがその収束のorderは$n^{-\frac{1}{\nu}}$になるから，$\hat{\mu}$の漸近効率が0になることには変りはない．

次にνが既知で，μ，αがともに未知の場合には，モーメント法による推定量は，2次の標本モーメントをM_2と表すと，$M_2 = \hat{\alpha}^2 \nu$となるから$\hat{\alpha} = (M_2/\nu)^{\frac{1}{2}}$となりその漸近分散は

$$(\nu+3)\alpha^2/2\nu n$$

となる．他方情報量行列は

$$n\begin{pmatrix} 1/\alpha^2(\nu-2) & 1/\alpha^2 \\ 1/\alpha^2 & \nu/\alpha^2 \end{pmatrix}$$

となるから，αの最尤推定量の漸近分散は$\alpha^2/2n$となり，従ってモーメント法による推定量の漸近効率は$\nu/(\nu+3)$となる．この場合も，$\nu > 2$ならば，正則条件が満されていないにもかかわらず，このような結論が正しいことは証明できる．

ところでFisherは$\nu \leqq 2$の場合でもαの最尤推定量の漸

近分散が $\alpha^2/2n$ で与えられるようにのべているが，それは誤まりである。ν，μ がともに既知のとき α の最小分散不偏推定量は $\hat{\alpha}^{**}=\Sigma(X_i-\mu)/\nu$ で与えられ，その分散は $\alpha^2/\nu n$ だから，μ が未知のとき α の推定量の漸近分散はこれより小さくはなり得ないから $\nu<2$ のときその漸近分散は $\alpha^2/2n$ よりは大きくならなければならない。実際 $\nu<2$ のときは $\hat{\mu}^*=\min X_i$ とすると その分散の order は n^{-1} より小さくなるから，$\hat{\alpha}^*=\Sigma(X_i-\hat{\mu}^*)/\nu$ の漸近分散は，μ が既知の場合のそれに一致する。従ってモーメント法の漸近効率は $\nu<2$ のとき $2/(\nu+3)$ に等しくなる。$\nu=2$ の場合には状況はより微妙になるが，α，μ の最尤推定量を $\hat{\alpha}^*$，$\hat{\mu}^*$ とすると $\sqrt{n\log n}(\hat{\mu}^*-\mu)$，$\sqrt{n}(\hat{\alpha}^*-\alpha)$ が漸近的に正規分布に従い，かつ $\hat{\alpha}^*$ の漸近分散は μ 既知の場合のそれに一致することが示される。

これは2つの母数のうち，1つに関して正則條件が成立しないとき，そのことがもう一つの母数の推定量の漸近効率にどのように影響するかを示す例であるということができる。

Fisher は更に進んで3つの母数がすべて未知である場合を考えている。この場合モーメント法による ν の推定量は

$$4/\hat{\nu}=M_3/M_2^{3/2}$$

で与えられ，その漸近分散は

$$6\nu(\nu+1)(\nu+5)/n$$

となる。他方情報行列は $\nu>2$ ならば

$$\frac{n}{\alpha^2}\begin{pmatrix} \frac{1}{\nu-2} & 1 & \frac{\alpha}{\nu-1} \\ 1 & \nu & \alpha \\ \frac{\alpha}{\nu-1} & \alpha & \alpha^2\frac{d^2}{d\nu^2}\log\Gamma(\nu) \end{pmatrix}$$

となるので ν の最尤推定量の漸近分散は

$$\frac{1}{n}\,\frac{1}{\frac{d^2}{d\nu^2}\log\Gamma(\nu)-\frac{1}{\nu-1}+\frac{1}{2(\nu-1)^2}}$$

となり，ν が大ならば，この値はほぼ

$$\frac{6}{n}(\nu-1)\left(\nu^2-2\nu+\frac{6}{5}\right)$$

となる。従って $\nu\to\infty$ ならばモーメント法の漸近効率は 1 に近づく。しかし ν が小さいときには，それは極めて小さいとして，Fisher はいろいろな ν に対応する相対効率の値を計算している。しかしここでは彼は $\nu<2$ の場合まで，同じ式で計算してしまっているが，これは勿論正しくない。この場合 μ の推定量の分散の order が n^{-1} よりも小さくなることに注意しているが，しかしそのことが ν の推定に及ぼす影

響については，正しく認識していなかったように思われる．

正しくはこの場合次のようになる．μを既知とすると，α，νの同時推定において，情報行列は

$$n\begin{pmatrix} \nu/\alpha^2 & 1/\alpha \\ 1/\alpha & \frac{d^2}{d\nu^2}\log\Gamma(\nu) \end{pmatrix}$$

となるから，νの最尤推定量の漸近分散は

$$\frac{1}{n}\ \frac{1}{\frac{d^2}{d\nu^2}\log\Gamma(\nu)-\frac{1}{\nu}}$$

となる．μが未知の場合にも$\nu\leqq 2$ならば，νの最尤推定量の漸近分散はこの式から計算される値に一致することが証明できる．

更にFisherはこの論文の中でPearson系の分布型でかつ位置，尺度が未知の分布についての母数推定を論じているが，その中には分布の両端が切れている非正則の場合もふくまれている．とくに一様分布については，分布の中心の推定量として

$$\hat{\mu}^*=(\min X_i+\max X_i)/2$$

を取ると，$n(\hat{\mu}^*-\mu)$が漸近的に両側指数分布に従うこと，すなわち分散のorderはn^{-2}になること，他方$\hat{\mu}^0=\bar{X}$の分散はn^{-1}にすぎないこと，従ってその漸近効率は0になるこ

とそのべているが その際 $\hat{\mu}^*$ の「よさ」については特に何もいっていない。

Fisher は 1925 年の論文で，最尤推定量が漸近的に有効であることを示したのみならず，更にその２次の漸近効率についても論じている．いま $X_1, X_2, \cdots X_n$ が互いに独立に同じ分布に従うとき，その情報量を $I(\theta)$，一方 $\hat{\theta}_n = \hat{\theta}_n(X_1, \cdots X_n)$ を推定量，$\hat{\theta}_n$ の分布の持つ情報量を $I_{\hat{\theta}_n}$ とするとき つねに

$$I_{\hat{\theta}_n} \leqq n I(\theta)$$

であるが $n \to \infty$ のとき

$$I_{\hat{\theta}_n} / nI \longrightarrow 1$$

となるとき $\hat{\theta}_n$ は（一次の）漸近有効性を持つという。これがふつうの意味の漸近有効性に一致することは証明できる．ところで Fisher は更に２次の漸近効率を考え，

$$\lim_{n\to\infty} nI - I_{\hat{\theta}_n}$$

を $\hat{\theta}_n$ の漸近情報量損失と呼んだ そうして正則な場合には漸近情報量損失が

$$I(\theta)\, \nabla \left\{ \frac{\partial^2}{\partial\theta^2} \log f(x,\theta) \,\middle|\, \frac{\partial}{\partial\theta} \log f(x,\theta) \right\}$$

で与えられることを示し，かつそれが一次漸近有効推定量の中で最小の情報量損失を与えることを示している．この辺り

のFisherの議論は極めて不明確であるが，その大筋の結論は誤まっていない。その論理は後にRaoによってはじめて明確にされたものである

ところでこのような最尤推定量の2次し或いはPfanzagl，Takeuchi-Akahiraの議論によれば3次）の有効性がなり立つためには，密度関数がθに関して2回微分可能であることが必要である。そうして2次微分可能性がなり立たないとき，最尤推定量の2次の漸近効率に関する上記の結論がなり立たないことを，Fisherは次の例で示している。

いま$X_1 \cdots X_n$が互いに独立に両側指数分布：

$$f(x,\theta)=\frac{1}{2}\exp -|x-\theta|$$

に従うとすると，θの最尤推定量は$\hat{\theta}^* = \mathrm{med}\, X_i$となる。そうしてこの場合

$$\frac{\partial}{\partial\theta}\log f(x-\theta) = -\mathrm{sgn}(x-\theta)$$

であるから，情報量は$I=1$であり$\sqrt{n}(\hat{\theta}^*-\theta)$は漸近的に平均0分散1の正規分布に従うから，$\hat{\theta}^*$は一次の漸近有効性を持っている。ところでこの場合$\hat{\theta}^*$の情報量損失は

$$l_n = V\left(\frac{\partial}{\partial\theta}L \mid \hat{\theta}^*\right)$$

で与えられるが，ここで$\partial L/\partial\theta = -\Sigma\, \mathrm{sgn}|X_i-\theta|$であるこ

とを考えると

$$l_n = 2V(\# X_i < \theta \mid \mathrm{med}\, X_i)$$

これは具体的に次の式で与えることができる

$$n - \frac{n(n+1)}{n-3}\left\{1 - \frac{(n-1)!}{\left(\frac{n-1}{2}!\right)^2 2^{n-2}}\right\}$$

そうしてFisherはこの量が漸近的に

$$\sqrt{n/\pi} - 4$$

になること、すなわちその大きさが正則の場合とは異なって $n^{\frac{1}{2}}$ の order になることを示した。Fisherはこの例を他の二三の論文でもくり返し引用している。

この場合、$\hat{\theta}^*$ は漸近情報量損失に関して2次の有効性を持たないこと、すなわち

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}}(nI - I_{\hat{\theta}_n})$$

を最小にするものでないことを示すことができる。それはこの場合 $\hat{\theta}_n^*$ の漸近分散が $1/(n-l_n)$ になることを示し他方

$$\hat{\theta}_n^{**} = \frac{1}{2}\left\{X\left[\frac{n}{2} - t\sqrt{n}\right] + X\left[\frac{n}{2} + t\sqrt{n}\right]\right\}$$

として、t を適当にえらべば、その漸近分散が $1/(n-c_n)$ という形で与えられ、かつ $\lim (l_n - c_n)/\sqrt{n} > 0$ となるようにすることができるからである。そうしてこのとき $\hat{\theta}_n^{**}$ の

情報量については $I_{\hat{\theta}_n^{**}} \geq 1/V(\hat{\theta}_n^{*}) = n - c_n$ となるから

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}} (I_{\hat{\theta}_n^{**}} - I_{\hat{\theta}_n^{*}}) > 0$$

となることが示されるのである。

このことはTakeuchi-Akahira [3]が示したように、この例については最尤推定量がふつうの意味の2次漸近有効性を持たないという事実に対応するものであるといえよう。

Fisherは非正則な場合の議論においても、鋭い直観と深い洞察力を持っていたことが知られるのである

References

[1] Akahira,M. and Takeuchi,K.(1981). Asymptotic Efficiency of Statistical Estimators: Concepts and Higher Order Asymptotic Efficiency. Lecture Notes in Statistics 7, Springer, New York.

[2] Fisher,R.A.(1950). Contributions to Mathematical Statistics. Wiley, New York.

[3] Takeuchi,K. and Akahira,M.(1976). On the second order asymptotic efficiencies of estimators. Proceedings of the Third Japan-USSR Symposium on Probability Theory, Lecture Notes in Mathematics 550, 608-638, Springer, Berlin.

[4] Woodroofe,M.(1972). Maximum likelihood estimation of a translation parameter of a truncated distribution. Ann. Math. Statist., 43, 113-122.