統計数理 (1999)
第 47 巻 第 1 号 71-79

特集「統計基礎理論」
[研究ノート]

標識実験における Paulik-Seber 推定に対する考察
── 局外母数が存在する時の補助性・十分性から ──

東京水産大学* 山田作太郎・北門 利英

(受付 1998 年 10 月 1 日；改訂 1998 年 11 月 24 日)

要　旨

不完全報告を含む標識実験による資源量の推定問題を考える．このモデルには，報告率および資源量の 2 つのパラメータが含まれる．この問題に対する既存の研究では，前者は条件付分布，後者は周辺分布を基に推定が議論されてきた．この論文では，これらの推定方法に対する正当化を，局外母数が存在する時の補助性・十分性という観点から行った．

キーワード：補助性，十分性，局外母数，不完全観測，標識実験．

1. 標識実験と問題の所在

水産資源解析における一つの有力な方法に標識実験法がある．魚に標識をつけて A 地点から放流し，その後 B 地点で捕獲されたとすると，その期間の A から B への移動がわかる．放流時に体長や体重を測定しておくと，捕獲までの期間におけるそれらの成長がわかる．本稿では標識実験による資源量推定を取り上げ，そこに現れる一つの統計的問題について考察を与える事を目的とする．

ある漁場にN尾の魚がいるとし，これを推定する事を考える．M尾捕獲して標識をつけて放流し，その後の漁獲機会に n 尾捕獲した時，その中にX尾の標識魚がいたとする．漁業を通しての捕獲データなら，n も確率変数のはずであるが，ここでは n は所与とみなし，Xのみを確率変数と考える．

次の諸条件を仮定する．

(ア) 母集団 (漁場にいる魚全体) は閉である．つまり，死亡，移入，逸散等による資源尾数の変動はない．
(イ) 特に標識をつける事で標識魚が死亡することはない．
(ウ) 標識の脱落はない．
(エ) 再捕された標識魚は全て試験機関等に報告される．
(オ) n 尾の捕獲は単純ランダムサンプリングである．

これらの諸条件の下で，Xの分布は超幾何分布

* 資源管理学科：〒108-8477 東京都港区港南 4-5-7.

$$P(X=x)=\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \tag{1.1}$$

で与えられる．よく知られている様に，Nの最尤推定値は

$$\hat{N}_{ML}=\left[\frac{nM}{x}\right] \tag{1.2}$$

である．ただし，[ ]はガウス記号である．[ ]を取った

$$\hat{N}=\frac{nM}{x} \tag{1.3}$$

はPetersen推定値と呼ばれている．

(1.2) または (1.3) が合理的な推定であるためには，上の(ア)〜(オ)が満たされている事が条件となる．しかし，実際にはどれも実現が難しい条件である．これらの制約のない時の推定理論が作れれば良いのだが，それは難しい．本稿では，条件(エ)がくずれた場合のみにおいて資源量の推定問題を考える．条件(エ)は，近年の遊漁者の増加による標識魚の持ち帰りや，水揚げされる市場での標識再捕魚の見落としによって，多くの場合において満たされない．

この問題に対して Paulik (1961) は捕獲を2つのグループに分け，第1のグループからは標識再捕魚の完全な報告が得られるが，第2のグループでは再捕された標識魚が報告される確率(報告率と呼ぶ)はすべての魚に共通で ρ とし，これを未知とした．グループ i $(i=1, 2)$ での捕獲尾数を n_i，その中の標識魚尾数を X_i，そのうちの報告尾数を R_i とする．但し，グループ1では報告が完全であるので $X_1=R_1$ となる．観測値 R_1, R_2 から報告率 ρ と資源尾数N（もしくは $\tau=M/N$ =母標識率）を推定する．ここに，Mは既知の放流した標識魚尾数である．Paulik (1961) と Seber (1982) は，R_1，R_2 は独立とし，

$$R_1\sim Po(n_1\tau), \qquad R_2\sim Po(n_2\tau\rho)$$

なるモデルを仮定し，$R=R_1+R_2$ が与えられた時の R_1 の条件付分布

$$P(R_1=r_1|R=r)=\binom{r}{r_1}\left(\frac{n_1}{n_1+n_2\rho}\right)^{r_1}\left(\frac{n_2\rho}{n_1+n_2\rho}\right)^{r-r_1} \tag{1.4}$$

から得られるモーメント法による推定方程式

$$\frac{r_1}{r}=\frac{n_1}{n_1+n_2\rho}$$

を ρ について解き，

$$\hat{\rho}=\frac{r_2/n_2}{r_1/n_1}$$

を得た．そして，観測値のない X_2 の実現値を $\hat{x}_2=r_2/\hat{\rho}$ で推定し，2つのグループのデータを合わせて Petersen 推定から

$$\hat{N}=\frac{(n_1+n_2)M}{r_1+r_2/\hat{\rho}}=\frac{n_1M}{r_1}$$

と推定した．$\hat{N}$ は2つのグループそれぞれからNの Petersen 推定を求め，捕獲尾数の割合によ

る重み付き平均

$$\frac{n_1\left(\dfrac{n_1 M}{x_1}\right)+n_2\left(\dfrac{n_2 M}{\widehat{x}_2}\right)}{n_1+n_2} \tag{1.5}$$

としても求まることに注意する．本稿では，この様な推定法を Paulik-Seber 推定，得られる推定量を Paulik-Seber 推定量と呼ぶことにする．なお，3 節ではNの代わりに主にτを用いて議論する為，Nとτの関係から得られる$\widehat{\tau}=r_1/n_1$をτの Paulik-Seber 推定量と呼ぶことにする．

ここで問題は，

(P-1)　何故Rで条件づけてρを推定するのか

(P-2)　Nの推定に完全な報告のあった第 1 のグループだけからのデータを用い，不完全ではあっても第 2 のグループからのデータを用いないのは何故か

という点である．なお，ρとNの最尤推定値は北門・山田 (1998) により

$$\widehat{\rho}_{ML}=\begin{cases}\dfrac{r_2/n_2}{r_1/n_1} & \text{if} \quad r_2/n_2<r_1/n_1\,,\\ 1 & \text{if} \quad r_2/n_2\geq r_1/n_1\,,\end{cases} \tag{1.6}$$

$$\widehat{N}_{ML}=\begin{cases}\dfrac{Mn_1}{r_1} & \text{if} \quad r_2/n_2<r_1/n_1\,,\\ \dfrac{M(n_1+n_2)}{r_1+r_2} & \text{if} \quad r_2/n_2\geq r_1/n_1\,,\end{cases} \tag{1.7}$$

で与えられる．

これらの問題へのアプローチとして，山田・北門 (1997) による「不完全データにおける粗雑化のメカニズムの無視可能性」という視点，および北門・山田 (1998) によるシミュレーションによる方法からの考察がある．本稿では「局外母数が存在する時の補助性・十分性」の視点からこの問題を考察する．

2. 局外母数が存在する時の補助性・十分性

局外母数が存在する時の補助性・十分性の問題は，推測の対象とする興味ある母数に対する補助性・十分性をいかに定義すべきかという点で古くから統計家の関心事である．

局外母数がある時の補助性・十分性の定義は密度関数の分解を通して定義される．つまり，ネイマン分解の拡張ないしは類似の形で定義される．十分性の定義では周辺密度が興味ある母数のみを含み，条件付密度は全母数を含むとする．局外母数がない時は，十分性の定義では条件付分布は (興味ある) 母数を含まない．この意味では局外母数が存在する時の十分性の定義は，それがない時の通常の十分性の定義の拡張になっていない．しかし，条件付密度に基づく統計モデルから興味ある母数に関する推測は合理的になされない，あるいはこのモデルが興味ある母数に関して情報を含まない事を条件としてつければ，結果的には充分に十分性の定義の資格が与えられるのである．従って，問題はこの条件付モデルが興味ある母数に関して持つ情報 (量) をどう定義するかである．この条件付モデルは興味ある母数の他に局外母数を含むからである．なお，例えば十分性の定義をしようとする時，周辺密度は興味ある母数のみを，条件付密度は局外母数のみを含む (尤度独立性) とするという考え (例えば Fraser (1956)) もあり

うるが少し強すぎると思われる．補助性の定義は条件付密度が興味ある母数のみを含むとし，周辺密度は興味ある母数と局外母数の両方を含むが，この周辺密度に基づく統計モデルは興味ある母数に関して情報を含まないと定義するのである．ここでは前節の問題に関係すると思われる Barndorff-Nielsen (1973, 1978), Godambe (1980) による定義を取り上げる．

一般に $p(x;\theta)$ は観測する確率変数Xの分布の確率密度関数または確率関数とする．母数 θ は $\theta=(\theta_1,\theta_2)\in\Omega_1\times\Omega_2=\Omega$ なる形（この時 θ_1 と θ_2 は distinct であるという）とする．ここに θ_1 は興味ある母数，θ_2 は局外母数とする．また，Tを統計量とする．まず，Barndorff-Nielsen による定義を与える．

定義 2.1. (Barndorff-Nielsen (1973)) 統計量Tに対して

(C-1) $p(x;\theta)=p(t;\theta)\,p(x;\theta_1|t)$ なる形に分解可能

(C-2) 任意の θ_1 に対して $\{p(t;\theta)=p(t;\theta_1,\theta_2);\theta_2\in\Omega_2\}$ は universal である，すなわち，任意の t に対してある θ_2 が存在して

$$p(t;\theta_1,\theta_2)\geq p(t';\theta_1,\theta_2)\qquad \text{for all } t' \tag{2.1}$$

が成り立つ時，Tは θ_1 に対して M-補助であるという．ただし，$p(t;\theta)$ はTの周辺密度，$p(x;\theta_1|t)$ は $T=t$ を与えた時のXの条件付密度である（以下この様なただし書きは省略する）．

この (C-2) は次のように解釈される．θ_2 が既知の時，$p(t;\theta_1,\theta_2)$ が一様分布でなければ，任意に固定した θ_1 に対して頻度の高い実現値と頻度の低い実現値が存在する．また，θ_1 の値の変化に対して，この様な頻度の高い実現値も変化し，逆にそれが θ_1 についての情報となって θ_1 に対する推測が可能となる．ところが θ_2 が未知の時 (C-2) の下では，どのような θ_1 に対しても，任意に固定した値 t が最も実現し易い値となるように θ_2 の値を選ぶ事ができる．仮に，Tが θ_1 に対して情報をもつとすれば，Tの実現値 t から見て適当な θ_1 の値と不適当な θ_1 の値を識別する事が可能で，そのような不適当な θ_1 に対しては $T=t$ の頻度も低くなると考えるのが自然であろう．しかし (2.1) の下では，すべての θ_1 に対して $T=t$ の頻度が最大となる．これは，Tが θ_1 に対して情報をもつ事に矛盾する．つまり，局外母数 θ_2 が θ_1 に関する推論を攪乱しているのである．従って，θ_2 が未知という状況の下では，$T=t$ を観測しても θ_1 に対する情報が得られない．これが (C-2) の意味することであり，Tが θ_1 に対して M-補助となる理由である．

従って，統計量Tが M-補助である時は，(C-1) から想像されるように，局外母数を含まない条件付モデル $\{p(x;\theta_1|t);\theta_1\in\Omega_1\}$ に基づき θ_1 の推論をすることが良いと期待される．

定義 2.2. (Barndorff-Nielsen (1973, 1978)) 統計量Tに対して

(C-3) $p(x;\theta)=p(t;\theta_1)\,p(x;\theta|t)$ なる形に分解可能

(C-4) 任意の θ_1 と t に対して $\{p(x;\theta|t)=p(x;\theta_1,\theta_2|t);\theta_2\in\Omega_2\}$ は universal が成り立つ時，Tは θ_1 に対して M-十分であるという．

一方，Godambe による定義では universality なる条件が完備性に置き換えられている．

定義 2.3. (Godambe (1980)) 統計量Tに対して

(C-1) $p(x;\theta)=p(t;\theta)\,p(x;\theta_1|t)$ なる形に分解可能

(C-5) 任意の θ_1 に対して $\{p(t;\theta)=p(t;\theta_1,\theta_2);\theta_2\in\Omega_2\}$ は完備

が成り立つ時，Tは θ_1 に対して Godambe の意味で補助（G-補助と呼ぶことにする）であると

いう.

Godambe は，この完備性の方が universality より数学的に取り扱い易いと述べ，さらに universality の統計的な弱点を示す例をあげている．Godambe のそれ以前の推定関数に関する研究 (Godambe (1976)) に基づいた定義でもあると考えるのが最も自然な様に見える．Godambe の定義の最大の特徴は，それを情報量に結びつけて議論できる事である．しかし，Fisher 情報量を考えるにしても局外母数が存在する時の興味ある母数に関する情報量をどの様に定義するかが問題である．Godambe は局外母数の値が既知 (θ_{20}) の時の興味ある母数 θ_1 に関する Fisher 情報量の類推から，

$$(2.2) \qquad \mathcal{I}(\theta_1) = \max_{G_1} \frac{\left\{E_\theta\left(\dfrac{\partial g}{\partial \theta_1}\right)\right\}^2}{E_\theta(g^2)}, \qquad G_1 = \{g(x\,;\,\theta_1)\,;\, E_\theta(g) = 0 \quad \forall\,\theta \in \Omega\}$$

により局外母数を無視して $p(x\,;\,\theta)$ の中での θ_1 に対する情報量を定義した．ただし，E_θ は $p(x\,;\,\theta)$ に関する期待値である．この時，Godambe は

(1)　T が θ_1 に対して G-補助ならば

$$(2.3) \qquad \mathcal{I}(\theta_1) = E_\theta(\mathcal{I}_t(\theta_1))$$

を示した．ここに，$\mathcal{I}_t(\theta_1)$ は $T = t$ を与えた時の条件付分布 $p(x\,;\,\theta_1|t)$ において (2.2) と同様に定義される θ_1 に関する情報量であり，E_θ は $p(t\,;\,\theta)$ に関する期待値である．(2.3) は，T が θ_1 に対して G-補助の時，T の条件付分布モデル $\{p(x\,;\,\theta_1|t)\,;\,\theta_1 \in \Omega_1\}$ から θ_1 を推論するときの情報量は平均的に $\mathcal{I}(\theta_1)$ に等しい事を意味する．従って，条件付分布のモデルから θ_1 の推論をしても「情報損失」は無いと直感的に想像される．このことをより強く主張するために，Godambe は

(2)　任意の $\theta_1 \in \Omega_1$ に対して $\{p(t\,;\,\theta_1, \theta_2)\,;\,\theta_2 \in \Omega_2\}$ が完備ならば，$\mathcal{I}^t \equiv 0$

を示した．ここに，$\mathcal{I}^t(\theta_1)$ は T の分布における，(2.2) と同様に定義される θ_1 に関する情報量である．

定義 2.4. (Godambe (1980)) 統計量 T は θ_1 に対して G-補助とする．また，T の任意に固定した値 t に対して，U_t は $\{p(x\,;\,\theta_1|t)\,;\,\theta_1 \in \Omega_1\}$ において通常の意味での十分統計量とする．この時，U_T の周辺分布が θ_1 のみに依存するならば，U_T は $\{p(x\,;\,\theta)\,;\,\theta \in \Omega\}$ において，θ_2 を無視して，θ_1 に対する十分統計量 (G-十分と呼ぶことにする) という．

T が θ_1 に対して G-補助とし，$T = t$ に対する十分統計量 U_t の確率分布における情報量を $\mathcal{I}_t^{u_t}(\theta_1)$ と表すと，

$$(3) \qquad \mathcal{I}(\theta_1) = E_\theta(\mathcal{I}_t^{u_t}(\theta_1))$$

が成立する．ここに，E_θ は T の分布に関する期待値である．(3)は (2.3) と U_t が十分統計量であることから明らかである．また，U_T が G-十分である時，U_T と T は独立となる (Godambe (1980))．従って，$U_T = U$ として，(3)より

$$(4) \qquad \mathcal{I}(\theta_1) = \mathcal{I}^u(\theta_1)$$

が成り立つ．

この様にみるとGodambeの定義の方が，Barndorff-Nielsenによる定義よりも情報量的考察が可能という点ですぐれている様にみえる．勿論，文献上では他にも多くの方法で局外母数が存在する時の補助性，十分性の定義が与えられている．これら全てを紹介し，比較検討する事は本稿の目的ではないが，少しだけ紹介しておく．

(ア) Fraser (1956) や Hájek (1967) による局外母数が存在する時の十分性の定義の下では，制約的ではあるが本質的完備類定理が示される．十分統計量の最も重要な性質は，その統計量に基づく決定関数のクラスが本質的完備類をつくる事であるのでこの結果は大切である．しかしそれ故にこの定義に課される条件は強いものとなる．

(イ) Godambeの補助性の定義の下では，周辺分布に基づく統計モデルは，θ_1 に対する情報を含まない ((2))．Liang (1983) では，指数型分布族に対して，この性質が局外母数が存在する時の補助性と同値になる様に，補助性と情報量の定義が与えられている．この情報量の定義はさらにGodambe (1984) によって拡張された．これらの情報量の関係，および多次元の母数への拡張はBhapkar (1989) で述べられている．

3. 補助性・十分性の観点からみたPaulik-Seber推定

1節で述べた様に観測値 (R_1, R_2) の分布は

$$p(r_1, r_2; \tau, \rho) = \frac{e^{-n_1\tau}(n_1\tau)^{r_1}}{r_1!}\frac{e^{-n_2\tau\rho}(n_2\tau\rho)^{r_2}}{r_2!} \tag{3.1}$$

である．本来，τ の定義域は

$$\left\{1, \frac{M}{M+1}, \frac{M}{M+2}, \ldots\right\} = \Omega_\tau \tag{3.2}$$

であるが，便宜上 $(0, 1]$ とする．

この時，いずれも簡単な証明により次が示される．

Case A. ρ が局外母数で τ が興味ある母数の時

(A-1) R_2 は τ に対して M-補助
(A-2) R_2 は τ に対して G-補助
(A-3) R_1 は τ に対して M-十分
(A-4) R_1 は τ に対して G-十分

Case B. τ が局外母数で ρ が興味ある母数の時

(B-1) R_1 は ρ に対して M-補助ではない
(B-2) R_1 は ρ に対して G-補助ではない
(B-3) R_2 は ρ に対して M-十分ではない
(B-4) $R = R_1 + R_2$ は ρ に対して M-補助
(B-5) $R = R_1 + R_2$ は ρ に対して G-補助
(B-6) R_1 または R_2 は $\{p(r_1, r_2; \rho|r); 0 < \rho \leq 1\}$ に対して通常の意味で十分

ただし，Case B の (B-5) で，τ の定義域を (3.2) とすると通常のポアソン分布の完備性は適用されず別途考えねばならない．また，Case A，B で十分とある所を Fraser (1956) や Hájek (1967) の意味での十分性にとると，それらは成立しない．

$R = R_1 + R_2$ の補助性により，ρ の推定は 1 節で述べた様に，r で条件付けられた (R_1, R_2) の分布を基に行うことが望ましい．これにより，ρ に対する Paulik - Seber 推定が正当化され，(P-1) が解けた．また，R_2 は τ に対して G-補助，および M-補助であり，しかも (R_1, R_2) が独立であることから，τ の推定は R_1 の分布を基に行うことが望ましい．従って，この場合も τ に対する Paulik - Seber 推定が正当化され，(P-2) が解けた．

これまでの議論には τ の定義域を $(0,1]$ と仮定した．しかし，τ の定義域を (3.2) とした場合においても，任意の ρ に対して

$$\{p(r\,;\,\tau,\,\rho)\,;\,\tau \in \Omega_\tau\} \tag{3.3}$$

が完備であれば，全く同様の結果が成り立つ．以下にこの完備性を示す．

いま，

$$\sum_{r=0}^{\infty} f(r) \frac{e^{-(n_1+n_2\rho)\tau}\{(n_1+n_2\rho)\,\tau\}^r}{r!} = 0 \qquad \text{for all } \tau \in \Omega_\tau \tag{3.4}$$

の成立を仮定する．上式は，

$$f(0) + \tau g(\tau) = 0 \tag{3.5}$$

と同値である．ここで，

$$g(\tau) = \sum_{r=1}^{\infty} f(r) \frac{(n_1+n_2\rho)^r \tau^{r-1}}{r!}$$

であり，この級数は収束する．この時，任意の $\varepsilon > 0$ に対して $\delta > 0$ が存在し，$|\tau| < \delta$ ならば $|\tau g(\tau)| < \varepsilon$ が成立する．従って，(3.5) より $|f(0)| < \varepsilon$ が成立し，ε が任意であることから，$f(0) = 0$ となる．以下，この論法を繰り返して，$0 = f(1) = f(2) = \cdots$ を得る．よって，(3.3) の完備性が示された．

以上の考察から，Paulik-Seber の推定量の良さが局外母数が存在する時の補助性・十分性の考察から，ある程度は示されたといえる．

謝　辞

貴重な御意見と御指摘を頂いた査読者の方に感謝申し上げます．

参 考 文 献

Barndorff-Nielsen, O. (1973). On M-ancillarity, *Biometrika*, **60**, 447-455.

Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory*, Wiley, New York.

Bhapkar, V. P. (1989). Conditioning on ancillary statistics and loss of information in the presence of nuisance parameters, *J. Statist. Plann. Inference*, **21**, 139-160.

Fraser, D. A. S. (1956). Sufficient statistics with nuisance parameters, *Ann. Math. Statist.*, **27**, 838-842.

Godambe, V. P. (1976). Conditional likelihood and unconditional optimum estimating equations, *Biometrika*, **63**, 277-284.

Godambe, V. P. (1980). On sufficiency and ancillarity in the presence of a nuisance parameter, *Biometrika*, **67**, 155-162.

Godambe, V. P. (1984). On ancillarity and Fisher information in the presence of a nuisance parameter, *Biometrika*, **71**, 626-629.

Hájek, J. (1967). On basic concepts of statistics, *Proceedings of Fifth Berkeley Symposium*, 139-162, University of California Press, Berkeley.

北門利英, 山田作太郎 (1998). 不完全報告を含む標識実験における推定問題について, 日本水産学会誌, **64**, 775-781.

Liang, K.-Y. (1983). On information and ancillarity in the presence of a nuisance parameter, *Biometrika*, **70**, 607-612.

Paulik, G. J. (1961). Detection of incomplete reporting of tags, *Journal of the Fisheries Research Board of Canada*, **18**, 817-832.

Seber, G. A. F. (1982). *The Estimation of Animal Abundance*, 2nd ed., Griffin, London.

山田作太郎, 北門利英 (1997). 局外母数が存在する時の十分性, 補助性 ― 不完全データに基づく推測の例から ―, 平成9年度科研費シンポジウム講演予稿集「統計的推測の基礎理論と応用の研究」, 1-14.

A Study on Paulik-Seber Estimation in a Tag Experiment — In Terms of Ancillarity and Sufficiency in the Presence of a Nuisance Parameter —

Sakutarō Yamada and Toshihide Kitakado

(Department of Fisheries Resource Management, Tokyo University of Fisheries)

The estimation problem of the population size of fish by a tag experiment including incomplete reports is considered. Two parameters in this problem, the reporting rate and the population size, were estimated based on the conditional and marginal distribution, respectively, in the previous works. In this paper, justification for these estimation methods is given through some notions of ancillarity and sufficiency in the presence of a nuisance parameter.

Key words: Ancillarity, sufficiency, nuisance parameter, incomplete observation, tag experiment.