目視による放流群の死亡係数推定における 統計モデルの検討

北 田 修 一

(1996年9月5日受付)

Examination on Statistical Models to Estimate the Mortality Rate of Released Seeds from Sighting Survey Data

Shuichi Kitada*1

This article examines statistical models to estimate the natural mortality rate, the sighting rate, and the dispersion parameter of released seeds from sighting survey data. Besides the multinomial and the log linear models already proposed, three additional models of a revised multinomial, a conditional binomial, and a weighted least squares type model are considered. The properties of these five models were examined by means of simulations. A total of 100 sets of simulated sighting data were generated, and point estimates were obtained by the five models for each data set, respectively. Mean values and SDs of the point estimates for the 100 simulations for the five models were then compared. The weighted least squares type model gave the best estimates among the five statistical models. Although the original log linear model underestimated the dispersion parameter, the model with a revised weight accurately estimated all parameters.

キーワード：種苗放流，目視，死亡係数，統計モデル，比較，最尤法，加重最小2乗法

放流後の連続的な目視観測によって得られた個体数から、放流種苗の生残率を推定する方法が提案されている。¹⁾ 熟練した同じ人間が継続して目視を行えば、目視によって確認される個体数は生残個体数の相対値として利用できる。この方法は近年技術開発試験が行われているシマアジの銅付け型栽培漁業²⁾において適用されている¹⁾ほか、マダイやトコブシなどでもその適用が検討されている。

この方法¹⁾では、時刻 $t=0$ で N 尾が放流され、任意の時刻 t_i (日, $i=1, 2, \dots, k$) で滞留している個体 N_i のうち n_i 尾が確率 P_i で目視される場合を取り扱う。目視観測個体数 n_i に多項分布を仮定し、over-dispersion²⁾を考慮した統計モデルが用いられている。死亡係数などのパラメータは最尤法あるいは加重最小2乗法によって推定される。¹⁾ しかし、ここで扱う目視データの場合、多項分布モデルで前提とする $n_{k+1}=N-n$, $P_{k+1}=1-\sum_{i=1}^k P_i$ が成り立たない。ここで、 $n=\sum_{i=1}^k n_i$ すなわち総目視個体数である。最小2乗法では $(k+1)$ 項目は計算に用いないのでこの点は問題にならないが、最尤法

では多項分布の適用が困難となる。このように、多項分布の仮定は適当ではないが、この問題に対する修正処理として、実際の計算では $n_{k+1}=N-n$, $P_{k+1}=1-P_k$ とモデルを変更してパラメータを推定してきた。このモデルでは、 $(k+1)$ 項目を $n_{k+1}=N-n$ としており、 k 回目の観測までの死亡が考慮されていない。

この方法を放流シマアジの目視データに適用した結果、死亡係数と目視率の点推定値は最尤法と最小2乗法ではほとんど同じであったが、dispersion パラメータは最尤法では大きく、加重最小2乗法では小さく推定された。¹⁾ 最近、2項分布の積を正規近似したモデルが考えられ、同じデータを解析した結果、死亡係数と目視率の点推定値は従来のモデルでの結果とはほとんど変わらないが、dispersion パラメータがさらに大きく推定された。²⁾ このように、特に dispersion パラメータの推定値が統計モデルによって大きく異なっており、これらの方法を適用しようとする場合に混乱が予想される。本論文ではこれらの統計モデルの妥当性を検討する。

*1 東京水産大学 (Tokyo University of Fisheries, Konan, Minato, Tokyo 108, Japan).

*2 平成元年度銅付け型栽培漁業技術開発報告書(1). 日本栽培漁業協会, p. 86.

方 法

目視確率と誤差 種苗放流を行った後、放流個体が比較的狭い範囲に滞留する状況では、筏などの飼付け基盤の上などの至近距離から目視が可能である。目視者は、目でみた尾数の情報からその時点での滞留尾数の相対値を決定する。同一の目視者が毎回同じ方法で観測するならば、目視できる放流個体の割合を一定とすることは許されよう。

観測期間中の逸散を含んだ自然死亡係数 M は一定と仮定し、すべての調査日において目視率 α は一定とすると、目視個体数 n_i の期待値は

$$E[n_i] = N\alpha e^{-Mt_i} = NP_i$$

で与えられる。目視個体数 n_i はこの期待値の周りでランダムに変動する。目視が単純ランダムサンプリングで行われる場合は、 n_i の誤差は $NP_i(1-P_i)$ であるが、実際にはその日の目視確率のふれも含まれたものになっており、目視個体数は通常大きく変動する。この変動は dispersion パラメータの中に含まれていると考えられる。

対数線形モデル 対数線形モデルは最小 2 乗法を念頭においたもので、モデルは

$$\log \left(\frac{n_i}{N} \right) = \log \alpha - Mt_i + \sqrt{\phi} E_i \quad (1)$$

である。¹⁾ ここで、 $E_i = \log(1 + \varepsilon_i / NP_i)$ はデータの確率変動を表す。 E_i の平均は $E(E_i) = 0$ 、 $(n_1, n_2, \dots, n_k, n_{k+1})$ が多項分布に従う時、分散共分散は近似的に $V(E_i) \simeq (1-P_i)/(NP_i)$ 、 $\text{Cov}(E_i, E_j) \simeq -1/N$ となるが、³⁻⁵⁾ over-dispersion を考慮したこのモデルでは分散共分散は ϕ 倍に膨らんでいる。 ϕ は dispersion パラメータで、前報¹⁾ では σ^2 としていたものである。 E_i の分散共分散の構造からわかるように N が大きい時は共分散 $-\phi/N$ は無視され、計算の簡単な加重最小 2 乗法 (WLS) を用いることができる。 E_i は NP_i が大きいとき (10 以上^{3,4)})、近似的に正規分布 $N(0, \phi/NP_i)$ に従う。

回帰直線の傾き M は

$$\hat{M} = \frac{\sum_{i=1}^k W_i(t_i - \bar{t})(y_i - \bar{y})}{\sum_{i=1}^k W_i(t_i - \bar{t})^2}$$

により、切片 $\log \alpha$ は M の推定値 $\hat{M}$ を用いて $\widehat{\log \alpha} = \bar{y} - \hat{M}\bar{t}$

で推定される。ここで、

$$y_i = \log(n_i/N),$$

$$\bar{t} = \frac{\sum_{i=1}^k W_i t_i}{\sum_{i=1}^k W_i}$$

$$\bar{y} = \frac{\sum_{i=1}^k W_i y_i}{\sum_{i=1}^k W_i}$$

W_i は重みである。¹⁾

$\hat{M}$ の分散は、

$$\hat{V}(\hat{M}) = \frac{\hat{\phi}}{\sum_{i=1}^k W_i(t_i - \bar{t})^2}$$

で推定され、切片 $\widehat{\log \alpha}$ の分散は

$$\hat{V}(\widehat{\log \alpha}) = \frac{\hat{\phi} \sum_{i=1}^k W_i t_i^2}{\sum_{i=1}^k W_i \sum_{i=1}^k W_i(t_i - \bar{t})^2}$$

で推定される。ここで dispersion パラメータ ϕ は

$$\hat{\phi} = \frac{1}{k-2} \sum_{i=1}^k W_i(y_i - \widehat{\log \alpha} - \hat{M}t_i)^2 \quad (2)$$

により推定される。^{1,3,4)} $\hat{\alpha}$ の分散は $\widehat{\log \alpha}$ に正規分布を仮定すれば、対数正規分布の性質から

$$\hat{V}(\hat{\alpha}) = e^{2\widehat{\log \alpha} + \hat{V}(\widehat{\log \alpha})} (e^{\hat{V}(\widehat{\log \alpha})} - 1)$$

で評価される。^{1,6)} 推定値の実際の計算では、重み W_i を NP_i の実現値 n_i で推定して計算する。^{1,3,4)}

最尤推定 最尤推定では、多項分布モデル、¹⁾ その修正モデル、条件付き 2 項分布モデル²⁾ の三つのモデルを検討する。多項分布モデルは over-dispersion を考慮し、これを正規近似したモデルが用いられている。尤度は

$$L = \frac{1}{(2\pi N\phi)^{k/2} \left(\prod_{i=1}^{k+1} P_i \right)^{1/2}} \times \exp \left\{ -\frac{1}{2} \sum_{i=1}^{k+1} \frac{(n_i - NP_i)^2}{NP_i \phi} \right\} \quad (3)$$

である。^{1,5-7)} はじめに述べたように、実際の計算では $n_{k+1} = N - n_k$ 、 $P_{k+1} = 1 - P_k$ としている。これをモデル 0 とする。

モデル 0 を修正し、 k 回目の観測までの死亡を考慮するモデルを修正モデル 0.1 とする。すなわち (3) 式において、 $(k+1)$ 項目を $n_{k+1} = Ne^{-Mt_k} - n_k = Ne^{-Mt_k}(1-\alpha)$ 、 $P_{k+1} = e^{-Mt_k}(1-\alpha)$ とする。

条件付き 2 項分布の積の尤度は

$$L = \prod_{i=1}^k \binom{N_i}{n_i} \alpha^{n_i} (1-\alpha)^{N_i - n_i}$$

である。^{2,8-10)} 種苗放流の場合 N は大きく、目視の場合には P も大きいので $NP \geq 5$ ($P \leq 0.5$ の場合) あるいは $N(1-P) \geq 5$ ($P > 0.5$ の場合) がどの i でも成り立つと考えられるから、ここではこれを正規近似して over-dispersion を考慮する。尤度は

$$L = (2\pi\phi)^{-k/2} \prod_{i=1}^k \{N_i\alpha(1-\alpha)\}^{-1/2} \times \exp \left\{ -\frac{1}{2} \sum_{i=1}^k \frac{(n_i - N_i\alpha)^2}{N_i\alpha(1-\alpha)\phi} \right\} \quad (4)$$

である。^{2,8,10)} これを2項分布モデルと呼ぶ。

統計モデルの比較 モデルの比較はシミュレーションによって行う。シマアジの事例¹⁾にあわせ、 $N=20,000$, $M=0.02/\text{日}$, $\alpha=0.8$, $\phi=600$, $k=30$ として、目視データを $n_i = N_i\alpha + U\sqrt{N_i\alpha(1-\alpha)\phi}$ で生成した。ここで、 U は標準正規乱数である。生成した同一のデータから、モデル0、修正モデル0.1、2項分布モデル、対数線形モデルの四つの統計モデルを用いてパラメータを推定することを100回繰り返し、それぞれの方法によって得た点推定値の平均と標準偏差を計算した。

結 果

シミュレーションの結果をTable 1にまとめて示した。 M と α の点推定値はモデル0、修正モデル0.1、対数線形モデルではほとんど遜色ないが、2項分布モデルではやや小さめに推定された。

ϕ については、対数線形モデルでは半分以下になっており過小推定となる。モデル0、修正モデル0.1では580前後ではほとんど変わらないが、2項分布モデルでは645とやや大きめに推定されている。

最大対数尤度の値は、モデル0で-279.50、修正モデル0.1で-279.11とほとんど同じ値となった。2項分布モデルの値が-276.50と最も大きかった。

点推定値の標準偏差については、 M と α については、

Table 1. Mean values of the point estimates and their SDs (in parentheses) obtained from 100 simulations for the different statistical models

	Statistical models			
	Model 0	Model 0.1	Conditional binomial	Log linear
M	0.01968 (0.00241)	0.01980 (0.00306)	0.01845 (0.00221)	0.01953 (0.00224)
α	0.79192 (0.02710)	0.79305 (0.03064)	0.78073 (0.02581)	0.79965 (0.02579)
ϕ	581.38 (82.68)	575.41 (91.09)	649.89 (205.78)	232.41 (75.02)
LL	-279.50 (2.36)	-279.10 (2.76)	-276.50 (7.66)	-0.8637* (0.0457)

N : number of seeds released, M : natural mortality rate, α : sighting rate, ϕ : dispersion parameter, LL : maximum log likelihood value

$N=20,000$, $M=0.02$, $\alpha=0.8$ and $\phi=600$. Model 0.1 is the revised model of Model 0. The asterisk indicates the correlation coefficient.

どのモデルでもほとんど変わらなかった。 ϕ については、モデル0で82.68、修正モデル0.1で91.08、2項分布モデルで205.78となった(対数線形モデルは ϕ が過小になるのでこの議論から除いた)。最大対数尤度の標準偏差の値はモデル0、修正モデル0.1がほぼ同じ値であったが、2項分布モデルではそれらの3倍近く大きかった。

考 察

対数線形モデルにおける ϕ の過小推定 対数線形モデルでは M と α については、ほぼ真の値が推定されたが、 ϕ が極端に過小となった。これにより、 $\hat{M}$ と $\hat{\alpha}$ の分散も過小推定することになる。目視のモデルは試験操業のモデル⁹⁾の拡張である。試験操業の場合に、 $N=20,000$, $M=0.1$, $\phi=2$, $F=0.01$ (目視の場合の α に対応する)および0.1で、対数線形モデルによる推定が検討されているが、いずれのパラメータの推定値も真の値に近く問題はなかった。⁹⁾ 試験操業の場合には、目視の場合と異なり、サンプリング強度 F が小さいことが普通である。そこで、 $M=0.02$, $\phi=600$ を固定し、 α を0.8, 0.6, 0.4と小さくして、各100回のシミュレーションを行ってみたところ、 ϕ の平均値は順に235.18, 316.27, 412.12となり、 α の値が小さくなるにつれ ϕ の推定値は真の値に近づいた。いくつかのシミュレーションの結果では、 α が大きいほど ϕ の過小推定が大きくなった。この原因について、以下のように考察した。

対数線形モデルの E_i は近似的に正規分布に従う。この時の全尤度を対数尤度でかけば

$$\log L = -\frac{k}{2} \log(2\pi\phi) - \frac{1}{2} \sum_{i=1}^k \log \left(\frac{1-P_i}{NP_i} \right) - \frac{1}{2} \sum_{i=1}^k \frac{NP_i [\log(n_i/N) - \log P_i]^2}{(1-P_i)\phi} \quad (5)$$

となる。(1)式のモデルの加重最小2乗法では残差平方和

$$SS = \sum_{i=1}^k n_i [\log(n_i/N) - \log P_i]^2 \quad (6)$$

を最小にすることで α と M を推定し、 ϕ はこれらの推定値を用いて(2)式によって別に推定するという方式をとっている。分散が一定の通常の最小2乗法(OLS)では最小2乗法と最尤法は同じであるが、ここで扱っている場合は分散が各 i で異なっており、分散がパラメータの関数になっているため最小2乗法と最尤法が一致しない。加重最小2乗法では(5)式の第1項と第2項を無視し、第3項目だけで α と M を推定している((6)式)。このとき、 ϕ を無視し、 $NP_i/(1-P_i) \approx n_i$ と近似していることがわかる。

(5)式の第1項と第2項を無視する推定方式は、

Table 1 のシミュレーション結果からもわかるように、 α と M の推定においては問題がない。これは、 α と M の推定では、 $\hat{M}$, $\hat{\alpha}$, $\hat{t}$ の分母分子の両方に重み W_i が含まれているので、重みの近似の影響については相殺されることによると考えられる。

α が大きくなると ϕ の過小推定をひきおこすことが問題である。これは、(2) 式には分子の方にしか重み W_i が入っていないので、 W_i が過小推定になると ϕ も過小推定になることによる。本来 W_i は分散の逆数であるが、ここでは分散を $(1-P_i)/(NP_i)=1/NP_i-1/N \approx 1/NP_i$ と近似し、重みを $(1/NP_i)^{-1}$ の実現値 n_i で推定して計算している。この近似の精度は N の大きさだけでなく P_i の大きさにも依存する。本来の重みは $NP_i/(1-P_i)$ である。 P_i が小さいときには近似は問題ないが、 P_i が大きくなると近似が悪くなることは容易に理解できる。 P_i が小さいことが一般的な、連続漁業の場合^{3,4)} や試験操業⁵⁾ の場合には問題がなかったのはこのためである。

100 回のシミュレーションでデータを生成し、(2) 式で重みを $W_i=n_i/(1-\hat{P}_i)$ として推定したところ、 ϕ の点推定値の平均は 604.41、 M のそれは 0.01976、 α では 0.80203 となり、ほぼ正しく推定できた。対数線形モデルで ϕ が過小推定になるのは、 P_i が大きい場合に分散の近似が悪くなることに起因することが確認された。目視に限らず標識再捕の場合をも含め、対数線形モデルでは、 ϕ の推定には(2)式で重みを $W_i=n_i/(1-\hat{P}_i)$ として推定すればよい。

統計モデル モデル 0 と修正モデル 0.1 は、推定値の真の値への近さでは 2 項分布モデルよりよい結果を与える。最大対数尤度の平均値から判断して、モデルのあてはめは 2 項分布の方がよいが、2 項分布では特に ϕ を過大推定し、 ϕ の変動も大きい。

はじめに述べたように、目視のデータの場合に多項分布を用いようとして、 $(k+1)$ 項目の扱いが変則的になったが、点推定値はむしろ 2 項分布よりもよかった。これは $(k+1)$ 項目の情報ほとんど推定に貢献していないことによると思われる。また対数線形モデルにおいて、多項分布を前提として E_i の分散を求めている点についても、重みを修正したシミュレーション結果から、問題とはならないことがわかる。対数線形モデルは $(k+1)$ 項目を使わず、分散共分散行列の共分散を無視したモデルである。これらを考えあわせると、 k 個の観測値のみを用い、それぞれの分散が異なる正規モデルをあてはめるアプローチが考えられる。これは加重最小 2 乗法のモデルに対応したモデルであり、尤度は

$$L = (2\pi\phi N)^{-k/2} \prod_{i=1}^k \{P_i(1-P_i)\}^{-1/2} \times \exp \left\{ -\frac{1}{2} \sum_{i=1}^k \frac{(n_i - NP_i)^2}{NP_i(1-P_i)\phi} \right\} \quad (7)$$

である。これをモデル 1 と呼ぶ。モデル 1 の良さを評価するため、修正モデル 0.1、2 項分布モデル、重みを修正した対数線形モデルを比較の対象として選び、同様に 100 回のシミュレーションで四つのモデルを比較した (Table 2)。

M と α の推定については、モデル 1、修正モデル 0.1、重みを修正した対数線形モデルでほとんど遜色なかったが、モデル 1 と重みを修正した対数線形モデルが真の値により近かった。2 項分布モデルでは過小推定の傾向が再現された。

ϕ については、モデル 1 がほぼ真の値に近く点推定の平均値は 598.30 であった。次に真の値に近いのは重みを修正した対数線形モデルの 607.52、修正モデル 0.1 は過小に、2 項分布モデルは過大になり、初めのシミュレーションの結果を再現した。 ϕ の標準偏差はモデル 1 で最も小さく 46.55、修正モデル 0.1 では 143.17、重みを修正した対数線形モデルでは 156.99、2 項分布モデルで 237.37 となり、モデル 1 が最も安定している。 ϕ の CV を計算すると、モデル 1 が 0.08、修正モデル 0.1 が 0.27、2 項分布モデルが 0.35、重みを修正した対数線形モデルでは 0.26 となる。 $(k-2)\phi$ は自由度 $(k-2)$ の χ^2 分布に従うから、 ϕ の変動係数の理論値は $CV(\phi) = \sqrt{2/(k-2)} \approx 0.27$ となる。修正モデル 0.1 と対数線形モデルではほぼこの値が得られている。これに比べてモデル 1 では CV は理論値の 1/3 より小さく、2 項分布

Table 2. Mean values of the point estimates and their SDs (in parentheses) obtained from 100 simulations for testing Model 1 and the log linear model with revised weight

	Statistical models			
	Model 1	Model 0.1	Conditional binomial	Log linear (revised weight)
M	0.01993 (0.00204)	0.01981 (0.00314)	0.01851 (0.00195)	0.01963 (0.00217)
α	0.79929 (0.02442)	0.79275 (0.03195)	0.77971 (0.02222)	0.79961 (0.02620)
ϕ	598.30 (46.55)	525.85 (143.17)	684.22 (237.37)	607.52 (156.98)
LL	-274.39 (3.57)	-278.78 (3.28)	-277.64 (7.04)	-0.8594* (0.0386)

$N=20,000$, $M=0.02$, $\alpha=0.8$ and $\phi=600$. The asterisk indicates the correlation coefficient.

モデルでは大きい。 ϕ の推定精度はモデル1が最もよく、2項分布モデルでは変動が大きいといえる。

最大対数尤度の値は、モデル1が-274.39、2項分布モデルで-277.64、修正モデル0.1では-278.78の順で、モデル1が最も大きかった。2項分布モデルの標準偏差はモデル1に比べて約2倍大きく、あてはめの安定性においてモデル1が優れている。

長谷川ら²⁾は目視尾数 n_i が放流尾数 N から確率 P_i で目視され、それが2項分布するモデル

$$L = \prod_{i=1}^k \binom{N}{n_i} P_i^{n_i} (1-P_i)^{N-n_i}$$

を加え、これを正規近似してシマアジのデータを解析している。この正規近似の尤度は(7)式(モデル1)と同じである。本論文では、前報¹⁾を基本に、相対的に良い結果を与える対数線形モデルの推定法から加重最小2乗法に対応する(7)式の尤度に至ったが、結果的に両者は一致した。

モデル1は、点推定値の真の値への近さ、データへのあてはめのよさや安定性などからみて、最も適当と考えられた。対数線形モデルでは多項分布を前提として $\log(n_i/N)$ の分散共分散が導かれたが、この影響はなく、重みを修正すれば ϕ も正しく推定できる。計算が簡単であり、実用的でもある。2項分布モデルは、データ発生モデルとしては自然であるにもかかわらず、あまり良い結果を示さなかった。(4)式で $N_i = Ne^{-M_{ti}}$ 、(7)式で $P_i = \alpha e^{-M_{ti}}$ として対数尤度をそれぞれ書き直すと、2項分布モデルの第2項と第3項の $(1-\alpha)$ がモデル1では $(1-e^{-M_{ti}}\alpha)$ になっているだけの違いである。2つ

のモデルは本質的に大きく違ってないにもかかわらず、モデル1の方が良かったが、この理由については不明である。

終りに、貴重なご意見をいただいた東京大学大学院総合文化研究科岸野洋久氏、遠洋水産研究所の平松一彦氏、静岡県水産試験場伊豆分場の長谷川雅俊氏および三重大学生物資源学部の白木原国雄氏に感謝申しあげる。

文 献

- 1) 北田修一：目視による飼付けシマアジ滞留尾数の推定。栽培技研, 21, 37-39 (1992)。
- 2) 長谷川雅俊, 高木康次, 岡本一利：放流群の目視データの解析法。栽培技研, 24, (1997), (印刷中)。
- 3) G. J. Paulik: Estimates of mortality rates from tag recoveries. *Biometrics*, 49, 28-57 (1963)。
- 4) R. W. Farebrother: Weighted least-squares estimates of mortality rates from single-release tagging studies. *J. Cons. Int. Explor. Mer* 42, 166-170 (1985)。
- 5) 北田修一：標識再捕に基づく種苗放流効果の評価に関する統計学的研究。栽培資源調査検討資料, 6, p. 89 (1991)。
- 6) 北田修一, 岸野洋久, 平松一彦：試験操業による放流魚の死亡係数の推定。日水誌, 58, 1399-1403 (1992)。
- 7) S. Kitada, K. Hiramatsu, and H. Kishino: Estimating mortality rates from tag recoveries: incorporating over-dispersion, correlation, and change points. *ICES J. mar. Sci.*, 51, 241-251 (1994)。
- 8) 平松一彦：最尤法による水産資源の統計学的研究, パラメータ推定とモデル選択。遠洋水研報, 29, 57-114 (1992)。
- 9) G. A. F. Seber: The Estimation of Animal Abundance and Related Parameters. Second ed., Griffin, London, 1982, p. 654。
- 10) T. Yamakawa, Y. Matsumiya, and S. Kitada: Comparison of statistical models for expanded DeLury's method. *Fisheries Sci.*, 60, 405-409 (1994)。