日本海ブロック試験研究集録
水産庁　日本海区水産研究所
第28号，85～86，1993

標識再捕調査によるピーターセン法の不偏推定量と区間推定

赤嶺達郎
（日本海区水産研究所）*

ピーターセン法は標識再捕調査によって資源尾数を推定する場合の基本的な手法である．しかし，点推定についてはさまざまな式が提唱されており，統一的な見解は示されていない．また区間推定についても統計学の教科書では明確な解法は示していない．ここではこれらの問題を検討して標準的な手法を提示する．

この問題は二項分布の「逆確率」問題で，ベイズの定理の原形であり，昔から議論されてきたものである．最終的にはベイズ統計が否定され，ネイマン流の区間推定論が確立されたのであるが，この問題で厳密にネイマン流の区間推定を行うのは容易ではない．この問題を単独で議論するよりも，二項分布と関係の深い「枠取り法」や「デルーリー法」と同時に議論した方が見通しがよくなる．これらの事情から1993年4月の日本水産学会のシンポジウム「水産資源解析の課題と展望－統計モデルと資源特性値の推定」において，「除去法における資源量推定とモデル選択」という演題で総合的に議論する予定である．これは10月に水産学シリーズの一冊として発行される予定なので，ここでは今回の発表の要旨を記録するだけにとどめる．

なお，以下では N：資源尾数，M：標識尾数，n：再捕尾数，x：標識魚の再捕尾数とする．

(1) ランダム・サンプリングであれば x は超幾何分布に従う．したがって，ピーターセン法は超幾何分布と同一モデルであり（統計学の教科書では超幾何分布の実例としてピーターセン法を紹介しているものが多い），超幾何分布において M, n, x の値より N を推定する問題である．

(2) 超幾何分布は $p=M/N$ が一定とみなせる場合には二項分布で近似できる．さらに p が小さくて $n\to\infty$, $np\to\lambda$ とみなせる場合にはポアソン分布で近似できる．実際のピーターセン法においては二項分布やポアソン分布で近似できる場合が多い．近似できない場合は調査そのものに問題がある（N が小さい訳だから調査する意味がない）．

(3) 超幾何分布における N の不偏推定量，二項分布における $1/p$ の不偏推定量，ポアソン分布における $1/\lambda$ の不偏推定量は存在しない．しかし，その近似値は計算できて，これらを用いると N の推定値としてそれぞれ $(n+1)(M+1)/(x+1)-1$，$(n+1)M/(x+1)$，$nM/(x+1)$ が得られる．

(4) 一方，最尤推定値は「比例配分」の値 nM/x とほとんど一致し，nM/x を越えない最大の整数となる（nM/x が整数の場合には $nM/x-1$ も最尤推定値となる）．

---

*現所属：中央水産研究所

(5) 以上のように N の推定値としてさまざまな式が提唱されてきた．実際の調査では n，M の値は大きいので，不偏推定値（の近似値）の3式はどれも大差ない．超幾何分布の式を用いれば精度が高くなるという訳ではない（二項分布やポアソン分布は超幾何分布とは独立に定義でき，実際の調査において x が超幾何分布に従うという保証はないから）．問題は分母の x と $x+1$ の差である．これにより最尤推定値と不偏推定値（の近似値）では比率で約 $1/x$ の差がある．x が小さい場合には非常に大きな差となる．x をできるだけ大きくする必要があるが，小さい場合には点推定よりも区間推定を行うべきである．

(6) どれが点推定値として優れているかは断定できない．最尤推定値と超幾何分布の不偏推定値（の近似値）を並記するのがもっとも一般的だろう．

(7) 区間推定を厳密に行うには帰無仮説 H_0 ： $N = N_0$ を立てて，各 N_0 についてしらみつぶしに x の信頼区間を求めなくてはならない．計算機を用いても実用的な方法とはいえない（このため教科書に明示されていないものと思われる）．

(8) 実用的には二項分布に近似しさらに正規分布に近似すればよい．この場合には p の2次方程式を解くことにより，N の信頼区間（の近似値）を容易に得ることができる．

(9) 正規分布に近似できない場合には（調査そのものに問題があるが），ベイズ統計を使う方法がある．二項分布では，p の事前分布として一様分布，超幾何分布では N の事前分布として $(M+1)/(N+2)(N+1)$ を採用すれば，厳密な方法と整合性の高い信頼区間を容易に得ることができる．ベイズ統計を否定する立場をとれば，この方法は近似的な簡便法と解釈できる．

(10) 実際の標識再捕調査ではランダム・サンプリングとなっていない場合が多く，上記のような手法は使えない．それは対象生物の分布様式がランダム分布よりもさらにバラツキの大きい集中分布に従っているためである．集中分布から単純に「枠取り法」で得られたデータについては，ここで示した信頼区間よりもはるかに大きな信頼区間となる．ここで示した手法を適用するためには，採集方法に工夫をこらしてランダム・サンプリングに持ち込む必要がある．