

# 階層ベイズモデルを使った兵庫県クマ生息数推定に於ける問題 (統計学の観点からの追記)

2016年7月31日

日本福祉大学経済学部 教授 山上俊彦 (応用経済学)

## 1. 背景

兵庫県はツキノワグマ保護(管理)計画に MCMC による階層ベイズモデルを用いたツキノワグマの生息数推定値を掲載している。また、推定のバックグラウンドペーパーは兵庫県から公表されている。

兵庫県による MCMC による階層ベイズモデルを用いたツキノワグマの生息数推定の問題点については、日本奥山学会掲載論文 Vol 4(2016年)において指摘したとおりである。(批判対象となった資料等は論文に掲載してある。)

<そこで述べた問題の要点>

捕獲数を用いて生息数を推定する際に、**捕獲率を階層ベイズ法で推定し、生息数を推定しているが、生息数が不明なのに捕獲率を推定することは論理的に不可能である。**

以下に、統計学上のさらなる問題点について追記する。

## 2. ベイズ統計学とは何か

ベイズ統計学

- **結果(y)から原因(θ)を探る。逆確率の推測。**
- ベイズの定理

$$P(\theta|y) = \frac{f(y|\theta)\pi(\theta)}{\int_{\Theta} \pi(\theta)f(y|\theta)d\theta} \propto f(y|\theta)\pi(\theta) \quad P(\theta|y): \text{事後確率}, f(y|\theta): \text{尤度}, \pi(\theta): \text{事前確率}$$

- $\pi(\theta)$  : 原因(θ)の事前確率(主観確率)
- $f(y|\theta)$  : θが与えられたときの、結果である客観的データ(y)の条件付き確率、あるいはθの尤度である。
- $P(\theta|y)$  : yを条件とした場合のθの事後確率。
- ベイズ統計学とは、自己の信念(θの確率)を客観的データyで修正するもの。

### 3. 問題点(その1)

#### <クマ生息数の推定に階層ベイズ法の名称を使用していることの問題点>

本来、ベイズ統計学においては事前情報の分布についてさらに分散の事前情報を与えることを階層ベイズ法と呼んでいる。しかし、兵庫県の推定では初年度生息個体数に自然増加率を乗じて生息数の推移を求めている。この階層構造になっているモデルのパラメータをベイズ統計学を用いて算定しているため、「階層ベイズモデル」と命名している。

従って、兵庫県の推定では、統計手法としての「階層ベイズ法」を必ずしも用いている訳ではない。つまり、兵庫県が言うところの「階層ベイズ法」は本来の統計学の用語とは定義がずれている。

### 4. 問題点(その2)

#### <生息数推定過程における問題点>

- 兵庫県では次のようにパラメータを推定している。
- $\theta$  : 自然増加率、捕獲率等
- $\pi(\theta)$  :  $\theta$  の事前確率(主観的な確率)は、過去の経験を基に自己の信念に基づく値に設定可能。
- $f(y|\theta)$  :  $\theta$  を条件としたデータ  $y$  (捕獲数)の確率分布から求められる。
- 捕獲率  $\theta$  が与えられたときの捕獲数  $y$  の条件付き確率は生息数が不明であるため、誰にも分からない。つまり、 $f(y|\theta)$  は設定不可能。
- 従って、 $\theta$  の事後確率  $P(\theta|y)$  は計算不可能。

以下、この問題を解決するために兵庫県が採用した方法

- 捕獲数  $y = \text{捕獲率}(\beta) \times \text{生息数}(N) = \beta N$  と読み替える。捕獲数  $y$  は平均  $\beta N$  の正規分布に従うと想定する。
- つまり  $f(y|\theta) = f(y|\beta, N)$  と設定する。
- これで  $\theta(\beta)$  の事後確率  $P(\theta|y)$  が求められる。
- 捕獲数さえ分かれば、捕獲率が推定されて、生息数が逆算で求められる。

→ **捕獲数から捕獲率、生息数が全て分かることなど、有り得ない。**

#### <どこがおかしいのかの説明>

- 捕獲数  $y$  の期待値は  $\beta N$  に等しくなる。
- $y = \beta N$
- $\beta$  と  $N$  の組み合わせは無限にある。
- $\beta$  を小さくすると  $N$  は大きくなる。生息数はいくらでも膨らませる。
- つまり  $\beta$  は特定の値に設定できない。

#### <統計学ではこのような問題にどう対応しているのか?>

- パラメータの候補値が無数にある場合を統計学では「識別不可能」と呼んでいる。

つまりモデルで求めたいパラメータ数と比較してデータの情報量が少なすぎる場合、パラメータを一意に決定できない。

- ここでは捕獲数しかデータがないので複数のパラメータ(捕獲率と生息個体数)の値を分離して導くことができない。
- 識別可能性を満たしていない場合、ベイズ統計学では尤度最大化は行わないので、形式的には統計式は成立するが、生態学モデルとしては成立していない。

→ **兵庫県モデルでは、捕獲数から生息数を求めて、求めた生息数から捕獲数を求めるという循環計算をしているに過ぎない。**

## 5. 識別可能性について

<識別可能性とは何か>

- 全ての  $y$  に対して、 $f(y|\theta)$ =一定となる複数の  $\theta$  が存在する場合、 $\theta$  は識別されない。
- つまり兵庫県の推定において、全ての捕獲数データに対して、捕獲率と生息数の組み合わせは無限に存在し、いずれの組み合わせも同一の尤度を与えるので捕獲率と生息数は識別されないことになる。
- 識別不可能な場合、尤度  $f(y|\theta)$  が同一となる。
- この場合、MCMC を用いて尤度を計算して最適なパラメータを求めようとしても、値が全て同一となり、収束条件が満たされず、捕獲率は求められないことになる。

<では、なぜ、捕獲率や生息数が求められたのか？>

- 本来、識別不可能で捕獲率が決定できないモデルにおいて、自然増加率が攪乱項となり尤度に影響を与えたため、見かけ上、MCMC の収束条件が満たされて、高い自然増加率と歩調を合わせた低い捕獲率が選択されたと考えられる。
- 自然科学系の研究者の多くは識別問題に鈍感であり、乱数等を用いて強引に解を求めようとする傾向がある。
- 識別可能性を満たしていない場合、ベイズ統計学を用いても、事前情報（研究者の信念）はデータによって修正されない。

→ つまり、作成者の思い込みが修正されずに成立する。

→ **ベイズ統計的推定と生態学のモデル構築は別次元の問題である。**

<正しいベイズモデルはどのようなモデルか？>

- 生態学の巨匠 Seber の標識再捕獲法を用いる場合のベイズモデル

$$f(N|data) \propto \int f(data|N, p)f(N, p)dp$$

$p$ : 捕獲率、 $N$ : 生息数、 $f(N, p)$ :  $N$  と  $p$  の結合事前分布、 $data$ :

データには標識個体に関するものが含まれる。

- Seber は、 $N$ : 生息数の事前分布を設定することが非常に難しいこと、生息数と捕獲率は独立であることを指摘している。
- 生息数と捕獲率の事前情報が揃っていなければパラメータとして推定できない。

Seber が指摘する作業を兵庫県は全く行っていない。

→ 兵庫県のバックグラウンドペーパーの問題点は、ここで Seber が提示して最も重要な核になるベイズ式を提示していないことである。

#### 6. 追加の問題点

- 兵庫県の推定では、目撃数や標識個体数の残存率等が推定対象とされている。しかし、これらのパラメータは捕獲率推定に有効に利用されておらず、結果としてこれらのパラメータは局外母数(nuisance parameter)として扱われている。そのため、生息数の推定に際しての制約条件として機能していない。
- 無暗にモデルを複雑化するのではなく、本来、標識個体数に関するパラメータをベイズ法を用いて推定し、その結果を用いて捕獲率や生息数を推定するという多段階の手続きを踏む必要がある。

#### 結語

- モデルの識別不可能性等の問題点は、生態学モデルとしての欠陥を意味している。
- ベイズ統計学として成立していても、モデル構築が正しいとは限らない。
- このような問題点は、ベイジアン計量経済学といったモデルを扱う高度な英語の教科書には記載されているものの、日本では本格的に紹介されていない。
- 学術論文は、外部の査読による点検が必要である。
- 研究者倫理の問題として、捕獲率の事前情報を変更した場合の結果等、途中経過を開示しなければならない。

#### 参考文献

- Bauwens,L.,M.Lubrano and J.-F.Richard(2000) “Bayesian Inference in Dynamic Econometric Models” Oxford University Press
- Lancaster,T.(2004)” Introduction to Modern Bayesian Econometrics” Wiley-Blackwell
- Seber,G.(1986) “A Review of Estimating Animal Abundance” Biometrics Vol.42, pp.267-292
- 山上俊彦(2016) 「最近のクマ類生息個体数を考える」日本奥山学会誌 Vol.4,No.1 pp.29-49