

4.1 標識再捕獲法に基づく密度推定手法の精度検証

松田裕之（横浜国大）・堀野真一（森林総研）・太田海香（横浜国大）・深澤圭太（自然研）

1. はじめに

DNAによる個体識別の技術の発展とコストダウンにより、ヘア・トラップによる生息密度推定はクマ類の保護管理の現場においても一般的なモニタリングツールとなりつつある。ヘア・トラップによる生息密度推定は、現地調査、DNAの分析、統計解析という3つの段階から成る。すでにヘア・トラップを用いた個体群のモニタリングが各地で行われているにもかかわらず、いずれの段階においても方法論にはさまざまなバリエーションがあり、最適な選択肢に関するコンセンサスは得られていない。

生息数推定においては、Lincoln-Petersen法などの閉鎖個体群モデルが一般に用いられてきた。しかし、この手法で推定された生息数はヘア・トラップからどの程度の範囲に生息する個体を対象としたものであるか自明ではなく、生息密度の算出のためには別途有効トラップ面積を計算する必要がある。従来の日本におけるヘア・トラップ調査においては、有効トラップ面積としてトラップが含まれるメッシュの面積やトラップから発生させたバッファの面積が用いられてきた。しかし、これらの方法はメッシュやバッファのサイズの決め方に科学的な根拠がなく、評価者の主観の影響を強く受ける点で不適切である。また、事前に面積が明らかな調査範囲を設定し、その範囲にトラップを可能な限り一様に設置した調査例もあるが、昨年度の報告書においてはそのような方法を用いても生息密度の推定に大きなバイアスがかかりうることを示した。

標識再捕獲法における有効トラップ面積は、対象生物の行動圏サイズとの関連が古くから議論されてきた（e.g. Dice 1938）。クマ類の行動圏サイズは堅果類の豊凶など、時空間的に変化する環境条件によって影響をうけ、それによって有効トラップ面積も大きく変動すると考えられる。標識再捕獲法における捕獲の空間情報は対象個体群の移動距離に関する情報を持つため、それを活用することで調査地・調査期間に特有な有効トラップ面積を明らかにし、客観的に生息密度を推定することができると考えられる。しかしながら、これまでに標識再捕獲法における空間情報から生息密度を推定する手法は複数考案されており、手法によって推定密度がどの程度変化するかについては明らかでない。

そこで、本サブテーマにおいては、ダミーデータと本プロジェクトで得られた岩手県北上山地における標識再捕獲データを用いた解析により、複数の生息密度推定手法の精度比較を行った。対象とした手法は下記の3つである。

- (i) 平均最大移動距離に基づく方法(Wilson & Anderson 1985)
- (ii) Inverse prediction (Program “DENSITY”, Efford 2004)
- (iii) ベイズ空間明示型標識再捕獲モデル(Gardner et al. 2009; Royle et al. 2009)

密度既知のダミーデータによりそれぞれの手法の推定値と真の値のずれを明らかにし、それと同様の傾向が野外で得られたデータに対する推定においても見られるかどうかを確かめることで、それぞれの手法の推定バイアスを明らかにする。

2. 方法

2-1. ダミーデータの生成

一般に、生物の行動は2次元ランダムウォークとして記述されるが、クマの場合はその軌跡が行動圏に従うパターンを示すことが多い。それをダミーデータ生成の際に再現するため、以下に示すようなアルゴリズムを考案した。

1) 確率論的行動シミュレーションモデルとトラップにおける捕捉

単純なランダムウォークにおいては、時点 t における個体の位置 \mathbf{u}_t は、 $t-1$ の位置 \mathbf{u}_{t-1} を中心とする2次元正規分布に従うと考えることができる。この分布は、単位時間当たりの個体の移動を規定する移動距離分布である。

$$p(\mathbf{u}_t; \mathbf{u}_{t-1}, \sigma^2) = N_2(\mathbf{u}_{t-1}, \sigma^2 I)$$

σ^2 は移動距離を決めるパラメータであり、大きな値を取るほど単位時間当たりの移動距離は長くなる。

ここで、このようなランダムウォークモデルに行動圏の制約を組み込むことを考える。一般に、行動圏は十分長い時間行動したときの個体位置の系列が従う分布としてみなされる。ここで、長い時間行動したときの個体位置が従う行動圏分布を任意の確率分布 $p_h(\mathbf{u})$ として定義すると、時点 t における個体の位置 \mathbf{u}_t の分布 $p(\mathbf{u}_t; \mathbf{u}_{t-1}, \sigma^2)$ は次のように記述することができる。

$$p(\mathbf{u}_t; \mathbf{u}_{t-1}, \sigma_m^2) \propto p_h(\mathbf{u}) \times N_2(\mathbf{u}_{t-1}, \sigma^2 I)$$

$t=0$ における個体の位置 \mathbf{u}_0 を与え、そこからこの分布に従う乱数列を逐次発生させることで、行動圏分布に制約される行動シミュレーションを行うことができる。このようなプロセスの模式図を図1に示した。図のように、時点 t における個体の位置が行動圏分布の外側にある場合、移動距離分布と行動圏分布の積として表わされる時点 $t+1$ の個体位置分布は行動圏分布の中心に偏ったものになる。その結果、個体の軌跡は行動圏分布の確率密度が高い範囲を動くことになる。なお、マルコフ連鎖の収束定理より、初期状態における個体の位置によらずこのプロセスの定常分布はホームレンジ分布 $p_h(\mathbf{u})$ に一致する。

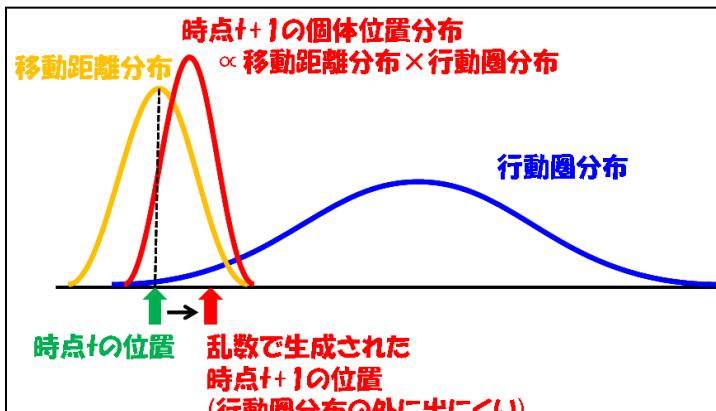


図1 行動圏分布と移動距離分布による行動シミュレーションの模式図

さらに、個体がトラップに接近したときのトラップの利用と検出についても確率モデルにより表現した。クマ個体とトラップの距離が近いほどトラップにおける検出率は高まるため、位置 $\mathbf{x}_j = (x_{1j}, y_{1j})$ にあるトラップ j に時点 $t+1$ で検出される確率 p_c を個体の位置とトラップの距離の関数として以下のように表わした。

$$p_c(\|\mathbf{u}_t - \mathbf{x}_j\|) = \exp(-(\|\mathbf{u}_t - \mathbf{x}_j\|/\sigma_c)^2)$$

この関数は正規分布と形状が同じであるが、距離 0 における検出率は 1 となる。シミュレーション

ンの各時点において最も近くにあるトラップに対して p_c を計算し、ベルヌイ乱数により検出の有無を決定した。

2) ダミーデータの生成

ダミーデータの生成においては、全てのトラップを 1 週間に 1 回、計 6 回の点検を行い、その期間内に検出された個体番号とそのトラップを記録する調査を想定したシミュレーションを行った。その結果、検出個体、トラップ、セッションそれぞれにおける検出の有無を示す 3 次元の行列が検出履歴として得られる。トラップは、2 km 格子の交点に $10 \times 10 = 100$ 個を設置したと仮定し、トラップ位置の凸包は 20 km 四方の正方形となる。

ダミーデータ生成の際のパラメータの設定は下記のとおりである。

- ・真の生息密度 : 0.2 個体/km²
- ・行動シミュレーションの時間ステップ : 10 ステップ/日
- ・1 ステップ当たりの移動距離分布のパラメータ σ : 0.18

※この値は、1 日の移動距離の平均を 1.28 km とすることに相当する。

- ・行動圏分布 $ph(s)$: 95% 確率円の面積が 4.85 km² となる 2 変量正規分布
- ・検出確率の距離依存性パラメータ σ_c : 0.02

※この値は、検出確率の変曲点が 20 m にあることを意味する。

個体の行動圏の中心は、トラップ設置範囲を中心とした 30×30 km の範囲にランダムに発生させた。シミュレーションの初期状態として、 $t=0$ における個体の位置を各個体の行動圏分布からランダムに発生させ、上記のアルゴリズムに従いデータ生成を行った。1 個体についてデータ生成を行った一例を図 2 に示した。シミュレートされた行動軌跡は行動圏分布の中にとどまることがわかる。また、一般的な標識再捕獲調査は短期間のうちに行われるため、その間で個体は行動圏分布全体をくまなく利用できるとは限らない。シミュレーションが生成する行動パターンはそのような性質も再現している。

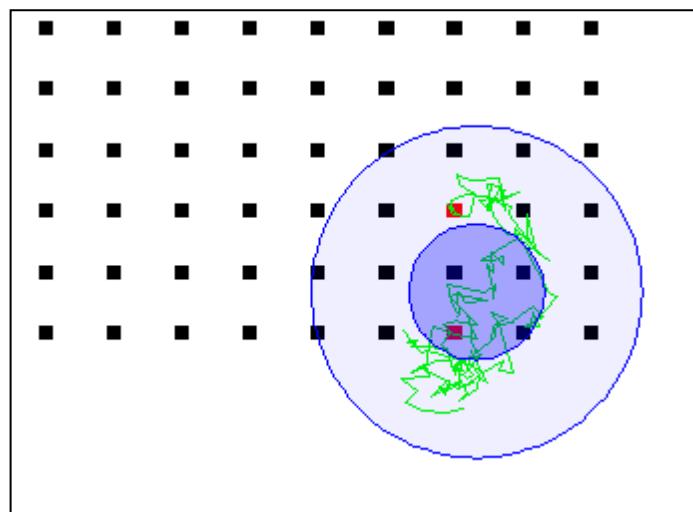


図 2 データ生成シミュレーションの 1 例。青色の同心円は内側が行動圏分布の標準偏差半径円、外側が 95% 確率円、緑色の折れ線はシミュレートされた行動軌跡、■は非検出トラップ、■は検出トラップを示す。

2-2. 岩手県北上山地におけるヘア・トラップ調査データ

実際に得られたデータによるモデル評価には、2010 年度に北上山地で得られたヘア・トラップによる標識再捕獲調査データを用いた。245 基のヘア・トラップによる 6 セッションの調査を行い、マイクロサテライトマーカーによる個体識別を行った結果、146 個体が識別された。本報告書「1.1 北上山地における大規模ヘア・トラップ調査」に調査方法の詳細が記述されている。個体ごとの検出トラップとセッションの分布を図 3 に示した。

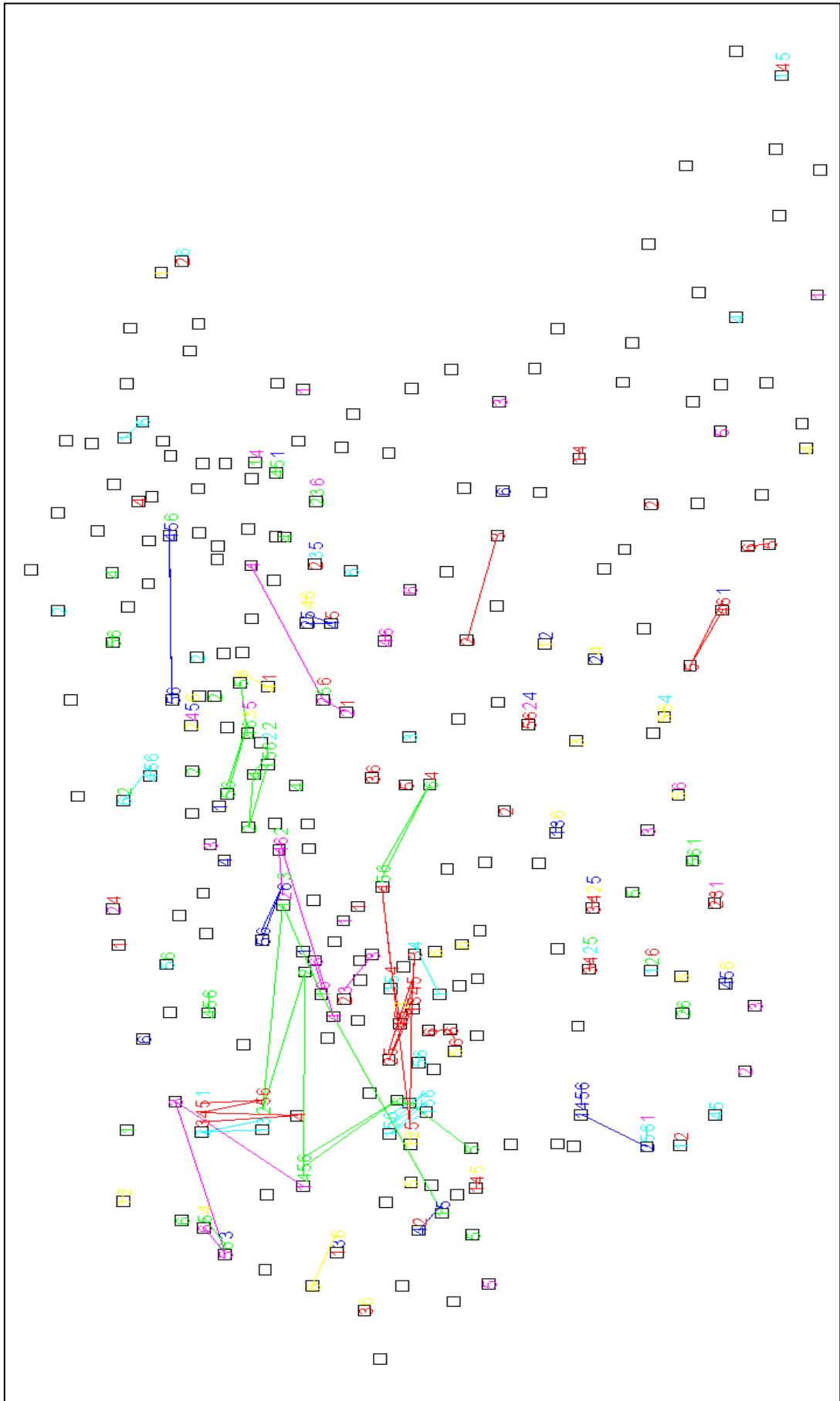


図3 北上山地ヘア・トラップ調査におけるサンプル採取位置とセッション。四角形はトラップの位置、数字はサンプルが採取されたセッションを示し、個体ごとに色分けされている。同一個体が複数回採取されている場合は、数字を実線でつなぎいる。

2-3. 密度推定手法の比較

本研究においては、平均最大移動距離法、Inverse prediction、ベイズ空間明示型標識再捕獲モデルの推定精度の比較を行った。ダミーデータを用いた比較においては、ベイズ空間明示型標識再捕獲モデルは40回、それ以外の手法については100回の繰り返しを行い、推定値と真の値のずれを相対バイアスと95%CI coverage（95%CIが真の値を含む割合）により評価した。下記にそれぞれの推定手法の概要を示した。

1) 平均最大移動距離法(Wilson & Anderson 1985)

この手法は、複数回検出された個体ごとに検出トラップ間の距離の最大値を計算し、その平均値の半分をホームレンジ円の半径 r とみなすものである。トラップから生成した半径 r のバッファに含まれる範囲の面積を有効トラップ面積とし、任意の閉鎖個体群モデル(例えば、Otis(1978)の捕獲率均一モデル M_0 や不均一モデル M_h)による生息数推定値とあわせて生息密度を計算する手法である。この方法は簡便であるが、有効トラップ面積の計算方法はアドホックであり統計学的な根拠に乏しいとされる。なお、今回の評価における生息数の推定には、捕獲率均一モデル M_0 を用いた。

2) Inverse Prediction (Program “DENSITY”, Efford 2004)

この手法は、多数の空間的なダミーデータに基づく密度推定手法である。生息密度 D 、距離 0 の捕獲率 p_0 、距離に対する捕獲率の減衰 σ の3つのパラメータを変化させながら与えられたトラップ配置の下でデータ生成シミュレーションを多数回行い、それぞれのダミーデータについて任意の閉鎖個体群モデルによる生息数推定値 \hat{N} 、捕獲率 \hat{p} 、各個体のトラップ間移動距離の平均値を計算する。すると、パラメータベクトル $\mathbf{x} = (D, p_0, \sigma)$ と観測値ベクトル $\mathbf{y} = (\hat{N}, \hat{p}, \hat{d})$ が対応する仮想的なデータセットが得られる。ここで、観測値ベクトルはパラメータベクトルに関する多変量重回帰モデルで説明されると仮定する。

$$\begin{aligned}\mathbf{y} &= \boldsymbol{\lambda} + \mathbf{B}\mathbf{x} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &\sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma})\end{aligned}$$

$\boldsymbol{\lambda}$ は切片ベクトル、 \mathbf{B} は回帰係数行列、 $\boldsymbol{\varepsilon}$ は観測誤差ベクトル、 $\boldsymbol{\Sigma}$ は分散共分散行列である。その逆関数は下記の通りになり、シミュレーションにより得られたパラメータベクトルと観測値ベクトルからそのパラメータを推定する。

$$\begin{aligned}\mathbf{x} &= \mathbf{B}^{-1}(\mathbf{y} - \boldsymbol{\lambda}) + \mathbf{B}^{-1}\boldsymbol{\varepsilon} \\ \mathbf{B}^{-1}\boldsymbol{\varepsilon} &\sim \text{MVN}(\mathbf{0}, \mathbf{B}^{-1}\boldsymbol{\Sigma}\mathbf{B}^{-1T})\end{aligned}$$

推定された逆関数を用いることで、閉鎖個体群モデルの生息数推定値など、実際のデータから簡単に計算できる値から生息密度を算出するための計算式を得ることができる。なお、今回は生息数推定 \hat{N} 値の計算においては Otis(1978)の M_0 を用いた。

3) ベイズ空間明示型標識再捕獲モデル(Gardner et al. 2009; Royle et al. 2009)

生息密度、各個体の行動圏の位置、トラップでの捕獲の関係を階層ベイズモデルとして記述し、

検出位置情報から各個体の行動圏の中心、捕獲率と距離の関係、生息密度推定を同時に推定する手法。行動圏の中心とトラップの距離が大きいほど捕獲される確率は低下すると考えられるため、距離 r と捕獲確率 p の関係を下記のように定式化している。

$$p = p_0 \times \exp(-(r/\sigma)^2)$$

このモデルでは、捕獲確率は距離に対して釣鐘型(ガウス型)に減衰すると仮定されている。 p_0 は距離 0 の時の捕獲率、 σ は釣鐘型曲線の変曲点までの距離で表わされる捕獲確率の減衰を決めるパラメータであり、これらはデータから推定される。T セッションの標識再捕獲調査は、この確率を期待値とする繰り返し T の 2 項過程とみなすことができる。

$$Y_{ij} \sim \text{Binomial}(p, T)$$

Y_{ij} は個体 i のトラップ j における捕獲回数である。

各個体 i の行動圏の中心 s_i は未知であり、これもデータから推定する必要がある。そこで、トラップを含む任意の長方形の空間 S を設定し、 s_i に対して S の範囲に関する一様事前分布を与える。調査において 1 度も検出されなかった個体に対しても s_i を定義するが、それによって s_i の長さは生息数によって変化することになる。推定には、パラメータ値を確率的探索的に変化させながら、データに対して当てはまりのよいパラメータ値のセットを得るマルコフ連鎖モンテカルロ法(MCMC 法)を用いる。それにより、少なくとも 1 回検出された個体の s_i はトラップの周囲を動き、1 度も検出されなかった個体の s_i は消失・生成を繰り返しながら S の中をほぼ自由に動き回ることになる。MCMC 法の過程で得られる s_i の数は、空間 S 内の生息数 N とみなすことができる。したがって、生息密度は N を S の面積で割った値となる。

3. 結果・考察

ダミーデータによる推定結果を図 4、それぞれのモデルの 95%CI 幅、相対バイアスと 95%CI coverage を表 1 に示した。平均最大距離法は生息密度を過小評価する傾向があったが、少数のケースで真の値よりも高い値を示していた。平均的に 95%CI の幅は小さく、coverage は 24% であった。Inverse Prediction は推定値の 25% 点が真の値を上回っており、過大評価の傾向が見られた。ベイズ空間明示型標識再捕獲モデルは 3 つのモデルの中でもっとも相対バイアスが小さく、精度の高い推定手法であることが明らかとなった。Inverse Prediction は 95%CI 幅が 3 つのモデルの中で最も大きく、coverage では Inverse Prediction とベイズ空間標識再捕獲モデルではほぼ同等の評価となった。

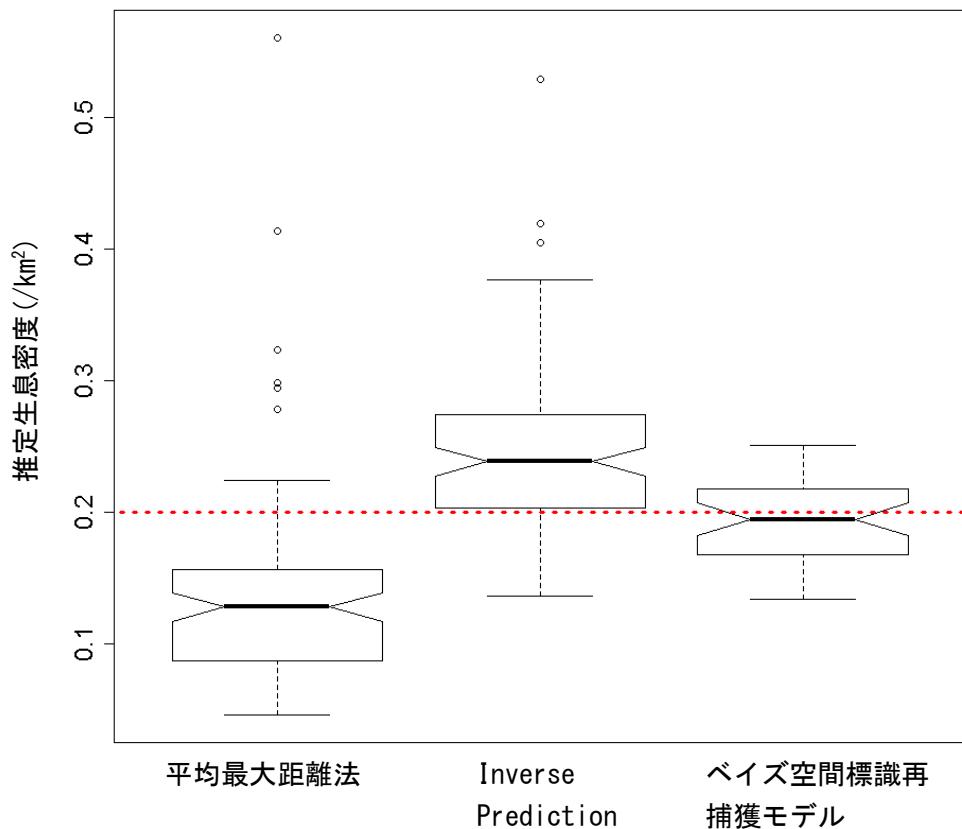


図4 ダミーデータによる生息密度推定モデルの推定結果。箱ひげ図は、内側から中央値、四分位点、四分位点×1.5の範囲の最大値を示し、○はそれより外側の値を示している。赤破線は真の生息密度 0.2 頭/km²を示す。

表1 ダミーデータ解析における各密度推定手法の 95%CI 幅、相対バイアスと 95%CI coverage

手法	平均 95%CI 幅 (95% 区間)	相対バイアス	95%CI coverage
平均最大距離法 (n=100)	0.45 (0.10-0.99)	-0.32	0.24
Inverse Prediction (n=100)	1.09 (0.82-1.57)	0.22	0.97
ベイズ空間標識再捕獲モデル (n=40)	0.67 (0.45-0.91)	-0.04	0.975

北上山地における3つのモデルの推定値を表2に示した。ダミーデータによる解析と同様、推定生息密度は平均最大距離法で小さな値を取った。Inverse Predictionとベイズ空間標識再捕獲モデルではほぼ同程度の値を示した。95%CI幅はベイズ空間標識再捕獲モデルの方がInverse Predictionよりも広かったが、その差はダミーデータの解析に比べて小さかった。

表2 北上山地のデータにおける各手法による推定生息密度

手法	推定生息密度 (95%CI) (頭/km ²)
平均最大距離法 (n=100)	0.225 (0.220-0.230)
Inverse Prediction (n=100)	0.434 (0.357-0.527)
ベイズ空間標識再捕獲モデル (n=40)	0.436 (0.359-0.518)

平均最大距離法による有効トラップ面積の計算は再捕獲の回数に影響を受けやすく、再捕獲が少ないと有効トラップ面積は小さくなり、密度を過大評価することがあると知られている(Wilson & Anderson 1985)。また、生息数の推定に用いた M_0 モデルは、個体ごとの行動圏の位置などに起因する検出率の違いがある状況では、生息数を過小評価する傾向がある。ダミーデータ解析の結果から明らかになった平均最大距離法の推定生息密度の過小評価およびばらつきの大きさは、これらのバイアス要因が複合的に作用した結果であると考えられる。Inverse Prediction は、距離に依存する捕獲率を明示的に扱っているモデルであるが、そのような空間構造は Inverse Predictionにおいては平均移動距離として集約されてしまい、各個体の捕獲位置に関する情報は考慮されない。そのことがダミーデータ解析における 95%CI 幅の広さにつながったと考えられる。その点で、捕獲位置に関する情報を最大限に活用しているベイズ空間標識再捕獲モデルはより優れていると考えられる。ただし、実際のデータに対する推定結果が示すように、大規模な調査データを用いる際には、その差はほとんど無視できる程度になる可能性もある。

その中で、Inverse Prediction を用いることにメリットがあるとすれば、それは計算時間の短さである。北上山地のデータを用いた場合、ベイズ空間標識再捕獲モデルは推定値を得るまでに 1 週間以上を要した(Core2 DUO E8600 搭載の PC を使用)。一方、Inverse Prediction では 2 分以内に結果を得ることができた。多数の場所、または広域を対象とした標識再捕獲調査を行い、推定に際してさまざまな試行錯誤を行う必要があるときには、ベイズ空間標識再捕獲モデルの適用が計算時間の面で非現実的になる場合があるかもしれない。その場合には、Inverse Prediction は比較的よい代替になりえると考えられる。

これらのことから、現時点では標識再捕獲法による生息密度推定には、ベイズ空間標識再捕獲モデルないしは Inverse Prediction を用いることが望ましいと考えられる。ベイズ空間標識再捕獲モデルはフリーソフト WinBUGS (<http://www.mrc-bsu.cam.ac.uk/bugs/>)により推定が可能であり、フリーの統計解析環境 R からデータおよびモデルコードの読み込み、推定結果の出力までの一連の操作を行うことができる。また、最近では R のパッケージ "SPACECAP" (<http://cran.r-project.org/web/packages/SPACECAP/index.html>)が公開され、GUI 上での操作で推定を実行することが可能となった。また、Inverse Prediction については、Program "DENSITY" (<http://www.otago.ac.nz/density/>)で実装されており、これについても GUI 上での操作が可能である。また、同様の機能をもつ R のパッケージ "secr" も公開されている。

引用文献

- Dice, L. R. (1938) Some census methods for mammals. *Journal of Wildlife Management* 2(2):119-130.
- Efford, M. G. (2004) Density estimation in live-trapping studies. *Oikos* 106:598-610.
- Gardner, B.; Royle, J. A.; Wegan, M. T. (2009) Hierarchical models for estimating density from DNA mark-recapture studies. *Ecology* 90(4):1106-1115.
- Otis, D. L.; Burnham, K. P.; White, G. C.; Anderson, D. R. (1978) Statistical inference from capture data on closed animal populations. *Wildlife Monographs* 62.
- Royle, J. A.; Karanth, K. U.; Gopalaswamy, A. N. Kumar, S. N. (2009) Bayesian inference in camera trapping studies for a class of spatial capture–recapture models. *Ecology* 90(11): 3233–3244.
- Wilson, K. R.; Anderson, D. R. (1985) Evaluation of two density estimators of small mammal population size. *Journal of Mammalogy* 66(1):13-21.