

# EM アルゴリズムと MCMC 法による 欠測時系列データからの個体群パラメータの最尤推定

箱山 洋<sup>1,2</sup>

水産総合研究センター<sup>1</sup>/東京海洋大学<sup>2</sup>

Hiroshi Hakoyama<sup>1,2</sup>

Fisheries Research Agency<sup>1</sup>/Tokyo University of Marine Science and Technology<sup>2</sup>

## 1 はじめに

個体数やバイオマスの時系列データには、途中で欠測値 (missing value) がある場合も多い (図 1)。天候・実体制・予算などの不確実性によって、野外調査を継続的に行うのは難しいものである。欠測値がある不完全な時系列データ (incomplete time-series data) に個体群モデルを当てはめて、パラメータを最尤推定する問題を考えてみる。ここでは、確率過程が離散時間のマルコフ過程に従う場合について、データが欠測値を含む不完全な場合の最尤推定について考察を行う。また、連続時間を仮定した個体群動態モデルの最尤推定についての短い考察も行う。

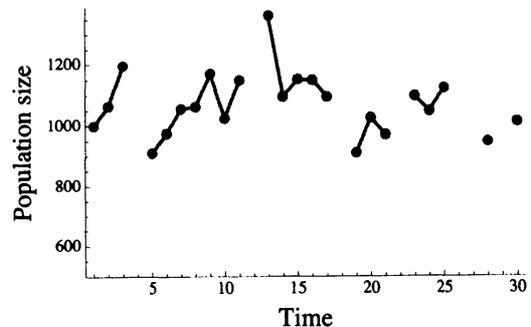


図 1 欠測値を含む不完全な時系列データ

## 2 離散時間のマルコフ過程モデル

離散時間  $t$  を仮定し ( $t = 1, 2, \dots, n$ )、個体数変動はマルコフ過程に従うとする。すなわち、時刻  $t$  の個体数  $X_t > 0$  の確率密度関数は  $f(x_t | X_{t-1} = x_{t-1}; \theta)$  であり、時刻  $t-1$  の個体数  $x_{t-1}$  の条件付き分布として陽に表されているとする。ただし、 $\theta = (\theta_1, \theta_2, \dots, \theta_s)$  はモデルの  $s$  個のパラメータのベクトルを表す。このとき、 $X_1, X_2, \dots, X_n$  の同時密度関数は、 $f(x_1, x_2, \dots, x_n; \theta) = \left( \prod_{t=1}^{n-1} f(x_{t+1} | x_t; \theta) \right) f(x_1; \theta)$  となる。 $f(x_1; \theta)$  は、初期個体数  $X_1$  の周辺密度である。

### 2.1 完全データの最尤推定量

欠測値のない完全な時系列データ  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  が得られたとき、その対数尤度は

$$\ln L^c(\theta; \mathbf{x}) = \ln f(\mathbf{x}; \theta) = \sum_{t=1}^{n-1} \ln f(x_{t+1} | x_t; \theta) + \ln f(x_1; \theta), \quad (1)$$

となる。ただし、 $L^c$  は完全データの尤度を表す。式 1 の右辺第二項は初期個体数の周辺分布から計算される

部分であるが、この部分のために尤度が煩雑になり計算が難しくなりやすい。そこで、個体数初期値の持つ情報を捨てて、式 1 の右辺第一項のみからなる条件付対数尤度関数 (conditional log-likelihood) で条件付最尤推定量を構成する方法もある。データ量  $n$  が大きければ、右辺第二項を無視しても影響は小さい。最尤推定量  $\hat{\theta}$  は、 $\hat{\theta} = \arg \max_{\theta} \ln L^c(\theta; \mathbf{x})$  と定義される。以下、解析的もしくは数値的に完全データに対して最尤推定値を計算することができる場合を考える。

## 2.2 不完全データの最尤推定量

欠測値を含む不完全な時系列データから個体群パラメータの最尤推定を行う場合、完全データ (complete data) の最尤推定量をそのまま使うことはできない。例として、データ  $\mathbf{x} = (x_1, x_2, \text{NA}, x_4, \text{NA}, \text{NA}, x_7)$  が得られた場合を考えてみる。NA は欠測値を表す。この場合、この不完全データの対数尤度は次のようになる：

$$\ln L(\theta; \mathbf{x}) = \ln f(x_2 | x_1; \theta) + \ln f(x_4 | x_2; \theta) + \ln f(x_7 | x_4; \theta) + \ln f(x_1; \theta), \quad (2)$$

ただし、 $L$  は観察した不完全データ (incomplete data) の尤度を表す。式 2 の右辺第一項はマルコフ過程モデルの定義ですでに与えられている簡単な式であるが、欠測値があるために右辺第二項と第三項はそれぞれ 2 年前と 3 年前の個体数の条件付き確率密度関数を計算しなければならない。これらは、チャップマン-コルモゴロフの方程式 (see Goel and Richter-Dyn, 1974) から次のような積分式として表すことができる：

$$f(x_4 | x_2; \theta) = \int_0^{\infty} f(x_4 | x_3; \theta) f(x_3 | x_2; \theta) dx_3, \quad (3)$$

$$f(x_7 | x_4; \theta) = \int_0^{\infty} \int_0^{\infty} f(x_7 | x_6; \theta) f(x_6 | x_5; \theta) f(x_5 | x_4; \theta) dx_5 dx_6. \quad (4)$$

すなわち、式 2 の対数尤度には、条件付き確率密度関数の積の一重積分と二重積分の式が含まれている。一般には、 $t$  年と  $t-s$  年のデータの間に  $s-1$  点の欠測値がある場合に条件付き確率密度関数  $f(x_t | x_{t-s}; \theta)$  は、条件付き確率密度関数の積の  $s$  重積分の形で表すことができる。このような多重積分の項を含む尤度関数のパラメータ  $\theta$  に関する最大化を数値的に行うのは困難である。このため、欠測値を含むデータを用いて最尤推定を行うためには、EM アルゴリズム (Dempster et al., 1977) を用いるなどの工夫が必要となる。

## 2.3 EM アルゴリズム

EM アルゴリズムは、完全データに対する対数尤度関数について、パラメータの推定値と観測データの条件付き期待値をとり、最大化することで、パラメータの推定値を逐次更新して行き、欠測値のある観測データに対するパラメータの最尤推定値を得る方法である (Dempster et al., 1977)。

個体数  $X_t$  に関するマルコフ確率過程モデルに対して、 $X_t$  の時系列  $X_1, \dots, X_n$  のうち観測値の確率変数ベクトルを  $\mathbf{X}_{obs}$ 、欠測値のそれを  $\mathbf{X}_{mis}$  とする。また、実現値ベクトルはそれぞれ  $\mathbf{x}_{obs}$ 、 $\mathbf{x}_{mis}$  とする。完全データは、 $\mathbf{x} = \{\mathbf{x}_{obs}, \mathbf{x}_{mis}\} = (x_1, \dots, x_n)$  である。このとき、EM アルゴリズムでは、 $k$  段階目のパラメータベクトル  $\theta$  の推定値  $\theta^{(k)}$  が得られたとして、観測値ベクトル  $\mathbf{x}_{obs}$  およびパラメータ  $\theta^{(k)}$  の条件のもとで、欠測値ベクトル  $\mathbf{x}_{mis}$  について対数尤度の条件付期待値  $Q$  を計算する (E ステップ, expectation step)：

$$Q(\theta|\theta^{(k)}) = \int_{\Omega} l^c(\mathbf{x}_{obs}, \mathbf{x}_{mis}; \theta) f_{X_{mis}}(\mathbf{x}_{mis}|\mathbf{x}_{obs}; \theta^{(k)}) d\mathbf{x}_{mis}, \quad (5)$$

ただし、 $l^c(\mathbf{x}; \theta) = \ln L^c(\mathbf{x}; \theta)$  は完全データに対する対数尤度関数、 $f_{X_{mis}}$  は  $\mathbf{x}_{mis}$  の条件付き確率密度関数である。 $\Omega$  は  $\mathbf{x}_{mis}$  の標本空間である。つぎに、Eステップで書き下すことができた  $Q(\theta|\theta^{(k)})$  を  $\theta$  に関して最大化することで  $\theta^{(k+1)}$  を計算する (Mステップ, maximization step) :

$$\theta^{(k+1)} = \arg \max_{\theta} Q(\theta|\theta^{(k)}). \quad (6)$$

このように、EM アルゴリズムでは E ステップと M ステップからなる手順を繰り返すことでパラメータの推定値を更新して行く。 $\theta^{(k+1)}$  は、 $L(\mathbf{x}_{obs}; \theta^{(k+1)}) \geq L(\mathbf{x}_{obs}; \theta^{(k)})$  を満たしており、イタレーションを繰り返すことで、観察した不完全データの尤度  $L$  は単調増加することが証明できる (Dempster et al., 1977)。さらに、尤度関数が凹で単峰であれば、パラメータ空間上の任意の初期値  $\theta^{(1)}$  から始めて十分なイタレーションを行うことで、 $\theta^{(k)}$  は最尤推定値に収束する (Wu, 1983)。

このように、EM アルゴリズムによる最尤推定では、煩雑で最大化の難しい不完全データの尤度 (式 2) は使わずに、扱いやすい完全データの尤度 (式 1) を用いて、不完全データに対する最尤推定値を得る。線形モデルの場合などでは (例えば、混合正規分布のパラメータ推定; see McLachlan and Krishnan, 2007)、積分の外れた簡単な式として  $Q(\theta|\theta^{(k)})$  が計算できて、比較的簡単に EM アルゴリズムの手順を構成することができる。

## 2.4 モンテカルロ積分

しかしながら、ここで考えている非線形なマルコフ過程モデルなど、一般には、式 5 の右辺の積分を解析的に評価して、 $Q(\theta|\theta^{(k)})$  を最大化することは難しい。解析的に完全データの最尤推定値を書き下すことができる場合でも、式 5 の右辺の積分が煩雑な形になるのが普通である。そこで、式 5 をモンテカルロ積分で置き換えることが考えられる :

$$\tilde{Q}(\theta|\theta^{(k)}) = \frac{1}{\eta} \sum_{i=1}^{\eta} l^c(\mathbf{x}_{obs}, \mathbf{x}_{mis}^{(i)}; \theta), \quad (7)$$

ただし、 $\mathbf{x}_{mis}^{(i)}$  は、確率密度関数  $f_{X_{mis}}(\mathbf{x}_{mis}|\mathbf{x}_{obs}, \theta^{(k)})$  からモンテカルロ法によって発生させる欠測ベクトルデータの  $i$  番目の値である。 $l^c$  が書き下された式であるとすれば、 $\tilde{Q}(\theta|\theta^{(k)})$  も書き下された式となり、多くの場合、 $\tilde{Q}(\theta|\theta^{(k)})$  を  $\theta$  について最大化することができる。このように、 $Q(\theta|\theta^{(k)})$  にモンテカルロ積分を用いて EM アルゴリズムを構成する方法は、MCEM (Monte Carlo EM) アルゴリズムと呼ばれている (Wei and Tanner, 1990)。

## 2.5 乱数の発生

離散時間の非線形マルコフ過程モデルに対する最尤推定値の計算に、MCEM アルゴリズムを用いるとして、次の問題は乱数の発生である。明示的に確率密度関数  $f_{X_{mis}}(\mathbf{x}_{mis}|\mathbf{x}_{obs}, \theta^{(k)})$  を書き下すことができて、それに従う乱数を発生させるのは、一般には簡単ではない。ある確率分布に従う乱数を発生させる方法には、分

布関数の逆関数が書き下せる場合の逆関数法や、確率密度関数を上から覆う関数が見つかる場合の棄却法なども考えられるが<sup>3</sup>(see Knuth, 1997)、ここではMCMC(Markov chain Monte Carlo)法 (Metropolis et al., 1953) を用いて、 $f_{X_{mis}}(x_{mis}|x_{obs}, \theta^{(k)})$  に従う疑似乱数を発生させることを考える。

### 2.5.1 Metropolis-Hastings algorithm

まず、 $x_{mis}$  を一度に発生させるのではなく、欠測が1点だけある場合に欠測値の分布から疑似乱数を発生させる問題を考える。三年間 ( $t = 1, 2, 3$ ) の不完全データ  $(x_1, NA, x_3)$  が得られたとき、欠測値  $x_2$  の条件付き確率分布は、条件付き確率の定義とチャップマン-コルモゴロフの方程式から次のようになる：

$$f(x_2 | x_3, x_1; \theta) = \frac{f(x_2, x_3 | b_1; \theta)}{f(x_3 | b_1; \theta)} = \frac{f(x_3 | x_2; \theta)f(x_2 | x_1; \theta)}{\int_0^\infty f(x_3 | x_2; \theta)f(x_2 | x_1; \theta)dx_2}. \quad (8)$$

$f(x_2 | x_3, x_1; \theta)$  に従う疑似乱数を発生させるには、MCMC法の一つであるMetropolis-Hastings algorithm (Metropolis et al., 1953) を使うことが考えられる。MCMC法は、計算機のパッケージソフトで簡単に疑似乱数を発生させることができる確率分布（一様分布や正規分布など；提案分布と呼ぶ）から疑似乱数を発生させ、目的の複雑な確率分布に従う疑似乱数を得る方法である。MCMC法では、まず提案分布から疑似乱数を発生させ、マルコフ過程の詳細釣り合い条件を満たすように決めたある棄却確率で、発生させた乱数を目的の分布からの乱数として採択する。ここでは、提案分布として目的の分布をできるだけ近似した固定した平均と分散を持つ切断正規分布（定義域が非負実数）を用いる独立連鎖を考える。切断正規分布である提案分布の決め方としては、例えば、 $f(x_2 | x_3, x_1; \theta)$  の分布が単峰形の場合、提案分布の平均  $m$  は  $f(x_2 | x_3, x_1)$  のモード、分散  $\sigma^2$  はモードにおける  $-\left(\frac{d^2 f(x_2 | x_3, x_1)}{dx_2^2}\right)^{-1}$  とする。 $f(x_2 | x_3, x_1; \theta)$  の分布が、二峰形（多峰形）の場合は単峰形の方法が使えないため、 $f(x_2 | x_3, x_1; \theta)$  の期待値と分散を提案分布の平均と分散とするなどが考えられる。連続して発生させる疑似乱数列の現在の値が  $x_2 = x_2^{(i)}$  で、新しく提案分布から発生された疑似乱数を  $x_2'$  とするとき、 $x_2'$  を次の確率で受容する：

$$\begin{aligned} \alpha(x_2, x_2') &= \min \left( \frac{f(x_2' | x_3, x_1; \theta)q(x_2)}{f(x_2 | x_3, x_1; \theta)q(x_2')}, 1 \right), \\ &= \min \left( \frac{f(x_3 | x_2'; \theta)f(x_2' | x_1; \theta) \exp\left(\frac{(x_2 - m)^2}{2\sigma^2}\right)}{f(x_3 | x_2; \theta)f(x_2 | x_1; \theta) \exp\left(\frac{(x_2' - m)^2}{2\sigma^2}\right)}, 1 \right), \end{aligned} \quad (9)$$

ただし、 $q(x_2)$  は提案分布の切断正規分布である。 $x_2'$  が受容されたとき、 $i + 1$  番目の乱数  $x_2^{(i+1)}$  は、 $x_2^{(i+1)} = x_2'$  とし、棄却した場合は  $x_2^{(i+1)} = x_2$  とする。式9において、うまく目的分布や提案分布の分母の積分項がキャンセルする。このようにして、MCMC法によって  $f(x_2 | x_3, x_1; \theta)$  に従う疑似乱数を発生させることができる。 $f$  の分布のほとんどが単峰形であれば、採択率も高くなり、効率よく疑似乱数を発生させることができる。

### 2.5.2 Metropolis sampling within Gibbs method

疑似乱数ベクトル  $x_{mis}^{(i+1)}$  は一度に発生させなくてもよい。ギブス・サンプラー (Geman and Geman, 1984) のように、 $i$  番目の欠測値ベクトルが  $x_{mis}^{(i)}$  が与えられているとき、 $i + 1$  番目の欠測値ベクトルのなかの  $j$  番目の疑似乱数を得るには、 $(x_{mis,1}^{i+1}, \dots, x_{mis,j-1}^{i+1}, \dots, x_{mis,j+1}^i, x_{mis,jmax}^i)$  が与えられている条件付きで  $x_{mis,j}^{i+1}$  を一つ一つ発生させればよい。すなわち、前節で考えた欠測値が1点だけある場合を連続して適用することで  $x_{mis}^{(i+1)}$  を発生させることができる。この方法はMetropolis sampling within Gibbs methodと呼ばれる。

## 2.6 MCEM アルゴリズムの手順

ここまでの手順をまとめる。まず初期値（1段階目の推定）については、

1. 初期パラメータはつながっているデータだけで推定
2. 初期欠測値は初期パラメータを仮定して、最尤推定

つぎに、 $k$  段階目の MCEM アルゴリズムの手順は、

1. 完全データの対数尤度の条件付き期待値  $\hat{Q}(\theta|\theta^{(k)})$  をモンテカルロ積分で計算する
2. モンテカルロ積分中の欠測値の疑似乱数は Metropolis within Gibbs method で発生
3. 完全データの最尤推定量（解析もしくは数値計算）を用いて、 $\hat{Q}(\theta|\theta^{(k)})$  を最大化して、 $\theta^{(k+1)}$  を得る

このようにして、離散時間のマルコフ過程モデルを個体数データに当てはめて、そのモデルのパラメータを推定する問題は、MCEM アルゴリズムを適用して解決できる。

## 2.7 分散の推定

最尤推定量の分散や信頼区間を評価する方法には、データ数が十分に大きいことを仮定する漸近的方法と、データ量が少ない場合のモンテカルロ法を用いた方法がある。

### 2.7.1 漸近的な分散の推定

データ量が十分大きく、完全データの場合には、よく知られているように観測情報行列  $J(\hat{\theta}, \mathbf{x})$  の逆行列によって、最尤推定量の分散を推定することができる。不完全データの観測情報行列については、完全観測情報行列から欠測情報行列を除する関係が成立しており、次のように表すことができる (Louis, 1982) :

$$J(\hat{\theta}, \mathbf{x}_{obs}) = E_{\hat{\theta}} \left[ -\frac{\partial^2 \log f_{\theta}(\mathbf{x})}{\partial \theta \partial \theta'} \Big|_{\theta=\hat{\theta}} \Big| \mathbf{X}_{obs} = \mathbf{x}_{obs} \right] \\ - E_{\hat{\theta}} \left[ \left( \frac{\partial \log f_{\theta}(\mathbf{x})}{\partial \theta} \Big|_{\theta=\hat{\theta}} \right) \left( \frac{\partial \log f_{\theta}(\mathbf{x})}{\partial \theta'} \Big|_{\theta=\hat{\theta}} \right)' \Big| \mathbf{X}_{obs} = \mathbf{x}_{obs} \right].$$

この不完全データの観測情報行列についても、モンテカルロ積分で計算できる (Wei and Tanner, 1990)。

### 2.7.2 Monte Carlo Bias Correction

ここでは詳しく述べないが、Hakoyama and Iwasa (2000) は、データ数が少ない場合に、モンテカルロ法を用いて最尤推定量のバイアスを修正する方法 (Monte Carlo Bias Correction, MCBC) や、MCBC と同様のアルゴリズムで漸近的に正確な信頼区間を求める方法を開発した。今回の不完全データから個体群パラメータを推定する問題にも適用可能である。

## 2.8 欠測の間隔と情報量

欠測の間隔が1年、2年と長くなって行くとき、パラメータ推定における情報量はどうなるだろうか？ 情報量の変化は、おそらくパラメータによって異なると考えられる。例えば、定常過程における平衡個体数の推定においては、欠測値があつて時間間隔が長い  $n$  点からなる時系列データと時間間隔が短い  $n$  点からなる完全

データでは情報量には差がないであろう。しかしながら、増殖率のようなパラメータは欠測の間隔が長くなるほど情報が失われていき、観測された2点の間がある欠測間隔以上では情報を持たないと予想される。

### 3 連続時間の確率過程モデルにおける最尤推定

連続時間を仮定し、個体群動態の確率過程モデルに確率微分方程式を採用すると、データはサンプルパスの一部のある時間間隔で観察したものと考えることができる。すなわち、連続時間のモデルの場合、サンプルパスのすべてを観察することはできないので、観測データは常に欠測値を含む不完全な時系列データである。仮定した確率微分方程式について、時刻  $t$  の個体数  $X_t$  の強解が陽に得られれば、さまざまな時間間隔で観察されたデータについて最尤推定量を構成することができ、モデルのパラメータを推定することができるはずである。しかしながら、陽な強解が得られるのは線形方程式や一部の非線形方程式の場合に限られていて、一般には解を得るのは難しい。また、仮に非線形確率微分方程式の強解が得られたとしても、解には積分項が含まれていて、数値的に尤度関数を最大化するのは難しい。例えば、環境変動とロジスティック式を仮定した確率微分方程式は個体群動態のモデルとして考慮に値するが、その強解には積分項が含まれている (Øksendal, 1998)。確率微分方程式の確率過程に対して、近似尤度に基づく推定が可能な場合もある (Hakoyama and Iwasa, 2000)。Hakoyama and Iwasa (2000) は、強解を陽に得ることが難しいモデルに対して、環境変動が小さい仮定を置き、準平衡個体数の周りで非線形モデルを線形化し、近似尤度に基づく推定法を構成した。結論として、連続時間の非線形な確率過程モデルを仮定すると、最尤推定量を構成して推定値を得ること自体が難しい。

### 参考文献

- Dempster, A. P., N. M. Laird, and D. B. Rubin (1977) "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society Series B-Methodological*, Vol. 39, No. 1, pp. 1-38.
- Geman, S and D Geman (1984) "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *Ieee Transactions On Pattern Analysis and Machine Intelligence*, Vol. 6, No. 6, pp. 721-741.
- Goel, N. S. and N. Richter-Dyn (1974) *Stochastic models in biology*: Academic Press.
- Hakoyama, H. and Y. Iwasa (2000) "Extinction risk of a density-dependent population estimated from a time series of population size," *Journal of theoretical Biology*, Vol. 204, No. 3, pp. 337-359.
- Knuth, DE (1997) *The Art of Computer Programming, Vol. 2: Seminumerical Algorithms*: Addison-Wesley Professional, 3rd edition.
- Louis, T. A. (1982) "Finding the Observed Information Matrix when Using the EM Algorithm," *Journal of the Royal Statistical Society Series B-Methodological*, Vol. 44, No. 2, pp. 226-233.
- McLachlan, G. J. and T. Krishnan (2007) *The EM algorithm and extensions*, Vol. 382: Wiley-Interscience.
- Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953) "Equation of state calculations by fast computing machines," *The journal of chemical physics*, Vol. 21, p. 1087.
- Øksendal, B. (1998) *Stochastic differential equations*: Springer.
- Wei, G. C. G. and M. A. Tanner (1990) "A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms," *Journal of the American Statistical Association*, Vol. 85, No. 411, pp. 699-704, September.
- Wu, C. F. J. (1983) "On the convergence properties of the EM algorithm," *Annals of Statistics*, Vol. 11, No. 1, pp. 95-103.