

# 有限状態マルコフ連鎖モデルにおける状態分布の将来予測と推定精度

広島大学 大学院医歯薬保健学研究院 疫学・疾病制御学  
秋田 智之、大久 真幸、坂宗 和明、田中 純子

## 1. はじめに

有限状態マルコフ連鎖モデル (以下マルコフモデル) は初期状態と確率推移行列のべき乗の積により未来の状態分布が記述できるという利便性から集団遺伝学、在庫管理、病態 stage の推移など様々な現象の記述やシミュレーションに利用されている。ところで、実際のデータから確率推移行列を推定するときに、標本サイズにより確率推移行列およびそれにより推定される未来の予測値の精度が大きく左右されると考えられる。そこで、2012 年度統計関連学会連合大会において状態分布の信頼区間の構成法について示した。本報告では、状態数や標本数を変えた時の精度についてシミュレーションを基に考察する。さらに漸近近似の次数を上げることで精度がどの程度改善するのかについて検討した。

## 2. マルコフモデルの定義および状態分布の点推定

離散型確率過程  $\{X_n\}$  の取り得る値 (状態) を  $\{1, 2, \dots, S\}$  とする。このとき、任意の  $n$ 、任意の履歴  $x_0, x_1, \dots, x_n, x_{n+1}$  に対して  $P(X_{n+1}=x_{n+1}|X_0=x_0, X_1=x_1, \dots, X_n=x_n)=P(X_{n+1}=x_{n+1}|x_n=x_n)$  が成立するとき、この確率過程を有限状態マルコフ連鎖モデルという。また、 $p_{jk}=P(X_{n+1}=k|x_n=j)$  を  $(j, k)$  成分とした行列  $\mathbf{P}$  を確率推移行列という。このとき、状態  $i$  から状態  $j$  への推移の回数を  $y_{ij}$  とおくと、推移確率の最尤推定量は  $\hat{p}_{ij}=y_{ij}/y_i$  が得られる。 $\hat{p}_{ij}$  を  $(i, j)$  成分とした推定確率推移行列  $\hat{\mathbf{P}}$  により、初期分布を  $\boldsymbol{\pi}_0$  としたとき  $n$  期後の状態確率分布が  $\boldsymbol{\pi}_n=\boldsymbol{\pi}_0\hat{\mathbf{P}}^n$  で推定される。

## 3. 信頼区間算出のためのアルゴリズム

最尤推定量の漸近正規性から推移確率  $\hat{p}_{ij}$  は漸近的に正規分布に従うので、それらの多項式である  $\boldsymbol{\pi}_n=(\boldsymbol{\pi}_{n1}(\mathbf{p}), \boldsymbol{\pi}_{n2}(\mathbf{p}), \dots, \boldsymbol{\pi}_{nS}(\mathbf{p}))$  の各項  $\boldsymbol{\pi}_{ni}(\mathbf{p})$  は漸近的に正規分布に従う。ここで  $\mathbf{p}$  は  $\hat{\mathbf{p}}$  を並べた確率ベクトルである。故に 95% 信頼区間は  $\boldsymbol{\pi}_{ni}(\mathbf{p}) \pm 1.96\text{SE}(\boldsymbol{\pi}_{ni}(\mathbf{p}))$  となる。状態数  $S=2$  のときのアルゴリズムは以下のようになる。

### $N$ 期後状態分布の信頼区間算出のためのアルゴリズム (状態数 2 のとき)

**Step 1.** データ行列  $\mathbf{X}$  から列和  $y_i$ 、確率推移行列の推定値  $\hat{\mathbf{P}}$  を算出する。

**Step 2.** 以下を  $n=1, 2, \dots, N$  だけ繰り返す。

$$\hat{\boldsymbol{\pi}}_{n+1} \leftarrow \hat{\boldsymbol{\pi}}_n \hat{\mathbf{P}}$$

$$\text{SE}(\hat{\boldsymbol{\pi}}_{n+1}) \leftarrow \sqrt{\left\{ \boldsymbol{\pi}_0 \left( \frac{\partial \mathbf{P}^n}{\partial \mathbf{p}_1} \Big|_{\mathbf{p}=\hat{\mathbf{p}}} \right) \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}^2 \frac{\hat{p}_1(1-\hat{p}_1)}{y_1} + \left\{ \boldsymbol{\pi}_0 \left( \frac{\partial \mathbf{P}^n}{\partial \mathbf{p}_2} \Big|_{\mathbf{p}=\hat{\mathbf{p}}} \right) \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}^2 \frac{\hat{p}_2(1-\hat{p}_2)}{y_2}}$$

$$\frac{\partial \mathbf{P}^{n+1}}{\partial \mathbf{p}_1} \Big|_{\mathbf{p}=\hat{\mathbf{p}}} \leftarrow \hat{\mathbf{P}} \frac{\partial \mathbf{P}^n}{\partial \mathbf{p}_1} \Big|_{\mathbf{p}=\hat{\mathbf{p}}} + \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} \hat{\mathbf{P}}^n$$

$$\frac{\partial \mathbf{P}^{n+1}}{\partial \mathbf{p}_2} \Big|_{\mathbf{p}=\hat{\mathbf{p}}} \leftarrow \hat{\mathbf{P}} \frac{\partial \mathbf{P}^n}{\partial \mathbf{p}_2} \Big|_{\mathbf{p}=\hat{\mathbf{p}}} + \begin{pmatrix} 0 & 0 \\ 1 & -1 \end{pmatrix} \hat{\mathbf{P}}^n$$

**Step 3.**  $N$  期後の状態 1 の推定値は次で得られる： $\boldsymbol{\pi}_{N1} \pm 1.96\text{SE}(\boldsymbol{\pi}_{N1})$

シミュレーションの詳細、改良方法については当日報告する。