時系列分析 3
1 変量時系列モデルの推定法

担当:　長倉　大輔
(ながくらだいすけ)

1 変量時系列モデルの推定法

- AR(p)モデルの推定

AR(p)モデルの推定法として代表的なものには
最小二乗法と**最尤推定法**がある。

- 最小二乗法

最小二乗法(Ordinary Least Square, 略してOLS) とは、残差を最小にするようにパラメーターを推定する方法である。

以下AR(1)モデルを用いて説明する。AR(p)モデルの場合も**考え方は同じ**である。

1 変量時系列モデルの推定法

- 最小二乗法

AR(1) モデル:

$$y_t = c + \phi_1 y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{W.N.}(\sigma^2)$$

において、未知パラメーターは c, ϕ_1, σ^2 の3つ。

まず c と ϕ_1 の推定について考えよう。これらの任意の推定値(量)を $\tilde{c}, \tilde{\phi}_1$ と書く事にしよう。この時

$$e_t = y_t - \tilde{c} - \tilde{\phi}_1 y_{t-1}$$

を**残差**という。これはモデルで説明できない部分を表している。

1 変量時系列モデルの推定法

- 最小二乗法

観測値 y_t $(t=1,\ldots,T)$ が与えられた時に、これら残差を二乗して $t=2$ から T まですべて足し合わせたもの:

$$SSR=\sum_{t=2}^{T}e_t^2=\sum_{t=2}^{T}(y_t-\tilde{c}-\tilde{\phi}_1 y_{t-1})^2$$

を**残差平方和**(sum of squared residuals: SSR)とよぶ。

最小二乗推定値 $\hat{c}$, $\hat{\phi}_1$ とはこの**残差平方和を最小にするような** c と ϕ_1 の推定値の事である。

1 変量時系列モデルの推定法

- 最小二乗法

最小二乗推定値はSSRを $\tilde{c}$ と $\tilde{\phi}_1$ で偏微分したものを 0 とおいた以下の**2つの方程式**を解くことによって求められる。

$$
\begin{cases}
\dfrac{\partial SSR}{\partial \tilde{c}} = -2\sum_{t=2}^{T}(y_t - \tilde{c} - \tilde{\phi}_1 y_{t-1}) = 0 \\
\dfrac{\partial SSR}{\partial \tilde{\phi}_1} = -2\sum_{t=2}^{T} y_{t-1}(y_t - \tilde{c} - \tilde{\phi}_1 y_{t-1}) = 0
\end{cases}
$$

これらを $\tilde{c}$, $\tilde{\phi}_1$ について解くと

1 変量時系列モデルの推定法

- 最小二乗法

$$\hat{\phi}_1 = \frac{\sum_{t=2}^{T}(y_t - \bar{y}_{2,T})(y_{t-1} - \bar{y}_{1,T-1})}{\sum_{t=2}^{T}(y_{t-1} - \bar{y}_{2,T})^2}, \quad \hat{c} = \bar{y}_{2,T} - \hat{\phi}_1 \bar{y}_{1,T-1}$$

という最小二乗推定量(値)を得る。ここで

$$\bar{y}_{r,s} = (s - r + 1)^{-1} \sum_{t=r}^{s} y_t$$

である。

1 変量時系列モデルの推定法

- 最小二乗法

σ^2の最小二乗推定量(値)はOLS残差：

$$\hat{e}_t = y_t - \hat{c} - \hat{\phi}_1 y_{t-1}$$

を用いて

$$\hat{\sigma}^2 = (T-3)^{-1} \sum_{t=2}^{T} \hat{e}_t^2$$

と定義される。

1 変量時系列モデルの推定法

- AR(p)モデル

AR(p) モデルの場合は未知パラメーターが $c,\ \phi_1,\dots,\ \phi_p$, σ^2 の$p+2$個になる。よって $\phi_1,\dots,\phi_p$を推定するための残差平方和に関して、その残差は

$$e_t = y_t - \tilde{c} - \tilde{\phi}_1 y_{t-1} \cdots - \tilde{\phi}_p y_{t-p}$$

となる。これより計算したSSRを最小化するように最小二乗推定する。

1 変量時系列モデルの推定法

また、この時 σ^2のOLS 推定量は残差平方和を

標本数 $(T-p)$ – 係数パラメーターの数$(p+1)$

で割ったもの、すなわち

$$\hat{\sigma}^2 = (T-2p-1)^{-1}\sum_{t=p+1}^{T}\hat{e}_t^2$$

と定義される(係数パラメーターとはOLSで推定した c, $\phi_1,\ldots,\phi_p$ のこと)。観測値が $t=1,\ldots,T$ のT 時点の場合には推定には $T-p$ 個のデータ(の組)しか使用できないため残差平方は $t=p+1$ から T までの和となる事に注意。

1 変量時系列モデルの推定法

例題 1

AR(1) モデル:

$$y_t = \phi_1 y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{W.N.}(1),\ t = 1,\ldots,T,$$

の ϕ_1 の最小二乗推定量を求めなさい。

例題 2

先ほどは y_t $(t = 1,\ldots., T)$ という T 個の観測値が得られた場合のOLSを考えた。今 y_t $(t = 0,\ldots,T)$ という $T + 1$ 個の観測値が得られている場合にはOLSはどのようになるだろうか?

1 変量時系列モデルの推定法

- 最小二乗法の利点

 (1) 計算が簡単である。

 (2) 一致性を持つ。

 (3) ε_t に 特定の分布を仮定する必要がない。

最小二乗法はAR(p)モデルの推定に最もよく使われる。MA(q)やARMA(p, q)モデルの推定には、次の最尤推定法が最もよく使われる。

1 変量時系列モデルの推定法

■ 最尤推定法

最小二乗法は ε_t に特定の分布を仮定しなかったが、
最尤法は ε_t の**分布を明示的に仮定する**。

ε_t の分布としてもっとも頻繁に仮定されるのは**正規分布**である。$\varepsilon_t \sim N(0, \sigma^2)$であるとするとその**密度関数**は

$$f(\varepsilon_t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\varepsilon_t^2}{2\sigma^2} \right]$$

となる。

1変量時系列モデルの推定法

- 最尤推定法

最尤推定法は**尤度関数を最大化する**ようなパラメーターを推定値とするような方法である。

以下AR(1)の場合について説明する(MAモデルやARMAモデルの場合も**考え方は同じ**である)。

1 変量時系列モデルの推定法

- AR(1)モデルの最尤推定

以下のAR(1) モデルのパラメーターを最尤法で推定する:

$$y_t = c + \phi_1 y_{t-1} + \varepsilon_t, \ \varepsilon_t \sim i.i.d.N(0, \sigma^2),$$

まず尤度関数を導出する。

尤度関数とは$\{y_1, \ldots, y_T\}$の**結合密度関数** $f(y_1, \ldots, y_T)$ (をパラメーターの関数としてみなしたもの)である。

1 変量時系列モデルの推定法

- AR(1)モデルの最尤推定

時系列分析では通常、このような結合密度関数を**条件付き密度関数の積**として表す。

一般に $(x_n,\dots,x_1)$ という n 変量の確率変数の結合密度関数 $f(x_n,\dots,x_{k+1},x_k,x_{k-1},\dots,x_1)$は任意の k に対して

$$
\begin{aligned}
&f(x_n,\dots,x_{k+1},x_k,x_{k-1},\dots,x_1)\\
&= f(x_k \mid x_n, x_{n-1},\dots,x_{k+1},x_{k-1},\dots,x_1)\\
&\quad\times f(x_n,x_{n-1},\dots,x_{k+1},x_{k-1},\dots,x_1)
\end{aligned}
$$

と表す事ができる。

1 変量時系列モデルの推定法

- AR(1)モデルの最尤推定

これより $\{ y_T, y_{T-1}, \ldots, y_1 \}$ の結合密度関数は

$$f(y_T, y_{T-1}, \ldots, y_1)$$
$$= f(y_T \mid y_{T-1}, y_{T-2}, \ldots, y_1)\, f(y_{T-1}, y_{T-2}, \ldots, y_1)$$

と表す事ができる。同様に$\{ y_{T-1}, y_{T-2}, \ldots, y_1 \}$ の結合密度関数は

$$f(y_{T-1}, y_{T-2} \ldots, y_1) = f(y_{T-1} \mid y_{T-2}, \ldots, y_1)\, f(y_{T-2}, \ldots, y_1)$$

と表す事ができる。

1 変量時系列モデルの推定法

- AR(1)モデルの最尤推定

これを繰り返す事によって

$$
\begin{aligned}
&f(y_T, y_{T-1},\dots, y_1) \\
&= f(y_T \mid y_{T-1},\dots, y_1)\, f(y_{T-1} \mid y_{T-2},\dots, y_1)\times \\
&\quad \cdots \times f(y_2 \mid y_1)\, f(y_1)
\end{aligned}
$$

と表す事ができる。

上記の式は(AR(1)モデルでなくても)一般的に(どのような時系列モデルに対しても)成り立つ

1 変量時系列モデルの推定法

■ AR(1)モデルの最尤推定

AR(1)モデルの場合は y_t が 1 時点前の y_{t-1} にしか依存しないので、

$$f(y_t \mid y_{t-1}, y_{t-2}, \ldots, y_1) = f(y_t \mid y_{t-1})$$ となる。

これより

$$\begin{aligned}
&f(y_T, y_{T-1}, \ldots, y_1) \\
&= f(y_T \mid y_{T-1}) f(y_{T-1} \mid y_{T-2}) \cdots f(y_2 \mid y_1) f(y_1) \\
&= \left(\prod_{t=2}^{T} f(y_t \mid y_{t-1}) \right) f(y_1)
\end{aligned}$$

となる。

1 変量時系列モデルの推定法

- AR(1)モデルの最尤推定

条件付き密度関数 $f(y_t \mid y_{t-1})$はどのような形をとるだろうか?

今 $\varepsilon_t \sim N(0, \sigma^2)$なので、$y_{t-1}$ が与えられた時の y_t の条件付き分布は $y_t \mid y_{t-1} \sim N(c+\phi_1 y_{t-1}, \sigma^2)$となる。

よってy_t の y_{t-1} という条件付き密度関数 $f(y_t \mid y_{t-1})$は

$$f(y_t \mid y_{t-1}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_t - c - \phi_1 y_{t-1})^2}{2\sigma^2} \right]$$

となる。

1 変量時系列モデルの推定法

- AR(1)モデルの最尤推定

またε_t が正規分布に従っている時、AR(1) モデルにおいて y_t の**定常(無条件)分布**は

$$y_t \sim N\left(\frac{c}{1-\phi_1}, \frac{\sigma^2}{1-\phi_1^2}\right)$$

となる事を示す事ができる。y_1はそれ以前の値が観測されないので、この定常分布に従うとすると、その密度関数は

$$f(y_1) = \frac{1}{\sqrt{2\pi[\sigma^2/(1-\phi_1^2)]}} \exp\left[-\frac{(y_1-[c/(1-\phi_1)])^2}{2[\sigma^2/(1-\phi_1^2)]}\right]$$

となる。これらを先ほどの式に代入するとAR(1)モデルの尤度関数が導ける。

1 変量時系列モデルの推定法

■ AR(1)モデルの最尤推定

実際の計算では尤度関数を直接最大化するのは、いろいろと数値計算上の問題があるので、その対数をとった**対数尤度関数**

$$\log L(c, \phi_1, \sigma^2) = \sum_{t=2}^{T} \log f(y_t \mid y_{t-1}) + \log f(y_1)$$

を最大化してパラメーターを推定する。

1 変量時系列モデルの推定法

- AR(1)モデルの条件付最尤推定法

先ほどの対数尤度関数から $\log f(y_1)$ を除いたもの、すなわち**条件付対数尤度関数**

$$\log L^C(c, \phi_1, \sigma^2) = \sum_{t=2}^{T} \log f(y_t \mid y_{t-1})$$

を最大化してパラメーターを求める方法を**条件付最尤推定法**という(y_1を確率変数として扱わずに所与として尤度関数を作ったのでこう呼ばれる)。

通常こちらの方が計算が簡単であり、**推定精度**も 観測数 T が十分大きければほとんど**変わらない**。

1 変量時系列モデルの推定法

- AR(p)モデルの最尤推定

尤度関数、すなわち$\{y_1,\ldots,y_T\}$の結合密度関数 $f(y_1,\ldots,y_T)$ を導く。この際 $f(y_1,\ldots,y_T)$ は条件付密度関数の積 として

$$f(y_T, y_{T-1},\ldots, y_1)$$
$$=\left[\prod_{t=p+1}^{T} f(y_t \mid y_{t-1}, y_{t-2},\ldots, y_{t-p})\right] f(y_p,\ldots, y_1)$$

と書き直す事ができるのでこれを利用する。

1 変量時系列モデルの推定法

- AR(p)モデルの最尤推定

尤度関数の対数、すなわち対数尤度関数

$$\log L(\boldsymbol{\Xi}) = \sum_{t=p+1}^{T} \log f(y_t \mid y_{t-1},\ldots,y_{t-p}) + \log f(y_p,\ldots,y_1)$$

を最大化するようなパラメーターの値を求める。
未知パラメーターは

$$\boldsymbol{\Xi} = (c,\phi_1,\ldots,\phi_p,\sigma^2)'$$

である。

1 変量時系列モデルの推定法

- AR(p)モデルの条件付最尤推定法

先ほどの対数尤度関数から $\log f(y_p,\ldots,y_1)$ を除いたもの、すなわち**条件付対数尤度関数**

$$\log L^C(\Xi) = \sum_{t=p+1}^{T} \log f(y_t \mid y_{t-1}, y_{t-2}, \ldots, y_{t-p})$$

を最大化してパラメーターを求める。ここでは $(y_p,\ldots,y_1)$ を確率変数として扱わずに、所与として尤度関数を作っている。

1 変量時系列モデルの推定法

- (一般の)最尤推定量の性質

(1) 最尤推定量は一般には不偏推定量ではない。
(2) 最尤推定量は一致推定量である。
(3) 最尤推定量は一致推定量の中で漸近的に最小の分散を持つ。

- 最尤推定法の利点

(1) 適用範囲が広い(いろいろなモデルを推定できる)
(2) 漸近的には(ある基準で)最もいい推定方法。

演習問題 (ARモデルの推定)

問題 1

次のAR(1)モデルの ϕ_1 の条件付最尤推定量を求めよ。
OLS推定量とどう違うか?

$$y_t = \phi_1 y_{t-1} + \varepsilon_t,\ \varepsilon_t \sim i.i.d.N(0, 1)$$

問題 2

次のAR(2) モデルの最小二乗推定量 ϕ_2 を求めよ。

$$y_t = y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t,\ \varepsilon_t \sim i.i.d.N(0, 1)$$

1変量時系列モデルの推定法

- MAモデルの最尤推定

基本的な考え方はまったく同じである。すなわち**対数尤度関数**を構成して、それを最大化するようにパラメーターを選ぶというものである。

ただし、MAモデルの場合には y_t が観測不能なノイズに依存しているため、正確な対数尤度関数を構成するのは難しい(不可能ではないが、少し高度)。以下、MA(1)モデルに関して、よく使われる簡便な方法についてみてみよう(ただし最近ではコンピューターの発達に伴って正確な最尤法もよく使われる)。

1 変量時系列モデルの推定法

- MA(1) モデルの(近似的な)最尤推定

以下のMA(1) モデルの(近似的な)最尤推定について考える。

$$y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1},\quad \varepsilon_t \sim \text{i.i.d. } N(0, \sigma^2)$$

(最尤推定法を適用するため、誤差項に正規分布を仮定する)。既に見たように尤度関数を構成するには y_t の条件付き密度関数

$$f(y_t \mid y_{t-1}, y_{t-2}, \ldots, y_1)$$

が分かればよい。

1変量時系列モデルの推定法

- MA(1) モデルの(近似的な)最尤推定

ここで簡便な方法として $\varepsilon_0 = 0$ という条件を仮定する方法がよく用いられる。この時 $y_1 = \mu + \varepsilon_1$ なので

$$y_1 \sim N(\mu, \sigma^2)$$

となり、$f(y_1)$ は

$$f(y_1) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_1 - \mu)^2}{2\sigma^2} \right]$$

となる。

1 変量時系列モデルの推定法

- MA(1) モデルの(近似的な)最尤推定

次に $f(y_2 \mid y_1)$ を構成する。しかし $\varepsilon_1 = y_1 - \mu$より

$$y_2 = \mu + \varepsilon_2 + \theta\varepsilon_1 = \mu + \varepsilon_2 + \theta y_1 - \theta\mu,$$

なので y_1 が分かっていれば $y_2 | y_1 \sim N(\theta y_1 + \mu - \theta\mu, \sigma^2)$ となる。よって

$$f(y_2 \mid y_1) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_2 - \theta y_1 - \mu + \theta\mu)^2}{2\sigma^2}\right]$$

となる。

1 変量時系列モデルの推定法

■ MA(1) モデルの最尤推定

次に $f(y_3 \mid y_2, y_1)$ を構成する。しかしながら

$$y_2 = \mu + \varepsilon_2 + \theta y_1 - \theta\mu \quad \Leftrightarrow \quad \varepsilon_2 = y_2 - \theta y_1 - \mu + \theta\mu$$

より y_1 および y_2 が分かっていれば、

$$y_3 = \mu + \varepsilon_3 + \theta\varepsilon_2 = \mu + \varepsilon_3 + \theta y_2 - \theta^2 y_1 - \theta\mu + \theta^2\mu$$

なので $y_3 | y_2, y_1 \sim N(\theta y_2 - \theta^2 y_1 + \mu - \theta\mu + \theta^2\mu, \sigma^2)$ となる。よって

$$f(y_3 \mid y_2, y_1) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_3 - \theta y_2 + \theta^2 y_1 - \mu + \theta\mu - \theta^2\mu)^2}{2\sigma^2} \right]$$

となる。

1 変量時系列モデルの推定法

- MA(1) モデルの最尤推定

以下同様の手順により

$$f(y_t \mid y_{t-1}, \ldots, y_1), t = 4, \ldots, T$$

を構成する事ができる(AR(1)モデルと異なり y_t の条件付き密度関数は過去の全ての y_t に依存している)。
これより対数尤度関数

$$\log L(\mu, \theta, \sigma^2) = \sum_{t=2}^{T} \log f(y_t \mid y_{t-1}, \ldots, y_1) + \log f(y_1)$$

を構成する事ができ、これをパラメータ μ, θ_1, σ^2 について最大化することにより、推定する。

1 変量時系列モデルの推定法

■ MA(q) モデルの最尤推定

MA(q) モデルの場合には

$$y_1 = \varepsilon_1 + \theta_1\varepsilon_0 + \theta_2\varepsilon_{-1} + \dots + \theta_q\,\varepsilon_{-q+1}$$

において、

$$\varepsilon_0 = \varepsilon_{-1} = \dots = \varepsilon_{-q+1} = 0$$

という条件を仮定して、同様の議論により尤度関数

$$\log L(\mu, \theta_1, \dots, \theta_q, \sigma^2) = \sum_{t=2}^{T} \log f(y_t \mid y_{t-1}, \dots, y_1) + \log f(y_1)$$

を構成する事ができる。

1 変量時系列モデルの推定法

- ARMA(p, q) モデルの最尤推定

ARMA(p, q) モデルの場合もMAモデルの部分が入っているので、MAモデルの場合と同様の理由により正確な尤度関数を構築するのは難しい(が不可能ではない)。

よく使われるのは、やはり $\varepsilon_0 = \varepsilon_{-1} = \ldots = \varepsilon_{-q+1} = 0$ という値を仮定して尤度関数を構成する(近似的な)条件付き最尤推定法である。これは先ほどのARおよびMAモデルに関する議論と同様の議論を適用することによって条件付き尤度関数を構成し、パラメーターを推定する方法である。

ARMAモデルの次数の選択

- ARMAモデルの次数の選択

データが与えられれば、ARMA(p, q)モデルは最尤推定法などで推定可能である事がわかった。しかしながら、真の次数 p と q は観測不可能である。仮に真のモデルがARMA(2, 2) モデルであるのにARMA(1, 1) モデルを仮定し、その推定結果に基づいて分析を行えば、誤った結論が導かれる可能性が高いであろう。

通常、様々な p と q の値についてARMA(p, q) モデルを推定し、その中から何らかの基準でデータに最もよく適合するモデル(つまり p とq) を選択し、そのモデルに基づいて、分析を行う。

ARMAモデルの次数の選択

- 情報量基準(Information Criteria)

近年、ARMAモデルの次数の選択に最もよく使われる基準は総称して**情報量基準** (Information Criteria)と呼ばれるものである。情報量基準の一般的な形は

$$IC = -2\log L(\hat{\boldsymbol{\theta}}) + p(T)k$$

で与えられる。ここで $\log L(\hat{\boldsymbol{\theta}})$ は対数尤度関数をそのパラメーターの最尤推定値 $\hat{\theta}$ で評価したもの(最大対数尤度) k は推定したパラメーターの数、T は標本数、$p(T)$ はT の何らかの関数(詳しくは後述する)である。情報量基準の定義はT で割って定義される事もある。

ARMAモデルの次数の選択

- 情報量基準(Information Criteria)

$p(T)$の形により様々な情報量基準が存在するが、ここでは応用上よく使われる **赤池情報量基準(AIC)**および**ベイズ情報量基準 (BIC)** (これはまたSchwartz 情報量基準(SIC) とも呼ばれる)についてのみ説明する。

ARMAモデルの次数の選択

- 赤池情報量基準 (AIC)

情報量基準で最もよく使われるのは**赤池情報量基準(AIC)**である。これは $p(T) = 2$ とする事によって得られる。

$$AIC = -2\log L(\hat{\boldsymbol{\theta}}) + 2k$$

- ベイズ情報量基準(BIC)

AICの次によく使われる情報量基準として**ベイズ情報量基準(BIC)**がある。これは $p(T) = \log(T)$ とする事によって

$$BIC = -2\log L(\hat{\boldsymbol{\theta}}) + \log(T)k$$

として得られる。

ARMAモデルの次数の選択

- 情報量基準の使い方

情報量基準に基づいてモデルを選択する場合、**小さい方がよいモデル**となる。つまり様々な p, q に対して、情報量基準を計算して、情報量基準が**最も小さくなるモデルを選択する**、という事になる。

ARMAモデルの次数の選択

- 情報量基準の解釈

情報量基準の定義を見てみると、第 1 項はモデルの尤度(に–2をかけたもの)であり、これはそのモデルとデータとのの適合度をみているのに対して、第 2 項(ペナルティーターームと呼ばれる)はモデルのパラメーターの数が多くなると大きくなる、すなわちモデルがより複雑になる事に対するペナルティーとみなす事ができる。

このように解釈してみると、情報量基準は**同じ適合度であれば、より単純なモデルほど望ましい**という事を示唆している。

ARMAモデルの次数の選択

- AICとBICの違い

AIC も BIC も第 1 項は同じである。よって違いは第 2 項のペナルティタームより生じる。

このペナルティータームは $\log(T) > 2$ すなわち $T \geqq 8$ であれば AIC の方が常に小さい。つまり AIC では次数が増えてもペナルティーは BIC 程には増えないという事であり、これは**BICの方がAICより小さい次数を選ぶ**傾向がある事を意味している。

ARMAモデルの次数の選択

- AIC とBIC の一致性

情報量基準における**一致性**とは標本数 T が大きくなる時に**正しい次数を選ぶ**という性質をいう。**BICは一致性を持つが、AICは一致性を持たない**事が知られている。

AICは標本数が大きくなるにつれ、真の次数より小さい次数を選択する確率は 0 に収束するが、大きい次数を選択する確率は 0 にはならない(よって一致性がない)。

ただし、次数を大きめに選択するという事は標本数が少ない時には必ずしも悪い事ではないのでAICもよく使われる。

ARMAモデルの次数の選択

- ARMAモデルの次数の選択についてのまとめ

ARMA (p, q) モデルの次数の選択の仕方をまとめると以下のようになる。

(1) まず p と q について最大いくつにするか決める(通常 3~6 くらい)。

(2) それら以下の p, q の組み合わせについて ARMAモデルを推定し、AIC と BIC を計算する。

(3) AIC および BICが一番小さいモデルを選択する。

(4) **AIC と BIC による選択結果は必ずしも一致しない**。この時は通常、分析者がやや主観的な理由でどちらかのモデルを選択する。