

# YUIMAにおけるモデル選択

江口 翔一

大阪大学 数理・データ科学教育研究センター

2018年12月2日

## 概要

本資料では、確率微分方程式モデルを対象としたモデル評価・選択について解説する。まず、現在モデル評価基準として広く認知されている2つの情報量規準の背景を述べる。その後、確率微分方程式モデルに適用可能な情報量規準について概説し、具体例を通して YUIMA パッケージにおける計算方法を紹介する。

## 1 情報量規準

観測データ  $\mathbf{x}_n = \{x_1, \dots, x_n\}$  が真のモデル  $g_n(x)$  に従って生成されているものとし、この真のモデルを捉えるためのパラメトリックモデルとして  $f_n(x, \theta) (\theta \in \Theta \subset \mathbb{R}^p)$  を想定する。一般に、最尤法によりパラメータ  $\theta$  の最尤推定量  $\hat{\theta}_n$  を求め、真のモデルを推定する。このとき考慮しなければならないのが、このように構築したモデル  $f_n(x, \hat{\theta}_n)$  のよさを評価することである。観測データに基づいて、候補となるモデル  $\mathcal{M}_1, \dots, \mathcal{M}_K$  と各モデル  $\mathcal{M}_i$  のパラメトリックモデル  $f_{i,n}(x, \theta_i) (\theta_i \in \Theta_i \subset \mathbb{R}^{p_i})$  を想定して、最もよいモデルを選択する。情報量規準とは、モデルのよさの評価とよいモデルの選択に利用することができる指標である。本節では、現在広く認知されている2つの情報量規準 AIC と BIC の背景について紹介する。以下、 $\mathbb{H}_{i,n}$  と  $\hat{\theta}_{i,n}$  をそれぞれモデル  $\mathcal{M}_i$  における対数尤度関数とパラメータ  $\theta_i$  の最尤推定量とする。

### 1.1 赤池情報量規準 AIC

赤池情報量規準 (Akaike information criterion; AIC) とは、モデルの評価を予測の観点から捉えることで導出されるモデル評価基準である。モデルの評価を予測の観点から捉えるとは、将来真のモデルから獲得したデータ  $z$  の従うモデル  $g_n(z)$  を、構築したモデル  $f_{i,n}(z, \hat{\theta}_{i,n})$  で予測したときの平均的なよさ (悪さ) を評価するということである。

2つの分布の近さを測る尺度として、Kullback-Leibler 情報量 (K-L 情報量) と呼ばれるものが用いられる。 $g_n(z)$  と  $f_{i,n}(z, \hat{\theta}_{i,n})$  を対象とした場合、その K-L 情報量  $KL(g_n; f_{i,n})$  は

$$\begin{aligned} KL(g_n; f_{i,n}) &= \mathbb{E}_Z \left( \log \frac{g_n(Z)}{f_{i,n}(Z, \hat{\theta}_{i,n})} \right) = \mathbb{E}_Z \{ \log g_n(Z) \} - \mathbb{E}_Z \{ \log f_{i,n}(Z, \hat{\theta}_{i,n}) \} \\ &= \int \{ \log g_n(z) \} g_n(z) dz - \int \{ \log f_{i,n}(z, \hat{\theta}_{i,n}) \} g_n(z) dz \quad (1) \end{aligned}$$

で与えられる。ここで、期待値  $\mathbb{E}_Z$  は未知の真のモデル  $g_n(z)$  に関してとったものである。K-L 情報量が、「 $KL(g_n; f_{i,n}) \geq 0$ 」、 $KL(g_n; f_{i,n}) = 0 \Leftrightarrow g_n = f_{i,n}$  という性質を持つことから、K-L 情報量が小さいほど構築したモデルが真のモデルに近いと考えられる。ここで、(1) の第一項はモデル  $g$  のみに依存する定数となるため、構築したモデルの比較・評価するためには第二項だけを考慮すればよく、第二項の値が大きいモデルほど真のモデルに近いといえる。しかし、(1) の第二項は未知である真の分布  $g$  に関して期待値を

取っており、その値を直接計算することができない。そこで、情報量規準の構成は、この第二項の有効な推定量を求めることが本質的となる。その推定量の一つがモデル  $f_{i,n}$  の最大対数尤度

$$\mathbb{H}_{i,n}(\hat{\theta}_{i,n}) = \sum_{j=1}^n \log f_{i,n}(x_j, \hat{\theta}_{i,n})$$

であり、 $n\mathbb{E}_Z\{\log f_{i,n}(Z, \hat{\theta}_{i,n})\}$  の推定量となっている。

これらのことから、候補モデルの中から最適なモデルを選択するためには、各モデルの最大対数尤度を比較すればよいように思える。しかし、最大対数尤度は  $n\mathbb{E}_Z\{\log f_{i,n}(Z, \hat{\theta}_{i,n})\}$  の推定量としてバイアスを持つため、最大対数尤度を比較するだけでは正確なモデル選択を行うことができない。例えば、 $f_{i,n}$  が真のモデル  $g_n$  を含んでいる（あるパラメータ  $\theta_{i,0}$  が存在して、 $g_n(\cdot) = f_{i,n}(\cdot, \theta_{i,0})$ ）とすると、 $n\mathbb{E}_Z\{\log f_{i,n}(Z, \theta_i)\}$  を最大とするパラメータは真のパラメータ  $\theta_{i,0}$  となる。一方、 $\hat{\theta}_{i,n}$  は対数尤度関数  $\mathbb{H}_{i,n}$  を最大とするパラメータである。したがって、本来の評価基準である  $\mathbb{E}_Z\{\log f_{i,n}(Z, \hat{\theta}_{i,n})\}$  においては、 $\mathbb{E}_Z\{\log f_{i,n}(Z, \hat{\theta}_{i,n})\} \leq \mathbb{E}_Z\{\log f_{i,n}(Z, \theta_{i,0})\}$  であるが、データから計算できる対数尤度においては  $\mathbb{H}_{i,n}(\hat{\theta}_{i,n}) \geq \mathbb{H}_{i,n}(\theta_{i,0})$  となり、その大小関係が逆転する。公平なモデル評価のためには、このバイアスを評価し、補正する必要がある。

最大対数尤度で  $n\mathbb{E}_Z\{\log f_{i,n}(Z, \hat{\theta}_{i,n})\}$  を推定したときのバイアスは

$$\mathbb{E}_{\mathbf{x}_n} \left[ \mathbb{H}_{i,n}(\hat{\theta}_{i,n}) - n\mathbb{E}_Z \left\{ \log f_{i,n}(Z, \hat{\theta}_{i,n}) \right\} \right] = \int \left[ \mathbb{H}_{i,n}(\hat{\theta}_{i,n}) - n \int \left\{ \log f_{i,n}(z, \hat{\theta}_{i,n}) \right\} g_n(z) dz \right] g_n(\mathbf{x}_n) d\mathbf{x}_n$$

で定義される。期待値  $\mathbb{E}_{\mathbf{x}_n}$  は同時分布  $g_n(\mathbf{x}_n)$  についてとったものである。候補モデルに真のモデルが含まれているとき、このバイアスを評価し、最大対数尤度のバイアスを補正することによって、

$$\text{AIC}_n^{(i)} = -2\mathbb{H}_{i,n}(\hat{\theta}_{i,n}) + 2p_i, \quad i = 1, \dots, K$$

が導出される ([1])。各候補モデルにおいて AIC を計算し、最小の値を取るモデルを最適なモデルとして選択する。

## 1.2 Bayes 情報量規準 BIC

Bayes 情報量規準 (Bayesian information criterion; BIC) とは、モデルの事後確率に基づくモデル評価基準であり、次のような考えに従って導出される。

パラメータ  $\theta_i$  の事前分布を  $\pi_{i,n}(\theta_i)$  とする。データ  $\mathbf{x}_n$  が観測されたとき、データに関するモデル  $\mathcal{M}_i$  の周辺尤度は

$$\begin{aligned} h_{i,n}(\mathbf{x}_n) &= \int_{\Theta_i} f_{i,n}(\mathbf{x}_n | \theta_i) \pi_{i,n}(\theta_i) d\theta_i \\ &= \int_{\Theta_i} \exp \{ \log f_{i,n}(\mathbf{x}_n | \theta_i) \} \pi_{i,n}(\theta_i) d\theta_i \\ &= \int_{\Theta_i} \exp \{ \mathbb{H}_{i,n}(\theta_i) \} \pi_{i,n}(\theta_i) d\theta_i \end{aligned}$$

で定義される。また、モデル  $\mathcal{M}_i$  の生起確率を  $\mathbf{p}_i$  とすると ( $\sum_{i=1}^K \mathbf{p}_i = 1$ )、モデル  $\mathcal{M}_i$  の事後確率は

$$\mathbb{P}(\mathcal{M}_i | \mathbf{x}_n) = \frac{h_{i,n}(\mathbf{x}_n) \mathbf{p}_i}{\sum_{i=1}^K h_{i,n}(\mathbf{x}_n) \mathbf{p}_i}, \quad i = 1, \dots, K$$

となる。この事後確率は、データ  $\mathbf{x}_n$  が得られたとき、それがモデル  $\mathcal{M}_i$  から生起するものである確率を表しており、 $K$  個の候補モデルの中から 1 つのモデルを選択するには、事後確率を最大とするモデルを選べばよ

いと考えられる。これは、事後確率の定義より、全ての候補モデルにおいて共通となるため、分子  $h_{i,n}(\mathbf{x}_n) \mathbf{p}_i$  を最大とするモデルの選択を意味する。さらに、 $\mathbf{p}_1 = \mathbf{p}_2 = \dots = \mathbf{p}_K$  とすると、データに関するモデルの周辺尤度  $h_{i,n}$  を最大とするモデルの選択を行うことになる。そこで、周辺尤度の自然対数をとった対数周辺尤度を使いやすい形で表現することができれば記述的モデル評価基準として用いることができる。このとき、最尤推定量の漸近正規性や Laplace 近似に基づき、モデル  $\mathcal{M}_i$  に対する古典的な BIC

$$\text{BIC}_n^{(i)} = -2\mathbb{H}_{i,n}(\hat{\theta}_{i,n}) + p_i \log n, \quad i = 1, \dots, K$$

が導出される ([3])。各候補モデルにおいて BIC を計算し、最小の値を取るモデルを最適なモデルとして選択する。

また、周辺尤度  $h_{i,n}$  に関して K-L 情報量を取ると、

$$\begin{aligned} KL(g_n; h_{i,n}) &= \mathbb{E}_X \left( \log \frac{g_n(X)}{h_{i,n}(X)} \right) \\ &= \mathbb{E}_X \{ \log g_n(X) - \log h_{i,n}(X) \} \end{aligned}$$

となることより、 $\log g_n(X) - \log h_{i,n}(X)$  は  $KL(g_n; h_{i,n})$  の不偏推定量となっている。ここで、 $\mathbb{E}_X$  は真のモデル  $g_n(x)$  に関する期待値である。真のモデル  $g_n$  のみに依存する部分を除いて考慮すると、対数周辺尤度  $\log h_{i,n}$  を最大とするモデルの選択は、真のモデル  $g_n$  と周辺尤度  $h_{i,n}$  の K-L 情報量を最小とするモデルの選択と言い換えることができる。

## 2 確率微分方程式モデルにおける情報量規準

本節では、確率微分方程式モデルを候補モデルとした際に、その数学的正当性が保証されている情報量規準を紹介する。特に、3 節で扱うソフトウェア R の `yuima` パッケージ上で実装されている関数 `IC` の計算対象として含まれている規準についてのみ記述している。

### 2.1 エルゴード的拡散過程

離散観測データ  $\mathbf{X}_n = (X_{t_j})_{j=0}^n$  が次の  $d$  次元確率微分方程式の解として与えられるとする：

$$dX_t = A(X_t)dt + B(X_t)dw_t, \quad t \in [0, T_n], \quad X_0 = x_0.$$

ドリフト項  $A: \mathbb{R}^d \rightarrow \mathbb{R}^d$ 、拡散項  $B: \mathbb{R}^d \rightarrow \mathbb{R}^d \otimes \mathbb{R}^d$ 、 $w$  は  $d$  次元標準 Wiener 過程である。また、 $t_j = jh_n$  であり、 $T_n = nh_n \rightarrow \infty$ 、 $nh_n^2 \rightarrow 0$  を満たす。このとき、候補モデル  $\mathcal{M}_{k,\ell}$  ( $k \leq K, \ell \leq L$ ) としてエルゴード的拡散過程モデル

$$dX_t = a_k(X_t, \alpha_k)dt + b_\ell(X_t, \beta_\ell)dw_t, \quad t \in [0, T_n], \quad X_0 = x_0 \quad (2)$$

を扱う。つまり、ドリフト項と拡散項の候補として  $a_1(x, \alpha_1), \dots, a_K(x, \alpha_K)$  及び  $b_1(x, \beta_1), \dots, b_L(x, \beta_L)$  が存在し、候補モデル  $\mathcal{M}_{k,\ell}$  がこれらの候補項の組み合わせで与えられる状況を考えており、候補モデルの総数は  $K \times L$  個となる。ここで、任意の  $k \leq K$  と  $\ell \leq L$  に対して、 $a_k: \mathbb{R}^d \times \Theta_{\alpha_k} \rightarrow \mathbb{R}^d$ 、 $b_\ell: \mathbb{R}^d \times \Theta_{\beta_\ell} \rightarrow \mathbb{R}^d \otimes \mathbb{R}^d$ 、 $\theta_{k,\ell} = (\alpha_k, \beta_\ell) \in \Theta_{\alpha_k} \times \Theta_{\beta_\ell} \subset \mathbb{R}^{p_{\alpha_k}} \times \mathbb{R}^{p_{\beta_\ell}}$  である。以下では、簡略化のため、候補モデルからモデルインデックス  $k, \ell$  を省略して記述する。

$S(x, \beta) = b(x, \beta)b^\top(x, \beta)$ 、 $f_{j-1}(\cdot) = f(X_{t_{j-1}}, \cdot)$ 、 $\Delta_j X = X_{t_j} - X_{t_{j-1}}$  とする。(2) のようなモデルを扱う場合、局所正規近似に基づき、疑似尤度関数は

$$\prod_{j=1}^n (2\pi h_n)^{-\frac{d}{2}} |S_{j-1}(\beta)|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2h_n} S_{j-1}^{-1}(\beta) \left[ (\Delta_j X - h_n a_{j-1}(\alpha)) \otimes 2 \right] \right\}$$

となる。疑似尤度関数の対数を取り、定数項となる項を除くことで、対数疑似尤度関数  $\mathbb{H}_n$  が以下の形で定義される:

$$\mathbb{H}_n(\theta) = -\frac{1}{2} \sum_{j=1}^n \left\{ \log |S_{j-1}(\beta)| + \frac{1}{h_n} S_{j-1}^{-1}(\beta) \left[ (\Delta_j X - h_n a_{j-1}(\alpha))^{\otimes 2} \right] \right\}.$$

また、疑似最尤推定量  $\hat{\theta}_n$  はこの対数疑似尤度関数  $\mathbb{H}_n$  を最大とするパラメータとして与えられる。つまり、

$$\{\hat{\theta}_n\} = \{(\hat{\alpha}_n, \hat{\beta}_n)\} = \underset{\theta}{\operatorname{argmax}} \mathbb{H}_n(\theta)$$

となる。エルゴード的拡散過程モデルを候補モデルとした際のモデル選択のため、1.1 節のような背景に基づき、[4] により AIC 型情報量規準となる Contrast-based information criterion

$$\text{CIC}_n = -2\mathbb{H}_n(\hat{\theta}_n) + 2(p_\alpha + p_\beta)$$

が導出された。加えて、1.2 節のような背景に基づき、[2] により BIC 型情報量規準となる BIC と quasi-BIC(QBIC)

$$\begin{aligned} \text{BIC}_n &= -2\mathbb{H}_n(\hat{\theta}_n) + p_\alpha \log T_n + p_\beta \log n, \\ \text{QBIC}_n &= -2\mathbb{H}_n(\hat{\theta}_n) + \log \left| -\partial_\alpha^2 \mathbb{H}_n(\hat{\theta}_n) \right| + \log \left| -\partial_\beta^2 \mathbb{H}_n(\hat{\theta}_n) \right| \end{aligned}$$

の2つが導かれた。 $\partial_\alpha$  は  $\alpha$  に関する偏微分  $\partial/\partial\alpha$  を意味する。この BIC と QBIC は、漸近的に同等な情報量規準となっている。各候補モデルごとにこれらの情報量規準を計算し、同じ規準同士でその値を比較し最小の値を取るものを最適なモデルとして選択する。

## 2.2 連続セミマルチンゲールにおけるボラティリティ推定

本節では、候補モデル  $\mathcal{M}_\ell (\ell \leq L)$  が

$$dY_t = a_t dt + b_\ell(X_t, \beta_\ell) dw_t, \quad t \in [0, T], \quad X_0 = 0$$

で与えられる場合を扱う。ここで、任意の  $\ell \leq L$  に対して、 $b_\ell: \mathbb{R}^d \times \Theta_{\beta_\ell} \rightarrow \mathbb{R}^m \otimes \mathbb{R}^r$ ,  $\beta_\ell \in \Theta_{\beta_\ell} \subset \mathbb{R}^{p_{\beta_\ell}}$ ,  $a$  と  $X$  はそれぞれ  $m$  次と  $d$  次の発展的可測過程、 $w$  は  $r$  次元標準 Wiener 過程、 $T$  はある正の値である。データセットは離散観測データ  $\mathbf{X}_n = (X_{t_j}, Y_{t_j})_{j=0}^n$  で与えられ、 $t_j = jh_n = jT/n$  を満たす。以下では、モデルインデックス  $\ell$  を省略して記述する。

$S(x, \beta) = b(x, \beta)b^\top(x, \beta)$ ,  $f_{j-1}(\cdot) = f(X_{t_{j-1}}, \cdot)$ ,  $\Delta_j Y = Y_{t_j} - Y_{t_{j-1}}$  とすると、対数疑似尤度関数  $\mathbb{H}_n$  は

$$\mathbb{H}_n(\beta) = -\frac{1}{2} \sum_{j=1}^n \left\{ \log |S_{j-1}(\beta)| + \frac{1}{h_n} S_{j-1}^{-1}(\beta) \left[ (\Delta_j Y)^{\otimes 2} \right] \right\}$$

で定義される。このとき、エルゴード的拡散過程の場合と同様にして、[2] により BIC 型情報量規準となる BIC と QBIC が以下の形で導出された:

$$\begin{aligned} \text{BIC}_n &= -2\mathbb{H}_n(\hat{\beta}_n) + p \log n, \\ \text{QBIC}_n &= -2\mathbb{H}_n(\hat{\beta}_n) + \log \left| -\partial_\beta^2 \mathbb{H}_n(\hat{\beta}_n) \right|. \end{aligned}$$

例えば、 $d = p$ ,  $m = r = 1$  とし、 $b(x, \beta) = \exp(x^\top \beta/2)$  の場合、QBIC は

$$\begin{aligned} \text{QBIC}_n &= \sum_{j=1}^n \left\{ X_{t_{j-1}}^\top \hat{\beta}_n + \frac{1}{h_n} (\Delta_j Y)^2 \exp \left( -X_{t_{j-1}}^\top \hat{\beta}_n \right) \right\} \\ &\quad + \log \left| \frac{1}{2h_n} \sum_{j=1}^n (\Delta_j Y)^2 \exp \left( -X_{t_{j-1}}^\top \hat{\beta}_n \right) X_{t_{j-1}} X_{t_{j-1}}^\top \right| \end{aligned}$$

で表される。

### 3 関数 IC の仕様

yuima パッケージに実装されている関数 IC の仕様について説明する. この関数は上記の確率微分方程式モデルにおける情報量規準に基づき作成されており, 候補となるモデルごとに (Q)BIC や CIC などの計算が可能である. 関数 IC は

```
IC(yuima, data, start, lower, upper, joint = FALSE, rcpp = FALSE, ...)
```

という形で定義されている. 関数に含まれる変数はそれぞれ以下の要素を表す:

- `yuima`: yuima object. 候補となるモデルの設定が含まれる.
- `data`: モデル選択に使用するデータ. データ構造はベクトルまたは行列
- `start`: パラメータの推定 (最適化) に用いる初期値. データ構造は名前付きのリスト.
- `lower`: パラメータの推定 (最適化) に用いる下限. データ構造は名前付きのリスト.
- `upper`: パラメータの推定 (最適化) に用いる上限. データ構造は名前付きのリスト.
- `joint`: パラメータに関して同時推定と二段階推定のどちらを行うかの選択. TRUE または FALSE で入力. デフォルトは FALSE.
- `rcpp`: C++ コードを使用するか否かの選択. TRUE または FALSE で入力. デフォルトは FALSE.

また, 得られる戻り値は次の 4 つである:

- `$par`: 候補モデルに含まれるパラメータの推定値.
- `$BIC`: BIC の値.
- `$QBIC`: QBIC の値.
- `$CIC`: CIC の値.

関数 IC は `method="L-BFGS-B"` とした関数 `qmlc` を内部で使用している. 以下では, 適当なデータ ( $X_t$ ) に対して, 候補モデルに

$$dX_t = \alpha_1 X_t dt + \sqrt{\beta_1} dw_t, \quad t \in [0, T_n]$$

を扱う場合の変数の準備及び関数 IC の使用に関する例とその計算結果を記載する. ただし, データ数  $n = 1000$ ,  $T_n = n^{1/3}$  とする.

```
### サンプルデータ生成
set.seed(123)
model <- setModel(drift = "alpha1*x", diffusion = "sqrt(beta1)")
sample <- setSampling(Terminal = 1000^(1/3), n = 1000)
yuima <- setYuima(model = model, sampling = sample)
simu <- simulate(yuima, xinit=1, true.parameter = list(alpha1 = -1, beta1 = 2))
Xt <- simu@data@original.data

### yuima object の生成
model1 <- setModel(drift = "alpha1*x", diffusion = "sqrt(beta1)")
sample1 <- setSampling(Terminal = 1000^(1/3), n = 1000)
yuima1 <- setYuima(model = model1, sampling = sample1) # yuima object
```

```

### 変数 start, lower, upper
init <- list(alpha1 = -1, beta1 = 0.5)      # 変数 start に対応
low  <- list(alpha1 = -3, beta1 = 0.001)  # 変数 lower に対応
upp  <- list(alpha1 = -0.001, beta1 = 3)   # 変数 upper に対応

### 関数 IC の使用
IC(yuima1, data = Xt, start = init, lower = low, upper = upp, rcpp = TRUE)

## $par
##      beta1      alpha1
##  1.965465 -1.583359
##
## $BIC
## [1] -1083.488
##
## $QBIC
## [1] -1086.631
##
## $CIC
## [1] -1088.698

```

### 注意 3.1.

- (1) 変数 `start`, `lower`, `upper` は計算対象のモデル (yuima object) に含まれる全てのパラメータについて設定されている必要がある。また、これらの変数で使われるパラメータの名前は対象のモデルに含まれるパラメータと同名でなければならない。上で例に挙げたモデルの場合、

```

init <- list(alpha1 = -1)
init <- list(gamma1 = -1, beta1 = 0.5)

```

などとするとエラーになる。

- (2) 変数 `start`, `lower`, `upper` が対象のモデルに含まれるパラメータに加え、余分な要素を含んでいる場合、関数 `IC` は必要な要素だけを抽出して計算を行う。つまり、上で例に挙げたモデルの場合、

```

init <- list(alpha1 = -1, alpha2 = -1, beta1 = 0.5, beta2 = 0.5)

```

としても `IC` は使用可能である。

- (3) 疑似尤度関数の  $-\partial_{\theta}^2 \mathbb{H}_n(\hat{\theta}_n)$  の行列式が 0 以下となるとき、関数 `IC` で計算される `QBIC` は `BIC` と同じ値を取る。

## 4 シミュレーション

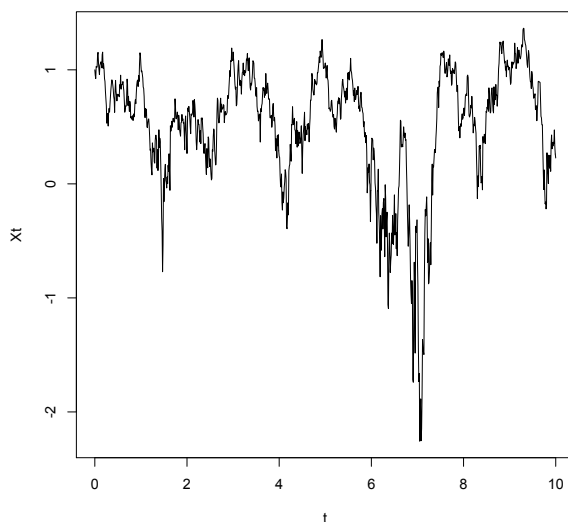
### 4.1 一次元データ

データ  $(X_{t_j})_{j=0}^n$  が従う真のモデルを

$$dX_t = -X_t dt + \exp\left\{\frac{1}{2}(-2 \sin X_t + 1)\right\} dw_t, \quad X_0 = 1, \quad t \in [0, T_n]$$

とする。ここで、データ数  $n = 1000$ ,  $t_j = jh_n = jn^{-2/3}$ ,  $T_n = n^{1/3}$  とする。データは以下のようにして生成される。

```
set.seed(123)
model <- setModel(drift = "alpha1*x", diffusion = "exp((beta1*sin(x)+beta2)/2)")
sample <- setSampling(Terminal = 1000^(1/3), n = 1000)
yuima <- setYuima(model = model, sampling = sample)
simu <- simulate(yuima, xinit = 1,
                 true.parameter = list(alpha1 = -1, beta1 = -2, beta2 = 1))
Xt <- simu@data@original.data
```



候補モデルとして

$$\text{Model 1 : } dX_t = \alpha_1 X_t dt + \exp\left\{\frac{1}{2}(\beta_1 \sin X_t + \beta_2)\right\} dw_t$$

$$\text{Model 2 : } dX_t = \alpha_1 X_t dt + \exp\left(\frac{1}{2}\beta_1 \sin X_t\right) dw_t$$

の2つのモデルを扱う。このとき、Model 1 が真のモデルとなる。各モデルの yuima object 生成、変数の設定及び関数 IC の使用を以下のように行う。パラメータの設定は、2つの候補モデルに関してまとめて行っているが、IC の使用に支障はない (注意 3.1 (2))。

```
### Model 1
model1 <- setModel(drift = "alpha1*x", diffusion = "exp((beta1*sin(x)+beta2)/2)")
sample1 <- setSampling(Terminal = 1000^(1/3), n = 1000)
yuima1 <- setYuima(model = model1, sampling = sample1)

### Model 2
model2 <- setModel(drift = "alpha1*x", diffusion = "exp(beta1*sin(x)/2)")
sample2 <- setSampling(Terminal = 1000^(1/3), n = 1000)
yuima2 <- setYuima(model = model2, sampling = sample2)

### パラメータ設定
init <- list(alpha1 = -1, beta1 = 0.5, beta2 = 0.5)
```

```

low <- list(alpha1 = -5, beta1 = -5, beta2 = -5)
upp <- list(alpha1 = -0.001, beta1 = 5, beta2 = 5)

### IC の使用
## Model 1
ic1 <- IC(yuima1, data = Xt, start = init, lower = low, upper = upp, rcpp = TRUE)
## Model 2
ic2 <- IC(yuima2, data = Xt, start = init, lower = low, upper = upp, rcpp = TRUE)

```

計算結果 ic1 と ic2 は,

```
ic1
```

```

## $par
##      beta1      beta2      alpha1
## -2.087821  1.022040 -1.085622
##
## $BIC
## [1] -1684.139
##
## $QBIC
## [1] -1687.36
##
## $CIC
## [1] -1694.257

```

```
ic2
```

```

## $par
##      beta1      alpha1
## -1.134591 -1.158270
##
## $BIC
## [1] -1292.771
##
## $QBIC
## [1] -1293.671
##
## $CIC
## [1] -1297.981

```

となる。BIC, QBIC, CIC の値をそれぞれ比較すると、このシミュレーションにおいては、全ての情報量規準において真のモデルと形式が一致する Model 1 が選択される。



## 4.2 多次元データ

ここでは、2次元データの場合を対象とする。データ  $(X_{1,t_j}, X_{2,t_j})_{j=0}^n$  が従う真のモデルを

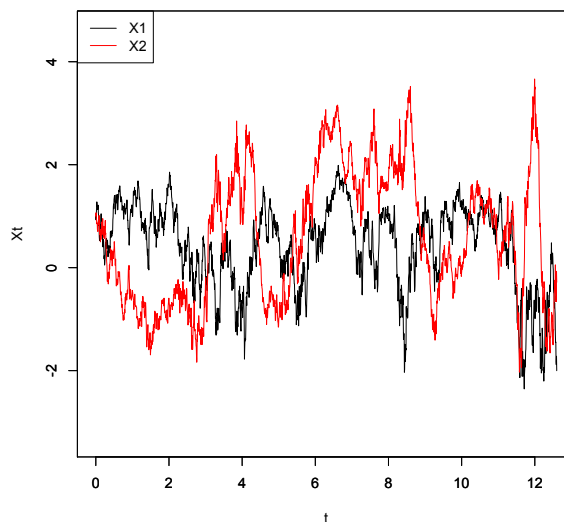
$$d \begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} -2X_{1,t} \\ -X_{2,t} \end{pmatrix} dt + \begin{pmatrix} -X_{1,t} + 2 & -1 \\ 1 & -X_{1,t} + 2 \end{pmatrix} d \begin{pmatrix} w_{1,t} \\ w_{2,t} \end{pmatrix}, \quad X_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad t \in [0, T_n]$$

とし、データ数  $n = 2000$ ,  $t_j = jh_n = jn^{-2/3}$ ,  $T_n = n^{1/3}$  の場合を扱う。このとき、データは以下で生成される。

```
set.seed(200)
### 拡散項の設定: 行列
diff.coef.matrix <- matrix(c("beta1*x1+beta3", "1", "-1", "beta1*x1+beta3"), 2, 2)

### ドリフト項の設定: ベクトル
drif.coef.vec <- c("alpha1*x1", "alpha2*x2")

### yuima object, データの生成
model <- setModel(drift = drif.coef.vec, diffusion = diff.coef.matrix,
                  state.variable = c("x1", "x2"), solve.variable = c("x1", "x2"))
sample <- setSampling(Terminal = 2000^(1/3), n = 2000)
yuima <- setYuima(model = model, sampling = sample)
simu <- simulate(yuima, xinit = c(1,1),
                 true.parameter = list(alpha1 = -2, alpha2 = -1, beta1 = -1, beta3 = 2))
Xt <- simu@data@original.data
```



候補モデルは

$$d \begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} \alpha_1 X_{1,t} \\ \alpha_2 X_{2,t} \end{pmatrix} dt + \begin{pmatrix} b(X_t, \beta) & -1 \\ 1 & b(X_t, \beta) \end{pmatrix} d \begin{pmatrix} w_{1,t} \\ w_{2,t} \end{pmatrix}$$

という形で構成し、拡散項の関数  $b(x, \beta)$  の部分に次の関数形を考える:

**Model 1** :  $b(X_t, \beta) = \beta_1 X_{t,1} + \beta_2 X_{t,2} + \beta_3$ ; **Model 2** :  $b(X_t, \beta) = \beta_1 X_{t,1} + \beta_2 X_{t,2}$ ;

**Model 3** :  $b(X_t, \beta) = \beta_1 X_{t,1} + \beta_3$ ; **Model 4** :  $b(X_t, \beta) = \beta_2 X_{t,2} + \beta_3$ ;

**Model 5** :  $b(X_t, \beta) = \beta_1 X_{t,1}$ ; **Model 6** :  $b(X_t, \beta) = \beta_2 X_{t,2}$ ; **Model 7** :  $b(X_t, \beta) = \beta_3$ .

真のモデルは、Model 3 の場合である。この例のように候補となる項が多数存在する場合、以下のようにして yuima object を生成することも可能である。

```
### 拡散項の候補を全てまとめたリスト (各行 Model 1 - Model 7 を表す)
cand.diff <- list(
  matrix(c("beta1*x1+beta2*x2+beta3", "1", "-1", "beta1*x1+beta2*x2+beta3"), 2, 2),
  matrix(c("beta1*x1+beta2*x2", "1", "-1", "beta1*x1+beta2*x2"), 2, 2),
  matrix(c("beta1*x1+beta3", "1", "-1", "beta1*x1+beta3"), 2, 2),
  matrix(c("beta2*x2+beta3", "1", "-1", "beta2*x2+beta3"), 2, 2),
  matrix(c("beta1*x1", "1", "-1", "beta1*x1"), 2, 2),
  matrix(c("beta2*x2", "1", "-1", "beta2*x2"), 2, 2),
  matrix(c("beta3", "1", "-1", "beta3"), 2, 2)
)

### ドリフト項の設定
cand.drif <- c("alpha1*x1", "alpha2*x2")

### 全候補モデルの yuima object をリスト形式で生成
cand.model <- NULL
for(i in 1:length(cand.diff)){
  cand <- setModel(drift = cand.drif, diffusion = cand.diff[[i]],
    state.variable = c("x1", "x2"), solve.variable = c("x1", "x2"))
  cand.samp <- setSampling(Terminal = 2000^(1/3), n = 2000)
  cand.yuima <- setYuima(model = cand, sampling = cand.samp)
  cand.model <- c(cand.model, list(cand.yuima))
}
```

上記の要領で生成した場合、`cand.model[[i]]` が候補としている Model  $i$  の yuima object に対応する。パラメータの設定と各候補モデルにおける関数 IC の使用は以下で行うことができる。

```
### パラメータ設定
init <- list(alpha1 = -1, alpha2 = -1, beta1 = 0.5, beta2 = 0.5, beta3 = 0.5)
low <- list(alpha1 = -3, alpha2 = -3, beta1 = -5, beta2 = -5, beta3 = -5)
upp <- list(alpha1 = -0.001, alpha2 = -0.001, beta1 = 5, beta2 = 5, beta3 = 5)

### 例: Model 1 に対する IC の使用
ic <- IC(yuima = cand.model[[1]], data = Xt, start = init, lower = low, upper = upp)

### 全候補モデルに対して IC を使用し結果を格納
paranum <- list(1:5, c(1,2,4,5), c(1,3,4,5), c(2,3,4,5), c(1,4,5), c(2,4,5), c(3,4,5))
results <- matrix(0, 7, 8)
colnames(results) <- c("beta1", "beta2", "beta3", "beta1", "beta2",
  "BIC", "QBIC", "CIC")
```

```

rownames(results) <- c("Model 1", "Model 2", "Model 3", "Model 4",
                      "Model 5", "Model 6", "Model 7")
for(i in 1:7){
  ic <- IC(yuima = cand.model[[i]], data = Xt, start = init, lower = low, upper = upp)
  results[i,paranum[[i]]] <- ic$par
  results[i,6:8] <- c(ic$BIC, ic$QBIC, ic$CIC)
}

```

results に全ての計算結果が格納されており、

```

results

##          beta1      beta2      beta3      beta1      beta2      BIC
## Model 1 -1.0583542  0.001811267  2.018310 -1.770173 -0.8747962 -4182.2607
## Model 2 -0.8069696  1.362905768  0.000000 -1.974062 -0.5502677 -2153.1532
## Model 3 -1.0586496  0.000000000  2.019548 -1.769961 -0.8739008 -4189.8550
## Model 4  0.0000000  0.059506439  1.732918 -2.336065 -1.0604995 -3230.4508
## Model 5 -2.1998290  0.000000000  0.000000 -1.851133 -1.4337826 -892.1917
## Model 6  0.0000000 -1.458342919  0.000000 -1.555513 -0.3980473 -2134.7236
## Model 7  0.0000000  0.000000000 -1.771589 -2.362494 -1.0457829 -3228.2303
##          QBIC      CIC
## Model 1 -4184.882 -4200.1307
## Model 2 -2158.122 -2165.4223
## Model 3 -4192.581 -4202.1241
## Model 4 -3232.794 -3242.7199
## Model 5 -896.013 -898.8598
## Model 6 -2137.536 -2141.3918
## Model 7 -3230.866 -3234.8985

```

となる。1-5列が各モデルにおけるパラメータの推定量、6-8列が各モデルにおける BIC, QBIC, CIC の値を示している。候補モデルが推定の対象となるパラメータを含んでいない場合 (Model 2 における  $\beta_3$  など), その要素は 0.000... となっている。各候補モデルの BIC, QBIC, CIC の値をそれぞれ比較すると、全ての情報量規準において Model 3 が選択される。

## 参考文献

- [1] H. Akaike. A new look at the statistical model identification. *IEEE Trans. Autom. Control.*, 19(6):716–723, 1974.
- [2] S. Eguchi and H. Masuda. Schwarz type model comparison for LAQ models. *Bernoulli*, 24(3):2278–2327, 2018.
- [3] G. Schwarz. Estimating the dimension of a model. *Ann. Statist.*, 6(2):461–464, 1978.
- [4] M. Uchida. Contrast-based information criterion for ergodic diffusion processes from discrete observations. *Ann. Inst. Statist. Math.*, 62(1):161–187, 2010.