

時系列解析 (1)

2019/4/10

東京大学

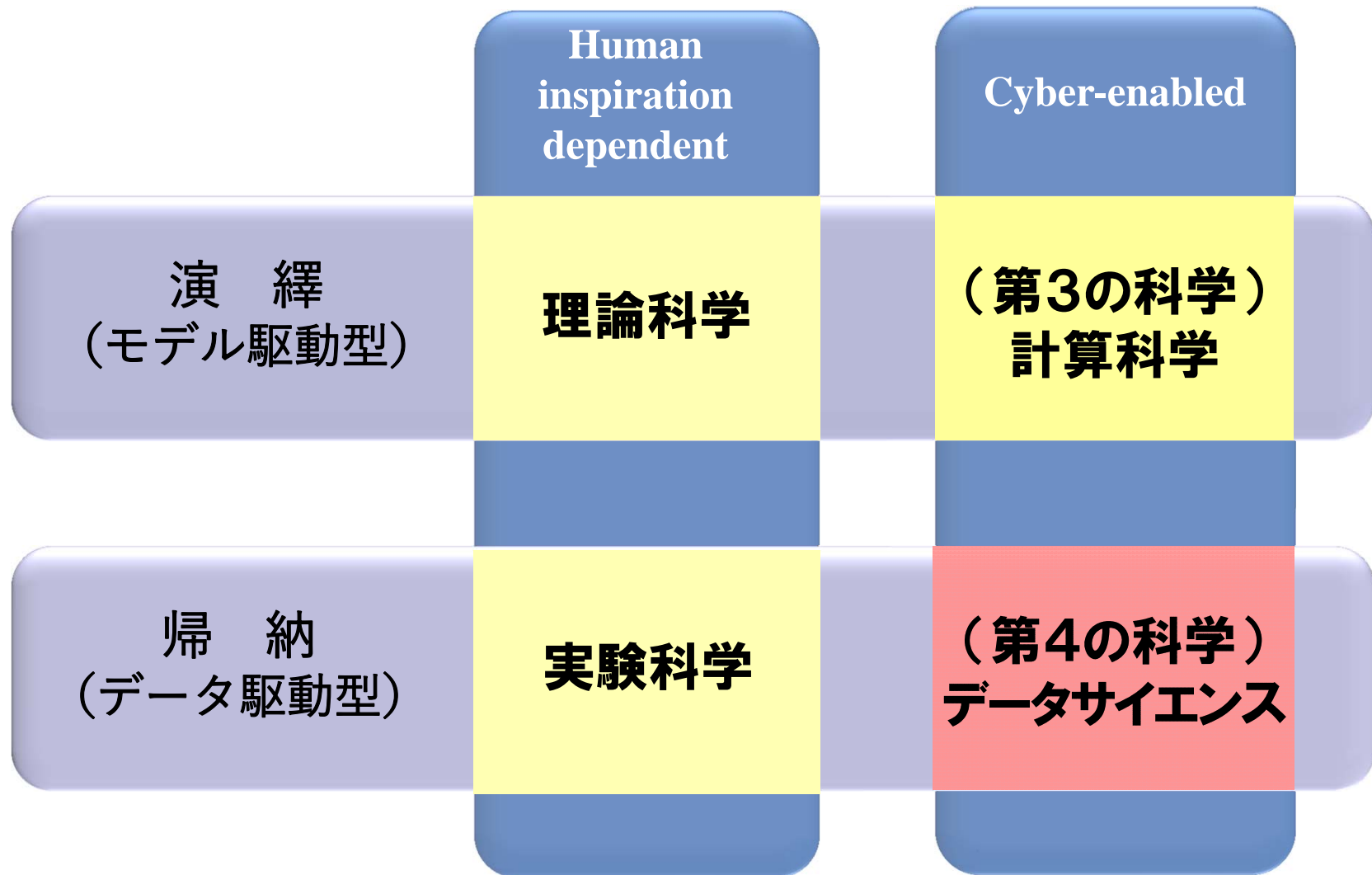
北川源四郎

講義の概要

学部横断型教育プログラム「数理・データサイエンス教育プログラム」

開講科目名： 数理手法VII（時系列解析）
時限： 水曜 5 限
単位数： 2
学年： B3/B4
担当教員： 北川源四郎
教室： 工 6 号館工 6 1 号講義室

第4の科学： データサイエンス



学部横断型プログラム「数理・データサイエンス教育プログラム」

数理・情報教育研究センター教員担当分

開講部局	授業科目	単位数	開講学期	開講部局	授業科目	単位数	開講学期
工学部	数理手法IV (確率論)	2	S1S2	理学部	コンピュータシステム概論	1	S1
工学部	数理手法VI (確率過程論)	2	A1A2	理学部	計算機実験I	1	S1S2
工学部	数理手法V (工学のための現代数学入門)	2	A1A2	理学部	計算機実験II	1	A1A2
工学部	数理手法III (最適化手法)	2	A1A2	工学部	数理手法VIII	2	A1A2
工学部	数理手法VII (時系列解析)	2	S1S2	教養学部	特殊講義「社会科学のための統計分析」	2	S1S2
理学部	数理科学統論I (統計データ解析I)	2	S1S2	経済学部	文科系のための解析・線形代数	4	S1S2
理学部	数理科学統論J (統計データ解析II)	2	A1A2	理学部	教養としてのアルゴリズムとデータ構造	2	S2
理学部	確率統計学基礎	2	S1S2	理学部	メディアプログラミング入門	1	S2A2
理学部	Pythonプログラミング入門	1	S1A1	工学部	データサイエンス超入門	1	S1A1
理学部	データマイニング入門	2	A1A2				

その他全学では 157科目が提供されている

http://www.mi.u-tokyo.ac.jp/mds-oudan/Data_Science_2019.pdf

講義の目的

時間とともに変動する現象を記録したデータが**時系列**である。時系列に基づき、複雑な現象を理解し、予測、制御や意思決定を行うための方法が**時系列解析**である。

この講義では、**時系列のモデリング**のための前処理や特徴の可視化、統計的モデリングの方法、線形・定常時系列モデル、状態空間モデルおよび非線形・非ガウス型モデルについて、実際の問題への応用含めつつモデリングの方法を中心に解説し、**現実の問題に対応して適切なモデリングができるようになることを目標とする。**

講義予定

1	4/10	時系列の前処理と可視化	イントロダクション, 時系列の前処理
2	4/17		共分散関数, スペクトルとピリオドグラム
3	4/24	モデリング	統計的モデリング・情報量規準AIC
4	5/8		最尤法, 最小二乗法, ベイズモデル
5	5/15	定常時系列モデル	ARMAモデルによる時系列の解析
6	5/22		ARモデルの推定・応用
7	5/29		局所定常ARモデル, 統計的制御
8	6/5	状態空間モデル	状態空間モデルによる時系列の解析
9	6/12		ARMAモデルの推定, トレンドの推定
10	6/19		季節調整モデル
11	6/26		ボラティリティ, 時変係数ARモデル
12	7/3	非線形・非ガウス型モデル	一般状態空間モデル, ガウス和フィルタ
13	7/10		粒子フィルタ, シミュレーション

講義の進め方, 単位認定方法

講義形態： PowerPoint中心

単位認定： レポートによる

- ✓ 3回程度
- ✓ 自分が興味を持つ対象に関する時系列データについて, 講義に関連した方法によってモデリングや時系列解析を行う

計算演習

- ✓ 授業内で演習は行わないが, 並行して自分が興味を持つ課題に関して実際に計算してみることを推奨する.
- ✓ 計算言語・ソフトは, R, Python, Matlabなど何でも良いが授業で示す例はなるべくRで再現できるようにする.

教科書・参考書

教科書

北川源四郎: 「時系列解析入門」 岩波書店(2005)

関連文献

1. 赤池弘次, 中川東一郎: 「ダイナミックシステムの統計的解析と制御」〔新訂版〕サイエンス社 (2000)
2. 北川源四郎: 「時系列解析プログラミング」 岩波書店 (1993)
3. 赤池弘次, 北川源四郎編: 「時系列解析の実際」 1, 2 朝倉書店 (1994, 1995)
4. 尾崎統, 北川源四郎編: 「時系列解析の方法」 朝倉書店 (1998)
5. G. Kitagawa and W. Gersch, *Smoothness Priors Analysis of Time Series*, Springer (1996)
6. G. Kitagawa, *Introduction to Time Series Modeling*, CRC-Chapman and Hall (2010)
7. 岩波データサイエンス刊行委員会編, 「時系列解析: 状態空間モデル・因果解析・ビジネス応用」 岩波データサイエンス 6, 岩波書店(2017)

注: 2.は教科書の旧版で、FORTRANソースコードを含む。6. は教科書の英語版

R パッケージについて

Rのインストール (Linux, Max OS X, Windows用)

<https://cran.r-project.org>

<http://cran.ism.ac.jp> (統計数理研究所のミラーサイト)

インストール方法はネットで検索

本講義関連時系列解析用のRパッケージ

- TSSS (Time Series analysis with State Space model)

<http://jasp.ism.ac.jp/ism/TSSS/>

参考文献2 (時系列解析プログラミング) に掲載されたプログラムを基に統計数理研究所が開発したR関数群。教科書 (時系列解析入門) のほとんどの例はTSSSで再現できる。開発者: 嵯峨優美 (統数研), 中野純司 (統数研/中大)

- Timsac (TIMSAC for R package)

<http://jasp.ism.ac.jp/ism/timsac/>

時系列解析統合プログラムパッケージTIMSACを基に統計数理研究所が開発したR関数群

その他の時系列解析Rパッケージ

- dlm (An R package for dynamic linear model)
- KFAS (Exponential family state space models in R)

WebDecompについて

本講義で取り上げる時系列解析法の一部をWebで実行できるようにしたもの。

オリジナルソース

統計数理研究所が開発したTIMSAC-84の中のプログラムDECOMPなど

- Web Decomp

<http://ssnt.ism.ac.jp/inets/inets.html>

ソフトウェアのインストールを行う必要がなく、データを貼り付けると計算結果を返してくれる。

開発してからかなり時間が経っている。

開発者： 佐藤整尚（東大・経済）

- RS-decomp

R-shinyによるWebアプリケーション。

現時点では非公開

開発者： 中野純司（統数研/中大）， 嵯峨優美（統数研）

講義資料

数理・情報教育研究センターHP

<http://www.mi.u-tokyo.ac.jp/index.html>

関連講義 > プログラムの詳細 > 数理手法VII(時系列解析) 授業HP

http://www.mi.u-tokyo.ac.jp/mds-oudan/講義資料一式%28数理手法VII%29/time_series_analysis.html

- PowerPoint資料
- Rコード
- 講義で使う時系列データ

時系列解析教材

学部横断型教育プログラム「数理・データサイエンス教育プログラム」
数理手法VII (時系列解析) 2019年版
北川源四郎

講義資料

- 第1回 (4月10日) [ppt資料](#) [Rコード](#)

講義で使用する時系列データ

- 一括ファイル [時系列データ.zip](#)
- [hakusan_new.csv](#)
- [sunspot_new.csv](#)
- [maxtemp.csv](#)
- [blsfood_new.csv](#)
- [whard_new.csv](#)
- [mye1f_new.csv](#)
- [nikkei225_new.csv](#)
- [haibara_new.csv](#)
- [lynx_new.csv](#)
- [rainfall_new.csv](#)

April 8, 2019

講義で使う時系列データ

時系列データ.zip： 下記のcsvファイルを含むzipファイル
(一部はRパッケージTSSSSに収録されている)

csvファイル	TSSSS	データ概要
hakusan_new.csv	HAKUSAN	船舶の方向角速度, 横揺れ, 縦揺れ, エンジン, 回転数(秒)
sunspot_new.csv	Sunspot	太陽黒点数 (年, 月)
maxtemp.csv	Temperature	東京の最高気温(日, 月)
blsfood_new.csv	BLSALLFOOD	アメリカの食品産業に従事する労働者人口 (月)
whard_new.csv	WHARD	ある商品の卸売高(月, 日)
mye1f_new.csv	MYE1F	地震波 の東西成分 (1/100秒)
nikkei225_new.csv		日経225株価 (日)
haibara_new.csv		地下水位と気圧 (分、時間)
Lynx-new.csv		Canadian Lynx捕獲数 (年)
rainfall_new.csv		降雨データ (東京, 日)

数理・情報教育研究センターHP (関連講義)

http://www.mi.u-tokyo.ac.jp/mds-oudan/講義資料一式%28数理手法VII%29/time_series_analysis.html

R_code (抜粋)

```
#####  
### データの読み込み ###  
#####  
  
# 船舶のデータ  
hakusan <- as.ts(read.csv("hakusan_new.csv"))  
hakusan1 <- hakusan[,1] # 方向角速度データ  
  
# 太陽黒点数データ  
sunspot <- as.ts(read.csv("sunspot_new.csv"))  
  
# 東京の日最高気温データ  
maxtemp <- as.ts(read.csv("maxtemp.csv"))  
  
# アメリカの食品産業に従事する労働者人口  
blsfood <- as.ts(read.csv("blsfood_new.csv"))  
  
#####  
### データのプロット ###  
#####  
par(mar=c(2,2,1,1)+0.1) # 余白を小さくする  
plot(hakusan1)  
plot(sunspot)  
plot(whard)  
plot(log(whard)) # log(whard)のプロット
```

```
#####  
### 日経225データの対数差分 ###  
#####  
x <- as.ts(nikkei225)  
y <- log(x)  
z <- diff(y)  
plot(z)  
# 前期比  
x <- as.ts(whard)  
plot(x/lag(x))  
  
# 前年同期比  
plot(x/lag(x,k=-12))  
  
#####  
### 移動平均と移動メディアン ###  
#####  
  
# 移動平均フィルタ (TTRパッケージの関数 SMA)  
plot(maxtemp,ylim=c(0,40))  
x <- SMA(maxtemp,17)  
lines(x,col=2,lwd=1)
```

統計データのいろいろ

1. 独立な観測データ

調査データ, 治験データ, 多変量解析データ

2. 時系列データ

株価データ, 地震データ, 制御系データ

3. 空間 (平面) データ

GIS(Geographic Information System)

4. 時空間データ

気象データ, 人口流動データ

時系列とは

時系列: Time Series

時間とともに不規則に変動する現象の記録

n, t	観測した時刻（本講義では通常 n を使う）
$y_n, y(n)$	時刻 n の観測値（本講義では通常 y_n を使う）
N, T	データ数（本講義では通常 N を使う）
y_1, \dots, y_N	時系列

- 世の中の興味あるデータの多くは時系列
- 数学的には、時系列は時刻をパラメータとする確率変数の実現値と考えるが、本講義では明確に区別しない

Y_n : 確率過程

y_n : 時系列

(本講義における) 時系列解析

- 可視化・特徴把握

時系列を図示したり, 基本的な記述統計量を用いて時系列の特徴や解析結果を簡潔に表現する

- モデリング

時系列の特徴を的確に捉えた時系列モデルをつくる

- 予測, シミュレーション

現在までに得られた情報から今後の変動を推測する

- 情報抽出

時系列から有用な情報を抽出する

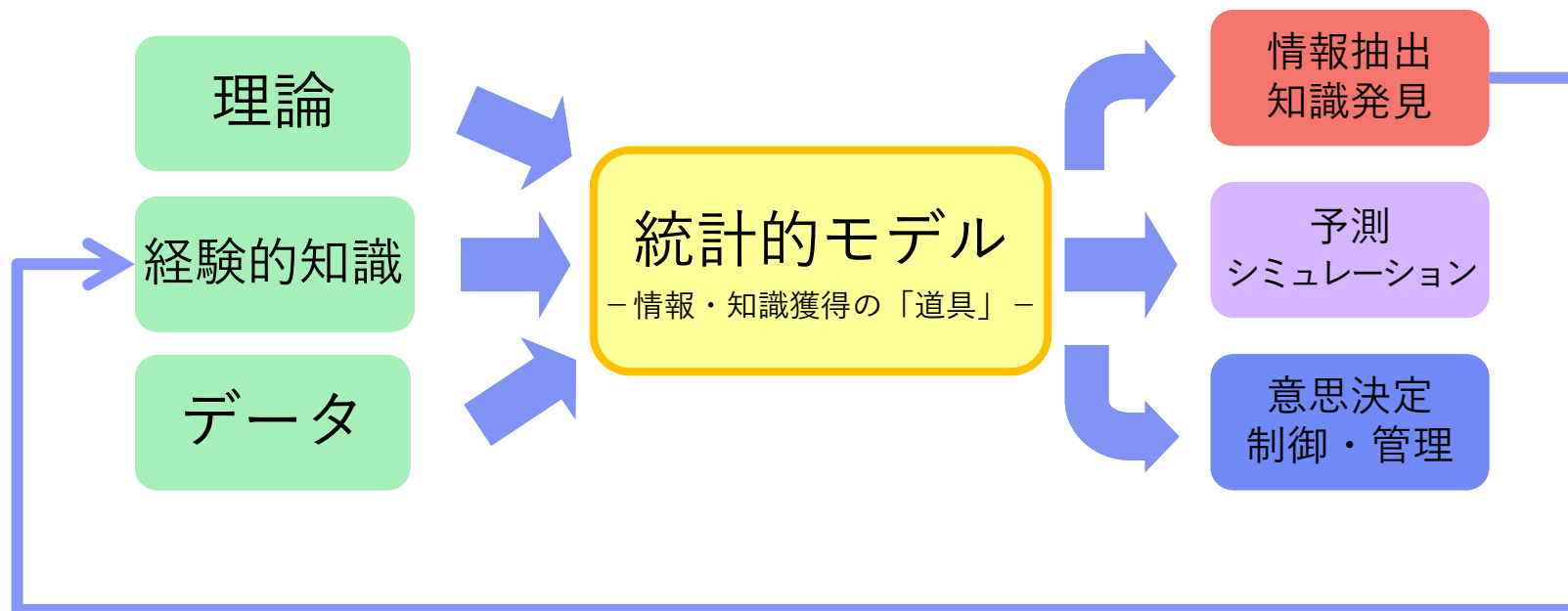
- 制御・意思決定

操作可能な変数を適当に変化させ目的とする制御変数が望ましい変動をするように制御すること

統計的モデリング

対象に関するあらゆる取得可能な情報（対象に関する理論，経験的知識，観測データ）およびモデリングの目的

事前情報とデータの持つ情報の統合 ➡ ベイズモデリング



時系列の可視化(特徴の抽出)



- データプロット (直接表示)
- 時間相関を可視化 (相関関数)
- 周期性を可視化 (スペクトル、ピリオドグラム)

データを読み込んで時系列化

```
# csvデータの読み込み
x <- read.csv("hakusan_new.csv")

# 時系列化する
hakusan <- as.ts(x)

# 多変量データの第1成分を取り出す
hakusan1 <- hakusan[,1]
```

```
# 船舶のデータ
hakusan <- as.ts(read.csv("hakusan_new.csv"))
hakusan1 <- hakusan[,1] # 方向角速度データ
```

```
# 太陽黒点数データ
sunspot <- as.ts(read.csv("sunspot_new.csv"))
```

```
# 東京の日最高気温データ
maxtemp <- as.ts(read.csv("maxtemp.csv"))
```

```
# アメリカの食品産業に従事する労働者人口
blsfood <- as.ts(read.csv("blsfood_new.csv"))
```

```
# 工業製品の卸売り高
whard <- as.ts(read.csv("whard_new.csv"))
```

```
# 地震データ（東西方向）
mye1f <- as.ts(read.csv("mye1f_new.csv"))
```

```
# 日経225データ
nikkei225 <-
as.ts(read.csv("nikkei225_new2.csv",header=TRUE))
```

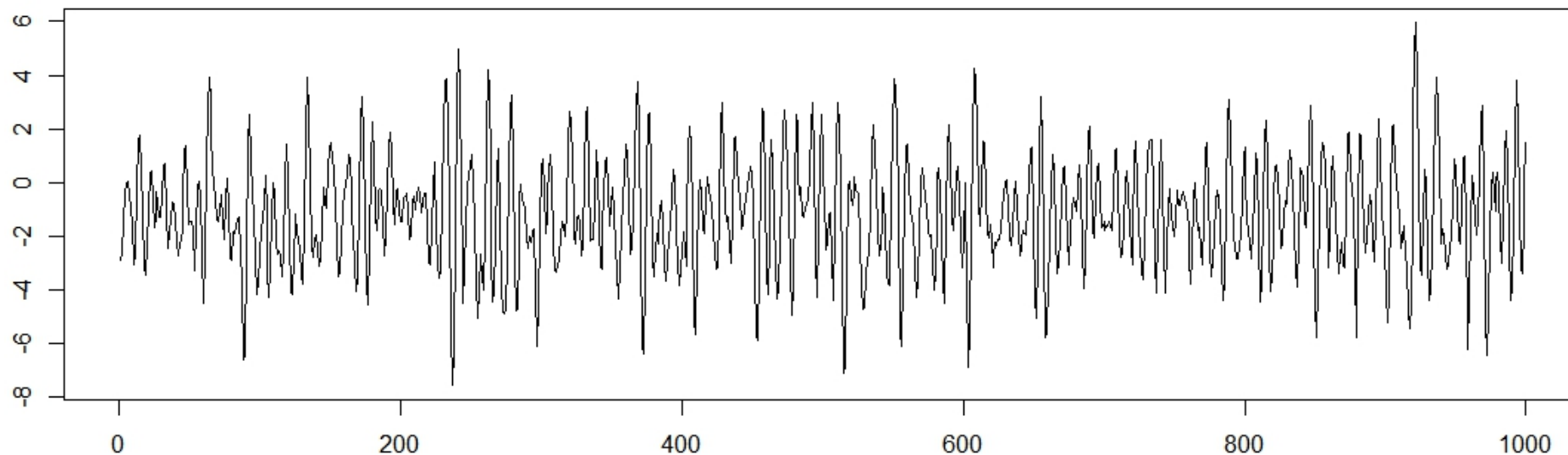
```
# 東京降雨データ（2年間の降雨日数）
rainfall <-
as.ts(read.csv("rainfall_new.csv",header=TRUE))
rainfall2 <- rainfall[,4]/2 # 2年間の平均
```

```
# 榛原地下水位データ
haibara <- as.ts(read.csv("haibara_new.csv"))
```

船舶の方向角速度データ

```
hakusan <- as.ts(read.csv("hakusan_new.csv"))  
plot(hakusan[,1])
```

$\Delta n = 1$ 秒



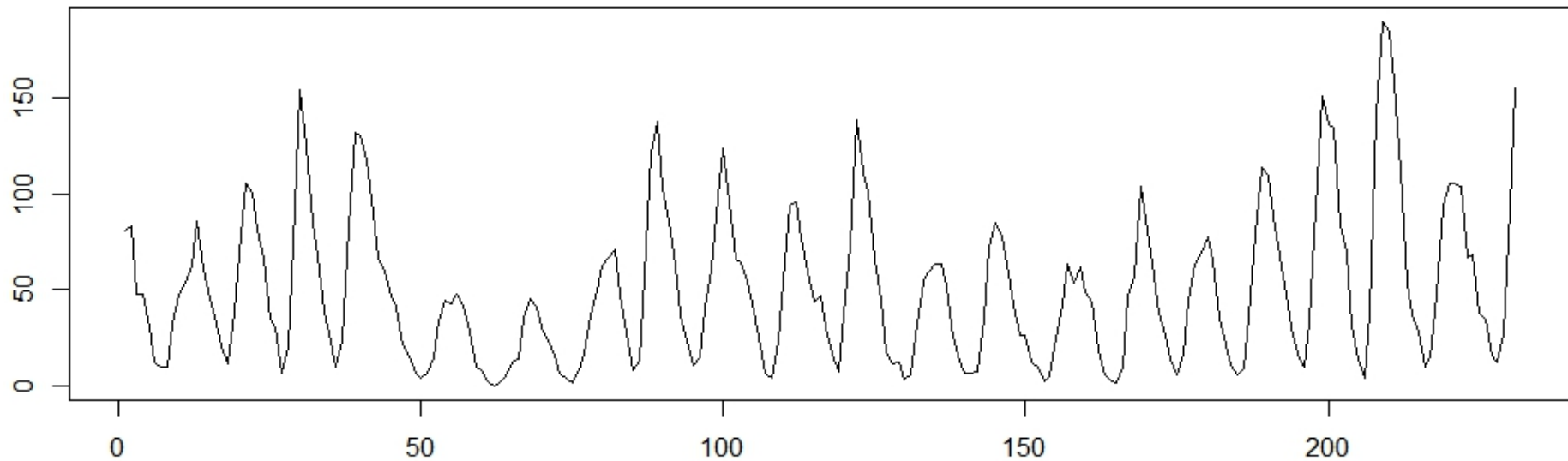
このデータの特徴

- 定常性
- 周期性

太陽黒点数データ

```
sunspot <- as.ts(read.csv("sunspot_new.csv"))  
plot(sunspot)
```

$\Delta n = 1$ 年

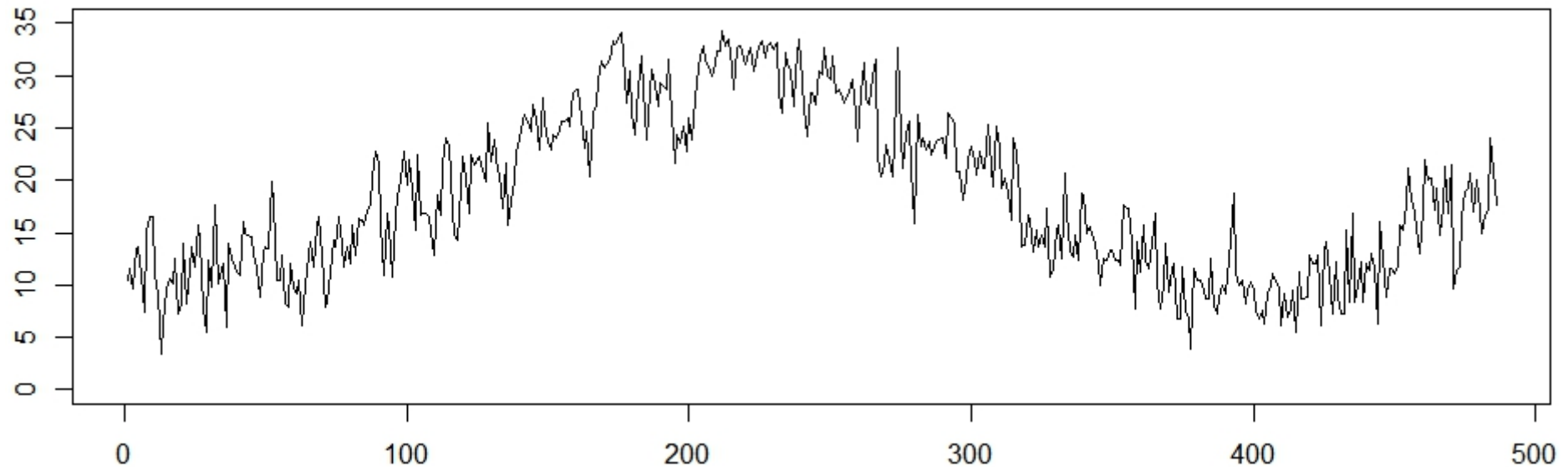


- 正值，上下非対称
- 擬似周期性
- 前後非対称性

日最高気温データ (東京)

```
maxtemp <- as.ts(read.csv("maxtemp.csv"))  
plot(maxtemp,ylim=c(0,35))
```

$\Delta n = 1$ 日

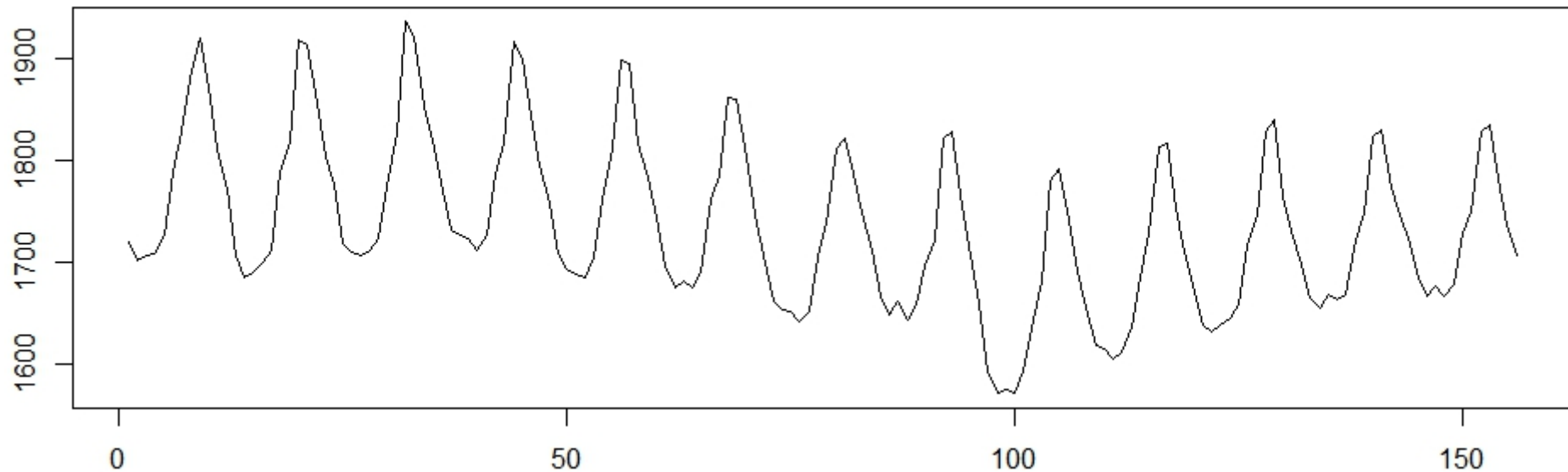


- トレンド (長期の周期性)
- トレンド周りほぼ定常

米国食品産業従事者数データ

```
blsfood <- as.ts(read.csv("blsfood_new.csv"))  
plot(blsfood)
```

$\Delta n = 1$ 月

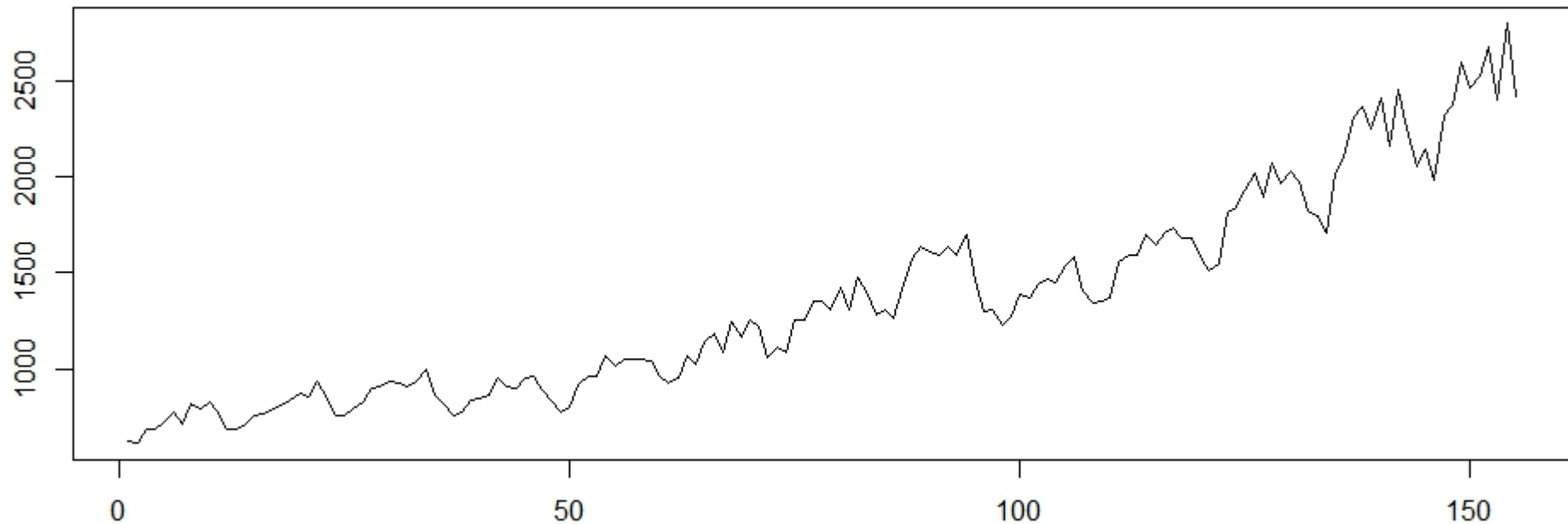


- 年周期性
- トレンド

卸売り高データ

```
whard <- as.ts(read.csv("whard_new.csv"))  
plot(whard)
```

$\Delta n = 1$ 月

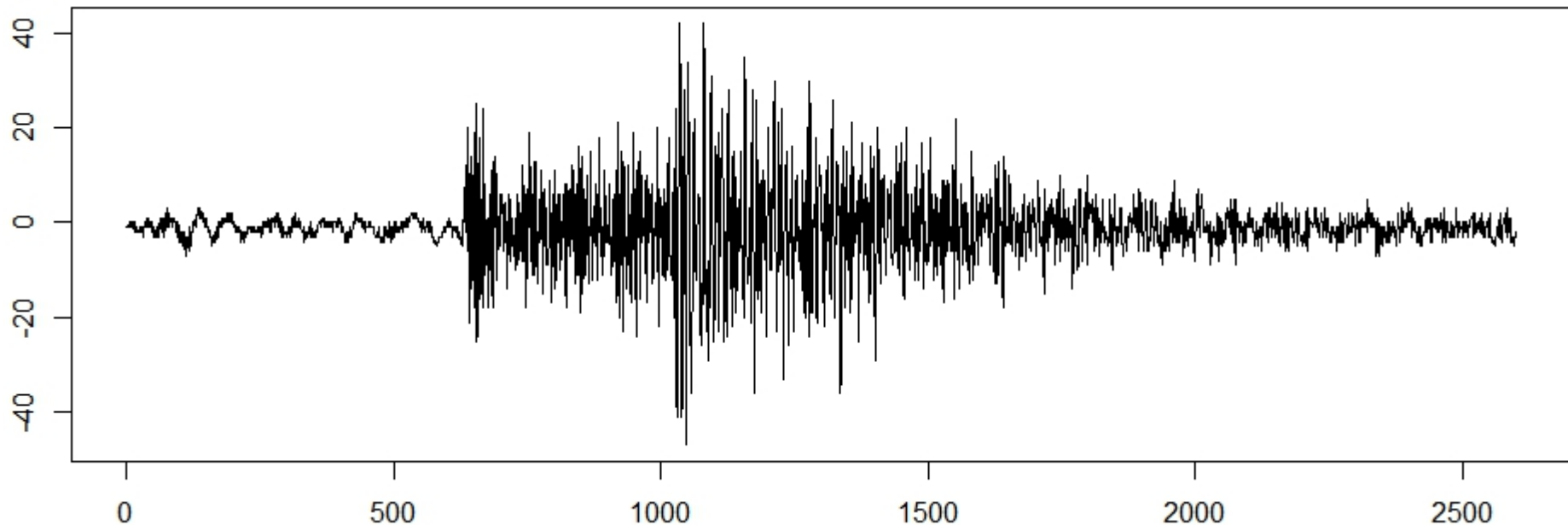


- 年周期性
- 正值性
- トレンドと変動分散の連動

MYE1F（東西方向地震波）データ

```
mye1f <- as.ts(read.csv("mye1f_new.csv"))  
plot(mye1f)
```

$\Delta n = 0.01$ 秒

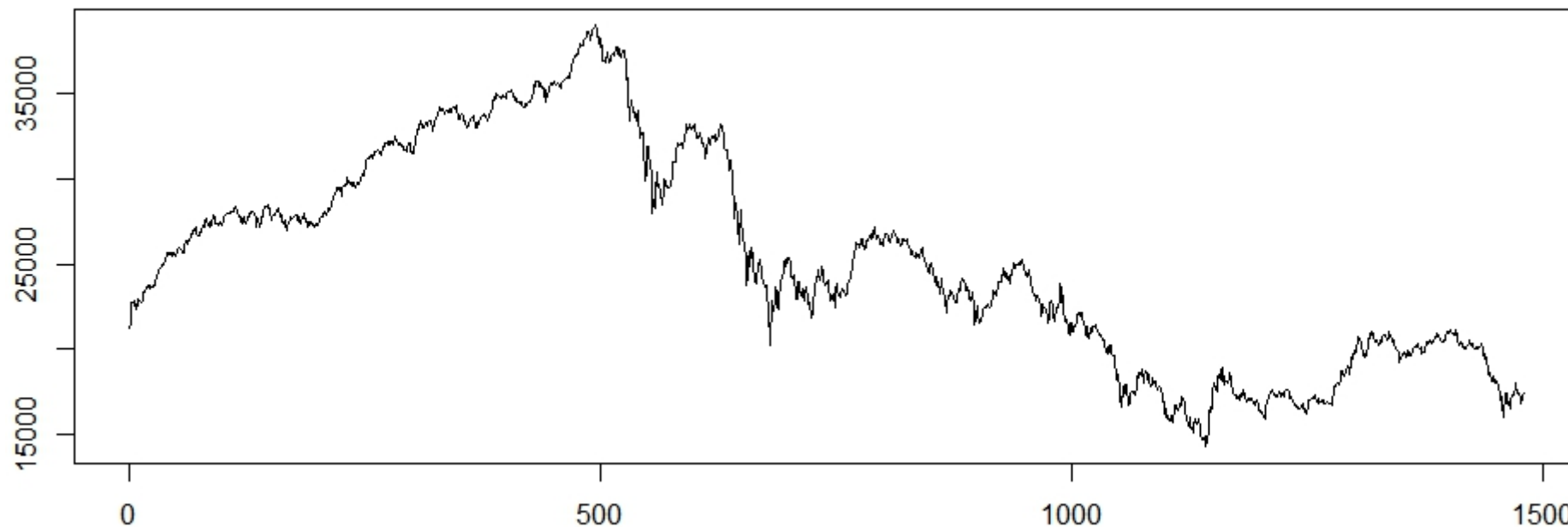


- トレンドなし
- 分散非定常, 共分散非定常
- 区分的定常

日経225平均株価データ

```
nikkei225 <- as.ts(read.csv("Nikkei225_new.csv"))  
plot(nikkei225)
```

$\Delta n = 1$ 日

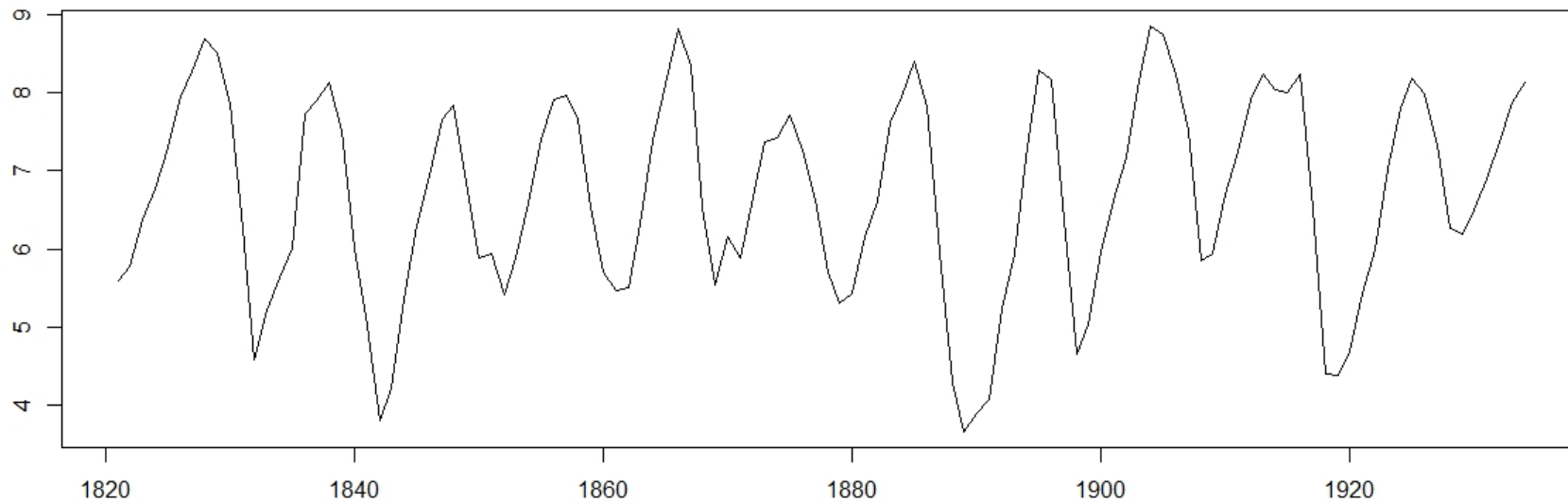


- トレンド + 分散変動
- トレンド下降時の分散増加

Canadian Lynx捕獲数

```
lynx <- as.ts(read.csv("Lynx-new.csv"))  
plot(lynx)  
plot(log(lynx))
```

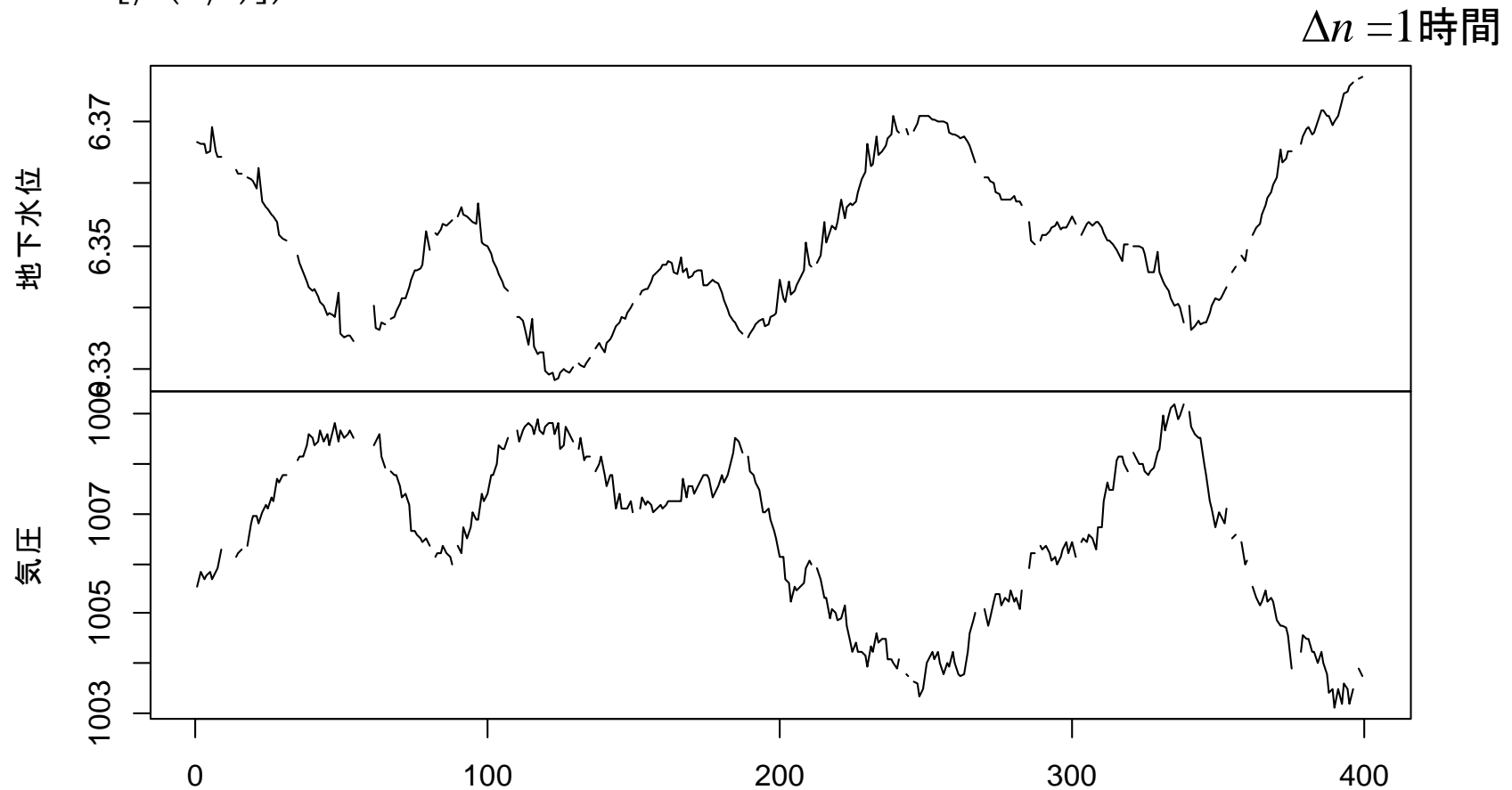
$\Delta n = 1$ 年



- 平均定常
- 擬似周期的
- 前後(左右)非対称

榛原地下水位・気圧データ

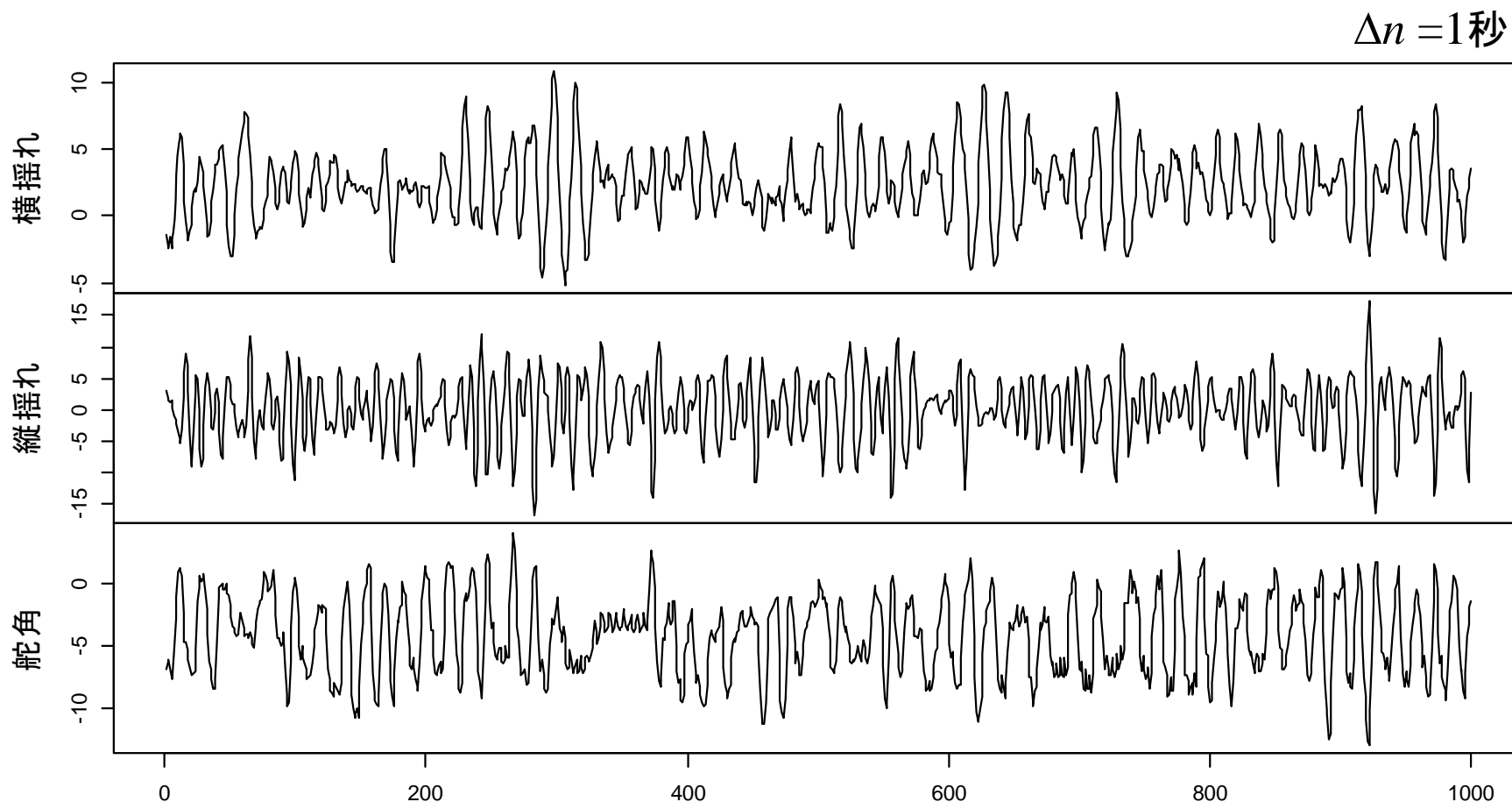
```
haibara <- as.ts(read.csv("haibara_new.csv"))  
plot(haibara[,c(2,4)])
```



- 2 変量データ, 逆相関
- 異常値, 欠測値を含む

船舶データ（横揺れ、縦揺れ、舵角）

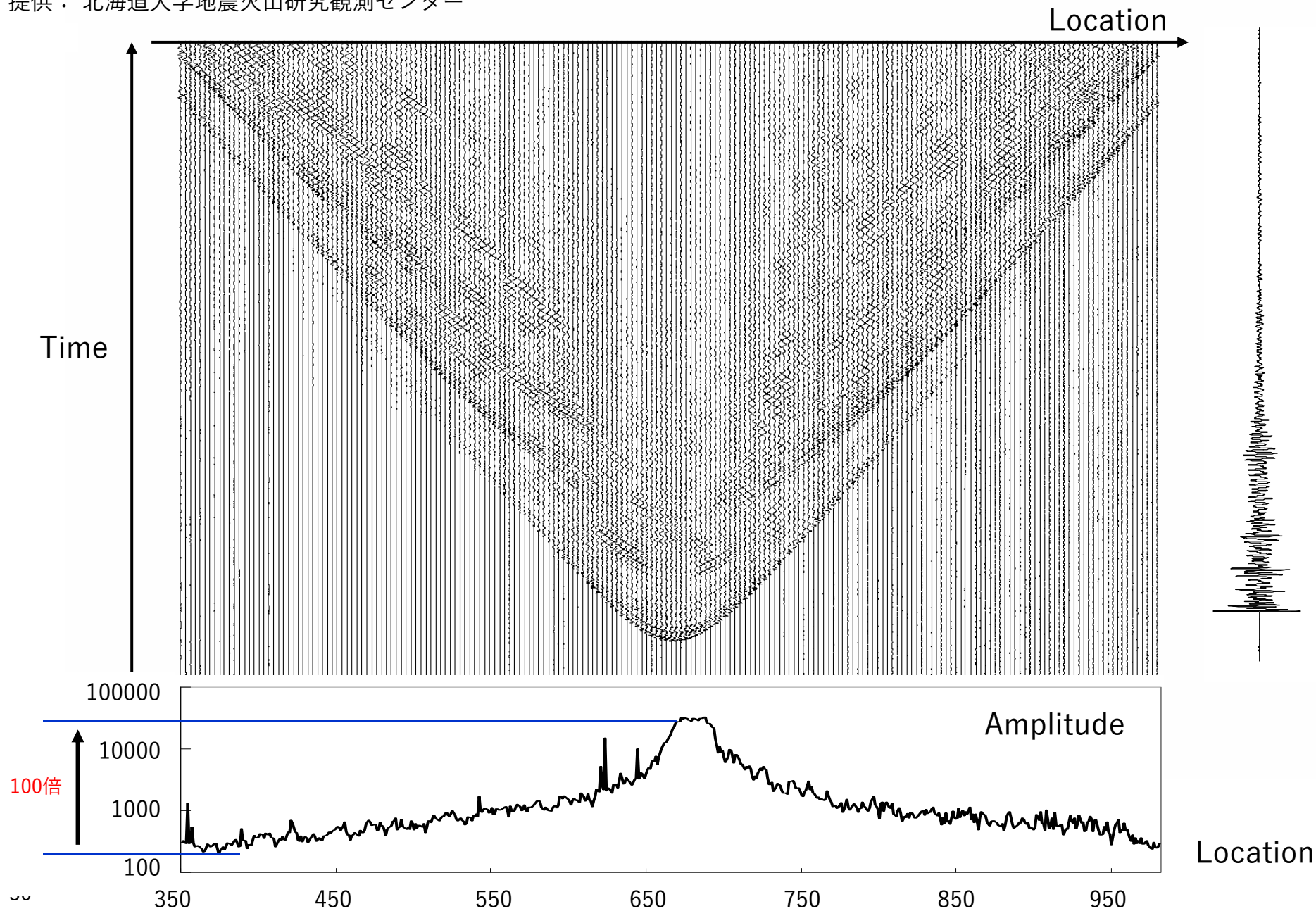
```
hakusan <- as.ts(read.csv("hakusan_new.csv"))  
plot(hakusan[,c(3,4,7)])
```



- 多変量制御系データ
- 舵角が入力，横揺れ・縦揺れは出力

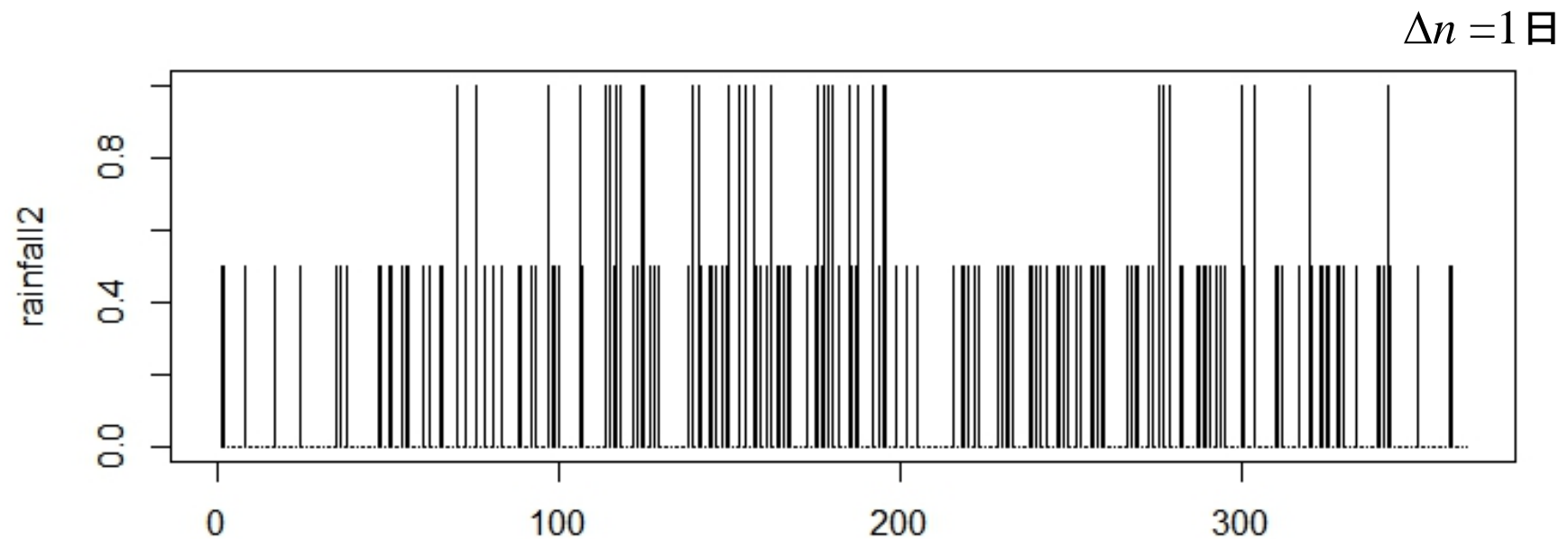
海底地震計データ (調整済, 982ch)

提供：北海道大学地震火山研究観測センター



東京降水量データ

```
rainfall <- as.ts(read.csv("rainfall_new.csv",header=TRUE))  
rainfall2 <- rainfall[,4]/2  
plot(rainfall2,type="h")
```



- 離散値
- 平均値(確率)非定常

- 連続時間時系列と離散時間時系列
 - 離散：等間隔時系列と不等間隔時系列
- 一変量時系列と多変量時系列
- 定常時系列と非定常時系列
 - 定常：性質が一定で時間的に変化しないもの
 - 非定常：性質が時間と共に変化するもの
- 線形時系列と非線形時系列
- ガウス型時系列と非ガウス型時系列

時系列固有の性質というよりは、モデリングにおける観点

時系列の前処理



- 変数変換
- 階差（差分），季節階差
- 前期比・前年比
- 移動平均・移動メディアン
- 欠測値・異常値

- ✓ モデリングを容易にするため
- ✓ 副作用もある

変数変換

目的

- 線形，定常，正規性などを仮定したモデリングを容易にするため
- パラメータ最適化を容易にするため

例

- 対数変換
- ロジット変換
- Box-Cox変換
- (階差)

変数変換

$$X \sim f(x) \quad \begin{array}{c} \xrightarrow{y = h(x)} \\ \xleftarrow{x = h(y)^{-1}} \end{array} \quad Y \sim g(y)$$

変数変換によって確率分布も変形される

$$g(y) = f(h^{-1}(y)) \left| \frac{dh^{-1}}{dy} \right|$$

$$f(x) = g(h(x)) \left| \frac{dh}{dx} \right|$$

定義域が変わることがある

$$h(x) = \begin{pmatrix} h_1(x_1, \dots, x_m) \\ \vdots \\ h_m(x_1, \dots, x_m) \end{pmatrix}$$

ヤコビアン (ヤコビ行列)

$$\frac{\partial h}{\partial x} = \begin{pmatrix} \frac{\partial h_1}{\partial x_1} & \dots & \frac{\partial h_1}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_m}{\partial x_1} & \dots & \frac{\partial h_m}{\partial x_m} \end{pmatrix}$$

対数変換

$$y = h(x) = \log x \quad (\text{対数変換})$$

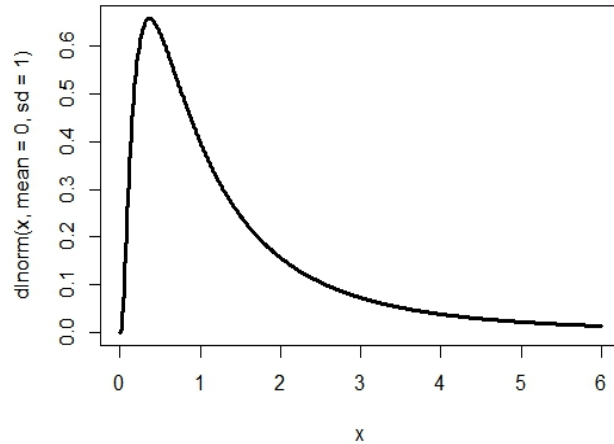
$$g(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \quad (\text{正規分布})$$

$$\frac{dh}{dx} = \frac{1}{x} \quad (\text{ヤコビアン})$$

(対数正規分布)

$$f(x) = g(h(x)) \left| \frac{dh}{dx} \right| = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}} \frac{1}{x}$$

対数変換



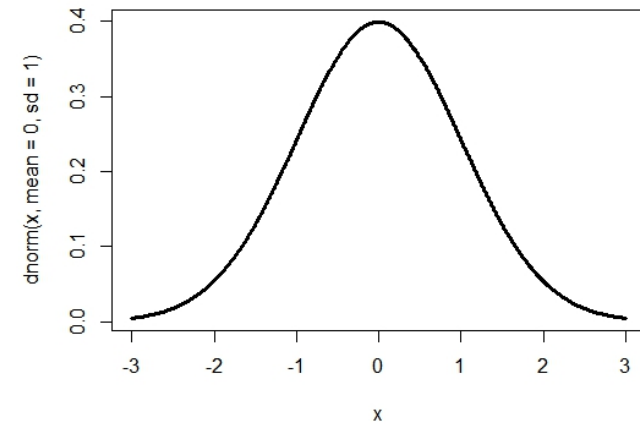
$0 < x < \infty$
対数正規分布

$$y = \log x$$

→

$$x = e^y$$

←



$-\infty < y < \infty$
正規分布

- 正規分布に近づけてからモデリングを行う

対数変換後のモデリング

$$y_n = \log x_n$$

$$y_n = t_n + w_n \quad \Rightarrow \quad x_n = T_n \cdot W_n$$

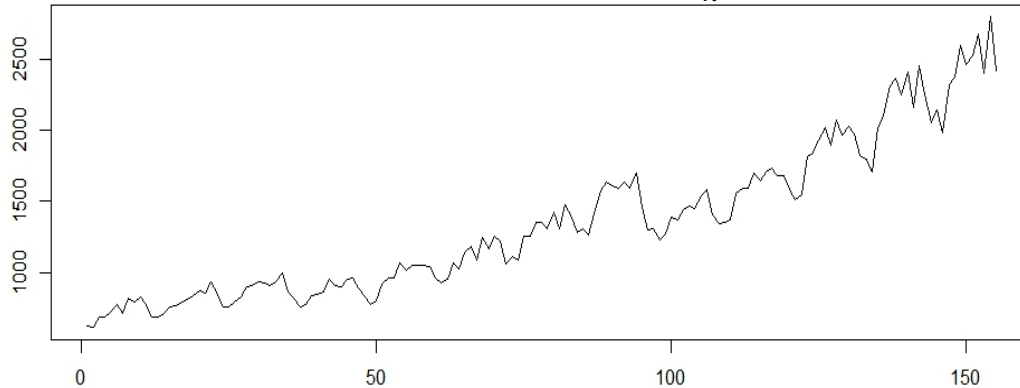
$$x_n = e^{y_n} = e^{t_n + w_n} = e^{t_n} \cdot e^{w_n} = T_n \cdot W_n$$

対数変換後に加法的モデリングを行うことは原データに関して**乗法型モデリング**を行うことに相当する。

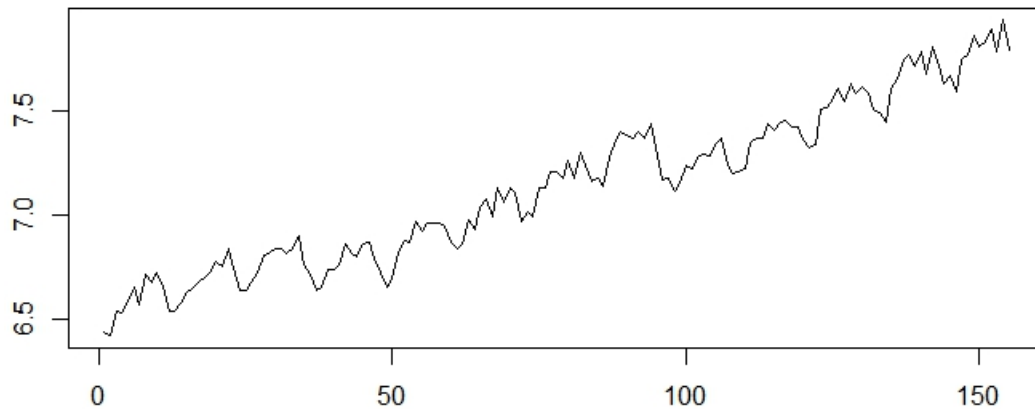
ただし、各成分の分布も変数変換を受ける
(モデリングの項で再考)

時系列の対数変換

Whardデータ x_n



$$y_n = \log x_n$$



- 分散一様化
- トレンドは消えない
- 正值時系列の無制約化
 $(0, \infty) \rightarrow (-\infty, \infty)$

`plot(whard)`
`plot(log(whard))`

Box-Cox変換

$$y_\lambda = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log x & \lambda = 0 \end{cases}$$

- λ によって様々なべき乗型変換が得られる
- データの分布を正規分布に近づけるために用いられる

$y_\lambda = x^\lambda$ のとき

$\lambda = 2$	$y = x^2$	二乗
$\lambda = 1$	$y = x$	不変
$\lambda = 0.5$	$y = \sqrt{x}$	平方根
$\lambda = 0$	$y = \log x$	対数
$\lambda = -1$	$y = 1/x$	逆数

λ の決定法？

ロジット変換

比率データ x をオッズ $x/(1-x)$ の自然対数に変換

x : 確率, 割合 $0 < x < 1$

$$y = \log\left(\frac{x}{1-x}\right)$$

$$(0, 1) \leftrightarrow (-\infty, \infty)$$

逆変換はシグモイド関数

$$x = \frac{e^y}{1 + e^y}$$

$$\begin{aligned} e^y = \frac{x}{1-x} &\Rightarrow (1-x)e^y = x \\ &\Rightarrow e^y = x(1+e^y) \end{aligned}$$

より一般の変換

$$y = \log\left(\frac{x-a}{b-x}\right)$$

$$(a, b) \leftrightarrow (-\infty, \infty)$$

逆変換

$$x = \frac{a + be^y}{1 + e^y}$$

$$e^y = \frac{x-a}{b-x} \Rightarrow (b-x)e^y = (x-a)$$

$$\Rightarrow a + be^y = x(1 + e^y)$$

$$\Rightarrow x = \frac{a + be^y}{1 + e^y}$$

階差 (差分)

y_n 時系列

$$\Delta y_n = y_n - y_{n-1} \quad \text{1階階差} \qquad \Delta^k y_n \quad k \text{階階差}$$

$$\Delta^2 y_n = \Delta y_n - \Delta y_{n-1} = y_n - 2y_{n-1} + y_{n-2}$$

- 1階階差 $y_n = a + bn$ (直線) のとき

$$\Delta y_n = y_n - y_{n-1} = b$$

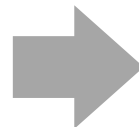
- 2階階差 $y_n = a + bn + cn^2$ のとき

$$\Delta y_n = y_n - y_{n-1} = (b - c) + 2cn$$

$$\Delta^2 y_n = 2c$$

注意

$$y_n = a + bn + \varepsilon_n$$



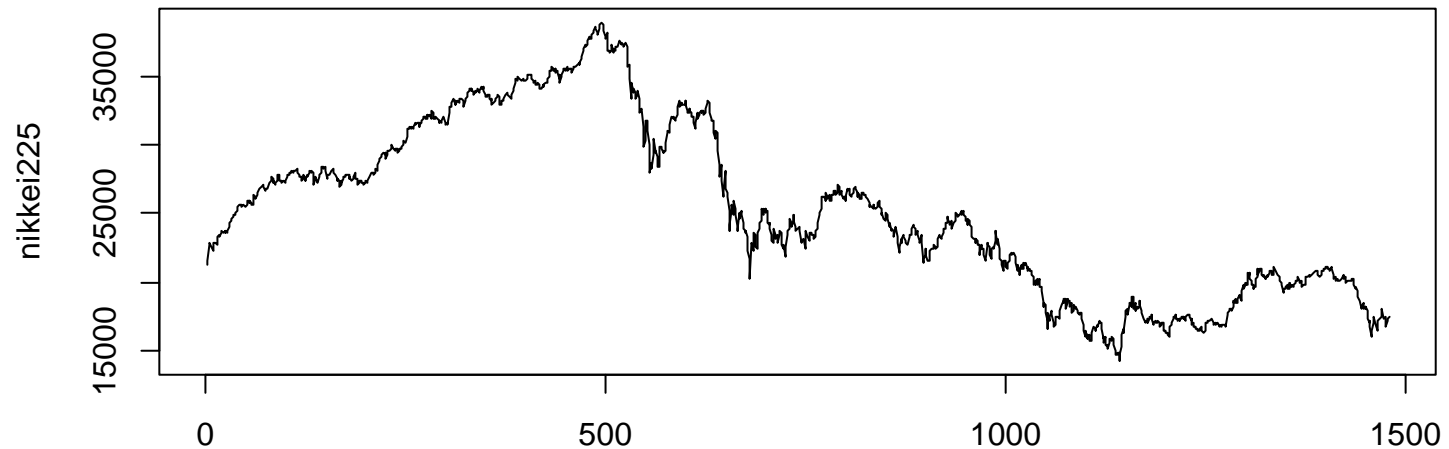
$$\Delta y_n = b + \varepsilon_n - \varepsilon_{n-1}$$

$$\Delta^2 y_n = \varepsilon_n - 2\varepsilon_{n-1} + \varepsilon_{n-2}$$

nikkei225データの（対数）差分

```
plot(nikkei225)
```

平均非定常



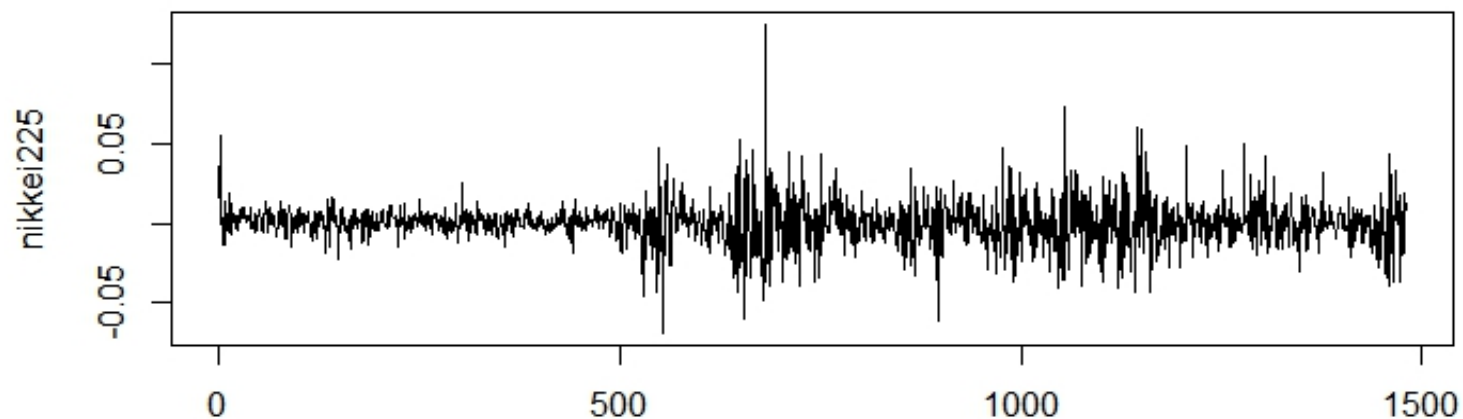
```
plot(diff(log(nikkei225)))
```

または

```
y <- log(nikkei225)
```

```
z <- diff(y)
```

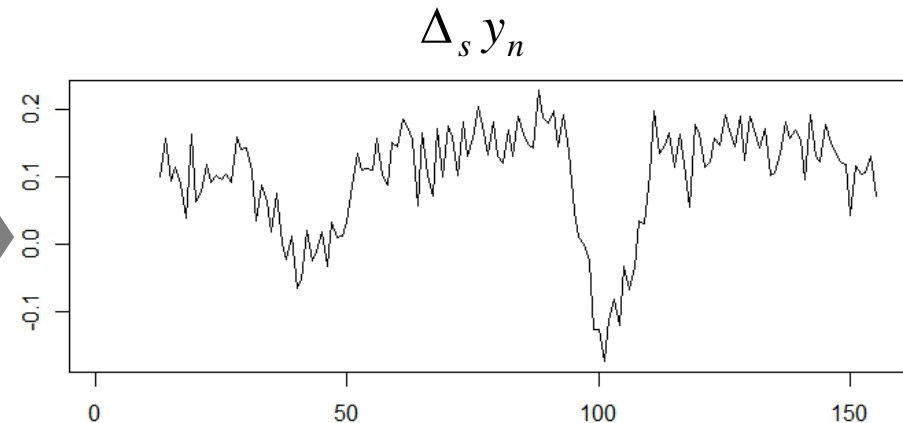
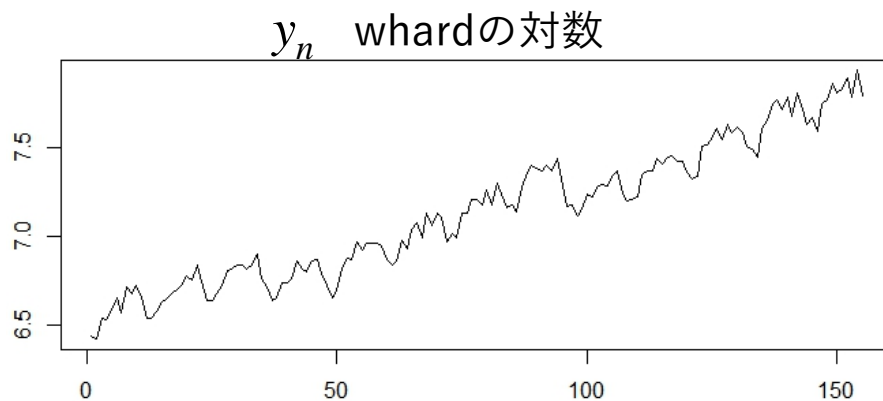
```
plot(z)
```



季節階差

s 季節(1周期)の長さ

$$\Delta_s y_n = y_n - y_{n-s} = (1 - B^s) y_n$$



季節的変動が除去できる

```
y<-log(whard)
z<-rep(NA,n)
n <- length(whard)
period <- 12
z[1:n] <- NA
for(i in period+1:n){
z[i] <- y[i]-y[i-period]}
plot(as.ts(z))
```

前期比

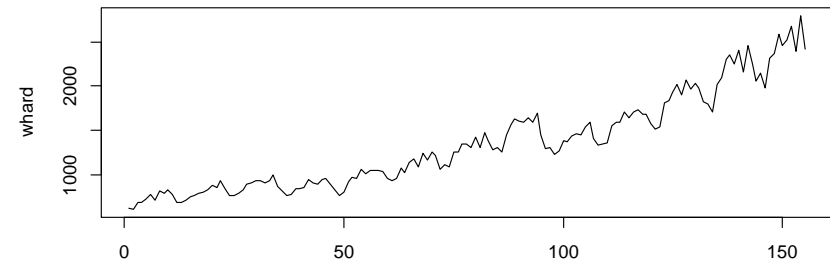
$$x_n = y_n / y_{n-1}$$

対数差分とほぼ同じ

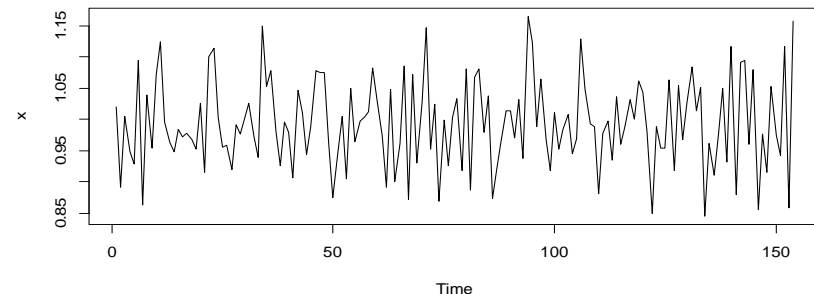
$$\begin{array}{l} y_n = T_n \\ T_n = (1 + \alpha)T_{n-1} \end{array} \quad \longrightarrow \quad x_n = \frac{T_n}{T_{n-1}} = \frac{(1 + \alpha)T_{n-1}}{T_{n-1}} = 1 + \alpha$$

実際は $y_n = T_n w_n$ または $T_n + w_n$

$$\begin{aligned} x_n &= \frac{y_n}{y_{n-1}} = \frac{T_n \cdot w_n}{T_{n-1} \cdot w_{n-1}} \\ &= \frac{(1 + \alpha)T_{n-1} \cdot w_n}{T_{n-1} \cdot w_{n-1}} = (1 + \alpha) \frac{w_n}{w_{n-1}} \end{aligned}$$



plot(whard/lag(whard))

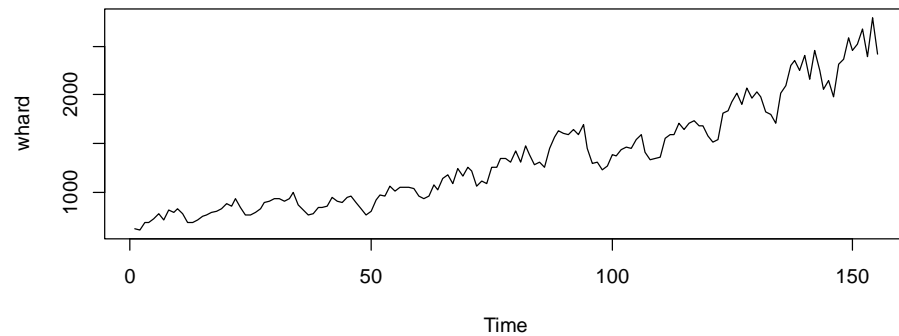


前年同期比

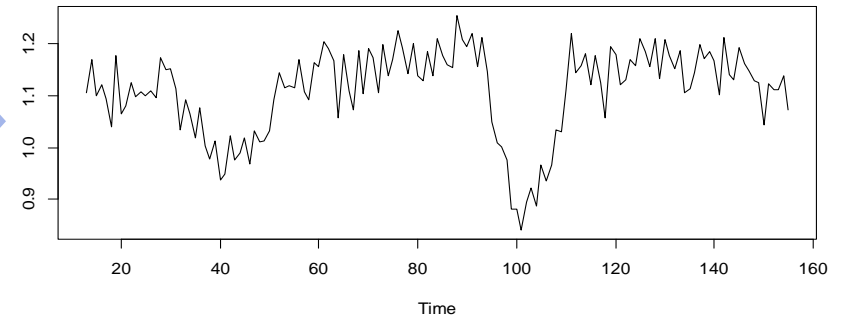
$$x_t = y_t / y_{t-p}$$

p 周期

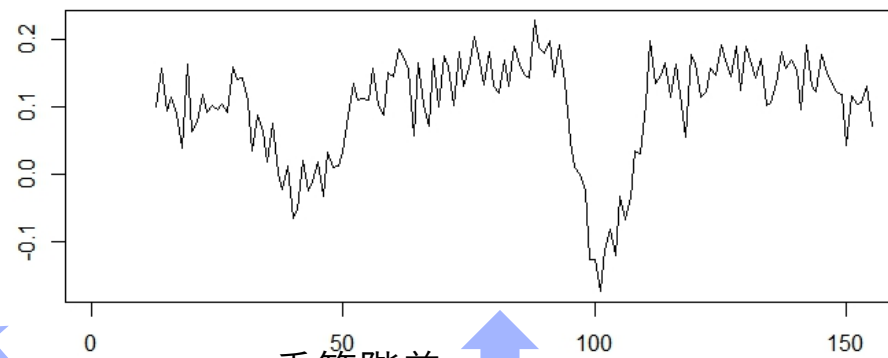
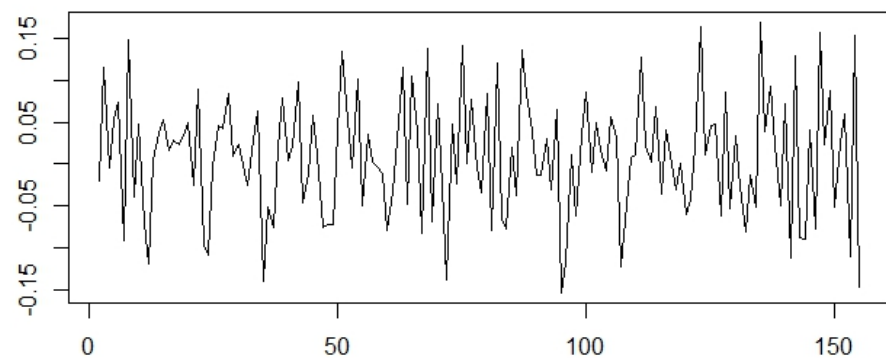
$$y_n = S_n \quad S_{nt} \cong S_{n-p} \quad \longrightarrow \quad x_n = \frac{y_n}{y_{n-p}} = \frac{S_n}{S_{n-p}} \cong 1$$



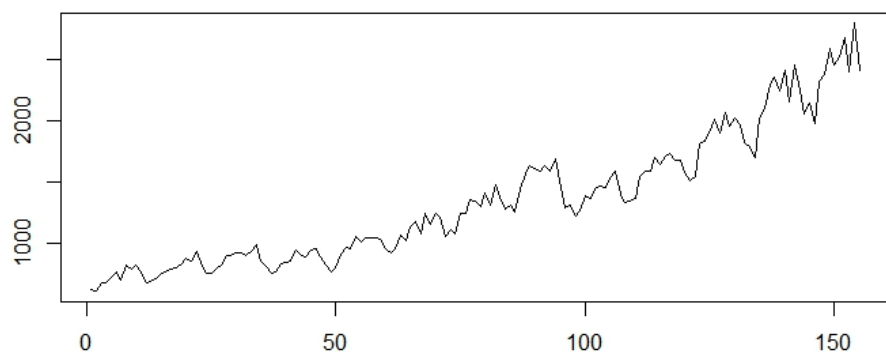
plot(whard/lag(whard,k=-12))



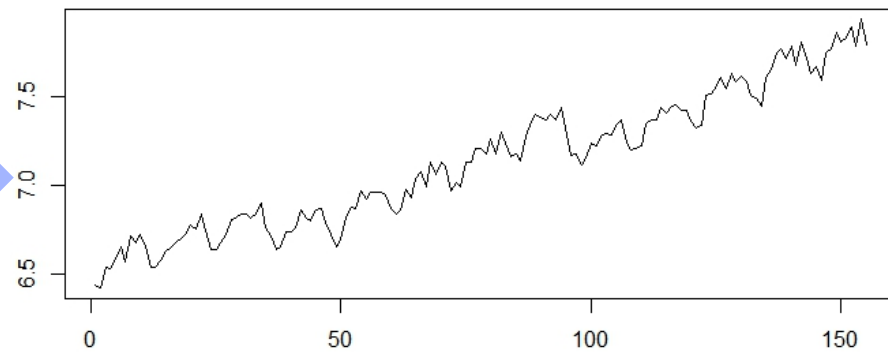
対数変換, 季節階差, 前期比, 前年同期比



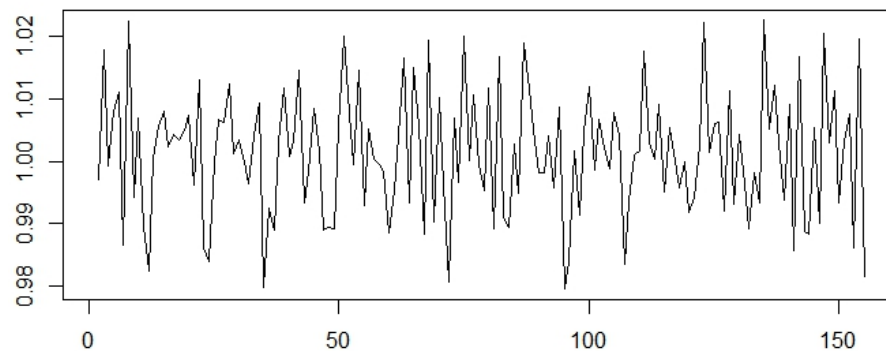
季節階差



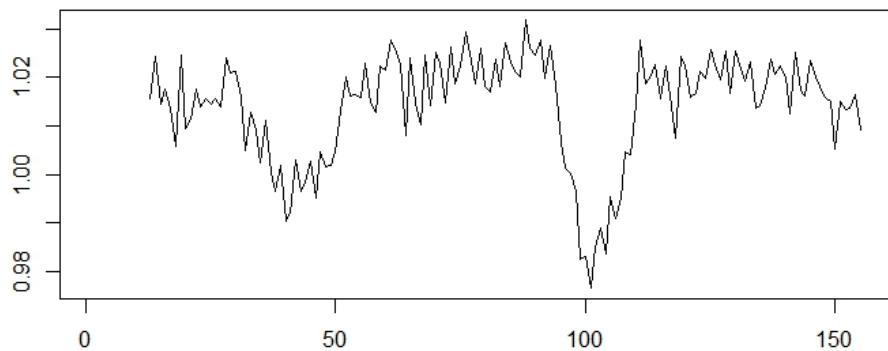
対数



前期比



前年同期比



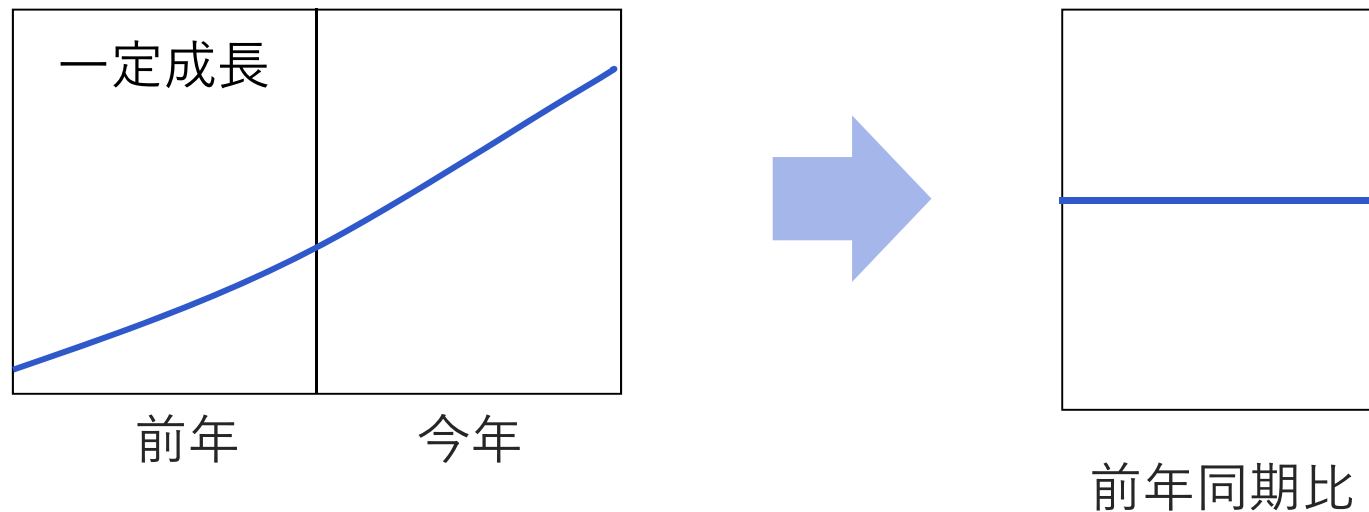
前年同期比の問題点

$$y_n = T_n \cdot S_n$$

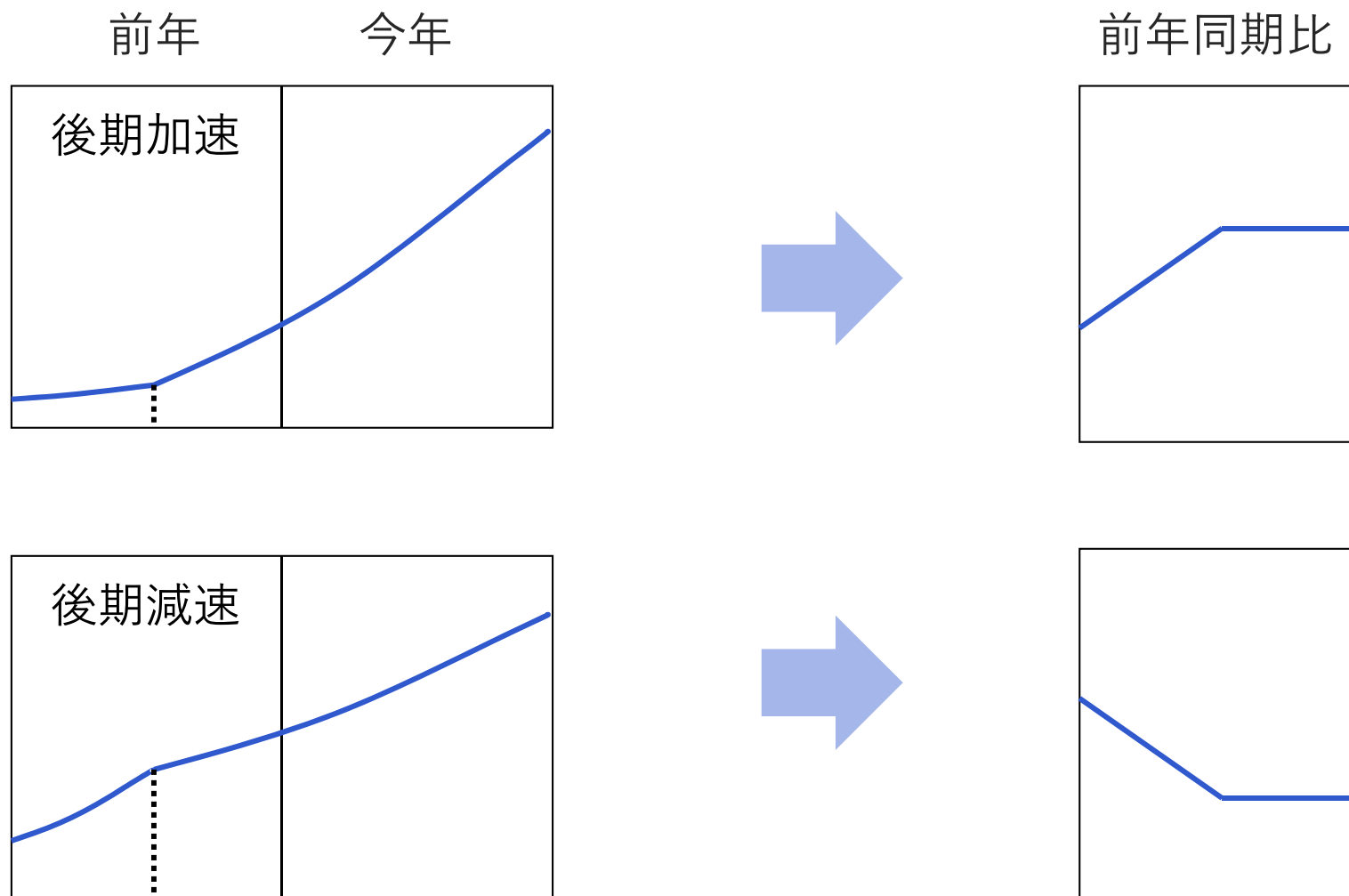
必ずしもその時点の変動に対応していない。

T_n トレンド S_n 季節成分

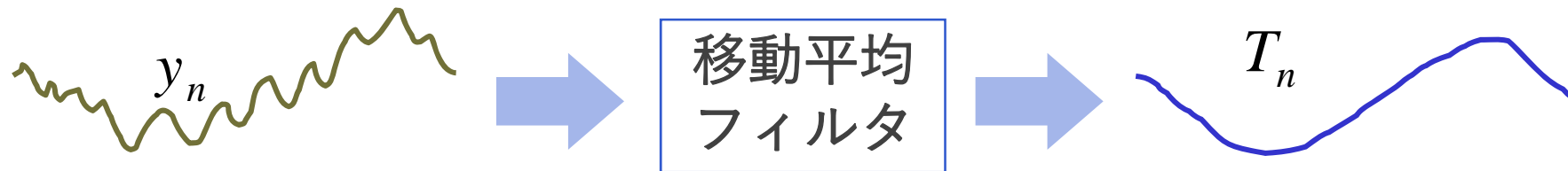
$$\frac{y_n}{y_{n-12}} = \frac{T_n \cdot S_n}{T_{n-12} \cdot S_{n-12}} \cong \frac{T_n}{T_{n-12}}$$



前年同期比の問題点



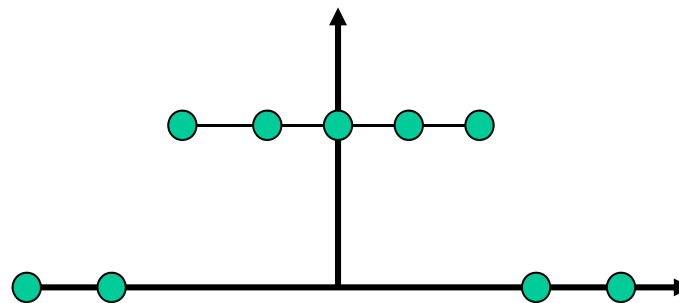
移動平均フィルタ



$$t_n = \frac{1}{2k+1} (y_{n-k} + \cdots + y_n + \cdots + y_{n+k})$$

$2k+1$: (移動平均フィルタの) 項数

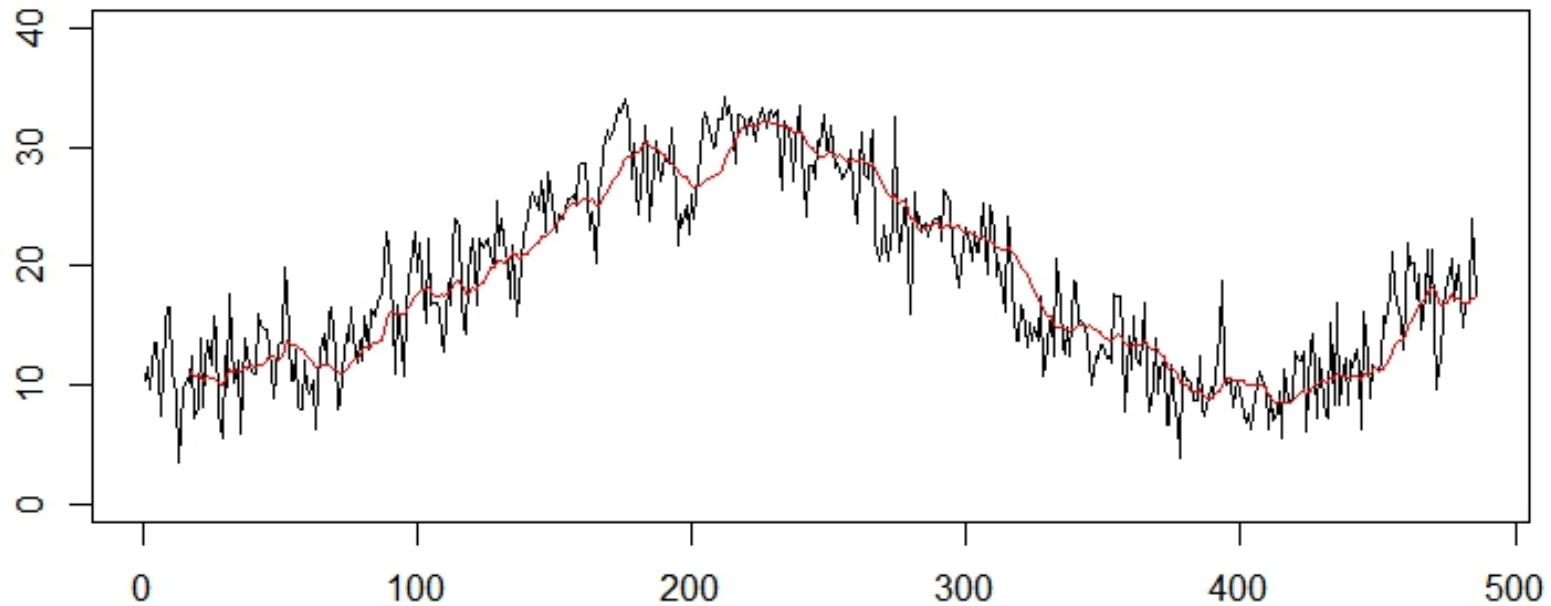
例: $k=2$



maxtemp(最高気温)データ

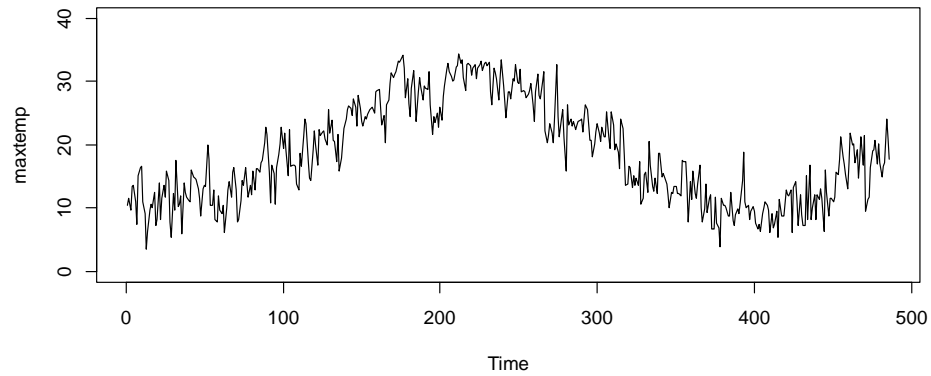
```
plot(maxtemp,ylim=c(0,40))  
x <- SMA(maxtemp,17)  
lines(x,col=2,lwd=1)
```

```
options(repos="http://cran.ism.ac.jp")  
install.packages("TTR")  
library(TTR)
```

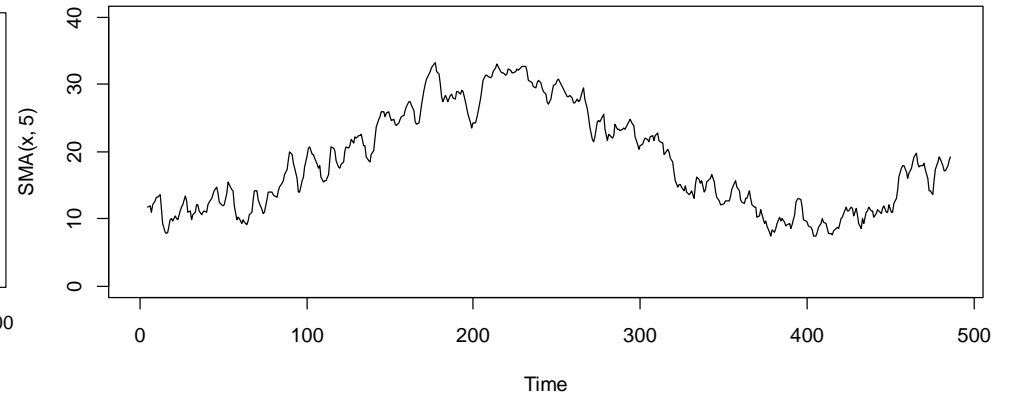


項数 k の影響

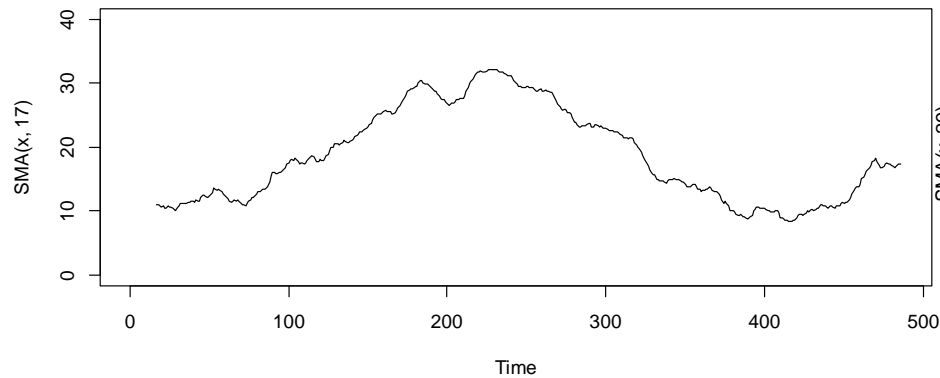
`plot(maxtemp,ylim=c(0,40))`



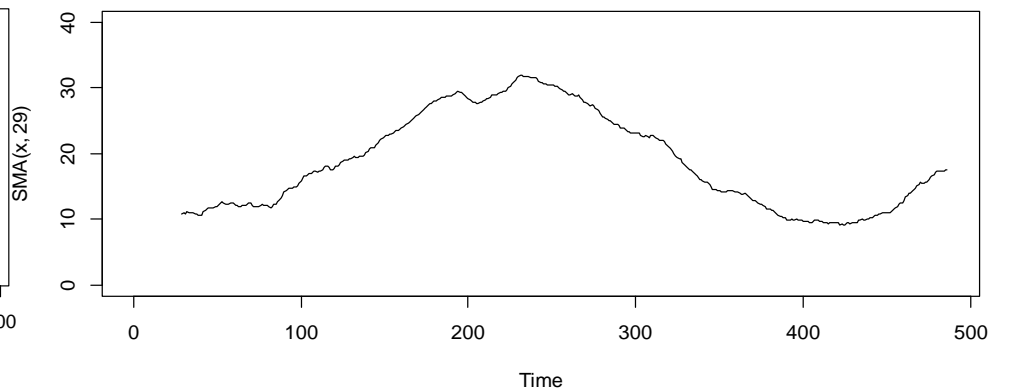
`plot(SMA(maxtemp,5),ylim=c(0,40))`



`plot(SMA(maxtemp,17),ylim=c(0,40))`



`plot(SMA(maxtemp,29),ylim=c(0,40))`



(2k+1) 項移動平均の性質

$y_n = T_n + w_n$, $T_n = a + bn$, $w_n \sim N(0, \sigma^2)$ の場合

平均：不変 分散： $\frac{\sigma^2}{2k+1}$

$$\begin{aligned} t_n &= \frac{1}{2k+1} (y_{n-k} + \cdots + y_n + \cdots + y_{n+k}) \\ &= \frac{1}{2k+1} (T_{n-k} + \cdots + T_n + \cdots + T_{n+k}) + \frac{1}{2k+1} (w_{n-k} + \cdots + w_n + \cdots + w_{n+k}) \\ &= T_n + \frac{1}{2k+1} (w_{n-k} + \cdots + w_n + \cdots + w_{n+k}) \end{aligned}$$

$$E(w_{n-k} + \cdots + w_n + \cdots + w_{n+k}) = E(w_{n-k}) + \cdots + E(w_n) + \cdots + E(w_{n+k}) = 0$$

$$\begin{aligned} E(w_{n-k} + \cdots + w_n + \cdots + w_{n+k})^2 \\ = E(w_{n-k}^2 + \cdots + w_{n+k}^2) + 2E(w_{n-k}w_{n-k+1} + \cdots + w_{n+k-1}w_{n+k}) = (2k+1)\sigma^2 \end{aligned}$$

$$E(w_n^2) = \sigma^2 \quad E(w_n w_m) = 0 \quad (n \neq m)$$

$$E\left(\frac{w_{n-k} + \cdots + w_n + \cdots + w_{n+k}}{2k+1}\right)^2 = \frac{1}{(2k+1)^2} E(w_{n-k} + \cdots + w_n + \cdots + w_{n+k})^2 = \frac{\sigma^2}{2k+1}$$

重みつき移動平均フィルタ

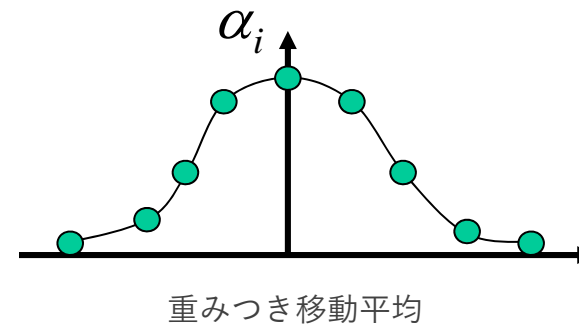
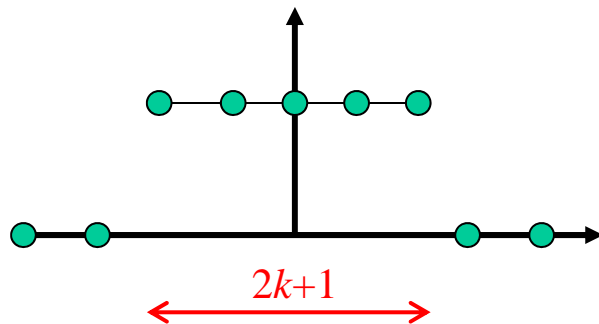
($2k+1$)項移動平均

$$t_n = \frac{1}{2k+1} (y_{n-k} + \cdots + y_n + \cdots + y_{n+k})$$

重みつき移動平均

$$t_n = \alpha_{-k} y_{n-k} + \cdots + \alpha_0 y_n + \cdots + \alpha_k y_{n+k}$$

α_i 重み係数



重み付き移動平均の性質

$y_n = T_n + w_n$, $T_n = a + bn$, $w_n \sim N(0, \sigma^2)$ の場合

平均：不変 分散： $\sigma^2 \sum_{j=-k}^k \alpha_j^2$

$$t_n = \sum_{j=-k}^k \alpha_j T_{n+j} + \sum_{j=-k}^k \alpha_j w_{n+j} = T_n + \sum_{j=-k}^k \alpha_j w_{n+j}$$

$$E\left(\sum_{j=-k}^k \alpha_j w_{n+j}\right) = \sum_{j=-k}^k \alpha_j E(w_{n+j}) = 0$$

$$E\left(\sum_{j=-k}^k \alpha_j w_{n+j}\right)^2 = \sum_{j=-k}^k \alpha_j^2 E(w_{n+j}^2) + \sum_{j=-k}^k \sum_{i \neq j} \alpha_j \alpha_i E(w_{n+j} w_{n+i}) = \sum_{j=-k}^k \alpha_j^2 \sigma^2$$

$$E(w_n^2) = \sigma^2 \quad E(w_n w_m) = 0 \quad (n \neq m)$$

移動メディアアン・フィルタ



$$t_n = \text{median}(y_{n-k}, \dots, y_n, \dots, y_{n+k})$$

$y_{(n-k)}, \dots, y_{(n+k)}$: y_{n-k}, \dots, y_{n+k} を小さい順に並べたもの (順序統計量)

$$\text{median}(y_1, \dots, y_n) = \begin{cases} y_{(\frac{n+1}{2})} & n : \text{奇数} \\ \frac{1}{2} \left(y_{(\frac{n}{2}-1)} + y_{(\frac{n}{2}+1)} \right) & n : \text{偶数} \end{cases}$$

長所 : 異常値, 急激な構造変化に対応できる

移動平均フィルタと移動メディアンフィルタ

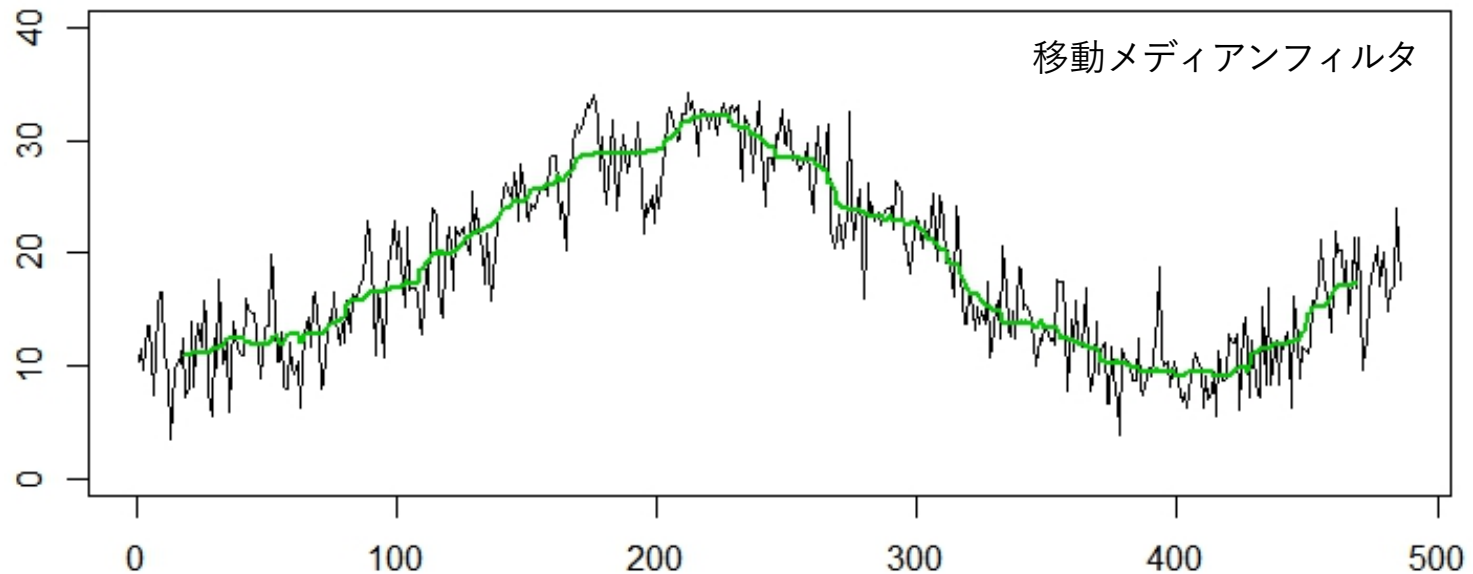
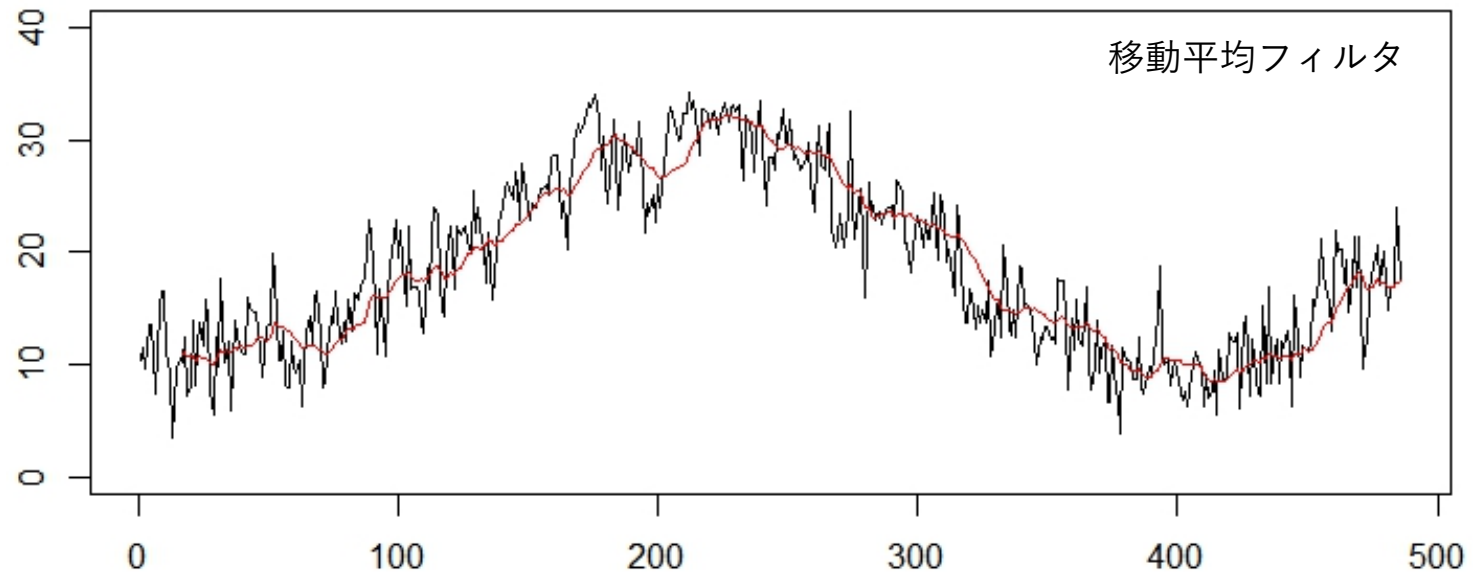
```
# 移動平均フィルタ
```

```
plot(maxtemp,ylim=c(0,40))
y <- maxtemp
ndata <- length(maxtemp)
y[1:ndata] <- NA
kfilter <- 17
n0 <- kfilter+1
n1 <- ndata-kfilter
for(i in n0:n1){
  i0 <- i-kfilter
  i1 <- i+kfilter
  y[i] <- mean(maxtemp[i0:i1])
}
lines(y,col=2,ylim=c(0,40))
```

```
# 移動メディアンフィルタ
```

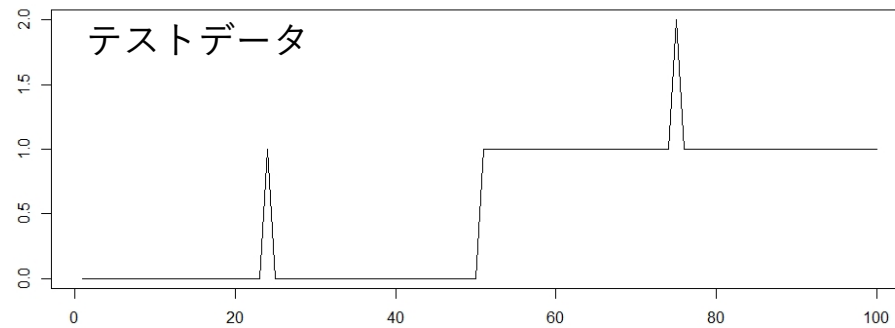
```
plot(maxtemp,ylim=c(0,40))
y <- maxtemp
ndata <- length(maxtemp)
y[1:ndata] <- NA
kfilter <- 17
n0 <- kfilter+1
n1 <- ndata-kfilter
for(i in n0:n1){
  i0 <- i-kfilter
  i1 <- i+kfilter
  y[i] <- median(maxtemp[i0:i1])
}
lines(y,col=3,ylim=c(0,40),lwd=2)
```

maxtemp (最高気温) データ

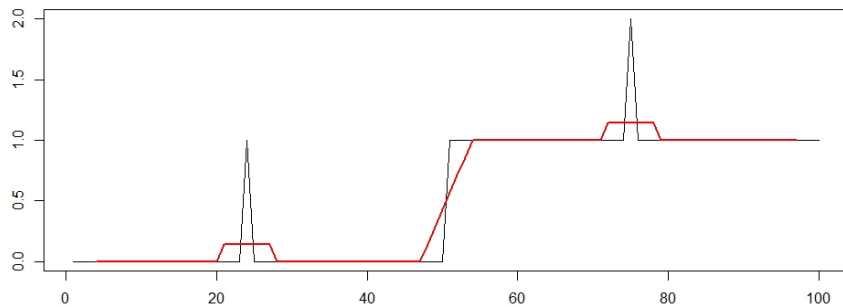


MAとMMの比較（ノイズなしの場合）

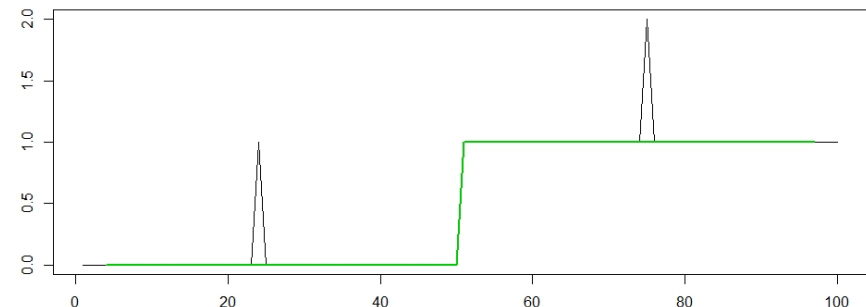
```
z<-rep(0,100)
z[51:100]<-1
z[25]<-1
z[75]<-2
z<-as.ts(z)
plot(z)
```



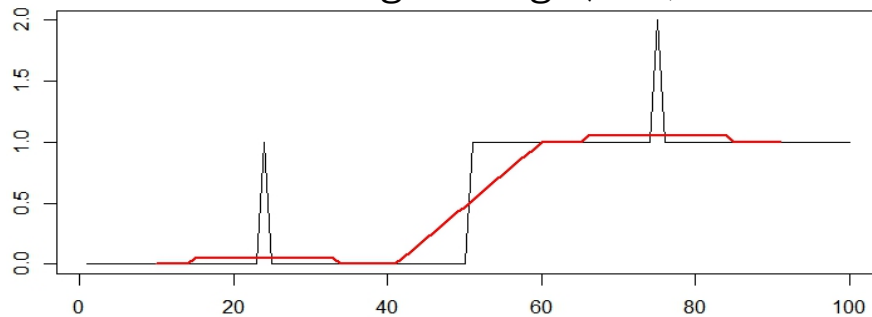
Moving-average($k=3$)



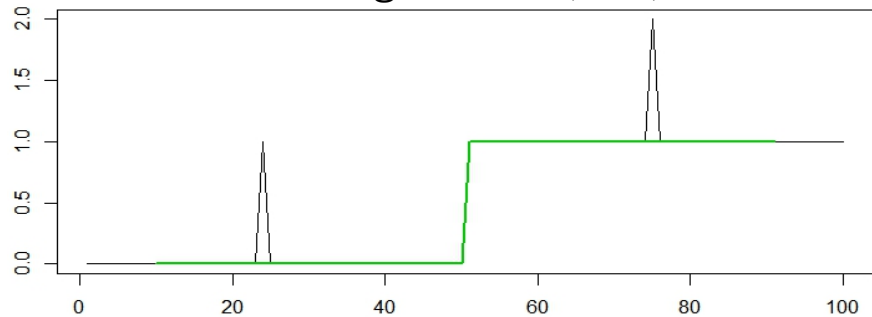
Moving-median($k=3$)



Moving-average($k=9$)

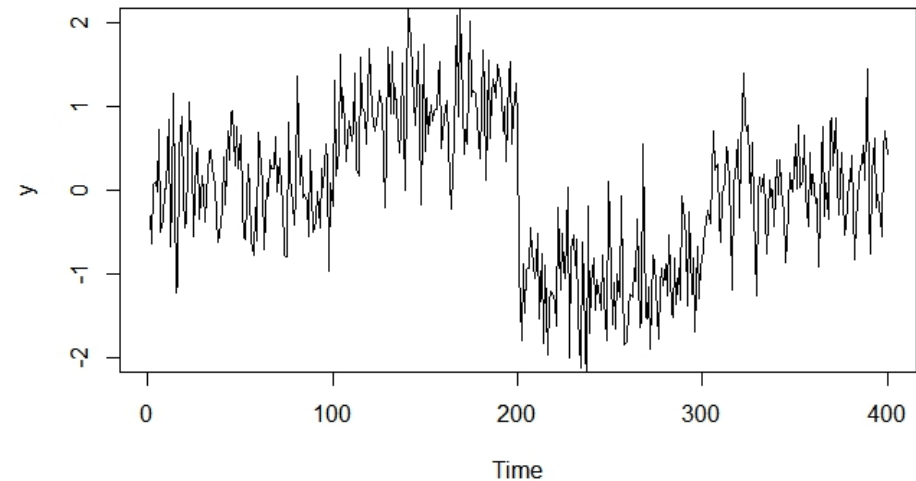


Moving-median($k=9$)

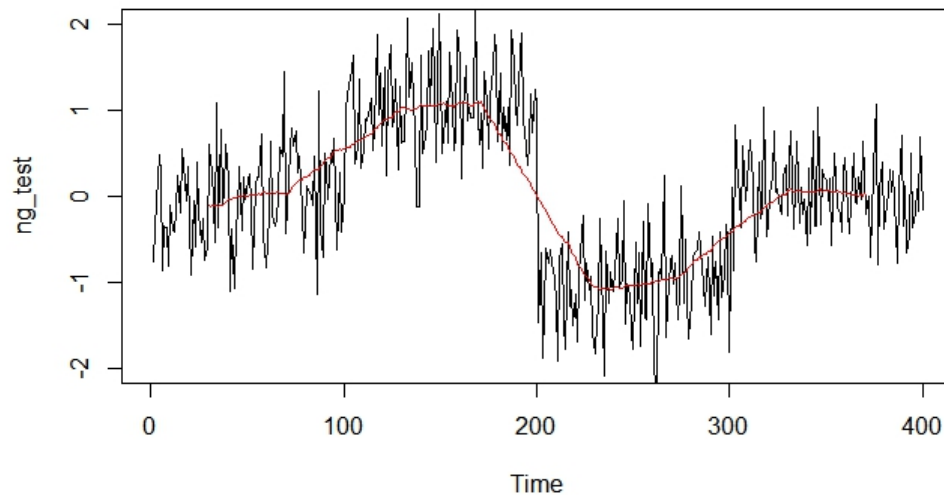


移動平均フィルタと移動メディアンフィルタ

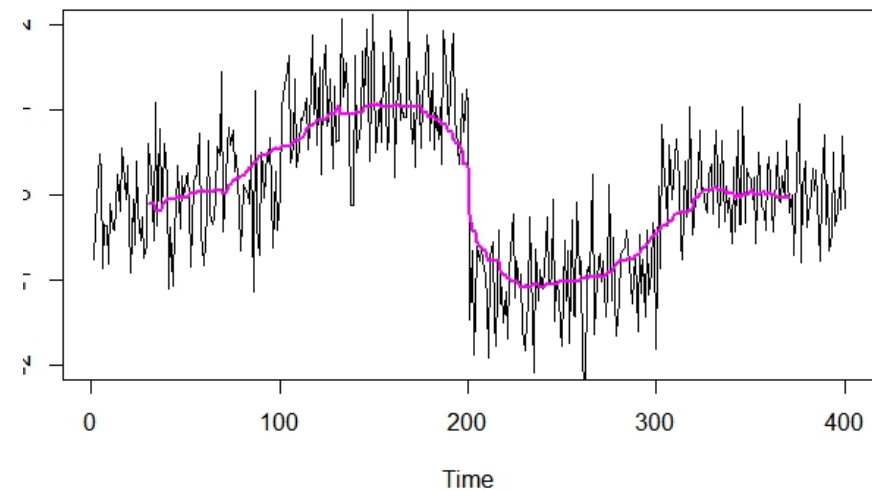
```
x <- rep(0,400)
x[101:200] <- 1
x[201:300] <- -1
y <- x + rnorm(400, mean=0, sd=0.5)
ng_test <- as.ts(y)
plot(ng_test)
```



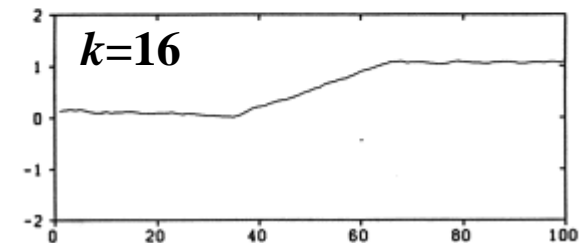
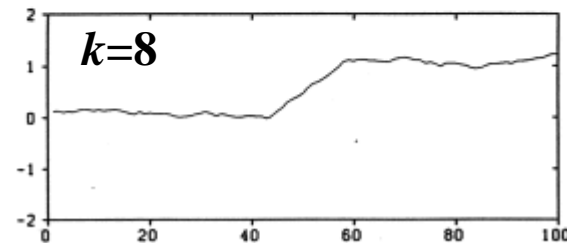
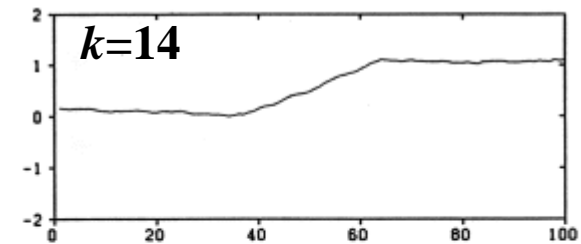
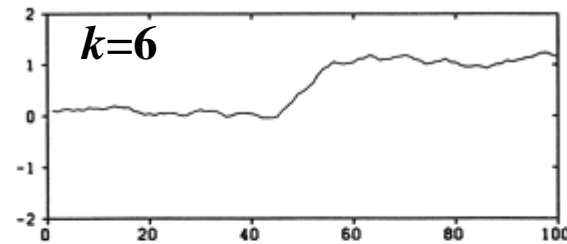
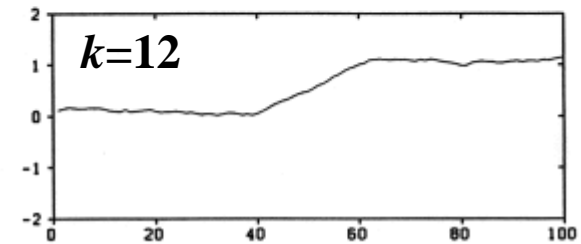
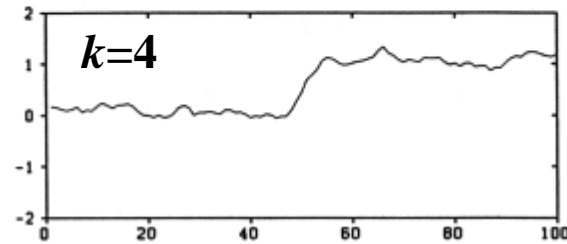
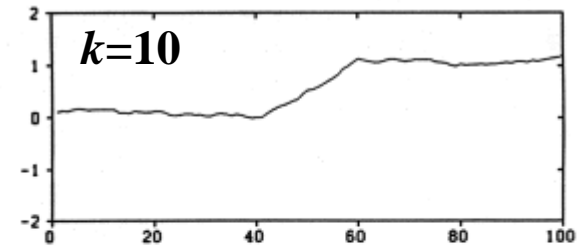
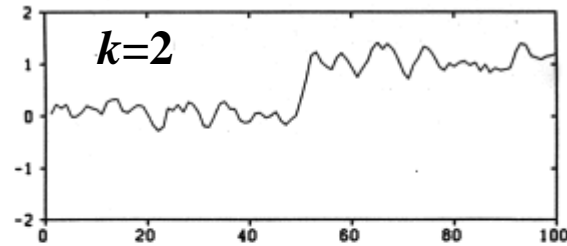
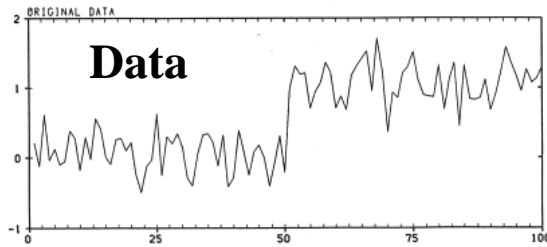
Moving-average(29)



Moving-median(29)



移動平均フィルタ



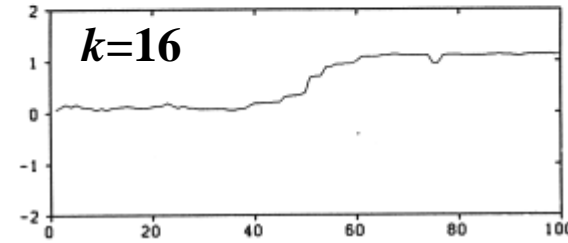
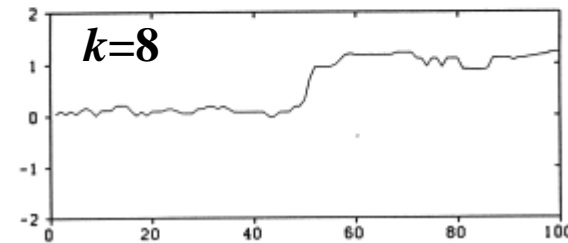
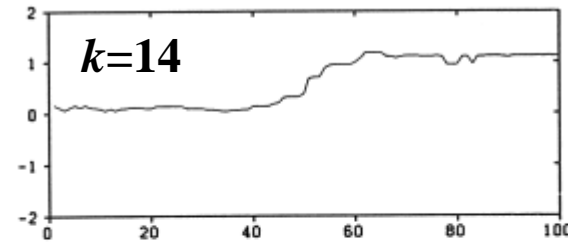
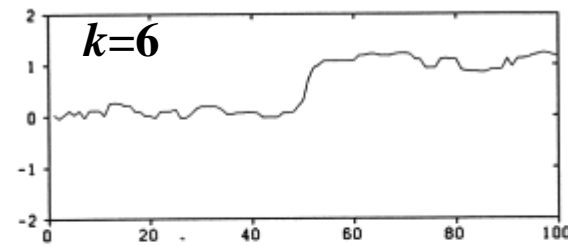
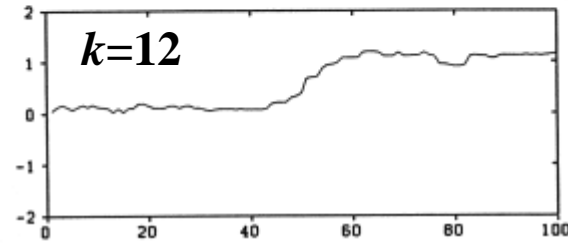
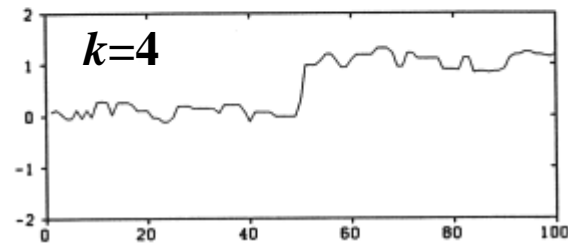
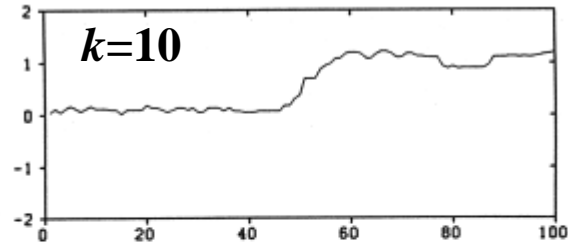
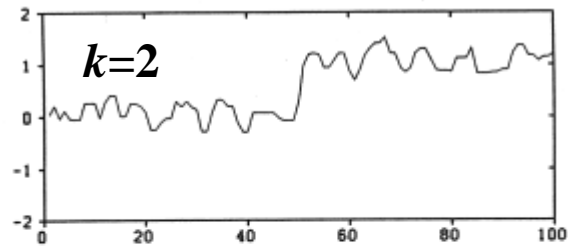
長所

- 滑らかな推定値

欠点

- 構造変化を正確に検出できない
- 異常値に敏感

移動メディアン



長所

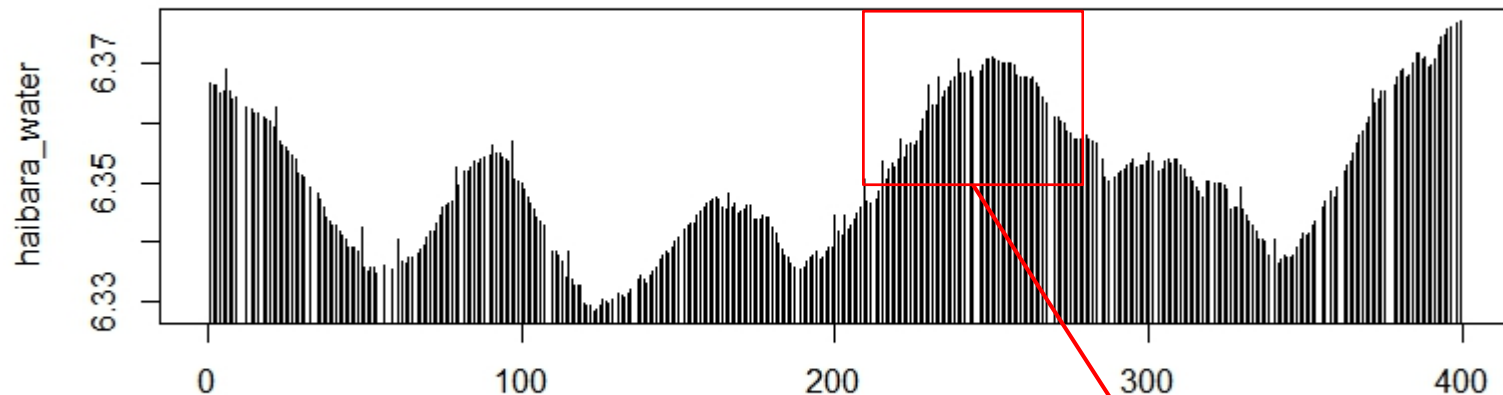
- 構造変化を検出できる
- 異常値に頑健

欠点

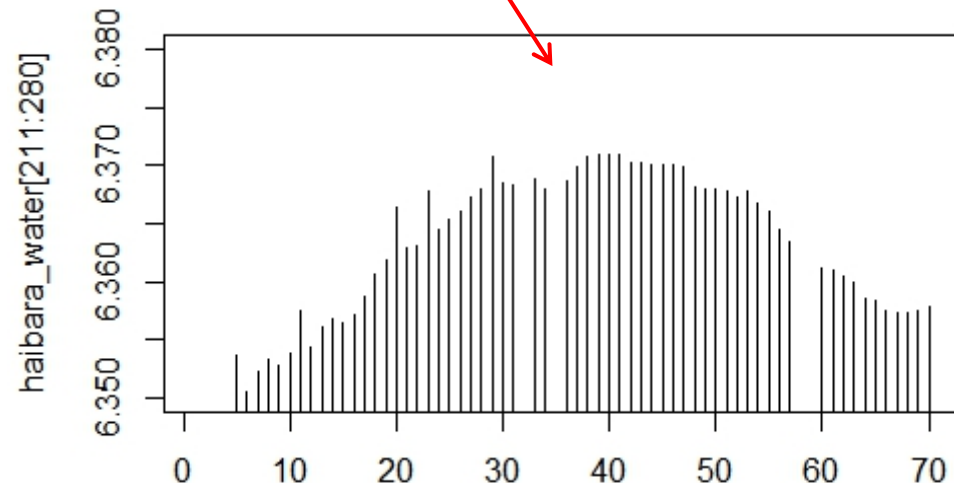
- 変動が大きい

欠測値と異常値（外れ値）

```
# 榛原 地下水位データ  
haibara <- as.ts(read.csv("haibara_new.csv"))  
haibara_water <- haibara[,2]  
plot(haibara_water,type="h")
```



```
plot(haibara_water[211:280],type="h",ylim=c(6.35,6.38))
```



処理方法は状態空間
モデリングのところで