時系列分析
ハンズオン#1

進行予定

第1回「時系列データをみる」

・時系列分析の考え方

・時系列データの構成要素

（自己相関，周期的変動，トレンド，外因性，ホワイトノイズ）

・定常過程とは何か，定常過程だと何がうれしいのか

・時系列データの変換

・時系列データをみる[ハンズオン]

・ホワイトノイズ，ランダムウォークなどを書いてみる[ハンズオン]

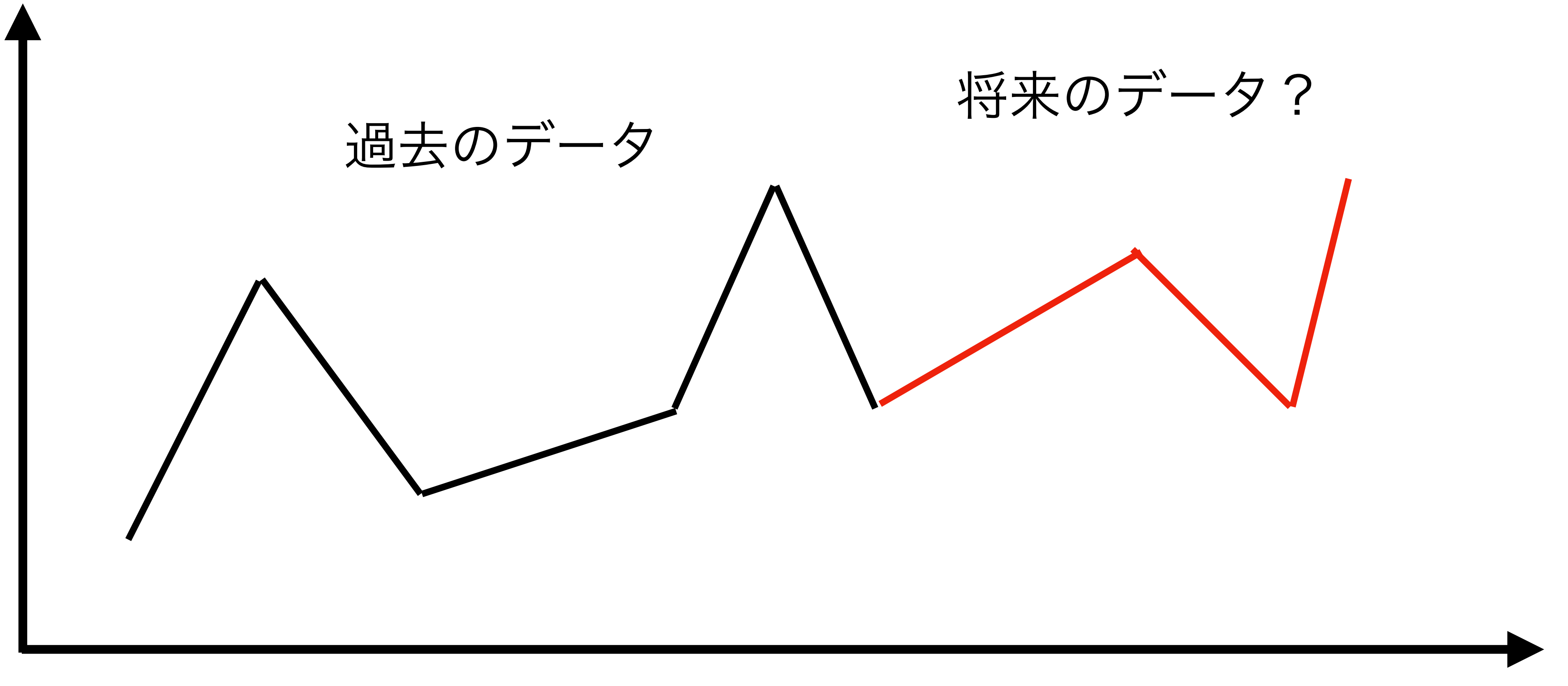
講師自己紹介

- 氏名：大久保　亮介
- 現在薬学部4年
- 研究予定例：シミュレーションによる漢方薬理の推定，時空間統計解析による「証」の分布推定　など
- 担当講義：基礎統計→CNN，高校数学

時系列分析の考え方

時系列分析とは？

過去のデータから,
将来のデータ変化を
予測する

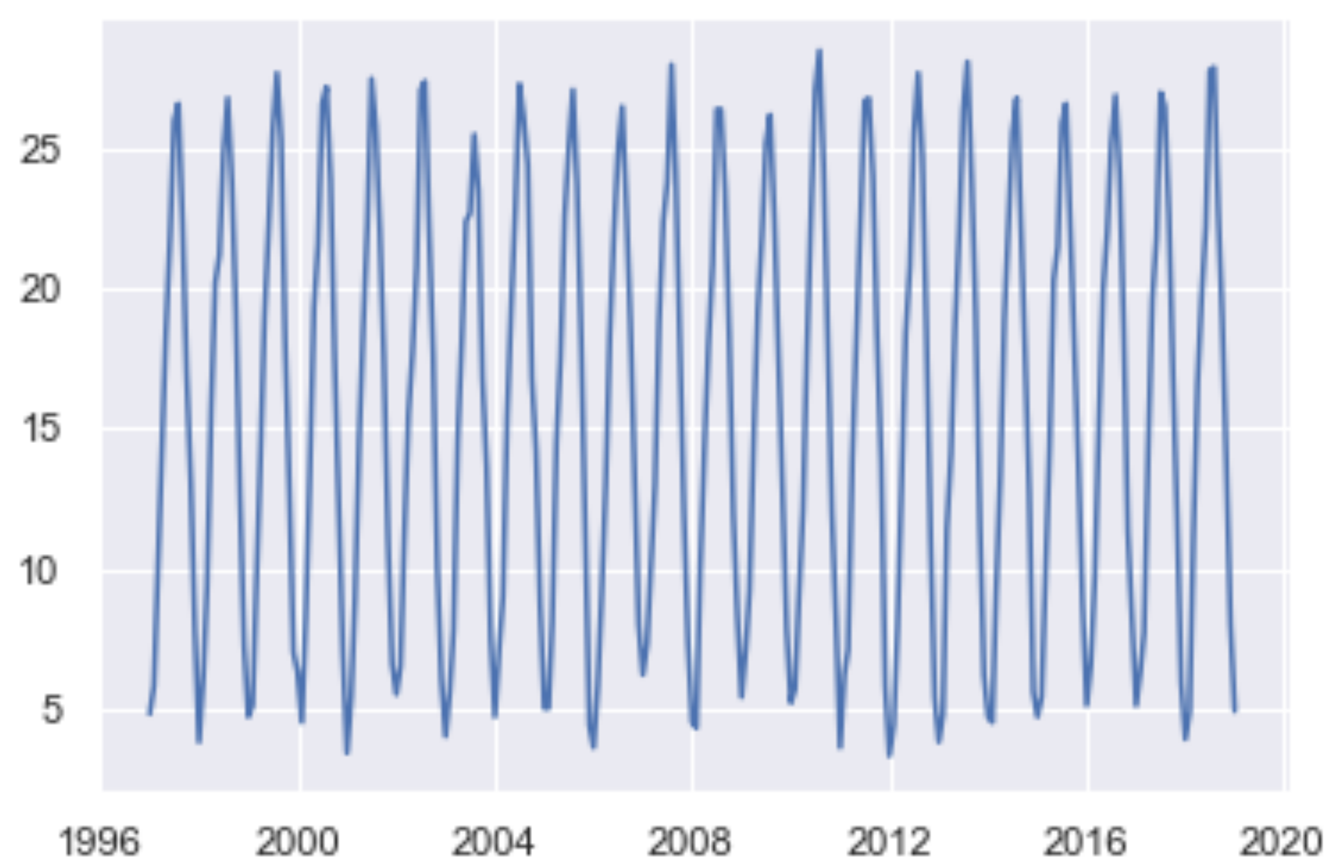


時系列データ

例：海老名市の平均気温

ebina_tuki

Month	kion
1997/1	4.8
1997/2	5.8
1997/3	9.7
1997/4	14.3
1997/5	18.5
1997/6	21.9
1997/7	26
1997/8	26.6
1997/9	22.3
1997/10	17
1997/11	13.1
1997/12	7.8
1998/1	3.8
1998/2	6.1
1998/3	9.3
1998/4	15.7
1998/5	20.2
1998/6	21.2

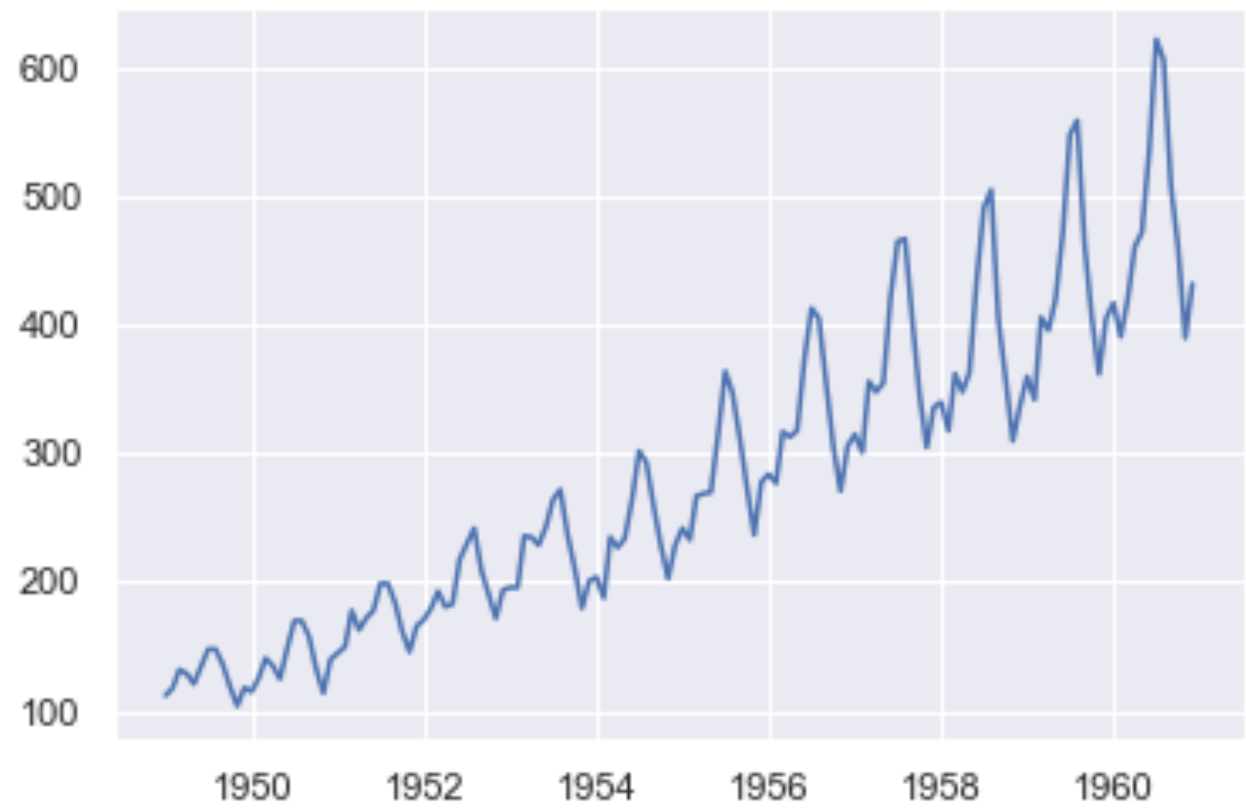


時系列データ

例：AirPassengers.csv

AirPassengers

Month	#Passengers
1949-01	112
1949-02	118
1949-03	132
1949-04	129
1949-05	121
1949-06	135
1949-07	148
1949-08	148
1949-09	136
1949-10	119
1949-11	104
1949-12	118
1950-01	115
1950-02	126
1950-03	141
1950-04	135
1950-05	125
1950-06	149



時系列データの特徴

観測される順序に
意味がある

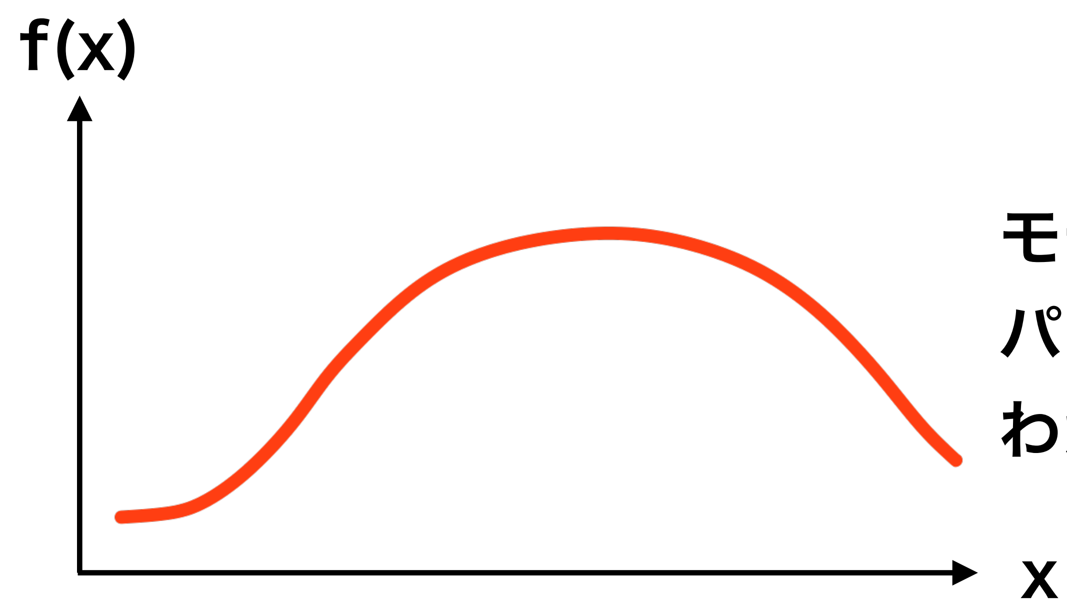
1日のデータは,
1日に一回しか手に入らない

→何が母集団？

確率分布

確率分布からの抽出

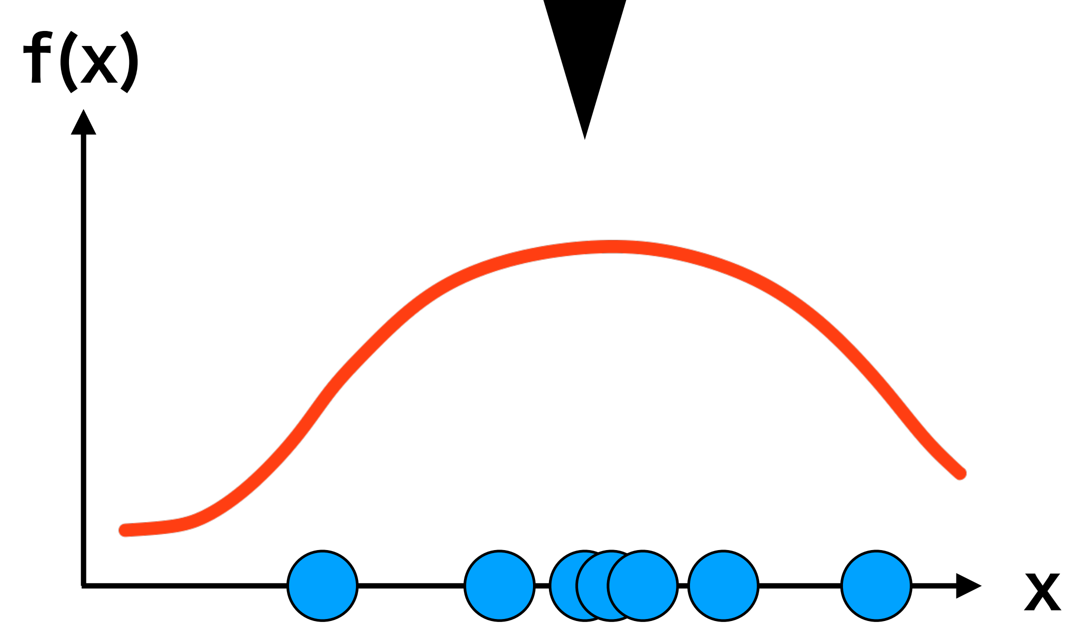
母集団の想定



モデルの形,
パラメータが
わかっている

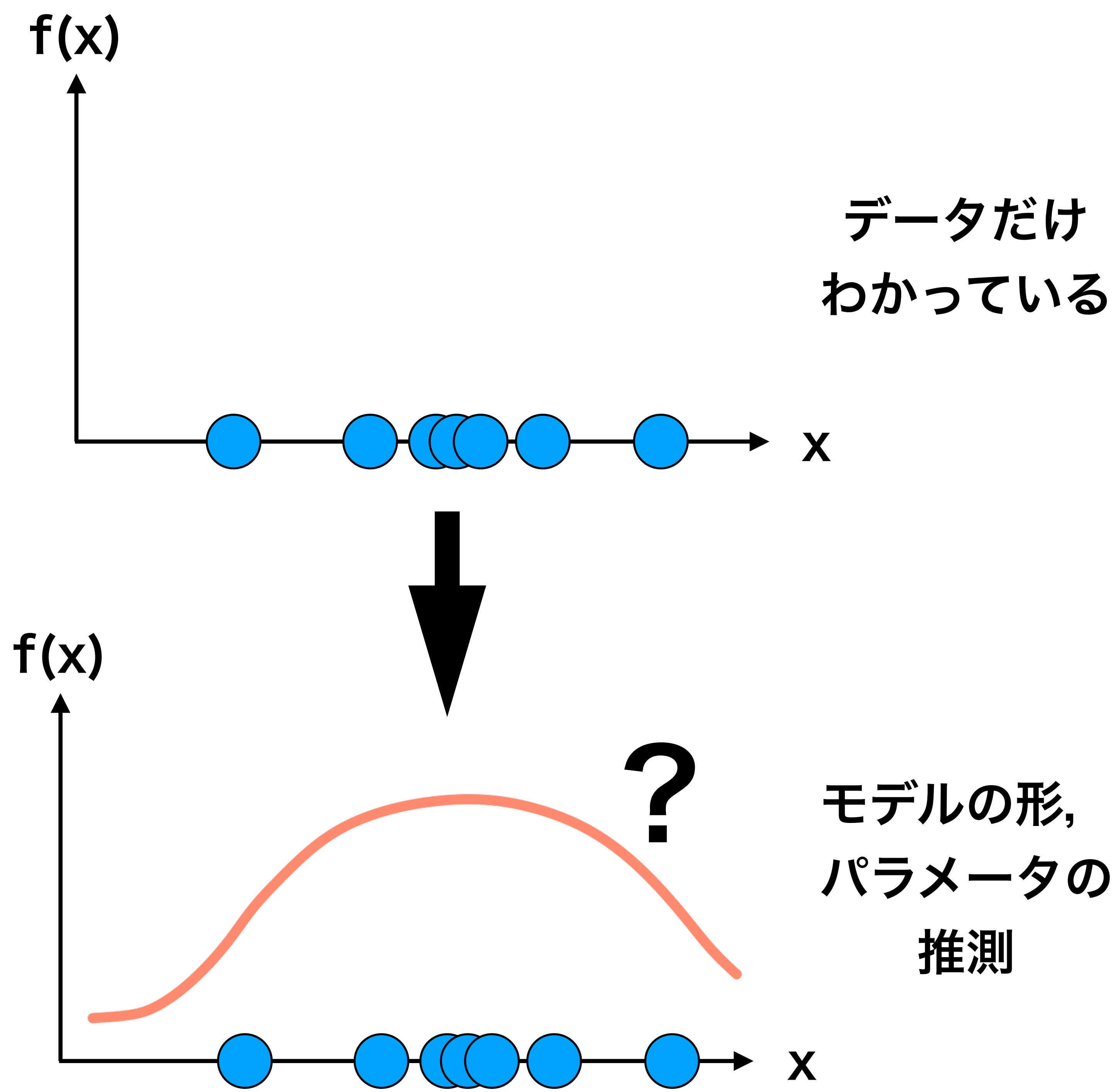
$$f(x\,|\,\theta)$$

標本の抽出



データの
発生

モデリングの考え方



時系列データの構成要素

基本的なパラメータ

もしt地点が無数に
あったときの平均と考える

期待値

$$\mu_t = E(y_t)$$

t地点におけるデータの期待値

分散

$$V(y_t) = E[(y_t - \mu_t)^2]$$

t地点におけるデータの分散

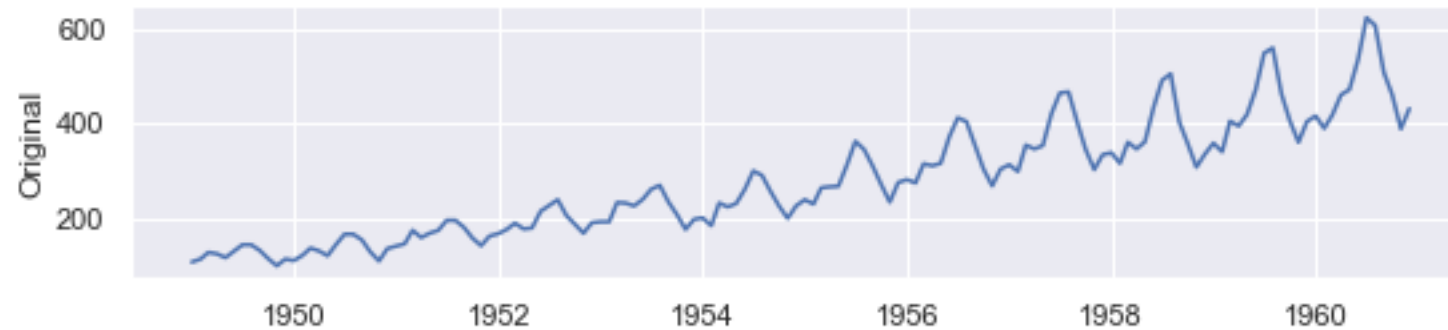
標準偏差（ボラティリティ）：
分散の平方根

構成要素

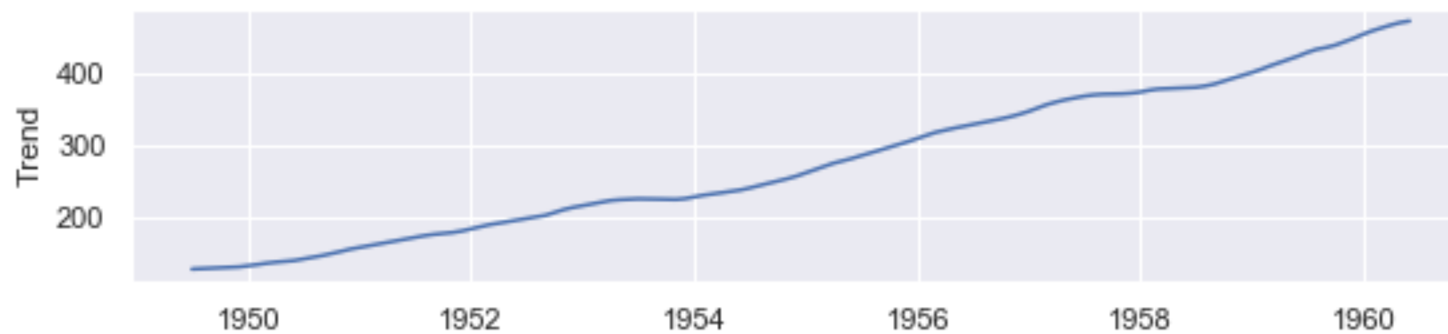
- 自己相関
- 周期的変動
- トレンド
- 外因性
- ホワイトノイズ

構成要素ごとに分解

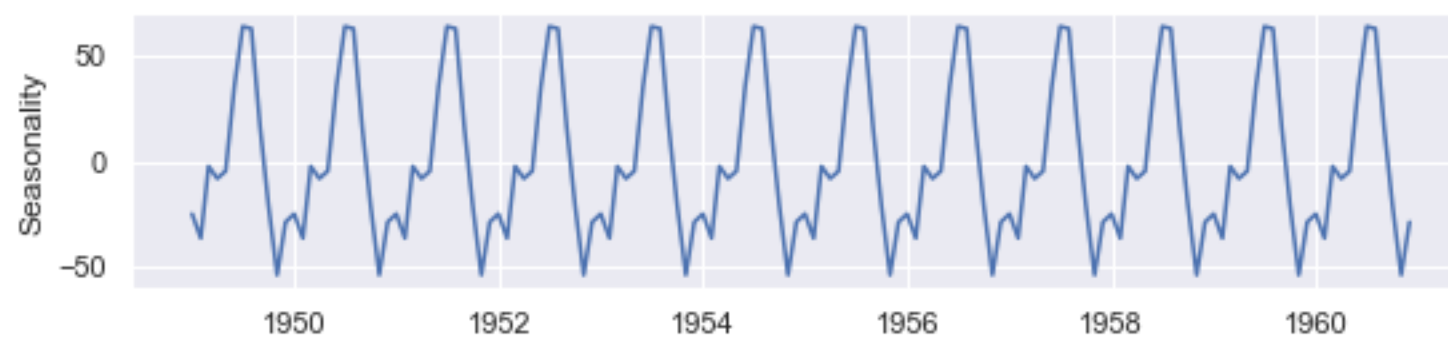
もとのグラフ



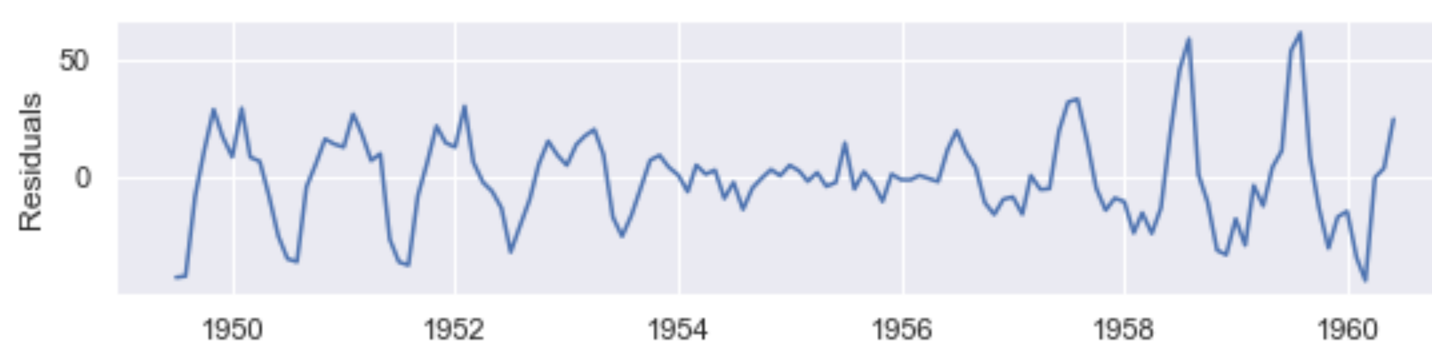
トレンド



季節性

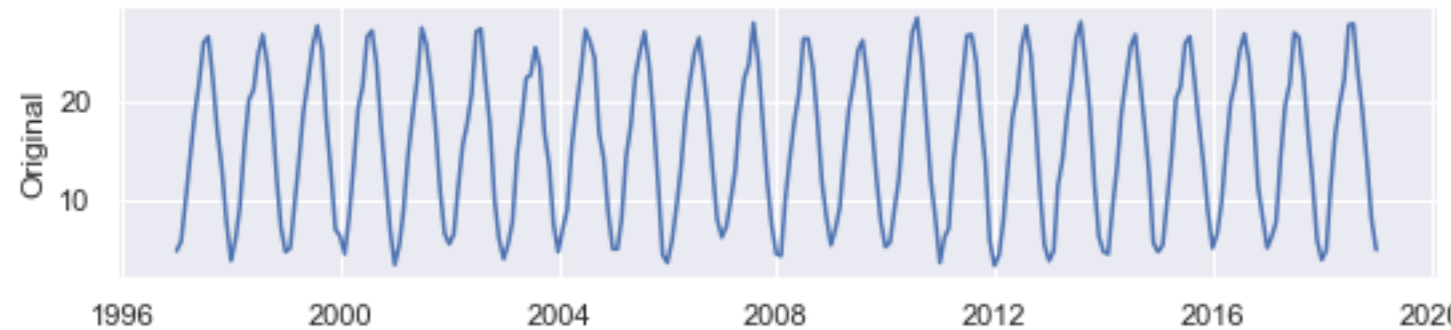


ホワイトノイズ



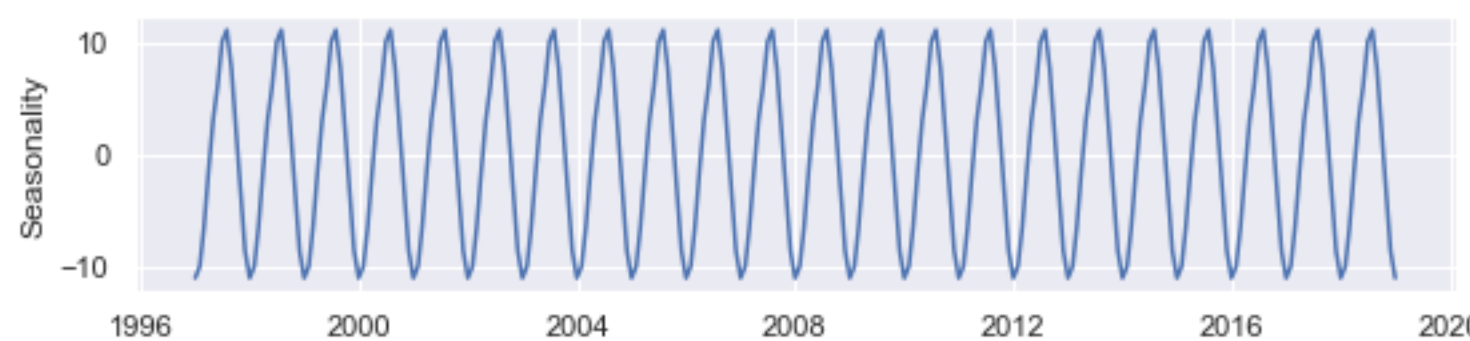
構成要素ごとに分解

もとのグラフ

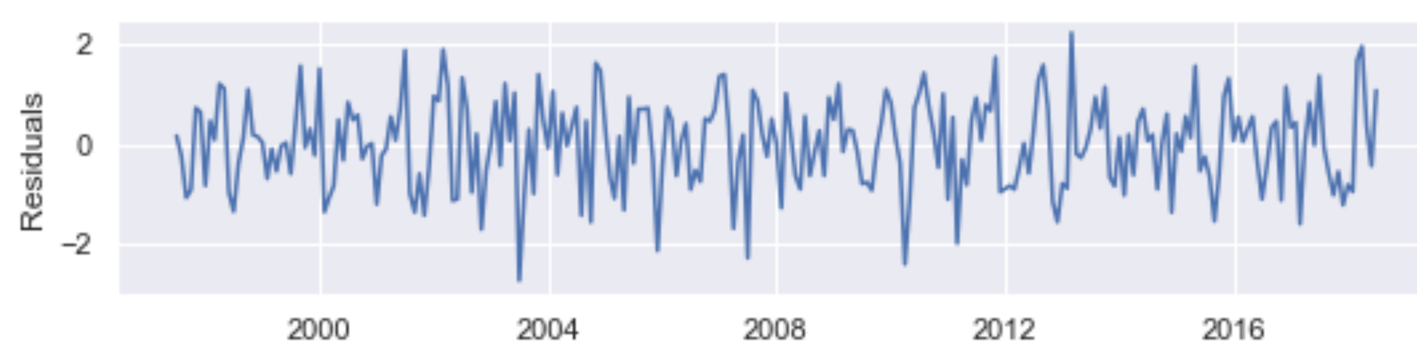


トレンド

季節性



ホワイトノイズ



自己相関

データの前後に，関係がある
→データ同士が独立でない

- 正の相関：昨日が高いと，今日も高い
- 負の相関：昨日が高いと，今日は低い

コレログラム：
何地点前と自己相関を起こしているか

パラメータ

自己共分散

$$\gamma_{kt} = Cov(y_t, y_{t-k})$$ k次の自己共分散

$$= E[(y_t - \mu_t)(y_{t-k} - \mu_{t-k})]$$

自己相関

$$\rho_{kt} = Corr(y_t, y_{t-k})$$ k次の自己相関

→t地点のデータと
t-k地点のデータとの
比較

$$= \frac{Cov(y_t, y_{t-k})}{\sqrt{V(y_t)V(y_{t-k})}}$$

パラメータ

偏自己相関

$$P_{tk} = \frac{Cov(y_t - \hat{y}_t, y_{t-k} - \widehat{y_{t-k}})}{\sqrt{V(y_t - \hat{y}_t)V(y_{t-k} - \widehat{y_{t-k}})}}$$

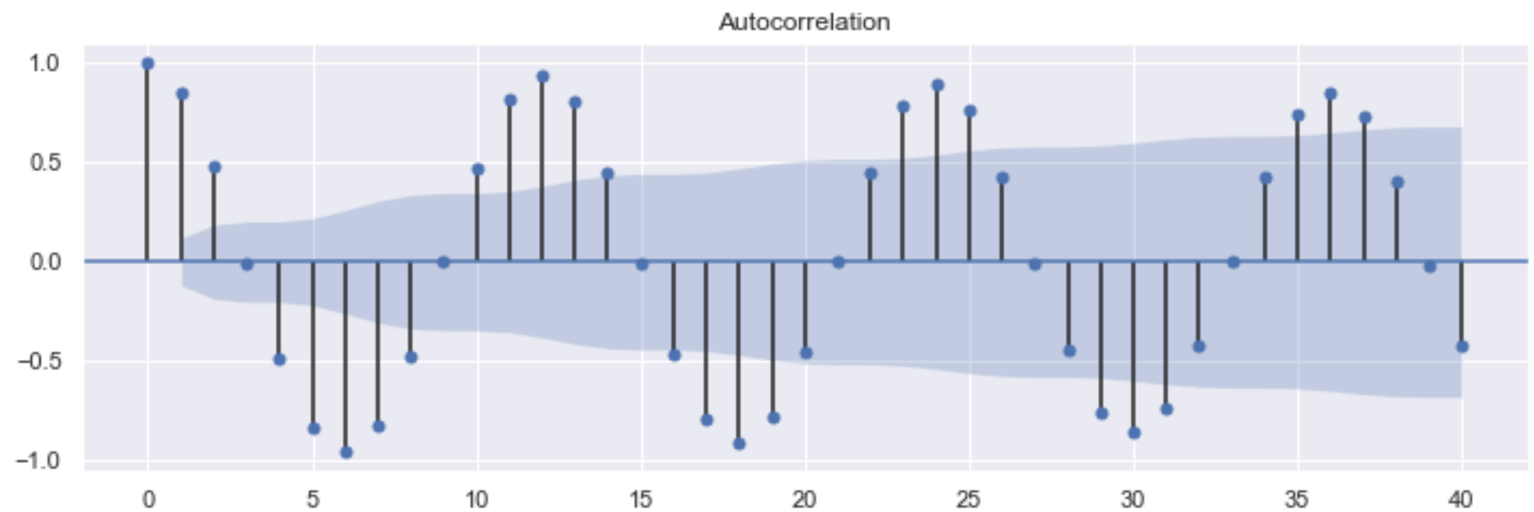
k次の偏自己相関

→k-1地点までの影響が除かれた自己相関

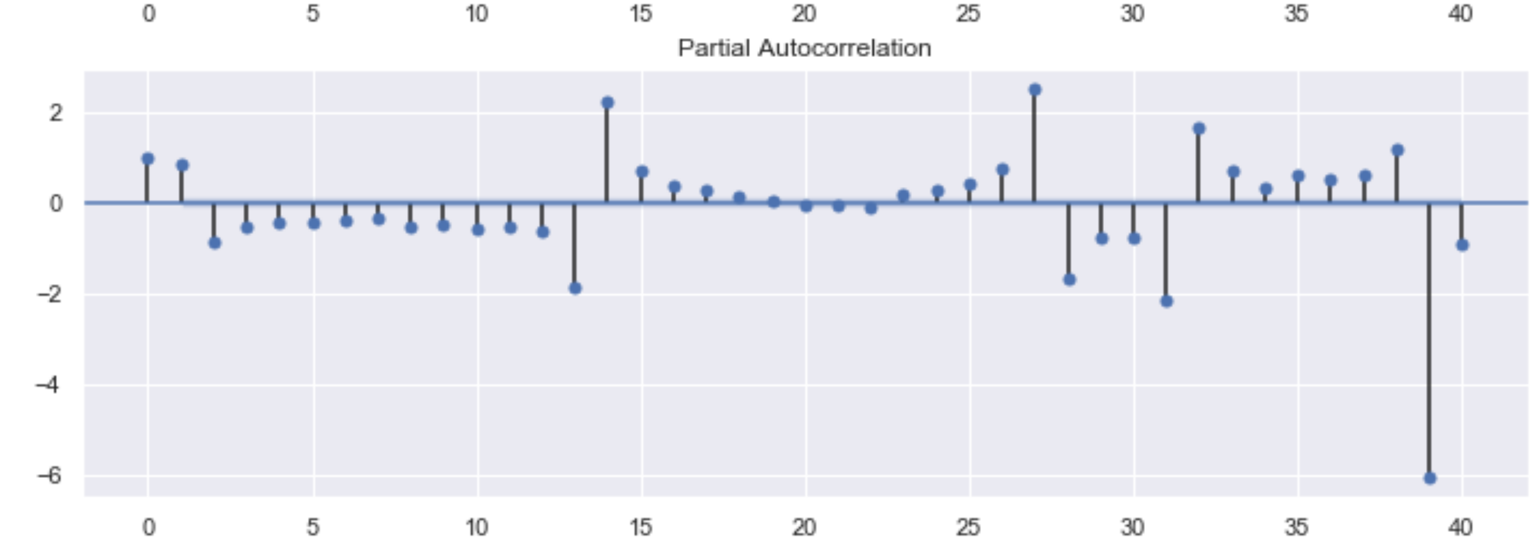
コレログラムの例

例：海老名市の平均気温

自己相関



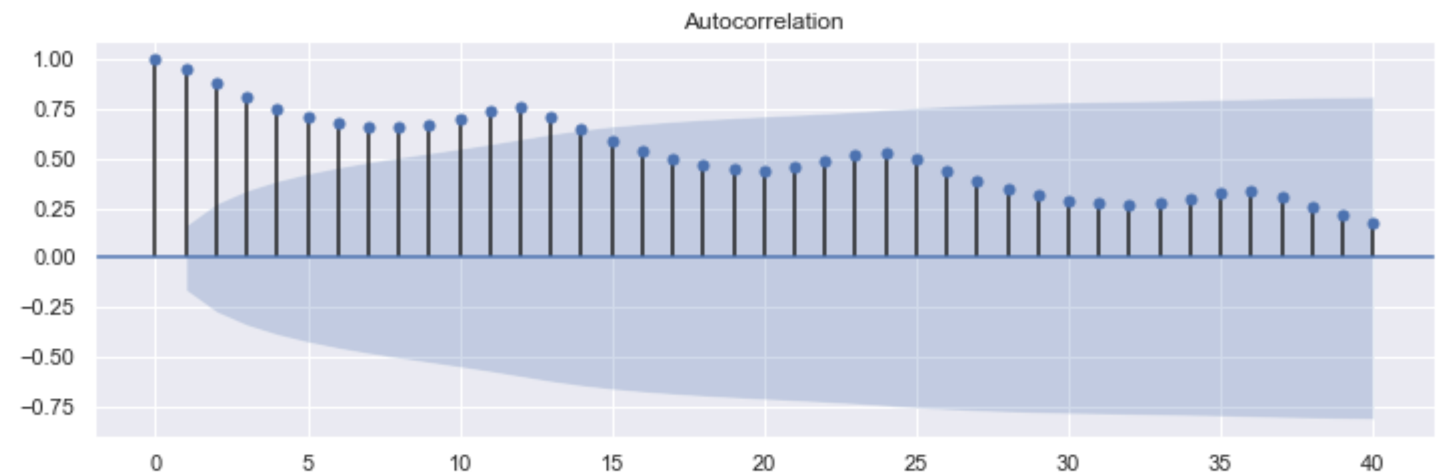
偏自己相関



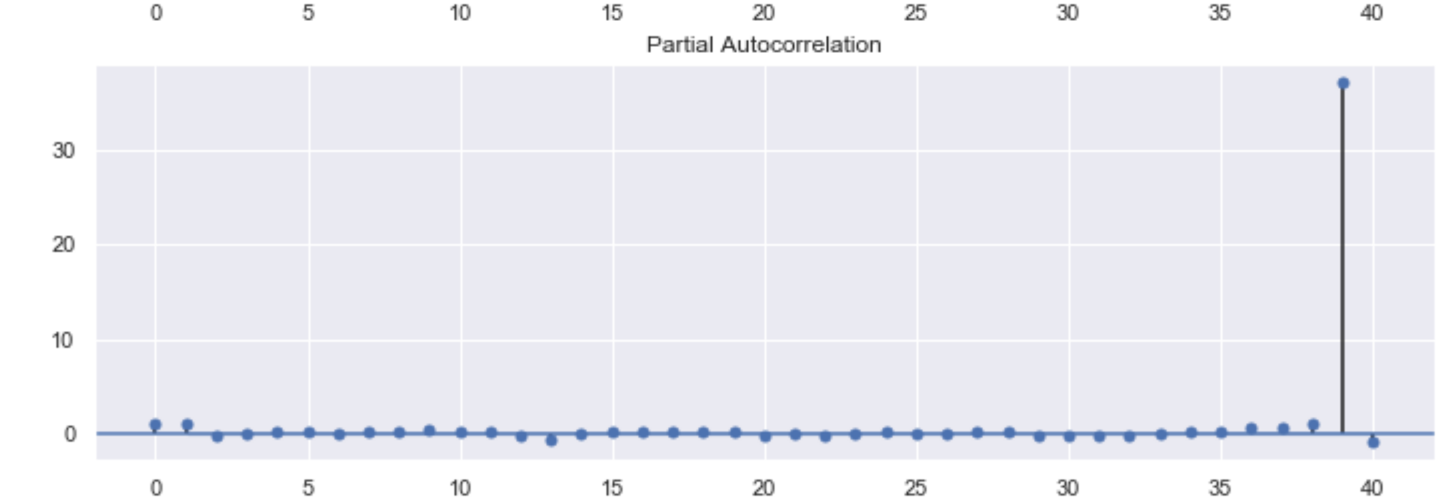
コレログラムの例

例：AirPassengers.csv

自己相関



偏自己相関



周期的変動

・気温の変化
→夏は高く，冬は低い

トレンド

・売り上げが毎月右肩上がり
→正のトレンドがある

外因性

・近くでイベントがあって売り上げが上がった

ホワイトノイズ

将来を予測する情報がない雑音

t地点のホワイトノイズ： ϵ_t

- 期待値は0
- 分散は一定
- 自己相関が0

$$E(\epsilon_t) = 0$$

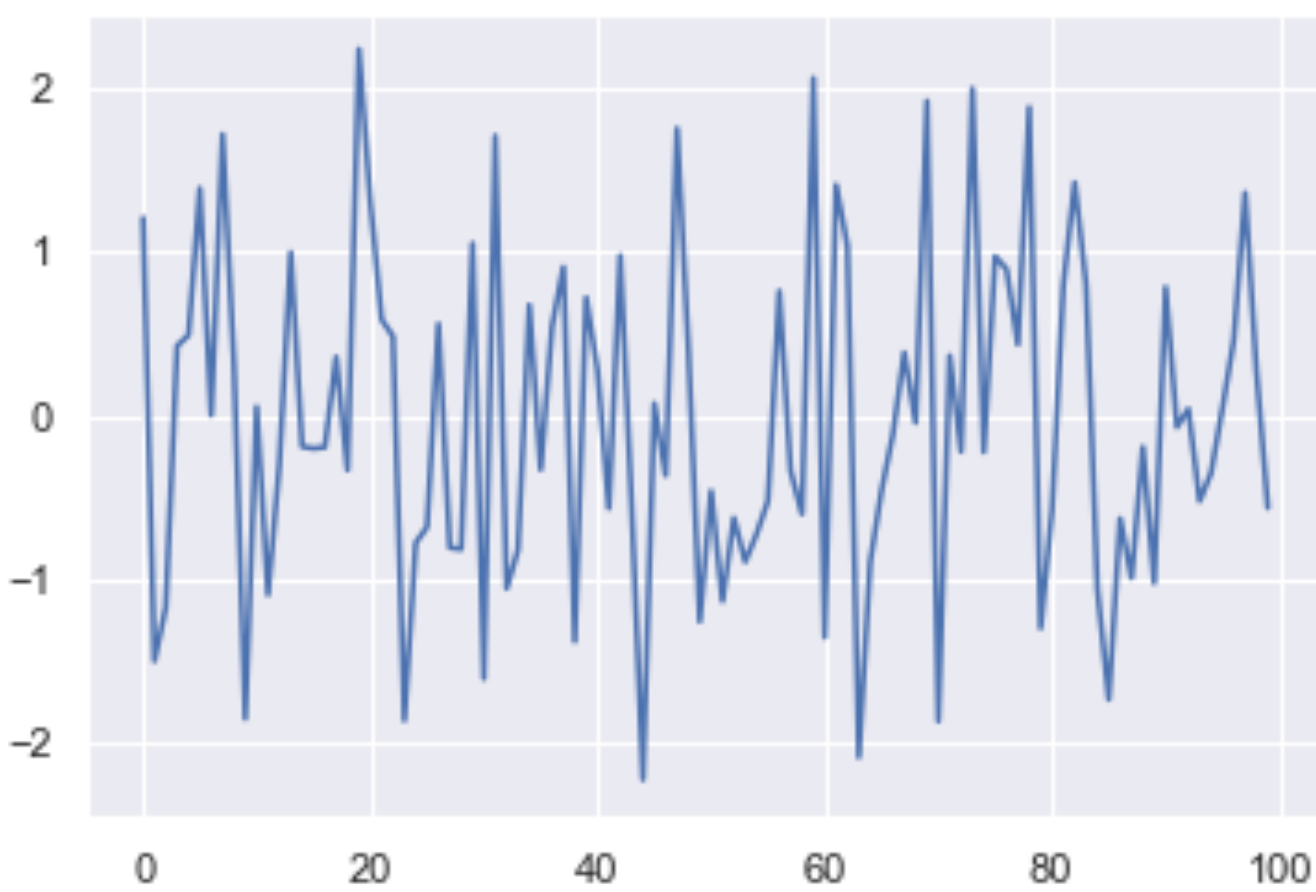
$$Cov(\epsilon_t, \epsilon_{t-k}) = \begin{cases} \sigma^2, & k = 0 \\ 0, & k \neq 0 \end{cases}$$

ホワイトノイズの例

実用的には正規分布に
したがうものを用いる

$$\epsilon_t \sim N(0, \sigma^2)$$

iid系列：
データが独立

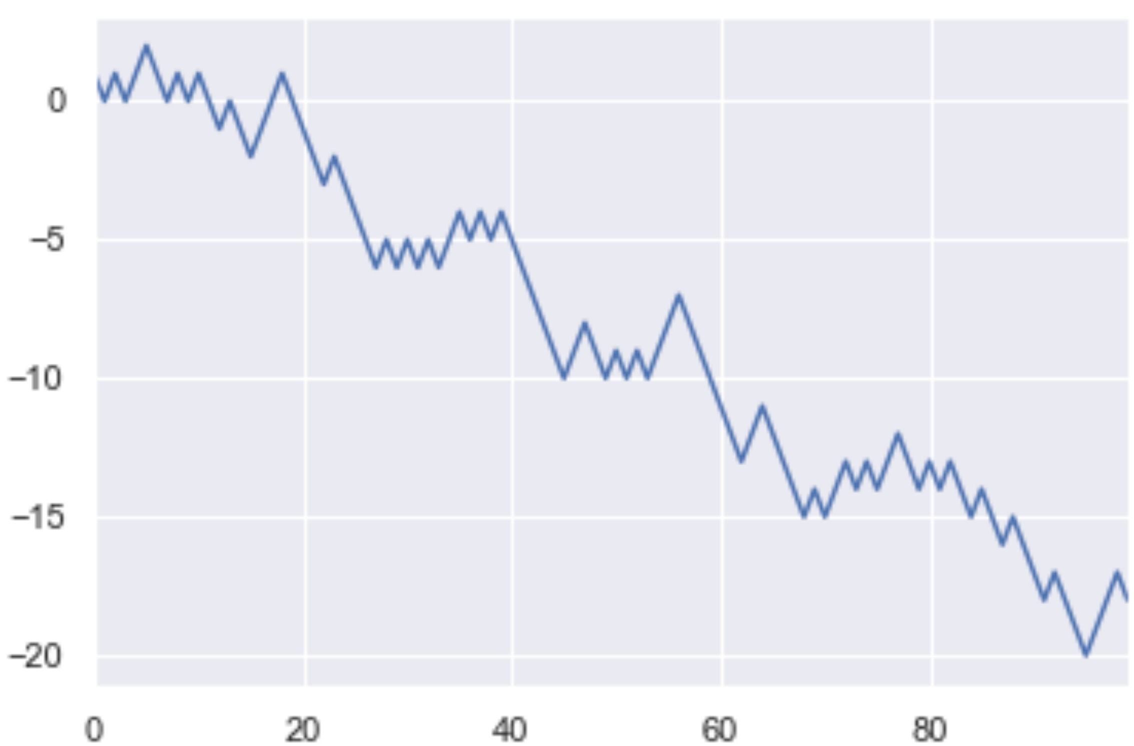


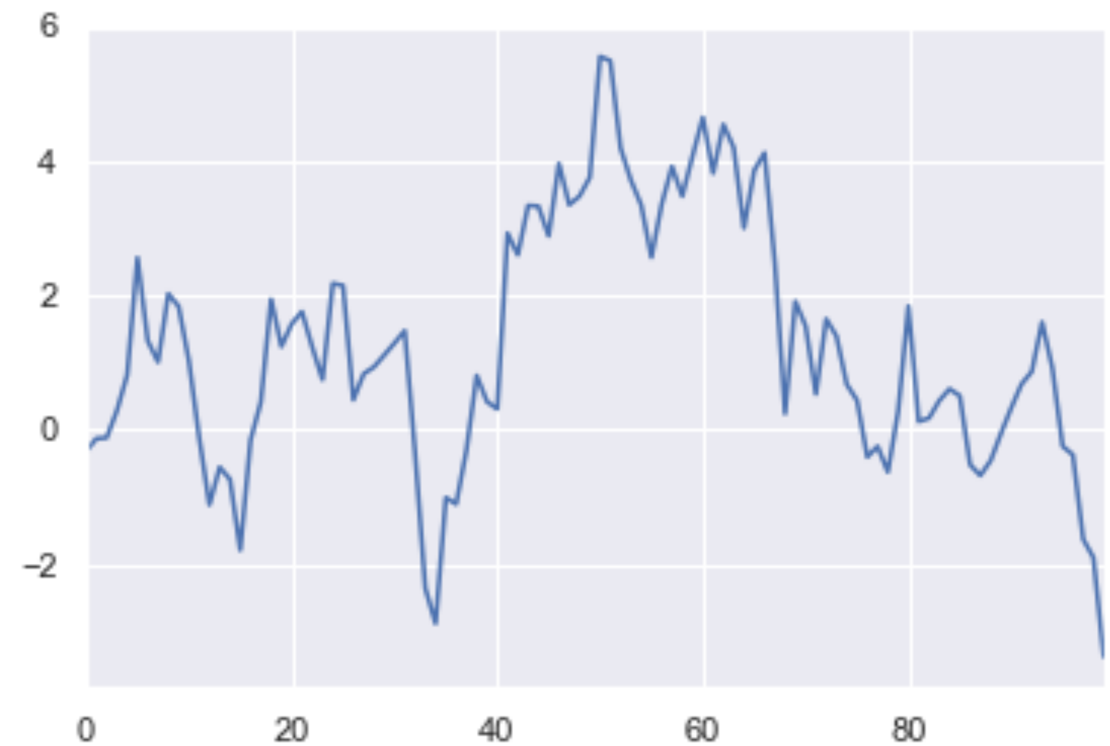
ランダムウォーク

iid系列の累積和

$$y_t = y_{t-1} + \epsilon_t$$

$$\epsilon_t = -1 \ or \ 1 \qquad \epsilon_t \sim N(0, \sigma^2)$$



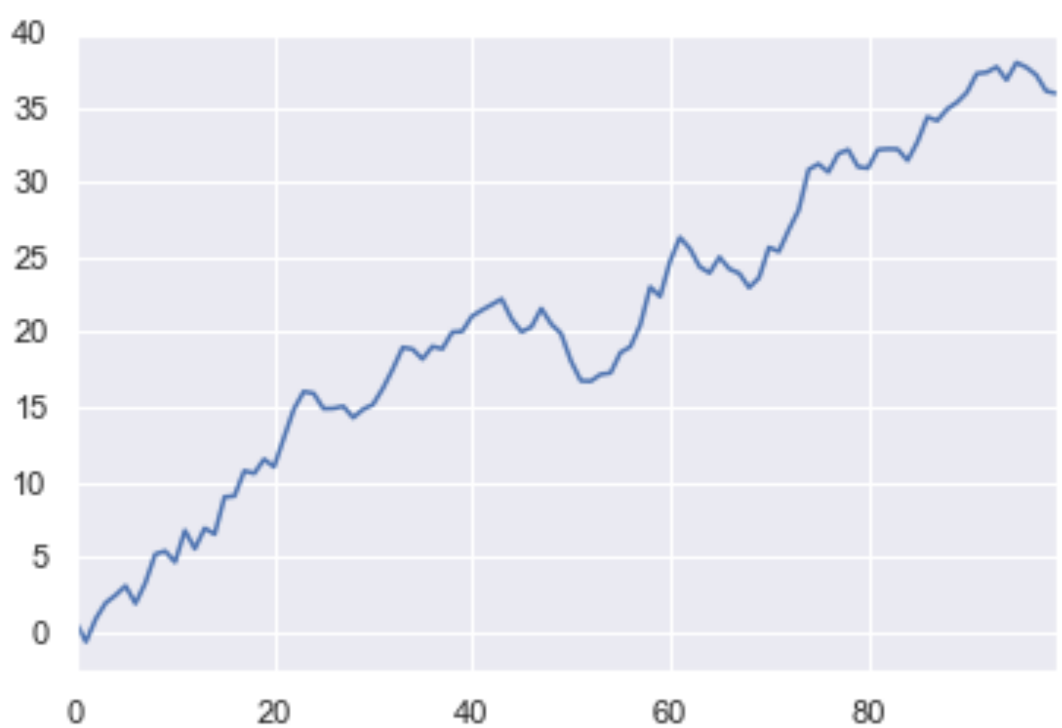


ランダムウォーク

ドリフト率　δ　→**トレンドを表せる**

$$y_t = \delta + y_{t-1} + \epsilon_t$$

$\delta = 0.5$



$\delta = -0.1$

定常過程とは何か,

定常過程だと何がうれしいのか

定常過程の定義

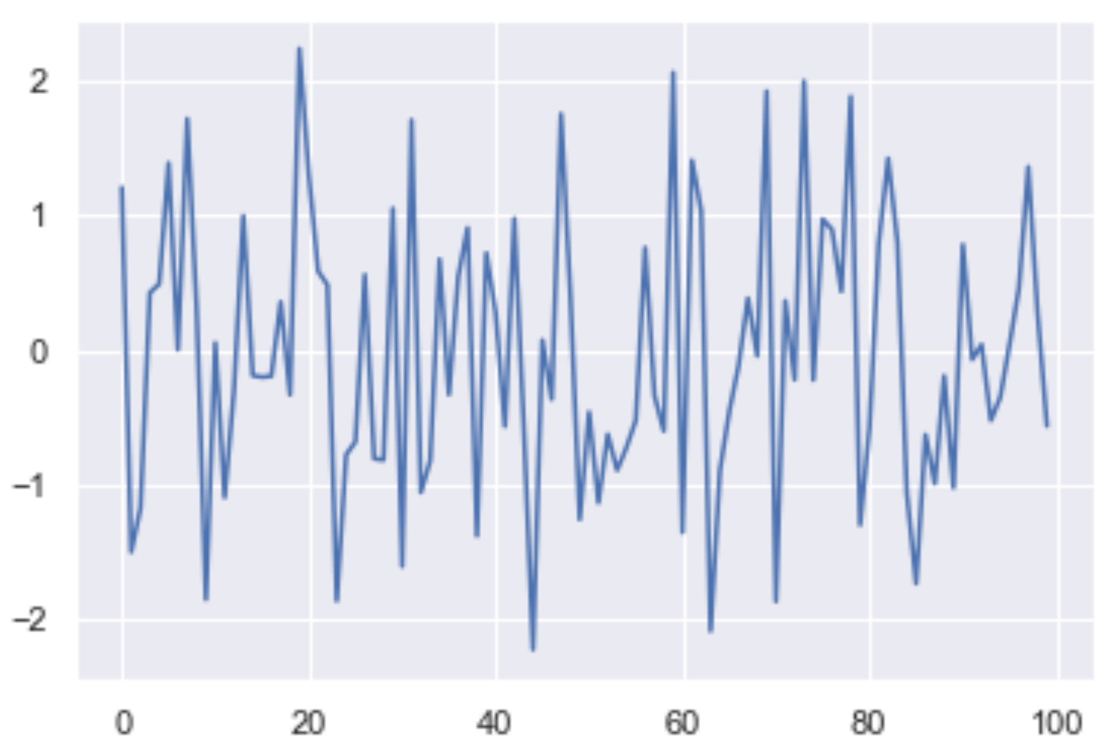
$\boxed{\text{期待値は時点によらず一定}}$ $E(y_t) = \mu$

$\boxed{\text{自己共分散は 時点によらず 時間差のみに依存}}$ $Cov(y_t, y_{t-k}) = E[(y_t - \mu)(y_{t-k} - \mu)]$

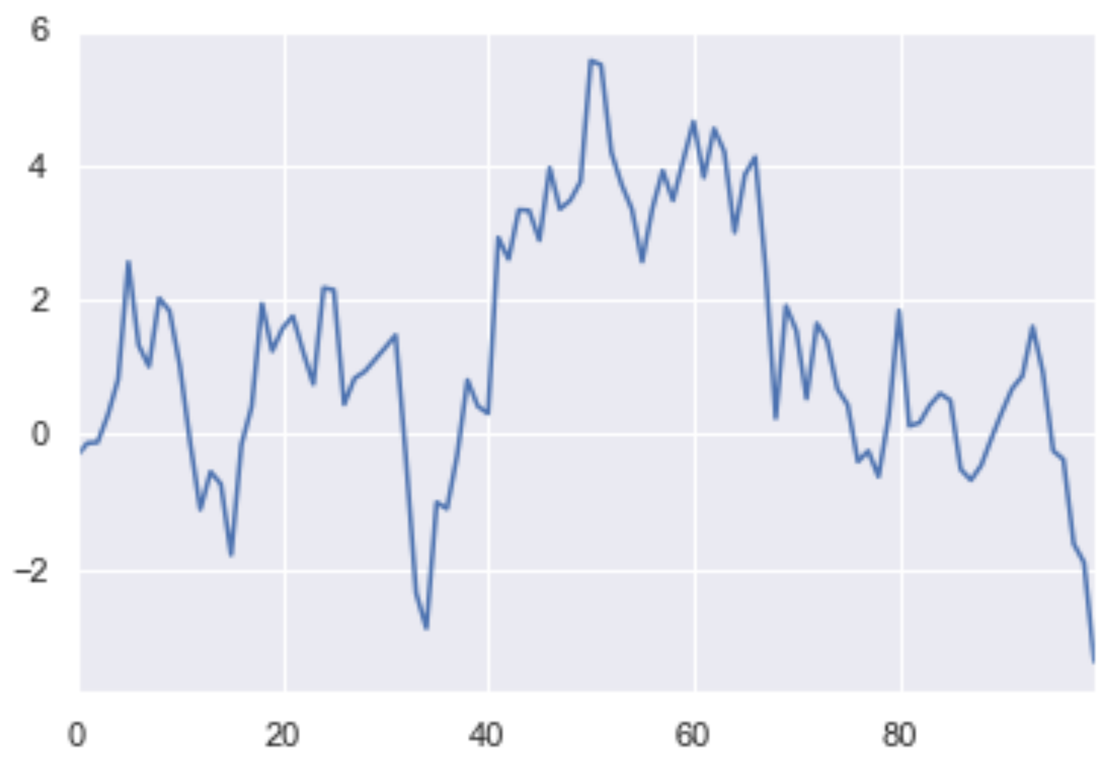
$$= \gamma_k$$

定常過程の例

定常過程



非定常過程



補足：定常過程の利点

パラメータが時点に依存せず一定

期待値 $$\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$$

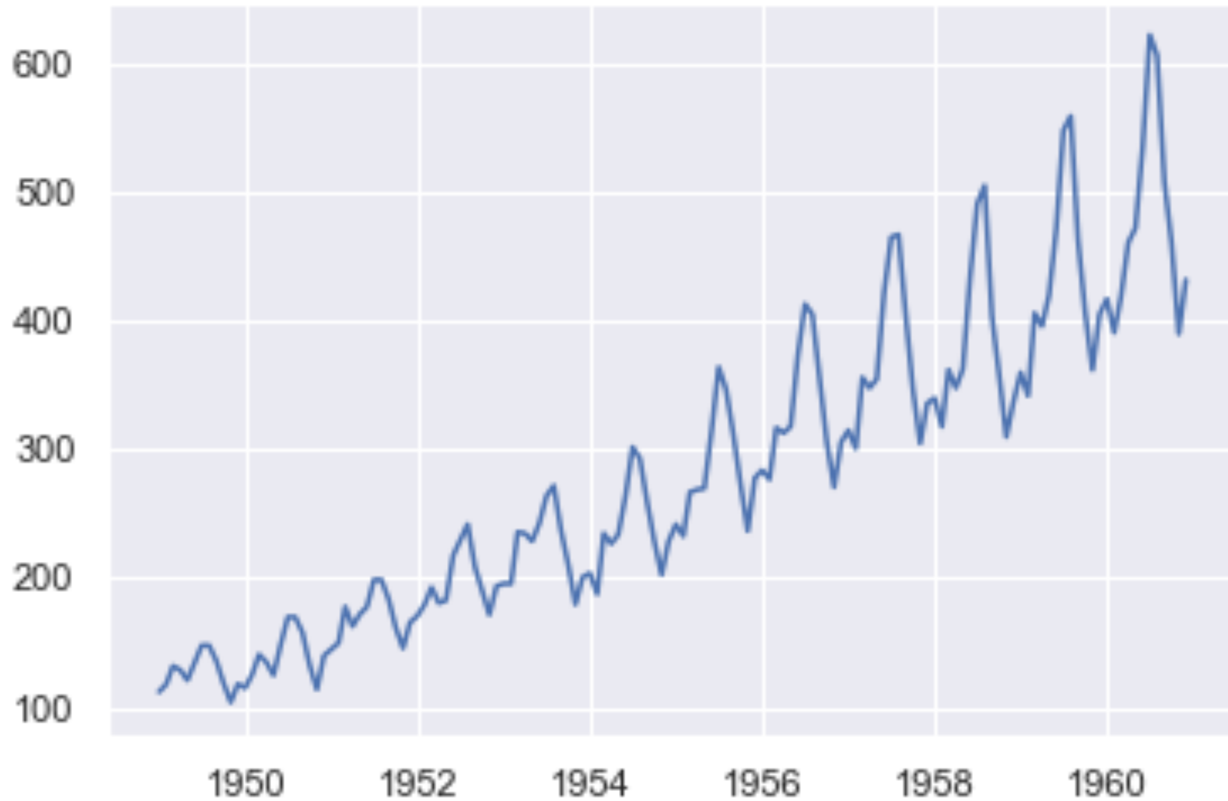
自己共分散 $$\hat{\gamma}_k = \frac{1}{T}\sum_{t=1}^{T} (y_t - \bar{y})(y_{t-k} - \bar{y})$$

自己相関 $$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}$$

推定値が簡単に得られる

時系列データの変換

原系列



・分散が一定でない

・非定常過程

など

→予測しにくい

・対数

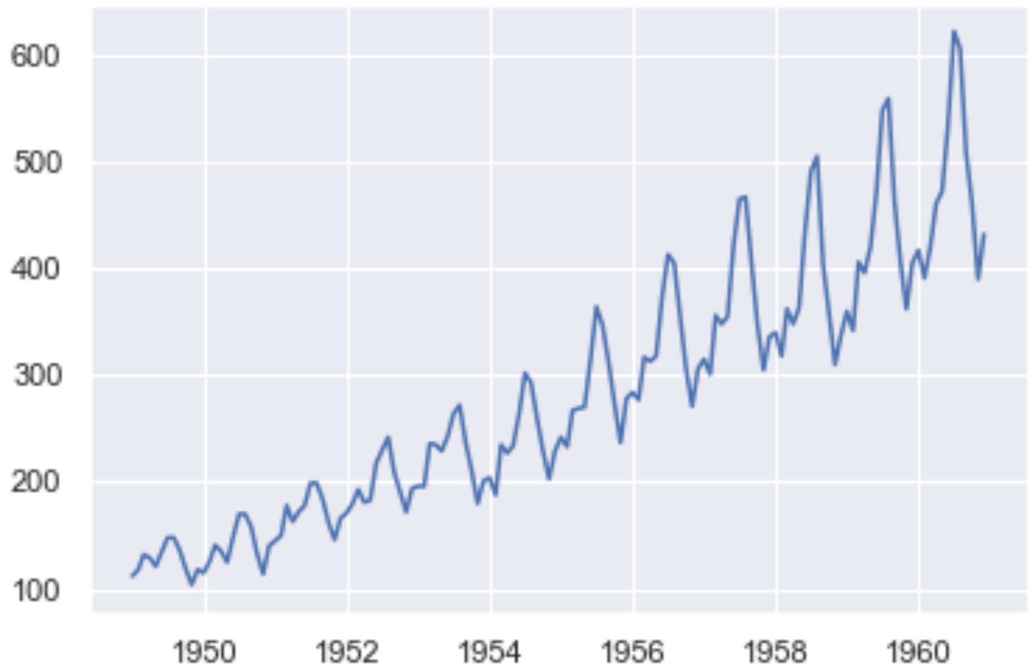
・差分

・対数差分

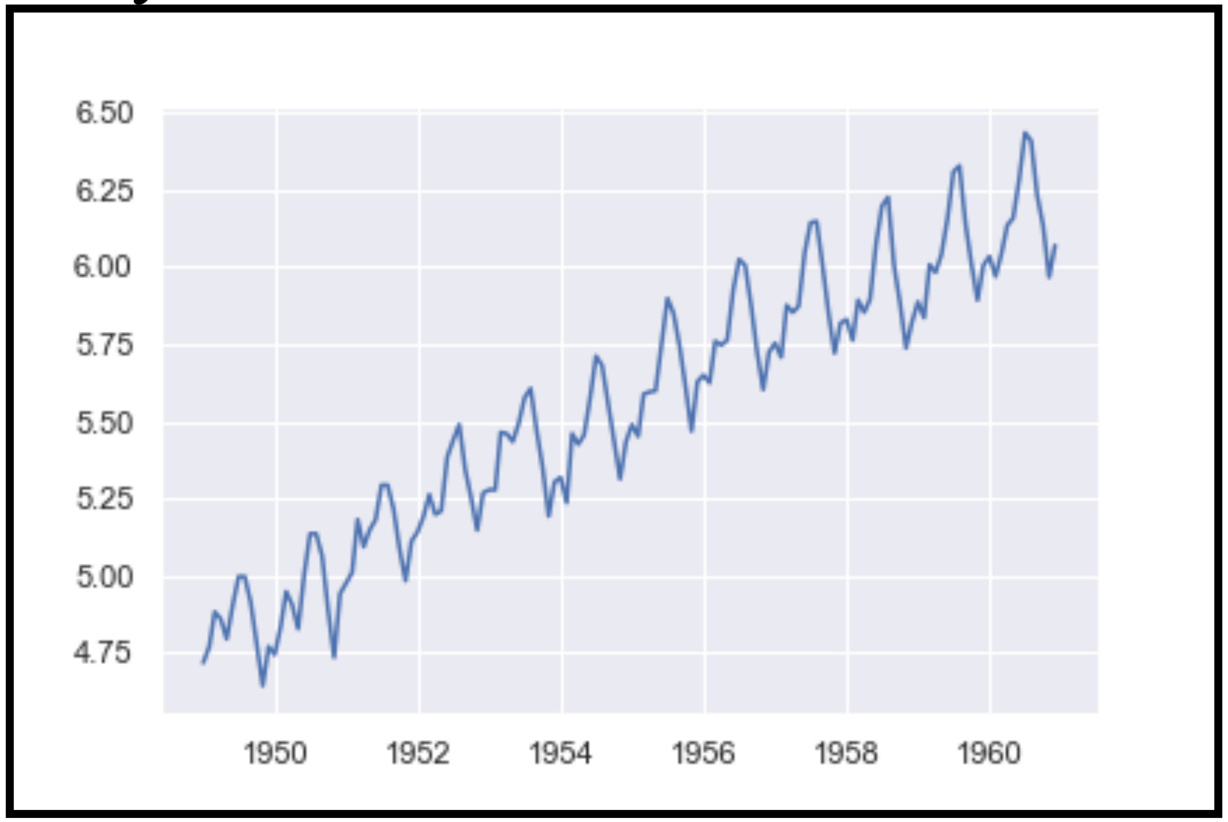
に変換

対数系列

$$\log y_t$$



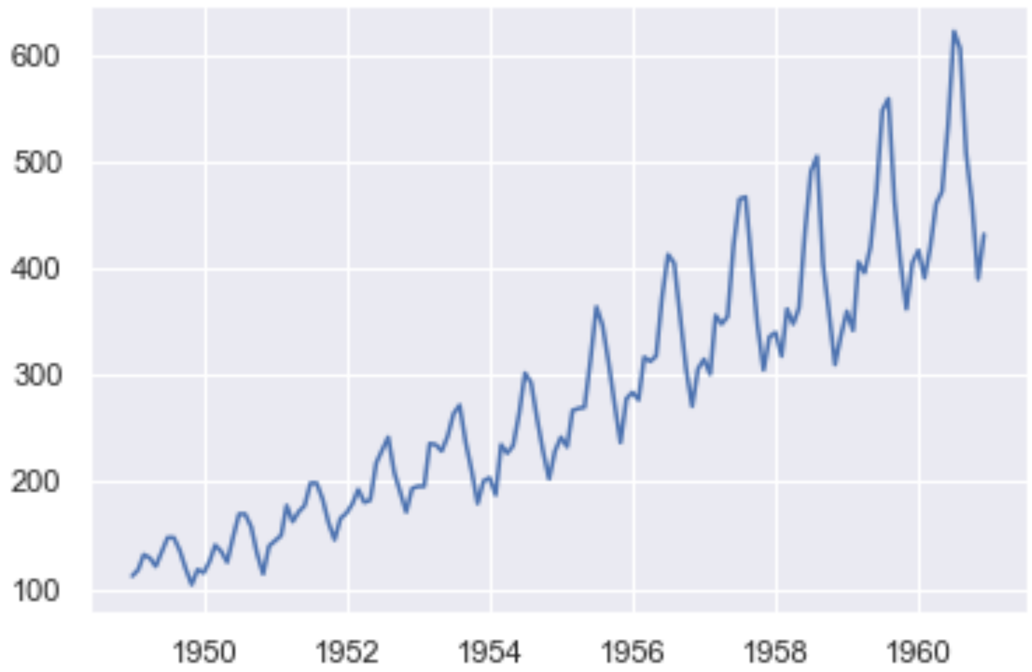
600
500
400
300
200
100
1950
1952
1954
1956
1958
1960

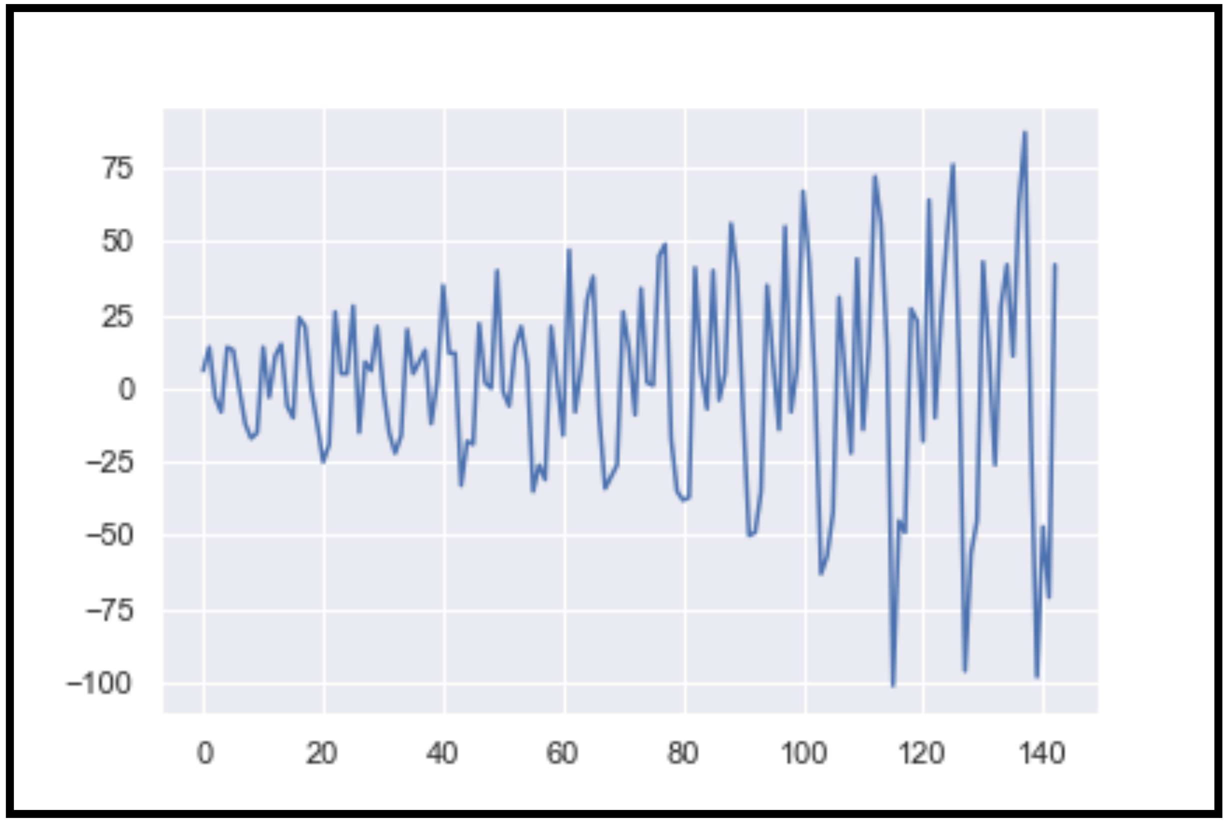


6.50
6.25
6.00
5.75
5.50
5.25
5.00
4.75
1950
1952
1954
1956
1958
1960

差分系列

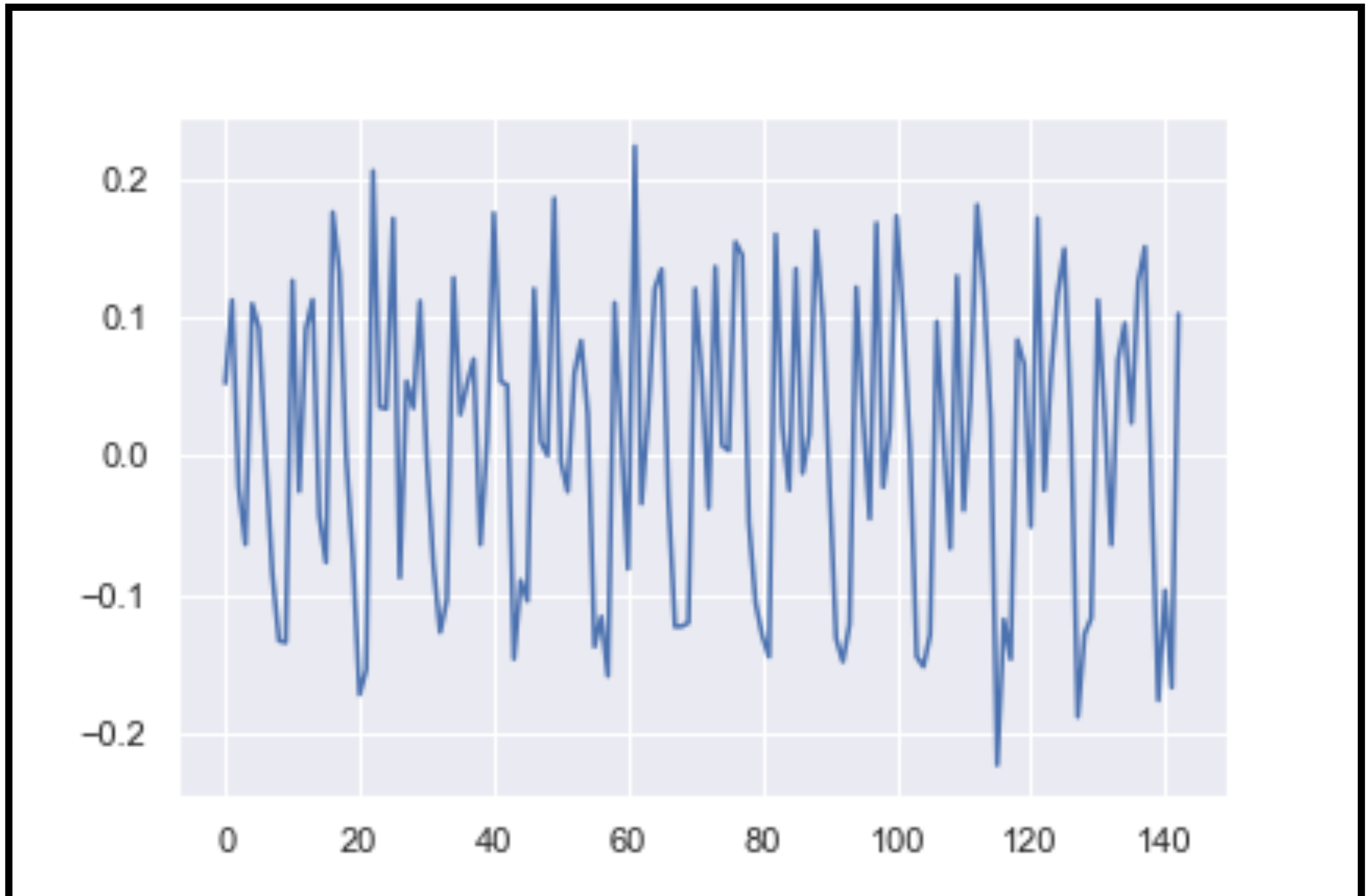
$$\Delta y_t = y_t - y_{t-1}$$





対数差分系列

$\Delta \log y_t$ 変化率に近似できる



・分散が一定

・定常過程

かもしれない

→直近なら

予測できそう

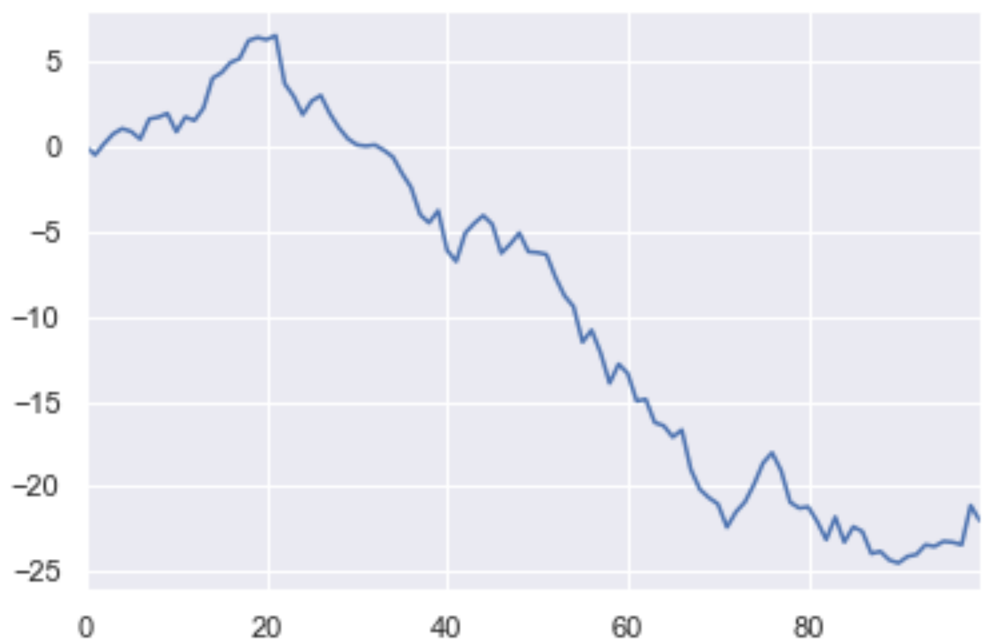
（本当か？）

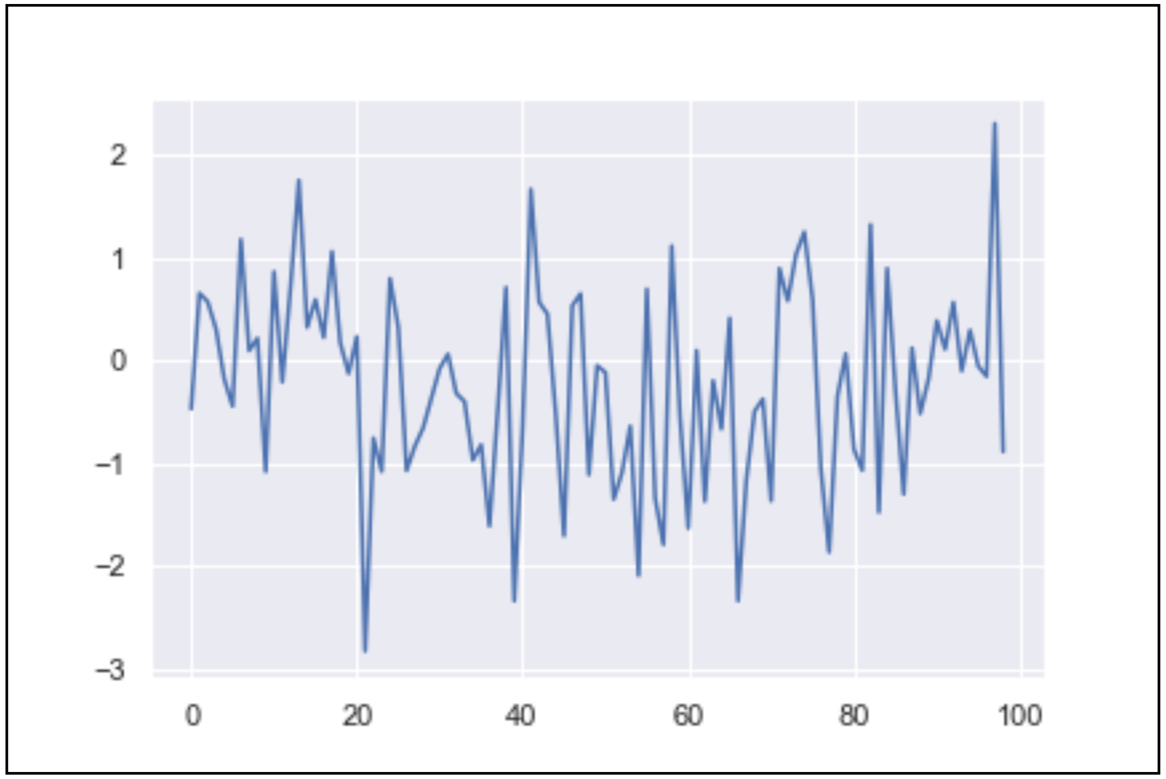
補足：単位根過程

もとの系列が非定常過程で,

差分系列が定常過程のもの

例：ランダムウォーク



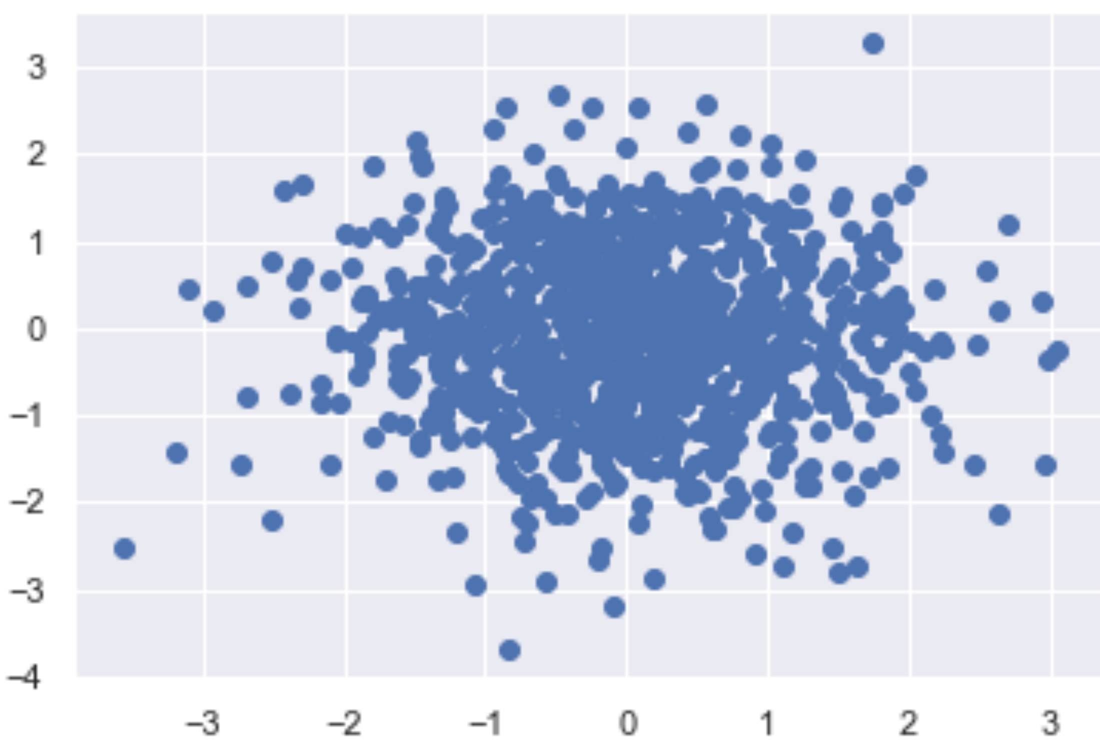


時系列データをみる[ハンズオン]

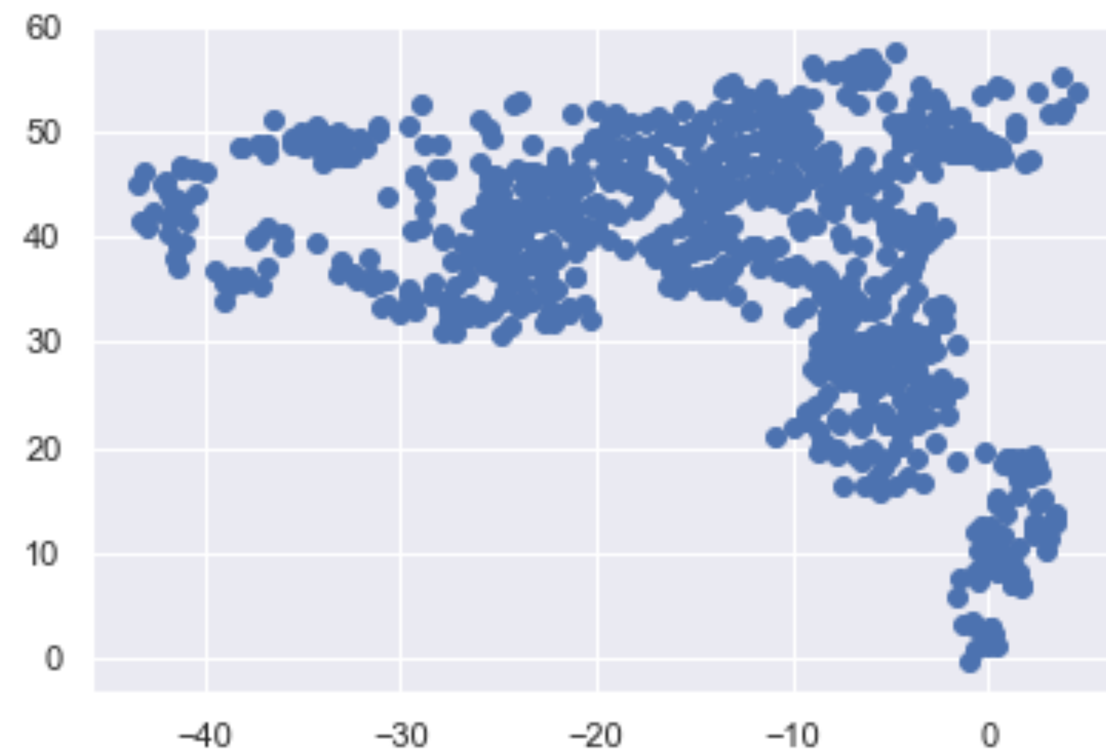
ホワイトノイズ，ランダムウォークなどを書いてみる[ハンズオン]

補足：見せかけの回帰

ホワイトノイズ



ランダムウォーク



回帰係数 -0.083 ~ 0.040 -0.436 ~ -0.314

?

なぜ？

回帰分析の前提

1. 系列無相関
2. 分散均一性
3. 説明変数との無相関
4. 正規性

時系列データには
自己相関がある
→係数に意味がない

確認：
Ljung-Box検定