時系列分析
ハンズオン#2

進行予定

第2回「ARIMAモデルを書いてみる」

・ARモデル
・MAモデル
・ARMAモデル
・ARIMAモデル
・優れたモデルとは何か
・それぞれのモデルを書いてみる[ハンズオン]
・モデルの評価を行う[ハンズオン]

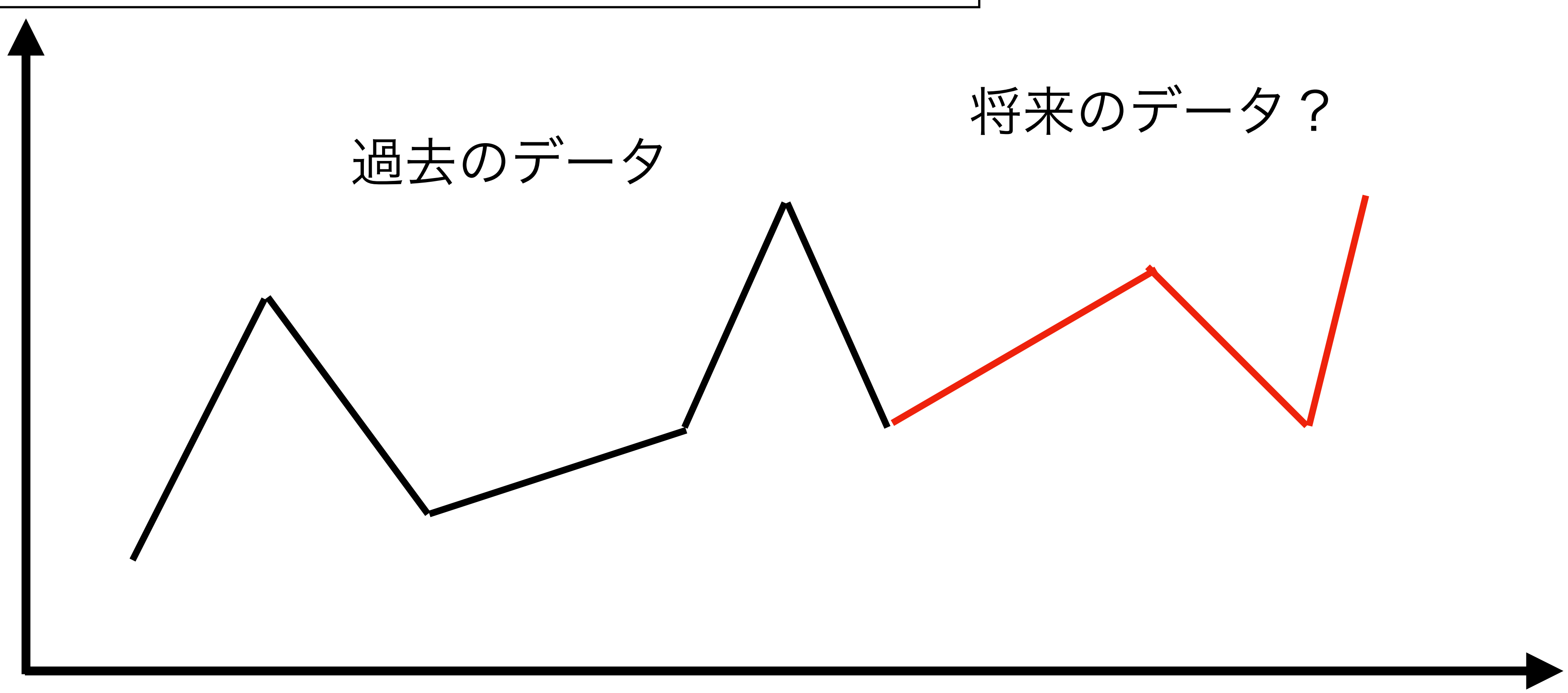
講師自己紹介

- 氏名：大久保　亮介
- 現在，薬学部4年　漢方薬専攻（予定）
- 担当講義：基礎統計→CNN，高校数学

第1回の復習

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

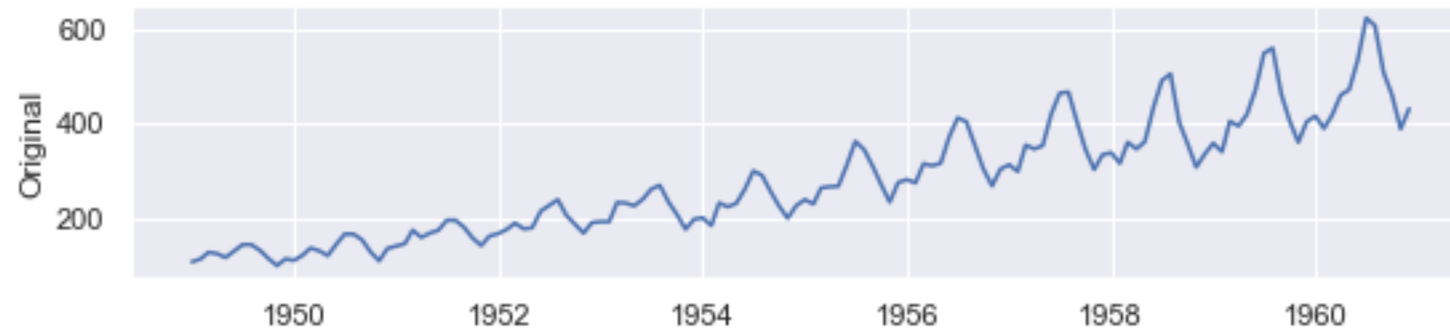


構成要素

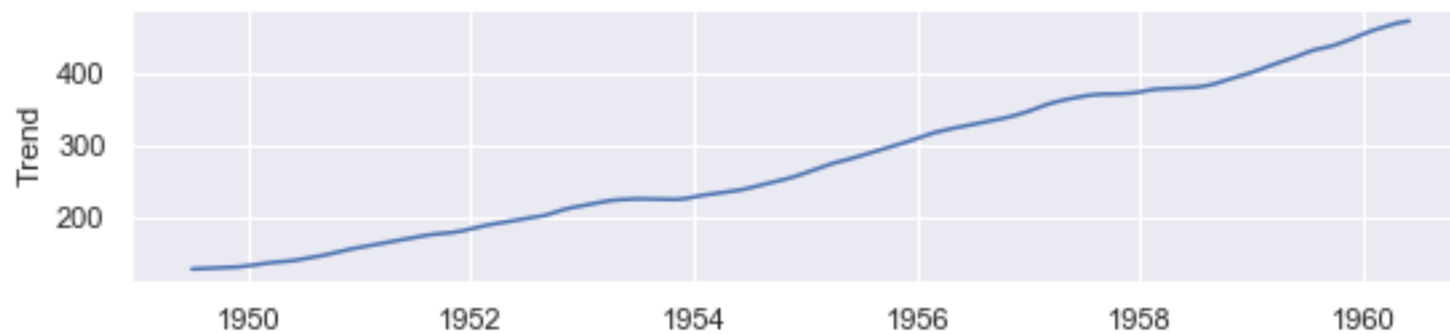
- 自己相関
- 周期的変動
- トレンド
- 外因性
- ホワイトノイズ

構成要素ごとに分解

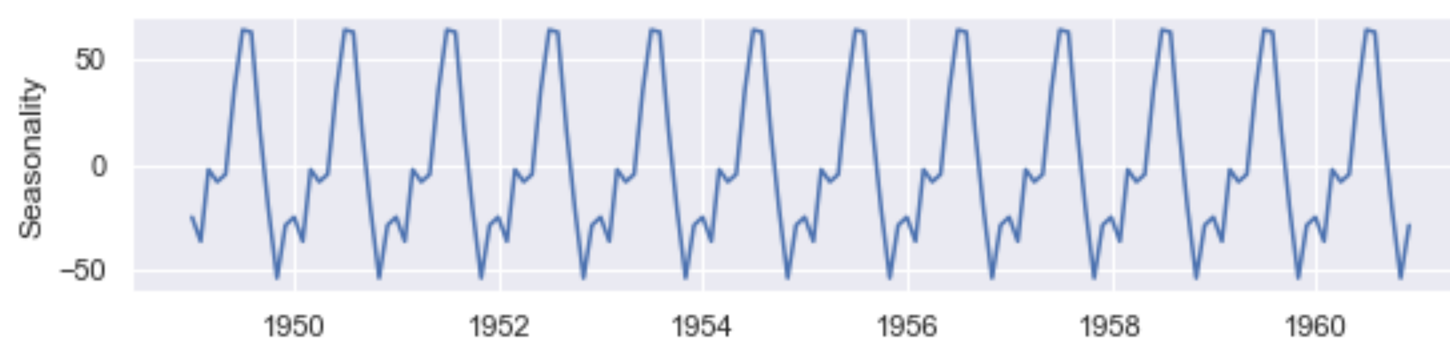
もとのグラフ



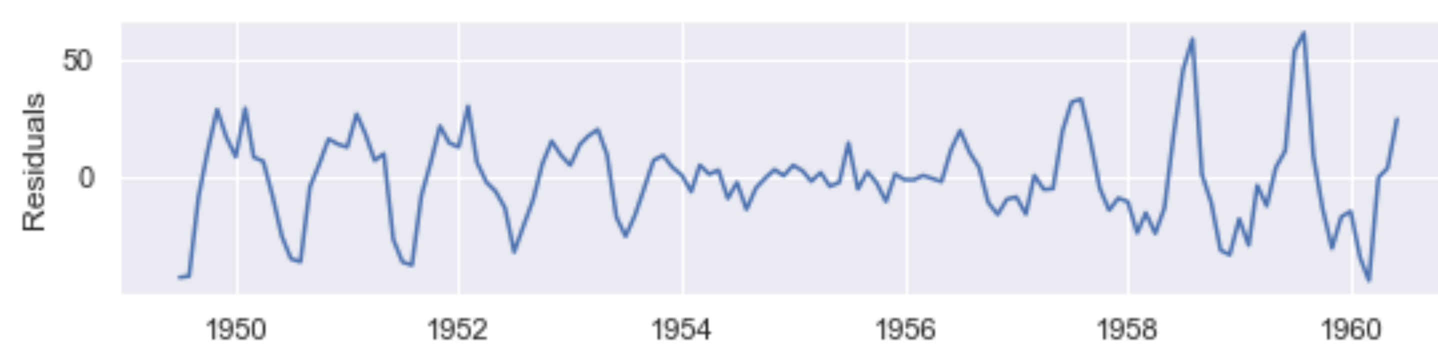
トレンド



季節性



ホワイトノイズ



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

$$E(\epsilon_t) = 0$$

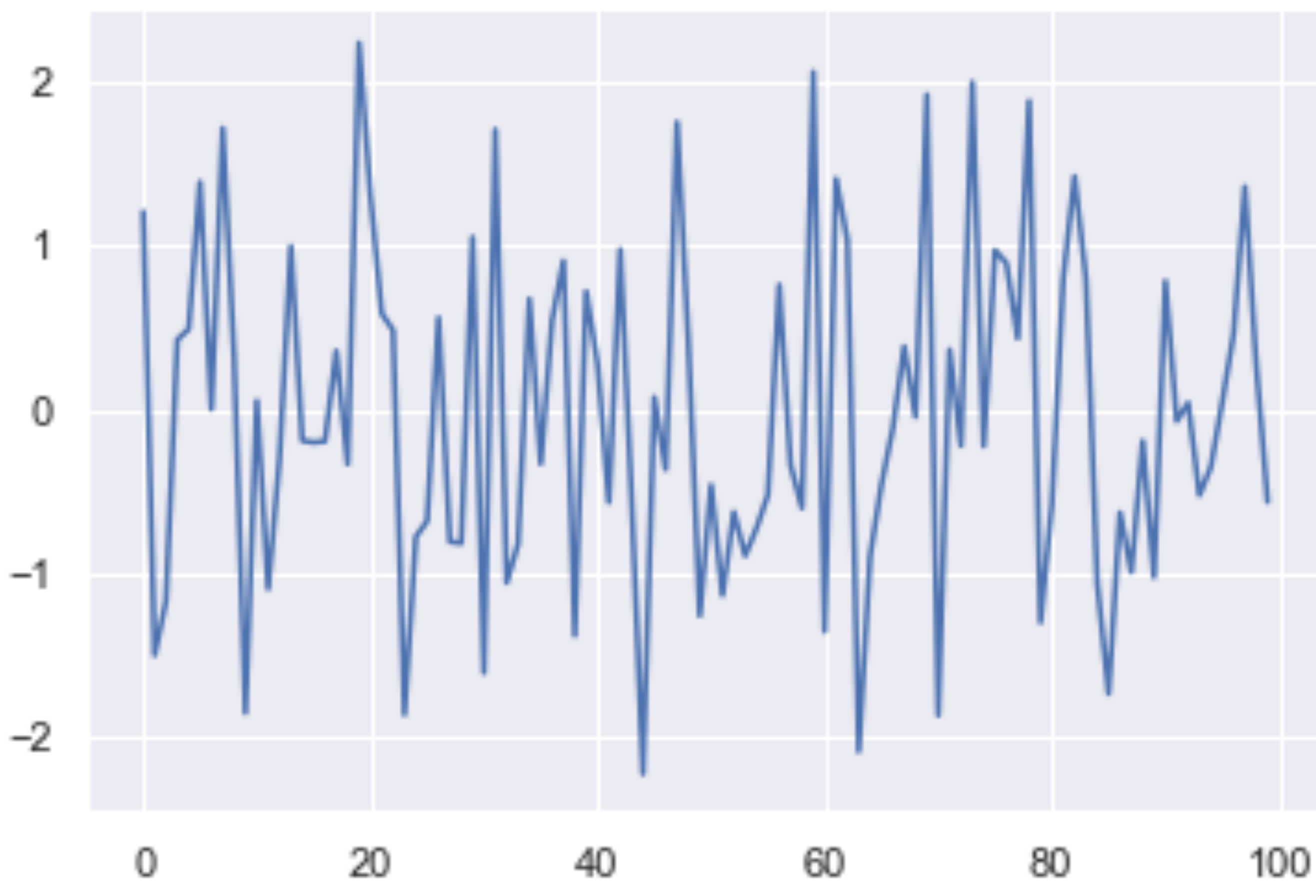
$$Cov(\epsilon_t, \epsilon_{t-k}) = \begin{cases} \sigma^2, k = 0 \\ 0\ , k \neq 0 \end{cases}$$

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

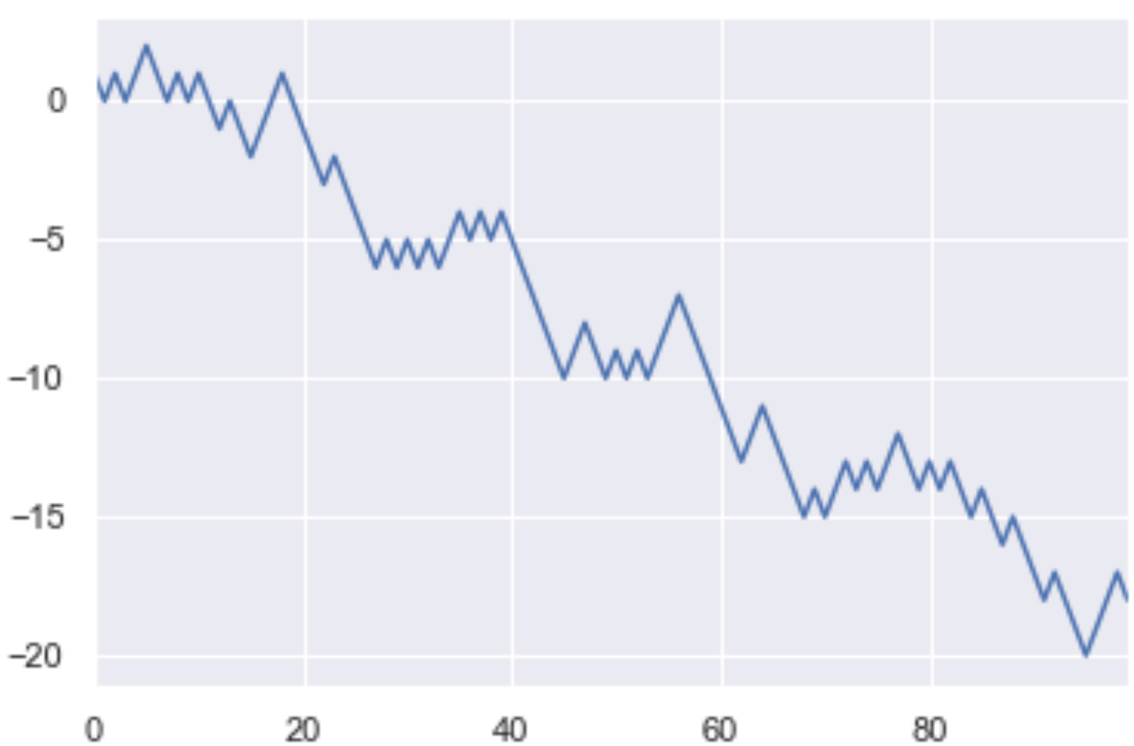


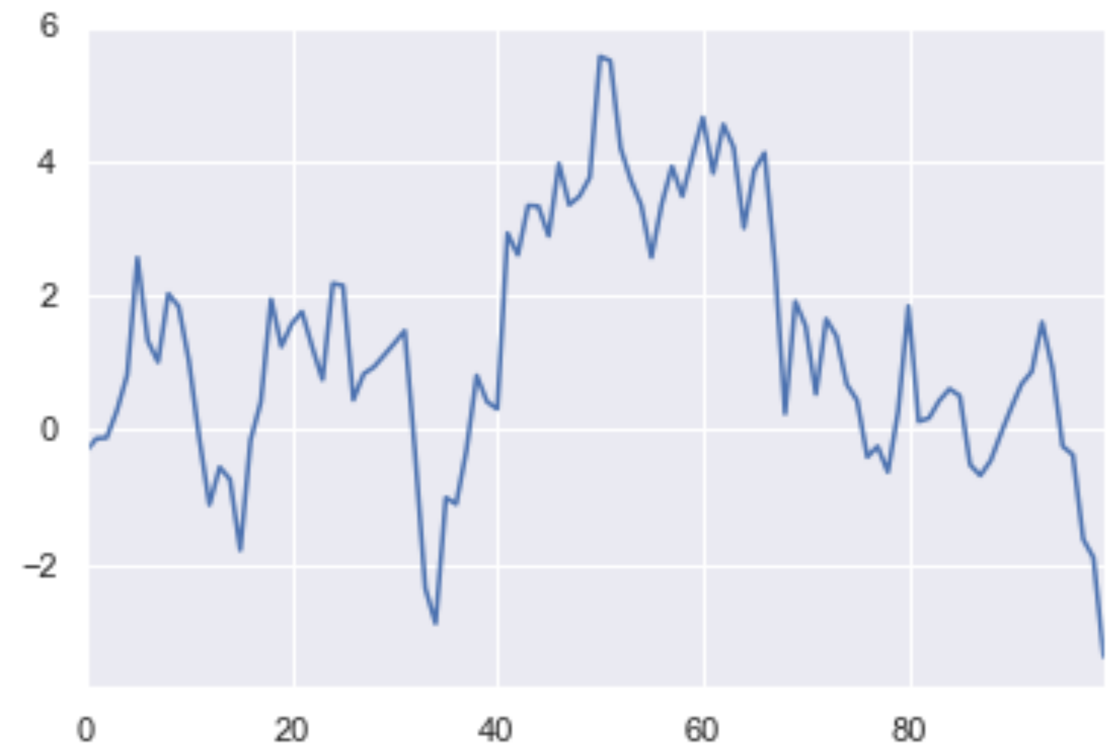
ランダムウォーク

iid系列の累積和

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

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





定常過程の定義

期待値は時点によらず一定

$$E(y_t) = \mu$$

自己共分散は
時点によらず
時間差のみに依存

$$Cov(y_t, y_{t-k}) = E[(y_t - \mu)(y_{t-k} - \mu)]$$

$$= \gamma_k$$

ARモデル

ARモデル

自己回帰モデル（AutoRegressive model）

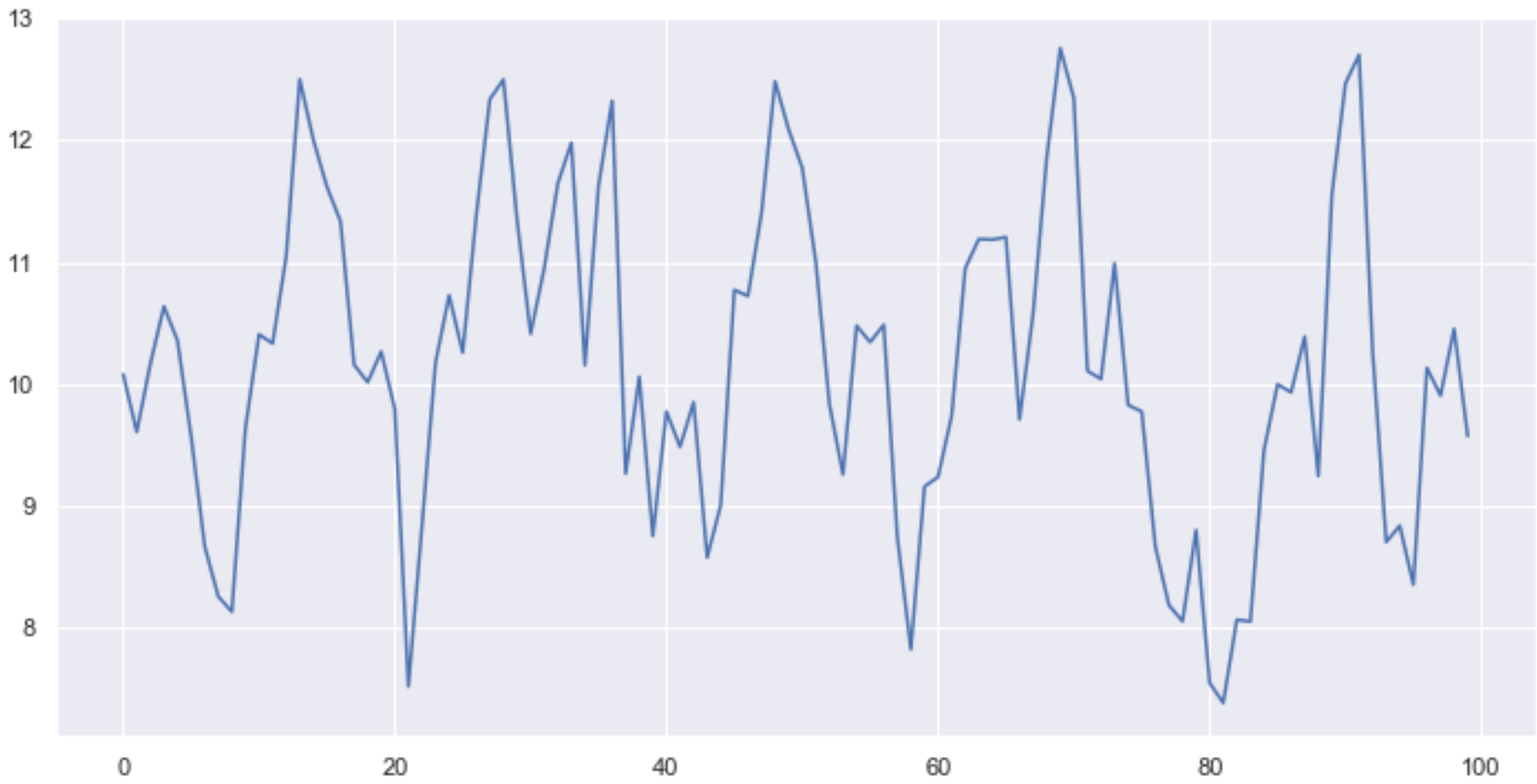
1次　AR(1)： $y_t = c + \phi_1 y_{t-1} + \epsilon_t \quad \epsilon_t \sim N(0,\sigma^2)$

例題： $y_t = 1 + 0.5 y_{t-1} + \epsilon_t$

$y_0 = 0,\ \epsilon_t = 0$ **のときの**
$y_1,\ y_2,\ y_3$ **を計算せよ**

p次　AR(p)： $y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t \quad \epsilon_t \sim N(0,\sigma^2)$

AR(1)のグラフ



AR(1)のパラメータ

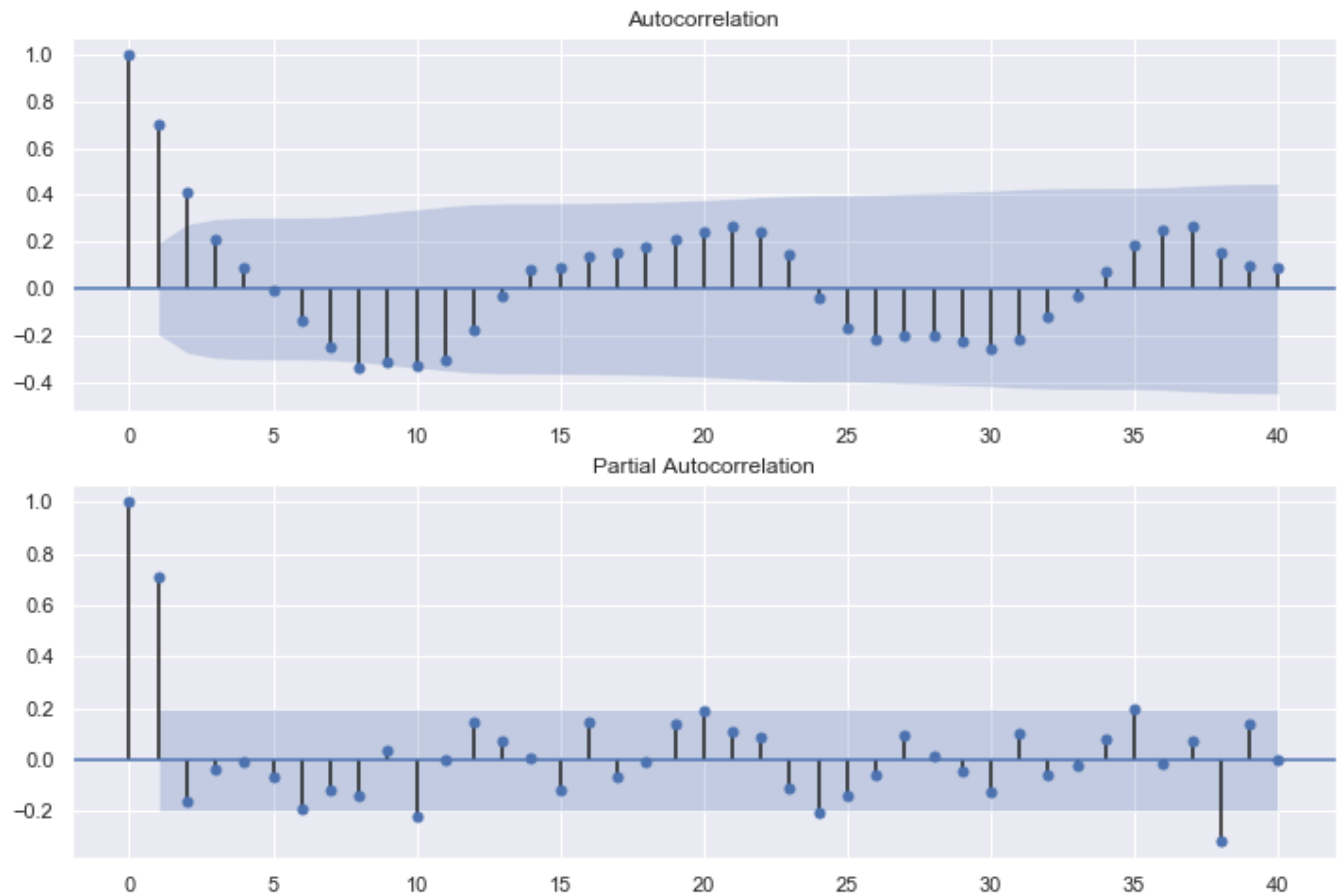
期待値 $\mu = \dfrac{c}{1 - \phi_1}$

分散 $\gamma_0 = \dfrac{\sigma^2}{1 - \phi_1^2}$

自己相関 $\rho_k = \phi_1 \rho_{k-1}$

Yule-Walker方程式

AR(1)のコレログラム



Autocorrelation
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
0
5
10
15
20
25
30
35
40
Partial Autocorrelation
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
5
10
15
20
25
30
35
40

MAモデル

MAモデル

移動平均モデル（Moving Average model）

1次　MA(1)： $y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1}$

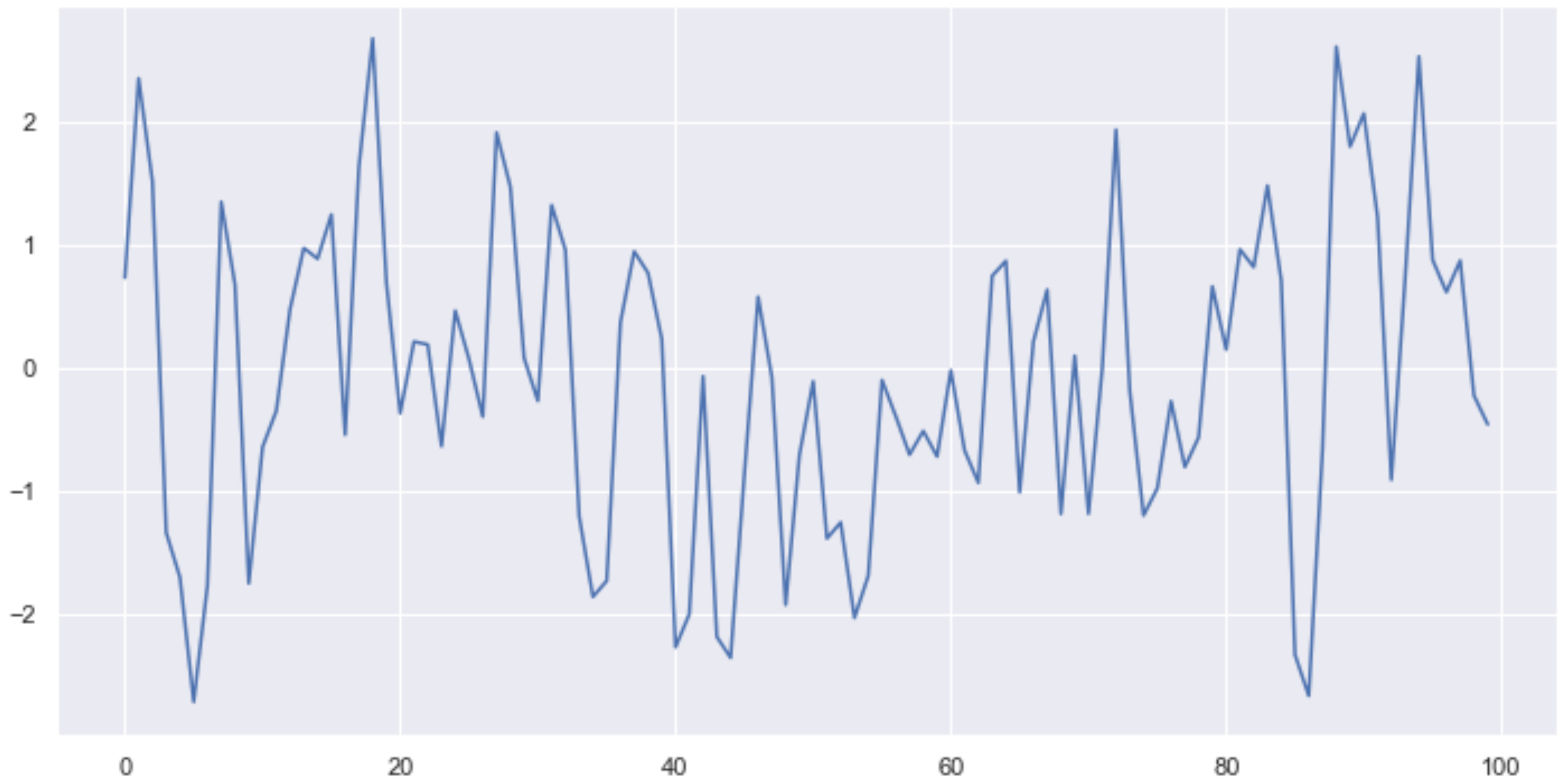
例題： $y_t = 1 + \epsilon_t + 0.5\epsilon_{y-1}$

$\epsilon_0 = 0.1,\ \epsilon_1 = 0.3,\ \epsilon_2 = 0.5$ **のときの**

$y_1,\ y_2$ **を計算せよ**

q次　MA(q)： $y_t = \mu + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t$

MA(1)のグラフ



MA(1)のパラメータ

期待値 $E(y_t) = \mu$

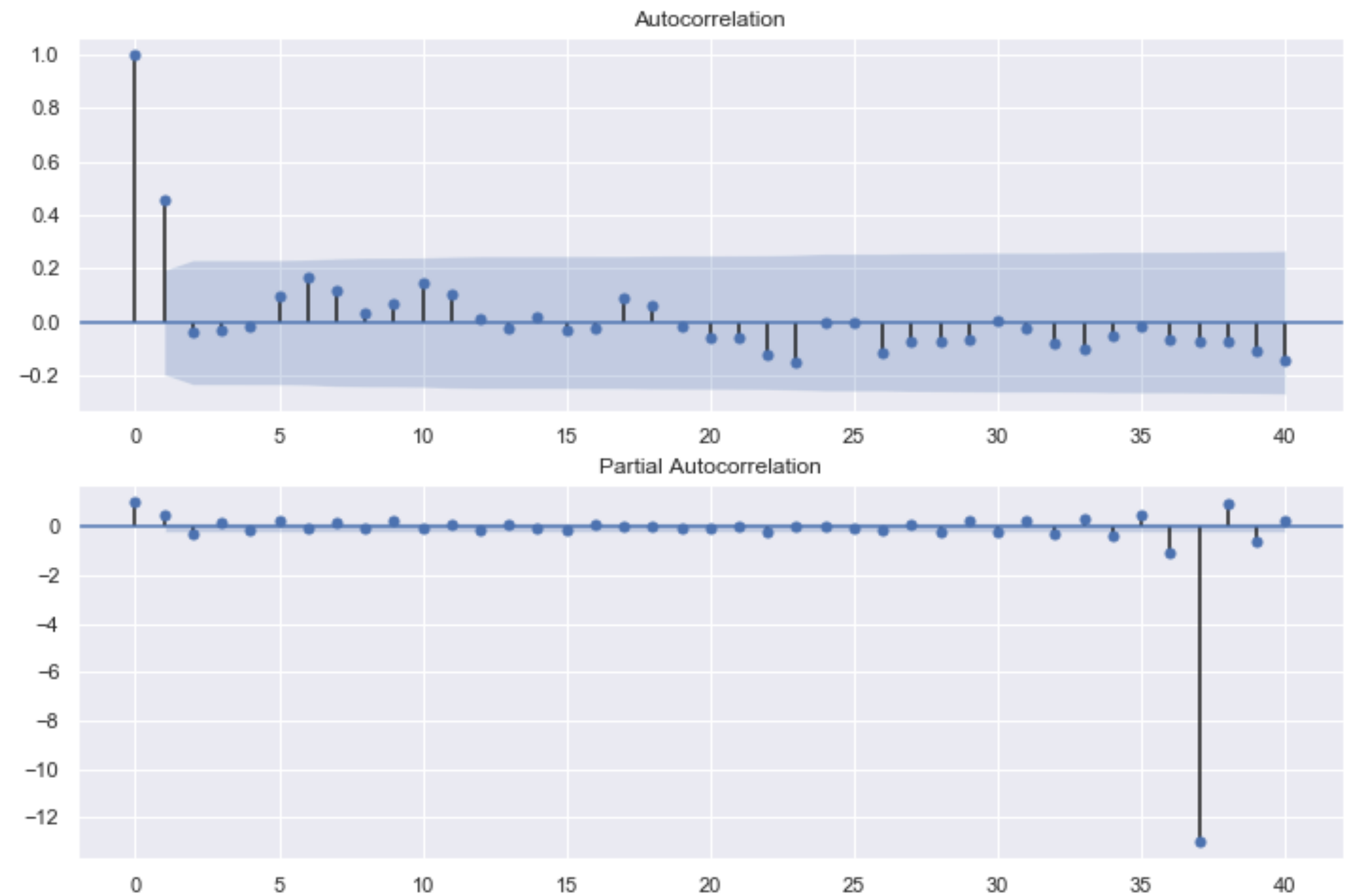
常に定常

分散 $\gamma_0 = (1 + \theta_1^2)\sigma^2$

自己相関 $\rho_1 = \dfrac{\theta_1}{1 + \theta_1^2}$

注意：理論上，2次以降の自己相関は0

MA(1)のコレログラム


Autocorrelation
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
5
10
15
20
25
30
35
40
Partial Autocorrelation
0
−2
−4
−6
−8
−10
−12
0
5
10
15
20
25
30
35
40

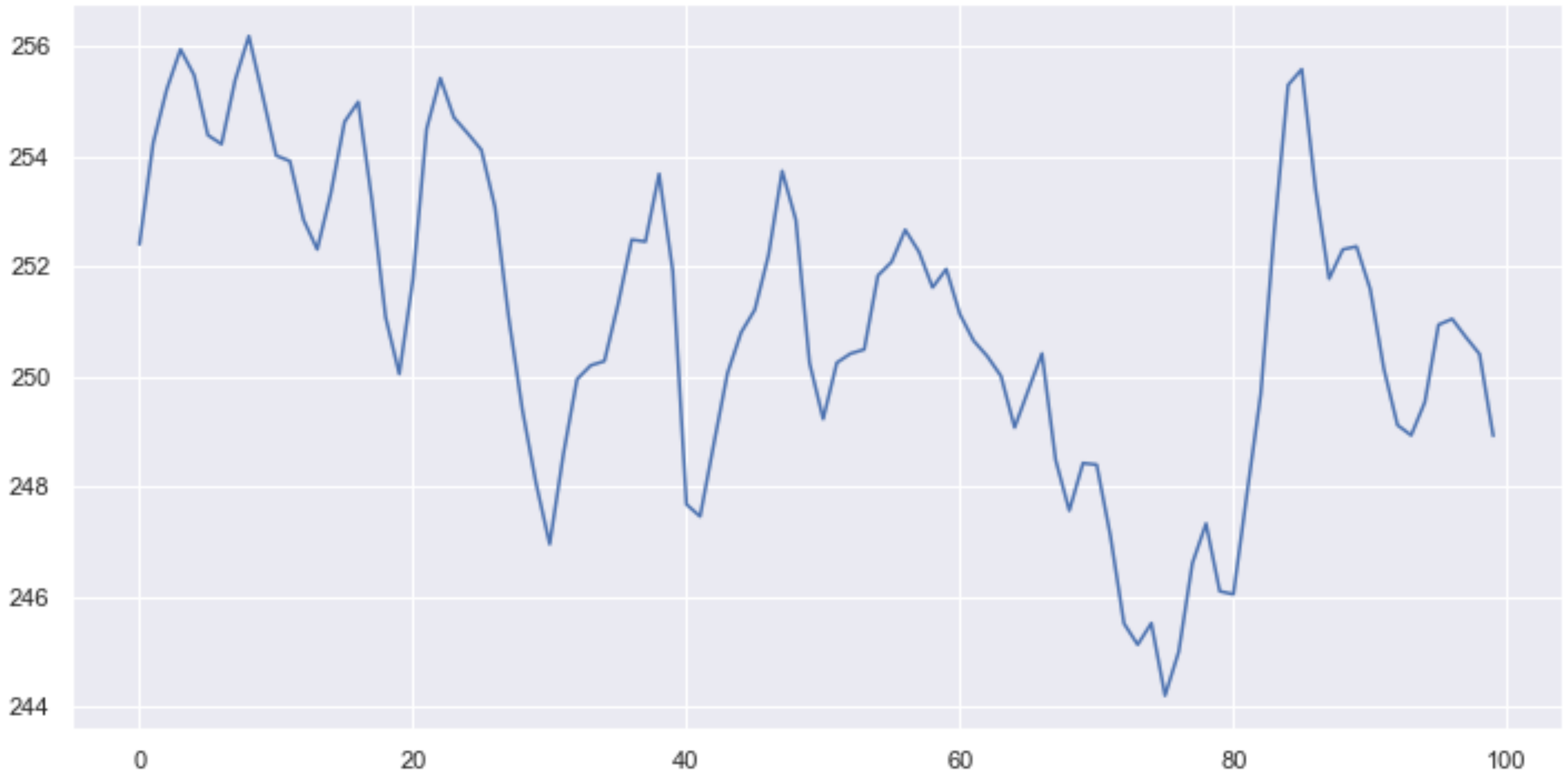
ARMAモデル

ARMAモデル

自己回帰移動平均モデル（AR + MA）

AR(p,q)： $$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t$$

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



補足：ARの定常条件

AR特性方程式

$$1 - \phi_1 z - \ldots - \phi_p z^p = 0$$

のすべての解の絶対値が1より大きい

例：AR(1)

$1 - \phi_1 z = 0$ **より** $z = \phi_1^{-1}$

$|\phi_1| < 1$ **のとき定常**

補足：MAの反転可能性

AR(∞)に書き直せるMA過程は反転可能

MA特性方程式

$$1 + \theta_1 z + \ldots + \theta_p z^p = 0$$

のすべての解の絶対値が1より大きい

例：MA(1)

$1 + \theta_1 z = 0$ **より** $|\theta_1| > 1$ **のとき反転可能**

補足：ARとMAとの関係

反転可能なMA(1)

$$y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} \ (\mu = 0)$$

$$\epsilon_t = -\theta_1 \epsilon_{t-1} + y_t$$

$$= (-\theta)^m \epsilon_{t-m} + \sum_{k=0}^{m-1} (-\theta)^k y_{t-k}$$

$$\to \sum_{k=0}^{\infty} (-\theta)^k y_{t-k} \ (m \to \infty)$$

AR(∞)

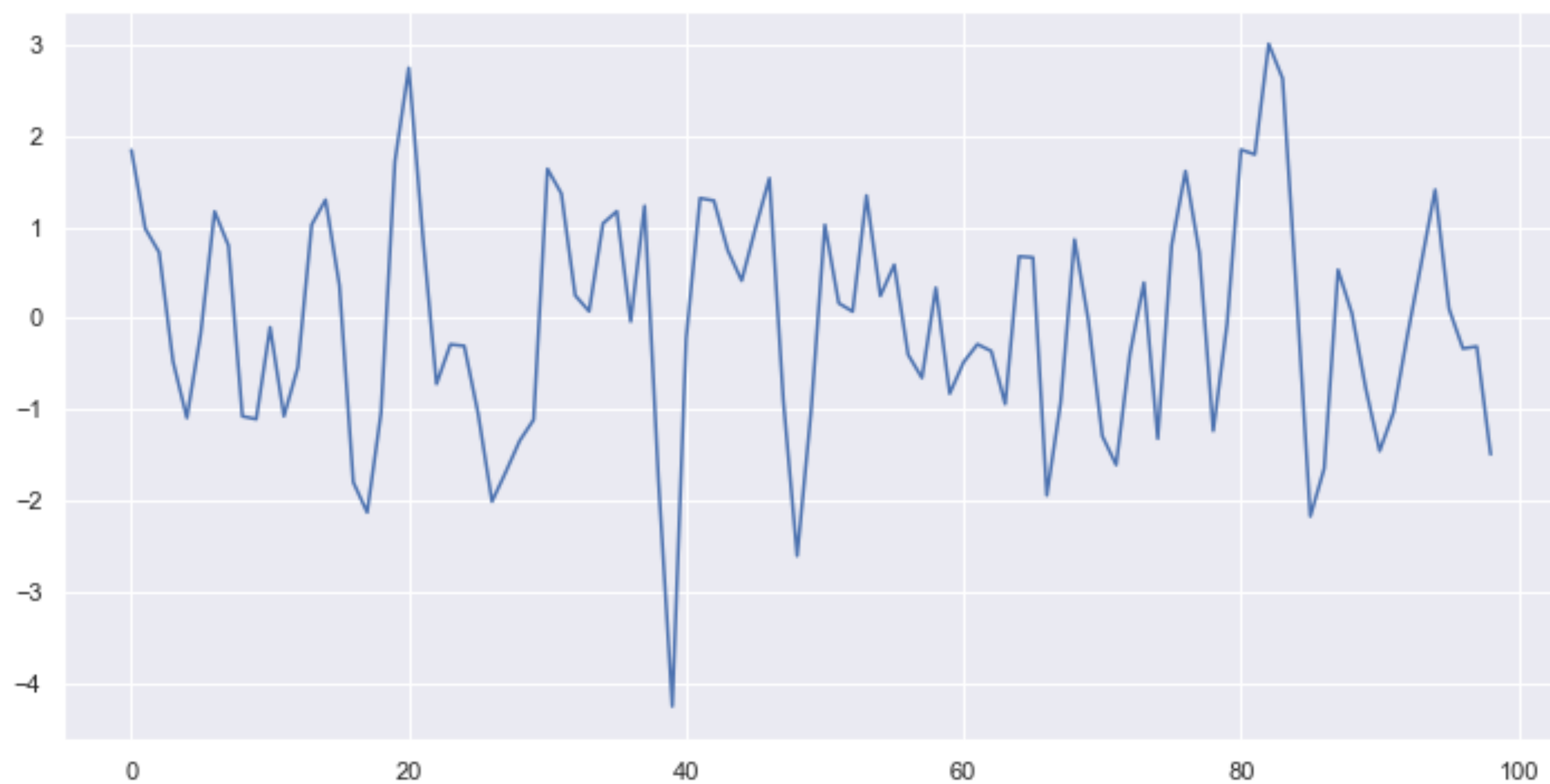
$$y_t = \sum_{k=0}^{\infty} (-\theta)^k y_{t-k} + \epsilon_t$$

ARIMAモデル

ARIMAモデル

自己回帰和分移動平均モデル（AR + Integrate +MA）

d階差分をとった系列が ARMA(p,q)にしたがう過程 →ARIMA(p,d,q)



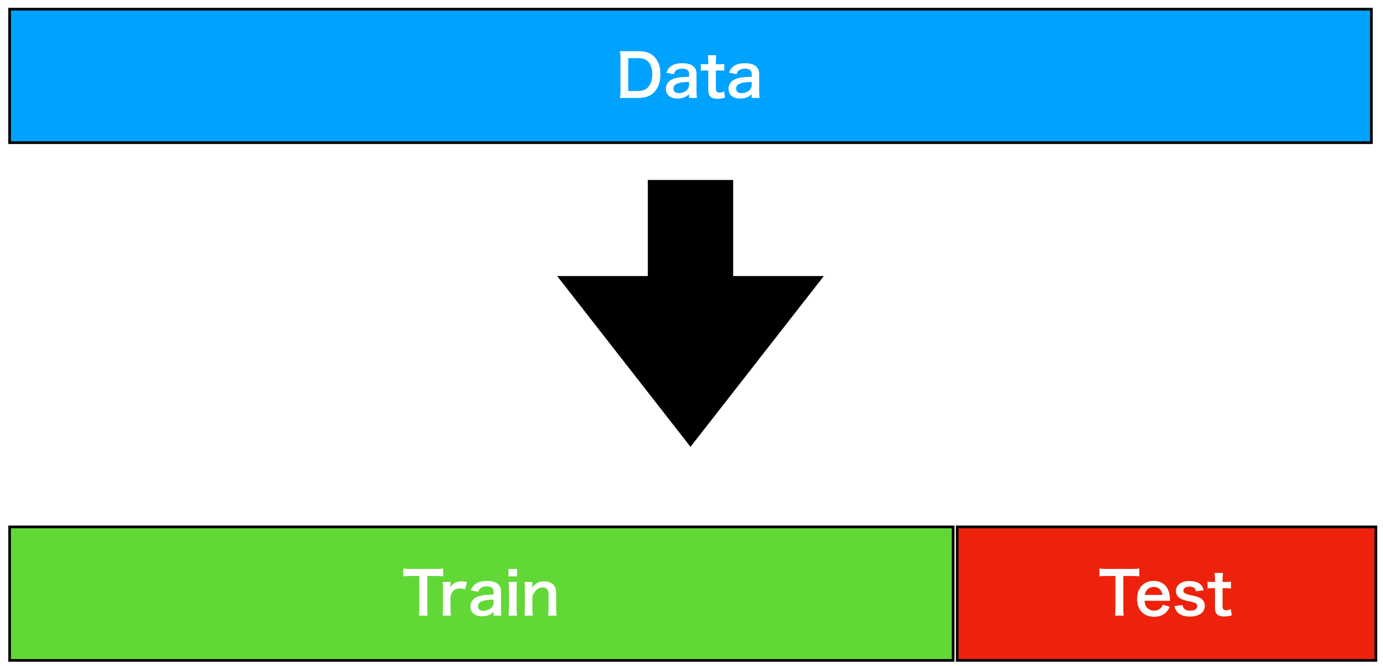
補足：ARIMAの派生

・SARIMA（Seasonal~）：
季節成分を加えたモデル

・ARIMAX（~with eXogenous variables）：
外因性を加えたモデル

優れたモデルとは何か

tast_dataによる予測



いいモデルの推測　　精度の算出

モデル選択

例：ARMA(p,q)の次数決定

手当たり次第に次数を当てはめて,
AICなどで評価する

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

最大化対数尤度　**パラメータの数**

→小さいほど，予測能が良い

モデル評価→同定

- 定常性：AR項

特性方程式について,
すべての解の絶対値が1より大きい

- 反転可能性：MA項
- 残差の自己相関 **Ljung-Box検定**
- 残差の正規性 **Jarque-Bera検定**

ナイーブ予測

その精度が高いと言えるか？

→単純なモデルと比較

- 過去の平均値を予測値として出す（ホワイトノイズ）

- 前地点の値を予測値として出す（ランダムウォークなど）

→予測モデル＞ナイーブ予測で

予測モデルの意味がある

それぞれのモデルを書いてみる[ハンズオン]

モデルの評価を行う[ハンズオン]