L14 自己回帰モデルAR(m), 移動平均モデルMA(k)

多変量解析☆演習

今日の目標

- 確率過程(母ナントカ)を時系列データ(標本ナントカ)と対比して説明できる.
- 確率過程のひとつである自己回帰モデルAR(m)を説明できる.
- 確率過程のひとつである移動平均モデルMA(m)を説明できる.

L13の補足

L13のQuizの略解

データ$\{X_t\}_{t=0,1,\ldots,L-1}$の周期性を知るには, 離散フーリエ変換も使える. フーリエ変換

$$\tilde{X}_s = \sum_{t=0}^{L-1} e^{-\frac{2\pi st}{L}} X_t$$

で, $|\tilde{X}_s|^2$ が大きいなら, 周期L/sの成分の寄与が大きい.

フーリエ解析

時系列の母ナントカ

時系列データは確率過程の標本

時系列データは確率変数の標本と思える. 為替相場が, 同じルール, 同じ状況でパラレルワールドでスタートしても別の値動きをしうる. どんな母ナントカの標本と思えるだろう?

確率過程(の特別なケース)
時刻 t に対した連続型確率変数の列 $\{X_t\}_{t=0,1,\cdots}$ を確率過程という. 大注意: X_tは独立でも同分布でもない! 1回サイコロを振ると(標本抽出すると), $X_0, X_1, \cdots$ がぜんぶでてくる.

単回帰でどう?

時系列データを見ると, 次のように単回帰分析したくなるかも.

- tを説明変数(いままでのx)
- Xを目的変数(いままでのy)

$$X_t = \beta_0 + \beta_1 \times t + \epsilon_t,$$
$$\epsilon_t \text{は独立同分布}, \quad \epsilon_t \sim \mathrm{N}(0, \sigma^2), .$$

これで済む場合もあるけど, つまらなくて単純. トレンドは$\beta_0 + \beta_1 t$, 周期なし, X_tは独立.

$$\begin{aligned} \mathrm{Cov}[X_1, X_2] =& \mathrm{Cov}[\epsilon_1, \epsilon_2] \quad \text{確率変数以外は共分散に関係しない} \\ =& 0 \end{aligned}$$

以下では, もっと複雑で面白い場合を扱う.

そこでは, tでなく x_{t-1} を説明変数にする(ようなもの)

1次の自己回帰モデルAR(1) auto regression

1次の自己回帰モデルAR(1)

- $\{\epsilon_t\}_{t=1,...,}$: 連続型確率変数の列
- $\{X_t\}_{t=0,1,...}$: 連続型確率変数の列
- ϕ_1: 定数

$$X_t = \phi_1 \cdot X_{t-1} + \epsilon_t. \quad (t \geq 1)$$

ただし, ϵ_t $t = 0, 1, 2, ...$は,

- WN1 $\mathrm{E}[\epsilon_t] = 0$
- WN1a $\mathrm{E}[\epsilon_t \epsilon_t] = \sigma^2 \quad \sigma$
- WN1b $\mathrm{E}[\epsilon_t \epsilon_s] = 0 \quad (t \neq s)$ 独立
- $\mathrm{E}[X(t)\epsilon_s] = 0 \quad (t > s)$

ϵ_tの「互いに無関係さ」に対しては, 独立同分布, $\epsilon_t \sim \mathrm{N}(0, \sigma^2)$のような分布ベースの強い仮定ではなく, 母共分散が0という母期待値ベースの弱い仮定を置いている. WN1, WN2 を満たすϵ_tを, ホワイトノイズ, white noise, 白色雑音という.

例1: 感染症の患者数

例2: $\phi_1 = 1$:ランダムウォーク

確率モデル及び演習

AR(1)の生成

漸化式のりで計算できる.

$$
\begin{aligned}
X_t =& \phi_1 X_{t-1} + \epsilon_t \\
=& \phi_1(\phi_1 X_{t-2} + \epsilon_{t-1}) + \epsilon_t \\
=& \phi_1(\phi_1(\phi_1 X_{t-3} + \epsilon_{t-2}) + \epsilon_{t-1}) + \epsilon_t \\
=& \cdots \\
=& \phi_1^t X_0 + \phi_1^{t-1}\epsilon_1 + \phi_1^{t-2}\epsilon_2 + \cdots + \phi_1^0 \epsilon_t \\
=& \phi_1^t X_0 + \sum_{k=0}^{t-1} \phi_1^k \epsilon_{t-k}
\end{aligned}
$$

時系列の定常性

ここに, 一般論, 時系列が定常であることの定義を挿入

時系列の(弱い意味での)定常性

時系列 $\{X_t\}$ が(弱い意味で)定常であるとは, 次の両方を満たすこと.

- $\mathrm{E}[X_t]$がtによらないこと.
- $\mathrm{Cov}[X_t, X_{t-s}]$が$t$によらないこと.

分布や高次のモーメントはtに依存してもよい(依存しない⇔強い意味での定常性)

$x_t = \phi_1^t X_0 + \sum_{k=0}^{t-1} \phi_1^k \epsilon_{t-k}$ は, $|\phi_1| \leq 1$なら$t \to +\infty$で定常になる望みがある.

AR(1)の母平均値, 母共分散

$$
\begin{aligned}
\mathrm{E}[X_t] =& \mathrm{E}[\phi_1^t X_0 + \sum_{k=0}^{t-1} \phi_1^k \epsilon_{t-k}] \\
=& \phi_1^t \mathrm{E}[X_0]
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{V}[X_t] =& \mathrm{V}[\phi_1^t X_0 + \sum_{k=0}^{t-1} \phi_1^k \epsilon_{t-k}] \\
=& \mathrm{V}[\phi_1^t X_0] + \sum_{k=0}^{t-1} \phi_1^k \mathrm{V}[\epsilon_{t-k}] \\
=& \phi_1^t \mathrm{V}[X_0] + \sum_{k=0}^{t-1} \phi_1^k \sigma^2 \\
=& \phi_1^t \mathrm{V}[X_0] + \frac{1-\phi_1^t}{1-\phi_1}\sigma^2
\end{aligned}
$$

$|\phi_1| < 1, t \to +0$で, 定常の望みあり, $\mathrm{V}[X_t]$も一定. 以下, 簡単のため$\mathrm{E}[X_0] = \mathrm{E}[X_t] = 0$と仮定.

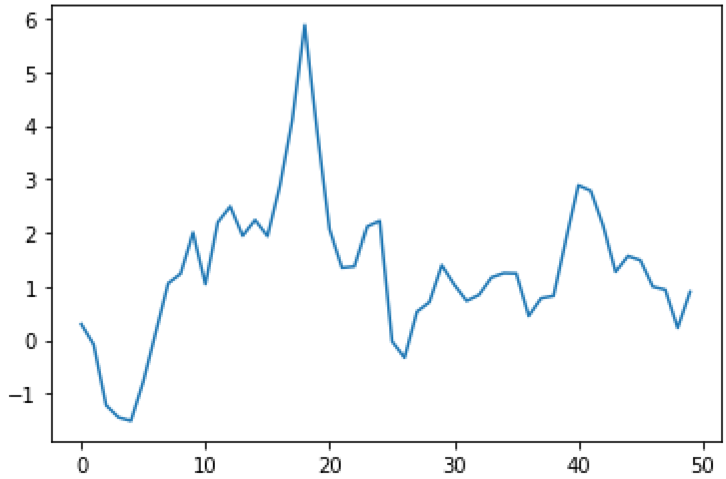
$$\begin{aligned}
&\mathrm{Cov}[X_t, X_{t-s}] \\
&\quad =\mathrm{E}[X_t X_{t-s}] \\
&\quad =\mathrm{E}[(\phi_1^t X_0 + \sum_{k=0}^{t-1} \phi_1^k \epsilon_{t-k}) \\
&\qquad \times (\phi_1^{t-s} X_0 + \sum_{\ell=0}^{t-s-1} \phi_1^\ell \epsilon_{t-s-\ell})]
\end{aligned}$$

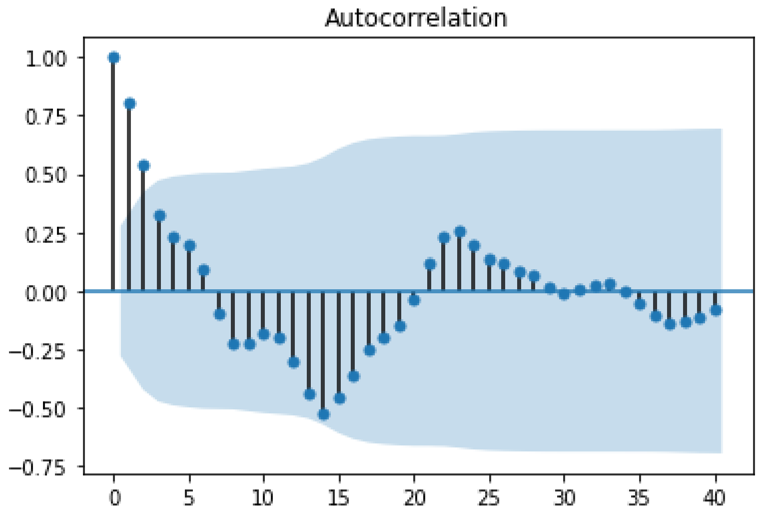
展開すると 0 にならないのは, $\mathrm{E}[X_0X_0]$, $\mathrm{E}[\epsilon_k\epsilon_k]=\sigma^2$のみ.

$$\begin{aligned}
\mathrm{Cov}[X_t, X_{t-s}] &=\phi_1^{2t-s}\mathrm{E}[X_0X_0] + (\phi_1^{2t-2s-2} + \phi_1^{2t-2s-4} + \cdots + \phi_1^0)\phi_1^s)\sigma^2 \\
&=\phi_1^{2t-s}\mathrm{E}[X_0X_0] + \frac{1-(\phi_1^2)^{t-s}}{1-\phi_1^2}\phi_1^s\sigma^2 \\
&\to \begin{cases} \frac{\phi_1^s\sigma^2}{1-\phi_1^2} & (|\phi_1|<1) \\ \approx t & (|\phi_1|=1) \\ \approx t\phi_1^{2t} & (|\phi_1|>1) \end{cases}
\end{aligned}$$

$|\phi_1|<1$のとき定常. このとき, ラグsの自己相関係数は $r(s)=\frac{\phi_1^s\sigma^2}{1-\phi_1^2} \quad (|\phi_1|<1)$ つまり等比数列. そうなってるやつあったでしょ? あれからϕ_1が読み取れる.

AR(1)の自己相関係数とコレログラム





m次の自己回帰モデルAR(m)

m次の自己回帰モデルAR(m)

m次の自己回帰モデルAR(m)

- $\{\epsilon_t\}_{t=1,...,}$: 連続型確率変数の列
- $\{X_t\}_{t=0,1,...}$: 連続型確率変数の列
- $\phi_0 = 1, \phi_1, \dots, \phi_m$ 定数

$$X_t = \phi_1 \cdot X_{t-1} + \cdots + \phi_m \cdot X_{t-m} + \epsilon_t. \quad (t \geq 1)$$

m個前の時刻のX_{t-s}から影響する.

Quiz L14-1

Quiz L14-1

AR(1)の生成

母集団からの標本抽出=確率過程のシミュレーション

シミュレーション及び実習(確率の現れないシミュレーション)

ARMAはAR(1)

を一般化したモデル.

```
from statsmodels.tsa.arima_process import ArmaProcess
ar=[1,-0.9,0.5]  # [a0,-a1,-a2,-a3,..]. a1以降は符号を逆にする
ma=[1]      # 現時点ではこうする
AR=ArmaProcess(ar,ma) # 自己回帰モデルを計算するオブジェクト
x=AR.generate_sample(nsample=50) # 時間の長さ=50
```

これは, $\phi_1 = 0.9, \phi_2 = -0.5$ という例.

リストxが$x_0, x_1, x_2, \ldots, x_{49}$に代入される.

手でやるならこう.

```
x=[2] # x1
t=1
phi1=1.0
for t in range(1,100,1):
  xnext=x[t-1]*phi1 + stats.norm(loc=0,scale=1).rvs(size=1)[0]
  x.append(ynext)
```

字下げした部分がブロック. Cでいうなら,

```
double x[101];
double xnext;
int t;
x[0]=2;
phi1=1.0;
for(t=1; t<=100; t+=1){
  xnext=x[t-1]*phi1+乱数();
  x[t]=xnext;
}
```

AR(m)モデルの推定と時系列の予測

```
from statsmodels.tsa import ar_model
model = ar_model.AR(x)  # modelは ARモデルでの予測を行うオブジェクト. xは時系列を収めた Series
result = model.fit(maxlag=12, ic='aic') # mが最大12までの範囲で よく予測するmと係数を探せ. 基準はAIC(魔法)
result.k_ar # k_ar とは mのこと
result.fittedvalues  # xと同じ範囲を予測した時系列
result.predict(t1,t2) # t1からt2までの期間を予測した時系列
```

k次の移動平均モデルMA(k)

k次の移動平均モデルMA(k)

k次の移動平均モデルMA(k)

- $\{\epsilon_t\}_{t=1,...,}$: 連続型確率変数の列
- $\{X_t\}_{t=0,1,...}$: 連続型確率変数の列
- $\theta_0 = 1, \theta_1, \ldots$: 定数

$$X_t = \beta_0 + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_k \epsilon_k. \quad (t \geq k)$$

ただし, ϵ_t $t = 0, 1, 2, \ldots$は,

- WN1 $\mathrm{E}[\epsilon_t] = 0$
- WN1a $\mathrm{E}[\epsilon_t \epsilon_t] = \sigma^2 \quad \sigma$
- WN1b $\mathrm{E}[\epsilon_t \epsilon_s] = 0 \quad (t \neq s)$ 独立
- $\mathrm{E}[X(t)\epsilon_s] = 0 \quad (t < s)$

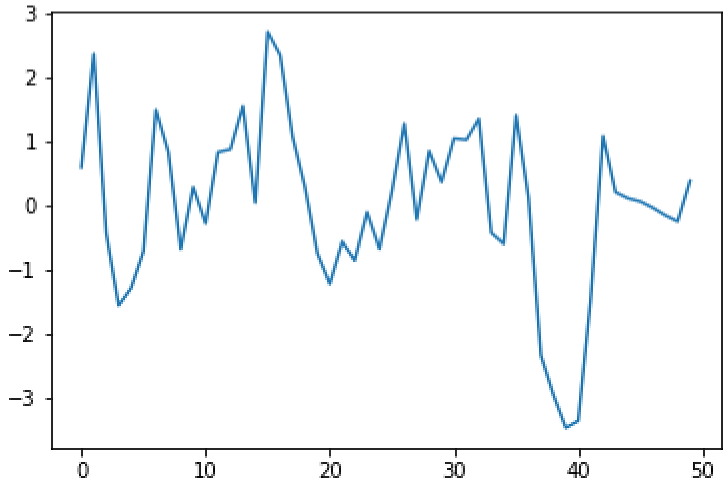
移動平均モデルMA(k)の母平均値・母共分散

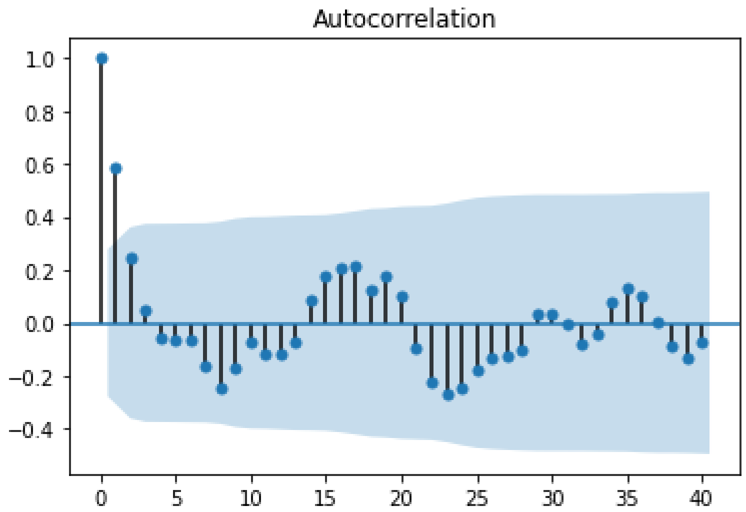
$$\mathrm{E}[X_t] = 0$$

$$\begin{aligned} \mathrm{V}[X_t] =& \mathrm{V}[\epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_k \epsilon_k] \\ =& \mathrm{V}[\epsilon_t] + \theta_1^2 \mathrm{V}[\epsilon_{t-1}] + \cdots + \theta_k^2 \mathrm{V}[\epsilon_k] \\ =& (1 + \theta_1^2 + \cdots + \theta_k^2)\sigma^2 \end{aligned}$$

$$\begin{aligned} \mathrm{Cov}[X_t, X_{t-s}] =& \mathrm{E}[(\epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_k \epsilon_{t-k})(\epsilon_{t-s} + \theta_1 \epsilon_{t-s-1} + \cdots + \theta_k \epsilon_{t-s-k})] \\ =& \cdots \\ =& \begin{cases} 0 & (k < s) \\ (\theta_0 \theta_s + \theta_1 \theta_{s+1} + \cdots + \theta_{k-s} \theta_k)\sigma^2 & (k \geq s) \end{cases} \end{aligned}$$

コレログラム





Quiz L14-2

[Quiz L14-2]

MA(k)の生成

```
from statsmodels.tsa.arima_process import ArmaProcess
ar=[1]           #
ma=[1,0.5,0.2]      # [theta_0,theta_1,theta_2]
MA=ArmaProcess(ar,ma) # 自己回帰移動平均モデルを計算するオブジェクト
x=MA.generate_sample(nsample=50) # 時間の長さ=50
```

バリエーション

- m, ℓ次の自己回帰移動平均モデルARMA(m, ℓ) 漸化式の右辺が, AR(m)とMA(ℓ)の和になっているもの.
- m, k次の自己回帰和分移動平均モデルARIMA(m, k).
  階差がARMA(m, k)にしたがう時系列. それ自身は非定常でありうる.
- ⋮