

時系列データと時系列データの解析

樋口さぶろお <https://hig3.net>

龍谷大学大学院 先端理工学研究科 数理・情報科学コース

計算科学特論 L09(2026-06-15 Mon)

最終更新: Time-stamp: "2026-08-08 Sat 07:20 JST hig"

今日の目標

- 時系列データの標本特徴量を説明できる
- 時系列モデルの例を挙げられる



L08-Q1

TA Prob and Sol: ポアソン分布

ある県では、交通死亡事故が、平均すると1日に3件起きるという。1日の事故の件数はパラメタ $\lambda = 3$ のポアソン分布に従う。

- ① 特定の1日に、交通死亡事故が0件である確率を求めよう。
- ② 特定の1日に、交通死亡事故が6件である確率を求めよう。
- ③ 1日に起きる交通死亡事故の件数の分散を求めよう。

略解

- ① $P(X = 0) = \frac{3^0}{0!} e^{-3} = e^{-3}.$
- ② $P(X = 6) = \frac{3^6}{6!} e^{-3} = \frac{81}{80} e^{-3}.$
- ③ $V[X] = 3.$

L08-Q2

TA Prob and Sol: ポアソン分布

あるサッカーチームは、90 分のゲームで平均 3 点得点できる。

- ① ハーフ 45 分間に 0 点である確率は?
- ② ハーフ (前半) 0 点 かつ ハーフ (後半) 3 点である確率は?
- ③ ゲーム 90 分で 3 点であるときに、ハーフ (前半) 0 点, ハーフ (後半) 3 点である確率は?

略解

ハーフの得点 X はパラメタ $\lambda = 1.5$ のポアソン分布にしたがう。

- ① $\frac{1.5^0}{0!} e^{-1.5} = e^{-1.5}$
- ② $\frac{1.5^0}{0!} e^{-1.5} \frac{1.5^3}{3!} e^{-1.5} = \frac{1.5^3}{6} e^{-3}.$

- ③ ゲームの得点 Y はパラメタ $\lambda = 3$ のポアソン分布にしたがう。条件付き確率を考えて,

$$\frac{P(X=0)P(X=3)}{P(X=0)P(X=3)+P(X=1)P(X=2)+P(X=2)P(X=1)+P(X=3)P(X=0)} = \frac{1}{8}.$$

$$\text{または, 再生性より, } \frac{P(X=0)P(X=3)}{P(Y=3)} = \frac{\frac{1.5^0}{0!}e^{-1.5} \frac{1.5^3}{3!}e^{-1.5}}{\frac{3^3}{3!}e^{-3}} = \frac{1}{8}.$$

L08-Q3

TA Prob and Sol: ポアソン分布の母平均値と母分散

離散型確率変数 X はポアソン分布 $\text{Po}(\lambda)$ にしたがう．すなわち，確率関数は

$$p(x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad (x = 0, 1, 2, \dots)$$

で与えられる ($\lambda > 0$ はパラメタ).

- ① $E[1] = 1$ を計算によって確かめよう．
- ② 母平均値 $E[X]$ を求めよう．
- ③ 母分散 $V[X]$ を求めよう．

略解

①

$$\begin{aligned} E[1] &= \sum_{x=0}^{\infty} 1 \cdot p(x) = \sum_{x=0}^{\infty} \frac{e^{-\lambda} \lambda^x}{x!} \\ &= e^{-\lambda} e^{\lambda} = 1. \end{aligned}$$

2

$$\begin{aligned} E[X] &= \sum_{x=0}^{\infty} x \frac{e^{-\lambda} \lambda^x}{x!} = \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \\ &= \lambda e^{-\lambda} e^{\lambda} = \lambda. \end{aligned}$$

3

$$\begin{aligned} E[X(X-1)] &= \sum_{x=0}^{\infty} x(x-1) \frac{e^{-\lambda} \lambda^x}{x!} \\ &= \lambda^2 e^{-\lambda} \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} = \lambda^2 \end{aligned}$$

となる．したがって， $X^2 = X(X-1) + X$ より，

$$\begin{aligned} V[X] &= E[X^2] - E[X]^2 \\ &= E[X(X-1)] + E[X] - E[X]^2 \\ &= \lambda^2 + \lambda - \lambda^2 = \lambda. \end{aligned}$$

L08-Q4

TA Prob and Sol: ロジスティック関数とロジット関数

ロジスティック関数 $p = \frac{1}{1+e^{-z}}$ を考える.

- ① 逆関数がロジット関数 $z = \text{logit}(p) = \log \frac{p}{1-p}$ であることを示そう
- ② ロジット関数グラフを描こう.

略解

- ① $0 < p < 1$ とする.

$$\begin{aligned} p = \frac{1}{1 + e^{-z}} &\iff \frac{1}{p} = 1 + e^{-z} \\ &\iff \frac{1-p}{p} = e^{-z} \\ &\iff z = \log \frac{p}{1-p}. \end{aligned}$$

したがって、逆関数は

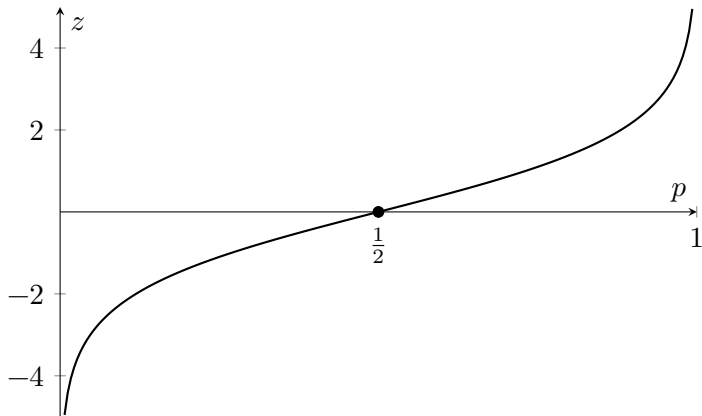
$$\text{logit}(p) = \log \frac{p}{1-p}$$

である.

- ② ロジット関数の定義域は $0 < p < 1$ で,

$$\lim_{p \rightarrow 0+} \text{logit}(p) = -\infty, \quad \text{logit}\left(\frac{1}{2}\right) = 0, \quad \lim_{p \rightarrow 1-} \text{logit}(p) = +\infty$$

となる. また, 単調増加である.



L08-Q5

TA Prob and Sol: 二項分布のポアソン分布への収束

二項分布 $B(n, p)$ は、極限 $n \rightarrow +\infty$, $np = \lambda$ (一定) でポアソン分布 $Po(\lambda)$ になることを示そう。

略解

$X_n \sim B(n, p)$ とし、 $p = \lambda/n$ とおく。固定した非負整数 x に対して、

$$\begin{aligned} P(X_n = x) &= \frac{n!}{x!(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\ &= \frac{\lambda^x}{x!} \frac{n(n-1)\cdots(n-x+1)}{n^x} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x}. \end{aligned}$$

ここで、 $n \rightarrow +\infty$ のとき、

$$\begin{aligned}\frac{n(n-1)\cdots(n-x+1)}{n^x} &= \prod_{j=0}^{x-1} \left(1 - \frac{j}{n}\right) \longrightarrow 1, \\ \left(1 - \frac{\lambda}{n}\right)^n &\longrightarrow e^{-\lambda}, \\ \left(1 - \frac{\lambda}{n}\right)^{-x} &\longrightarrow 1\end{aligned}$$

だから、

$$\lim_{n \rightarrow +\infty} P(X_n = x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad (x = 0, 1, 2, \dots).$$

右辺はポアソン分布 $\text{Po}(\lambda)$ の確率関数なので、この極限で二項分布 $B(n, p)$ はポアソン分布 $\text{Po}(\lambda)$ に収束する。

ここまで来たよ

8 一般化線形モデルとベイズ推定

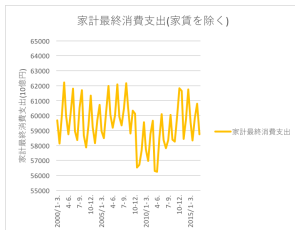
9 時系列データと時系列データの解析

- 時系列データ
 - 時系列データの標本特徴量
 - 時系列モデルの例：ランダムウォークモデル
 - 時系列モデルの例：自己回帰モデル $AR(m)$

時系列解析 Time Series Analysis

時系列 時間 t に依存する量の列 $x(0), x(1), x(2), \dots, x(t), \dots$
 以前の値が今の値に影響. $x(t)$ $t = 0, 1, 2, 3, \dots$ は独立でない.
 例

- 特定の銘柄の毎日の株価のデータ
- 週ごとの売上のデータ
- 1分おきの気温のデータ
- 1年ごとの太陽黒点の個数のデータ
- 時刻 t のランダムウォーカーの座標 $X(t)$



母ナントカ 時系列モデル $\simeq$ 確率過程

標本ナントカ 時系列データ (を時系列解析)

時系列解析 Time-series analysis 時系列データを解析する手法群. 経済統計学でさかん.

目的 時系列を再現する. $t \leq T$ のデータから $t > T$ を予測する.
 時系列データを解析 $\rightarrow$ 時系列モデルを作成 $\rightarrow$ 再現・予測

時系列データの形式

多変量時系列データ $\mathbf{x}(t)$.

日時	気温	気圧
2021-12-18 Sat 10:00	10.1	1010
2021-12-18 Sat 11:00	10.5	1005
⋮	⋮	⋮

時間等間隔が前提で、具体的な日時でなく通し番号が使われることも多い.

日時には実質的な意味がある. 日時を消して順序を変えてはいけない.

日時 i のデータは日時 $j (< i)$ のデータの影響を受けることがある.

多変量データ

便宜的通し番号	100m 走	棒高跳び
1	10.1	2.30
2	10.4	2.12
⋮	⋮	⋮

(同じ分布からとってきたデータ点かもしれないが) 選手 i のデータが選手 j のデータの影響を受けるようなことはない.

以下では, 等間隔, 日時順に並んでいるデータだとして, 日時 **index** を省いて考える

数理のカリキュラムで見る，時系列データ分析手法

以下の表は，「ここで学ぶ」という意味でなく，「時系列解析のこの概念は，多変量解析で言えばこの辺りで学ぶことに相当」的なこと

学年	主流の科目	時系列解析
学部 1	データ分析	標本移動平均，標本自己相関
学部 2	確率統計	
学部 2	多変量解析	フーリエ級数，マルコフ連鎖，自己回帰モデル，移動平均モデル，
学部 3	機械学習	状態空間モデル
学部 4		
大学院		ディープラーニング，RNN(GRU,LSTM), seq2seq, Transformer

時系列解析，時系列モデルの応用

- 確率的に時間発展する現象
 - ▶ ランダムウォーク，天気，株価，経済データ，学習教育データ
- 時間とは限らない順序あり系列一般（後が前の影響を受ける）
 - ▶ 自然言語処理

順序なし系列の手法も使える

確率過程

$X(t)$, $t = 0, 1, 2, \dots$ を, 状態空間 $S \subset \mathbb{R}$ に値をとる確率変数とする. t は確率変数につけた番号 $X_1, X_2, \dots$. '離散時間' パラメタ $t \in \mathbb{Z}$. これを**確率過程** stochastic process という.

$X_1, X_2, \dots$ というと独立同分布や標本を想像するが, ここでは, 互いに独立でないもの, 特に $X(t+1)$ が $X(t')$ ($t' \leq t$) の影響を受けるようなものに興味がある

X が離散型だとして, $X(t)$ の確率関数を, $P(X(t) = x) = p(x, t)$ と略記する. このとき, $\forall t \quad \sum_x p(x, t) = 1$.

例

- 猫の状態 $x = 1, 2, 3 = n$. $\leftrightarrow$ 食う, 寝る, 遊ぶの状態
- 株価
- 気温

確率モデル

多変量解析

ここまで来たよ

8 一般化線形モデルとベイズ推定

9 時系列データと時系列データの解析

- 時系列データ
- 時系列データの標本特徴量
- 時系列モデルの例：ランダムウォークモデル
- 時系列モデルの例：自己回帰モデル $AR(m)$

時系列の3要素

現実の時系列データは次の3つの重ね合わせになっていることが多い

$$x(t) = \boxed{\text{トレンド}} + \boxed{\text{周期的変動}} + \boxed{\text{ランダム成分}}$$

- **トレンド**（長期的傾向）期間を通して時間に比例して増減する傾向.
あるいは増減しない定数値.
- 短い周期の**周期的変動** 季節, 週, 月, 年

フーリエ級数解析

- **ランダム成分**（ノイズ）時刻ごとに独立な確率変数の実現値

移動平均 Moving Average

時系列 $x(t)$ から平滑化 (smoothing) した別の時系列 $y(t)$ を作る手法

$(2\ell + 1)$ 次の移動平均

$$y_{2\ell+1}(t) = \frac{1}{2\ell + 1} \sum_{t'=t-\ell}^{t+\ell} x(t')$$

$x(0) \quad x(1) \quad x(2) \quad x(3) \quad x(4) \quad x(5) \quad x(6) \quad x(7) \quad x(8) \quad x(9) \quad x(10) \quad \dots$

2ℓ 次の移動平均

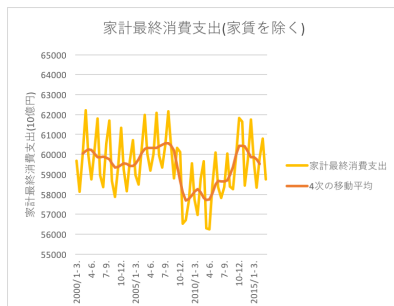
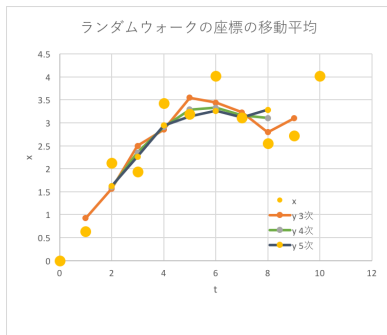
$$y_{2\ell}(t) = \frac{1}{2\ell} \left(\frac{1}{2} \cdot x(t-\ell+1) + x(t-\ell+2) + \dots + x(t) + \dots + x(t+\ell-2) + \frac{1}{2} x(t+\ell-1) \right)$$

$x(0) \quad x(1) \quad x(2) \quad x(3) \quad x(4) \quad x(5) \quad x(6) \quad x(7) \quad x(8) \quad x(9) \quad x(10) \quad \dots$

離散フーリエ級数の第 0 項

移動平均の性質

- 次数が高くなるほど滑らかになる
- 元のデータの「真ん中へん」を通る.
- 時間帯の端までは描けない



移動平均の性質 2

- **トレンド** (長期的傾向) → 移動平均ではっきり見えるようになる
- 短い周期の**周期的変動** → (周期くらいの) 移動平均で消える. (もとのデータ)-(移動平均) ではっきり見える
- **ランダム成分** (ノイズ) → 中心極限定理により, 移動平均で小さくなる

L09-Q1

Quiz(移動平均)

次の時系列データから, 3,4 次の移動平均を求めよう.

t	0	1	2	3	4	5	6	7	8	9
x	1.8	-1.6	2.6	-1.2	3.0	-1.2	3.4	-0.8	3.8	0.0
y_3										
y_4										

離散フーリエ変換

データ $\{X_t\}_{t=0,1,\dots,L-1}$ の周期性を知るには、**離散フーリエ変換**が使える。
離散フーリエ変換

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

で、 $\frac{1}{L} \cdot |\tilde{X}_s|$ は周期 L/s 成分の寄与の大きさ。

フーリエ解析

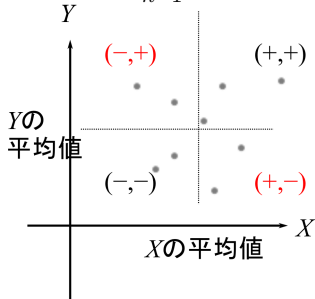
- **トレンド**（長期的傾向）期間を通して時間に比例して増減する傾向。
あるいは増減しない定数値。
- 短い周期の**周期的変動** 季節, 週, 月, 年
- **ランダム成分**（ノイズ）がないかのように考えている

復習: 標本共分散と標本相関係数

標本共分散 (covariance)

$$x, y \text{ の共分散 } C_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}) \times (y_i - \bar{y})$$

(不偏なら $\frac{1}{n-1}$)



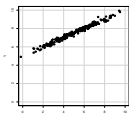
$(+, -) = (x_i - \bar{x} \text{ の符号}, y_i - \bar{y} \text{ の符号})$.

相関係数

$$\text{標本相関係数 } r = \frac{C_{xy}}{s_x s_y}$$

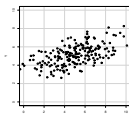
s_x, s_y : 標本標準偏差

分子分母で $\frac{1}{n}$, 不偏 $\frac{1}{n-1}$ をそろえる.



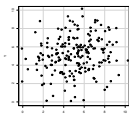
強い正の相関

$$r = 0.99$$



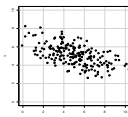
弱い正の相関

$$r = 0.55$$



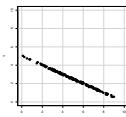
無相関

$$r = 0$$



弱い負の相関

$$r = -0.55$$



強い負の相関

$$r = -0.99$$

相関

‘正の相関’: x が大きい $\Leftrightarrow y$ が**大きい**

‘負の相関’: x が大きい $\Leftrightarrow y$ が**小さい**

強い/弱い: 傾向がはっきりしている/していない

標本自己共分散, 標本自己相関係数

k 次の標本自己共分散・標本自己相関係数

時間 t を, **ラグ** k だけずらした $x(t), x(t-k)$ を 2 変量データだと思って, 標本共分散・標本自己相関係数を考えたもの

$k = 1$ の例

x	y
$x(1)$	—
$x(2)$	$x(2-1)$
$x(3)$	$x(3-1)$
$\vdots$	$\vdots$
$x(T-1)$	$x(T-1-1)$
$x(T)$	$x(T-1)$
—	$x(T)$

k の例

x	y
$x(1)$	—
$\vdots$	
$x(k+1)$	$x(1)$
$\vdots$	$\vdots$
$x(T)$	$x(T-k)$
$\vdots$	
—	$x(T)$

定常な時系列データに対する標本ナントカ

‘時系列分析’では、定常過程について、1個の時系列データを、一定の長さに分割して複数個のデータからなる標本のように扱うことが行われる。

横: t 縦: 標本内のデータ番号. 標本自己相関係数を求めるとき

	A	B	C	D	E	F	G	H	I	J
1	n \ t	0	1	2	3	4	5	6	7	8
2	1	2	-0.2522	-1.4293	0.41339	-0.4521	-0.3585	0.05638	1.16461	0.77765
3	2	2	0.40156	-0.1803	1.02945	-0.1411	-0.1659	1.1405	0.1365	0.27561
4	3	2	1.74901	0.20238	-0.7695	-0.6431	-0.1627	0.66943	0.10917	1.30642
5	4	2	0.45349	-0.3762	-0.6185	-0.1757	-0.461	-1.4453	0.09464	1.55283
6	5	2	1.01492	0.94089	-0.5959	-0.8685	-0.30052	0.5173	0.11447	1.38334
7	6	2	1.78265	1.70528	2.07552	-0.06083	0.429	0.10902	-0.7164	-1.1645
8	7	2	0.69062	1.27761	1.15377	1.44955	-0.4663	0.24885	-1.323	-1.1454
9	8	2	0.74554	0.27013	1.05844	-0.1876	0.06592	-0.0232	1.09015	1.91198
10	9	2	1.21618	-0.2882	0.81639	1.84743	0.18484	-1.1622	-1.5981	-2.3211
11	10	2	1.20421	-0.487	-0.2274	0.34252	0.62605	1.46621	0.13736	0.13426

本当はこういう標本が欲しい

	A	B	C	D	E	F	G	H	I	J
1	n \ t	0	1	2	3	4	5	6	7	8
2	1	2	-0.2522	-1.4293	0.41339	-0.4521	-0.3585	0.05638	1.16461	0.77765
3	2	2	0.40156	-0.1803	1.02945	1.31571	-0.3059	1.1405	0.1365	0.27561
4	3	2	1.74901	0.20238	-0.7695	-1.3943	-0.0067	0.66943	0.10917	1.30642
5	4	2	0.45349	-0.3762	-0.6185	-1.7157	-0.461	-1.4453	0.09464	1.55283
6	5	2	1.01492	0.94089	-0.5959	-0.8685	0.30052	0.5173	0.11447	1.38334
7	6	2	1.78265	1.70528	2.07552	0.06083	0.429	0.10902	-0.7164	-1.1645
8	7	2	0.69062	1.27761	1.15377	1.44955	-0.4663	0.24885	-1.323	-1.1454
9	8	2	0.74554	0.27013	1.05844	-0.1876	0.06592	-0.0232	1.09015	1.91198
10	9	2	1.21618	-0.2882	0.81639	1.84743	0.18484	-1.1622	-1.5981	-2.3211
11	10	2	1.20421	-0.487	-0.2274	0.34252	0.62605	1.46621	0.13736	0.13426

定常ならこれでもいいじゃん

	A	B	C	D	E	F	G	H	I	J
1	n \ t	0	1	2	3	4	5	6	7	8
2	1	2	-0.2522	-1.4293	0.41339	0.93494	1.350	0.9	0.6	0.7
3	2	2	0.40156	-0.1803	1.02945	1.31571	-0.3059	1.1405	0.1365	0.27561
4	3	2	1.74901	0.20238	-0.7695	-1.3943	-0.0067	0.66943	0.10917	1.30642
5	4	2	0.45349	-0.3762	-0.6185	-1.7157	-0.461	-1.4453	0.09464	1.55283
6	5	2	1.01492	0.94089	-0.5959	-0.8685	0.30052	0.5173	0.11447	1.38334
7	6	2	1.78265	1.70528	2.07552	0.06083	0.429	0.10902	-0.7164	-1.1645
8	7	2	0.69062	1.27761	1.15377	1.44955	-0.4663	0.24885	-1.323	-1.1454
9	8	2	0.74554	0.27013	1.05844	-0.1876	0.06592	-0.0232	1.09015	1.91198
10	9	2	1.21618	-0.2882	0.81639	1.84743	0.18484	-1.1622	-1.5981	-2.3211
11	10	2	1.20421	-0.487	-0.2274	0.34252	0.62605	1.46621	0.13736	0.13426

Excel 的にはこうやると楽

k 次の標本自己共分散 autocovariance

$$C(k) = \frac{1}{T-k} \sum_{t=k+1}^T (x(t) - \bar{x})(x(t-k) - \bar{x})$$

$$\text{ただし標本平均値 } \bar{x} = \frac{1}{T} \sum_{t=1}^T x(t)$$

k 次の標本自己相関係数（標本自己相関関数） autocorrelation

$$r(k) = \frac{\text{自己共分散}}{\sqrt{\text{分散}}\sqrt{\text{分散}}} = \frac{C(k)}{C(0)}$$

$C(0)$ は $x(t)$ をサイズ T の標本と思ったときの分散。

コレログラム correlogram

横軸 ラグ k , 縦軸 k 次の自己相関係数 のグラフ.

- 多くの定常モデル (自己回帰モデルなど) では, k が大きいほど (遠い時刻ほど), 標本自己相関係数 $r(k)$ の絶対値は小さくなる.

Python tsa=Time Series Analysis 時系列解析

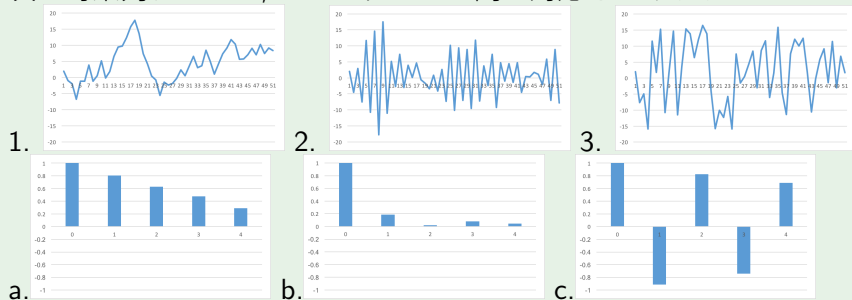
自己相関係数 `statmodels.tsa.stattools.acf(df['x'])`

コレログラム `statmodels.graphics.tsa.plot_acf(df['x'])`

L09-Q2

Quiz(コレログラム)

次の時系列データと、コレログラムの間の対応をつけよう。



ここまで来たよ

8 一般化線形モデルとベイズ推定

9 時系列データと時系列データの解析

- 時系列データ
- 時系列データの標本特徴量
- 時系列モデルの例：ランダムウォークモデル
- 時系列モデルの例：自己回帰モデル $AR(m)$

ランダムウォークモデル

定義 (ランダムウォークモデル)

$t = 0, 1, 2, \dots$: 時刻

$\{\epsilon_t\}_{t=1, \dots}, \{X_t\}_{t=0, 1, \dots}$: 確率変数の列. 時刻 t における位置 $X_t \in \mathbb{R}$.
 $X_t (t \geq 1)$ は次のように定まる.

$$X_{t+1} = X_t + \epsilon_{t+1}. \quad (t \geq 1)$$

ただし ϵ_t は次を満たす.

$$E[\epsilon_t] = 0 \quad (\text{WN1})$$

$$E[\epsilon_t \epsilon_s] = \sigma^2 \delta_{ts} \quad (\text{WN2})$$

$$E[X_t \epsilon_s] = 0 \quad (t < s) \quad (\text{PI})$$

たとえば, $\frac{1}{2}(\epsilon_t + 1) \sim B(1, \frac{1}{2})$, i.i.d. なら, $\mathbb{Z}$ 上のランダムウォークで, 推移確率が

$$P(X_{t+1} = y \pm 1 | X_t = y) = \frac{1}{2}.$$

たとえば, $\epsilon_t \sim N(0, \sigma^2)$, i.i.d. なら, $\mathbb{R}$ 上のランダムウォークで, 推移確率**密度**が,

$$P(X_{t+1} = x | X_t = y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-x)^2}{2\sigma^2}}.$$

定義 (白色雑音 (white noise))

WN1, WN2 を満たす ϵ_t を **白色雑音** (white noise) という.

形容詞 **white** は特徴がないことを表現.

たとえば $\epsilon_t \sim N(0, \sigma^2)$ (独立同分布) なら WN1, WN2 を満たす.

分布を具体的に指定せず, 母期待値のみを制限する弱い条件になっている. 例えば $E[\epsilon_t^4]$ が様々な値である可能性を残している.

PI はジャンプ幅が場所によらないという条件 (マルコフ連鎖ですべての状態が同等).

ランダムウォークの漸化式の計算

$$\begin{aligned}X_t &= X_{t-1} + \epsilon_t \\&= X_{t-2} + \epsilon_{t-1} + \epsilon_t \\&= \cdots \\&= X_0 + \epsilon_1 + \epsilon_2 + \cdots + \epsilon_t\end{aligned}$$

ランダムウォークの母平均値, 母分散

$$\begin{aligned} E[X_t] &= E[X_0 + \sum_{k=1}^t \epsilon_k] \\ &= E[X_0] \end{aligned}$$

時間とともに変化しない.

$$\begin{aligned} V[X_t] &= V[X_0 + \sum_{k=1}^t \epsilon_k] \\ &= V[X_0] + \sum_{k=1}^t V[\epsilon_k] \quad (\text{WN2,PI}) \\ &= V[X_0] + \sum_{k=1}^t \sigma^2 \quad (\text{WN2}) \\ &= V[X_0] + t\sigma^2 \end{aligned}$$

時間に比例して分散が大きくなる.

ここまで来たよ

8 一般化線形モデルとベイズ推定

9 時系列データと時系列データの解析

- 時系列データ
- 時系列データの標本特徴量
- 時系列モデルの例：ランダムウォークモデル
- 時系列モデルの例：自己回帰モデル $AR(m)$

1 次の自己回帰モデル $AR(1)$

確率過程の単純な例

定義 (1 次の自己回帰モデル $AR(1)$)

$t = 0, 1, 2, \dots$: 時刻

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

$\phi_1 \in \mathbb{R}$: パラメタ

$X_t (t \geq 0)$ は次のように定まる.

$$X_{t+1} = \phi_1 \cdot X_t + \epsilon_{t+1}. \quad (t \geq 0)$$

ただし ϵ_t は WN1, WN2, PI を満たす.

例

- $t = \text{シーズン}$, $X_t = \text{タイガースの勝率の 5 割との差}$, $-1 < \phi < 0$.
- ランダムウォーク
- $t = \text{日}$, $X_t = \text{気温の, 年平均気温との差}$, $0 < \phi < 1$.

AR(1) の漸化式の計算

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

AR(1) の母平均値, 母分散

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

$$\begin{aligned} V[X_t] &= V[\phi_1^t X_0 + \sum_{k=0}^{t-1} \phi_1^k \epsilon_{t-k}] \\ &= V[\phi_1^t X_0] + \sum_{k=0}^{t-1} V[\phi_1^k \epsilon_{t-k}] \quad (\text{WN2, PI}) \\ &= \phi_1^{2t} V[X_0] + \sum_{k=0}^{t-1} \phi_1^{2k} \sigma^2 \quad (\text{WN2}) \\ &= \phi_1^{2t} V[X_0] + \frac{1 - \phi_1^{2t}}{1 - \phi_1^2} \sigma^2 \quad (|\phi_1| < 1 \text{ のとき}) \end{aligned}$$

$|\phi_1| < 1$ ならば, $t \rightarrow +\infty$ で, $E[X_t] \rightarrow 0$, $V[X_t] \rightarrow \frac{1}{1-\phi_1^2} \sigma^2$ となり時間によらなくなる可能性. 以下, 簡単のため $E[X_0] = E[X_t] = 0$ と仮定.

時系列モデルの自己共分散

定義 (母自己共分散, 母自己相関係数)

時系列 $\{X_t\}$ に対して, s 次の

- 母自己共分散 (auto covariance) $\text{Cov}[X_t, X_{t-s}]$.
- 母自己相関係数

$$(\text{autocorrelation}) r_t(s) = \frac{\text{Cov}[X_t, X_{t-s}]}{\sqrt{V[X_t]V[X_{t-s}]}} \stackrel{\text{定常なとき}}{=} \frac{\text{Cov}[X_t, X_{t-s}]}{\text{Cov}[X_t, X_t]}.$$

$s = 0$ 次の母自己共分散はふつうの母分散, 母自己相関係数は 1.
 s を **ラグ** (lag) という.

定義 (k 次のインパルス応答関数)

$$\delta_k = \frac{\text{Cov}[\epsilon_t, X_{t+k}]}{V[\epsilon_t]}$$

時系列の定常性

定義 (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** (stationary) であるとは、次の両方を満たすこと。

- 母平均値 $E[X_t]$ が t によらないこと。
 - 各ラグ s に対して、 s 次の母**自己共分散** (auto covariance) $\text{Cov}[X_t, X_{t-s}]$ 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 $\Leftrightarrow$ 強い意味での定常性)

AR(1) の母自己共分散

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

展開すると 0 にならないのは, $E[X_0 X_0]$, $E[\epsilon_k \epsilon_k] = \sigma^2$ のみ.

$$\begin{aligned}\text{Cov}[X_t, X_{t-s}] &= \phi_1^{2t-s} E[X_0 X_0] + (\phi_1^{2t-2s-2} + \phi_1^{2t-2s-4} + \dots + \phi_1^0) \phi_1^s \sigma^2 \\ &= \phi_1^{2t-s} E[X_0 X_0] + \frac{1 - (\phi_1^2)^{t-s}}{1 - \phi_1^2} \phi_1^s \sigma^2 \\ &\rightarrow \begin{cases} \frac{\phi_1^s \sigma^2}{1 - \phi_1^2} & (|\phi_1| < 1) \\ \approx t & (|\phi_1| = 1) \\ \approx t \phi_1^{2t} & (|\phi_1| > 1) \end{cases} \quad \begin{array}{l} \\ \text{1 行前を修正} \\ \end{array}\end{aligned}$$

$|\phi_1| < 1$ のとき, $t \rightarrow +\infty$ で定常.

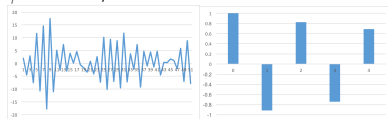
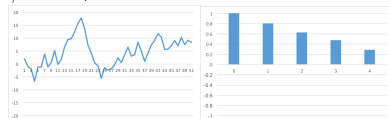
このとき s 次の母自己相関係数は $r(s) = \frac{\phi_1^s \sigma^2}{1 - \phi_1^2}$ ($|\phi_1| < 1$) つまり等比数列.

AR(1) モデルの母自己相関係数

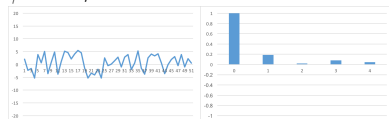
k 次の母自己相関係数

$$r(s) = \frac{\text{Cov}[X_t, X_{t-s}]}{\sqrt{V[X_t]}\sqrt{V[X_{t-s}]}} \stackrel{\text{定常}}{=} \frac{E[X_t X_{t-s}]}{E[X_t X_t]} = \phi_1^s.$$

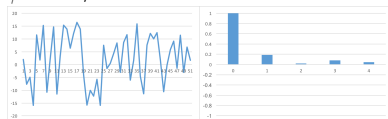
サンプルパスの t - X_t グラフと、下で説明する方法で推定した s - $r(s)$ グラフ (コレログラム)
 $\phi = 0.9, \sigma = 1$
 $\phi = -0.9, \sigma = 1$



$\phi = 0.2, \sigma = 1$



$\phi = 0.2, \sigma = 3$



L09-Q3

Quiz(自己回帰モデル $AR(2)$ の例)

2 次の自己回帰モデル $AR(2)$ モデルを考える (授業中に使った記号で). 初期条件として, X_0, X_1 は定数 0 に等しいとしてよい ($P(X_0 = 0) = P(X_1 = 0) = 1$).

- ① 確率変数 X_2 を, ϵ_t と ϕ_k で表そう.
- ② 母期待値 $E[X_3^2]$ を σ と ϕ_k で表そう.
- ③ 母自己共分散 $Cov[X_3, X_4]$ を σ と ϕ_k で表そう.