

時系列解析

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 L12(2025-12-10 Wed)

最終更新: Time-stamp: "2025-12-10 Wed 08:09 JST hig"

今日の目標

- 時系列データを説明できる
- ランダムウォーク, $AR(m)$ モデルを説明できる
- 移動平均, 自己共分散, 自己相関係数を説明・計算できる



L11-Q1

Quiz 解答: マルコフ連鎖の時間発展

- ① $\vec{p}(1) = M\vec{p}(0) = \frac{1}{3} \begin{pmatrix} 1 \\ 2 \end{pmatrix}$.
- ② 転置推移確率行列 M の固有値 1 の固有ベクトル $\vec{u}$ を (あるなら) 求めればよい. $M\vec{u} = \vec{u}$ を解いて, $\vec{u} = \begin{pmatrix} 3 \\ 4 \end{pmatrix} s$ ($s \neq 0$). 定常分布は, 規格化された $\vec{u} = \frac{1}{7} \begin{pmatrix} 3 \\ 4 \end{pmatrix}$.
- ③ 転置推移確率行列 M の固有値 (絶対値の大きさの順に) λ_1, λ_2 , 対応する固有ベクトル $\vec{u}_1, \vec{u}_2$ を求めると,

$$\lambda_1 = 1, \lambda_2 = -\frac{1}{6}, \quad \vec{u}_1 = \begin{pmatrix} 3 \\ 4 \end{pmatrix} s, \vec{u}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} s \quad (s \neq 0).$$

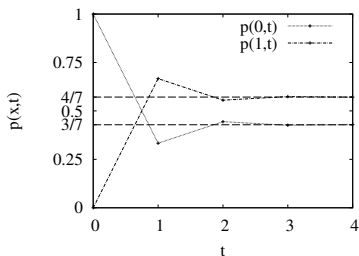
固有方程式には解 $\lambda_1 = 1$ があることが最初からわかってるから, 因数分解は楽.

$\vec{u}_1, \vec{u}_2$ とも $s = 1$ に固定する (他の取り方をしても最終的には同じ $\vec{p}(t)$ が求まる). このとき, $\vec{p}(0) = a\vec{u}_1 + b\vec{u}_2$ で係数 a, b を決めると, $a = \frac{1}{7}, b = \frac{4}{7}$.

$$\begin{aligned}\vec{p}(t) &= M^t \vec{p}(0) \\ &= M^t (a\vec{u}_1 + b\vec{u}_2) \\ &= \lambda_1^t a\vec{u}_1 + \lambda_2^t b\vec{u}_2 \\ &= \frac{1}{7} \begin{pmatrix} 3 \\ 4 \end{pmatrix} + \frac{4}{7} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \left(-\frac{1}{6}\right)^t.\end{aligned}$$

$\vec{u}_1$ の係数が $\frac{1}{7}$ である (=この項が確率ベクトルである) ので, $t \rightarrow +\infty$ でも $\vec{p}(t)$ が確率ベクトルでありつづけることが確認できる. (固有値が $\lambda = 1, -\frac{1}{6}$ となった時点で, 第 1 項が確率ベクトルになるはず, ということから, 1 次方程式を解かずに $a = \frac{1}{7}$ と決められたはずだった).

時間変化. $|\lambda_2| = \frac{1}{6} < 1$ なので, $\vec{p}(t) \rightarrow \frac{1}{7} \begin{pmatrix} 3 \\ 4 \end{pmatrix} \quad (t \rightarrow +\infty)$

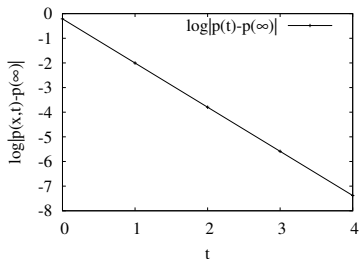


- ④ 極限分布 $\vec{p}(\infty) = \frac{1}{7} \begin{pmatrix} 3 \\ 4 \end{pmatrix}$ との差の大きさ (ベクトルの長さ) を考えると,

$$\begin{aligned} |\vec{p}(t) - \vec{p}(\infty)| &= \left| \frac{4}{7} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \left(-\frac{1}{6}\right)^t \right| \\ &= \frac{4\sqrt{2}}{7} \left(\frac{1}{6}\right)^t \end{aligned}$$

$$\log |\vec{p}(t) - \vec{p}(\infty)| = t \times \log \frac{1}{6} + \log \frac{4\sqrt{2}}{7}$$

よって, t 対 $\log |\text{差}|$ のグラフを描くと, 傾き $\log |\text{第2固有値}| (< 0)$ の直線になるはず.



状態空間が3点以上からなり、(第2固有値より絶対値が小さい) 第3固有値以降がある場合も、 t が大きいところではこの直線に近づいていく。

さらに t を大きくすると、 $|\vec{p}(t) - \vec{p}(\infty)|$ が桁落ちしたり、数値的に0とみなされるなどで、直線から外れる。

5

$$\vec{p}(t) = \frac{1}{7} \begin{pmatrix} 3 \\ 4 \end{pmatrix} + \frac{1}{14} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \left(-\frac{1}{6}\right)^t$$

L11-Q2

Quiz 解答: マルコフ連鎖の定常分布と時間発展

- ① 固有値 $\lambda = 1$ の固有ベクトルである確率ベクトルは $\frac{1}{3} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ のみであり、これが唯一の定常分布

②

$$\vec{p}(t) = \frac{1}{3} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \frac{1}{6} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \left(\frac{3}{4}\right)^t + \frac{1}{6} \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \left(\frac{1}{4}\right)^t$$

- ③ 式でまともに計算すると、固有ベクトルの内積 $\vec{u}_2 \cdot \vec{u}_3$ とか出てきてたいへん、だけど、 $t \rightarrow +\infty$ では、 $\vec{p}(t) - \vec{p}(\infty)$ の第3項が(第2項と比べて)無視できて、

$$\begin{aligned} \log |\vec{p}(t) - \vec{p}(\infty)| &= \log \left| -\frac{1}{6} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \left(\frac{3}{4}\right)^t \right| + \text{もっと小さい項} \\ &= t \log \frac{3}{4} + \log \frac{\sqrt{2}}{6} + \text{もっと小さい項} \end{aligned}$$

となる.

L11-Q3 L11-Q4

Quiz 解答: 可約なマルコフ連鎖の定常分布と時間発展

固有値 $\lambda = 1$ (重解) に対応する固有ベクトルは,

$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} s + \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} k$ ($s \neq 0$ または $k \neq 0$). 固有値 1 なので, 線形独立な確率ベクトルを 1 組選ぶと便利 (他の選び方でも最終的な結果は変わらない) で, $\vec{u}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\vec{u}_2 = \frac{1}{2} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$ とすると, $s\vec{u}_1 + (1-s)\vec{u}_2$ ($0 \leq s \leq 1$) は定常分布.

固有値 $\lambda = \frac{1}{3}$ に対する固有ベクトルは, $\begin{pmatrix} 0 \\ +1 \\ -1 \end{pmatrix} s$ ($s \neq 0$). 確率ベクトルにはなりえないので, 適当に非零ベクトルをひとつ選んで (他の選び方でも最終的な結果は変わらない), $\vec{u}_3 = \begin{pmatrix} 0 \\ +1 \\ -1 \end{pmatrix}$.

$\vec{p}(0)$ が一般の場合の時間発展は

$$\vec{p}(t) = a\vec{u}_1 1^t + b\vec{u}_2 1^t + c\vec{u}_3 \cdot \left(\frac{1}{3}\right)^t.$$

- ① $\vec{p}(0) = \frac{1}{2} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ から a, b, c を定めて,

$$\vec{p}(t) = \frac{1}{2}\vec{u}_1 + \frac{1}{2}\vec{u}_2 + \frac{1}{4}\vec{u}_3 \cdot \left(\frac{1}{3}\right)^t$$

極限分布 $\vec{p}(\infty) = \frac{1}{4} \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$ は定常分布のひとつ.

- ② $\vec{p}(0) = \frac{1}{3} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ から a, b, c を定めて,

$$\vec{p}(t) = \frac{1}{3}\vec{u}_1 + \frac{2}{3}\vec{u}_2 + 0\vec{u}_3 \cdot \left(\frac{1}{3}\right)^t$$

極限分布 $\vec{p}(\infty) = \frac{1}{3} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ は定常分布のひとつ.

- ③ $\{0\}$ と $\{1, 2\}$ が分かれた推移図. すなわち, 可約なマルコフ連鎖である.

L11-Q5

Quiz 解答: 周期的な状態のあるマルコフ連鎖の定常分布と時間発展

- ① 推移確率行列 T の固有値 λ , 固有ベクトルを求めると,

$$\lambda = 1, \omega, \omega^2$$

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} s, \begin{pmatrix} 1 \\ \omega \end{pmatrix} s, \begin{pmatrix} 1 \\ \omega^2 \end{pmatrix} s \quad (s \neq 0)$$

定常分布は $\lambda = 1$ に対する固有ベクトルで, 確率ベクトルになるように s を定めると, $\vec{u}_1 = \frac{1}{3} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$.

- ② 他の固有値 $\lambda = \omega, \omega^2$ は, $|\omega| = |\omega^2| = 1$ を満たす. よって, 一般には極限分布は存在しない.

極限分布が存在するのは, $\vec{p}(0) = \vec{u}_1 = \vec{p}(t)$ のように最初から定常分布であったときに限られる.

- ③ 3 個の状態が 1 方向に回るような推移図.

ここまで来たよ

11 マルコフ連鎖の時間発展・極限分布・周期的・可約なマルコフ連鎖

12 時系列解析

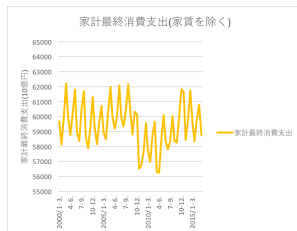
- 時系列データ
- 時系列データの分析方法 (1) 移動平均と離散フーリエ解析
- ランダムウォークモデル
- 自己回帰モデル $AR(m)$
- 時系列データの分析方法 (2) 標本自己相関係数

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．

時系列の3要素

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

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

ここまで来たよ

11 マルコフ連鎖の時間発展・極限分布・周期的・可約なマルコフ連鎖

12 時系列解析

- 時系列データ
- 時系列データの分析方法 (1) 移動平均と離散フーリエ解析
- ランダムウォークモデル
- 自己回帰モデル $AR(m)$
- 時系列データの分析方法 (2) 標本自己相関係数

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

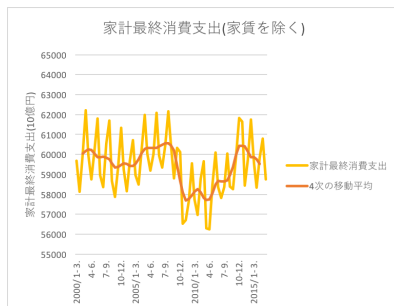
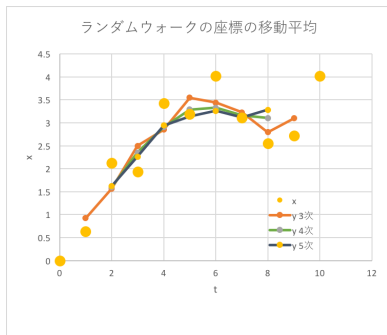
$$y_{2\ell}(t) = \frac{1}{2\ell} \left(\frac{1}{2} \cdot x(t-\ell+1) + x(t-\ell+2) + \dots + x(t) + \dots + x(t+\ell-2) + \frac{1}{2} x(t+\ell-1) \right)$$

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

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

L12-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）ないかのように考えている

ここまで来たよ

11 マルコフ連鎖の時間発展・極限分布・周期的・可約なマルコフ連鎖

12 時系列解析

- 時系列データ
- 時系列データの分析方法 (1) 移動平均と離散フーリエ解析
- ランダムウォークモデル
- 自己回帰モデル $AR(m)$
- 時系列データの分析方法 (2) 標本自己相関係数

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.

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

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

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

これはマルコフ連鎖, マルコフ過程.

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

11 マルコフ連鎖の時間発展・極限分布・周期的・可約なマルコフ連鎖

12 時系列解析

- 時系列データ
- 時系列データの分析方法 (1) 移動平均と離散フーリエ解析
- ランダムウォークモデル
- 自己回帰モデル $AR(m)$
- 時系列データの分析方法 (2) 標本自己相関係数

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

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

$s = 0$ 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)

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

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

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

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

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

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

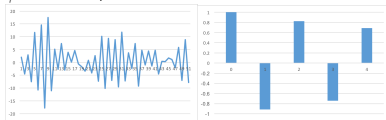
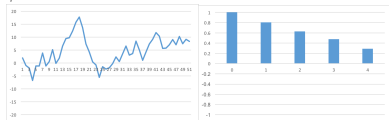
このとき s 次の母自己相関係数は $r(s) = \frac{\phi_1^s \sigma^2}{1 - \phi_1^2}$ ($|\phi_1| < 1$) つまり等比数列.

AR(1) モデルの母自己相関係数

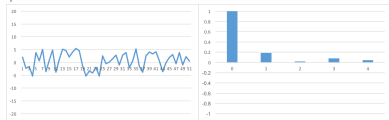
k 次の母自己相関係数

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

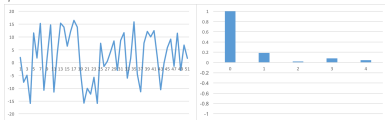
サンプルパスの $t-X_t$ グラフと、下で説明する方法で推定した $s-r(s)$ グラフ（コレログラム）
 $\phi = 0.9, \sigma = 1$
 $\phi = -0.9, \sigma = 1$



$\phi = 0.2, \sigma = 1$



$\phi = 0.2, \sigma = 3$



L12-Q2

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 $\text{Cov}[X_3, X_4]$ を σ と ϕ_k で表そう.

ここまで来たよ

11 マルコフ連鎖の時間発展・極限分布・周期的・可約なマルコフ連鎖

12 時系列解析

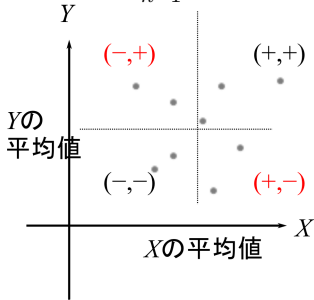
- 時系列データ
- 時系列データの分析方法 (1) 移動平均と離散フーリエ解析
- ランダムウォークモデル
- 自己回帰モデル $AR(m)$
- 時系列データの分析方法 (2) 標本自己相関係数

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 $\frac{1}{n-1}$)



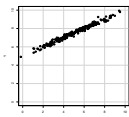
$(+, -) = (x_i - \bar{x} \text{ の符号}, y_i - \bar{y} \text{ の符号})$.

相関係数

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

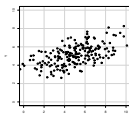
s_x, s_y : 標本標準偏差

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



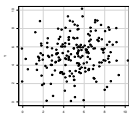
強い正の相関

$$r = 0.99$$



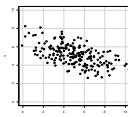
弱い正の相関

$$r = 0.55$$



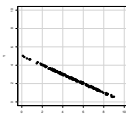
無相関

$$r = 0$$



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

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'])`

L12-Q3

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。

