全学教育「コンピュータが創る世界」
2013年1月11日(金)16:20
川内北キャンパスA205

マルコフ連鎖
Markov Chains

情報科学研究科
尾畑伸明

2013/1/11

時間とともに状態がランダムに変化

- 毎日の天気
  晴れ、曇り、雨、雪
  仙台市の天気 2012年1ー2月 同年11-12月
- 市バスに乗っている乗客数
  0人、1人、2人、、、
- 人気のラーメン屋
  空席有、空席無

2013/1/11

2状態の場合

状態	0	1
スイッチ	OFF	ON
空席	無	有
コイン	裏	表
立ち寄るカフェ	スタバ	ドトール
支持政党	A党	B党
原子のエネルギー	基底状態	励起状態

2013/1/11

0と1の配列

状態の時間変化を記録する → {0,1} の配列

【例 1】（仙台市の天気 2012年1月）
0101101100100111111000111101110

【例 2】（コイン投げ）
00001000110110011010110111110011011

規則的な配列 vs ランダムな配列

【例3】
0101010101010101010101010101010101

0,1 が交互に現れる

【例4】
0100101001010010100101001010100101001

01001 をひとかたまりとして繰り返す

規則的な配列 vs ランダムな配列

【例 2】 コイン投げ
00001000110110011010110111110011011

Excel による乱数
- RAND() [0,1] 上の一様乱数
- 表が出る確率が p であるコインのシミュレーション
  =IF(RAND()<p,1,0)

【例 5】 3のべき乗の最高位の偶奇
11000100111110101111111000100001111

3のべき乗の最高位の偶奇

Excel 計算

3	1	43046721	0
9	1	129140163	1
27	0	387420489	1
81	0	1162261467	1
243	0	3486784401	1
729	1	10460353203	1
2187	0	31381059609	1
6561	0	94143178827	1
19683	1	2.8243E+11	0
59049	1	8.47289E+11	0
177147	1	2.54187E+12	0
531441	1	7.6256E+12	1
1594323	1	2.28768E+13	0
4782969	0	6.86304E+13	0
14348907	1	2.05891E+14	0

2013/1/11

規則的な配列 vs ランダムな配列

【例 3】0,1 が交互に現れる

0101010101010101010101010101010101?

0→1　1→0　$n \geq 2$

【例 4】01001 をひとかたまりとして繰り返す

01001010010100101001010010100101001?

1→0　00→1　1010→0　0010→1　$n \geq 5$

2013/1/11

配列の生成機構に注目：規則的な配列

$x_1\ x_2\ x_3\ x_4\ x_5\ x_6\ x_7\ x_8\ \cdots\ x_{n-1}\ x_n \rightarrow x_{n+1}$

決定的
一意的に決まる

【例 3】0,1 が交互に現れる

0101010101010101010101010101010101?

$x_n \rightarrow x_{n+1}$

【例 4】01001 をひとかたまりとして繰り返す

01001010010100101001010010100101001?

$x_{n-3}x_{n-2}x_{n-1}x_n \rightarrow x_{n+1}$

2013/1/11

配列の生成機構に注目：ランダムな配列

$x_1\ x_2\ x_3\ x_4\ x_5\ x_6\ x_7\ x_8\ \cdots\ x_{n-1}\ x_n \rightarrow x_{n+1}$

一意的に決まらず、
確率的に決まる

- コイン投げ
  過去の履歴にまったく無関係
- マルコフ連鎖
  直前の状態によって次の状態が確率的に決まる

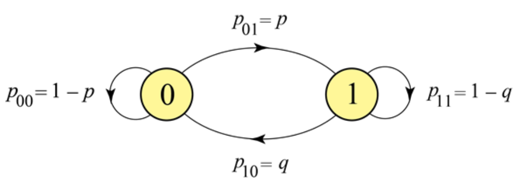
マルコフ連鎖

直前の状態によって次の状態が確率的に決まる

$x_1\ x_2\ x_3\ x_4\ x_5\ x_6\ x_7\ x_8\ \cdots\ x_{n-1}\ x_n \rightarrow x_{n+1}$

$x_n \setminus x_{n+1}$	0	1
0	$1-p$	p
1	q	$1-q$

状態推移図



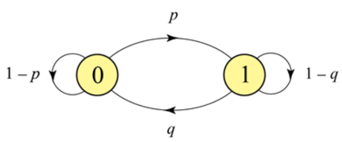
$P(i \rightarrow j)$	0	1
0	$1-p$	p
1	q	$1-q$

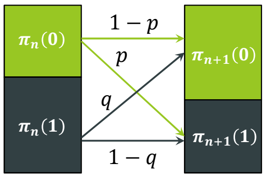
推移確率行列

$$P = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$$

$P(i \rightarrow j) = P(X_{n+1} = j | X_n = i)$　　条件付き確率

推移確率行列





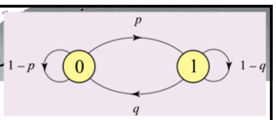
$$\begin{cases}\pi_{n+1}(0) = (1-p)\pi_n(0) + q\pi_n(1) \\ \pi_{n+1}(1) = p\pi_n(0) + (1-q)\pi_n(1)\end{cases}$$

参考:行列による表示

$$\pi_{n+1} = \pi_n P \qquad \pi_n = [\pi_n(0)\ \pi_n(1)],\quad P = \begin{bmatrix}1-p & p \\ q & 1-q\end{bmatrix}$$

2013/1/11

定常分布への収束



$$\begin{cases}\pi_{n+1}(0) = (1-p)\pi_n(0) + q\pi_n(1) \\ \pi_{n+1}(1) = p\pi_n(0) + (1-q)\pi_n(1)\end{cases}$$

終状態　$\pi(0) = \lim_{n\to\infty} \pi_n(0) \qquad \pi(1) = \lim_{n\to\infty} \pi_n(1)$

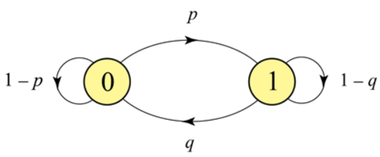
$0 < p + q < 2$　のときに極限が存在する(要証明)　シミュレーション

$$\begin{cases}\pi(0) = (1-p)\pi(0) + q\pi(1) \\ \pi(1) = p\pi(0) + (1-q)\pi(1)\end{cases}$$

定常状態　$\pi(0) = \dfrac{q}{p+q} \qquad \pi(1) = \dfrac{p}{p+q}$

2013/1/11

$\pi_n = [\pi_n(0)\ \pi_n(1)]$ を求める



$$\begin{cases}\pi_{n+1}(0) = (1-p)\pi_n(0) + q\pi_n(1) \\ \pi_{n+1}(1) = p\pi_n(0) + (1-q)\pi_n(1)\end{cases}$$

線形差分方程式(連立漸化式)

線形代数学、特に固有値と固有ベクトル

2013/1/11

線形差分方程式(漸化式)

$$ax_{n+2} + bx_{n+1} + cx_n = 0$$

$$\boxed{a\lambda^2 + b\lambda + c = 0}$$

特性方程式

$$\begin{cases} x_{n+1} = ax_n + by_n \\ y_{n+1} = cx_n + dy_n \end{cases}$$

$$\boxed{\begin{vmatrix} \lambda - a & -b \\ -c & \lambda - d \end{vmatrix} = (\lambda - a)(\lambda - d) - bc = 0}$$

特性根	$\lambda_1 \neq \lambda_2$ (異なる2根)	$\lambda_1 = \lambda_2$ (重根)
一般解	$x_n = C_1\lambda_1^n + C_2\lambda_2^n$	$x_n = (C_1 + C_2 n)\lambda_1^n$

2013/1/11

$\pi_n = [\pi_n(0)\ \pi_n(1)]$ を求める

$$\begin{cases} \pi_{n+1}(0) = (1-p)\pi_n(0) + q\pi_n(1) \\ \pi_{n+1}(1) = p\pi_n(0) + (1-q)\pi_n(1) \end{cases}$$

特性方程式と特性根

$$\begin{vmatrix} \lambda - (1-p) & -q \\ -p & \lambda - (1-q) \end{vmatrix}$$
$$= \lambda^2 - (2-p-q)\lambda + (1-p)(1-q) - pq = 0$$

$$\lambda = 1,\quad 1 - p - q \equiv r$$

重根でない場合($r \neq 1 \Leftrightarrow p + q > 0$)

$$\pi_n(0) = \frac{1}{p+q}\{q + (p\pi_0(0) - q\pi_0(1))r^n\}$$

$$\pi_n(1) = \frac{1}{p+q}\{p + (-p\pi_0(0) + q\pi_0(1))r^n\}$$

重根の場合
($r = 1 \Leftrightarrow p + q = 0$)

$$\pi_n(0) = \pi_0(0)$$

$$\pi_n(1) = \pi_0(1)$$

2013/1/11

応用:顧客生涯価値 (CLV)

【例:携帯電話の契約】
一定期間ごとに契約を更新するが、その時、他社に乗り換えることがある。

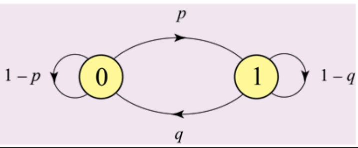
A社 → B社 → B社 → A社 → C社 →

A社から見ると

顧客 → 非顧客 → 非顧客 → 顧客 → 非顧客 →

マルコフ連鎖によるモデル化

1= 顧客 (customer)
0= 非顧客 (former)



2013/1/11

応用：顧客生涯価値 (CLV)

【マーケティング】 顧客の獲得には費用がかかる。
COCA: 顧客獲得に要する費用 (cost of customer acquisition)
CLV: 顧客が将来にわたりもたらしてくれる売上 (customer lifetime value)
CLV – COCA : その顧客が将来にわたってもたらしてくれる利益

$$CLV = \sum_{n=1}^{\infty} \boxed{\left(\frac{1}{1+d}\right)^{n-1}} \boxed{aP(X_n = 1|X_0 = 0)}$$

$\{X_n\}$: 初期条件 $X_n = 0$ のマルコフ連鎖
$a > 0$: 顧客がもたらす売上（定額）
時刻 n で顧客がもたらす売り上げの期待値

$$a \times P(X_n = 1|X_0 = 0) + 0 \times P(X_n = 0|X_0 = 0)$$

CLV では, 時刻 n で顧客がもたらす売り上げの期待値は割り引いて考える

2013/1/11

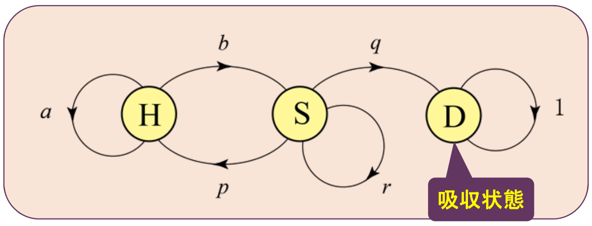
応用：顧客生涯価値 (CLV)

$$CLV = \sum_{n=1}^{\infty} \left(\frac{1}{1+d}\right)^{n-1} aP(X_n = 1|X_0 = 0)$$
$$= \sum_{n=1}^{\infty} \left(\frac{1}{1+d}\right)^{n-1} a \times \frac{1}{p+q}(q + pr^n)$$
$$= \frac{aq}{p+q} \sum_{n=1}^{\infty} \left(\frac{1}{1+d}\right)^{n-1} + \frac{apr}{p+q} \sum_{n=1}^{\infty} \left(\frac{r}{1+d}\right)^{n-1}$$
$$= \frac{aq(1+d)}{(p+q)d} + \frac{apr(1+d)}{(p+q)(1+d-r)}$$

2013/1/11

3状態のマルコフ連鎖

Healthy（健康） Sick（病気） Dead（死亡）

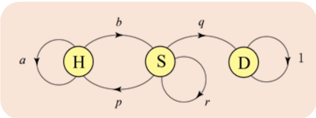


$$\begin{cases} a + b = 1 \\ p + q + r = 1 \end{cases} \qquad P = \begin{bmatrix} a & b & 0 \\ p & r & q \\ 0 & 0 & 1 \end{bmatrix}$$

2013/1/11

応用：平均寿命 (Average Life Time)

H_n: 健康な人が時刻 n で死亡する確率
S_n: 病気の人が時刻 n で死亡する確率



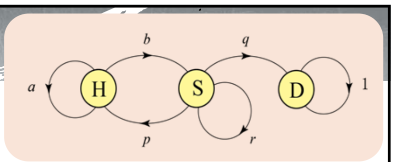
$$\begin{cases} H_n = aH_{n-1} + bS_{n-1} \\ S_n = pH_{n-1} + rS_{n-1} \end{cases} \quad n \geq 2$$

$$H_1 = 0, \qquad S_1 = q$$

特性根は相異なることがわかる：$\lambda_1 \neq \lambda_2$

$$H_n = \frac{bq(\lambda_1^{n-1} - \lambda_2^{n-1})}{\lambda_1 - \lambda_2} \qquad S_n = \frac{q\left((r-\lambda_2)\lambda_1^{n-1} - (r-\lambda_1)\lambda_2^{n-1}\right)}{\lambda_1 - \lambda_2}$$

平均寿命



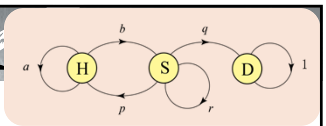
$H_n = \dfrac{bq(\lambda_1^{n-1} - \lambda_2^{n-1})}{\lambda_1 - \lambda_2}$: 健康な人が時刻 n で死亡する確率

$S_n = \dfrac{q\left((r-\lambda_2)\lambda_1^{n-1} - (r-\lambda_1)\lambda_2^{n-1}\right)}{\lambda_1 - \lambda_2}$: 病気の人が時刻 n で死亡する確率

平均寿命

$$E[H] = \sum_{n=1}^{\infty} nH_n \qquad E[S] = \sum_{n=1}^{\infty} nS_n$$

平均寿命の計算



健康な人の平均寿命

$H_n = \dfrac{bq(\lambda_1^{n-1} - \lambda_2^{n-1})}{\lambda_1 - \lambda_2}$: 健康な人が時刻 n で死亡する確率

$$E[H] = \sum_{n=1}^{\infty} nH_n = \frac{bq}{\lambda_1 - \lambda_2} \sum_{n=1}^{\infty} n(\lambda_1^{n-1} - \lambda_2^{n-1})$$

無限級数の公式

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x} \qquad \sum_{n=0}^{\infty} nx^n = \frac{x}{(1-x)^2}$$

$$E[H] = \sum_{n=1}^{\infty} nH_n = \frac{b+p+q}{bq} \qquad E[S] = \sum_{n=1}^{\infty} nS_n = \frac{b+p}{bq}$$

コンピュータが創る世界　第 11 回			講義年月日： 平成 25 年 1 月 11 日 講師：尾 畑 伸 明
学籍 番号		氏名	

【課題】マルコフ連鎖について思うことを述べなさい。

（裏面に続けて結構です。）