4. 有限マルコフ連鎖

来嶋 秀治

滋賀大学 データサイエンス学部

1. 有限マルコフ連鎖

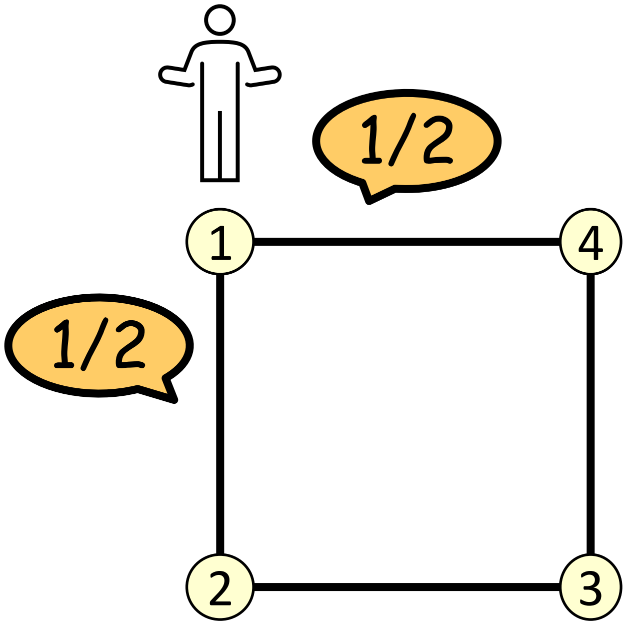
1. 有限マルコフ連鎖

■ 有限斉時マルコフ連鎖$X_0, X_1, X_2, \ldots$

$\Omega = \{1,2,\ldots,n\}$: 有限

$$P_{ij} = \Pr[X_{t+1} = j | X_t = i]$$

$P = (P_{ij})$: 遷移確率行列 (transition matrix)



O. Haggstrom, “Finite Markov chains and algorithmic applications” より

1. 有限マルコフ連鎖

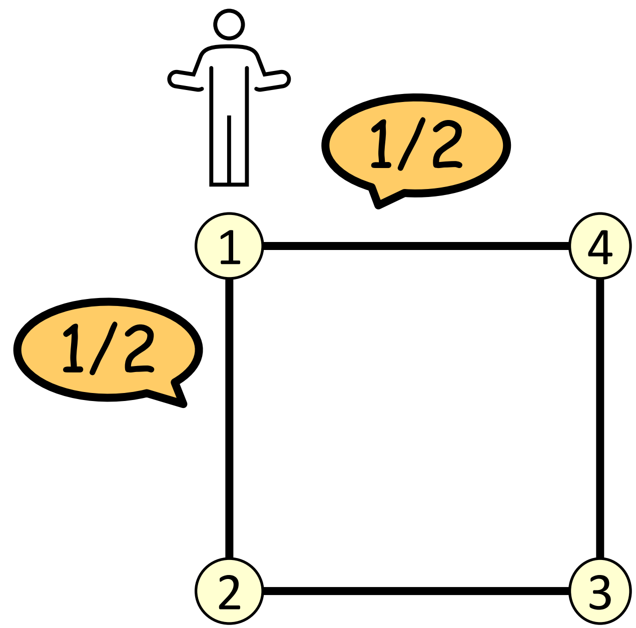
□ 有限斉時マルコフ連鎖$X_0, X_1, X_2, \ldots$

$\Omega = \{1,2,\ldots,n\}$: 有限

$$P_{ij} = \Pr[X_{t+1} = j | X_t = i]$$

$P = (P_{ij})$: 遷移確率行列 (transition matrix)

$$P = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}$$



O. Haggstrom, “Finite Markov chains and algorithmic applications” より

1. 有限マルコフ連鎖

■ 有限斉時マルコフ連鎖$X_0, X_1, X_2, \ldots$

$\Omega = \{1,2,\ldots,n\}$: 有限

$$P_{ij} = \Pr[X_{t+1} = j \mid X_t = i]$$

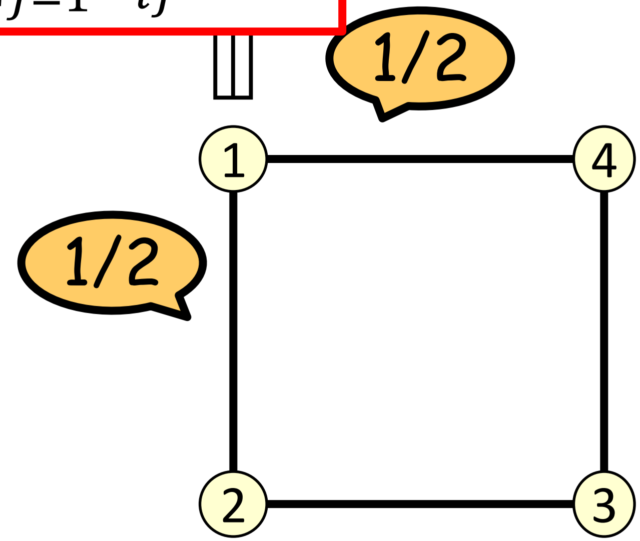
$P = (P_{ij})$: 遷移確率行列 (transition matrix)

このような行列を**確率行列**(stochastic matrix)という

定理.

$P \in \mathbb{R}_{\geq 0}^{n \times n}$が遷移確率行列 $\Leftrightarrow \forall i \in \Omega, \sum_{j=1}^{n} P_{ij} = 1$

$$P = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}$$



O. Haggstrom, “Finite Markov chains and algorithmic applications” より

1. 有限マルコフ連鎖

■ 有限斉時マルコフ連鎖$X_0, X_1, X_2, \dots$

$\Omega = \{1,2,\dots,n\}$: 有限

$$P_{ij} = \Pr[X_{t+1} = j | X_t = i]$$

$P = (P_{ij})$: 遷移確率行列 (transition matrix)

このような行列を**確率行列**(stochastic matrix)という

定理.

$P \in \mathbb{R}_{\geq 0}^{n\times n}$が遷移確率行列 $\Leftrightarrow \forall i \in \Omega, \sum_{j=1}^{n} P_{ij} = 1$

証明.

(⇒)

$$\sum_{j=1}^{n} P_{ij} = \sum_{j=1}^{n} \Pr[X_{t+1} = j | X_t = i] = \sum_{j=1}^{n} \frac{\Pr[(X_{t+1} = j) \wedge (X_t = i)]}{\Pr[(X_t = i)]}$$
$$= \frac{\sum_{j=1}^{n} \Pr[(X_{t+1} = j) \wedge (X_t = i)]}{\Pr[(X_t = i)]} = \frac{\Pr[(X_t = i)]}{\Pr[(X_t = i)]} = 1$$

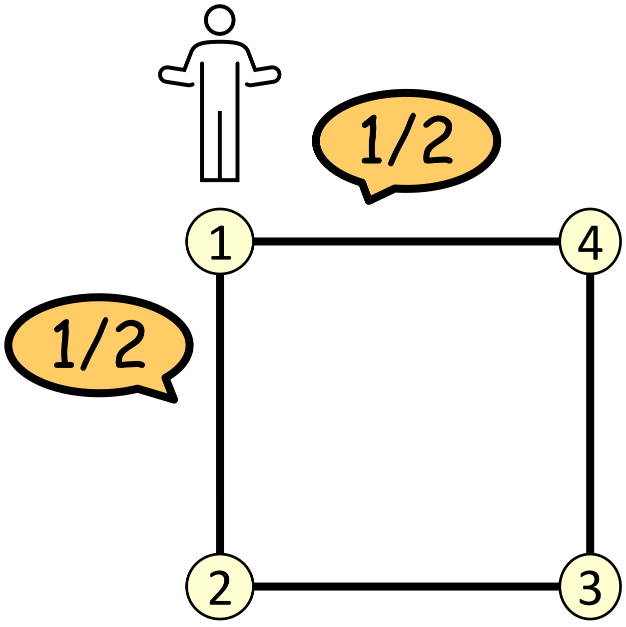
(⇐) ほぼ同じ.

2. マルコフ連鎖の確率分布

Ω上の確率分布: $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n) \in \mathbb{R}_{\geq 0}^n$

$$\|\boldsymbol{\mu}\|_1 = \sum_{i=1}^{n} \mu_i = 1$$

$\mu_0 = (1,0,0,0)$



O. Haggstrom, “Finite Markov chains and algorithmic applications” より

2. マルコフ連鎖の確率分布

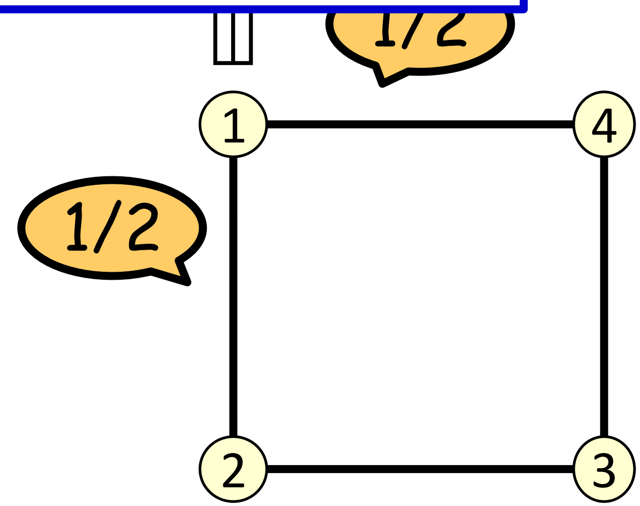
Ω上の確率分布: $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n) \in \mathbb{R}^n_{\geq 0}$

$$\|\boldsymbol{\mu}\|_1 = \sum_{i=1}^{n} \mu_i = 1$$

命題.
マルコフ連鎖$X_0, X_1, X_2, \dots$の状態空間を$\Omega = \{1, \dots, n\}$,
遷移確率行列を$P \in \mathbb{R}^{n \times n}$とする．$\boldsymbol{\mu} \in \mathbb{R}^n$はΩ上の確率分布とし,
$\boldsymbol{\mu}' = \boldsymbol{\mu} P$とする．$\Pr[X_t = i] = \mu_i$とすると，$\Pr[X_{t+1} = j] = \mu'_j$.

$$\mu_0 = (1,0,0,0)$$

$$\mu_1 = (1,0,0,0) \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix} = \left(0, \frac{1}{2}, 0, \frac{1}{2}\right)$$



O. Haggstrom, “Finite Markov chains and algorithmic applications” より

2. マルコフ連鎖の確率分布

Ω上の確率分布: $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n) \in \mathbb{R}^n_{\geq 0}$

$$\|\boldsymbol{\mu}\|_1 = \sum_{i=1}^{n} \mu_i = 1$$

命題.
マルコフ連鎖$X_0, X_1, X_2, \dots$の状態空間を$\Omega = \{1, \dots, n\}$,
遷移確率行列を$P \in \mathbb{R}^{n \times n}$とする．$\boldsymbol{\mu} \in \mathbb{R}^n$はΩ上の確率分布とし，
$\boldsymbol{\mu}' = \boldsymbol{\mu}P$とする．$\Pr[X_t = i] = \mu_i$とすると，$\Pr[X_{t+1} = j] = \mu'_j$.

証明.

$$\Pr[X_{t+1} = j] = \sum_{i=1}^{n} \Pr[X_t = i] \Pr[X_{t+1} = j | X_t = i]$$

$$= \sum_{i=1}^{n} \mu_i P_{ij} = (\mu_1, \mu_2, \dots, \mu_n) \begin{pmatrix} P_{1j} \\ P_{2j} \\ \vdots \\ P_{nj} \end{pmatrix} = (\boldsymbol{\mu}P)_j$$

3. マルコフ連鎖の定常分布

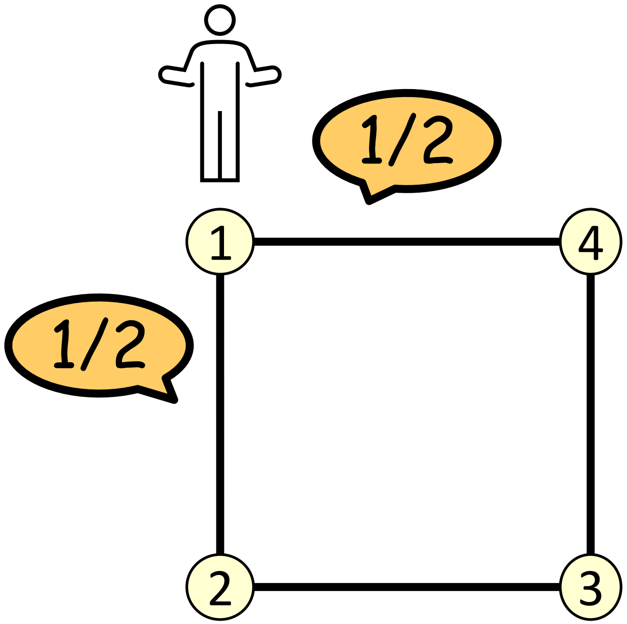
有限マルコフ連鎖の状態空間をΩ，遷移確率行列をPとする.

$\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_n) \in \mathbb{R}^n_{\geq 0}$は$\Omega$上の確率分布とする.

$\boldsymbol{\pi} P = \boldsymbol{\pi}$のとき，$\boldsymbol{\pi}$をマルコフ連鎖の**定常分布**(stationary distribution)という.

$$\pi = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$$

$$\pi P = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right) \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix} = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$$



O. Haggstrom, “Finite Markov chains and algorithmic applications” より

3. マルコフ連鎖の定常分布

有限マルコフ連鎖の状態空間をΩ，遷移確率行列をPとする.

$\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_n) \in \mathbb{R}^n_{\geq 0}$はΩ上の確率分布とする.

$\boldsymbol{\pi} P = \boldsymbol{\pi}$のとき，$\boldsymbol{\pi}$をマルコフ連鎖の**定常分布**(stationary distribution)という.

命題.
遷移確率行列$P \in \mathbb{R}^{n \times n}$は少なくとも一つの定常分布をもつ.

注意: 定常分布は一つとは限らない.
例 $P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

証明.
$\mathbf{1} \in \mathbb{R}^n$は$P\mathbf{1} = \mathbf{1}$

$$(P\mathbf{1})_i = (P_{i1}, P_{i2}, \dots, P_{in}) \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \sum_{j=1}^{n} P_{ij} = 1$$

つまりPは固有値1を持ち，その**右**固有ベクトルは$\mathbf{1}$.
さらに，「固有値1の左固有ベクトルに非負のものが存在」も成り立つ.
（**Perron-Frobeniusの定理**; 証明はやや複雑．）
非負の左固有ベクトルを正規化すれば，定常分布を得る.

[復習1: 線形代数] 固有値固有ベクトル

例 1.

$$P=\begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}$$ とすると

$$\left(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4}\right)\begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}=\left(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4}\right)$$

$\left(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4}\right)$は$P$の固有値1に対応する固有ベクトル.

[復習1: 線形代数] 固有値固有ベクトル

例 2.

$$P = \begin{pmatrix} 0.2 & 0.4 & 0.4 \\ 0.2 & 0.2 & 0.6 \\ 0.1 & 0.3 & 0.6 \end{pmatrix}$$とすると

$$\left(\frac{1}{7}, \frac{2}{7}, \frac{4}{7}\right) \begin{pmatrix} 0.2 & 0.4 & 0.4 \\ 0.2 & 0.2 & 0.6 \\ 0.1 & 0.3 & 0.6 \end{pmatrix} =$$

[復習1: 線形代数] 固有値固有ベクトル

例 2.

$P = \begin{pmatrix} 0.2 & 0.4 & 0.4 \\ 0.2 & 0.2 & 0.6 \\ 0.1 & 0.3 & 0.6 \end{pmatrix}$とすると

$$\left(\frac{1}{7}, \frac{2}{7}, \frac{4}{7}\right)\begin{pmatrix} 0.2 & 0.4 & 0.4 \\ 0.2 & 0.2 & 0.6 \\ 0.1 & 0.3 & 0.6 \end{pmatrix} = \left(\frac{1}{7}, \frac{2}{7}, \frac{4}{7}\right)$$

$\left(\frac{1}{7}, \frac{2}{7}, \frac{4}{7}\right)$は$P$の固有値1に対応する固有ベクトル

[復習1: 線形代数] 固有値固有ベクトル

例 3.

$P=\begin{pmatrix} 1/3 & 1/2 & 1/6 \\ 1/2 & 1/6 & 1/3 \\ 1/6 & 1/3 & 1/2 \end{pmatrix}$とすると

$$\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)\begin{pmatrix} 1/3 & 1/2 & 1/6 \\ 1/2 & 1/6 & 1/3 \\ 1/6 & 1/3 & 1/2 \end{pmatrix}=\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$$

$\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$は$P$の固有値1に対応する固有ベクトル

4. マルコフ連鎖の極限分布

- マルコフ連鎖が既約(irreducible)

$$\forall i, j \in \Omega, \exists t > 0, \Pr[X_t = j | X_0 = i] > 0$$

状態iから出発して，いつか状態jにたどり着く確率は0でない.

- マルコフ連鎖の周期

$$\gcd\{t > 0 | \Pr[X_t = i | X_0 = i] > 0\}$$

  - ✓ 例$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$の周期は2
  - ✓ 周期が1のマルコフ連鎖を非周期的(apperiodic)という.

4. マルコフ連鎖の極限分布

- マルコフ連鎖が既約(irreducible)

$$\forall i, j \in \Omega, \exists t > 0, \Pr[X_t = j | X_0 = i] > 0$$

状態iから出発して，いつか状態jにたどり着く確率は0でない．

- マルコフ連鎖の周期

$$\gcd\{t > 0 | \Pr[X_t = i | X_0 = i] > 0\}$$

  - ✓ 例$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$の周期は2
  - ✓ 周期が1のマルコフ連鎖を非周期的(apperiodic)という．

定理 4.1.

Pが既約で非周期的⇒ Pの定常分布は一意．

4. マルコフ連鎖の極限分布

- マルコフ連鎖が既約(irreducible)

$$\forall i, j \in \Omega, \exists t > 0, \Pr[X_t = j | X_0 = i] > 0$$

状態iから出発して，いつか状態jにたどり着く確率は0でない.

- マルコフ連鎖の周期

$$\gcd\{t > 0 | \Pr[X_t = i | X_0 = i] > 0\}$$

  - ✓ 例$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$の周期は2
  - ✓ 周期が1のマルコフ連鎖を非周期的(apperiodic)という.

定理 4.1.

Pが既約で非周期的$\Rightarrow$ Pの定常分布は一意.

定理 4.2.

Pが既約で非周期的$\Leftrightarrow$

ある分布πが存在して，任意の分布μに対して，$\mu P^{\infty} = \pi$.

Pが既約で非周期的なとき，エルゴード的(ergodic)という.

[復習2: 線形代数] 対角化 (diagonalization)

正方行列Pに対して，正則行列Uと対角行列Dが存在して，
$D = UPU^{-1}$が成り立つ時，これを**対角化**という．

対角行列

$$D = \begin{pmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{pmatrix}$$

[復習2: 線形代数] 対角化 (diagonalization)

正方行列Pに対して，正則行列Uと対角行列Dが存在して，
$D = UPU^{-1}$が成り立つ時，これを**対角化**という．

対角行列

$$D = \begin{pmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{pmatrix}$$

このとき，

$$P = U^{-1}DU$$

より，

$$\begin{aligned} P^n &= (U^{-1}DU)(U^{-1}DU)(U^{-1}DU)\cdots(U^{-1}DU)(U^{-1}DU) \\ &= U^{-1}DUU^{-1}DUU^{-1}DU\cdots U^{-1}DUU^{-1}DU \\ &= U^{-1}D(UU^{-1})D(UU^{-1})D(U\cdots U^{-1})D(UU^{-1})DU \\ &= U^{-1}D^nU \\ &= U^{-1}\begin{pmatrix} d_1^n & & & \\ & d_2^n & & \\ & & \ddots & \\ & & & d_n^n \end{pmatrix}U \end{aligned}$$

線形代数の定理.

d_jはPの**固有値**．Uの第j列はその**固有ベクトル**.

例題

例 3.

$$P = \begin{pmatrix} 1/3 & 1/2 & 1/6 \\ 1/2 & 1/6 & 1/3 \\ 1/6 & 1/3 & 1/2 \end{pmatrix}$$ のP^nを求めよ.

例題

例 3.

$$P=\begin{pmatrix}1/3 & 1/2 & 1/6\\ 1/2 & 1/6 & 1/3\\ 1/6 & 1/3 & 1/2\end{pmatrix}$$ のP^nを求めよ.

Pを対角化する．特性方程式を立てると.

$$\left(\frac{1}{3}-\lambda\right)\left(\frac{1}{6}-\lambda\right)\left(\frac{1}{2}-\lambda\right)+\frac{1}{2}\cdot\frac{1}{3}\cdot\frac{1}{6}+\frac{1}{6}\cdot\frac{1}{2}\cdot\frac{1}{3}-\left(\frac{1}{3}-\lambda\right)\frac{1}{3}\cdot\frac{1}{3}-\frac{1}{2}\cdot\frac{1}{2}\left(\frac{1}{2}-\lambda\right)-\frac{1}{6}\left(\frac{1}{6}-\lambda\right)\frac{1}{6}$$

$$=-\lambda^3+\lambda^2-\frac{11}{36}\lambda+\frac{1}{36}+\frac{1}{36}+\frac{1}{36}-\frac{1}{9}\left(\frac{1}{3}-\lambda\right)-\frac{1}{4}\left(\frac{1}{2}-\lambda\right)-\frac{1}{36}\left(\frac{1}{6}-\lambda\right)$$

$$=-\lambda^3+\lambda^2+\frac{1}{12}\lambda-\frac{1}{12}=(1-\lambda)\left(\lambda^2-\frac{1}{12}\right)=(1-\lambda)\left(\frac{\sqrt{3}}{6}-\lambda\right)\left(-\frac{\sqrt{3}}{6}-\lambda\right)$$

固有値と右固有ベクトルは$\left(1,\begin{pmatrix}1\\1\\1\end{pmatrix}\right),\left(\frac{\sqrt{3}}{6},\begin{pmatrix}2\\ \sqrt{3}-1\\ -\sqrt{3}-1\end{pmatrix}\right),\left(-\frac{\sqrt{3}}{6},\begin{pmatrix}2\\ -\sqrt{3}-1\\ \sqrt{3}-1\end{pmatrix}\right)$

$$\begin{pmatrix}\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}}\\ \frac{2}{\sqrt{12}} & \frac{\sqrt{3}-1}{\sqrt{12}} & \frac{-\sqrt{3}-1}{\sqrt{12}}\\ \frac{2}{\sqrt{12}} & \frac{-\sqrt{3}-1}{\sqrt{12}} & \frac{\sqrt{3}-1}{\sqrt{12}}\end{pmatrix}\begin{pmatrix}\frac{1}{3} & \frac{1}{2} & \frac{1}{6}\\ \frac{1}{2} & \frac{1}{6} & \frac{1}{3}\\ \frac{1}{6} & \frac{1}{3} & \frac{1}{2}\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{3}} & \frac{2}{\sqrt{12}} & \frac{2}{\sqrt{12}}\\ \frac{1}{\sqrt{3}} & \frac{\sqrt{3}-1}{\sqrt{12}} & \frac{-\sqrt{3}-1}{\sqrt{12}}\\ \frac{1}{\sqrt{3}} & \frac{-\sqrt{3}-1}{\sqrt{12}} & \frac{\sqrt{3}-1}{\sqrt{12}}\end{pmatrix}=\begin{pmatrix}1 & 0 & 0\\ 0 & \frac{\sqrt{3}}{6} & 0\\ 0 & 0 & \frac{-\sqrt{3}}{6}\end{pmatrix}$$

例題

例 3.

$$P=\begin{pmatrix}1/3 & 1/2 & 1/6\\ 1/2 & 1/6 & 1/3\\ 1/6 & 1/3 & 1/2\end{pmatrix}$$ のP^nを求めよ.

$$\begin{pmatrix}\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}}\\ \frac{2}{\sqrt{12}} & \frac{\sqrt{3}-1}{\sqrt{12}} & \frac{-\sqrt{3}-1}{\sqrt{12}}\\ \frac{2}{\sqrt{12}} & \frac{-\sqrt{3}-1}{\sqrt{12}} & \frac{\sqrt{3}-1}{\sqrt{12}}\end{pmatrix}\begin{pmatrix}\frac{1}{3} & \frac{1}{2} & \frac{1}{6}\\ \frac{1}{2} & \frac{1}{6} & \frac{1}{3}\\ \frac{1}{6} & \frac{1}{3} & \frac{1}{2}\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{3}} & \frac{2}{\sqrt{12}} & \frac{2}{\sqrt{12}}\\ \frac{1}{\sqrt{3}} & \frac{\sqrt{3}-1}{\sqrt{12}} & \frac{-\sqrt{3}-1}{\sqrt{12}}\\ \frac{1}{\sqrt{3}} & \frac{-\sqrt{3}-1}{\sqrt{12}} & \frac{\sqrt{3}-1}{\sqrt{12}}\end{pmatrix}=\begin{pmatrix}1 & 0 & 0\\ 0 & \frac{\sqrt{3}}{6} & 0\\ 0 & 0 & \frac{-\sqrt{3}}{6}\end{pmatrix}$$

したがって,

$$P^n=\begin{pmatrix}\frac{1}{\sqrt{3}} & \frac{2}{\sqrt{12}} & \frac{2}{\sqrt{12}}\\ \frac{1}{\sqrt{3}} & \frac{\sqrt{3}-1}{\sqrt{12}} & \frac{-\sqrt{3}-1}{\sqrt{12}}\\ \frac{1}{\sqrt{3}} & \frac{-\sqrt{3}-1}{\sqrt{12}} & \frac{\sqrt{3}-1}{\sqrt{12}}\end{pmatrix}\begin{pmatrix}1^n & 0 & 0\\ 0 & \left(\frac{\sqrt{3}}{6}\right)^n & 0\\ 0 & 0 & \left(\frac{-\sqrt{3}}{6}\right)^n\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}}\\ \frac{2}{\sqrt{12}} & \frac{\sqrt{3}-1}{\sqrt{12}} & \frac{-\sqrt{3}-1}{\sqrt{12}}\\ \frac{2}{\sqrt{12}} & \frac{-\sqrt{3}-1}{\sqrt{12}} & \frac{\sqrt{3}-1}{\sqrt{12}}\end{pmatrix}$$

$n\to\infty$を考えると.

$$P^\infty=\begin{pmatrix}1/3 & 1/3 & 1/3\\ 1/3 & 1/3 & 1/3\\ 1/3 & 1/3 & 1/3\end{pmatrix}$$

4. マルコフ連鎖の極限分布

定理 4.2.
Pが既約で非周期的⇔
ある分布πが存在して，任意の分布μに対して，$\mu P^{\infty} = \pi$.

証明略.

- Perron Frobeniusの定理 (線形代数)
- Coupling lemma (確率論)

まとめ: マルコフ連鎖がエルゴード的なら,

- **極限分布は一意**，そして
- その分布はPの**定常分布**.

マルコフ連鎖の極限分布

定理 4.2.
Pが既約で非周期的$\Leftrightarrow$
ある分布πが存在して，任意の分布μに対して，$\mu P^{\infty} = \pi$.

証明.

- Perron Frobeniusの定理 (線形代数)
- Coupling lemma (確率論) <=今日はこっち

まとめ: マルコフ連鎖がエルゴード的なら，
- **極限分布は一意**，そして
- その分布はPの**定常分布**.

MCMC法

可逆マルコフ連鎖 (reversible Markov chain)

定理 4.3.

有限集合Ω上の非負関数fに対して，推移確率行列Pが，任意の$i, j \in \Omega$について

$$P_{ij} f(i) = P_{ji} f(j)$$

を満たすとする．このとき，Pの定常分布$\boldsymbol{\pi}$は

$$\pi_i = \frac{f(i)}{\sum_{j \in \Omega} f(j)}$$

を満たす.

可逆マルコフ連鎖 (reversible Markov chain)

定理 4.3.

有限集合Ω上の非負関数fに対して，推移確率行列Pが，任意の$i, j \in \Omega$について

$$P_{ij} f(i) = P_{ji} f(j)$$

詳細釣り合いの式
(detailed balance equation)

を満たすとする．このとき，Pの定常分布$\boldsymbol{\pi}$は

$$\pi_i = \frac{f(i)}{\sum_{j \in \Omega} f(j)}$$

を満たす．

可逆マルコフ連鎖

可逆マルコフ連鎖 (reversible Markov chain)

定理 4.3.

有限集合Ω上の非負関数fに対して，推移確率行列Pが，任意の$i, j \in \Omega$について

$$P_{ij} f(i) = P_{ji} f(j)$$

詳細釣り合いの式
(detailed balance equation)

を満たすとする．このとき，Pの定常分布$\boldsymbol{\pi}$は

$$\pi_i = \frac{f(i)}{\sum_{j \in \Omega} f(j)}$$

を満たす.

可逆マルコフ連鎖

証明.

Ω上の分布$\boldsymbol{\pi}$が$\boldsymbol{\pi} \propto f$を満たすとき，詳細釣り合いの式から

$\boldsymbol{\pi} P = \boldsymbol{\pi}$ (つまり$\boldsymbol{\pi}$は定常分布)が導かれることを示す.

$$(\pi P)_k \propto \sum_{i=1}^{n} f(i) P_{ik} = \sum_{i=1}^{n} f(k) P_{ki} = f(k) \sum_{i=1}^{n} P_{ki} = f(k) \propto \pi_k.$$

可逆マルコフ連鎖/MCMC法の例

1. 一様分布
2. 無向グラフ上の単純ランダムウォーク
3. メトロポリス法
4. スライスサンプリング
5. ギブス分布/ボルツマン分布/Isingモデル

例1. 一様分布

例1. 一様分布

Ω上の非負関数fは任意のiに対して$f(i) = 1$とする.

このとき，詳細釣り合いの式は

$$P_{ij} = P_{ji}$$

となり，すなわちPは対称行列である.

このときΩ上の一様分布はPの定常分布.

例2. グラフ上のランダムウォーク

- 無向グラフ$G = (V, E)$上の単純ランダムウォークは

$$p_{uv} = \Pr[X_{t+1} = v | X_t = u] = \frac{1}{d(u)} \quad (\{u, v\} \in E)$$

と定める.

ただし$d(v) := |\{u \in V | \{u, v\} \in E\}|$を頂点$v$の次数(degree)という.

- あきらかに

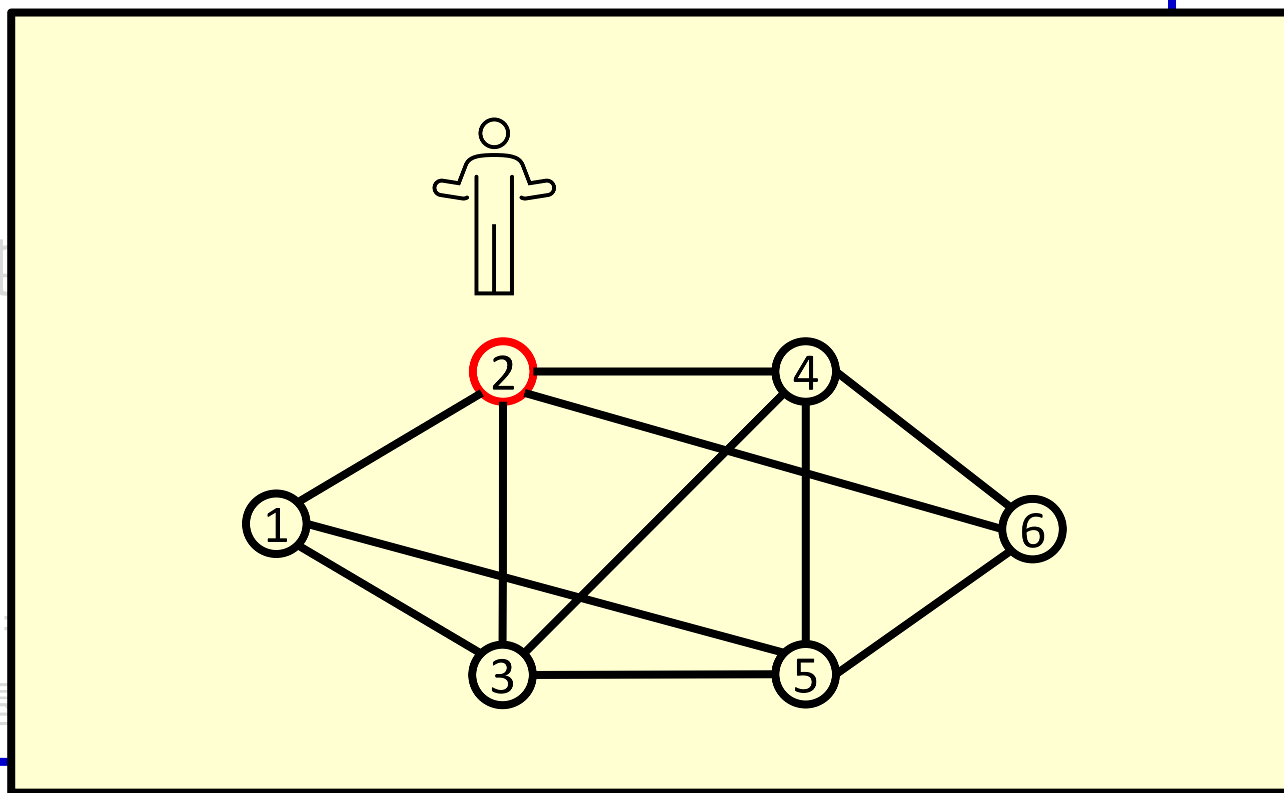
が成り立つ.

- すなわち，単純

となる.

ここで$\sum_{v \in V} d(v)$

はグラフ理論の基



例2. グラフ上のランダムウォーク

- 無向グラフ$G = (V, E)$上の単純ランダムウォークは

$$p_{uv} = \Pr[X_{t+1} = v | X_t = u] = \frac{1}{d(u)} \quad (\{u, v\} \in E)$$

と定める.

ただし$d(v) := |\{u \in V | \{u, v\} \in E\}|$を頂点$v$の次数(degree)という.

- あきらかに

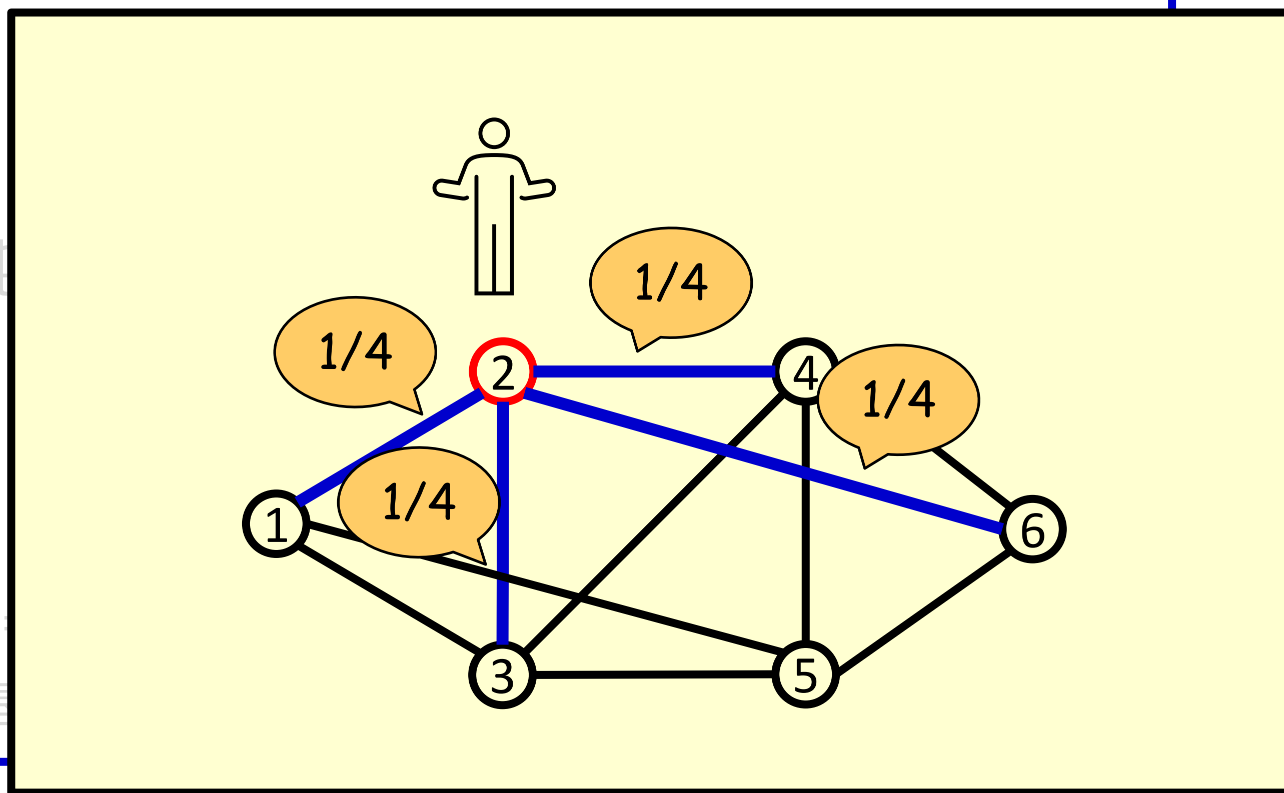
が成り立つ.

- すなわち，単純

となる.

ここで$\sum_{v \in V} d(v)$

はグラフ理論の基



例2. グラフ上のランダムウォーク

- 無向グラフ$G = (V, E)$上の単純ランダムウォークは

$$p_{uv} = \Pr[X_{t+1} = v | X_t = u] = \frac{1}{d(u)} \quad (\{u, v\} \in E)$$

と定める.

ただし$d(v) := |\{u \in V | \{u, v\} \in E\}|$を頂点$v$の次数(degree)という.

- あきらかに

$$p_{uv} d(u) = p_{vu} d(v)$$

が成り立つ.

- すなわち，単純ランダムウォークの定常分布は

$$\frac{d(v)}{\sum_{v \in V} d(v)} = \frac{d(v)}{2|E|}$$

となる.

ここで$\sum_{v \in V} d(v) = \sum_{v \in V} \sum_{u \in V} 1_{\{u,v\} \in E} = \sum_{(u,v) \in V^2} 1_{\{u,v\} \in E} = 2|E|$はグラフ理論の基本的な定理である.

例2. グラフ上のランダムウォーク

- 無向グラフ$G=(V,E)$上の単純ランダムウォークは

$$p_{uv}=\Pr[X_{t+1}=v|X_t=u]=\frac{1}{d(u)}\quad(\{u,v\}\in E)$$

と定める.

ただし$d(v):=|\{u\in V|\{u,v\}\in E\}|$を頂点$v$の次数(degree)という.

- あきらかに

$$p_{uv}d(u)=p_{vu}d(v)$$

が成り立つ.

- すなわち，単純ランダムウォークの定常分布は

$$\frac{d(v)}{\sum_{v\in V}d(v)}=\frac{d(v)}{2|E|}$$

となる.

ここで$\sum_{v\in V}d(v)=\sum_{v\in V}\sum_{u\in V}1_{\{u,v\}\in E}=\sum_{(u,v)\in V^2}1_{\{u,v\}\in E}=2|E|$はグラフ理論の基本的な定理である.

メトロポリス法 (基本形)

- Ω上に正値関数$f: \Omega \to \mathbb{R}_{>0}$が与えられ，単純無向グラフ$G = (\Omega, E)$が与えられたとする．すなわち$E \subseteq \Omega^2$は$\Omega$の要素の隣接関係を表す．
-

$$p_{xy} = \Pr[X_{t+1} = y | X_t = x] = \begin{cases} \dfrac{1}{d(x)} \min\left\{\dfrac{f(y)}{f(x)}, 1\right\} & \{x, y\} \in E \\ 0 & \{x, y\} \notin E, y \neq x \\ 1 - \displaystyle\sum_{y:\, y \neq x} p_{xy} & y = x \end{cases}$$

と定める．

- このとき，

$$p_{xy} f(x) = p_{yx} f(y)$$

が成り立つ．

  - 証明は$\min\left\{\frac{f(y)}{f(x)}, 1\right\} = \frac{\min\{f(y), f(x)\}}{f(x)}$から容易．

メトロポリス法 (発展形)

- Ω上に非負値関数$f : \Omega \to \mathbb{R}_{>0}$が与えられ，単純無向グラフ$G = (\Omega, E)$が与えられたとする．すなわち$E \subseteq \Omega^2$は$\Omega$の要素の隣接関係を表す．
- $f(x) \neq 0$のxに対して，

$$p_{xy} = \Pr[X_{t+1} = y | X_t = x] = \begin{cases} \dfrac{1}{d(x)} \min\left\{\dfrac{f(y)}{f(x)}, 1\right\} & \{x, y\} \in E \\ 0 & \{x, y\} \notin E, y \neq x \\ 1 - \displaystyle\sum_{y:\, y \neq x} p_{xy} & y = x \end{cases}$$

と定める．$f(x) = 0$の時は便宜的に$p_{xx} = 1$とする．

- このとき，

$$p_{xy} f(x) = p_{yx} f(y)$$

が成り立つ．

  - 証明は$\min\left\{\frac{f(y)}{f(x)}, 1\right\} = \frac{\min\{f(y), f(x)\}}{f(x)}$から容易．
  - とくに$f(y) = 0$のとき，任意のxについて$p_{xy} = 0$となる．

スライスサンプリング (基本形)

- スライスサンプリングは厳密には連続空間(もしくは連続/離散の混合空間)上のランダムウォークだが，大変有用なテクニックなので紹介する.
- $\mathbb{R}$上の非負有限可測関数$f: \mathbb{R} \to \mathbb{R}_{\geq 0}$が与えられたとする.

  スライスサンプリング(基本形)は2段階からなる.

  1. $\xi \sim \mathrm{U}\left(0, f(X_t)\right)$ を選ぶ(連続一様分布)
  2. $X_{t+1} \sim U\{y \in \mathbb{R} | f(y) \geq \xi\}$
- このとき，$f(x) > 0$となるxを初期値とすると，極限密度は$\frac{f}{\int f(t)\mathrm{d}t}$となる.
  - ✓ スライスサンプリングの状態遷移確率密度は

    $$p_{xy} = \frac{1}{f(x)} \int_0^{\min\{f(x), f(y)\}} \frac{1}{g(t)} dt$$

    となる．ただし$g(t) = \{y \in \mathbb{R} | f(y) \geq t\}$.
  - ✓ このとき，

    $$p_{xy} f(x) = p_{yx} f(y)$$

    が成り立つ.

スライスサンプリング (rejection 変形)

- スライスサンプリングは厳密には連続空間(もしくは連続/離散の混合空間)上のランダムウォークだが，大変有用なテクニックなので紹介する．
- $\mathbb{R}$上の非負有限可測関数$f: \mathbb{R} \to \mathbb{R}_{\geq 0}$が与えられたとする．
  スライスサンプリング(パラメータ$\alpha > 0$)は2段階からなる．
  1. $\xi \sim \mathrm{U}(0, f(X_t))$ を選ぶ(連続一様分布)
  2. $Y \sim U\{y \in \mathbb{R} \mid |y - x| < \alpha\}$
  3. $X_{t+1} = \begin{cases} Y & f(Y) \geq \xi \\ X_t & \text{それ以外} \end{cases}$
- このとき，$f(x) > 0$となるxを初期値とすると，極限密度は$\frac{f}{\int f(t)\mathrm{d}t}$となる．
  - ✓ スライスサンプリングの状態遷移確率密度は

$$p_{xy} = \frac{1}{f(x)} \int_0^{\min\{f(x), f(y)\}} \frac{1}{2\alpha} dt.$$

  - ✓ このとき，

$$p_{xy} f(x) = p_{yx} f(y)$$

  が成り立つ．

ギブスサンプリング/ボルツマン分布/Ising模型

- **ギブス(Gibbs)分布**，**ボルツマン(Boltzmann)分布**，Isingモデル，Pottsモデルは統計物理のモデルに由来する用語である．文脈に依存して，その語の指す定義は異なるが，数学的な抽象化レベル(一般化)の違いであることが多い．
- $\Omega \subseteq \mathbb{Z}^n$とする．$h\colon \Omega \to \mathbb{R}$をエネルギー関数と呼び，

$$f(x) = \exp\bigl(-\beta h(\boldsymbol{x})\bigr) \qquad (\boldsymbol{x} \in \Omega)$$

　をボルツマン分布と呼ぶ．定数$\beta > 0$は逆温度(inverse temperature).
- 通常，ボルツマン分布，ギブス分布では，グラフ$G = (\{1,2, \dots, n\}, E)$を仮定し，

$$h(x) = \sum_{\{i,j\} \in E} a_{ij} x_i x_j + \sum_{i=1}^{n} b_i x_i$$

　である．
  - Gは2次元，3次元格子グラフが標準的．d次元格子グラフもある．
  - 完全グラフの場合は平均場近似と呼ばれるものに対応するらしい．
  - 「グラフは一般」と書かれることも多いが，本当に一般のグラフで成り立つ議論かどうかは文脈をよく確認する必要がある．
    - 数学，理論計算機科学では，額面通りのことが多い．
    - 物理，計算科学等では，しばしば “一般のグラフ” がunit disc graphを指す．本来の物理モデルが$\mathbb{R}^2, \mathbb{R}^3$で，計算のためのグリッド化に由来．
- $\Omega = \{-1,1\}^n$の時，Isingモデルという．$-\beta a_{ij} > 0$の時を強磁性(ferromagnetic)，$-\beta a_{ij} < 0$の時を反磁性(anti-feromagnetic)という．
- Potts模型は $h(x) = \sum_{\{i,j\} \in E} a_{ij} \delta(x_i, x_j) + \sum_{i=1}^{n} b_i x_i$ で $\delta(s,t) = \begin{cases} 1 & s = t \\ 0 & s \neq t \end{cases}$

ギブスサンプリング

- **ギブス(Gibbs)分布**が特別視されるのは，主に格子グラフモデルで，しばしば，$x = (x_1, \ldots, x_n)$の半数近い変数を一度に変更する(大近傍; large neighbor)手法が用いられるからである.
  - これをギブスサンプラー(Gibbs sampler)という.

    ※ 大近傍を使わない場合でも，しばしばギブスサンプラーと呼ぶ.

おわり