日本オペレーションズリサーチ学会

March 17, 2022

思う

私の数理工学

来嶋秀治

九州大学 大学院システム情報科学研究院

研究略歴

略歴

2002. 3. 東大 工 計数工学科 **数理工学**コース 卒

2004. 3. 同院 情報理工 数理情報学専攻 修士了

2007. 3. 同 博士了

博士論文「MCMC法における近似精度保証と完璧サンプリング法」

2007. 4.– 2010.3. 京大 数理解析研 助教

2010. 4.- 現在 九大 システム情報科学研究院 准教授

2016.10 – 2020.3 JSTさきがけ研究員

領域「社会的課題の解決に向けた数学と諸分野の協働 (國府総括)」

研究分野

数理工学, 理論計算機科学

キーワード

- ✓ 確率過程: ランダムウォーク, MCMC（マルコフ連鎖モンテカルロ）
- ✓ 離散数学: 劣モジュラ関数，グラフ
- ✓ 分散計算: 自律分散ロボット，自己安定

機関誌「オペレーションズ・リサーチ」，
34巻, 11号, 616頁, 1989. (伏見先生)

研究室だより ◆

東京大学 工学部 計数工学科

当学科のＯＲ関係の研究室は３年半ほど前に紹介ずみ

それらにもとづく対象への働きかけ（制御）までの全体

数理工学者・技術者は，広い工学的技能・常識，古典的近代的数学の基本的な考え方に通じているにとどまらず，これを**自由に**駆使して工学の基礎的な問題にも現実の実践的な問題にも深く切り込むことができなければならない.

---伏見正則

年生でも共通の部分が多く，互いに補い合って教育を行なっています．工学部の他の学科が主として産業別の区分になっているのに対して，両学科は横断的な教育・研究をめざす学科であるといえましょう．

数理工学コースの目標はおよそ次のようになっています．――「数理」というのは単なる「数学」ないしはその「応用」のことではない．数理工学者・技術者は，広い工学的技能・常識，古典的近代的数学の基本的な考え方に通じているにとどまらず，これを自由に駆使して工学の基礎的な問題にも現場の実践的な問題にも深く切り込むことができなければならない．――これはＯＲの目標とよく似ているのではないでしょうか．

数理コースの科目の中には，位相・代数・線形・解析

というのは無限の彼方のように思われるらしく，のんびり構えていて最後にあわてふためく者が少なからずいるのは嘆かわしいことです．

年によって違いはありますが，学部卒業生の４～５割程度が修士課程に進学します．そのさいも，出身コースや卒業論文で配属された研究室とは別の研究室へ行くことを歓迎しています．修士課程修了者の大部分は就職しますが(学部卒を含む)，就職先は，学科の性格を反映して特定の業種に偏ることがなく，きわめて多方面にわたります．近頃話題になっている“工学部出身者の金融業への就職”は，当学科にとっては目新しいことではなく，学科発足当時から見られている現象です．

教官組織を詳しく述べる余裕がなくなりましたが，11

「数理工学 ------ 「工学」諸問題を数理する！」(文部科学省科学技術政策研究所 編著：「数学イノベーション」 第７章(2007年)の 元原稿) (藤重先生)

数理工学 —— 「工学」諸問題を数理する！

藤重 悟　(京都大学数理解析研究所)

「数理工学」とは，端的に述べると，**「工学」諸問題を数理する学問**である.

---藤重悟

1. 数理工学とは

数理工学について語るには，その言葉の定義から始めなければならないが，数理工学に関わっている人たちの間で，はっきりとした共通の定義が了解されているとは言いがたいのが実情であろう. したがって，本小文で述べる「数理工学」はあくまでも私見に過ぎないことを始めにお断りしておきたい.

さて，私見であるが,「**数理工学**」とは，端的に述べると，**「工学」諸問題を数理する学問**である.

ここで，括弧付きの「工学」とは，(括弧付きの)「**もの作り**」である. ここで言う「もの作り」とは，従来の伝統的な工学の分野で対象としてきた物理的な実在としての物（たとえば，自動車，飛行機，建築物，電気製品，コンピュータ，道路，通信網など）の物作りは勿論，それらの物を動かし運用するシステム作りや設計・配置，さらには計画作り・スケジューリングなどの産業諸活動におけるソフトな「もの作り」をも意味し，もっと広く企業・自治体や人間などの組織や個人の意思決定に関

今日の話題

I. 私の数理工学: 理論編 (20分: 研究紹介)
II. 私の数理工学: 実践編 (15分)
III. 私の数理工学: 開拓編 (15分)

I. 私の数理工学: 理論編

良い研究をして、良い論文を書く。

私の数理工学: 理論編 (研究宣伝)

- 完璧サンプリング法
  - ✓ 分割表 [RSA 2006]
  - ✓ 待ち行列ネットワーク [SICOMP 2008, AOR 2008]
- メディアン安定マッチング [MathOR 2012]
- 確率と計算
  - ✓ 決定性RW [**RSA 2015**, TCS 2017, SIDMA 2018]
  - ✓ 高次元体積の決定性近似 [Algorithmica 2016, **TCS 2020**]
- 対数優モジュラ分布
  - ✓ 劣モジュラ関数の変数変換 [**SIDMA 2020**]
  - ✓ 任意の有限分配束は$M^\natural$凹関数の最小化元族に同型 [**OR lett. 2021**]
- 動的環境下のアルゴリズム論
  - ✓ 増えるクーポン収集 [**SODA 2021**]

私の数理工学: 理論編 (研究宣伝)

◘ 完璧サンプリング法 (学生～数理解析研)

- ✓ **分割表** [RSA 2006]
- ✓ 待ち行列ネットワーク [SICOMP 2008, AOR 2008]

◘ メディアン安定マッチング [MathOR 2012]

[1] S. Kijima and Tomomi Matsui, Polynomial time perfect sampling algorithm for two-rowed contingency tables, Random Structures & Algorithms, 29:2 (2006), 243--256.
[2] S. Kijima and Tomomi Matsui, Approximation algorithm and perfect sampler for closed Jackson networks with single servers, SIAM Journal on Computing, 38:4 (2008), 1484--1503.
[3] S. Kijima and Tomomi Matsui, Randomized approximation scheme and perfect sampler for closed Jackson networks with multiple servers, Annals of Operations Research, 162 (2008), 35--55.
[4] Tomomi Matsui, Mitsuo Motoki, Naoyuki Kamatani and S. Kijima, Polynomial time approximate or perfect samplers for discretized Dirichlet distribution, Japan Journal of Industrial and Applied Mathematics, 27:1 (2010), 91--123.
[5] S. Kijima, Sampling from log-super/submodular distributions, in Proc. 7th Hungarian-Japanese Symposium on Discrete Mathematics and Its Applications (2011), 227--236.

例: 2行分割表のランダム生成

2元分割表

- ✓ 各セルには非負整数が入る
- ✓ （与えられた）周辺和を満たす

						12
						18
5	4	3	7	5	6	30

5	4	3	0	0	0	12
0	0	0	7	5	6	18
5	4	3	7	5	6	30

状態A

4	3	1	3	1	0	12
1	1	2	4	4	6	18
5	4	3	7	5	6	30

状態B

0	0	0	1	5	6	12
5	4	3	6	0	0	18
5	4	3	7	5	6	30

状態C

問題

Given: 周辺和

出力: 分割表の一様ランダム生成

例: 2行分割表のランダム生成

2元分割表

- ✓ 各セルには非負整数が入る
- ✓ （与えられた）周辺和を満たす

						12
						18
5	4	3	7	5	6	30

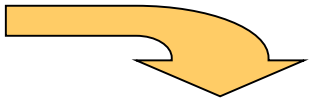
与えられた周辺和を満たす**2行**分割表の**個数を求める問題**

⇒#P完全 (NP困難) [‘97 Dyer, Kannan, & Mount]

問題

Given: 周辺和

出力: 分割表の一様ランダム生成

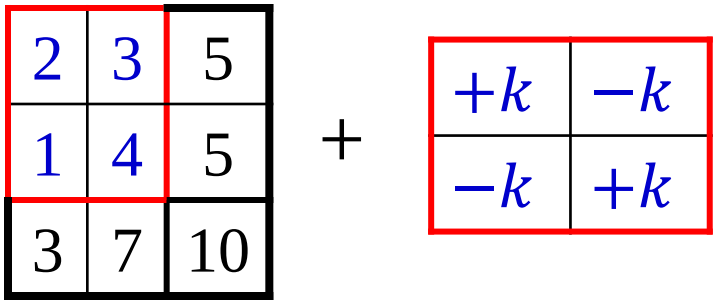


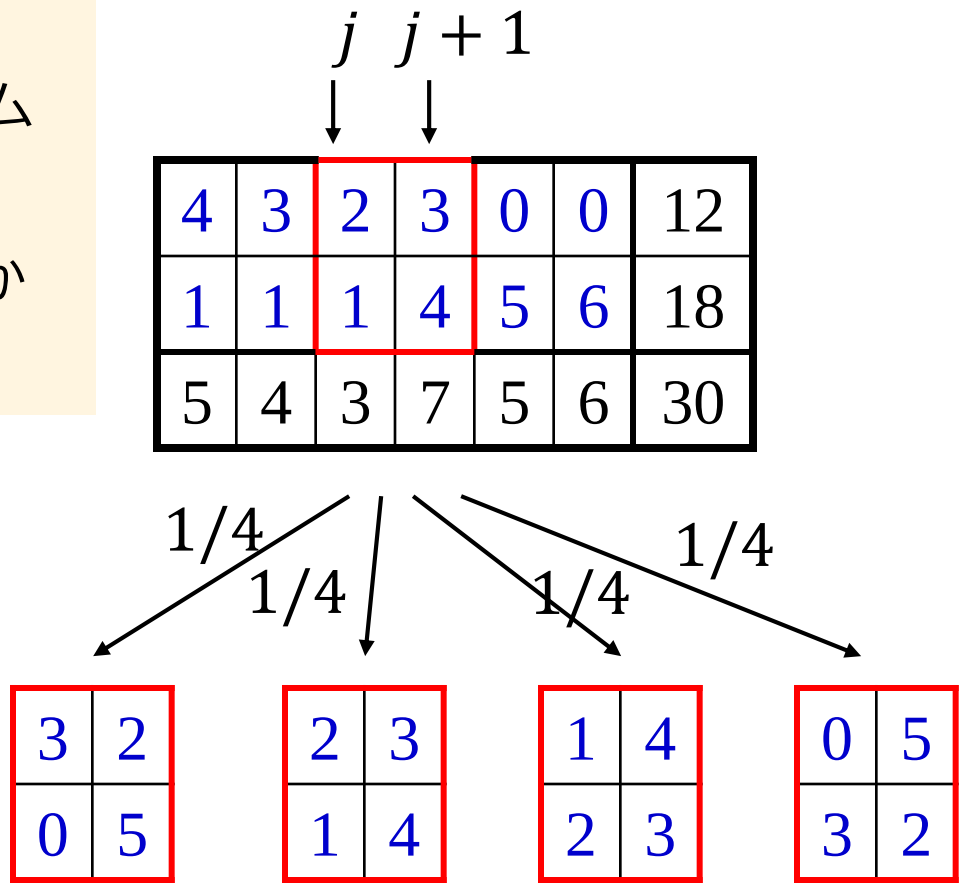
マルコフ連鎖を用いた
サンプリング法

例: 2行分割表に対するマルコフ連鎖 [K & Matsui '06]

K-M chain

1. $j \in \{1, \ldots, n-1\}$ を一様ランダムに選ぶ.
2. $\begin{pmatrix} x_{1,j} & x_{1,j+1} \\ x_{2,j} & x_{2,j+1} \end{pmatrix}$ を実行可能な中から一様ランダムに選ぶ.



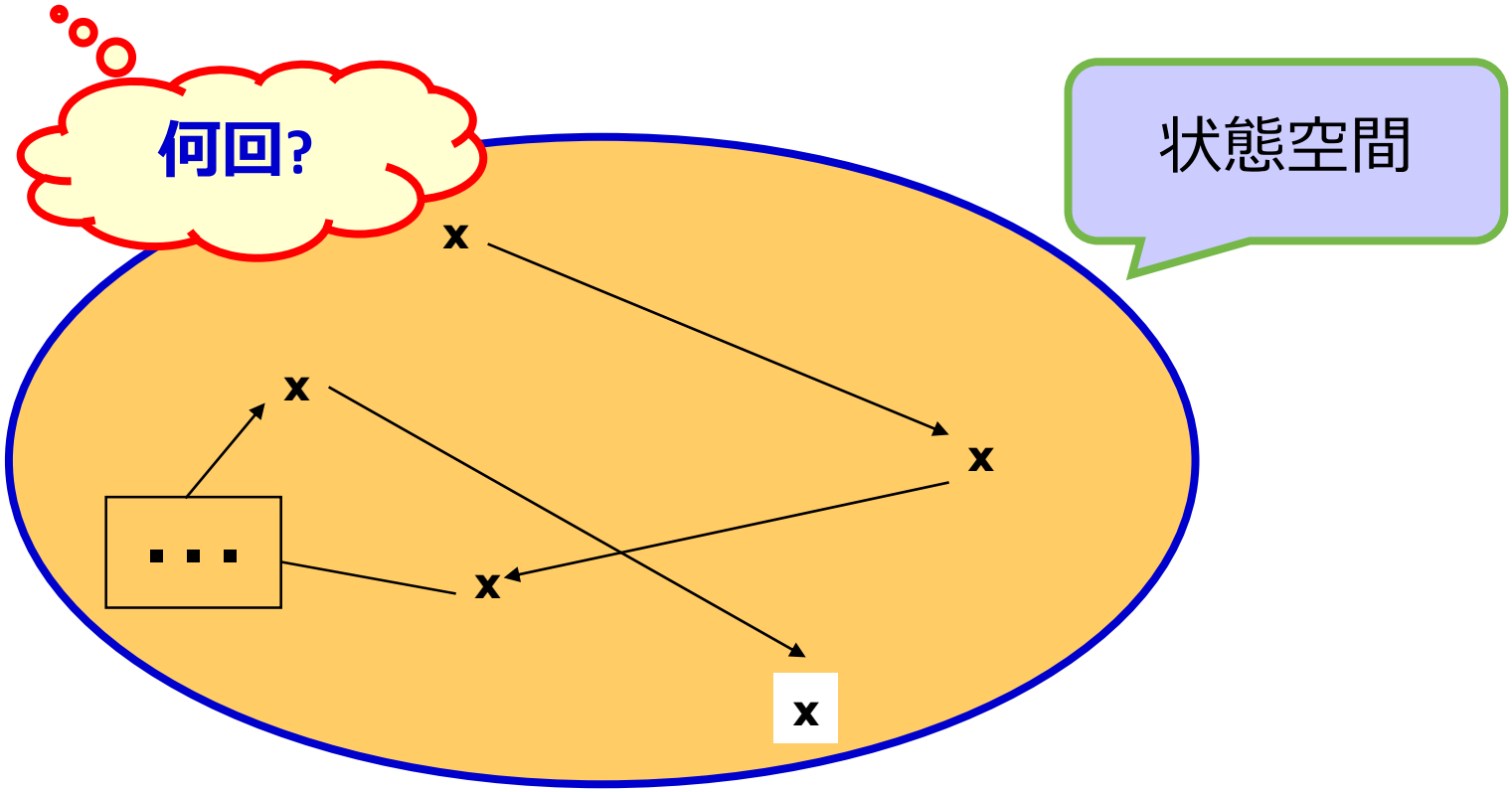


定理

このマルコフ連鎖の定常分布は一様分布.

MCMC法のアイデア

1. 所望の分布を定常分布にもつマルコフ連鎖を設計する.
2. 十分な回数推移させて，定常分布からサンプリングする.



⇒ (漸近的に)所望の分布に従うサンプリングを実現

問題点は何か？

「何回推移させれば十分か？」

近似サンプリング法

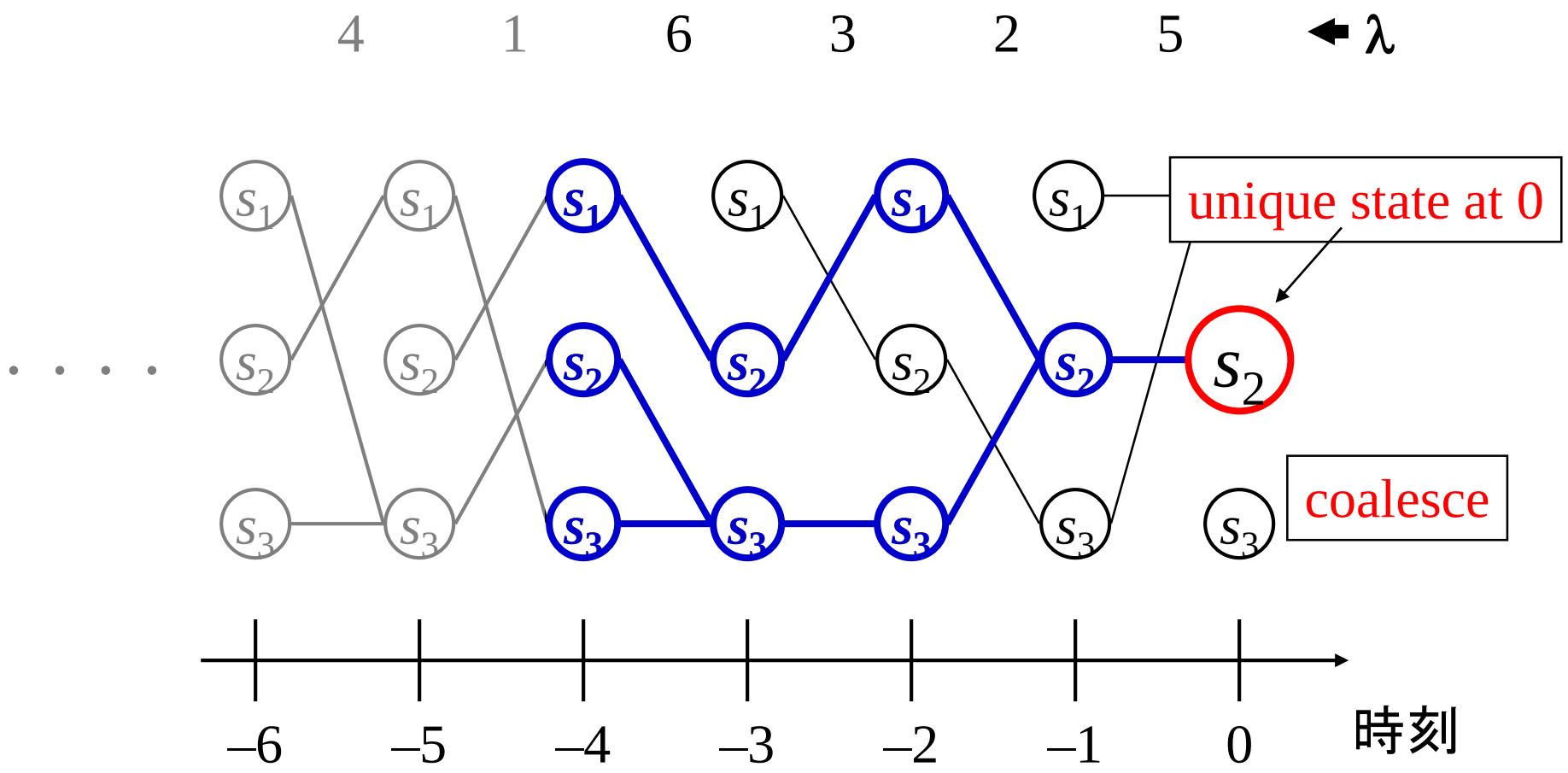
- ✓ 収束スピード(mixing time)の算定

完璧サンプリング法

- ✓ マルコフ連鎖の**推移シミュレーションを工夫**
- ✓ **無限回の推移の結果**を出力 (⇒ 定常分布に**厳密に**従う)
  - **過去からのカップリング** [Propp & Wilson 1996]

過去からのカップリング[Propp & Wilson 1996]

初期状態を知る必要がない !!



2行分割表の完璧サンプリング法

2元分割表

- ✓ 各セルには非負整数が入る
- ✓ （与えられた）周辺和を満たす

						12
						18
5	4	3	7	5	6	30

[K & Matusi 2006]
2行分割表の一様分布に対して，過去からのカップリングを用いた$O(n^3 \log N)$時間の完璧サンプリング法を設計.

[1] S. Kijima and Tomomi Matsui, Polynomial time perfect sampling algorithm for two-rowed contingency tables, Random Structures & Algorithms, 29:2 (2006), 243--256.

問題

Given: 周辺和

出力: 分割表の一様ランダム生成

私の数理工学: 理論編 (研究紹介)

◘ 完璧サンプリング法 (学生～数理解析研)

- ✓ 分割表 [RSA 2006]
- ✓ **待ち行列ネットワーク** [SICOMP 2008, AOR 2008]

◘ メディアン安定マッチング [MathOR 2012]

[1] S. Kijima and Tomomi Matsui, Polynomial time perfect sampling algorithm for two-rowed contingency tables, Random Structures & Algorithms, 29:2 (2006), 243--256.

[2] S. Kijima and Tomomi Matsui, Approximation algorithm and perfect sampler for closed Jackson networks with single servers, SIAM Journal on Computing, 38:4 (2008), 1484--1503.

[3] S. Kijima and Tomomi Matsui, Randomized approximation scheme and perfect sampler for closed Jackson networks with multiple servers, Annals of Operations Research, 162 (2008), 35--55.

[4] Tomomi Matsui, Mitsuo Motoki, Naoyuki Kamatani and S. Kijima, Polynomial time approximate or perfect samplers for discretized Dirichlet distribution, Japan Journal of Industrial and Applied Mathematics, 27:1 (2010), 91--123.

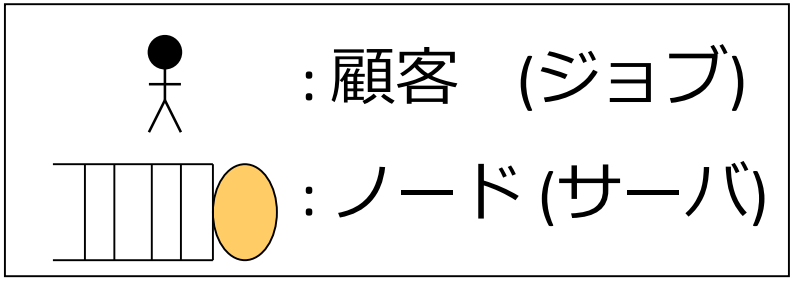
[5] S. Kijima, Sampling from log-super/submodular distributions, in Proc. 7th Hungarian-Japanese Symposium on Discrete Mathematics and Its Applications (2011), 227--236.

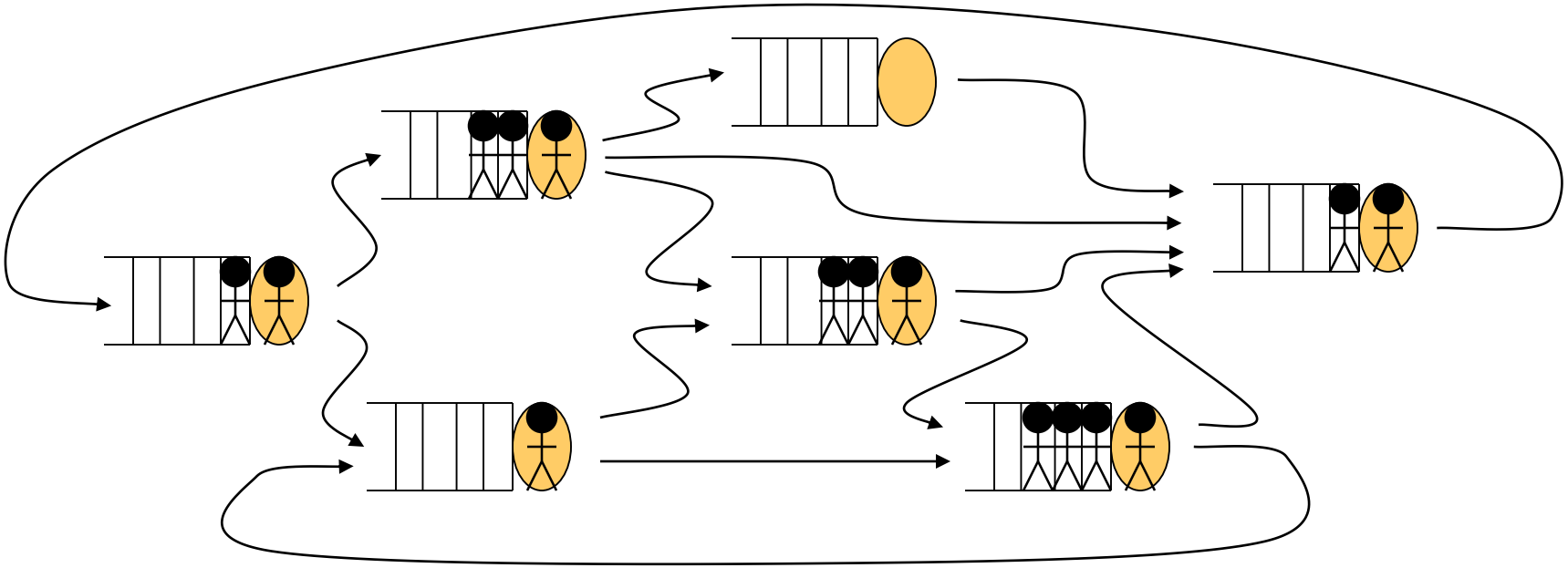
待ち行列ネットワーク

- 顧客はノードでサービスを受ける
- サービス終了後，顧客は次のノードへ確率的に移動する．

どれくらい待つのだろう?

- 待ち行列の長さ
- ノードでの滞在時間





Everybody waits for Space Mountain, the Haunted Mansion, Jungle cruise, and so on …

閉ジャクソンネットワーク (単一サーバ)

K: ネットワーク中のジョブの総数
$\{1, 2, \ldots, n\}$: ノードの集合
$\mu = (\mu_1, \mu_2, \ldots, \mu_n), \boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_n)$: パラメータ

1. **閉**ネットワーク (= ジョブの総数 K は一定)
2. 各ノードでは **先着順 (First Come First Served)**.
3. 各ノードには**1個**のサーバ (**単一サーバ**モデル).
4. ノード i でのサービス時間は期待値μ_i の**指数分布**.
5. ノード i でサービス終了したジョブは, **確率 P_{ij}** でノード j へ.

セントラルサーバモデルなどに応用

閉ジャクソンネットワーク (単一サーバ) の均衡分布

K: ネットワーク中のジョブの総数
$\{1, 2, \ldots, n\}$: ノードの集合
$\mu = (\mu_1, \mu_2, \ldots, \mu_n), \boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_n)$: パラメータ

ネットワーク中のジョブの均衡分布

単体中の整数格子点

状態空間:
$$\Xi \stackrel{\text{def.}}{=} \{(x_1, \ldots, x_n) \in \mathbb{Z}^n \mid x_i \geq 0,\ \textstyle\sum_{i=1}^n x_i = K\}$$

確率関数:
$$J(\boldsymbol{x}) \stackrel{\text{def.}}{=} \frac{1}{G(K)} \prod_{i=1}^n \alpha_i(x_i)$$

積形式解
(product form solution)

ただし
$$\alpha_i(z) \stackrel{\text{def.}}{=} \left(\frac{\theta_i}{\mu_i}\right)^z$$
$$G(K) \stackrel{\text{def.}}{=} \textstyle\sum_{\boldsymbol{x} \in \Xi} \prod_{i=1}^n \alpha_i(x_i)$$

正規化定数

閉ジャクソンネットワーク (単一サーバ) の均衡分布

K: ネットワーク中のジョブの総数
$\{1, 2, \ldots, n\}$: ノードの集合
$\mu = (\mu_1, \mu_2, \ldots, \mu_n), \boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_n)$: パラメータ

単体中の整数格子点

ネットワーク中のジョブの均衡分布

[K & Matusi 2008]
閉ジャクソンネットワークの積形式解に対して，過去からのカップリングを用いた$O(n^3 \log N)$時間の完璧サンプリング法を設計.

ただし

$$\alpha_i(z) \stackrel{\text{de}}{=}$$

$$G(K) \stackrel{\text{de}}{=} \sum_{\boldsymbol{x} \in \Xi} \prod_{i=1}^{n} \alpha_i(x_i)$$

正規化定数

[2] S. Kijima and Tomomi Matsui, Approximation algorithm and perfect sampler for closed Jackson networks with single servers, SIAM Journal on Computing, 38:4 (2008), 1484--1503.
[3] S. Kijima and Tomomi Matsui, Randomized approximation scheme and perfect sampler for closed Jackson networks with multiple servers, Annals of Operations Research, 162 (2008), 35--55.

私の数理工学: 理論編 (研究紹介)

ロ 完璧サンプリング法 (学生～数理解析研)

- ✓ 分割表 [RSA 2006]
- ✓ 待ち行列ネットワーク [SICOMP 2008, AOR 2008]

ロ メディアン安定マッチング [MathOR 2012]

[1] S. Kijima and Tomomi Matsui, Polynomial time perfect sampling algorithm for two-rowed contingency tables, Random Structures & Algorithms, 29:2 (2006), 243--256.
[2] S. Kijima and Tomomi Matsui, Approximation algorithm and perfect sampler for closed Jackson networks with single servers, SIAM Journal on Computing, 38:4 (2008), 1484--1503.
[3] S. Kijima and Tomomi Matsui, Randomized approximation scheme and perfect sampler for closed Jackson networks with multiple servers, Annals of Operations Research, 162 (2008), 35--55.
[4] Tomomi Matsui, Mitsuo Motoki, Naoyuki Kamatani and S. Kijima, Polynomial time approximate or perfect samplers for discretized Dirichlet distribution, Japan Journal of Industrial and Applied Mathematics, 27:1 (2010), 91--123.
[5] S. Kijima, Sampling from log-super/submodular distributions, in Proc. 7th Hungarian-Japanese Symposium on Discrete Mathematics and Its Applications (2011), 227--236.

完璧サンプリングと単調優モジュラ分布

定理 [5]
分配束上の reversible Hasse walk が 単調な更新関数 を持つ
⇔ 定常分布が**対数優モジュラ**

[1--4]よりも素朴なマルコフ連鎖.
特に[4](離散化Dirichlet分布)は(一般に)
対数優モジュラ**ではない**.

[5] S. Kijima, Sampling from log-super/submodular distributions, in Proc. 7th Hungarian-Japanese Symposium on Discrete Mathematics and Its Applications (2011), 227--236.

私の数理工学: 理論編 (研究宣伝)

- 完璧サンプリング法
  - ✓ 分割表 [RSA 2006]
  - ✓ 待ち行列ネットワーク [SICOMP 2008, AOR 2008]
- **メディアン安定マッチング** [MathOR 2012]
- 確率と計算
  - ✓ 決定性RW [**RSA 2015**, TCS 2017, SIDMA 2018]
- 劣モジュラ関数の変数変換 [**SIDMA 2020**]
- 任意の有限分配束は$M^\natural$凹関数の最小化元族に同型 [**OR lett. 2021**]
- 動的環境下のアルゴリズム論
  - ✓ 増えるクーポン収集 [**SODA 2021**]

[6] S. Kijima and Toshio Nemoto, On randomized approximation for finding a level ideal of a poset and the generalized median stable matchings, Mathematics of Operations Research, 37:2 (2012), 356--371.

安定結婚問題 [Gale & Shapley '62]

2012年ノーベル経済学賞

Shapley and Roth

a. Preference list

Men's lists (favor)					
m_1	w_1	w_2	w_4	w_3	w_5
m_2	w_2	w_1	w_3	w_4	w_5
m_3	w_3	w_4	w_2	w_5	w_1
m_4	w_4	w_3	w_5	w_1	w_2
m_5	w_5	w_3	w_1	w_2	w_4

Women's lists (favor)					
w_1	m_2	m_1	m_3	m_4	m_5
w_2	m_3	m_1	m_2	m_4	m_5
w_3	m_5	m_4	m_3	m_2	m_1
w_4	m_1	m_3	m_4	m_5	m_2
w_5	m_4	m_5	m_1	m_2	m_3

men

1

2

3

4

5

women

1

2

3

4

5

安定マッチング

“駆け落ち”を誘引しないマッチング

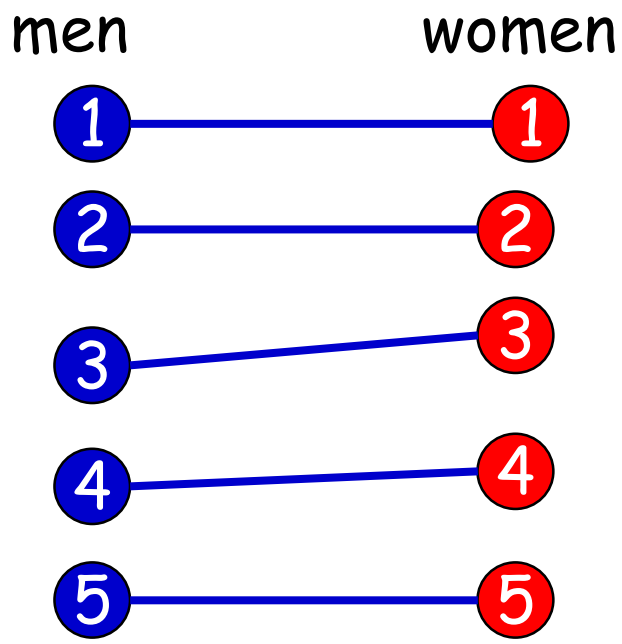
a. Preference list

Men’s lists

m_1	w_1	w_2	w_4	w_3	w_5
m_2	w_2	w_1	w_3	w_4	w_5
m_3	w_3	w_4	w_2	w_5	w_1
m_4	w_4	w_3	w_5	w_1	w_2
m_5	w_5	w_3	w_1	w_2	w_4

Women’s lists

w_1	m_2	m_1	m_3	m_4	m_5
w_2	m_3	m_1	m_2	m_4	m_5
w_3	m_5	m_4	m_3	m_2	m_1
w_4	m_1	m_3	m_4	m_5	m_2
w_5	m_4	m_5	m_1	m_2	m_3



安定マッチングは一つとは限らない

a. Preference list

Men's lists

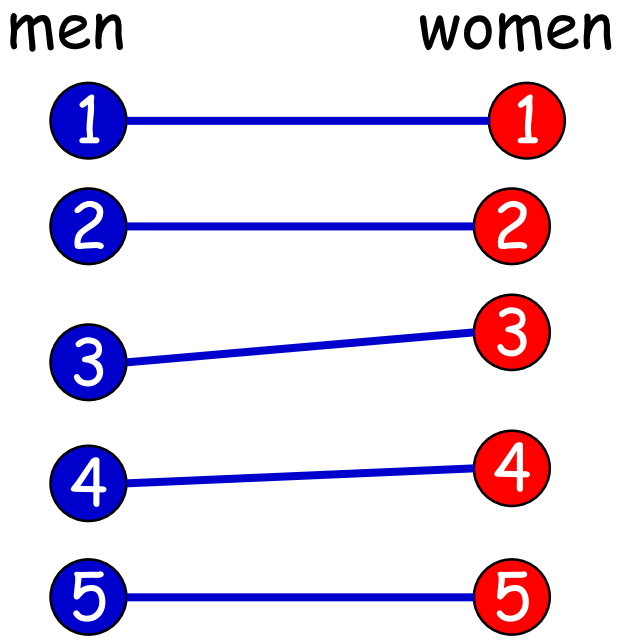
m_1	w_1	w_2	w_4	w_3	w_5
m_2	w_2	w_1	w_3	w_4	w_5
m_3	w_3	w_4	w_2	w_5	w_1
m_4	w_4	w_3	w_5	w_1	w_2
m_5	w_5	w_3	w_1	w_2	w_4

Women's lists

w_1	m_2	m_1	m_3	m_4	m_5
w_2	m_3	m_1	m_2	m_4	m_5
w_3	m_5	m_4	m_3	m_2	m_1
w_4	m_1	m_3	m_4	m_5	m_2
w_5	m_4	m_5	m_1	m_2	m_3

b. Stable matchings (from men's view)

	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7	μ_8
m_1	**1**	**2**	**1**	**2**	**1**	**4**	**2**	**4**
m_2	**2**	**1**	**2**	**1**	**2**	**1**	**1**	**1**
m_3	**3**	**3**	**4**	**4**	**4**	**2**	**4**	**2**
m_4	**4**	**4**	**3**	**3**	**5**	**3**	**5**	**5**
m_5	**5**	**5**	**5**	**5**	**3**	**5**	**3**	**3**

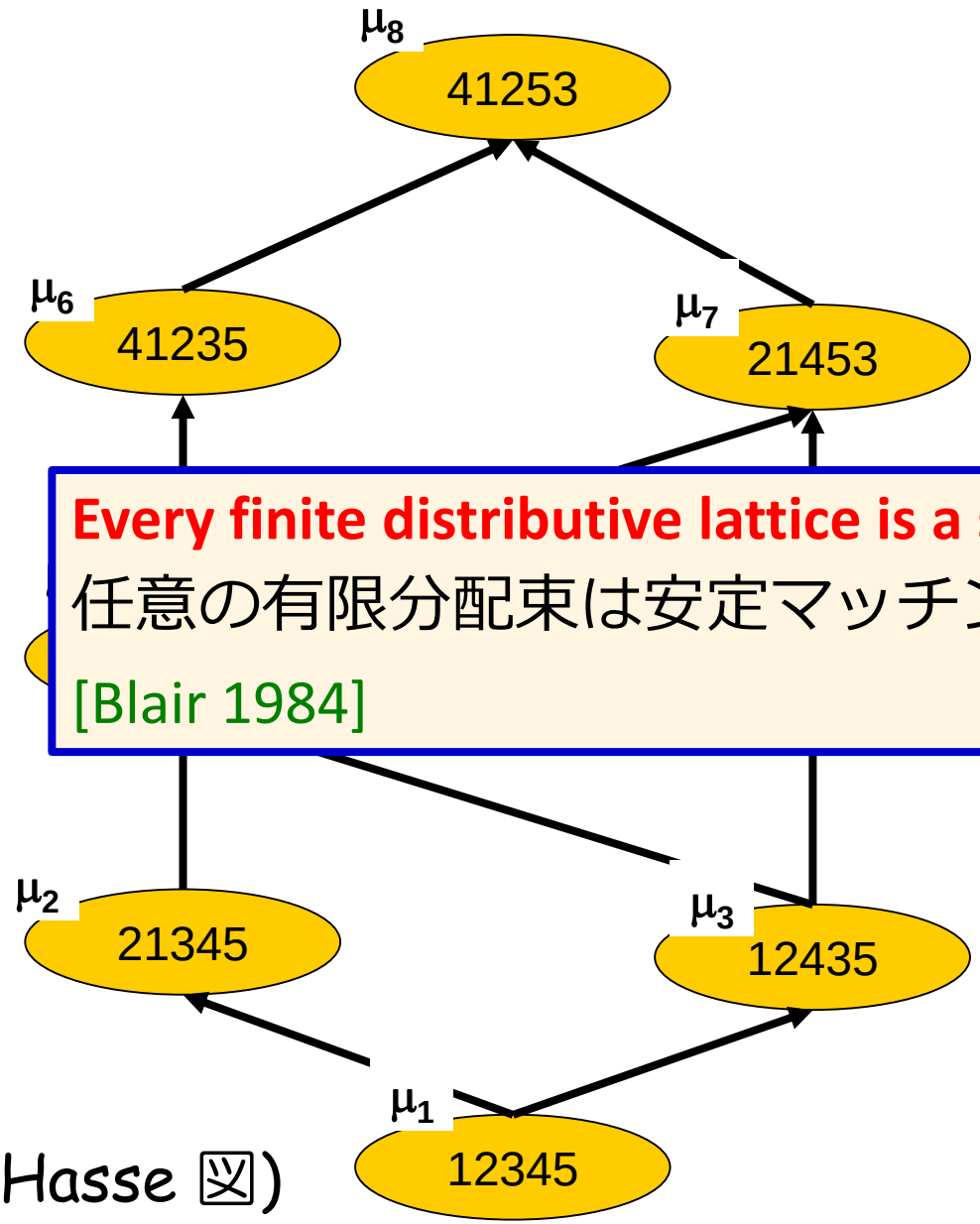


[Conway (Knuth '77,'91)]

安定マッチングは分配束をなす

b. Stable matchings (from men's view)

	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7	μ_8
m_1	1	2	1	2	1	4	2	4
m_2	2	1	2	1	2	1	1	1
m_3	3	3	4	4	4	2	4	2
m_4	4	4	3	3	5	3	5	5
m_5	5	5	5	5	3	5	3	3



Every finite distributive lattice is a set of stable matchings

任意の有限分配束は安定マッチングに同型

[Blair 1984]

favor

favo

男女の間で
の"公平性"

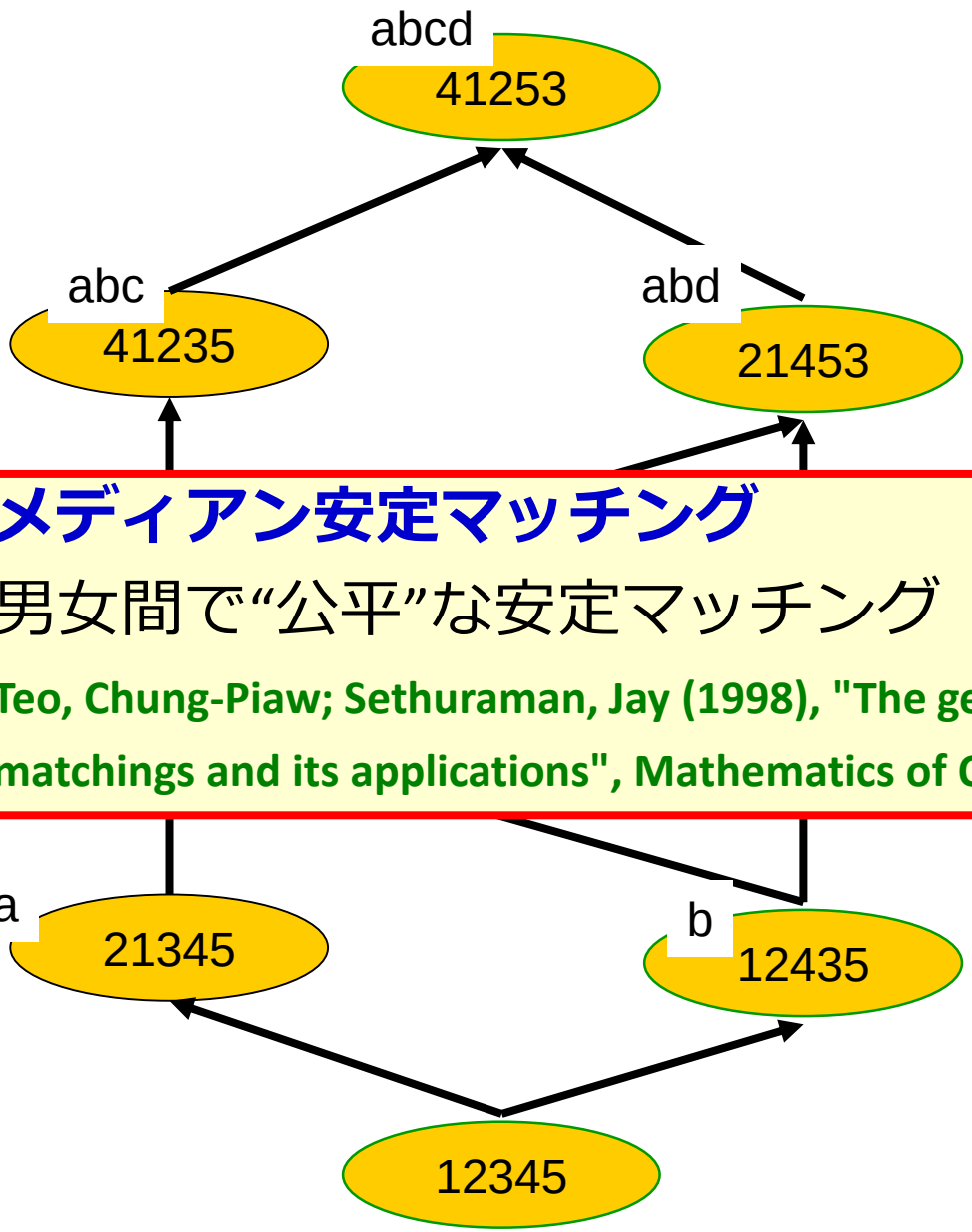
(Hasse 図)

[Teo & Sethuraman 1998]

メディアン安定マッチング

c. GMSMs

m_1	1	1	1	2	2	2	4	4
m_2	2	2	2	1	1	1	1	1
m_3	3	3	4	4	4	4	2	2
m_4	4	4	3	3	3	5	5	5
m_5	5	5	5	5	5	3	3	3
	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8
						μ_1	μ_3	μ_4	μ_7	μ_8



メディアン安定マッチング

男女間で“公平”な安定マッチング

Teo, Chung-Piaw; Sethuraman, Jay (1998), "The geometry of fractional stable matchings and its applications", Mathematics of Operations Research, 23 (4): 874–891.

favo

男女の間での“公平性”

メディアン安定マッチングの計算

[6] S. Kijima and Toshio Nemoto, On randomized approximation for finding a level ideal of a poset and the generalized median stable matchings, Mathematics of Operations Research, 37:2 (2012), 356--371.

定理 [Cheng '08]

i = Ω(N) の時，「i 番目の LI は要素 e を含むか？」は**#P困難**.

定理 [Cheng '08]

i = O(log |P|) の時，i 番目の LI は**多項式時間**で見つかる.

NP問題？

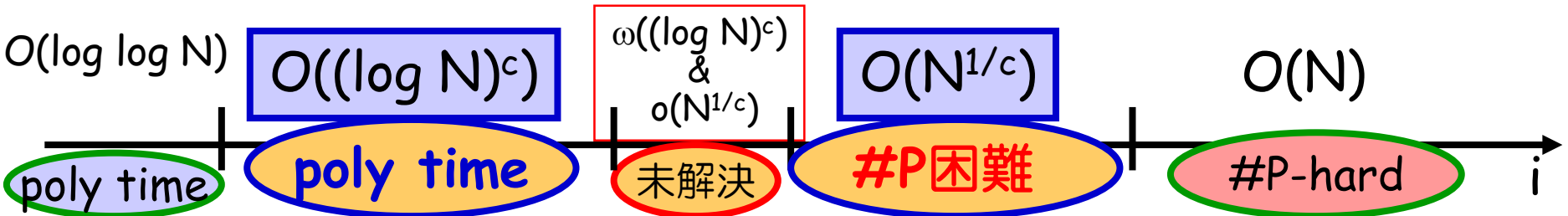
定理 [K, Nemoto '12]

「イデアル S は i 番目の LI か？」は，

i = $O(N^{1/c})$の場合でも **#P困難**．但し $c \geq 1$ は任意の定数.

定理 [K, Nemoto '12]

i = $O(|P|^c)$ の時，i 番目の LI は**多項式時間**で見つかる.



私の数理工学: 理論編 (研究宣伝)

- 完璧サンプリング法
  - ✓ 分割表 [RSA 2006]
  - ✓ 待ち行列ネットワーク [SICOMP 2008, AOR 2008]
- メディアン安定マッチング [MathOR 2012]
- **確率と計算** (2010-)
  - ✓ **決定性RW** [**RSA 2015**, TCS 2017, SIDMA 2018]
  - ✓ 高次元体積の決定性近似 [Algorithmica 2016, **TCS 2020**]
- 対数値エジュー二分布

[7] S. Kijima, Kentaro Koga and Kazuhisa Makino, Deterministic random walks on finite graphs, Random Structures & Algorithms, 46:4 (July 2015), 739--761.

[8] Takeharu Shiraga, Yukiko Yamauchi, S. Kijima and Masafumi Yamashita, Deterministic random walks for rapidly mixing chains, SIAM Journal on Discrete Mathematics, 32:3 (2018), 2180--2193.

[9] Takeharu Shiraga, Yukiko Yamauchi, Shuji Kijima and Masafumi Yamashita, Total variation discrepancy of deterministic random walks for ergodic Markov chains, Theoretical Computer Science, 699 (7 November 2017), 63--74.

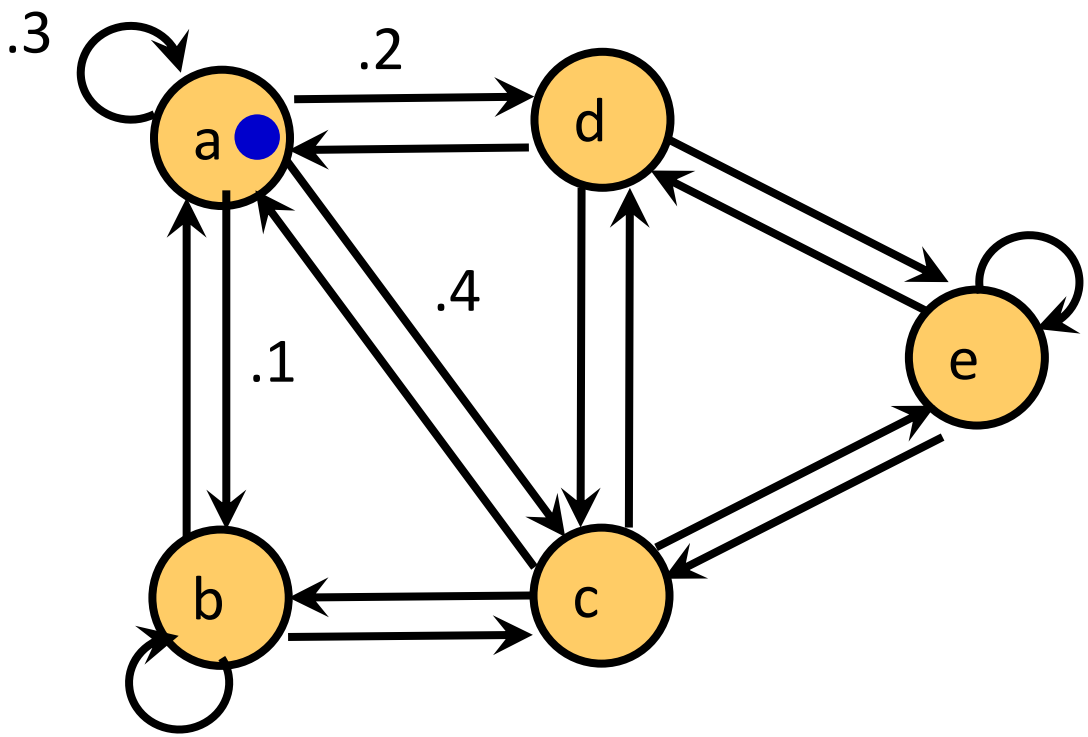
確率と計算

「計算にランダムネスは必要か？」

ランダムウォーク

トークンがグラフ上をランダムウォークする.

- ✓ π^0: 初期分布 (トークンは確率π^0_vで頂点vに居る)
- ✓ P: 推移確率行列 (確率P_{uv}で頂点u頂点vに移動)
- ✓ 時刻tの確率分布$\pi^t := \pi^0 P^t$ (頂点vに居る確率 $(\pi^t)_v$)



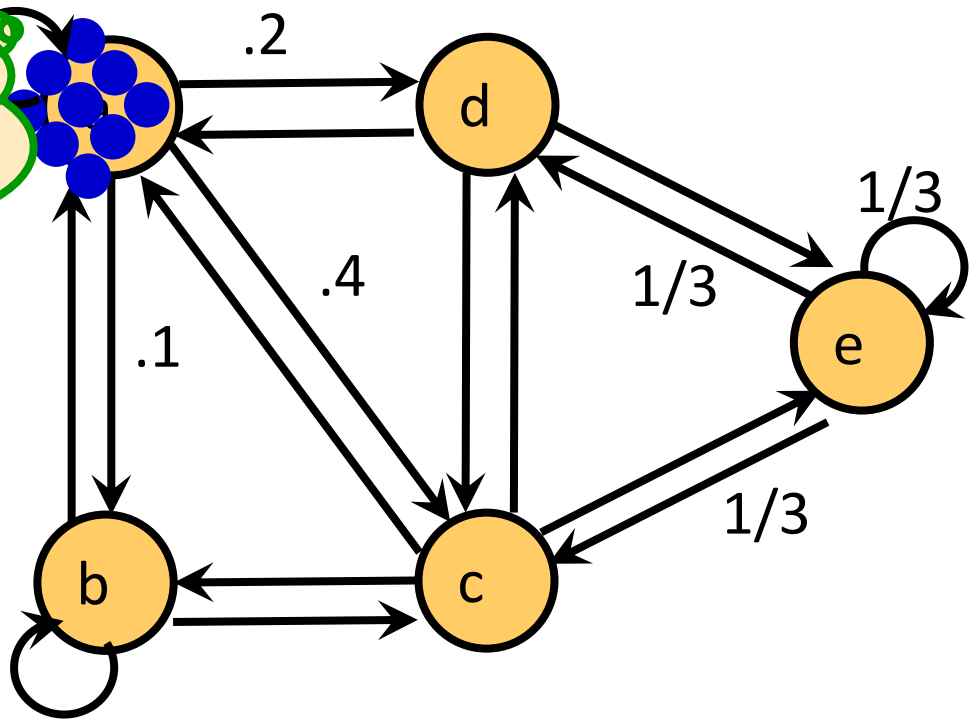
ランダムウォーク (複数トークンによる分布の近似)

N個のトークンがグラフ上を独立にランダムウォークする.

- ✓ μ^0: 初期配置 ($\pi^0 \approx \mu^0/N$)
- ✓ P: 推移確率行列 (確率P_{uv}で頂点u頂点vに移動)
- ✓ 時刻tの期待配置$\mu^t := \mu^0 P^t$ ($\pi^t \approx \mu^t/N$)

(Nが非常に大きければ,)
トークンを約2:4:1:3の割合で
隣接点にばらまいている.

決定性過程でも大差ない?

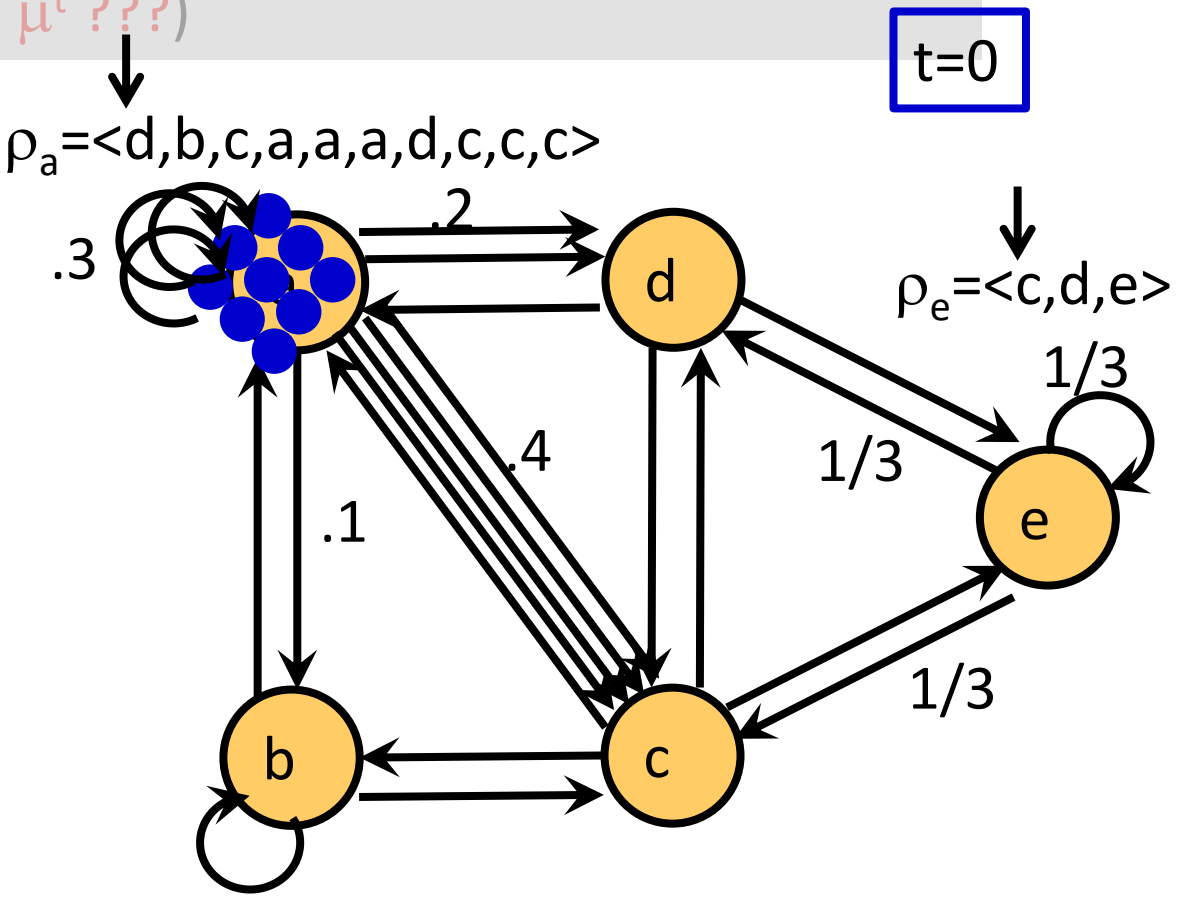


決定性RW (Propp機械; rotor-router)

N個のトークンがグラフ上を移動する.

省略

- ✓ χ^0: 初期配置 ($\chi^0 = \mu^0$)
- ✓ ρ: "**rotor router**" (比率P_{uv}で頂点u頂点vに"順次"移動)
- ✓ 時刻tの配置χ^t ($\chi^t \approx \mu^t$???)



主結果

$\chi_w^{(T)}$: #tokens @ $w \in V$, @ time $T > 0$ in detRW

$\mu_w^{(T)}$: E[#tokens] @ $w \in V$, @ time $T > 0$ in RW

省略

定理 [K, Koga, Makino 10+]

対応する推移確率行列 P の固有値がすべて非負ならば,
任意の多重有向グラフ, 任意の初期状態, 任意のrotor-router,
任意の頂点 w, 任意の時刻 t について,

$$\left|\chi_w^{(T)} - \mu_w^{(T)}\right| \leq 4mn - O(m)$$

但し, n, m は多重グラフの頂点数,枝数.

Remark

「推移確率行列Pの固有値がすべて非負」という条件.

- reversible lazy Markov chainはこの条件を満たす.
  - ✓ MCMC法で使われるマルコフ連鎖
- +Pが対称 ⇒ Pは半正定値行列

主結果

$\chi_w^{(T)}$: #tokens @ $w \in V$, @ time $T > 0$ in detRW

$\mu_w^{(T)}$: E[#tokens] @ $w \in V$, @ time $T > 0$ in RW

定理 [K, Koga, Makino 10+]

対応する推移確率行列 P の固有値がすべて非負ならば,

任

省略

トークン数を増やせば,

deterministic random walk (決定性過程)は

ランダムウォークをよく模倣する.

➢ +Pが対称 ⇒ Pは半正定

$\frac{\mu^{(T)}}{N} \simeq \pi^{(T)}$: 時刻 T の分布

私の数理工学: 理論編 (研究宣伝)

- 完璧サンプリング法
  - ✓ 分割表 [RSA 2006]
  - ✓ 待ち行列ネットワーク [SICOMP 2008, AOR 2008]
- メディアン安定マッチング [MathOR 2012]
- **確率と計算** (2010-)
  - ✓ 決定性RW [**RSA 2015**, TCS 2017, SIDMA 2018]
  - ✓ **高次元体積の決定性近似** [Algorithmica 2016, **TCS 2020**]
- 対数優モジュラ分布

[10] Ei Ando and S. Kijima, An FPTAS for the volume computation of 0-1 knapsack polytopes based on approximate convolution, Algorithmica, 76:4 (December 2016), 1245--1263.

[11] Ei Ando and S. Kijima, An FPTAS for the volume of some V-polytopes --- It is hard to compute the volume of the intersection of two cross-polytopes, Theoretical Computer Science, 833:12 (September 2020), 87--106.

高次元多面体の体積計算の決定性近似

Lovász [1986]

“membership oracleで与えられたn次元凸体の体積に対して，

近似比1.999^nの**決定性**多項式時間アルゴリズムは**存在しない**．”

乱択近似 (1980～)

- ✓ Jerrum, Valiant, Vazirani (1986): 近似数え上げ≃ランダムサンプリング
- ✓ Toda (1991): $\mathrm{PH} \subseteq \mathrm{P}^{\#\mathrm{P}}$ (Gödel賞)
- ✓ Dyer, Frieze, Kanna (1991): 高次元凸体体積の乱択近似可能 (Fulkerson賞)
- ✓ Jerrum, Sinclair, Vigoda (2004): パーマネントの乱択近似可能 (Fulkerson賞)

決定性近似 (話者の研究 2014～)

- ✓ ナップサック多面体の体積は決定性近似可能(FPTAS) [Algorithmica16]
  - H多面体: 線形不等式系で記述される多面体
- ✓ ‘双対’ナップサック多面体の体積は決定性近似可能(FPTAS) [TCS 20]
  - V多面体: 端点集合で記述される多面体
  - n次元L_1球2個の交差領域(resp. union)の体積計算は#P困難

私の数理工学: 理論編 (研究宣伝)

- 完璧サンプリング法
  - ✓ 分割表 [RSA 2006]

[12] Junpei Nakashima, Yukiko Yamauchi, S. Kijima and Masafumi Yamashita, Finding submodularity hidden in symmetric difference, SIAM Journal on Discrete Mathematics, 34:1 (2020), 571--585.

[13] Tomohito Fujii and Shuji Kijima, Every finite distributive lattice is isomorphic to the minimizer set of an $M^\natural$-concave set function, Operations Research Letters, 49:1 (January 2021), 1--4.

- **対数優モジュラ分布** (2016-2020)
  - ✓ 劣モジュラ関数の変数変換 **[SIDMA 2020]**
  - ✓ 任意の有限分配束は$M^\natural$凹関数の最小化元族に同型 **[OR lett. 2021]**
- 動的環境下のアルゴリズム論
  - ✓ 増えるクーポン収集 **[SODA 2021]**

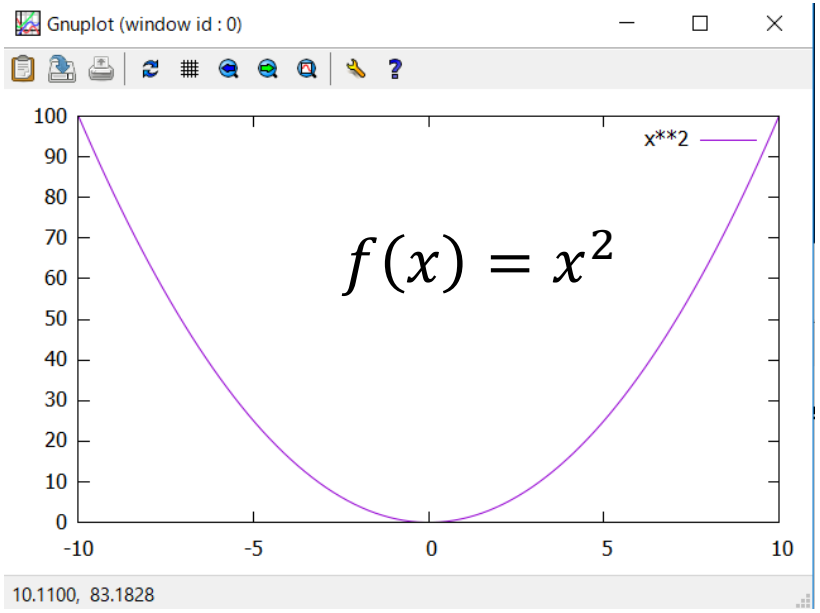
凸性はアフィン変換で保存される

定義：凸関数

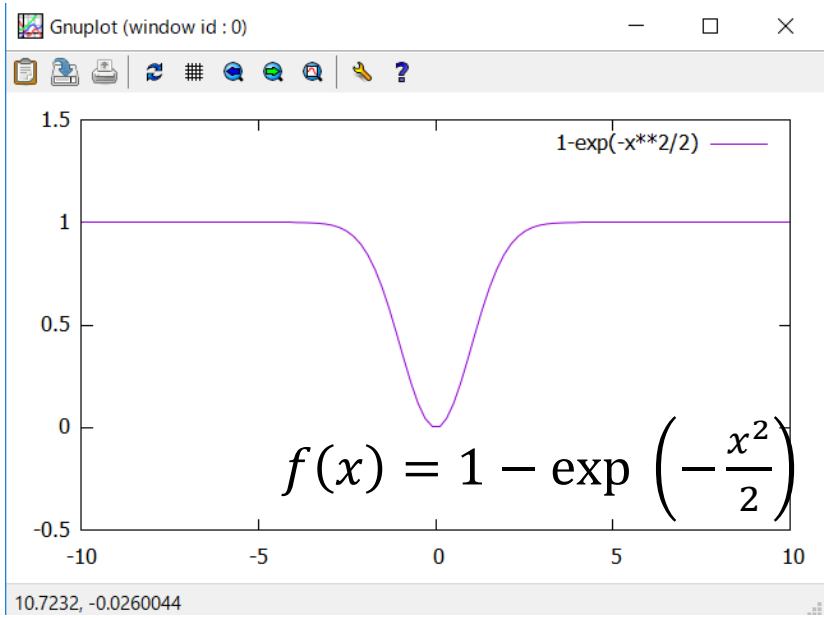
関数$f: \mathbb{R}^n \to \mathbb{R}$に対し，任意の$\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$と任意の$\lambda \geq 0$に対して

$$\lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y}) \geq f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y})$$

が成り立つとき，fは凸関数という．



凸関数



凸ではない関数

問題

$$f(x,y) = 2x^2 + 2xy + 5y^2$$

は凸関数か？

問題

$$f(x,y) = 2x^2 + 2xy + 5y^2$$

は凸関数か？

解答

変数変換

$$x := s + 2t$$
$$y := s - t$$

を施す.

$$\begin{aligned} g(s,t) &= f(s+2t, s-t) \\ &= 2(s+2t)^2 + 2(s+2t)(s-t) + 5(s-t)^2 \\ &= \left(2s^2 + 8st + 8t^2\right) + \left(2s^2 + 2st - 4t^2\right) + \left(5s^2 - 10st + 5t^2\right) \\ &= 9s^2 + 9t^2 \end{aligned}$$

$\boldsymbol{g(s,t)}$は凸.

問題

$$f(x,y) = x^2 + 4xy + 3y^2$$

は凸関数か？

問題

$$f(x,y) = x^2 + 4xy + 3y^2$$

は凸関数か？

解答

変数変換

$$x := s + 3t$$
$$y := s - t$$

を施す.

$$\begin{aligned} g(s,t) &= f(s+3t, s-t) \\ &= (s+3t)^2 + 4(s+3t)(s-t) + 3(s-t)^2 \\ &= \left(s^2 + 6st + 9t^2\right) + \left(4s^2 + 8st - 12t^2\right) + \left(3s^2 - 6st + 3t^2\right) \\ &= 8s^2 + 8st \end{aligned}$$

$g(s,t)$は凸ではない.

e.g., $g(1,-2) = 8 - 16 = -8$, $g(-1,2) = 8 - 16 = -8$,

$$g\left(\frac{1}{2}(1,-2) + \frac{1}{2}(-1,2)\right) = g(0,0) = 0 > \frac{1}{2}g(1,-2) + \frac{1}{2}g(-1,2).$$

凸性はアフィン変換で保存される

定義：凸関数

関数$f: \mathbb{R}^n \to \mathbb{R}$に対し，任意の$\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$と任意の$\lambda \geq 0$に対して

$$\lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y}) \geq f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y})$$

が成り立つとき，fは凸関数という.

定理 cf. [Rockafellar]

$h: \mathbb{R}^n \to \mathbb{R}^n$はアフィン写像とする.

関数$f: \mathbb{R}^n \to \mathbb{R}$**が**凸のとき$g \coloneqq f \circ h$は凸関数.

(i.e., $h = A\boldsymbol{x} + \boldsymbol{b}$: $A \in \mathbb{R}^{n \times n}$, $\boldsymbol{b} \in \mathbb{R}^n$ のとき,

$g(\boldsymbol{x}) = f(A\boldsymbol{x} + \boldsymbol{b})$は凸関数)

凸性はアフィン変換で保存される

定義：凸関数

関数$f: \mathbb{R}^n \to \mathbb{R}$に対し[illegible]対して

[illegible]が成り[illegible]

離散関数のアナロジー？

⇒劣モジュラ関数

定理 c[illegible]

$h: \mathbb{R}^n \to$[illegible]

関数$f: \mathbb{R}^n \to \mathbb{R}$[illegible]数.

(i.e., $h = A\boldsymbol{x} + \boldsymbol{b}$: $A \in$[illegible], $\boldsymbol{b} \in \mathbb{R}^n$のとき,

$g(\boldsymbol{x}) = f(A\boldsymbol{x} + \boldsymbol{b})$は凸関数)

定義: 劣モジュラ

集合関数 $f:\{0,1\}^V \to \mathbb{R}$ に対して，$\Phi_f:\{0,1\}^V \times \{0,1\}^V \to \mathbb{R}$を

$$\Phi_f(X,Y) \coloneqq f(X) + f(Y) - f(X \cup Y) - f(X \cap Y)$$

とする.

- 集合関数fが劣モジュラとは，任意の$X, Y \in \{0,1\}^V$に対して $\Phi_f(X,Y) \geq 0$が成り立つこととする.

定義: 対称差変換

対称差写像$\sigma_S:\{0,1\}^V \to \{0,1\}^V$ を

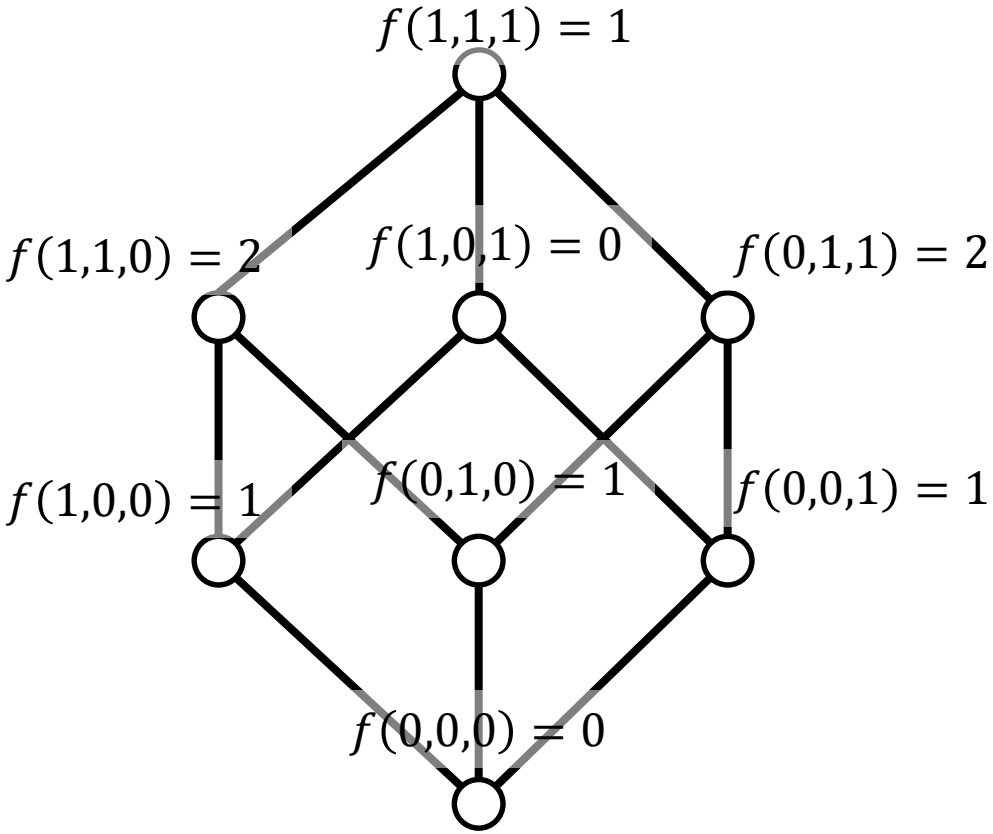
$\sigma_S(X) \coloneqq X \oplus S \ \ (X \in \{0,1\}^V)$

とする.

集合関数$g:\{0,1\}^V \to \mathbb{R}$ を

$$g \coloneqq f \circ \sigma_S$$

を S による f の**対称差変換**と呼ぶ.



定義: 劣モジュラ

集合関数 $f:\{0,1\}^V \to \mathbb{R}$ に対して，$\Phi_f:\{0,1\}^V \times \{0,1\}^V \to \mathbb{R}$を

$$\Phi_f(X,Y) \coloneqq f(X) + f(Y) - f(X \cup Y) - f(X \cap Y)$$

とする.

- 集合関数fが劣モジュラとは，任意の$X, Y \in \{0,1\}^V$に対して $\Phi_f(X,Y) \geq 0$が成り立つこととする.

$g \coloneqq f \circ \sigma_{(1,0,0)}$の場合

定義: 対称差変換

対称差写像$\sigma_s:\{0,1\}^V \to \{0,1\}^V$ を

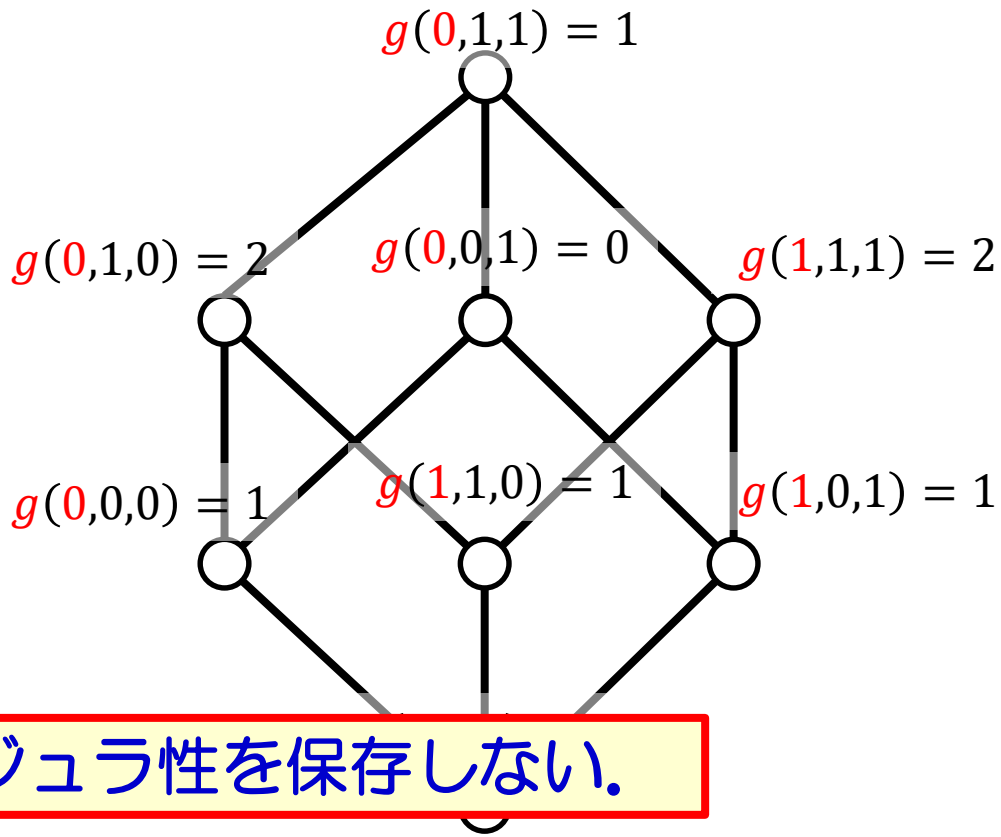
$\sigma_s(X) \coloneqq X \oplus S \ \ (X \in \{0,1\}^V)$

とする.

集合関数$g:\{0,1\}^V \to \mathbb{R}$ を

$$g \coloneqq f \circ \sigma_S$$

を S による f の**対称差変換**と呼ぶ.



一般に，対称差変換は劣モジュラ性を保存しない.

定義: 劣モジュラ

集合関数 $f:\{0,1\}^V \to \mathbb{R}$ に対して，$\Phi_f:\{0,1\}^V \times \{0,1\}^V \to \mathbb{R}$を

$$\Phi_f(X,Y) \coloneqq f(X) + f(Y) - f(X \cup Y) - f(X \cap Y)$$

とする.

- 集合関数fが劣モジュラとは，任意の$X, Y \in \{0,1\}^V$に対して $\Phi_f(X,Y) \geq 0$が成り立つこととする.

$g \coloneqq f \circ \sigma_{(1,0,1)}$の場合

定義: 対称差変換

対称差写像$\sigma_s:\{0,1\}^V \to \{0,1\}^V$ を

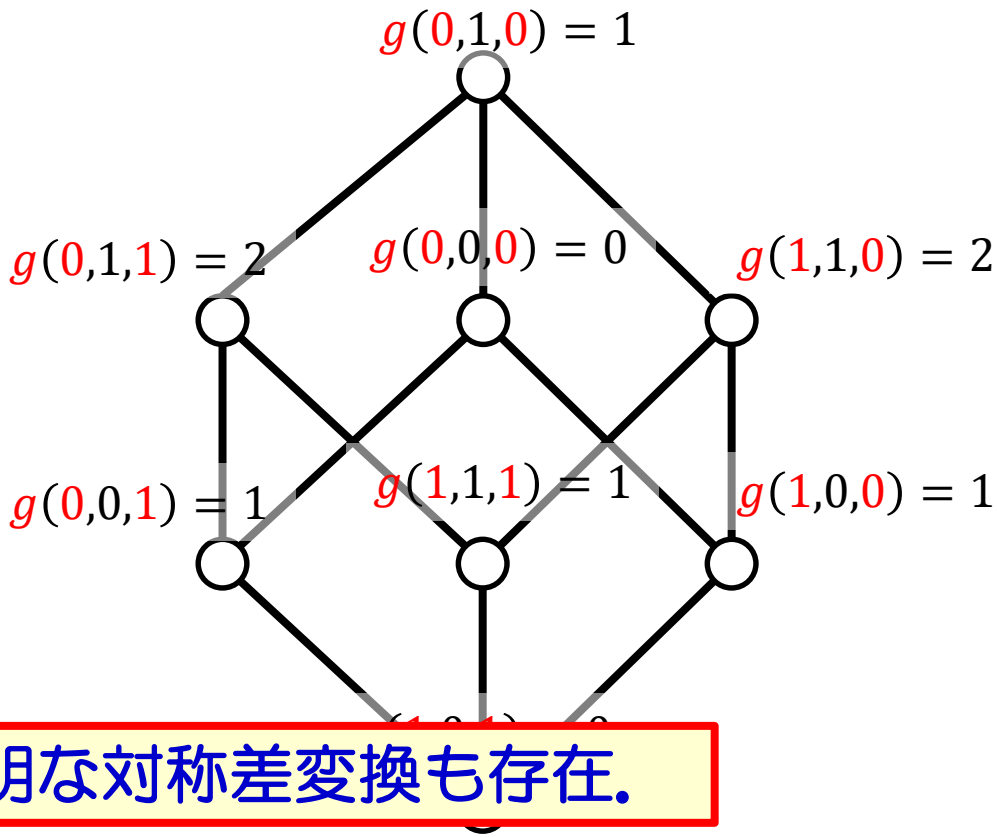
$\sigma_s(X) \coloneqq X \oplus S \ \ (X \in \{0,1\}^V)$

とする.

集合関数$g:\{0,1\}^V \to \mathbb{R}$ を

$$g \coloneqq f \circ \sigma_S$$

を Sによるfの**対称差変換**と呼ぶ.



劣モジュラ性を保存する非自明な対称差変換も存在.

定義: 劣モジュラ

集合関数 $f:\{0,1\}^V \to \mathbb{R}$ に対して，$\Phi_f:\{0,1\}^V \times \{0,1\}^V \to \mathbb{R}$を

$$\Phi_f(X,Y) \coloneqq f(X) + f(Y) - f(X \cup Y) - f(X \cap Y)$$

とする.

- 集合関数fが劣モジュラとは，任意の$X, Y \in \{0,1\}^V$に対して $\Phi_f(X,Y) \geq 0$が成り立つこととする.

$g \coloneqq f \circ \sigma_{(1,0,1)}$の場合

定義

対称

$\sigma_S(X) \coloneqq X \oplus S$ (λ

とする.

集合関数$g:\{0,1\}^V \to \mathbb{R}$ を

$$g \coloneqq f \circ \sigma_S$$

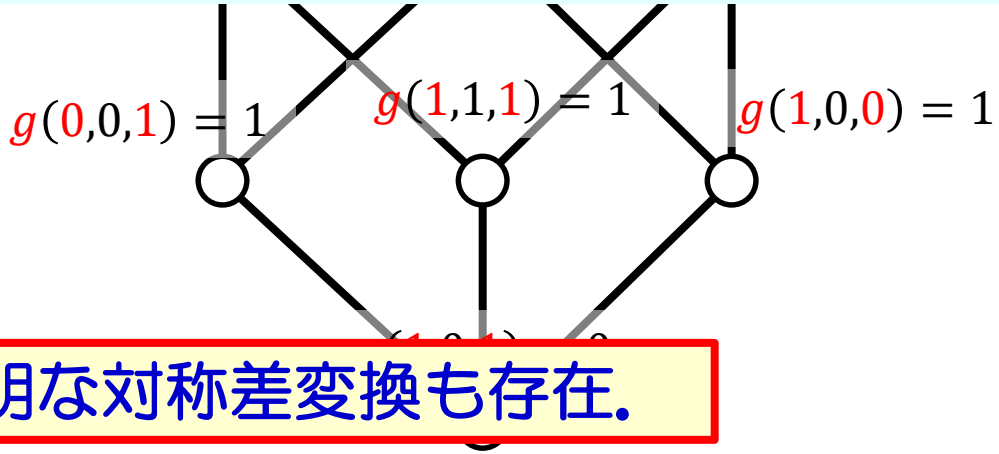
を Sによるfの**対称差変換**と呼ぶ.

[Nakashima et al.]

劣モジュラ性を保存する対称差変換の特徴づけ.

狭義劣モジュラ関数なら$O(d)$時間で発見.

[12] Junpei Nakashima, Yukiko Yamauchi, S. Kijima and Masafumi Yamashita, Finding submodularity hidden in symmetric difference, SIAM Journal on Discrete Mathematics, 34:1 (2020), 571--585.



劣モジュラ性を保存する非自明な対称差変換も存在.

Operations Research Letters 49 (2021) 1–4

Contents lists available at ScienceDirect

Operations Research Letters

journal homepage: www.elsevier.com/locate/orl

Every finite distributive lattice is isomorphic to the minimizer set of an $M^\natural$-concave set function

Tomohito Fujii, Shuji Kijima *

Graduate School of Information Science and Electronic Engineering, Kyushu University, Japan

任意の有限分配束は安定マッチングに同型

ARTICLE INFO

Accepted 28 October 2020
Available online 3 November 2020

Keywords:
$M^\natural$-concave
Valuated matroid
Submodular
Distributive lattice

ABSTRACT

… the class of $M^\natural$-concavity is … minimizers of a submodular function forms a distributive lattice, where every finite distributive lattice is possible to appear. It is a natural question whether every finite distributive lattice appears as the minimizer set of an $M^\natural$-concave set function. This paper affirmatively answers the question.

© 2020 Elsevier B.V. All rights reserved.

答え: 対数M#凸分布でもサンプリングは依然難しい.

1. Introduction

A set function $f: 2^N \to \mathbb{R}$ for a finite set N is *submodular* if

$$f(X)+f(Y) \geq f(X \cup Y)+f(X \cap Y)$$

holds for any $X, Y \in 2^N$. Submodularity is often regarded as a discrete analog of convexity [9,19,21,26]. A submodular function is

Proposition 2 (*See e.g., [9]*). *Given a finite poset $\mathcal{P} = (N, \preccurlyeq)$, let $f: 2^N \to \mathbb{R}$ be defined by*

$$f(X) = |\{j \in N \setminus X : \exists i \in X \text{ such that } j \prec i\}| \quad (1)$$

for any $X \in 2^N$, where $j \prec i$ denotes $j \preccurlyeq i$ and $j \neq i$. Then f is submodular, and it satisfies

$$f(X) \begin{cases} = 0 & \text{if } X \in \mathcal{I}(\mathcal{P}), \end{cases}$$

II. 私の数理工学: 実践編

使ってこその数理工学

応用(お手伝い)研究でも、貢献してきました。

ロ 学習理論 [ISAAC2011, ALT2012, 他]

[14] Daiki Suehiro, Kohei Hatano, S. Kijima, Eiji Takimoto and Kiyohito Nagano, Online prediction under submodular constraints, Proc. ALT 2012, 260--274.

[15] Shota Yasutake, Kohei Hatano, S. Kijima, Eiji Takimoto and Masayuki Takeda, Online linear optimization over permutations, Proc. ISAAC 2011, 534--543.

ロ 分散計算論 [JACM 2017, TOCT 2021, 他]

[16] Akihiro Monde, Yukiko Yamauchi, Shuji Kijima and Masafumi Yamashita, Can a skywalker localize the midpoint of a rope?, ACM Transactions on Computation Theory, 13:3 (2021), Article 17, 1--23.

[17] Yukiko Yamauchi, Taichi Uehara, Shuji Kijima and Masafumi Yamashita, Plane formation by synchronous mobile robots in the three dimensional Euclidean space, Journal of the ACM, 64:3 (2017), Article 16.

いつも論文化がうまくいくわけではない。

- 単調CFTP ⇔ 対数優モジュラ分布
  - [5] S. Kijima, Sampling from log-super/submodular distributions, in Proc. 7th Hungarian-Japanese Symposium on Discrete Mathematics and Its Applications (2011), 227--236.
- 最大ポピュラーマッチングの構成員は一意
  - Mizuki Hirakawa, Yukiko Yamauchi, *S. Kijima, Masafumi Yamashita, On the structure of popular matchings in the stable marriage problem – who can join a popular matching?, MATCH-UP 2015, Glasgow, UK, April 16-18, 2015 (April 17).
  - **平川瑞樹, The Family of Vertex Sets of Popular Matchings in The Stable Marriage Problem, 九大システム情報 修士論文, 2015.**
- 大規模系統樹の作成 (マルウェア分類)
  - *米巧，山内由紀子, 来嶋秀治, 山下雅史, 大規模マルウェアデータ群に対する系統樹推定, 日本オペレーションズ・リサーチ学会 2016年春季研究発表会, 慶應大学矢上キャンパス（東京都新宿区）, 2016年3月17日-18日(3月18日).
  - **米巧, 大規模マルウェアデータ群に対する系統樹推定, 九大システム情報 修士論文, 2016.**
- ストリーミングデータの$O(\log\log n)$領域サンプリング
  - [18] Masatora Ogata, Yukiko Yamauchi, Shuji Kijima and Masafumi Yamashita,A randomized algorithm for finding frequent elements in streams using O(loglog N) space,in Proc. ISAAC 2011, 514--523.

産官学連携のープロジェクトのーチームのー研究課題

大規模マルウェアデータ群に対する系統樹推定

[1]九州大学 システム情報科学府 / [2]九州先端科学技術研究所
〇米 巧[1,2], 山内 由紀子[1], 来嶋 秀治[1,2], 山下 雅史[1,2]

謝辞
本研究は, 総務省「国際連携によるサイバー攻撃の予知技術の研究開発」の支援を受けた.

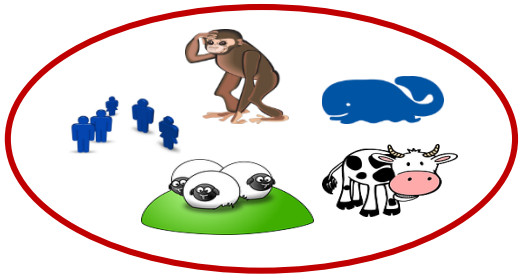
数理工学: 実践編

Cilibrasi & Vitanyiでマルウェア分類とかどうでしょう？

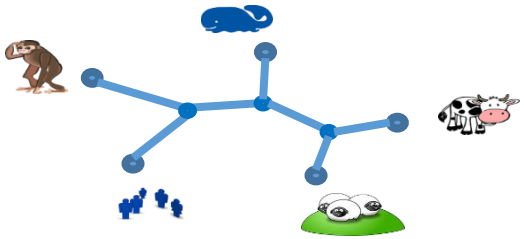
正規圧縮距離を用いたデータ分類

R. Cilibrasi, P. Vitanyi, Clusterring by compression, IEEE Trans Information Theory, 51(2005), 1523--1545.

CilibrasiとVitanyiによるデータ分類概要



	クジラ	チンパンジー	ウシ	人間	ヒツジ
クジラ	0.013152	0.768324	0.599067	0.771675	0.600764
チンパンジー	0.768324	0.013269	0.758214	0.289562	0.800126
ウシ	0.599067	0.758214	0.012968	0.806397	0.437699
人間	0.771675	0.289562	0.806397	0.013047	0.784512
ヒツジ	0.600764	0.800126	0.437699	0.784512	0.012524



データ群

↓

データ間の距離を定義

↓

距離の木距離近似
(系統樹作成)

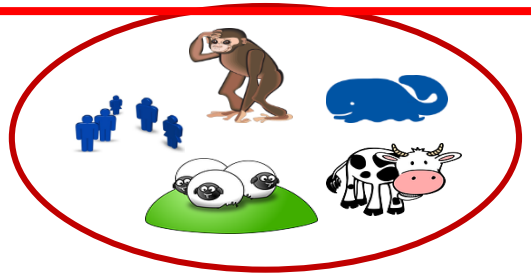
正規圧縮距離

4点法(quartet method)

正規圧縮距離を用いたデータ分類

R. Cilibrasi, P. Vitanyi, Clusterring by compression, IEEE Trans Information Theory, 51(2005), 1523--1545.

CilibrasiとVitanyiによるデータ分類概要



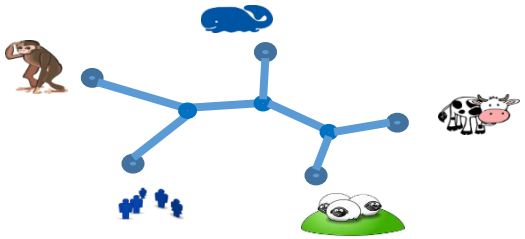
データ群

正規圧縮距離

	クジラ	チンパンジー	ウシ	人間	ヒツジ
クジラ	0.013152	0.768324	0.599067	0.771675	0.600764
チンパンジー	0.768324	0.013269	0.758214	0.289562	0.800126
ウシ	0.599067	0.758214	0.012968	0.806397	0.437699
人間	0.771675	0.289562	0.806397	0.013047	0.784512
ヒツジ	0.600764	0.800126	0.437699	0.784512	0.012524

データ間の距離を定義

4点法(quartet method)



距離の木距離近似
(系統樹作成)

正規圧縮距離

R. Cilibrasi, P. Vitanyi, Clusterring by compression, IEEE Trans Information Theory, 51(2005), 1523--1545.

データ間の距離(類似度)を圧縮度に基づいて定義

Cf. コルモゴロフ複雑度

M.Li, X. Chen, X.Li, B. Ma, and P. Vitanyi, The similarity metric, IEEE Transactions on Information Therory, 50, 3250--3264, 2004

$$\mathrm{NCD}(x,y) = \frac{\max\{c(xy) - c(x),\ c(yx) - c(y)\}}{\max\{c(x),\ c(y)\}}$$

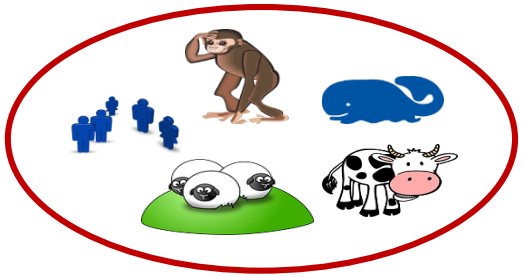
$c(x)$: 文字列xを圧縮プログラムCを用いて圧縮した文字列長
xy : 文字列xと文字列yを連結した文字列

↳ 文字列で表現されるデータであれば適用可能

正規圧縮距離を用いたデータ分類

R. Cilibrasi, P. Vitanyi, Clusterring by compression, IEEE Trans Information Theory, 51(2005), 1523--1545.

CilibrasiとVitanyiによるデータ分類概要



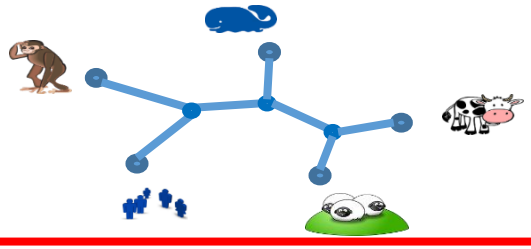
データ群

正規圧縮距離

	クジラ	チンパンジー	ウシ	人間	ヒツジ
クジラ	0.013152	0.768324	0.599067	0.771675	0.600764
チンパンジー	0.768324	0.013269	0.758214	0.289562	0.800126
ウシ	0.599067	0.758214	0.012968	0.806397	0.437699
人間	0.771675	0.289562	0.806397	0.013047	0.784512
ヒツジ	0.600764	0.800126	0.437699	0.784512	0.012524

データ間の距離を定義

4点法(quartet method)



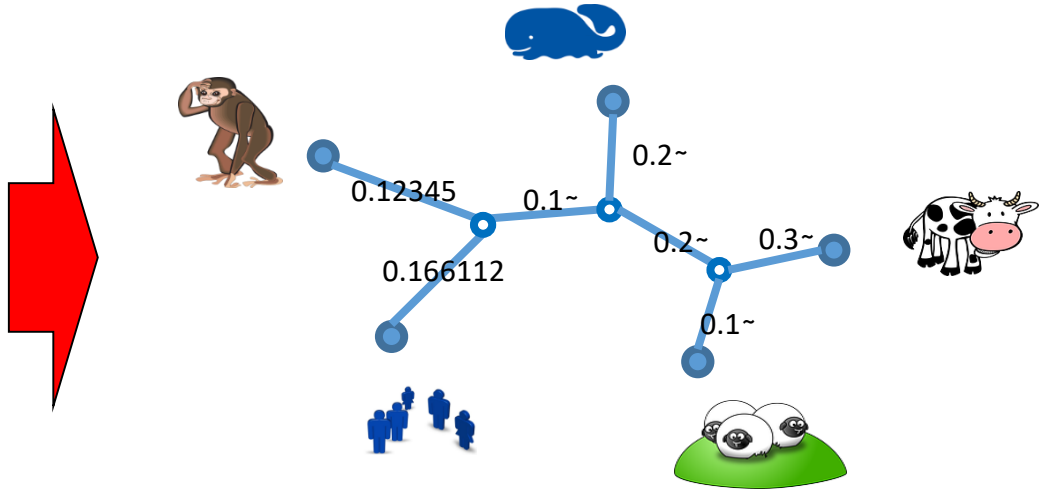
距離の木距離近似
(系統樹作成)

系統樹推定

R. Cilibrasi, P. Vitanyi, Clusterring by compression, IEEE Trans Information Theory, 51(2005), 1523--1545.

距離行列を**木距離**で近似する.

	クジラ	チンパンジー	ウシ	人間	ヒツジ
クジラ	0.013152	0.768324	0.599067	0.771675	0.600764
チンパンジー	0.768324	0.013269	0.758214	0.289562	0.800126
ウシ	0.599067	0.758214	0.012968	0.806397	0.437699
人間	0.771675	0.289562	0.806397	0.013047	0.784512
ヒツジ	0.600764	0.800126	0.437699	0.784512	0.012524

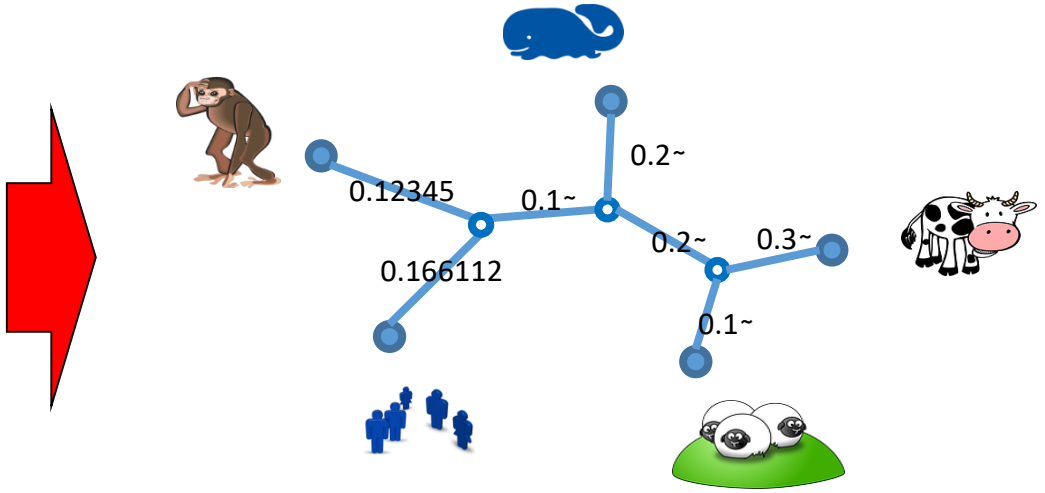


系統樹推定

R. Cilibrasi, P. Vitanyi, Clusterring by compression, IEEE Trans Information Theory, 51(2005), 1523--1545.

入力：データ間の距離行列D
出力：各データに対応する葉をもつ系統樹T

	クジラ	チンパンジー	ウシ	人間	ヒツジ
クジラ	0.013152	0.768324	0.599067	0.771675	0.600764
チンパンジー	0.768324	0.013269	0.758214	0.289562	0.800126
ウシ	0.599067	0.758214	0.012968	0.806397	0.437699
人間	0.771675	0.289562	0.806397	0.013047	0.784512
ヒツジ	0.600764	0.800126	0.437699	0.784512	0.012524

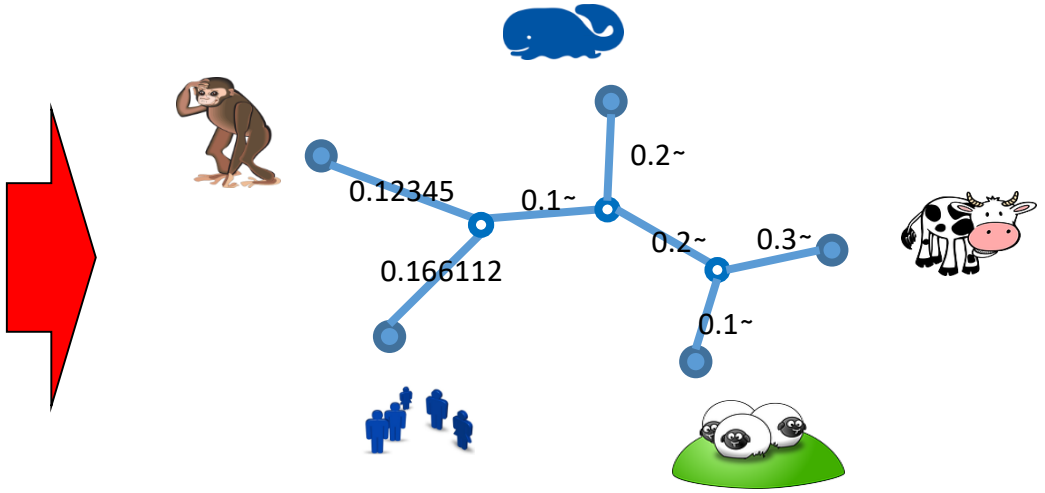


系統樹推定

R. Cilibrasi, P. Vitanyi, Clusterring by compression, IEEE Trans Information Theory, 51(2005), 1523--1545.

入力：データ間の距離行列D
出力：各データに対応する葉をもつ系統樹T

	クジラ	チンパンジー	ウシ	人間	ヒツジ
クジラ	0.013152	0.768324	0.599067	0.771675	0.600764
チンパンジー	0.768324	0.013269	0.758214	0.289562	0.800126
ウシ	0.599067	0.758214	0.012968	0.806397	0.437699
人間	0.771675	0.289562	0.806397	0.013047	0.784512
ヒツジ	0.600764	0.800126	0.437699	0.784512	0.012524



マルウェア分類に使ってみた.

数理工学: 実践編

Cilibrasi & Vitanyiでマルウェア分類とかどうでしょう？

こんな感じです。とりあえず50件でやってみました。

データ数$n = 5 \times 10^4$.

あのね、データは5万件あるんだよ。

- 距離行列作成の時間/領域計算量 $O(n^2) = 2.5 \times 10^9$

  ※ 便宜上，ここでは文字列圧縮にかかる時間は1とする.

- 系統樹作成の時間計算量
  - ✓ 4点法 [Cilibrasi &Vitányi 05]: ヒューリスティクス (NP困難な最小化問題)
  - ✓ 近隣結合法[Saitou & Nei 87]: $O(n^3) = 1.25 \times 10^{14}$

[Yone et al.]

距離行列計算+系統樹作成: $O(s^3 + sn) = 10^9 + 5 \times 10^7$

$s = 10^3$はサンプルサイズ(パラメータ)

高速化のための提案手法

概要

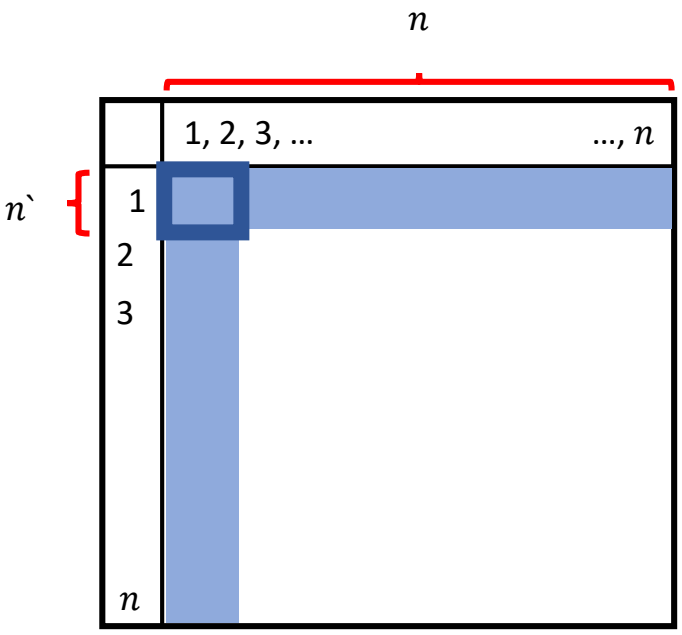
1. 一部のデータに対し, 系統樹を作成.
   n’ : ベースサイズ (n’≪ n)

2. 残りのデータをベース系統樹上で分割.

3. 距離行列再計算.

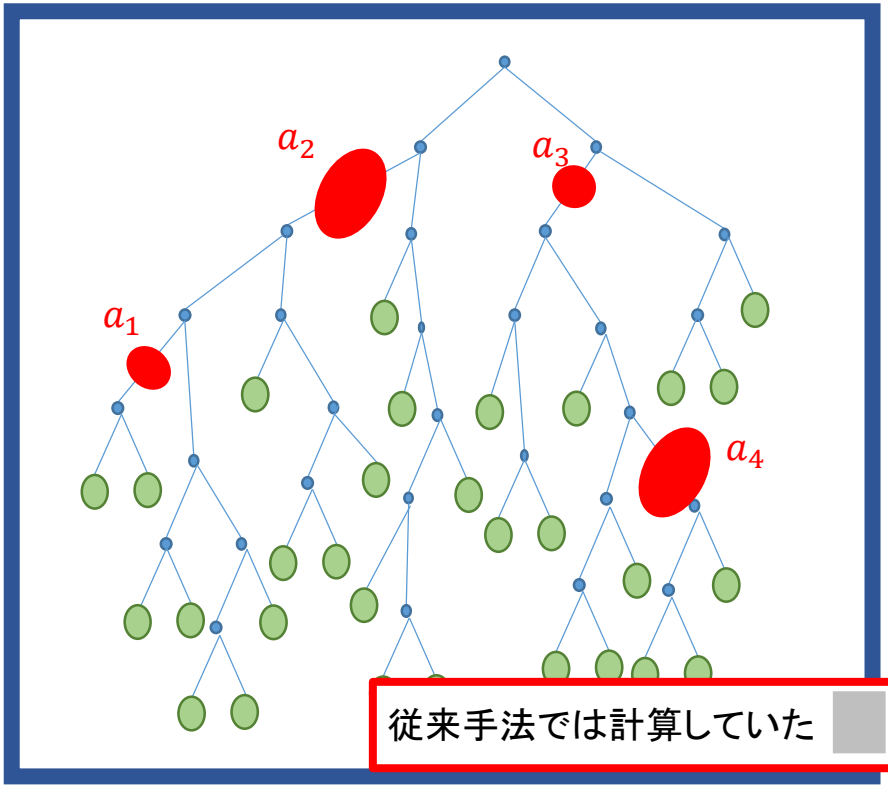
4. 頂点の縮約.

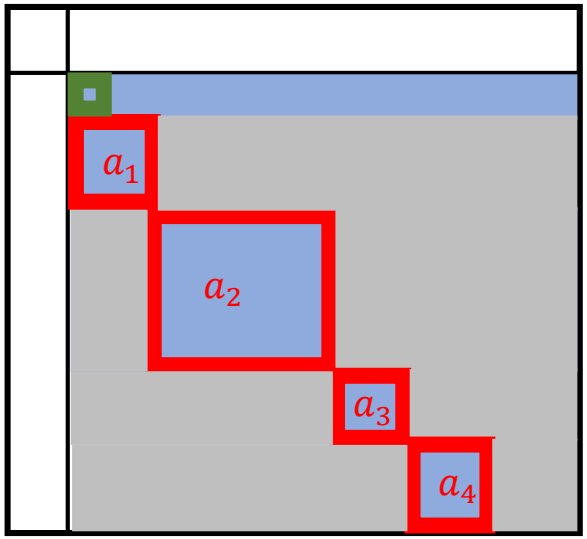
5. 系統樹作成.



提案手法のイメージ

の部分で系統樹を作成，残りのデータをその系統樹上で分割し，ブロックごとに系統樹を作成する．





従来手法では計算していた　の部分を計算せずに系統樹を作成．

私の思う数理工学: 実践編

一論文一主張

実践編を論文にするのは難しい.

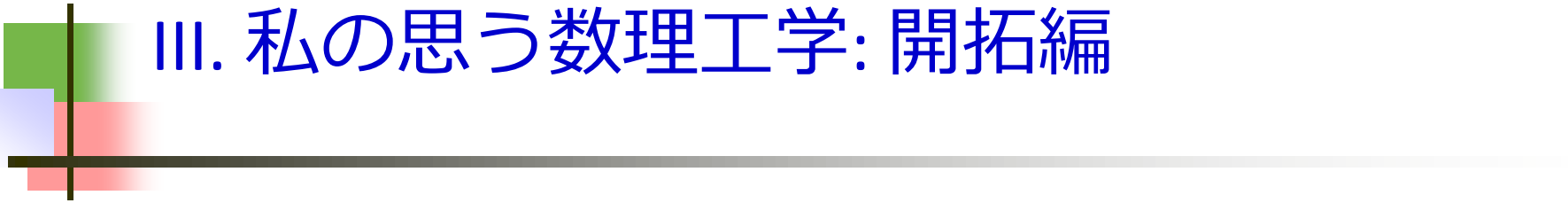
- 「論文」は「ポイント(=原理)」が明確な方が評価されやすい.
  - マルウェア分類はまだ論文に近づいて行った方.
- **一言で解決した話も多い** (暗黙知).
  - **本当は，こちらが理想なのでは？**
- 実践は実際に使われることで貢献する方が健全なのでは?
  - 成果発表: 特許，ハード/ソフト開発，起業 etc.
  - 評価指標: DL数，売上高

実践編で私が大事に思うこと.

- サイズ感覚.
  - データサイズ，要望応答時間，精度(粒度).

    ⇒ **使うべき技術が変わる**.
- 易しいアルゴリズム.
  - 根拠は深く．アルゴリズム解析自体は難しくてよい.

数理工学の出番



III. 私の思う数理工学: 開拓編

拓く数理工学

私の思う数理工学: 開拓編

〇〇をどうにかしたい。
だけど、何をすればよいのかわからない。

私の数理工学: 理論編 (研究宣伝)

- 完璧サンプリング法
  - ✓ 分割表 [RSA 2006]
  - ✓ 待ち行列ネットワーク [SICOMP 2008, AOR 2008]
- メディアン安定マッチング [MathOR 2012]
- 確率と計算

[20] S. Kijima, Nobutaka Shimizu, Takeharu Shiraga, How many vertices does a random walk miss in a network with moderately increasing the number of vertices?, Proc. SODA 2021, 106--122,

- 対数優モジュラ分布
  - ✓ 劣モジュラ関数の変数変換 [**SIDMA 2020**]
  - ✓ 任意の有限分配束は$M^\natural$凹関数の最小化元族に同型 [**OR lett. 2021**]
- **動的環境下のアルゴリズム論**
  - ✓ **増えるクーポン収集** [**SODA 2021**]

1. 学術的背景 (未発達のシーズ)

動的ネットワーク上のアルゴリズム解析

起点: **「現実のネットワークは刻々と変化する」**

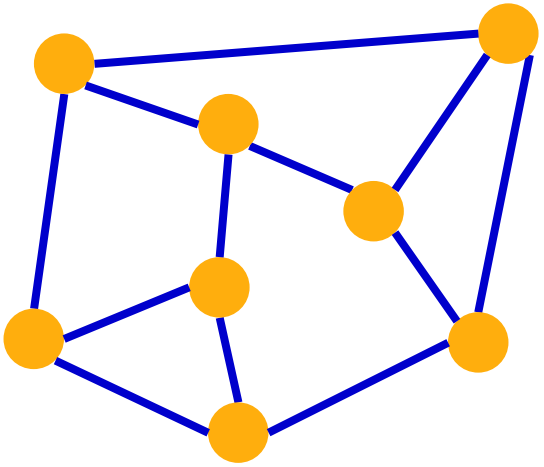
- ネットワーク工学で特に注目
  - [Kuhn & Oshman SIGACT News 2011]
  - [Augustine et al. SIGACT News 2016]
  - [Michail & Spirakis CACM 2018]

「しかし、理論的なことはほとんどわかっていない。」

- 海外の大型プロジェクト
  - “NETWORKS” (蘭 2014-2023)
  - “DYNAMIC MARCH” (ERC Starting Grant 2016-2021)

⇨ 今日は動的グラフ上のRWの話を中心に

ランダムウォークとアルゴリズム



☐ RWとアルゴリズム

- ✓ ネットワーク探索

  cf. 深さ優先探索，幅優先探索，

  universal traversal sequence

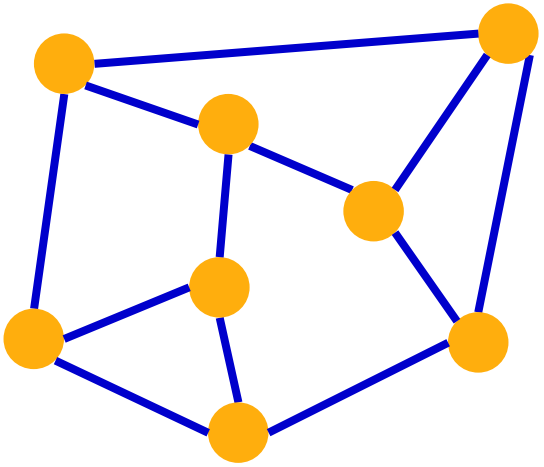
☐ Cover time

- ✓ 全探索にかかる(期待)時間
  - ➢ [Aleliunas et al. **1979**]
    - ➢ **単純RW** $t_{\mathrm{cov}} = \mathrm{O}(mn)$ $\Omega(n^3)$: lollipop
  - ➢ [Ikeda et al. **2003**, Nonaka et al. 2010]
    - ➢ "$\boldsymbol{\beta}$**-RW**", Metropolis walk $t_{\mathrm{cov}} = \mathrm{O}(n^2 \log n)$ $\Omega(n^2 \log n)$: spider
  - ➢ [David & Feige **2017**]
    - ➢ **local-min-degree RW** $t_{\mathrm{cov}} = \mathrm{O}(n^2)$

⇒ グラフが時間変化するとき、全探索にかかる時間は？

動的グラフ上のランダムウォーク（2010—2015頃の来嶋）

木場孝輔，山内由紀子, 来嶋秀治, 山下雅史, 動的グラフ上のランダムウォークの到達時間と全訪問時間, 数理解析研講究録, 1849, 100—104, 2013.



- ✓ 解析が難しい。
  想像はつくけど、面倒そう。
  - ➢ [L. Saloff-Coste and J. Zuniga 2009, 2011]
- ✓ **頂点集合が変化する場合はどうなんだろう？**
  - ➢ [C. Cooper and A. Frieze 2003]

    c stepごとに1頂点増える => $\frac{c}{c+1}n$ をcover

■ 動的グラフ上のRWの黎明期

- ➢ [Avin, et al. 2008], [Clementi et al. 2008], [Baumann et al. 2009] etc.
- ➢ Workshop "Random Walks on Random Graphs and Applications," Eindhoven, The Netherlands, April 14-16, **2015**.
  - ✓ （当時のマイブームは**決定性ランダムウォーク**）

動的ネットワーク上のアルゴリズム解析 --- なぜ難しい？

- 解析には**高い技術**が必要.
  - 静的ネットワーク上のアルゴリズム解析より“難しい”.
- 「**何を**」解析すれば良い？: **数理モデリング**
  - モデルができる前には「何もない」
  - モデルができてみれば「当たり前」

数理工学

解析可能な**数理モデル**

増えるクーポン収集問題

*来嶋秀治 (九大), 清水伸高(東工大)，白髪丈晴 (東工大)

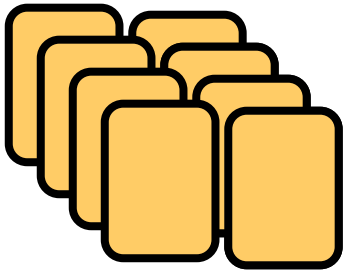
*本研究はJSPS科研費 JP17K19982の助成を受けたものです。

クーポン収集

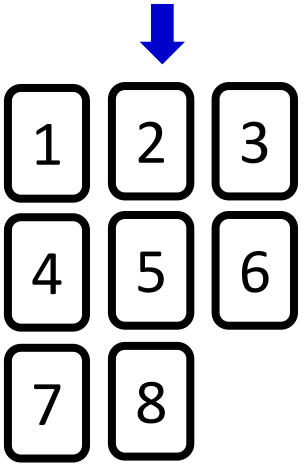
- ビックリマンシール
- プロ野球チップス
- ガチャ

- P●KEM●Nカード100種類．コンプまでに何枚引けば良い？

- n種類のカードが裏返しにおいてある．1枚選んでめくり，種類を確認したら元に戻して，よく混ぜる．全ての種類を確認するまでに何回めくればよいか？



- 年賀状のお年玉くじ下2桁．00から99まで
  そろえるには年賀状を何枚もらえばよいか？



- 一万円札を千円札10枚に両替する．千円札の
  記番号下2桁を00から99までそろえるには，
  いくら両替すればよいか？

クーポン収集

- ✓ $X_i \in \{1, \dots, n\}\ (i = 1,2, \dots)$は独立とし，$\Pr[X_i = k] = \frac{1}{n}(k = 1, \dots, n)$.
- ✓ $T = \min\{t | \{X_1, \dots, X_t\} = \{1, \dots, n\}\}$をコンプ回数(completion time)と呼ぶ.

Q. $\mathrm{E}[T]$を求めよ.

クーポン収集

- ✓ $X_i \in \{1, \dots, n\}$ $(i = 1,2, \dots)$は独立とし，$\Pr[X_i = k] = \frac{1}{n} (k = 1, \dots, n)$.
- ✓ $T = \min\{t | \{X_1, \dots, X_t\} = \{1, \dots, n\}\}$をコンプ回数(completion time)と呼ぶ.

Q. $\mathrm{E}[T]$を求めよ.

$T_k := \min\{t \mid |\{X_1, \dots, X_t\}| = k\}$:$k$種コンプ回数

$S_k := T_k - T_{k-1}$:$k-1$種コンプしてから新種が出るまでの回数

Claim. $\mathrm{E}[S_k] = \frac{n}{n-k+1}$.

- ✓ $k-1$種コンプした状態で，新種の出る確率 $p_k = \frac{n-(k-1)}{n}$.
- ✓ 従って$\mathrm{E}[S_k] = \frac{1}{p_k} = \frac{n}{n-k+1}$ (幾何分布の期待値).

$T = T_n = \sum_{k=1}^{n} S_k$に注意して，

$$\mathrm{E}[T] = \mathrm{E}[T_n] = \sum_{k=1}^{n} \mathrm{E}[S_k] = \sum_{k=1}^{n} \frac{n}{n-k+1} = n \sum_{k'=1}^{n} \frac{1}{k'} \simeq n(\ln n + 0.577)$$

≃519 ($n = 100$の時)

クーポン収集

- ✓ $X_i \in \{1, \dots, n\}$ $(i = 1,2, \dots)$は独立とし，$\Pr[X_i = k] = \frac{1}{n} (k = 1, \dots, n)$.
- ✓ $T = \min\{t | \{X_1, \dots, X_t\} = \{1, \dots, n\}\}$をコンプ回数(completion time)と呼ぶ.

Q. $\mathrm{E}[T]$を求めよ.

Q. $\Pr[T \geq 2n \ln n]$はどれくらいか？

クーポン収集

- ✓ $X_i \in \{1, \dots, n\}\ (i = 1,2, \dots)$は独立とし，$\Pr[X_i = k] = \frac{1}{n}(k = 1, \dots, n)$.
- ✓ $T = \min\{t | \{X_1, \dots, X_t\} = \{1, \dots, n\}\}$をコンプ回数(completion time)と呼ぶ.

Q. $\mathrm{E}[T]$を求めよ.

Q. $\Pr[T \geq 2n \ln n]$はどれくらいか？

Chevyshevの不等式を使うと

$$\Pr\left[X \geq 2\mathrm{E}[X]\right] \leq \frac{\mathrm{Var}[X]}{\mathrm{E}[X]^2} \leq \frac{\frac{n^2\pi^2}{6}}{(n \ln n)^2} = \frac{\pi^2}{6(\ln n)^2}$$

$\simeq 0.078\ (n = 100$の時)

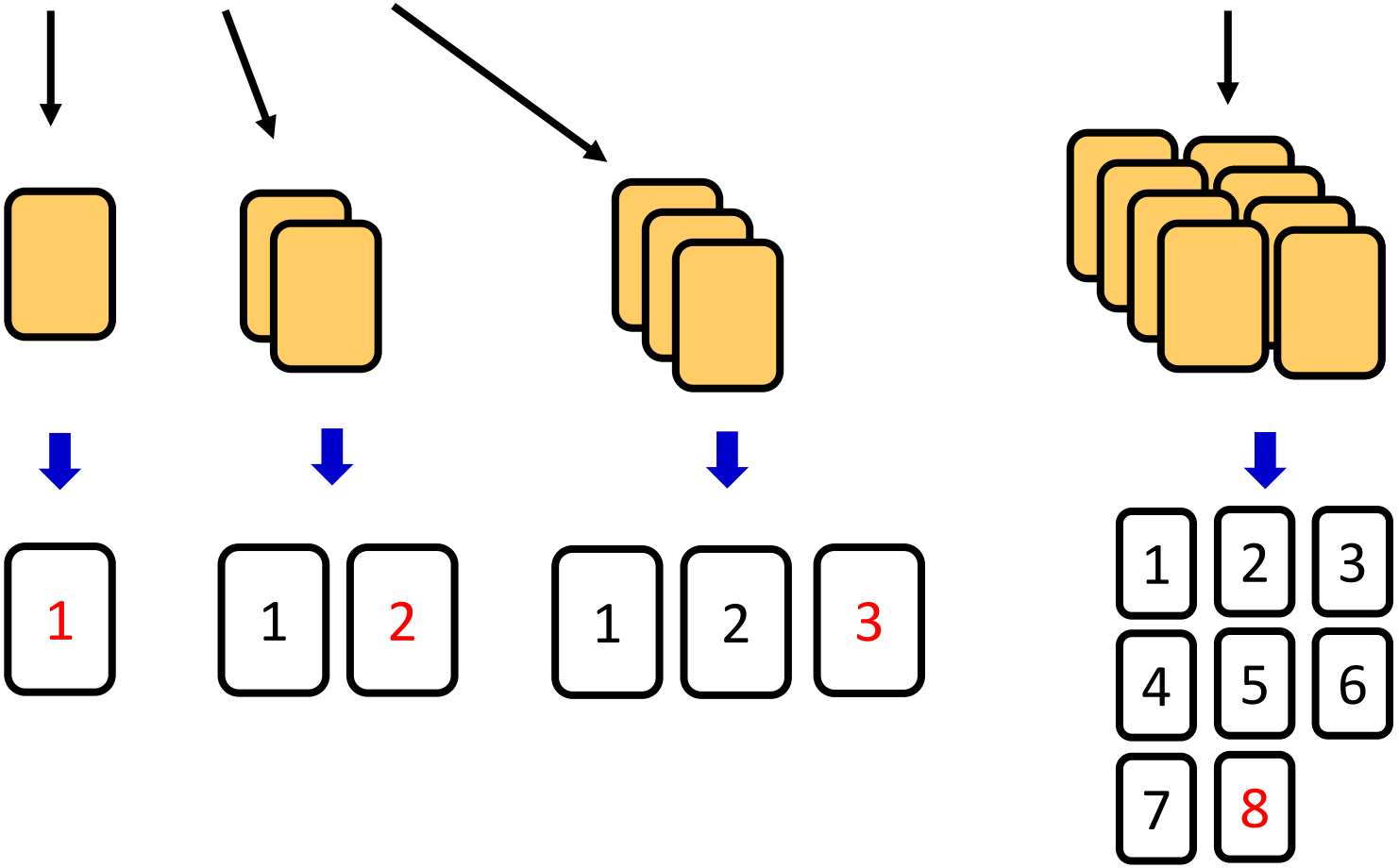
増えるクーポン収集

毎日1回ガチャを引く.

- 毎日1種ずつ増えるとき，コンプ回数は？

増えるクーポン収集

Day (i)	1日目	2日目	3日目	4日目	5日目	6日目	7日目	8日目	9日目
種類	1	2	3	4	5	6	7	8	9
$\Pr[X_i = k]$	$\frac{1}{1}$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{7}$	$\frac{1}{8}$	$\frac{1}{9}$



増えるクーポン収集

毎日1回ガチャを引く．

- 毎日1種ずつ増えるとき，コンプ回数は？
  - コンプ不可能

増えるクーポン収集

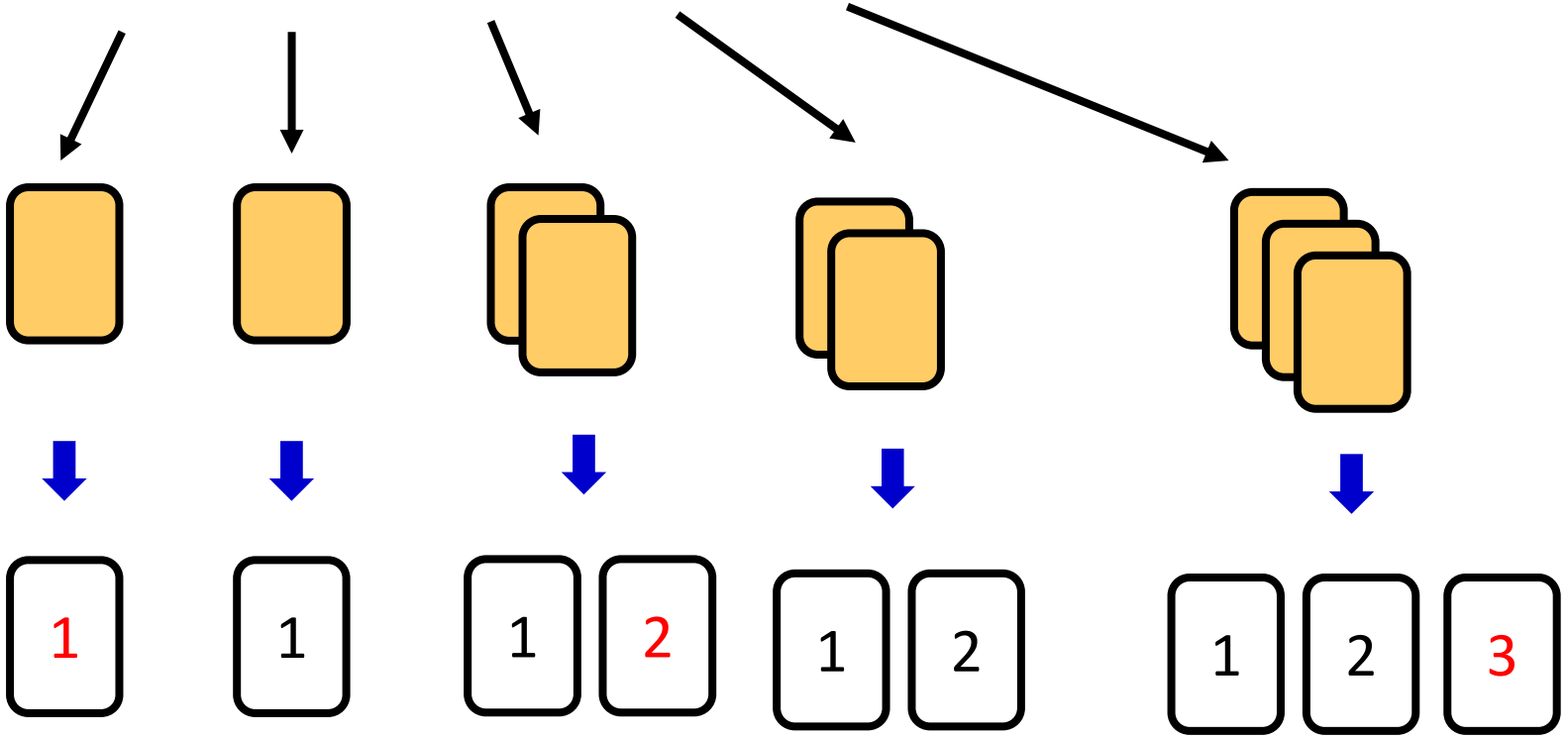
毎日1回ガチャを引く．

- 毎日1種ずつ増えるとき，コンプ回数は？
  - コンプ不可能

- 2日に1回新種リリースされたら，コンプ回数は？

増えるクーポン収集

Day	1日目	2日目	3日目	4日目	5日目	6日目	7日目	8日目	9日目
種類	1	1	2	2	3	3	4	4	5
$\Pr[X_i = k]$	$\frac{1}{1}$	$\frac{1}{1}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{5}$



増えるクーポン収集

毎日1回ガチャを引く．

- 毎日1種ずつ増えるとき，コンプ回数は？
  - コンプ不可能

- 2日に1回新種リリースされたら，コンプ回数は？
  - コンプ不可能

増えるクーポン収集

毎日1回ガチャを引く.

- 毎日1種ずつ増えるとき，コンプ回数は？
  - コンプ不可能
- 2日に1回新種リリースされたら，コンプ回数は？
  - コンプ不可能
- 10日に1回リリースされたら？
  - コンプ不可能(種類数1000の時を考えてみよ)

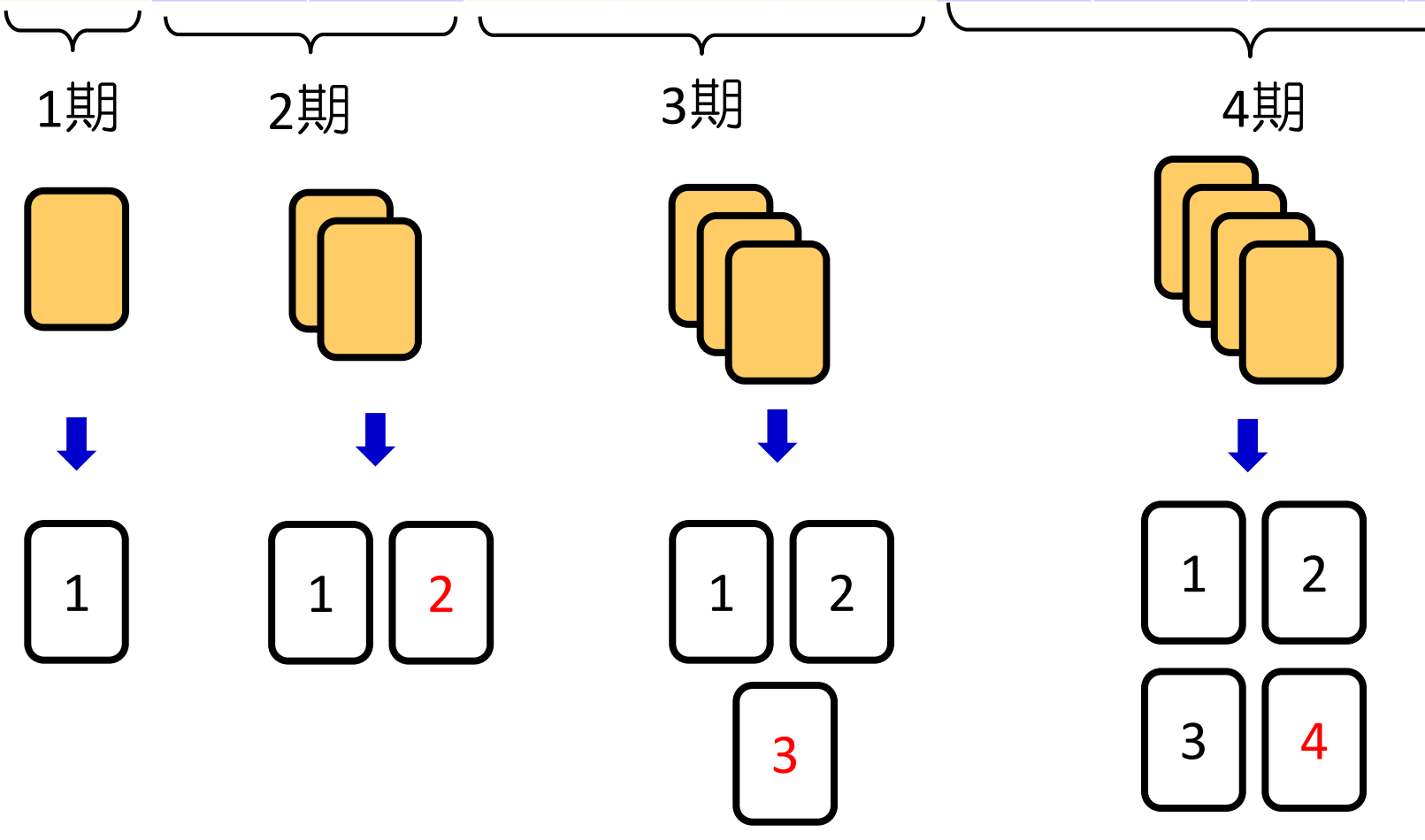
ゆっくりと増えるクーポン収集

毎日1回ガチャを引く.

- n種類目がリリースされてn日後に新種がリリース.
  - Q. コンプ回数は？

ゆっくりと増えるクーポン収集

Day	1日目	2日目	3日目	4日目	5日目	6日目	7日目	8日目	9日目
種類	1	2	2	3	3	3	4	4	4
$\Pr[X_i = k]$	$\frac{1}{1}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$



ゆっくりと増えるクーポン収集

毎日1回ガチャを引く.

- n種類目がリリースされてn日後に新種がリリース.
  - Q. コンプ回数は？
    - ✓ A. コンプ不可能(新種は出続けるから)

ゆっくりと増えるクーポン収集

毎日1回ガチャを引く.

- n種類目がリリースされてn日後に新種がリリース.
  - Q. コンプ回数は？
    - ✓ A. コンプ不可能(新種は出続けるから)

  - **Q. n期末にはどのくらいコンプされる？**

    1. $\mathrm{o}(n)$
    2. cn
    3. $n - c\sqrt{n}$
    4. $n - c\log n$
    5. $n - c$

cは適当な定数

ゆっくりと増えるクーポン収集

毎日1回ガチャを引く.

- n種類目がリリースされてn日後に新種がリリース.
  - Q. コンプ回数は？
    - ✓ A. コンプ不可能(新種は出続けるから)

  - **Q. n期末にはどのくらいコンプされる？**

    1. $\mathrm{o}(n)$
    2. cn
    3. $n - c\sqrt{n}$
    4. $n - c\log n$
    5. **$n - c$ (正解)**

cは適当な定数

ゆっくりと増えるクーポン収集

毎日1回ガチャを引く(一様ランダム)

$\mathfrak{d}(n)$: n期の日数

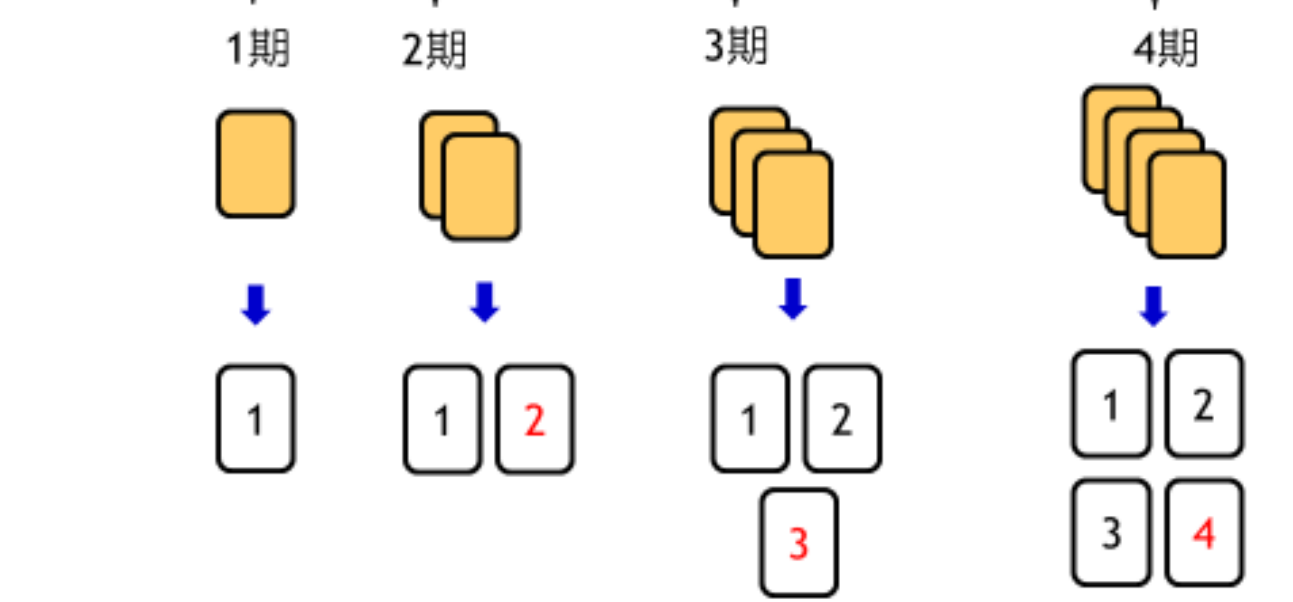
U_n: n期末の未収集アイテムの種類数

命題

$\mathfrak{d}(n) = n$のとき，$\mathrm{E}[U_n] < 1$.

ゆっくりと増えるクーポン収集

Day	1日目	2日目	3日目	4日目	5日目	6日目	7日目	8日目	9日目
種類	1	2	2	3	3	3	4	4	4
$\Pr[X_i = k]$	$\frac{1}{1}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$



ゆっくりと増えるクーポン収集

毎日1回ガチャを引く(一様ランダム)

$\mathfrak{d}(n)$: n期の日数

U_n: n期末の未収集アイテムの種類数

命題

$\mathfrak{d}(n) = n$のとき，$\mathrm{E}[U_n] < 1.$

証明

✓ $\mathcal{E}_{i,n} \coloneqq \begin{cases} 1 & (n\text{期末にアイテム } i \text{ が未収集}) \\ 0 & (n\text{期末にアイテム } i \text{ が収集済み}) \end{cases}$ $\quad (i = 1,2,\dots,n)$

✓ $U_n = \sum_{i=1}^{n} \mathcal{E}_{i,n}$

✓ アイテムnがn期末に未収集の確率

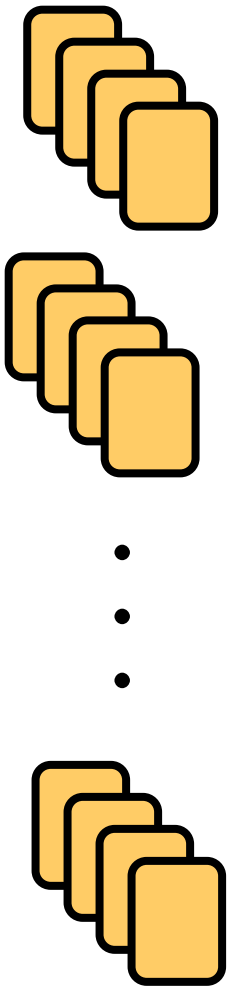
$$\Pr[\mathcal{E}_{n,n} = 1] = \left(1 - \frac{1}{n}\right)^n < \mathrm{e}^{-1}$$

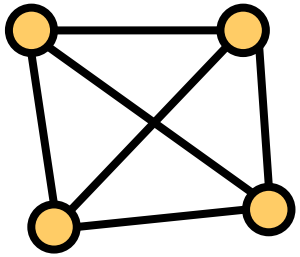
✓ アイテム i $(i \le n)$が n期末に未収集の確率

$$\Pr\left[\mathcal{E}_{i,n} = 1\right] = \left(1 - \frac{1}{i}\right)^i \left(1 - \frac{1}{i+1}\right)^{i+1} \dots \left(1 - \frac{1}{n}\right)^n < \left(\frac{1}{\mathrm{e}}\right)^{n+1-i}$$

✓ $\mathrm{E}[U_n] = \sum_{i=1}^{n} \Pr\left[\mathcal{E}_{i,n}\right] < \sum_{i=1}^{n} \left(\frac{1}{\mathrm{e}}\right)^{n+1-i} = \frac{1}{\mathrm{e}} + \frac{1}{\mathrm{e}^2} + \cdots + \frac{1}{\mathrm{e}^n} < \frac{\frac{1}{\mathrm{e}}}{1-\frac{1}{\mathrm{e}}} = \frac{1}{\mathrm{e}-1} < 0.582$

クーポン収集からランダムウォークへ





クーポンの**completion time** = 完全グラフ上のRWの**cover time**

頂点の増えるグラフ上のランダムウォーク [SODA 2021]

RWoGG: $\left(\mathfrak{d}, \left(G^{(i)}\right)_{i=1}^{\infty}, \left(P^{(i)}\right)_{i=1}^{\infty}\right)$

定理(一般上界)

$\mathfrak{d}(i) \geq c t_{\mathrm{hit}}(i)$ $(c \geq 1)$のとき，$\mathrm{E}[U] = O(1)$.

さらに$\frac{\mathfrak{d}(i)}{t_{hit}(i)} \overset{i\to\infty}{\longrightarrow} \infty$なら，$\mathrm{E}[U_n] \overset{n\to\infty}{\longrightarrow} 0$.

定理(lazy reversible)

$P^{(i)}$はlazy reversibleとする.

$\frac{t_{\mathrm{hit}}(i)}{t_{\mathrm{mix}}(i)} \geq \frac{i^{\gamma}}{c}$かつ$\mathfrak{d}(i) \geq \frac{3 c t_{\mathrm{hit}}(i)}{i^{\gamma}}$ $(c > 0)$のとき，$\mathrm{E}[U_n] \leq \frac{8 n^{\gamma}}{c} + 32$.

S. Kijima, N. Shimizu, T. Shiraga, How many vertices does a random walk miss in a network with moderately increasing the number of vertices?, in Proc. SODA 2021, 106—122.

増えるクーポン収集問題まとめ

参入お待ちしてます.

- ✓ **増えるクーポン収集問題**
- ✓ 頂点がゆっくり増加するグラフ上のRWの非被覆具合
  - ➢ 期間$\mathfrak{d}$と未訪問数$\mathrm{E}[U_n]$の関係

今後の課題

- ✓ $\mathbf{E}[U_n] = \mathbf{O}(\mathbf{1})$**の必要条件:** $\mathfrak{d}(i) =^{?} \Omega\left(t_{\mathrm{hit}}(i)\right)$
- ✓ “Cover time”以外の指標
- ✓ ランダムウォークからアルゴリズムへ

□「動的環境下のアルゴリズム解析」

- ➢ とは？ “入力”が静的でない???

計算機科学では「問題」は「入力」と「質問」からなる:

例: 素数判定問題

入力: $z \in \mathbb{N}$

質問: zは素数か?

まとめ

私の数理工学

I. 私の数理工学: 理論編
   - 良い研究をして、良い論文を書く

II. 私の数理工学: 実践編
   - 使ってこその数理工学

III. 私の数理工学: 開拓編
   - 数理で工学を拓く

これからも精進いたします．ご鞭撻のほど，お願いいたします．
理論，実践，開拓，いずれでも，ご興味のある方，
お声がけください．ぜひ，一緒に研究しましょう．

Fin.