

## 科学研究費助成事業 研究成果報告書

平成 26 年 6 月 16 日現在

機関番号：32629

研究種目：若手研究(B)

研究期間：2011～2013

課題番号：23700025

研究課題名(和文) サンプルングアルゴリズムの新提案

研究課題名(英文) Proposing new sampling algorithms

研究代表者

山本 真基 (Masaki, Yamamoto)

成蹊大学・理工学部・准教授

研究者番号：50432414

交付決定額(研究期間全体)：(直接経費) 1,600,000円、(間接経費) 480,000円

研究成果の概要(和文)：本研究で行えたことは主に次の二つである．1: グラフのパスおよびサイクルの数え上げが #P 完全(計算困難な問題)である．2: グラフのパスおよびサイクルの数え上げが(RP ≠ NPのもと)近似不可能(近似計算困難な問題)である．パスやサイクルはグラフ構造として木に近い性質を持っており、それゆえ、この結果は Tutte 多項式の近似計算可能性・不可能性において、未解決となっている領域である木の数え上げの近似可能性・不可能性に向けて何かしらの知見を与えていると期待できる．

研究成果の概要(英文)：There are two main outcomes in this research project. We prove that, 1: counting paths and cycles are #P-hard (i.e., it is hard to exactly compute the numbers.), 2: they are inapproximable (under RP is not NP). Since the structure of a path and a cycle is close to that of a tree, we hope that our results give some insights into approximately counting trees which is a subproblem of approximately computing Tutte polynomials.

研究分野：情報学

科研費の分科・細目：情報学基礎

キーワード：サンプルングアルゴリズム マルコフ連鎖モンテカルロ法 近似数え上げ

## 1. 研究開始当初の背景

### 【学術的背景】

マルコフ連鎖モンテカルロ法( Markov chain Monte Carlo methods, 以降, MCMC 法と略す)とは, 所望の確率分布を定常分布として持つようなマルコフ連鎖を用いたサンプリングアルゴリズムのことである。(より正確には, そのようなサンプリングアルゴリズムを用いて, 所望の値を求める計算手法のことである.)

MCMC 法は, 1950年代に統計物理(Statistical Mechanics)の分野で導入され, 1990年代に入ってから, 情報科学, 計算機科学の分野でも盛んに利用されるようになった。理論計算機科学(Theoretical Computer Science)においては, MCMC 法は, 問題の解の上の一様分布を生成するためのサンプリングアルゴリズムとして利用され, 解の個数を近似するアルゴリズムに利用されている。(理論計算機科学における MCMC 法については文献 [1,2] を参照。)例えば, 次のような, 基本的かつ重要な組合せ問題の解の個数の近似を求めることに利用されている:

1. 頂点彩色問題
2. 頂点独立集合問題
3. マッチング問題

更に, このような, グラフの数え上げ問題以外にも, 数学・物理の分野で重要な関数の値の近似を求めることにも利用されている:

4. 行列のパーマネント(permanent)
5. 分配関数(partition function)
6. Tutte 多項式(Tutte polynomial)

理論計算機科学における MCMC 法の研究の中心は, (1) サンプリングアルゴリズムを考案し, (2) そのアルゴリズムの効率のよさ, つまり, 混合時間(mixing time)を算定する, ことの2つである。冒頭の概要に掲げた2つの項目は, これらの2つに対応している。サンプリングアルゴリズムは, メトロポリス・ヘイスティングスアルゴリズム(Metropolis-Hastings algorithm)やギブスサンプリング(Gibbs sampling)といったランダムウォークを用いた方法が代表的である。また, 近年では, 焼きなまし(simulated annealing)法という, より進んだ方法の研究も盛んである。混合時間を算定するための理論的解析には, カップリング(coupling)法, それを応用したパス・カップリング(path coupling)法, コンダクタンス(conductance)を利用したより強力な方法がある。以降で挙げられるように, 理論計算機科学における MCMC 法には重要な未解決問題がたくさん残されており, それを解決するためのサンプリングアルゴリズム, 及び, 混合時間を算定するための解析手法の開発が要求されている。

## 2. 研究の目的

### 【本研究で解明すること】

以下に, 理論計算機科学の MCMC 法における, 2つの代表的な未解決問題を挙げる。本研究では, これらの解決に向けて, 新たなサンプリングアルゴリズムを提案する。それと同時に, そのようなサンプリングアルゴリズムの混合時間を算定するための新たな理論的解析手法を提案する。

**研究課題1:** 統計物理の分野には, イジング模型(Ising model)やポッツ模型(Potts model)のように, 粒子の動きを記述する確率モデルがいくつか提案されている。そのような確率モデルの正規化定数は「分配関数(partition function)」と呼ばれ, 熱平衡状態の粒子の統計的性質を表す重要な値となっている。イジング模型やポッツ模型の分配関数は, 「Tutte 多項式」で表されることが知られている。理論計算機科学では, Tutte 多項式の計算困難性が, 更に, 最近では, 近似可能性・不可能性が議論されるようになった。(文献[3]を参照。)特に, Tutte 多項式の(一部の)近似可能性は, イジング模型の分配関数が近似可能であるという事実から導かれている。Tutte 多項式の「計算困難性」については完全に解決された研究テーマであるが, 近似可能性・不可能性については, まだ完全には解決していない。(文献 [3] を参照。)

Tutte 多項式の近似可能性・不可能性において, 未解決となっている部分の解決を図る。

**研究課題2:** MCMC 法は, 問題の解の上の一様分布を生成するためのサンプリングアルゴリズムとして利用され, 解の個数を近似するアルゴリズムに利用されている。これまで, 様々な問題(#P-困難な問題)に対して, (多項式時間)近似可能性・不可能性が議論されてきた。その一方で, まだそのどちらとも証明されていない問題がいくつかある。半順序集合(poset)のイデアル(ideal)の個数を求める問題が代表的である。この問題は, #2-SAT 問題の部分問題(#1P1N-2CNF 問題), 更には, 2部グラフの頂点独立集合の数え上げ問題と(近似可能性に関して)等価であるということも知られている。(文献[4]を参照。)

半順序集合のイデアルの数え上げ問題が近似可能かどうかを証明する。

### 【学術的意義】

上で述べた2つの研究課題は, 理論計算機科学の MCMC 法において, 代表的な未解決問題となっている。これらの解決には, 新しいアルゴリズムを考案することが, 更に, これまでの混合時間の算定方法にはない, 新たな

理論的解析手法を開発する必要があることが予想される。これらの新たなアルゴリズムと解析を提案することで、理論計算機科学における MCMC 法が深化し、当該分野が発展することが大いに期待できる。また、MCMC 法が様々な分野で応用されていることから、本研究で提案されたサンプリングアルゴリズムが多く分野を発展させることが期待される。

[1] M. Jerrum, “Mathematical foundations of the markov chain monte carlo method”, in Probabilistic Methods for Algorithmic Discrete Mathematics, Habib et al., Ed. 1998, pp. 116-165, Springer-Verlag, Germany.

[2] R. Montenegro and P. Tetali, “Mathematical Aspects of Mixing Times in Markov Chains”, Foundations and Trends in Theoretical Computer Science, vol. 1 (3), pp. 237-354, 2006.

[3] L. A. Goldberg and M. Jerrum, “Inapproximability of the Tutte polynomial”, Information and Computation 206, pp. 908-929, 2008.

[4] M. Dyer, L. A. Goldberg, M. Jerrum, The relative complexity of approximate counting problems, Algorithmica 38, pp. 471-500, 2004.

### 3. 研究の方法

理論研究を効率よく進めるには、専門家と直に会って密に議論することが、有効な方法であると感じている。本研究では、来嶋秀治(九州大学)、松井泰子(東海大学)、渡辺治(東京工業大学)、岩間一雄(京都大学)、西関隆夫(関西学院大学)らと密に議論することを考えている。特に、来嶋氏、松井先生とは、これまでに MCMC 法について様々な議論をしてきている。その中で、文献[1]は、理論計算機科学では主要な国際会議(COCOON2009)に採択された。更に、その国際会議の selected paper として雑誌への掲載が決まっており、国際的にみて一定の評価を頂いている。

### 4. 研究成果

#### 【23年度研究成果】

この年度の具体的な研究目標は、Tutte 多項式と呼ばれる関数の近似可能性・不可能性を解決することであった。Tutte 平面上で既に近似可能・不可能が証明された領域は少ししかない。少しでも多くの領域について近似可能性・不可能性を解明することが、この年度の研究計画であった。この研究計画に関連した研究実績は大きく3つ挙げられる: 1. パス、サイクルの近似数え上げが不可能であることを示した、2. 木のランダムサンプリ

ングが通常のマルコフ連鎖では不可能であることを示した、3. 3-SAT 問題を解く最速のアルゴリズムを提案した。このうち、最初の2つが当研究課題に直接に関係するものである。Tutte 平面に、森の数え上げ問題に相当する領域がある。(この部分の近似可能性・不可能性を示すことは研究課題の一部にもなる。)上の2つの不可能性及びその証明は、森の数え上げ問題について何らかの手がかりを与えるものであると思っている。特に、2つ目の木のランダムサンプリングがありきたりのマルコフ連鎖からは近似数え上げが不可能であることは、同じく、森のランダムサンプリングについても同様の状況である証拠を与えている。3つ目の研究成果は、直接的な関連性は少ないにしても、アルゴリズムの解析手法が当研究課題を解明するのに役立つのではないかと考えている。それは、解析した 3-SAT 問題を解くアルゴリズムが、ランダムウォークタイプのアルゴリズムであり、マルコフ連鎖の一種であるからである。

#### 【24年度研究成果】

この年度の具体的な研究目標は、半順序集合のイデアルの数え上げ問題の近似可能性・不可能性を解決することであった。この問題は、#2-SAT 問題の部分問題(#1P1N-2CNF 問題)、更には、2部グラフの頂点独立集合の数え上げ問題と(近似可能性に関して)等価な問題であるということが知られている。また、前回の実施状況報告書から、Tutte 多項式に関して、木の数え上げ問題及び森の数え上げ問題の近似可能性・不可能性を解決することも今年度の課題の一つであった。これらの研究課題に関連した研究実績は、木の数え上げ問題において、次数が高々3であるようなグラフに限った場合に、木のランダムサンプリングが可能であることから、木の数え上げ問題が近似可能であることを示したことである。昨年度の研究結果を合わせると次のようなことが分かった。グラフの基本的な構造に、パス、サイクル、木の三つがある。この三つに対して、1. パスの数え上げ及びサイクルの数え上げは近似不可能である。2. 次数が高々3であるグラフに限れば、木の数え上げは近似可能である。また、半順序集合のイデアルの数え上げ問題と関連のある、グラフの頂点独立集合の数え上げ問題が近似不可能であるという論文が発表され、その論文の調査を行った。これは、2部グラフの頂点独立集合の数え上げ問題、つまり、半順序集合のイデアルの数え上げ問題の近似可能性・不可能性の解明に大きく関係する事実である。

#### 【25年度研究成果】

具体的な研究課題として以下の二つをあげていた。1: Tutte 多項式の近似可能性・不可能性において、未解決となっている領域の解決を図る。2: 半順序集合のイデアルの数え上げ問題が近似可能かどうかを証明する。

これに対して実際の研究では以下のことができた。1: パスおよびサイクルの数が  $\#P$  完全である。2: パスおよびサイクルの数が  $(RP, NP)$  のもと) 近似不可能である。この結果は上の一つ目の研究課題に関連するものである。Tutte 多項式の近似可能性・不可能性において未解決となっている領域に「木」の数の上げに相当するものがある。パスやサイクルはグラフ構造として木に近い性質を持っており、そのため、そういったグラフの数の上げの近似不可能性を示したことは、木の数の上げの近似可能性・不可能性に向けて何かしらの知見を与えていると期待できる。

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[1] Masaki Yamamoto, Approximately counting paths and cycles in a graph, Inference, Computation, and Spin Glasses (ICSG2013), Sapporo, Japan (A satellite meeting of STATPHYS 25, Seoul, Korea) (2013)

[2] Kazuhisa Makino, Suguru Tamaki, Masaki Yamamoto: Derandomizing the HSSW Algorithm for 3-SAT. Algorithmica 67(2): 112-124 (2013)

[3] Masaki Yamamoto: A Combinatorial Analysis for the Critical Clause Tree. Theory Comput. Syst. 52(2): 271-284 (2013)

[4] Yuichi Yoshida, Masaki Yamamoto, Hiro Ito: Improved Constant-Time Approximation Algorithms for Maximum Matchings and Other Optimization Problems. SIAM J. Comput. 41(4): 1074-1093 (2012)

[5] Tomonori Ando, Yoshiyuki Kabashima, Hisanao Takahashi, Osamu Watanabe, Masaki Yamamoto: Spectral Analysis of Random Sparse Matrices. IEICE Transactions 94-A(6): 1247-1256 (2011)

[6] Tatsuya Akutsu, Avraham A. Melkman, Takeyuki Tamura, Masaki Yamamoto: Determining a Singleton Attractor of a Boolean Network with Nested Canalizing Functions. Journal of Computational Biology 18(10): 1275-1290 (2011)

[7] Masaki Yamamoto, Shuji Kijima, Yasuko Matsui: A polynomial-time perfect sampler for the Q-Ising with a vertex-independent noise. J. Comb. Optim. 22(3): 392-408 (2011)

[8] Kazuhisa Makino, Suguru Tamaki, Masaki Yamamoto: An exact algorithm for the Boolean connectivity problem for k-CNF. Theor. Comput. Sci. 412(35): 4613-4618 (2011)

[9] Kazuhisa Makino, Suguru Tamaki, Masaki Yamamoto: Derandomizing HSSW Algorithm for 3-SAT. COCOON 2011: 1-12

[10] Masaki Yamamoto: A tighter lower bound on the circuit size of the hardest Boolean functions. Electronic Colloquium on Computational Complexity (ECCC) 18: 86 (2011)

以上、合計 10 件の内訳:

[雑誌論文](計 7件)

[学会発表](計 3件)

#### 6. 研究組織

(1) 研究代表者

山本真基 (成蹊大学理工学部)

研究者番号: 50432414

(2) 研究分担者

( )

研究者番号:

(3) 連携研究者

( )

研究者番号: