

令和 6 年 6 月 14 日現在

機関番号：62603

研究種目：基盤研究(C)（一般）

研究期間：2020～2023

課題番号：20K03742

研究課題名（和文）代数的従属性をもつ離散確率構造のダイレクトサンプリング

研究課題名（英文）Direct sampling of discrete random structures with algebraic dependencies

研究代表者

間野 修平（Mano, Shuhei）

統計数理研究所・統計基盤数理研究系・教授

研究者番号：20372948

交付決定額（研究期間全体）：（直接経費） 2,400,000円

研究成果の概要（和文）：代表者は、GKZ超幾何系と呼ばれる線形偏微分方程式系が定める線形制約下の整数値ベクトルに値をとる確率分布の族について、計算代数を用いたサンプリング（所与の確率分布に独立に従う確率変数を生成すること）のアルゴリズムを考案した。本研究では、アルゴリズムの定式化を通して効率化と計算複雑性の評価を行い、性質の良いグラフが定めるクラスにおけるアルゴリズムの挙動を超幾何関数の和公式を通して議論する一方、アルゴリズムに動機づけられた超幾何関数を提案し、関連するいくつかの確率分布のサンプリングについても議論した。

研究成果の学術的意義や社会的意義

統計モデルに基づくデータ解析においては確率的な量を計算機で数値的に実現する必要がある。その計算は近似的に行われることも多いが、本研究ではモデルの数学的表現に基づく厳密なサンプリングを対象とした。実用上の意義だけでなく、超幾何関数論の新たな応用を見出し、新たな超幾何関数を提案したことにも意義があると考えている。

研究成果の概要（英文）：The investigator proposed an algorithm for sampling (generation of random variables independently follows a given distribution) from a class of probability distributions. In the class of distributions, a random variable takes its value on the integer-valued vector that obeys linear constraints determined by a system of linear partial differential equations called the GKZ-hypergeometric system. In this project, through a formulation of the algorithm, efficient implementation was discussed, and the computational complexity was evaluated. The algorithm's behavior for a class determined by a graph with some nice properties was discussed via a sum formula of hypergeometric functions. On the other hand, a class of hypergeometric functions motivated by the algorithm was proposed. Samplings of some related probability distributions were also discussed.

研究分野：統計科学・応用確率論

キーワード：サンプリング アルゴリズム 離散確率構造 計算代数 データ解析

1. 研究開始当初の背景

統計モデルに基づくデータ解析においては確率的な量を計算機で数値的に実現する必要がある。所与の確率分布に独立に従う確率変数（スカラーとは限らない）を計算機で生成することをサンプリングという。サンプリングは近似的に行われることも多いが、本研究は厳密なサンプリングを対象とした。

離散指数型分布族は整数値ベクトルに値をとる確率変数を扱う統計モデルの基本的で広いモデルであり、分割表やグラフィカルモデルと呼ばれるモデルを含む。離散指数型分布族のデータに基づく条件付き確率分布のサンプリングは統計学において成立当初からの課題である。ここで、条件は観察を特徴付ける量（統計学では十分統計量と呼ばれる）の間の従属性を表す線形制約として表される。

厳密なサンプリングではないが、マルコフ連鎖を所与の確率分布に収束させることに基づくサンプリングのアルゴリズムがあり、メトロポリス連鎖と呼ばれる（マルコフ連鎖モンテカルロと呼ばれることもある）。収束には既約な（任意の2つの状態の間の推移がある時間後に正の確率で起こること）マルコフ連鎖が必要である。Diaconis と Strumfels (1998)は、推移の基底としてマルコフ基底を定義し、マルコフ基底がマルコフ連鎖が既約になるための必要十分条件を与えた。それは、マルコフ基底に対応する二項式の有限集合が、上記の線形制約に現れる行列が定める多項式環のトーリックイデアルの生成系であることである。後者はグレブナー基底として得られることが保証される。この結果は代数統計学と呼ばれる分野が発展する契機になり、様々な統計モデルに関する研究が行われてきた。

しかし、厳密なサンプリングが可能であれば、もちろんそれが望ましい。メトロポリス連鎖について言えば、一般には収束の評価が難しく、独立なサンプリングでもない。ここで、離散指数型分布族の条件付き確率分布の規格化定数は、上記の線形制約を表す行列が定める GKZ 超幾何系と呼ばれる線形偏微分方程式系の解である。研究代表者は、GKZ 超幾何多項式間の漸化式を用いて推移確率を定めたマルコフ連鎖により厳密なサンプリングが可能であることを示し、そのアルゴリズムを 2017 年、2018 年に論文と書籍に発表した。この推移確率は GKZ 超幾何系が定める微分作用素環のイデアルのグレブナー基底により得られることが保証される。

2. 研究の目的

- (1) 上記のサンプリングのアルゴリズムの効率化を図る。
- (2) 離散指数型分布族以外の離散確率構造に展開する。

3. 研究の方法

- (1) アルゴリズムを一般的な形に定式化し、効率化し、計算複雑性を評価する。
- (2) 個々の離散確率構造について検討する。

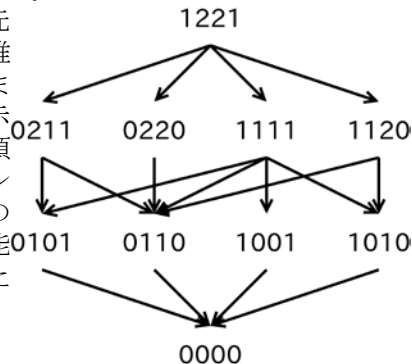
4. 研究成果

(1) アルゴリズムの定式化と効率化

各行の和が 1, 2、各列の和が 2, 1 の分割表を、空欄を非負の整数で埋めて作成することを考える（右の表）。可能な埋め方は左上から右下に (1,0,1,1) または (0,1,2,0) であるが、これを多項式で $x_1x_3x_4 + x_2x_3^2/2$ のように表す。これを各変数で微分すると 4通りの多項式 $x_3x_4, x_3^2/2, x_1x_4 + x_2x_3, x_1x_3$ が得られるが、それらを可能な分割表の集合とみなすと、行と列の和は 0211, 0220, 1111, 1120 である。これを繰り返すと、右図に示す有界束が得られる。

		1
		2
2	1	3

アルゴリズムは、束の各元を状態とし、最大元から最小元に推移するマルコフ連鎖として表せる。推移確率は可能な推移先の多項式の値に比例する（変数の値はモデルにより定まる）。ここで、多項式は GKZ 超幾何多項式であり、陽な表示がない場合でも、微分に関する漸化式を用いて推移先の多項式を計算することができる。この漸化式は必要な微分をグレブナー基底に関する標準形により表すことで得られる。この定式化により、アルゴリズムの効率化を議論することが可能になり、離散指数型分布族以外の離散確率構造に展開するための足掛かりを得た。



(2) 計算複雑性の評価

(1) で効率化したアルゴリズムの計算複雑性は、推移確率が陽に求まる場合は（例えば次項で述べる三角化されたグラフが定める GKZ 超幾何多項式はそうである）、行列のサイズ（行数と列数の積）と多項式の次数の積で抑えられる。そうでない場合は推移確率を計算するのに時間

がかり、行列の標準体積（行列が定める多面体の体積）が行列の列数より大きい場合は、標準体積の2乗と多項式の次数の積で抑えられる。

計算代数システム **Risa/Asir** 上に本アルゴリズムとメトロポリス連鎖を実装し、計算にかかる時間を測定した。メトロポリス連鎖が生成する確率変数には自己相関があるため、メトロポリス連鎖で一定の数の確率変数を生成し、自己相関に基づき定義される有効数を求めて、その数を本アルゴリズムで生成した。標準体積が小さい例としてポアソン回帰と呼ばれるモデルを実装したところ、メトロポリス連鎖の半分の時間で生成できた。標準体積が大きい例としては二変量の分割表を実装した。標準体積は行列のサイズで定まる組合せの数になり大きくなりがちである。推移確率が陽に与えられる独立モデルでは10分の1の時間で生成できる一方、そうでないモデルでは800倍ほどかかった。計算時間の点でメトロポリス連鎖を上回るの難しいことが分かった。

(3) アルゴリズムと多変数超幾何関数

① 三角化されたグラフが定める GKZ 超幾何多項式におけるガウスの和公式

三角化されたグラフ（長さ4以上の巡回路はそれを構成する頂点の間に辺をもつグラフ）が定めるモデル（統計学では分解可能グラフィカルモデルと呼ばれる）については、グラフが大域マルコフ性と呼ばれる性質を満たす場合、規格化定数は陽に求まることが知られている。これはグラフィカルモデルにおける目覚ましい結果の一つであるが、ガウスの超幾何関数における特殊値における和公式の一般化であることを指摘し、GKZ 超幾何関数のオイラー型積分表示とグラフの分解を用いて導出した。

② 古典一変数超幾何関数の一般化

古典一変数超幾何関数は ${}_nF_{n-1}$, $n = 2, 3, \dots$ と表記される超幾何関数のクラスであり、ガウスの超幾何関数は $n = 2$ の場合である。青本-ゲルファントの超幾何関数 ((m, n) 型超幾何関数と表記される) はガウスの超幾何関数の多変数への拡張とみなせるが、二変量の分割表の確率分布の規格化定数を与える。その類似として、多変量の分割表の確率分布の規格化定数を与えるような古典一変数超幾何関数 ${}_2F_{2l-1}$, $l = 3, 4, \dots$ の多変数への拡張を定義した。

以上の(1-3)は高山信毅氏と共同で行った研究の成果であり、学術雑誌に論文を投稿したが、本報告書執筆時点で査読中である（プレプリントは [arXiv:2110.14992](https://arxiv.org/abs/2110.14992)）。ただし、(1)の有界束については雑誌「数理科学」2024年3月号の特集において紹介した。

(4) 有限グラフの頂点を台とする確率分布

離散指数型分布族に含まれる確率分布に整数分割上の確率分布があり、単体上の確率分布に従う標本の確率分布とみなせる。ディリクレ分布は単体上の確率分布として基本的であるが、単体上の測度値拡散過程の定常分布である。この拡散過程を有限グラフ上の頂点を台とする確率分布に値をとる拡散過程とみなして一般化し、その定常分布がグラフの独立集合が定める面を台とする確率分布であること、線形な浮動項を加えて得られる定常分布が一意であることなどを示した。後者の定常分布は完全グラフの場合にディリクレ分布に帰着するが、一般に密度関数を陽に与えることはできない。しかし、双対な粒子系のマルコフ過程を用いて浮動項のパラメタに関する尤度を評価するアルゴリズムを構成し、グラフの独立集合を発見するアルゴリズムの待ち時間を評価した。これらの成果は学術雑誌 **Mathematics** の離散確率構造に関する特集に発表した。

(5) ランダム量子回路の出力

Google が 2019 年に発表した量子超越（何らかのタスクについて何らかの規準の下で古典計算が量子計算を優越すること）の実証は、ランダムに生成した量子回路の出力（2進数上に値をとる確率変数）を古典計算機で生成するタスクについて行われた（ここでの量子計算はゲート型と呼ばれるものを指す）。現状の量子計算機は非力なので量子計算機に有利な設定で行われ、量子回路の出力が従う確率分布関数の計算は計算複雑性の意味で古典計算機には難しいことに依拠する。ところが、このタスクは、出力の交換可能性（確率変数列の分布が順序の置換に関して不変であること）を仮定すれば、古典計算機で容易に実行できることを指摘した。国内研究集会と国際会議において発表し、論文を準備中である。

(6) その他

アルゴリズムに関しては、量子計算に基づくモンテカルロ法により GKZ 超幾何多項式を評価すること（国際会議において発表）など、離散確率構造に関しては、数え上げ組合せ構造に関するもの、有限群に関するものなどを検討したが、本報告書執筆時点で論文化するには至っていない。

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 1件）

1. 著者名 Shuhei Mano	4. 巻 2022
2. 論文標題 Measure-on-Graph-Valued Diffusion: A Particle System with Collisions and Its Applications	5. 発行年 2022年
3. 雑誌名 Mathematics, Special Issue Random Combinatorial Structures	6. 最初と最後の頁 21pp
掲載論文のDOI（デジタルオブジェクト識別子） 10.3390/math10214081	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 間野修平	4. 巻 729
2. 論文標題 グレブナー基底とサンプリング	5. 発行年 2024年
3. 雑誌名 数理科学 特集「グレブナー基底のすすめ」	6. 最初と最後の頁 31-38
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計11件（うち招待講演 4件/うち国際学会 7件）

1. 発表者名 間野修平
2. 発表標題 ランダム確率測度と量子超越の実証
3. 学会等名 研究集会「ベイズ統計学の最近の展開」（招待講演）
4. 発表年 2023年

1. 発表者名 間野修平
2. 発表標題 グラフ上の確率測度に値をとる拡散過程について
3. 学会等名 研究集会「無限粒子系，確率場の諸問題XVII」
4. 発表年 2023年

1. 発表者名 間野修平, 高山信毅
2. 発表標題 コーダグラフが定める行列が導くGKZ超幾何多項式の特特殊値における和公式
3. 学会等名 日本数学会2022年度秋季総合分科会
4. 発表年 2022年

1. 発表者名 Shuhei Mano
2. 発表標題 Algebraic algorithm for direct sampling from toric models
3. 学会等名 Algebraic Statistics 2022 (国際学会)
4. 発表年 2022年

1. 発表者名 間野修平
2. 発表標題 コーダグラフが定める行列が導くA超幾何多項式の特特殊値における和公式について
3. 学会等名 確率・統計・行列ワークショップ彦根2021
4. 発表年 2021年

1. 発表者名 Shuhei Mano
2. 発表標題 An algorithm for direct sampling from toric models with computational algebra
3. 学会等名 Bernoulli-IMS One world Symposium 2020 (国際学会)
4. 発表年 2020年

1. 発表者名 Shuhei Mano
2. 発表標題 Sampling from toric models and hypergeometric functions
3. 学会等名 Random Structures & Algorithms 2023 (国際学会)
4. 発表年 2023年

1. 発表者名 Shuhei Mano
2. 発表標題 Modeling exchangeable sequence by mixture of mixture and its application
3. 学会等名 6th International Conference on Econometrics and Statistics (招待講演) (国際学会)
4. 発表年 2023年

1. 発表者名 Shuhei Mano
2. 発表標題 Sampling problems and hypergeometric functions
3. 学会等名 10th International Congress on Industrial Mathematics (国際学会)
4. 発表年 2023年

1. 発表者名 Shuhei Mano
2. 発表標題 A measure-on-graph-valued diffusion: a particle system with collisions and its applications
3. 学会等名 6th Institute of Mathematical Statistics Asia-Pacific Regional Meeting (招待講演) (国際学会)
4. 発表年 2024年

1. 発表者名 Shuhei Mano
2. 発表標題 Sampling from probability distributions associated with toric variety via computational algebra
3. 学会等名 Monash Probability & Statistics Seminar (招待講演) (国際学会)
4. 発表年 2024年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

https://www.ism.ac.jp/~smano/ Algebraic algorithm for direct sampling from toric models, with Nobuki Takayama. arXiv: 2110.14922

6. 研究組織		
氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関		
オーストラリア	Monash University		