

科学研究費助成事業 研究成果報告書

令和 5 年 6 月 10 日現在

機関番号：34315

研究種目：若手研究

研究期間：2018～2022

課題番号：18K18016

研究課題名（和文）計算統計のための擬似乱数と準モンテカルロ法の研究

研究課題名（英文）Research on pseudorandom number generation and quasi-Monte Carlo methods for computational statistics

研究代表者

原瀬 晋（Harase, Shin）

立命館大学・理工学部・講師

研究者番号：80610576

交付決定額（研究期間全体）：（直接経費） 1,600,000円

研究成果の概要（和文）：マルコフ連鎖モンテカルロ(MCMC)法のための準乱数を開発した。短い周期のTausworthe発生法の一周期分の出力を準乱数と見なし、一様性の指標であるt-値の観点からパラメータを全数探索した。さらに、実装を行い、階層モデルや回帰モデルなどのベイズ推定に適用し、擬似乱数よりも収束性が向上することを数値実験で示した。

擬似乱数の64ビット化の進展に合わせて、32ビットMT法の出力を倍精度浮動小数点数に変換した際に現れる問題点を解析し、ラグを付けた統計的検定で棄却されることを示した。また、金融工学で用いられるSobol'列について、t-値と数値積分の収束性の関係を調べ、比較研究を行った。

研究成果の学術的意義や社会的意義

MCMC法は、統計科学において必要不可欠な道具となっているが、収束が非常に遅い。しかるに、MCMC法では、従来の準乱数はそのまま適用できない。ここで、CUD列と呼ばれる特殊な点列を用いると、準モンテカルロ法による期待値計算に適用できることが理論的に示されているが、具体的な点列の構成は不十分であった。本研究課題では、擬似乱数と準モンテカルロ法の手法を駆使して、MCMC法のための新しい準乱数を開発した。実際に、ベイズ統計学に現れるMCMC法に適用して、収束性の向上を確認した。この結果は、統計計算において、非常に広範な応用を持つことが期待される。

研究成果の概要（英文）：We developed quasi-random point sets for Markov chain Monte Carlo. We proposed a method to use the output values generated by short-period Tausworthe generators as quasi-random points and conducted an exhaustive search in terms of the t-value, which is a measure of uniformity. We implemented our new generators and applied them to Bayesian computation in practice. We demonstrated the effectiveness in numerical examples, such as hierarchical models and regression models.

Motivated by recent progress on 64-bit pseudorandom number generators, we analyzed the conversion from 32-bit Mersenne Twister to 53-bit double-precision floating-point numbers. From this point of view, we presented that MT19937 with a specific lag set fails several statistical tests. As another research, we conducted a comparative study of Sobol' quasi-random sequences in financial applications; in particular, we investigated the relationship between the t-value and the rate of convergence in numerical integration.

研究分野：統計科学

キーワード：擬似乱数 モンテカルロ法 準モンテカルロ法 マルコフ連鎖モンテカルロ法 統計計算 ベイズ統計学

1. 研究開始当初の背景

乱数を用いた、確率現象シミュレーションや数値積分などの数値計算法を総称して、モンテカルロ法と呼ぶ。モンテカルロ法は、計算機性能の飛躍的向上に伴って、科学技術計算において重要な役割を果たしているが、収束が非常に遅い問題点を有する。そこで、乱数をより高い一様性を有した準乱数に置き換えて高速化を図る準モンテカルロ法が知られている。1990年代以降、準乱数を用いると、金融派生商品の価格計算に現れる高次元数値積分を大幅に高速化できることが分かり、特に、金融工学の分野を中心に、盛んに用いられている。

一方、ベイズ統計学では、マルコフ連鎖モンテカルロ(MCMC)法が必要不可欠な道具となっている。しかるに、一般的な準乱数はMCMC法にはそのまま適用できない。近年、スタンフォード大学統計学部のA.B. Owen教授とその指導学生を中心に、CUD列(completely uniformly distributed sequence)と呼ばれる特殊な1次元点列を用いると、MCMC法に適用できるという理論が示された。ここで、理論研究のみが先行し、具体的な点列の構成が不十分であり、ユーザーが気軽に使える形の実装もなされていない問題点があった。

2. 研究の目的

本研究の主たる目的は、MCMC法のための準乱数の開発とベイズ統計学の数値計算への応用である。ここで、解くべき問題は、ある多変数関数 f と目標分布 π に対して、期待値 $E[f(X)]$ を通常のMCMC法より高速・高精度計算することである。この問題を解決するために、研究代表者のもつ擬似乱数と準乱数のテクニックを用いて、新しいCUD近似点列の設計を目指す。また、実装を行い、MCMC法の様々な数値計算に適用し、有用性を検証する。

関連する研究課題として、モンテカルロ法用の擬似乱数発生法の性能評価に関する研究を行う。特に、計算機の64ビット化が進展した一方、擬似乱数発生法の開発環境は64ビット化に対応しきれていない。そのため、考案されている発生法が、十分に安全性が解析されているとは言い難い。そこで、64ビット化に伴う、擬似乱数の理論的・統計的な性能評価方法を模索する。

3. 研究の方法

先行研究として、Chen-松本-西村-Owen(2012)は、短い周期のTausworthe型擬似乱数発生法を準備し、出力の一周期分を使い切った際に現れる格子構造を逆手に取って利用して、CUD近似点列として用いる方法を提案した。Chenらは、擬似乱数発生法の分野で用いられてきた、均等分布性と呼ばれる比較的粗い一様性の評価指標を用いて、パラメータの選択を行った。本研究では、準モンテカルロ法の分野で広く用いられてきた、より精緻な一様性の評価指標である (t, m, s) -netの t -値に置き換えて、点列の高性能化を狙う。前・科研費研究課題「マルコフ連鎖準モンテカルロ法の実装」において、正則連分数展開に基づいて、2次元の t -値が最適値0を達成する、見通しの良いパラメータ探索方法の着想を得た。この方法について、実際にプログラミングを行い、パラメータを探索する。また、様々なベイズ統計モデルに適用し、実データをを用いたMCMC法の実証研究を試みる。

4. 研究成果

(1) マルコフ連鎖準モンテカルロ法のための短い周期のTausworthe発生法

本研究課題の中核をなす研究成果である、マルコフ連鎖準モンテカルロ法のための短い周期のTausworthe発生法の論文が完成し、2021年3月に、Journal of Computational and Applied Mathematicsより出版された。

Tausworthe発生法は、二元体 F_2 上の原始多項式の法 $P(x)$ と乗数の多項式 $Q(x)$ の組 $(P(x), Q(x))$ をパラメータにもつが、2次元の t -値は $Q(x)/P(x)$ の正則連分数展開の部分商の次数がすべて1の時、その時に限り、最適値0をもつ。そこで、本研究では、手塚-伏見(1993)のFibonacci多項式の理論に帰着させ、正則連分数展開を用いて、2次元の t -値が最適値0を満たす多項式の組 $(P(x), Q(x))$ を数学的に絞っておき、その中から3次元以上でなるべく小さい t -値をもつ多項式の組を効率的に探索する方策を得た。実際に、アルゴリズムの細部をつめて、プログラムを作成し、周期 2^m-1 のTausworthe発生法に対して、 $m=10, \dots, 32$ までの全数探索を行い、パラメータ表を作成した。また、点列生成のプログラムを、Tausworthe発生法の定義式から、 F_2 -線形擬似乱数発生法の状態空間表現を用いたアルゴリズムに書き直して、生成速度の高速化を図った。この方法を実装して、いくつかのギブスサンプリングに適用した。

図1は、平均ベクトルが0、共分散行列が単位行列 I となる2次元正規分布に対するギブスサンプリングの結果を可視化した散布図である。左の図(New)は本研究の準乱数、右の図((Chen et al. (2012)))は先行研究の準乱数であり、4096個の点をプロットしたものである。先行研究の準

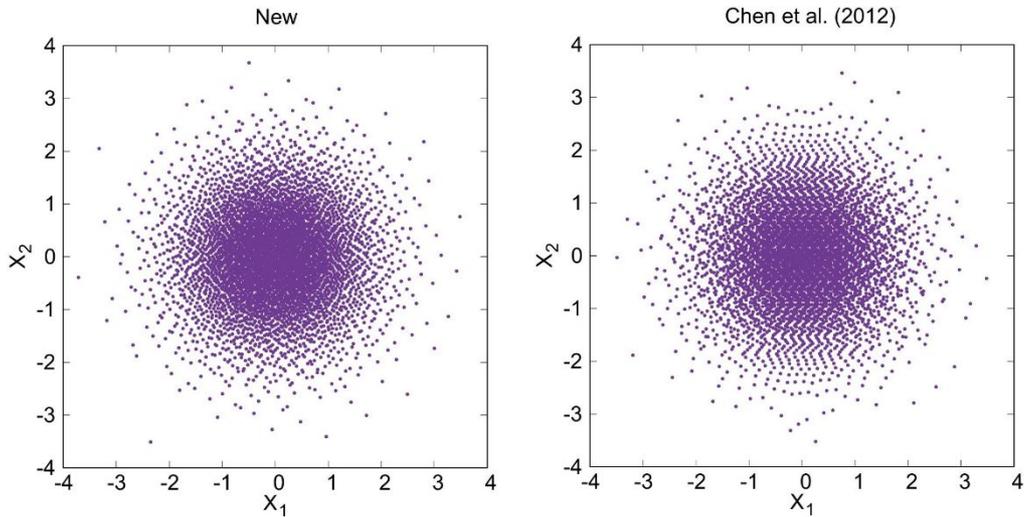


図1 2次元正規分布のギブスサンプリング

乱数は、縞模様が出ている一方、本研究の準乱数は、非常に効率よくサンプリングできており、高次元かつ相関がある場合にも有効なように設計した。

さらに、実データを用いた数値例として、Gelfand-Smith(1990)の10台のポンプの故障データに当てはめ、ギブスサンプリングにより、事後分布を計算し、ポアソン分布のパラメータの期待値 $E[\lambda_1], \dots, E[\lambda_{10}]$ を求めた。下の表は、初期値を変えて、300回試行した際に得た、期待値の標本分散である。通常の乱数と比べ、分散が大幅に小さくなり、収束性が向上した。

	λ_1	λ_2	λ_3	λ_4	λ_5
乱数	2.48e-09	3.21e-08	7.49e-09	3.47e-09	3.88e-07
Chen らの方法	2.50e-14	1.05e-12	5.12e-14	2.58e-14	1.86e-11
提案手法	2.41e-15	7.97e-14	5.60e-15	1.86e-15	1.72e-12
	λ_6	λ_7	λ_8	λ_9	λ_{10}
乱数	8.30e-08	1.34e-06	1.64e-06	1.65e-06	6.65e-07
Chen らの方法	1.39e-12	7.52e-11	1.83e-10	9.83e-11	1.68e-11
提案手法	7.84e-14	4.23e-11	8.99e-11	4.72e-11	3.80e-12

(2) 一般の有限体 F_b 上の Tausworthe 発生法への拡張と MCMC 法への応用

前述の二元体 F_2 上の Tausworthe 発生法の設計方法を一般の素数べき位数 b の有限体 F_b に拡張した。 F_2 上の Tausworthe 発生法では、2次元の t -値は最適値0を達成するが、3次元での t -値の完全な最適値(すなわち、 t -値が0)には至らなかった。3次元の t -値が0となる準乱数の構成は学術的にも応用面からも大変興味深い。

そこで、二元体 F_2 上の Tausworthe 発生法の設計方法を一般の有限体 F_b に拡張し、探索アルゴリズムを書き直して、3次元の t -値が最適値0となる Tausworthe 発生法の設計を試みた。 F_2 の場合、1つの既約多項式 $P(x)$ に対して、 $Q(x)/P(x)$ の正則連分数展開の部分商の次数がすべて1となる $Q(x)$ は2つしか存在しないことが証明されている(ここで、 $\deg(Q(x)) < \deg(P(x))$ とする)。一方、数値実験によれば、一般の F_b の場合、上記の $Q(x)$ の個数が大幅に増えて、パラメータの自由度が高くなる。特に、計算機で扱いやすい F_4 上の短周期 Tausworthe 発生法を中心に探索を行い、3次元の t -値が最適値0となるパラメータが多数見つかった。また、この発生法を実装し、MCMC 法の数値例として、3次元正規分布ギブスサンプリング、待ち行列、線形回帰モデルのベイズ推定に適用した。その結果、本研究課題を通じて開発した準乱数は、 F_2 と F_4 のどちらについても、Chen ら(2012)の従来の準乱数と比較して、計算精度が向上することを確認した。この結果を論文にまとめ、学術雑誌に投稿した。プレプリントは arXiv より入手可能である(arXiv:2303.10622)。

(3) 32 ビット MT 法の倍精度浮動小数点変換に現れる問題点とその解析

計算機環境の64ビット化に伴って、本来であれば、64ビット整数出力の擬似乱数発生法を準備して使用することが望ましい。しかしながら、32ビット版メルセンヌツイスタ(MT)法の整数出力を連結して53ビット浮動小数点出力に変換する方法がしばしば実装されている。MT法は、安全性の保障として、高次元均等分布性に基づいてパラメータが最適化されているが、整数出力の連結を行って、53ビットないし64ビットに変換して使用した場合、上位ビットの均等分布次元(均等分布性を満たす最大の次元)が下がり、乱数性が低下することが分かった。この原因を、メルセンヌツイスタ法の格子構造、とりわけ、出力から構成した格子に対して、その最短ベクトル

ルと高次元出力のビット間に現れる F_2 -線形関係式の関係を調べることにより、解析した。また、合わせて、ラグをつけた MT 法の連結出力が、通常の 32 ビット出力に比べ、より低い次元において、複数の統計的検定(系列検定・行列ランク検定など)で棄却されることを示した。この結果は、MT 法が棄却される統計的検定が明らかになったことに加え、一般論として、安易に出力の連結を行ってビット数を増やす実装方法に警鐘を鳴らす意義があるものと思われる。論文は、2019 年 7 月に Mathematics and Computers in Simulation より出版された。

(4) 金融工学の数値計算における Sobol' 列の比較

数年来行ってきた、金融工学における Sobol' 列の比較研究を完成させた。金融工学では、数百次元 ~ 1000 次元以上の数値積分が頻繁に現れる。これに対して、準モンテカルロ法を適用する際、準乱数の 1 つである Sobol' 列が用いられる。Sobol' 列は、direction number と呼ばれるパラメータをもち、その選び方には自由度がある。そのため、様々な実装がなされており、包括的な比較が望まれている。この分野の最新の結果として、Faure-Lemieux(2019)は、 F_2 上の Niederreiter 列の生成行列の行を入れ替えて、行列の上三角化を行うと、パラメータの探索を行わなくても、すでに t-値が最適化された Sobol' 型の点列が得られることを示した。この点列を実装して、本研究に取り込み、先行研究の論文では触れられていなかった金融工学の様々な数値積分に適用して、比較を行い、有効性を確認した。合わせて、2次元プロジェクションの t-値の頻度分布(特に、外れ値)を精査し、数値積分の収束性との関連性を見出した。この結果は、2019 年 3 月に Monte Carlo Methods and Applications より、論文が出版された。

その他、研究代表者が、木本貴光氏と共著で 2018 年に ACM TOMS 誌より論文を発表した、64 ビット最適均等分布 F_2 -線形擬似乱数発生法について、GitHub 上のプログラムを整理するとともに、分かりやすい説明をつけて、ユーザーの便宜を図った。

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 3件／うち国際共著 0件／うちオープンアクセス 0件）

1. 著者名 Shin Harase	4. 巻 384
2. 論文標題 A table of short-period Tausworthe generators for Markov chain quasi-Monte Carlo	5. 発行年 2021年
3. 雑誌名 Journal of Computational and Applied Mathematics	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.cam.2020.113136	査読の有無 有
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 -

1. 著者名 Shin Harase	4. 巻 25
2. 論文標題 Comparison of Sobol' sequences in financial applications	5. 発行年 2019年
3. 雑誌名 Monte Carlo Methods and Applications	6. 最初と最後の頁 61-74
掲載論文のDOI（デジタルオブジェクト識別子） 10.1515/mcma-2019-2029	査読の有無 有
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 -

1. 著者名 Shin Harase	4. 巻 161
2. 論文標題 Conversion of Mersenne Twister to double-precision floating-point numbers	5. 発行年 2019年
3. 雑誌名 Mathematics and Computers in Simulation	6. 最初と最後の頁 76-83
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.matcom.2018.08.006	査読の有無 有
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 0件／うち国際学会 3件）

1. 発表者名 原瀬晋
2. 発表標題 Fibonacci 格子の多項式類似と準乱数生成への応用
3. 学会等名 日本応用数理学会2022年度年会
4. 発表年 2022年

1. 発表者名 Shin Harase
2. 発表標題 A Construction of Short-Period Tausworthe Generators with Low Discrepancies over F_b
3. 学会等名 13th International Conference on Monte Carlo Methods and Applications (MCM 2021) (国際学会)
4. 発表年 2021年

1. 発表者名 原瀬晋
2. 発表標題 マルコフ連鎖準モンテカルロ法の使い方
3. 学会等名 2021年度統計関連学会連合大会
4. 発表年 2021年

1. 発表者名 Shin Harase
2. 発表標題 Implementing short-period Tausworthe generators for Markov chain quasi-Monte Carlo
3. 学会等名 14th International Conference in Monte Carlo & Quasi-Monte Carlo Methods in Scientific Computing (MCQMC 2020) (国際学会)
4. 発表年 2020年

1. 発表者名 原瀬晋
2. 発表標題 マルコフ連鎖準モンテカルロ法の応用
3. 学会等名 日本応用数学会2020年度年会
4. 発表年 2020年

1. 発表者名 原瀬晋
2. 発表標題 マルコフ連鎖準モンテカルロ法のための短い周期のTausworthe発生法
3. 学会等名 JCCA-2019・離散数学とその応用研究集会2019
4. 発表年 2019年

1. 発表者名 原瀬晋
2. 発表標題 正則連分数展開に基づく短い周期のTausworthe発生法
3. 学会等名 日本応用数理学会2019年度年会
4. 発表年 2019年

1. 発表者名 原瀬晋
2. 発表標題 マルコフ連鎖準モンテカルロ法のための短い周期のTausworthe発生法
3. 学会等名 2019年度統計関連学会連合大会
4. 発表年 2019年

1. 発表者名 原瀬晋
2. 発表標題 マルコフ連鎖準モンテカルロ法のための短い周期のTausworthe発生法
3. 学会等名 第1回「乱数・準乱数の数学」研究集会
4. 発表年 2019年

1. 発表者名 原瀬晋
2. 発表標題 マルコフ連鎖準モンテカルロ法のための短い周期のTausworthe発生法
3. 学会等名 立命館大学数理ファイナンスセミナー
4. 発表年 2020年

1. 発表者名 Shin Harase
2. 発表標題 An Implementation of Short-Period Tausworthe Generators for Markov Chain quasi-Monte Carlo Methods
3. 学会等名 13th International Conference in Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing (MCQMC2018) (国際学会)
4. 発表年 2018年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

64ビット最適均等分布F2-線形擬似乱数発生法 https://github.com/sharase/melg-64 マルコフ連鎖準モンテカルロ法のための短い周期のTausworthe発生法のパラメータテーブル https://github.com/sharase/cud Sobol'-Niederreiter列のパラメータテーブル https://github.com/sharase/niederreiter-nut
--

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
---------	---------