

令和 2 年 6 月 16 日現在

機関番号：62603

研究種目：基盤研究(C)（一般）

研究期間：2016～2019

課題番号：16K00345

研究課題名（和文）時間逆転シミュレーション法のレアイイベントサンプリングへの応用とデータ解析への展開

研究課題名（英文）Applications of time reverse Monte Carlo method and its relation to data analysis

研究代表者

伊庭 幸人（Iba, Yukito）

統計数理研究所・モデリング研究系・教授

研究者番号：30213200

交付決定額（研究期間全体）：（直接経費） 1,400,000円

研究成果の概要（和文）：前回の科研費で提案した「時間逆転モンテカルロ法」による珍しい現象の確率の計算法をさらに発展させた。提案手法では、ゴールとなる状態から初期状態に向かって逆に経路を追いかけることで、与えられた確率差分方程式モデルのもとでの事象の生起確率を計算する。本研究では、(1)経路の分裂・消去を含む逐次モンテカルロ法の導入、(2)高次近似にもとづく逆方向の遷移確率の改良、(3)確率の高い初期状態のほうに経路を誘導する「ガイド場」の導入、の3つの改良を試み、これらがバイアスを生ぜずに計算効率を改善することを示した。

研究成果の学術的意義や社会的意義

確率が極端に低い事象（たとえば、東京に台風が直撃する）について通常のシミュレーションの繰り返しで生起確率を計算すると多くの計算量を要する。これに対し、本研究では、前回の科研費で提案した「出发点から経路を逆にたどる」という発想にもとづいた手法を発展させ、いくつかの重要な改良を加えた。こうしたタイプの問題はさまざまな分野で応用があり、提案手法の学術的・社会的意義は大きいと考えられる。

研究成果の概要（英文）：We improved "Time reverse Monte Carlo method" for calculating rare event probabilities, which is proposed in our previous study. In this method, the probability of rare events under the assumed stochastic difference equation is calculated by stochastic dynamics that traces a time-reversed path from the target region to the initial state. We tested the following improvements of the algorithm and showed that they improve the efficiency of computation without introducing bias of the estimated probabilities: (1) Use of the sequential Monte Carlo method that splits/erases paths according to the weight of paths, (2) Improvement of the transition probability for generating time-reversed paths, (3) Introduction of "guiding field" that controls the direction of paths.

研究分野：統計学，統計物理学

キーワード：時間逆転 モンテカルロ法 逐次モンテカルロ法 確率過程 レアイイベント

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

(1) 確率台風モデルの研究を通じて「台風が東京を直撃する確率」など、滅多に起きない現象の生起確率をシミュレーションで効率的に求める方法が必要とされた。類似の問題はそれ以外にもいろいろな分野で存在すると思われた。

(2) このような問題に対して「時間逆転シミュレーション」によるアプローチを考えた。上の例でいえば、ゴールである東京から逆方向のパスをたどって発生源の南海上に至る経路を多数生成し、それを利用して確率を計算するというアイデアである。もしゴールの領域のほうが十分小さければ、この方法によって効率よく確率が計算できることが予想される。

(3) 一般の確率的力学系について、このような動機から時間逆転シミュレーションの方法を考えたものはこれまでの研究には見当たらなかった。理論的・数学的な時間逆転についての考察はあったが、実用的な計算手法についての考察は十分なされていなかった。

(4) そこで、時間逆転シミュレーションの具体的な方法について検討したところ、単純な時間逆転では正しい答えが得られないこと、また、逆関数の計算に多くのリソースが必要になるため、計算効率が低下することがわかった。

(5) こうした問題点を解決するシンプルな方法として「時間逆転モンテカルロ法」を発案したが、方法の妥当性を完全に検証して査読付き論文として発表するには至らなかった。また、問題設定によっては、計算効率の上で不十分な点が認められた。

2. 研究の目的

(1) 提案手法がバイアスのない確率の推定値を与えることを数値計算で精度よく検証する。また、一定の条件下で順方向のシミュレーションより効率的であることを確かめる。それらの結果を査読付き論文として出版する。

(2) これまでに考案したシンプルな手法が不十分な場合がいくつかあるが、そうした状況に対応できるような手法を開発する。

(3) 時間逆転した経路を導く「ガイド場」を導入することで、手法の改良を行う。

(4) より広い視野から提案手法の位置づけと応用を考える。

3. 研究の方法

(1) 以下の3種類のモデルについて提案する各手法と順方向でのシミュレーションによる数値計算を行い、バイアスの有無と計算効率を比較する。以下で、 ϵ_x 、 ϵ_y 、 ϵ_t は正規分布に従う乱数である。

2次元の確率差分方程式。

$$\begin{aligned}x_{t+1} &= x_t + (x_t + \lambda_x y_t^2) \delta t + \epsilon_x \sqrt{\delta t} \\y_{t+1} &= y_t + (y_t + \lambda_y x_t) \delta t + \epsilon_y \sqrt{\delta t}\end{aligned}$$

確率台風モデル。統計数理研究所の中野慎也氏が作成されたものを利用した。 $V(x_t)$ は実際の台風の動きを表現するようにデータから推定された場である。

$$\begin{aligned}x_{t+1} &= x_t + v_t \\v_{t+1} &= V(x_{t+1}) + w(v_t - V(x_t)) + \epsilon_t\end{aligned}$$

ローレンツ 96 モデル (気象学で用いられる連立非線形常微分方程式系) に雑音を加えたもの。本研究では9変数のモデルを使用した。

$$\frac{dx_k}{dt} = -x_{k-2}x_{k-1} + x_{k-1}x_{k+1} - x_k + F + \epsilon_t \sqrt{\delta t}$$

(2) 国際研究集会を開催し、海外の研究者に研究成果を発表するとともに、広く関連する話題を議論する。

(3) 高柳慎一氏 (総研大博士課程) との共同研究として行う。

4. 研究成果

(1) 時間逆転モンテカルロ法
順方向のシミュレーションに対応する式を

$$x_{t+1} = g(x_t) + \epsilon_t$$

とすると、時間逆転シミュレーションでは、これを逆に解いた

$$x_t = g^{-1}(x_{t+1} - \epsilon_t)$$

を計算すればよいように思われる。この方法の問題点のひとつは逆関数 f^{-1} を求めるのに高次元では多くの計算量を要することである。それだけでなく、この方法では f が保測変換 (高次元体積を保存する変換) でない場合には一般には正しい確率の値が得られない。正しい確率の値を得

ためにはヤコビ行列式の計算が必要であるが、これはさらに計算量を増大させる。

これを避けるには、計算しやすい「近似的な逆方向のダイナミクス」 $p(x_t \rightarrow x_{t-1})$ を導入して、それが定義するパスの確率密度と順方向のシミュレーションの定義する確率密度の比（以下では補正因子と呼ぶ）をパスの重みとして、多数回の時間逆転シミュレーションについて重み付きの期待値を計算すればよい。すなわち、補正因子を重みとする逐次インポートランス・サンプリング法を用いるわけである。これまでに行った研究では、具体的な逆方向のダイナミクスについては、順方向のダイナミクスが雑音 η_t を含む確率差分方程式

$$x_{t+1} = x_t + f(x_t) + \epsilon_t$$

で表現される場合について

$$x_{t+1} = x_t + f(x_{t+1}) + \epsilon_t$$

のように、関数 f の引数を1ステップ分ずらすと

$$x_t = x_{t+1} - f(x_{t+1}) - \epsilon_t$$

となり、時間を逆転したダイナミクスが逐次的に解けることを利用して、これで定義される $p(x_t \rightarrow x_{t-1})$ を近似的な逆方向のダイナミクスとすることを提案した。今回の研究では、この手法がモデルでバイアスのない確率の推定値を与えること、また状況により順方向のシミュレーションより高い効率を持つことを精密に確かめることができた（刊行論文参照）。

（2）高次近似による改良

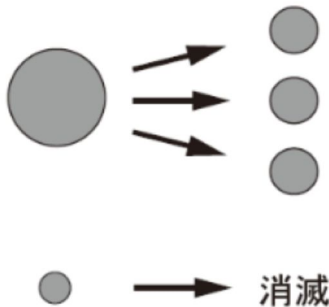
上記の手法の問題点として、雑音 ϵ_t が小さい場合に、関数 f の引数を1ステップ分ずらす近似が雑音の強さに相対的に精度が悪くなり、計算効率が悪化するという問題がある。今回の研究では、次のような高次近似による改良法を与えた。（1）で述べた方法では、 $x_{t+1} = x_t + f(x_t) + \epsilon_t$ の $f(x_t)$ を $f(x_{t+1})$ で置き換えた。これに対して、改良法では雑音とステップ幅が小さければ、近似的に $x_t \approx x_{t+1} - f(x_{t+1})$ となることを利用して、 $f(x_{t+1})$ を $f(x_{t+1} - f(x_{t+1}))$ で置き換える。すると、近似的ダイナミクスとして、

$$x_t = x_{t+1} - f(x_{t+1} - f(x_{t+1})) - \epsilon_t$$

が得られる。（1）で述べた方法を1次、ここで述べた方法を2次として、同様の考え方でN次の近似を考えることも可能である。数値実験により、2変数の確率微分差分方程式モデルで雑音が小さい場合に効率が向上することを確かめることができた。

（3）粒子の淘汰 - SISからSMCへ

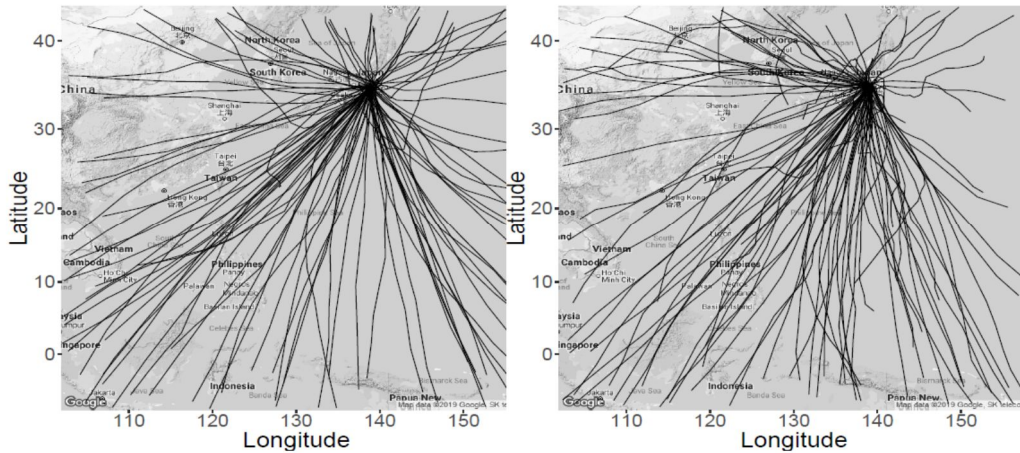
逐次インポートランス・サンプリング法(SIS)に共通の問題として、ステップ数が大きくなると、各回のシミュレーションの重みが片寄ってしまい、少数のサンプルに重みが集中してしまう問題がある。これを防ぐ方法として、多数のシミュレーションを並列に行い、重みが大きくなったものを途中の時刻で枝分かれさせ、重みが小さいものは消す方法がある。一般にこれを逐次モンテカルロ法(SMC)と呼ぶ。



今回の研究では、この方法を取り入れることで、ステップ数の多い場合に期待値のバイアスを生ぜずに計算効率が改善することを示した。例としてはローレンツ96モデルを用いた。

（4）ガイド場の導入とバイアスの検出

提案手法の問題点として、ゴールから出発した時点でスタート方向とは違う方向に進むパスが一定の割合で生じることがある。これは、補正因子に初期分布の情報が含まれていないことに起因している。これらのパスがあっても原理的に期待値のバイアスは生じないが、計算効率は下がる。この問題を軽減させるため、初期値の領域にひびくような仮想的な「ガイド場」を含めてシミュレーションを行い、その影響を重みでキャンセルすることでバイアスのない期待値を求めることが考えられる。今回の研究では、確率台風モデルの例についてこの手法を実装し、計算効率が改善することを示した。ガイド場が強すぎると、有限回のシミュレーションではガイド場の影響が重みでキャンセルできなくなりバイアスが生じる可能性があるが、それを診断する方法についても考察した。図は、確率台風モデルについて時間逆転シミュレーションで求めたパスの例を示す（高柳慎一氏の博士論文より許可を経て引用）。図の左がガイド場なし、右がありに対応しているが、東京から初期状態と反対方向の北方に行く台風の数が減っていることがわかる。



(5) 補正因子の解釈

時間逆転シミュレーションでバイアスのない期待値を得るために必要な重み因子は、連続時間の極限で

$$\exp(-\int \text{div } f(x_t) dt)$$

というシンプルな形になる。これはたとえば経路に沿った状態空間の体積変化(エントロピー生成)として解釈することもできる。刊行論文ではこの因子を、ヤコビ行列式の連続時間極限、および、提案手法の SIS の重みの連続時間極限、の 2 つの方法で導出したが、第 3 の解釈として、確率微分方程式に随伴するフォッカープランク方程式とコルモゴロフ後向き方程式の比較から導出することも可能である。このことは、提案手法がコルモゴロフ後向き方程式の逐次モンテカルロ法による解法としても捉えられることを示している。

(6) 粒子平滑化法との関連

時間逆転モンテカルロ法を開発した動機は「生成確率の低い現象を時間逆転によってシミュレートする」ということにあり、データ解析とは直接の関係がないが、数理的な面で関連する研究(逐次モンテカルロ法による粒子平滑化)が、時系列データ解析の分野で行われている。データ解析の文脈では、本研究で扱った場合は「時系列の最終時刻にのみ観測値がある」という特別なケースに相当する。ただし、時系列データ解析の場合には、通常は途中の時刻に多くの観測値があり、フィルタリングの過程でその情報が取り込めるため、本研究で扱った例とはかなり条件が異なっている。その意味では別の問題と考えられるが、ガイド場の導入の部分は平滑化問題における知見を参考にした部分もある。粒子平滑化との関連についてお教え頂いた Arnaud Doucet 氏(Oxford 大学)に感謝する。

(7) 最尤推定への応用

時間逆転モンテカルロ法では「ゴールに到達する確率の計算」を目的としたが、「ゴール」を「ある確率過程によって生成された対象」と考えれば、提案手法を「順方向のシミュレーションの定義する統計モデルのもとでの対象の尤度を計算する方法」と読み替えることも可能である。研究計画書にも記載したように、集団遺伝学のモデルではこれに相当する手法の先行研究があるが、原理的にはさまざまなモデルの尤度計算に役立つ可能性がある。今回の研究では、空間における点配置の問題について、提案手法を応用することを試みた。空間点配置の統計モデルとしてはギブス点過程を用いる方法がよく知られているが、点配置が時間的に逐次生成されるモデルを考えることもできる。最終的な配置のみがデータとして与えられる場合には、生成順序について和をとることが必要になるが、これは一般には計算量が大で、効率的な計算法が必要とされる。本研究では、簡単な例について具体的なアルゴリズムを考え、予備的なプログラムの作成までを行ったが、有意義な結果を得るには至っていない。この方向の発展は今後の課題としたい。

5. 主な発表論文等

〔雑誌論文〕 計1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 0件）

| | |
|--|----------------------|
| 1. 著者名 Shinichi Takayanagi and Yukito Iba | 4. 巻 87 |
| 2. 論文標題 Backward Simulation of Stochastic Process Using a Time Reverse Monte Carlo Method | 5. 発行年 2018年 |
| 3. 雑誌名 Journal of the Physical Society of Japan | 6. 最初と最後の頁 124003 |
| 掲載論文のDOI（デジタルオブジェクト識別子） https://doi.org/10.7566/JPSJ.87.124003 | 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3件（うち招待講演 0件 / うち国際学会 1件）

| |
|--------------------------------------|
| 1. 発表者名 伊庭幸人 |
| 2. 発表標題 タイムトラベルとカオス（及びビリヤードの計算練習） |
| 3. 学会等名 学際計算物理学研究会（阪大） |
| 4. 発表年 2018年 |

| |
|---|
| 1. 発表者名 伊庭幸人 |
| 2. 発表標題 時間逆転シミュレーションによるレアイベント・サンプリング |
| 3. 学会等名 「非線形科学とその周辺」セミナー |
| 4. 発表年 2017年 |

| |
|---|
| 1. 発表者名 Iba, Y. and Takayanagi, S. |
| 2. 発表標題 Rare event sampling using multicanonical MCMC and time-reversed simulation |
| 3. 学会等名 Simulations Encounter with Data Science（国際学会） |
| 4. 発表年 2017年 |

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

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 |
|-------|------------------------------------|---|----|
| 研究協力者 | 高柳 慎一 (Takayanagi Shinichi) | 総合研究大学院大学・複合科学研究科統計科学専攻・大学院生(博士課程) (12702) | |