

PDMP によりスパイク付きの非絶対連続分布からもサンプリングが可能になる

司馬博文（総合研究大学院大学先端学術院統計科学コース 5 年一貫博士課程 2 年）

2/25/2025

Abstract

区分確定的マルコフ過程 (PDMP) とはマルコフ連鎖、拡散過程に続いて、近年活発にモンテカルロ法に利用されつつある連続時間マルコフ過程のクラスである。より速く収束するサンプラーが構成しやすいこと、モンテカルロ推定量にバイアスを導入しないようなサブサンプリング（バッチ実行）が可能であることから、近年特に統計・機械学習の分野でも注目が高まっている。

本ポスター発表では PDMP と MCMC の違いを明確にすると同時に、PDMP のシミュレーション法に紙面を割き、最新のアルゴリズムと筆者開発のパッケージの実装の強みを紹介する。加えて「 δ -部分を持った非絶対連続確率分布からも正確なサンプリングが可能」という PDMP に基づくモンテカルロ法の第 3 の利点に焦点を当て、特にベイズ変数選択を応用として取り上げる。

1 導入：新たなモンテカルロ法の出現

マルコフ連鎖モンテカルロ法 (MCMC: Markov Chain Monte Carlo) や逐次モンテカルロ法 (SMC: Sequential Monte Carlo) などのモンテカルロ法は、21 世紀に入って計算機が普及してのち、ベイズ統計の応用範囲を爆発的に拡大する立役者となった (Gilks et al., 1996), (Doucet et al., 2001), (Martin et al., 2024). 近年提案された**区分確定的マルコフ過程** (PDMP: Piecewise Deterministic Markov Process) を利用した新たなモンテカルロ法も、この系列に続くものと理解できる。

MCMC と SMC はいずれも離散時間のマルコフ過程 $(X_n)_{n=1}^{\infty}$ のシミュレーションに基づく。ここでは特に MCMC に焦点を当てる。ベイズ統計では事後分布 π を導くことで推論を実行するが、事前分布や尤度の選択範囲を大きく制限しない限り、 π は既知の分布とはならない。たとえ π の関数形がわかっている場合でも、事後分布 π の平均を求めるなど、これを後続の統計解析に供することは必ずしも簡単ではない。そんな中でも、事後分布 π を平衡分布にもつエルゴード的なマルコフ連鎖 (X_n) を設計することができれば、大数の強法則

$$\frac{1}{N} \sum_{n=1}^N f(X_n) \xrightarrow{N \rightarrow \infty} \int f(x) \pi(dx) \text{ a.s.} \quad f \in \mathcal{L}^1(\pi) \quad (1)$$

が成り立つ (Kulik, 2018, pp. 175–176). この事実を用いて事後分布 π を近似する枠組みが MCMC であり、条件を満たす (X_n) の構成法に Gibbs サンプラー (Geman and Geman, 1984) やランダムウォーク Metropolis-Hastings 法 (Hastings, 1970) などがある。

多くの MCMC 法は物理学に由来する。例えば Gibbs サンプラーは Ising モデル (Glauber, 1963), Metropolis-Hastings 法は状態方程式 (Metropolis et al., 1953) のシミュレーションにおいて最初に提案された。MCMC という名前は (Geman and Geman, 1984) や (Hastings, 1970) などでは統計学における確率分布からのサンプリングの問題に応用されて初めて付いたものである。

連続時間のマルコフ過程である PDMP を用いた新しいモンテカルロ法も、やはり物理学における研究 (Bernard et al., 2009), (Peters and de With, 2012) で最初に使われ始め、初の氷の液相転移の全原子シミュレーションも成功させている (Faulkner et al., 2018). これに倣い統計でも Bouncy Particle Sampler (BPS) (Bouchard-Côté et al., 2018) や Zig-Zag Sampler (Bierkens et al., 2019), Forward PDMP (Michel et al., 2020) などの汎用アルゴリズムが提案され、この手法群に PDMP に由来する**区分確定的モンテカルロ法** (PDPMC: Piecewise Deterministic Monte Carlo) の名前がついた。

1.1 ポスターと本稿の構成

本ポスターでは紙面を3つにわけ、本稿もこれに対応する3つの節を用意し、それぞれを詳説する。初めに PDMC と MCMC (従来法) の違いを、シミュレートされるマルコフ過程の違いとアルゴリズムの違いの両面からみる (第2節)。続いて Poisson 剪定をはじめとする PDMP のシミュレーションの自動化法を解説し、筆者が開発したパッケージ PDMPFlux.jl (Shiba, 2025) ではどのようなアプローチを取っているかを解説する (第3節)。最後に MCMC では不可能であったが PDMC で可能になることとして、デルタ部分を持つ非絶対連続分布

$$p(dx) = \prod_{i=1}^d \left(\omega_i p_i(x_i) dx_i + (1 - \omega_i) \delta_0(dx_i) \right) \quad (2)$$

からのサンプリング法を紹介し、ベイズ変数選択への応用に触れる (第4節)。

2 PDMC と MCMC の違い

既存の PDMC 法はいずれも、

$$\pi(x) \propto e^{-U(x)}, \quad U \in C^1(\mathbb{R}^d), \quad (3)$$

という表示を持つ確率分布を対象に扱う。定数の差を除いて負の対数密度 $-\log \pi$ にあたる U をポテンシャルとも呼び、PDMC アルゴリズムはその勾配 ∇U のみを入力に取る。さらに $\nabla U(x)$ の計算コストが高い場合は、 $\nabla U(x)$ の不偏推定量を代わりに使うだけでも、Poisson 剪定 (命題 3.1) を通じて正確に π に収束する PDMP のシミュレーションが可能である (Bierkens et al., 2019)。MCMC では各イテレーションでデータの一部のみを用いたバッチ実行を、収束先 π を変えずに行うことが難しかったために、PDMC は特に大規模データを用いたベイズ推論への応用が期待されている。

PDMC と MCMC の最大の違いはアルゴリズムであるが、これはシミュレートしようとしているマルコフ過程の性質の違いに起因する。例えば最も広く使われている MCMC 法の1つである Metropolis-adjusted Langevin 法 (MALA, Besag, 1994) は、ポテンシャル U が定める Langevin 拡散過程

$$dL_t = -\nabla U(L_t) dt + \sqrt{2} dB_t, \quad (4)$$

に基づくが、 (L_t) は正確なシミュレーションが困難であり、Euler-Maruyama 法などの離散化が必要になる。この段階で数値誤差が入り、このままでは大数の法則 (1) が成り立たなくなる。これを防ぐためには Metropolis-Hastings 様の棄却法を用いる必要がある。こうして、最終的にシミュレートされるマルコフ過程は離散時間過程になり、さらに人工的な対称性が導入され、収束が遅くなる。

PDMP とは式 (4) のような拡散過程から、拡散項 dB_t を除いた代わりに、ランダムなジャンプ項を加えたものを言う。このクラスのマルコフ過程は、ジャンプの時刻を棄却法によりシミュレートすることで、離散化誤差を伴わない正確なシミュレーションが可能である。その結果人工的な対称性が少なく、一般にその分収束を速くすることができる (Andrieu and Livingstone, 2021)。

式 (3) で与えられる確率分布 π を平衡分布にもつ PDMP (X_t, V_t) は、次の2ステップの繰り返しでシミュレートできる。まず、

$$m(t) = \left(v \cdot \nabla U(x_0 + tv_0) \right), \quad i = 1, 2, \dots \quad (5)$$

を強度関数に持つ非一様 Poisson 点過程の最初の到着時刻 T_1 をシミュレートし、時刻 T_1 まで初期速度 v_0 で決定論的に運動する。このイベント発生地点を新しい $x_1 := x_0 + T_1 v_0$ とし、新たな速度 v_1 をサンプリングして同じことを繰り返す。前述の BPS, Zig-Zag Sampler, Forward PDMC などの既存手法はいずれもこの枠組みの下で設計されており、新たな速度 v_1 の決め方のみが違う。

式 (5) の強度を持つ非一様 Poisson 点過程のシミュレーションでは、**Poisson 剪定** (thinning, Lewis and Shedler, 1979) と呼ばれる棄却法が主流であるが、その実行には m の上界 M を構成する必要がある (命題 3.1)。ポテンシャル U はモデルの尤度を含み、大変複雑である。従って ∇U が有界になるロジスティックモデルなど、PDMC の応用先は限られると考えられていた。

3 自動 Poisson 剪定と PDMPflux パッケージの紹介

命題 3.1 (Poisson 剪定). $m \leq M$ を関数とする. M を強度にもつ Poisson 点過程の点 T_1, T_2, \dots のそれぞれについて, 確率 $m(T_1)/M(T_1), m(T_2)/M(T_2), \dots$ のみで採択し, 他を取り除いて残る点群は, m を強度にもつ Poisson 点過程に従う.

$[0, t_{\max}]$ 上の点と $[t_{\max}, 2t_{\max}]$ 上の点とは互いに独立に決まるため, $t_{\max} > 0$ はアルゴリズムのハイパーパラメータとし, 最初の点 T_1 が採択されるまで幅 t_{\max} の区間上での剪定を繰り返せば良い. (Andral and Kamatani, 2024) にて, 一般の関数 m に対して, その $[0, t_{\max}]$ 上での上界を自動的に求めながら剪定を実行する方法が提案された. その方法とは次の通りである. まずグリッドの数 n を固定し, $[0, t_{\max}]$ の n 等分点 $\{0 = t_0 < t_1 < \dots < t_n = t_{\max}\}$ を考え, 各点 t_i 上での接線を自動微分 (Baydin et al., 2017) を用いて計算する. 続いてそれらの接線同士の交点を求め, その y 座標と両端点 $m(t_i), m(t_{i+1})$ の3点の最大値を, 小区間 $[t_i, t_{i+1}]$ 上での m の上界とする.

こうして得られた $[0, t_{\max}]$ 上の区分定数関数 M は, $t_{\max} \rightarrow 0$ または $n \rightarrow \infty$ の極限で m の上界を与えるが, m が $[0, t_{\max}]$ 上で激しく変化する場合, m の上界になっていない場合がある. そこで (Andral and Kamatani, 2024) は, Poisson 剪定の最中に評価される比 $m(t)/M(t) > 1$ が1を超えた場合, t_{\max} の値を縮めて $[0, t_{\max}]$ 上での剪定をやり直す. Andral は Python パッケージ `pdmp_jax` にこのアルゴリズムを実装し, BPS, Zig-Zag Sampler, Forward PDMP を含む5つのサンプラーを利用可能にした.

しかし, この方法は $M(t) < m(t)$ が発生してしまっている点 t を検出できずに, $t < T_1$ を満たす到着時刻 T_1 を提案してしまった場合に, シミュレートしている PDMP に漸近的に消えないバイアスを導入してしまう. 従って U の性質に対してグリッド数 n を小さく, または t_{\max} を大きく取りすぎた場合, M の誤りを検出できず, 使用者が知らないうちに大数の法則 (1) が成り立たなくなってしまう可能性がある.

筆者が開発したパッケージ PDMPflux (Shiba, 2025) ではこの問題の解決を図った. $M(t) < m(t)$ が小区間 $t \in [t_i, t_{i+1}]$ で発生するためには, m が $[t_i, t_{i+1}]$ 上で1つ以上の変曲点を持つ必要があることに注目する. そこでグリッド $\{t_i\}_{i=0}^N$ 上で接線だけでなく2階微分係数も計算し, m が $[0, t_{\max}]$ 上で高々1つしか変曲点を持たないように t_{\max} を調整することを提案した. このことにより, U の Lipschitz 係数に関する一定の仮定の下で, $t_{\max} \rightarrow 0, N \rightarrow \infty$ の極限を考えずとも, $m < M$ の成立を保証できるようになる.

こうしてパッケージ PDMPflux はポテンシャル U を与えるだけで π からの正確な PDMP サンプリングを実行することのできるほとんど唯一のパッケージとなった. 加えて, 第4節で後述する Sticky Zig-Zag を加えた6つのアルゴリズムが利用可能である. MCMC ベースのベイズ推論エンジンである Stan に替わる, 新たなバックエンドとしての利用も可能である.

4 スパイク付きの非絶対連続分布からのサンプリング

MCMC では難しく, PDMP で真に新たに可能になったこととして, 式 (2) で与えられるような, δ -部分を持った非絶対連続確率分布 p からのサンプリングが挙げられる. この事実は次のようにも理解できる. 式 (2) で与えられるような δ_0 -部分を持った分布 p は, δ_0 -部分を分散が小さい正規分布 $N(dx_i; 0, \epsilon^2)$ に近似した

$$p_\epsilon(dx) = \prod_{i=1}^d \left(\omega_i p_i(x_i) dx_i + (1 - \omega_i) N(dx_i; 0, \epsilon^2) \right) \quad (6)$$

という絶対連続分布の, 分散が小さくなる極限 $\epsilon \rightarrow 0$ と理解できる. 多くの標準的な MCMC は p_ϵ からのサンプリングは可能でも, $\epsilon \rightarrow 0$ の極限で誤った分布 $\otimes_{i=1}^d p_i \neq p$ に収束するマルコフ過程に収束してしまう (未発表原稿). さらに $\epsilon > 0$ が小さいほど, 長時間の相関が強くなり, 非現実的なほど長い時間アルゴリズムを実行しない限り, 大きなバイアスを持つてしまう.

一方で標準的な PDMP は $\epsilon \rightarrow 0$ の極限でもまた別の PDMP になり, これは正確に p に収束する. この極限過程を直接シミュレーションすることで, p からのサンプリングを効率的に実現する手法 Sticky PDMP が (Bierkens et al., 2023) により提案された.

Sticky PDMP はベイズ変数選択に大きな応用を持つ。(Mitchell and Beauchamp, 1988) で提案された spike-and-slab 事前分布はまさに式 (2) で与えられる形をしているが、従来の MCMC では直接の事後分布サンプリングが難しかったため、モデルの空間 $\{0, 1\}^d$ 上で動く Gibbs サンプラーと組み合わせてサンプリングをしたり (George and McCulloch, 1993)、連続近似 (6) を代わりに用いたりされることが多かった。しかし、連続近似をベイズ変数選択に用いた場合は、説明変数 x_i がモデルに含まれる事後確率 (PIP: Posterior Inclusion Probability) を正確に計算するために追加の処理が必要になる (Hahn and Carvalho, 2015)。

筆者は Sticky PDMP のダイナミクスを分析すると同時に、Sticky Zig-Zag アルゴリズムを PDMPLux に実装することで、 δ -部分を持った非絶対連続分布 p からのサンプリングが可能であることの、ベイズ変数選択以外への応用も模索している。例えば、特定の密度推定の問題においてミニマックス最適なベイズ事後密度推定量を与える事前分布は離散分布になる (Gangopadhyay and Mukherjee, 2021) ように、非絶対連続事前分布を用いることで正確に計算可能なベイズ推定量の幅が広がる可能性がある。

5 参考文献

- Andral, C., and Kamatani, K. (2024). [Automated techniques for efficient sampling of piecewise-deterministic markov processes](#).
Andrieu, C., and Livingstone, S. (2021). [Peskun–Tierney ordering for Markovian Monte Carlo: Beyond the reversible scenario](#). *The Annals of Statistics*, 49(4), 1958–1981.
Baydin, A. G., Pearlmutter, B. A., Radul, A. A., and Siskind, J. M. (2017). [Automatic differentiation in machine learning: A survey](#). *J. Mach. Learn. Res.*, 18(1), 5595–5637.
Bernard, E. P., Krauth, W., and Wilson, D. B. (2009). [Event-chain monte carlo algorithms for hard-sphere systems](#). *Phys. Rev. E*, 80, 056704.
Besag, J. E. (1994). [Comments on “Representations of Knowledge in Complex Systems” by U. Grenander and M. I. Miller](#). *Journal of the Royal Statistical Society. Series B (Methodological)*, 56(4), 591–592.
Bierkens, J., Fearnhead, P., and Roberts, G. (2019). [The Zig-Zag Process and Super-Efficient Sampling for Bayesian Analysis of Big Data](#). *The Annals of Statistics*, 47(3), 1288–1320.
Bierkens, J., Grazi, S., Meulen, F. van der, and Schauer, M. (2023). [Sticky PDMP Samplers for Sparse and Local Inference Problems](#). *Statistics and Computing*, 33(1), 8.
Bouchard-Côté, A., Vollmer, S. J., and Doucet, A. (2018). [The bouncy particle sampler: A nonreversible rejection-free markov chain monte carlo method](#). *Journal of the American Statistical Association*, 113(522), 855–867.
Doucet, A., Freitas, N., and Gordon, N. (Eds.). (2001). [Sequential Monte Carlo Methods in Practice](#). Springer New York.
Faulkner, M. F., Qin, L., Maggs, A. C., and Krauth, W. (2018). [All-atom computations with irreversible markov chains](#). *The Journal of Chemical Physics*, 149(6), 064113.
Gangopadhyay, U., and Mukherjee, G. (2021). [On discrete priors and sparse minimax optimal predictive densities](#). *Electronic Journal of Statistics*, 15(1), 1636–1660.
Geman, S., and Geman, D. (1984). [Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images](#). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6), 721–741.
George, E. I., and McCulloch, R. E. (1993). [Variable selection via gibbs sampling](#). *Journal of the American Statistical Association*, 88(423), 881–889.
Gills, W. R., Richardson, S., and Spiegelhalter, D. (Eds.). (1996). [Markov Chain Monte Carlo in Practice](#). Chapman & Hall.
Glauber, R. J. (1963). [Time-Dependent Statistics of the Ising Model](#). *Journal of Mathematical Physics*, 4(2), 294–307.
Hahn, P. R., and Carvalho, C. M. (2015). [Decoupling shrinkage and selection in bayesian linear models: A posterior summary perspective](#). *Journal of the American Statistical Association*, 110(509), 435–448. doi: 10.1080/01621459.2014.993077.
Hastings, W. K. (1970). [Monte Carlo Sampling Methods Using Markov Chains and Their Applications](#). *Biometrika*, 57(1), 97–109.
Kulik, A. (2018). [Ergodic behavior of markov processes: With applications to limit theorems](#), Vol. 67. De Gruyter: Berlin, Boston.
Lewis, P. A. W., and Shedler, G. S. (1979). [Simulation of nonhomogeneous poisson processes by thinning](#). *Naval Research Logistics Quarterly*, 26(3), 403–413.
Martin, G. M., Fraizier, D. T., and Robert, C. P. (2024). [Computing bayes: From then ‘til now](#). *Statistical Science, Advanced Publication*, 1–17.
Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). [Equation of state calculations by fast computing machines](#). *The Journal of Chemical Physics*, 21(6), 1087–1092.
Michel, M., Durmus, A., and Sénécal, S. (2020). [Forward event-chain monte carlo: Fast sampling by randomness control in irreversible markov chains](#). *Journal of Computational and Graphical Statistics*, 29(4), 689–702.
Mitchell, T. J., and Beauchamp, J. J. (1988). [Bayesian variable selection in linear regression](#). *Journal of the American Statistical Association*, 83(404), 1023–1032.
Peters, E. A. J. F., and de With, G. (2012). [Rejection-free monte carlo sampling for general potentials](#). *Physical Review E*, 85(2).
Shiba, H. (2025). [PDMPLux v0.3.3](#). Package Release.