

或る区分確定的マルコフ過程のスケーリング極限

統計科学コース 5年一貫制博士課程3年

司馬 博文

主任指導教員： 鎌谷 研吾

1 区分確定的マルコフ過程の歴史：OR → 計算物理 → ベイズ統計

区分確定的モンテカルロ法 (PDMP: Piecewise Deterministic Markov Process) とは、はじめ計算物理学分野で提案されたモンテカルロ法（例えば Bernard et. al. (2009) のイベント連鎖法や Peters & de With (2012) の棄却フリー法）に対する、数理的な観点からの総称である。事実上ほとんどの PDMP 手法は、目標分布に従う区分線型な軌道を生成することでサンプリングを達成する。そのため区分線型モンテカルロ法と呼ばれるのも良いかもしれないが、OR 分野に古くから存在した用語 Davis (1984) を踏襲して PDMP と呼ぶ。

PDMP の前身たるイベント連鎖法は計算物理で不動の地位を占めている。例えば2次元の水の液相転移で、 10^6 オーダーの全原子シミュレーションに成功し、途中で六角形構造が創発することを初めて観察できた Bernard & Krauth (2011)。これは剛体系などの近距離相互作用系では、従来の state-of-the-art であった HMC (Hamiltonian Monte Carlo) を超える圧倒的なスケーラビリティが達成できることに因る。

この事実を統計的な語彙で表現し、統計コミュニティに輸入したのが Bouchard-Côté et. al. (2018) である。さらに BPS (Bouncy Particle Sampler) というかわいい名前もつけた。HMC は次元 d に対して $O(d^{5/4})$ の計算複雑性を持つ一方、PDMP はスパースな Markov 確率場で $O(d)$ スケーリングを達成する。他にも PDMP はデータサイズに対するスケーリングも良い。確率的勾配を用いた実装で、有効な control variate が利用可能な状況では、データ数に依存しない計算量を達成できる Bierkens et. al. (2019)。

2 スケーリング解析を通じた PDMP 手法の性能比較

しかしベイズ統計において汎用的に使える PDMP の開発は鈍化しているように思われる。実際、以上の結果は全て「適切な実装を作用した場合」という但し書き付きであり、HMC に比べて汎用性に劣る。計算物理の場合と違い、「PDMP を採用してまで正確にシミュレーションしたいモデル」が手頃にあるわけでもない。

何より、手法の比較の枠組みが不足していると言える。例えば、イベント連鎖法の開発者率いる Michel et. al. (2020) は FECMC (Forward Event-chain Monte Carlo) を開発し、計算物理での知見をフルに活かして、BPS より大きく効率的なアルゴリズムを提案した。実験的にその有効性は明らかであったが、理論はほとんどつけられておらず、原著の読みにくさもあり、応用が限られていた。

FECMC が BPS よりも「速い」ことは次のように定量化できる。簡単のため目標分布 π^d は d 次元の標準正規分布とすると、アルゴリズムの出力はいずれも連続時間の確率過程 $(X_t^d)_{t \geq 0}$ になる。このアルゴリズムの探索が最も遅い方向の1つに、(負の) 対数尤度 $h^d(x) := -\log \pi^d(x)$ がある。標準 Gauss でいえば動径の自乗 $h^d(x) = \|x\|_2^2/2$ である。この方向への射影 $\tilde{Y}_t^d := h^d(X_t^d)$ は探索すべき独立な方向の数 d が多ければ多いほどほとんど動かなくなってしまう。そこで我々は $O(d)$ の加速を施した正規化過程

$$Y_t^d := \frac{h^d(X_t^d) - \mathbb{E}[h(X_t^d)]}{\sqrt{\text{Var}[h(X_t^d)]}}, \quad t \geq 0,$$

の $d \rightarrow \infty$ におけるスケーリング極限を考える． $X_0^d \sim \pi^d$ という平衡分布からのスタートを仮定すると，FECMC も BPS も，全く同じ関数形での極限を持つ：

$$(1) \quad dY_t = -\frac{\sigma^2}{2} Y_t dt + \sigma dB_t.$$

ただし，拡散係数 σ だけ，FECMC と BPS で値が違うのである．仮に σ_F と σ_B と表すとすると， $\sigma_F > \sigma_B$ が成り立つ．(1) は Ornstein-Uhlenbeck 過程であり，ドリフト関数の違いは時間変換 $t \mapsto (\sigma_F^2/\sigma_B^2)t$ で写り合う．そのため $\sigma_F > \sigma_B$ は FECMC の方が確率過程として速いということを意味する．この事実はそのまま， h の平均推定に FECMC を用いた場合，BPS の場合よりも $\sigma_F^2/\sigma_B^2 \approx 1.77$ 倍だけ漸近分散を削減することができることを意味する．

以上は純粋に確率過程に関する議論であったが，アルゴリズムの CPU Time も加味すると， h の平均推定において有効サンプル数に実に約 15 倍の違いが， $d \approx 10$ 次元以上では安定して現れる．

3 PDMP の収束判定への応用

我々の証明を精査すると， Y_t^d の時間微分の過程 $R_t^d := (X_t^d | V_t^d)$ が鍵を握っていることがわかる． V_t^d というのはサンプラーの速度ベクトルであり，従来は純粋な補助変数としてアルゴリズムの実行後は直ちに捨てられるものであった．しかし V_t^d にも重要な情報が眠っており，これを利用すれば次元 d に依らない精度で h の Monte Carlo 推定量の漸近分散推定が可能であることを示す．これは一般に， h のような探索が困難な方向でサンプラーが $O(d)$ で減速していくような状況でも，その事実自体は $h(X_t^d)$ などの量の時間微分を見ることで $O(1)$ で検出可能であることを示唆する．PDMP アルゴリズムは内部で h の時間微分を用いるので， h の場合は追加の計算はほとんど発生しない．この方向で将来的には PDMP（果てには HMC のような速い補助変数を持つアルゴリズム）に効率的な収束判定指標が開発できるかもしれない．

参考文献

- Bernard, E. P., Krauth, W., & Wilson, D. B. (2009). Event-chain Monte Carlo Algorithms for Hard-sphere Systems, *Physical Review E*, **80**, 056704.
- Peters, E. A. J. F., & de With, G. (2012). Rejection-Free Monte Carlo Sampling for General Potentials, *Physical Review E*, **85**, 026703.
- Davis, M. H. A. (1984). Piecewise-Deterministic Markov Processes: A General Class of Non-Diffusion Stochastic Models, *Journal of the Royal Statistical Society: Series B (Methodological)*, **46**(3), 353-376.
- Bernard, E. P., & Krauth, W. (2011). Two-Step Melting in Two Dimensions: First-Order Liquid-Hexatic Transition, *Physical Review Letters*, **107**, 155704.
- Bouchard-Côté, A., Vollmer, S. J., & 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.
- Bierkens, J., Fearnhead, P., & Roberts, G. (2019). The Zig-Zag Process and Super-Efficient Sampling for Bayesian Analysis for Big Data, *The Annals of Statistics*, **47**(3), 1288-1320.
- Michel, M., Durmus, A., & 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.