大学共同利用機関法人 情報・システム研究機構
統計数理研究所
The Institute of Statistical Mathematics
SOKENDAI

PDMP によりスパイク付き非絶対連続分布からもサンプリングが可能になる

総合研究大学院大学　司馬博文

モンテカルロ法に使う確率過程の進化

拡散過程 MCMC → PDMP 区分確定的マルコフ過程

$$dX_t = a(X_t)\,dt + b(X_t)\,dB_t$$

➡ ジャンプ過程へ

$$dX_t = a(X_t)\,dt + \int_{u\in\mathbb{R}^d} c(X_{t-},u)\eta(dtdu)$$

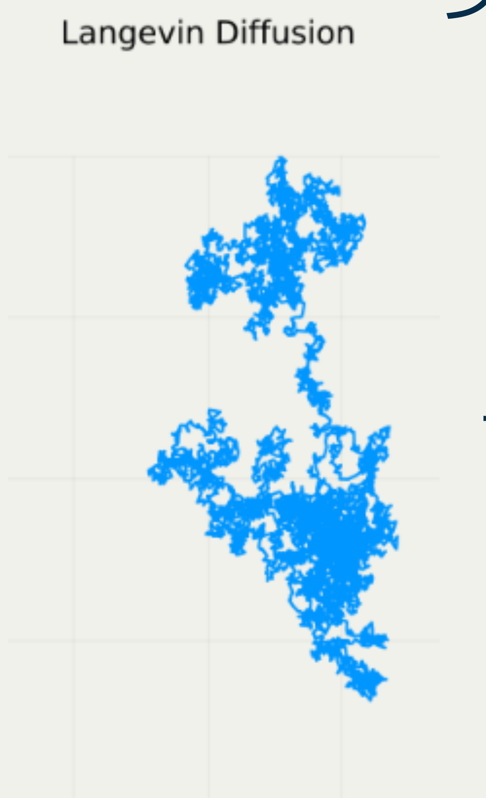
MCMC の枠組み

① 離散化
Euler-丸山など

② 棄却法
Metropolis-Hastings など

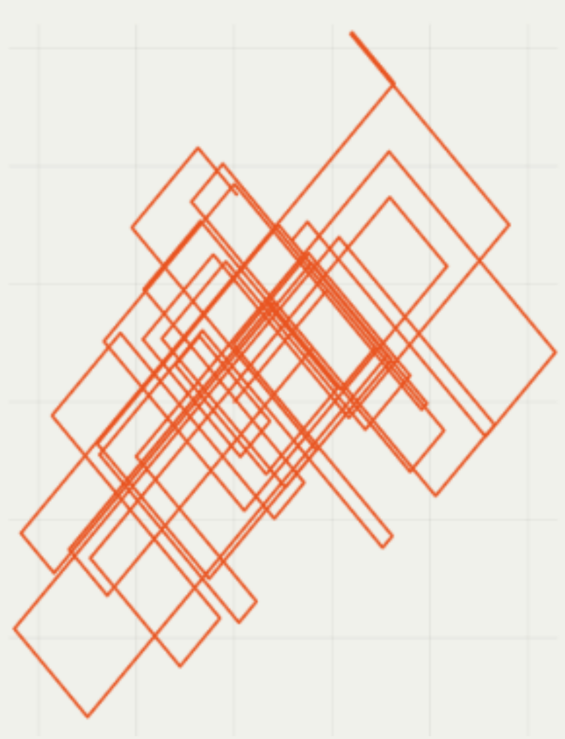
→ 元は連続時間だが, **離散時間**に帰着してしまう.

Langevin Diffusion



➡ アルゴリズムも変わる

Zig-Zag Sampler



PDMP の枠組み

① 棄却法
Poisson 剪定など

② 離散化(必要に応じて)
ほとんどの場合すごく簡単

→ 本質的には連続時間のまま
→ **連続時間 MCMC** とも呼ばれる

Poisson 剪定

(Lewis & Shedler, 1979)

$$\mathrm{P}[T_1^{(i)} \ge s] = \exp\left(-\int_0^s \underbrace{\left(v^{(i-1)}\cdot\nabla U(x^{(i-1)}+sv^{(i-1)})\right)_+}_{=:m(s)} ds\right)$$ という生存関数を持つ $T_1^{(i)}$ のシミュレーション法

＊ $U(x) = -\log\pi(x)$ はエネルギーともいう

$m \le M$ を満たす上界 M があれば……

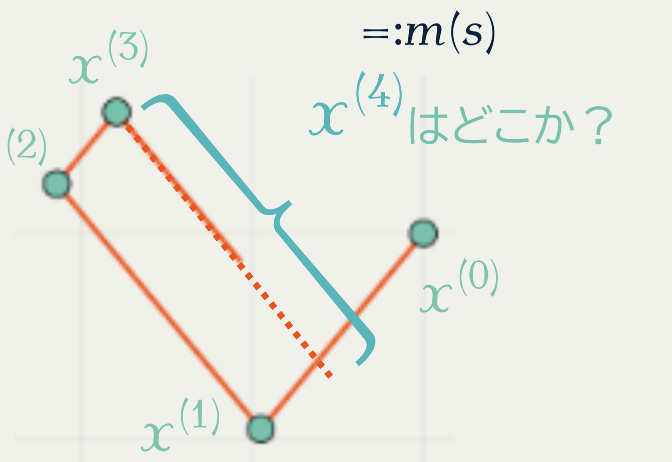
$x^{(i-1)}$　T_1　T_2　T_3　T_4　$\sim \mathrm{Pois}(M)$

$T_0 = 0$

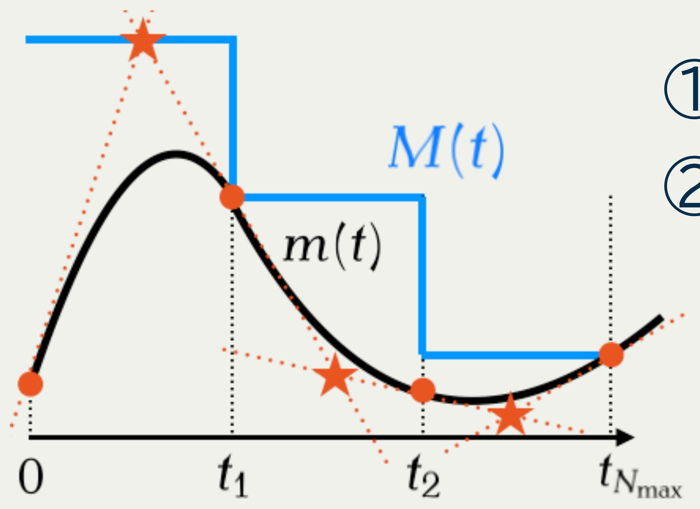
$\frac{m(T_1)}{M(T_1)}, \frac{m(T_2)}{M(T_2)}, \frac{m(T_3)}{M(T_3)}, \cdots$ の確率で採択／棄却

$x^{(i-1)}$　$\cancel{T_1}$　T_2　$\cancel{T_3}$　T_4

$T_0 = 0$　T_1　T_2　$\sim \mathrm{Pois}(m)$

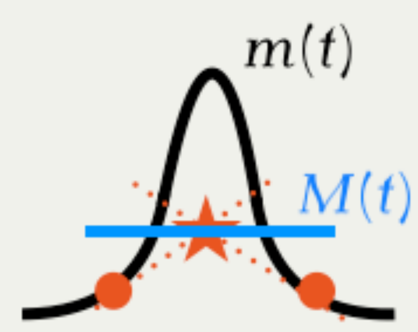


上界 M の発見をどう自動化するか？



① グリッド上で自動微分
② 2端点と交点の最大値
$M(t) := \max\{●●★\}$
でグリッド毎に定める

＊失敗する例



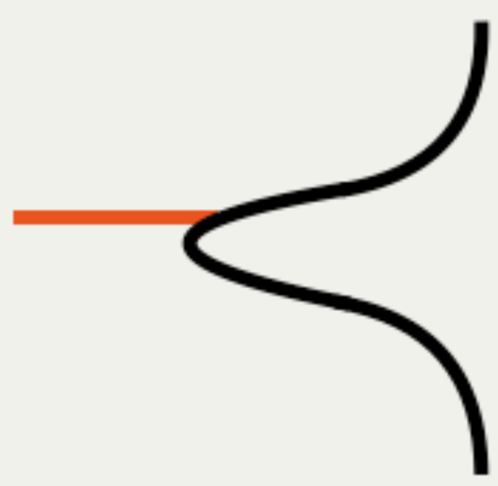
非絶対連続分布からのサンプリング

PDMP ＋ Stick テクニックで対応可能

Spike-and-slab prior (Mitchell & Beauchamp, 1988)

$$\mathrm{prior}(dx) = \prod_{i=1}^d \Big(\omega_i\pi_i(x_i)\,dx_i + (1-\omega_i)\delta_0(dx_i)\Big)$$

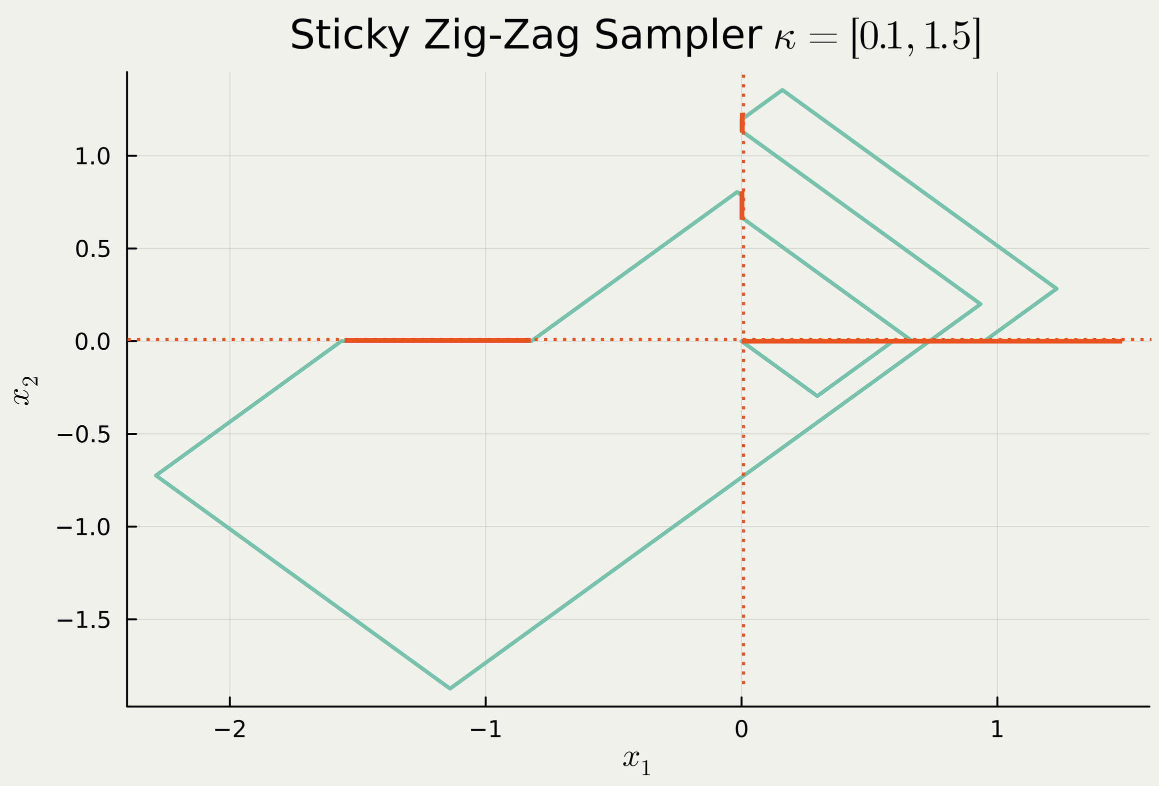
という形の事前分布. 変数選択に用いる.



→ 事後分布も **スパイク付き非絶対連続分布** になる

$$\mathrm{posterior}(dx) \propto Ce^{-U(x)}\prod_{i=1}^d\Big(\widetilde{\pi}_i(x_i)\,dx_i + \kappa_i^{-1}\delta_0(dx_i)\Big)$$

Sticky Zig-Zag Sampler $\kappa = [0.1, 1.5]$



＊発表者のパッケージ `PDMPFlux.jl` からの出力

$x_i = 0$ となった瞬間, i 番目の変数を凍結する

＝ 軸にぶつかるたびに, 軸にくっつく

➡ 脱出するまでの時刻 $\tau \sim \mathrm{Exp}(\kappa_i)$ を計算

募集　変数選択以外の非絶対連続分布, ないですか？