

マルコフ連鎖モンテカルロ法によるサロゲートデータ生成

伊庭 幸人 モデリング研究系 准教授

「モンテカルロ法」とは乱数を用いて、さまざまな量を計算する手法の総称です。計算機の中で擬似的にサイコロを投げることで複雑な計算を行うことができます。なかでも、統計物理に起源を持つ「マルコフ連鎖モンテカルロ法」(MCMC)は、高次元の確率分布を扱うための強力な道具です。データ解析、珍しい事象のサンプリング、組み合わせ問題、非線形科学、統計物理など、さまざまな分野でMCMCの新しい応用を探求しています。なかでも、困難な問題で力を発揮するMCMCの技法「レプリカ交換モンテカルロ法」「マルチカノニカル法」の応用に特に力を入れています。



■ マルコフ連鎖モンテカルロ法によるサロゲートデータ生成

【サロゲーション】

データのうち注目する統計量のみが(近似的に)等しくなるような仮想データを生成し帰無仮説とみなす
→別の統計量を検定統計量としてもとのデータを検定

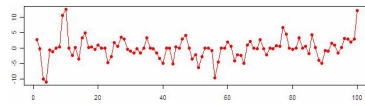
従来法1 問題に応じた工夫
→ 複雑になると手に負えない
従来法2 アニール(最適化)
→ 複数のサンプルを公平に作る手法でない
提案法 マルチカノニカルMCMCを利用
→ 多数のサンプルを公平に生成できるはず

【具体例】非線形時系列

データの順番を入れ替える
(⇒各時点での周辺分布を変えない)
2次の相関が(近似的に)一致
するような「仮想データ」を作る

→ これを帰無仮説として検定などに使う

$\{x^{orig}(t)\}$ もとの時系列 (長さ N)



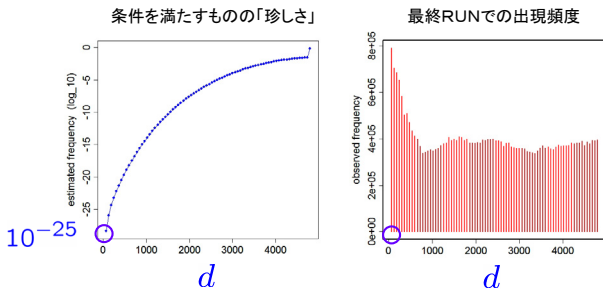
時間順序変更

$$C^{orig}(\tau) = \frac{1}{N} \sum_{t=1}^{N-\tau} x_t^{orig} x_{t+\tau}^{orig} \quad C(\tau, \{x_t\}) = \frac{1}{N} \sum_{t=1}^{N-\tau} x_t x_{t+\tau}$$

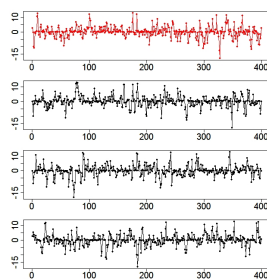
$$d(\{x_t\}) = \sum_{\tau=1}^T |C^{orig}(\tau) - C(\tau, \{x_t\})|$$

以下では $T = 8$

【マルチカノニカルMCMCを使う】

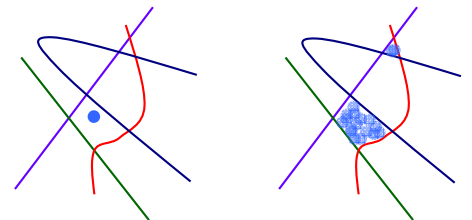


結果(一番上はもとの時系列)



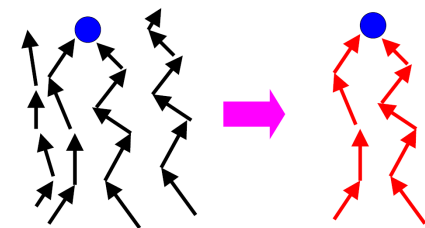
1個だけ選ぶなら
アニール(最適化)

制約を満たす範囲から
満遍なくサンプリング



■ そのほか進行中・企画中の共同研究など

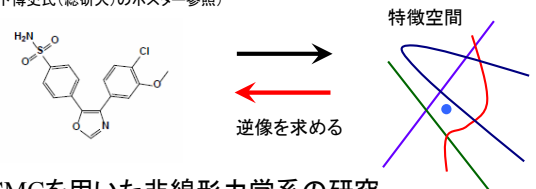
■ 気象モデルなどの大自由度非線形系におけるMCMCによるレアイベント生成と確率評価



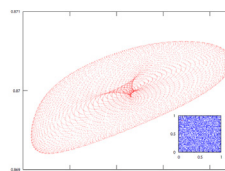
ターゲット●(レアイベント)に到達する経路をサンプリング

■ MCMCを用いたプレイメージ生成と創薬への応用

(山下博史氏(総研大)のポスター参照)



■ MCMCを用いた非線形力学系の研究



マルチカノニカル法による初期値探索で見つかったカオスの海に棲む微小なトラス