

Z120a MCMC 法における尤度計算量の削減手法

服部公平 (統計数理研究所)

モデルパラメータを MCMC 法によって推定する場合、モデルの尤度を多数回計算する必要がある。モデルの尤度の計算コストが大きい場合、この尤度計算は MCMC 法を用いたベイズ推定のボトルネックとなる。そしてこのボトルネックは、天体数 N が飛躍的に増大する天文学のビッグデータ時代には大きな障壁となる可能性がある。例えば、独立した N 個の天体のデータを用いてモデルの対数尤度を計算する場合を考える。この場合、モデルの対数尤度 $\ln L = \sum_i \ln L_i$ は個々の天体データからの対数尤度の寄与 $\ln L_i$ を足し合わせたものとなるため、天体数 N が増大すれば、それに応じてモデルの対数尤度 $\ln L$ の計算コストも増加する。さらに厄介な問題は、 $\ln L_i$ に数値誤差 ϵ が付随している場合、 $\ln L$ には $\epsilon \times N$ の誤差が付随することである。 N が小さい場合は実用上問題にならないが、 N が非常に大きい場合には $\epsilon \times N$ が $O(1)$ 以上に大きくなってしまい、正しくベイズ推定を行えなくなってしまう。この問題を避けるためには ϵ を $1/N$ よりも十分小さくする必要があるが、 N が大きいほど ϵ を小さくする必要があり、1 天体あたりの計算コストが増大する。その結果、天体数 N の増加は MCMC の計算コストを二重に増大させてしまう。本講演では、この問題の解決策として尤度の精度と尤度の計算コストのバランスをとる手法を紹介する。また、尤度の数値誤差の影響で MCMC 鎖が局所解にトラップされないための工夫として、尤度を 2 通りの方法で近似計算し、尤度が小さい方を尤度として採用するという手法についても紹介する。これらの手法の応用例として、Gaia DR2 で得られたハロー星の運動データを用いて銀河系恒星ハローの分布関数モデル (力学平衡モデル) と銀河系の暗黒物質の 3 次元密度分布を同時に推定する問題を取り扱う。