

11.5.2 ハイブリッドモンテカルロアルゴリズム

前節ではランダムウォークの振る舞いを低減するために、ハミルトン力学の枠組みを取り入れるサンプリング手法を扱った。この手法では、リープフロッグ離散化とよばれる時間積分の近似を行うことで、一定時間後の位置変数と運動量変数のサンプリングを行う。ただし、離散化による近似は積分の誤差を伴い、得られるサンプルにバイアスがかかるという問題がある。

本節で扱うハイブリッドモンテカルロアルゴリズムは、リープフロッグアルゴリズムと Metropolis アルゴリズムを組み合わせることで、バイアスを取り除く手法である。リープフロッグアルゴリズムで得られた状態候補を Metropolis 規準に従って受理または棄却することで、詳細釣り合い条件が成り立つことを示す。

おさらいとして、リープフロッグ離散化は時間 τ 後のサンプルを得るために、ステップサイズ ε を設定して位置変数と運動量変数を次の式で順に更新する方法であった。

$$\hat{r}_i(\tau + \varepsilon/2) = \hat{r}_i(\tau) - \frac{\varepsilon}{2} \frac{\partial E}{\partial z_i}(\hat{z}(\tau)) \quad (11.64)$$

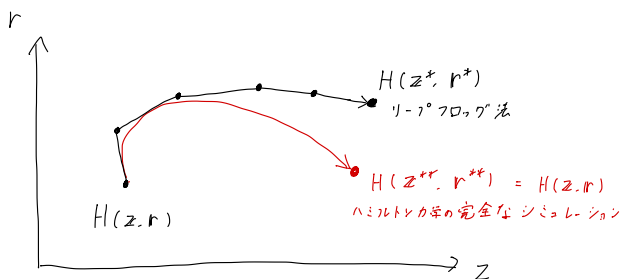
$$\hat{z}_i(\tau + \varepsilon) = \hat{z}_i(\tau) + \varepsilon \hat{r}_i(\tau + \varepsilon/2) \quad (11.65)$$

$$\hat{r}_i(\tau + \varepsilon) = \hat{r}_i(\tau + \varepsilon/2) - \frac{\varepsilon}{2} \frac{\partial E}{\partial z_i}(\hat{z}(\tau + \varepsilon)) \quad (11.66)$$

リープフロッグ法の初期状態を (z, r) とし、積分後の状態を (z^*, r^*) とするとき、ハイブリッドモンテカルロアルゴリズムにおける Metropolis 規準は次のように定められる。

$$\min(1, \exp\{H(z, r) - H(z^*, r^*)\}) \quad (11.67)$$

もしリープフロッグ積分がハミルトン力学を完全にシミュレーションするとしたら、 H の値は不変なのですべての候補は常に受理される。



上記の Metropolis 規準によって得られるサンプルが真の分布に従うことを確かめるために、詳細釣り合い条件を満たすことを示す。証明の方針としては、位相空間の微小領域 R から R' への遷移の釣り合い条件を考える。その上で、リープフロッグが R' から R への遷移も可逆であることを示し、また、遷移により体積が変化しないことを示す。これらを用いて詳細釣り合い条件の式が成り立つことを示す。

まず、リープフロッグ積分を開始する前に、時間が進む方向（ステップサイズ ε ）に積分するか、時間が戻る方向（ステップサイズ $-\varepsilon$ ）に積分するかを、同一確率でランダムに選択する。リープフロッグ積分スキーム (11.64)、(11.65) および (11.66) は時間の増加分 ε に対して可逆であるため遷移確率 \mathcal{P} を用いて次の式が成り立つ。

$$\mathcal{P}(R' | R) = \frac{1}{2} \quad (\text{ステップサイズ } \varepsilon \text{ が選ばれる確率})$$

$$\mathcal{P}(R | R') = \frac{1}{2} \quad (\text{ステップサイズ } -\varepsilon \text{ が選ばれる確率})$$

次に、リープフロッグ法により領域の体積が変化しないことを示す。リープフロッグスキームの各ステップが、 z_i 変数もしくは r_i 変数のどちらかを、もう一方の変数に依存する量だけ加算的に更新することから導かれる。このときの R 、 R' の体積を δV と表現し、次の図を用いて説明する。

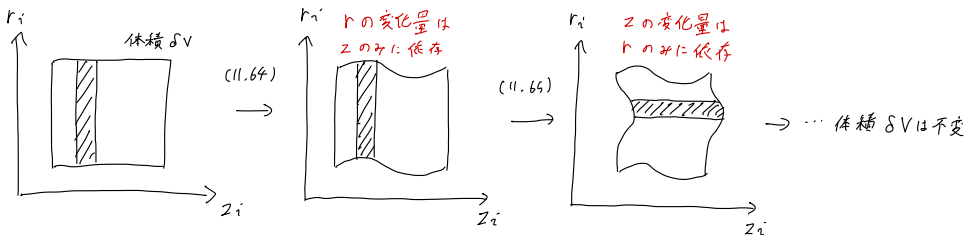


図 11.14

以上の結果を用いて、詳細釣り合い条件 (11.40) が満たされることを示す。

$$\begin{aligned}
 (\text{L.H.S. of (11.40)}) &= P(R) T(R, R') \\
 &= P(R) g(R' | R) \cdot A(R', R) \\
 &= \frac{1}{Z_H} \exp(-H(R)) \delta V \cdot \frac{1}{2} \cdot \min \left\{ 1, \exp(H(R) - H(R')) \right\}
 \end{aligned}$$

同様に、

$$(\text{R.H.S. of (11.40)}) = \frac{1}{Z_H} \exp(-H(R')) \delta V \cdot \frac{1}{2} \cdot \min \left\{ 1, \exp(H(R') - H(R)) \right\}$$

これらが等号で結ばれることを示す。

i) $H(R) = H(R')$ のとき、

$$\begin{aligned}
 (\text{L.H.S. of (11.40)}) &= \frac{1}{Z_H} \exp(-H(R)) \cdot \delta V \cdot \frac{1}{2} \\
 &= \frac{1}{Z_H} \exp(-H(R')) \cdot \delta V \cdot \frac{1}{2} \\
 &= (\text{R.H.S. of (11.40)})
 \end{aligned}$$

ii) $H(R) < H(R')$ のとき、

$$\begin{aligned}
 (\text{L.H.S. of (11.40)}) &= \frac{1}{Z_H} \exp(-H(R)) \cdot \delta V \cdot \frac{1}{2} \cdot \exp\{H(R) - H(R')\} \\
 &= \frac{1}{Z_H} \exp(-H(R)) \exp(H(R)) \exp(-H(R')) \delta V \cdot \frac{1}{2} \\
 &= \frac{1}{Z_H} \exp(-H(R')) \cdot \delta V \cdot \frac{1}{2} \cdot 1 \\
 &= \frac{1}{Z_H} \exp(-H(R')) \cdot \delta V \cdot \frac{1}{2} \min \left\{ 1, \exp(H(R') - H(R)) \right\} \\
 &= (\text{R.H.S. of (11.40)})
 \end{aligned}$$

iii) $H(R) > H(R')$ のとき、

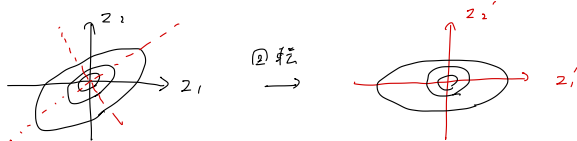
$H(R) < H(R')$ のときと同様に等号が成立立つ。

269頁の残りでは、多変量ガウス分布へ適用することでハイブリッドモンテカルロアルゴリズムの優れた点を考察する。具体的には、Metropolis アルゴリズムではランダムウォーク的な振る舞いにより、独立なサンプルを得るまでのステップが $(\sigma_{\max} / \sigma_{\min})^2$ のオーダーであったのに対して、ハイブリッドモンテカルロアルゴリズムでは $\sigma_{\max} / \sigma_{\min}$ のオーダーで済む。(読み合わせとします)

次の点がいまいちわからなかったので、確認したい。

「ハイブリッドモンテカルロアルゴリズムは回転等方性を持つので、我々の結論は相関がある要素を持つガウス分布においても同様に当てはまる」

> これは、 z_i 要素の相関が大きい場合に、相関が大きい方向に軸を持つように座標を回転させれば同じ議論が成り立つ、という認識であってそう？



「これを実現するには、リープフロッグ積分を $\sigma_{\max} / \sigma_{\min}$ のオーダーの繰り返し回数だけ続ける必要がある。」

> この理由については特に述べられていないよね？

11.6 分配関数の推定

これまで扱ってきた多くのサンプリングアルゴリズムでは、正規化係数を除いた分布の関数形だけを必要としていた。よって、

$$P_E(z) = \frac{1}{Z_E} \exp(-E(z))$$

と書いたとき、分配関数 Z_E はサンプリングにおいて不要であった。

しかし、ベイズモデルの比較においては分配関数がモデルエビデンスそのものを表すため、異なるモデル間の分配関数の比は重要な役割を持つ。ベイズモデル比較については上巻161頁を参照する。

$$\text{モデルエビデンス} : P(D|M_i) = \int P(D|w, M_i) p(w, M_i) dw \quad (3.68)$$

$$\text{事後分布} : P(w|D, M_i) = \frac{P(D|w, M_i) P(w|M_i)}{P(D|M_i)} \quad (3.69)$$

(
モデルエビデンスが正規化係数 (= 分配関数) となっている。

2つのモデルの分配関数の比率について考える。比率を求めるために、それぞれのベイズモデルを当てはめたときの事後分布のエネルギー関数を導入する。2つベイズモデル M_E, M_G をそれぞれ当てはめた結果得られた事後分布 $P_E(z), P_G(z)$ のエネルギー関数を $E(z), G(z)$ とする。分配関数の比は次のように求まる。

$$\begin{aligned} \frac{Z_H}{Z_G} &= \frac{\sum_z \exp(-E(z))}{\sum_z \exp(-G(z))} \\ &= \frac{\sum_z \exp(-E(z) + G(z)) \exp(-G(z))}{\sum_z \exp(-G(z))} \\ &= \sum_z \exp(-E(z) + G(z)) \cdot p_G(z) \\ &= E_{G(z)} [\exp(-E + G)] \\ &\simeq \frac{1}{L} \sum_j \exp \{ -E(z^{(j)}) + G(z^{(j)}) \} \end{aligned}$$

ここで、 $\{z^{(l)}\}$ は $p_q(z)$ で定義される分布から抽出されたサンプルである。もし分布 $p_q(z)$ がガウス分布のように分配関数が解析的に計算できるものであれば、 Z_E そのものの値を得ることができる。なぜなら、(11.72) により分配関数の比率を近似することができ、さらに Z_q の値も解析的に求まるためである。

この近似は、 $p_q(z)$ から得たサンプルを用いて分配関数の比を重点サンプリングしていることに他ならない。したがって、比率 $p_E(z) / p_q(z)$ が大きな変動をしない場合に精度のよい結果をもたらす。しかし、このような重点サンプリングに適した分布は簡単には見つからない。

そこで、マルコフ連鎖から得られるサンプルを用いて重点サンプリングを行うことができる。少し気になったのは、元々2つベイズモデル M_E, M_q を比較することが目的であったのに、2つの事後分布同士が近いことが求められていることである。ベイズモデル比較ではどのようなモデル同士でも比較できることが求められるはずなのに、 $p_q(z)$ は $p_E(z)$ に近いように選ぶべきであるという主張がよくわからなかった。また、ベイズモデルが異なるとパラメータ数も変わるので、 M_E と M_q で z の次元も変わりうることも考えられるが今回は無視されているのも気になる。サンプル集合が $z^{(1)}, \dots, z^{(L)}$ で与えられるとすると、サンプリング分布は

$$\frac{1}{Z_q} \exp(-G(z)) = \frac{1}{L} \sum_{l=1}^L T(z^{(l)}, z) \quad (11.73)$$

のように書き表すことができ、これを直接 (11.72) に用いることができる。ここが何を言っているかわからなかった。マルコフ連鎖から得られたサンプルを使ってそのまま (11.72) に代入するのではダメなのか。

残りの連鎖に関する部分は読み合わせにします。むずすぎ。