目次

1 Metropolis-Hastings(MH) 法

C を確率変数の集合とする。Markov 過程 $P(C \to C')$ が、詳細釣り合い条件

$$e^{-S(C)}P(C \to C') = e^{-S(C')}P(C' \to C) \tag{1}$$

を満たすとき、十分な回数の遷移を経た後、$P(C \to C')$ は

$$P_{\rm eq}(C) = \frac{e^{-S(C)}}{\sum_C e^{-S(C)}} \tag{2}$$

に収束する。これを不変分布と呼ぶ。

サンプリングを行いたい確率分布を目標分布 (target distribution) という。我々が興味のある物理の問題の場合、この目標分布とは Boltzmann 分布 e^{-S} である。MH 法は、提案分布 (proposal distribution) とよばれる確率分布をもちいて目標分布からのサンプリングを行う手法である。目標分布を $\pi(C)$、提案分布（遷移確率を与える条件付き確率分布）を $q(C \to C')$ とする。MH 法の具体的なアルゴリズムは以下で与えられる。

1. 初期値 $C^{(0)}$ を決める。
2. $C^{(n)}(n \geq 0)$ が与えられているとする。遷移確率 $q(C \to C')$ に従って、$C^{(n)}$ から C' を発生させる。
3. r を一様分布 $U(0,1)$ に従って発生させる。$r < \min\left\{1, \frac{\pi(C')q(C\to C')}{\pi(C)q(C'\to C)}\right\}$ のとき、$C^{(n+1)} = C'$ とし（採択)、それ以外のとき、$C^{(n+1)} = C^{(n)}$ とする（棄却)。
4. ステップ 2 へ戻る。

MH 法の中でも提案分布が $q(C \to C') = q(C' \to C)$ を満たすようなものを用いるものが、もともとの Metropolis 法である。MH 法の実効性は、提案分布をどのように設定するかに強く依存する。現在広く用いられている方法は、分子動力学を用いて詳細釣り合いを満たす確率過程を実現する、Hybrid Monte Carlo 法と呼ばれる手法である。

2 Hybrid Monte Carlo(HMC) 法

ϕ を力学変数、$S(\phi)$ を作用とする。また、π を ϕ に共役な運動量とする。このとき、Hamiltonian

$$H[\phi,\pi] = \frac{1}{2}\pi^2 + S(\phi) \tag{3}$$

によって決まる次の力学系

$$\dot{\phi}(t) = \frac{\partial H}{\partial \pi(t)} = \pi(t) \tag{4}$$

$$\dot{\pi}(t) = -\frac{\partial H}{\partial \phi(t)} = -\frac{\partial S}{\partial \phi(t)} \tag{5}$$

を分子動力学という。これを Leep-frog 法（1 次の差分スキーム）で差分化すると次のようになる。$t = N\epsilon$ とおき、$\phi(n\epsilon)$ のことを単に $\phi(n)$ などと書くことにする。ϕ, π の初期条件をそれぞれ、$\phi_{\mathrm{ini}}, \pi_{\mathrm{ini}}$ とする。

$$\phi(0) = \phi_{\mathrm{ini}} \tag{6}$$

$$\pi(0) = \pi_{\mathrm{ini}} \tag{7}$$

ただし、π_{ini} は確率分布 $e^{-\pi^2/2}$ に従って生成する。次に π を 1 ステップだけ進める。

$$\pi(1/2) = \pi(0) - \frac{1}{2}\epsilon\frac{\partial S}{\partial \phi(0)} \tag{8}$$

ステップ $n = 0, 1, \ldots, N-2$ は次のようにする。

$$\phi(n+1) = \phi(n) + \epsilon\pi(n+1/2) \tag{9}$$

$$\pi(n+3/2) = \pi(n+1/2) - \epsilon\frac{\partial S}{\partial \phi(n+1)} \tag{10}$$

最終ステップは次で与える。

$$\phi(N) = \phi(N-1) + \epsilon\pi(N-1/2) \tag{11}$$

$$\pi(N) = \pi(N-1/2) - \frac{1}{2}\epsilon\frac{\partial S}{\partial \phi(N)} \tag{12}$$

この過程を時間 ｔ＝1 だけ進めて得られた配位 $\phi(N), \pi(N)$ を Metropolis テストにかける。すなわち、r を一様分布 $U(0,1)$ に従う確率変数として、

$$r < \min\left\{1, e^{-\Delta H}\right\} \tag{13}$$

$$\Delta H = H(\phi(N), \pi(N)) - H(\phi(0), \pi(0)) \tag{14}$$

ならば、$\phi(N)$ を採択する。この確率過程は詳細釣り合い条件をみたすことが知られている。

3　誤差解析

平均 μ、分散 σ^2 の母集団から、独立に N 個のデータ (標本)$O_k(k = 1, \ldots, N)$ をサンプリングしたとする。標本平均を

$$\bar{O} = \frac{1}{N}\sum_{i=1}^{N} O_k \tag{15}$$

としたとき、大数の法則より、

$$E(\bar{O}) = \mu, \quad V(\bar{O}) = \frac{\sigma^2}{N} \tag{16}$$

が成り立つ。

観測量 O の独立な測定データ $O_k(k=1,\ldots,N)$ があったとする。N が十分大きいなら、平均と誤差は次のように推定される。

平均値（標本平均）

$$\langle O\rangle = \frac{1}{N}\sum_{k=1}^{N} O_k \tag{17}$$

不偏標準偏差

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{k=1}^{N}(O_k - \langle O\rangle)} \tag{18}$$

誤差（標準誤差）

$$\delta\langle O\rangle = \frac{1}{\sqrt{N}}\sigma = \sqrt{\frac{\langle O^2\rangle - \langle O\rangle^2}{N-1}} \tag{19}$$

4　誤差伝播

一般的な誤差伝播の計算方法

$$\delta f(X,Y) = \sqrt{\left(\frac{\partial f}{\partial X}\right)^2(\delta X)^2 + \left(\frac{\partial f}{\partial Y}\right)^2(\delta Y)^2} \tag{20}$$

例 (1): $f = X+Y$ のとき、

$$\delta f(X,Y) = \sqrt{(\delta X)^2 + (\delta Y)^2} \tag{21}$$

例 (2): $f = X/Y$ のとき、

$$\delta f(X,Y) = \sqrt{\frac{1}{Y^2}(\delta X)^2 + \frac{X^2}{Y^4}(\delta Y)^2} = \sqrt{\frac{Y^2(\delta X)^2 + X^2(\delta Y)^2}{Y^4}} \tag{22}$$

例 (3): $f = XY$ のとき、

$$\delta f(X,Y) = \sqrt{Y(\delta X)^2 + X(\delta Y)^2} \tag{23}$$

ただし、X,Y に相関がある場合、この公式に従って誤差を計算すると過大評価してしまう。そのようなときは、jackknife や bootstrap を用いるべきである。

5　Jackknife 法

データの集合 $X=\{X_1,\ldots,X_N\}$ があったとする。まず bin サイズ 1 の Jackknife 法 (delete-1 jackknife) について述べる。X から i 番目のデータを取り除いたものを考え、これを Jackknife sample vector という。

$$X_{(i)} = \{X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_N\} \tag{24}$$

$\hat{\theta}(X)$ を推定関数 (estimator) とする。これにフルのデータ・セットを入れる代わりに、Jackknife sample を代入したものを考える。

$$\hat{\theta}_{(i)} = \hat{\theta}(X_{(i)}) \tag{25}$$

また、この平均を

$$\hat{\theta}_{(\cdot)} = \frac{1}{N}\sum_{i=1}^{N}\hat{\theta}_{(i)} \tag{26}$$

で表す。これらを用いて、jackknife estimate と誤差は、

$$\hat{\theta}_{\text{jack}} = \hat{\theta} - \widehat{bias} = N\hat{\theta} - (N-1)\hat{\theta}_{(\cdot)} \tag{27}$$

$$\delta\hat{\theta}_{\text{jack}} = \sqrt{\frac{N-1}{N}\sum_{i=1}^{N}(\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)})^2} \tag{28}$$

で与えられる。ここで、

$$\widehat{bias} = (N-1)\left(\hat{\theta}_{(\cdot)} - \hat{\theta}\right) \tag{29}$$

を jackknife bias という。

評価関数が標本平均のとき、$\hat{\theta}_{(\cdot)}$ は標本平均に、Jackknife 誤差は標準誤差に一致することが示せる。実際、

$$\hat{\theta}_{(i)} = \frac{1}{N-1}\sum_{k\neq i}x_k = \frac{N\bar{x} - x_i}{N-1} \tag{30}$$

$$\hat{\theta}_{(\cdot)} = \frac{1}{N}\sum_{k=1}^{N}\frac{N\bar{x} - x_i}{N-1} = \frac{N\bar{x} - \bar{x}}{N-1} = \bar{x} \tag{31}$$

$$\bar{x} = \frac{1}{N}\sum_{k=1}^{N}x_k \tag{32}$$

だから、

$$\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)} = \frac{(N\bar{x} - x_i) - (N\bar{x} - \bar{x})}{N-1} = \frac{\bar{x} - x_i}{N-1} \tag{33}$$

となるので、

$$\delta\hat{\theta} = \sqrt{\frac{N-1}{N}\sum_{i=1}^{N}\left(\frac{\bar{x} - x_i}{N-1}\right)^2} = \sqrt{\frac{1}{N-1}\frac{1}{N}\sum_{i=1}^{N}(\bar{x} - x_i)^2} = \sqrt{\frac{\bar{x^2} - \bar{x}^2}{N-1}} \tag{34}$$

を得る。これは確かに標準誤差である。また、このときの bias はゼロである。

delete-d jackknife では、X から d 個のデータを削除したサンプルを用いる。d 個削除するやりかたは、$\begin{pmatrix} N \\ d \end{pmatrix}$ 通りある。この時の誤差の評価は

$$\delta\hat{\theta} = \sqrt{\frac{N-d}{d\begin{pmatrix} N \\ d \end{pmatrix}}\sum_{z}(\hat{\theta}_{(z)} - \hat{\theta}_{(\cdot)})^2} \tag{35}$$

で与えられる。d は $\sqrt{n} < d < n-1$ となるように選ぶ。

X が時系列データの場合は、次のようにする。N 個のデータをサイズが k の M 個のブロックに分割する。$N = Mk$。もし N が割り切れない場合は、割り切れるように余分なデータを適宜捨てて N を再定義することにする。このようにして、各データ X_i にブロック番号を付与する。例えば j 番目のブロックに属すならば $X_i^{(k)}$ と書く。

$$X = \{X_1^{(1)}, \dots, X_k^{(1)}, X_{k+1}^{(2)}, \dots, X_{2k}^{(2)}, \dots, X_{(M-1)k+1}^{(M)}, \dots, X_{Mk}^{(M)}\} \tag{36}$$

jackknife sample は、ブロックをひとつ取り除いたものとして定める。

$$X_{(i)} = \{X_1^{(1)}, \dots, X_k^{(1)}, \dots, X_{\dots}^{(i-1)}, \dots, X_{\dots}^{(i-1)}, X_{\dots}^{(i+1)}, \dots, X_{\dots}^{(i+1)}, \dots, X_{(M-1)k+1}^{(M)}, \dots, X_{Mk}^{(M)}\} \tag{37}$$

あとは同様である。

$$\hat{\theta}_{(i)} = \hat{\theta}(X_{(i)}) \tag{38}$$

$$\hat{\theta}_{(\cdot)} = \frac{1}{M}\sum_{i=1}^{M} \hat{\theta}_{(i)} \tag{39}$$

estimate と誤差は

$$\hat{\theta}_{\text{jack}} = \hat{\theta} - \widehat{bias} \tag{40}$$

$$\delta\hat{\theta}_{\text{jack}} = \sqrt{\frac{M-1}{M}\sum_{i=1}^{M}(\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)})^2} \tag{41}$$

$$\widehat{bias} = (M-1)\left(\hat{\theta}_{(\cdot)} - \hat{\theta}\right) \tag{42}$$

で与えられる。ブロックサイズ k が、データの相関長 τ より十分大きくなるようにとれば、適切な評価を与える。（また逆に、k 依存性を調べれば相関長が分かる。）