

再重法 (Reweighting Technique)

山内 仁喬

2021 年 10 月 31 日

1 単ヒストグラム再重法 (Single-Histogram Reweighting Techniques)

単ヒストグラム最重法 [1, 2] を用いることで, あるパラメーターにおけるシミュレーショントラジェクトリーから, 他のパラメータにおける平均値を求めることができる. 例えば, 温度 T のシミュレーションから温度 T' における平均値を算出したり, マルチカノニカルシミュレーションの結果からある温度 T における物理量の平均値を求めることが可能となる.

1.1 一般的な定式化

ある重み $W(\lambda)$ に基づく物理系を考える. パラメータ λ は例えばポテンシャルエネルギーなどであり, カノニカルアンサンブルの場合, $W(\lambda)$ はボルツマン因子 $W_B(U) = e^{-U/k_B T} = e^{-\beta U}$ となる. パラメータ λ に関する系の状態密度 $n(\lambda)$ が分かっている場合, λ に関する確率分布と任意の物理量 $A(\lambda)$ の期待値は,

$$P(\lambda) = \frac{n(\lambda)W(\lambda)}{\int d\lambda n(\lambda)W(\lambda)} = \frac{1}{Z}n(\lambda)W(\lambda) \quad (1)$$

$$\langle A \rangle = \frac{\int d\lambda A(\lambda)n(\lambda)W(\lambda)}{\int d\lambda n(\lambda)W(\lambda)} = \frac{1}{Z} \int d\lambda A(\lambda)n(\lambda)W(\lambda) \quad (2)$$

と計算される. ここで, Z は分配関数である. 系の状態密度が分かっているならば, 確率分布や物理量の期待値を得ることができる. しかし, 一般には系の状態密度は厳密には分かっているため, 確率分布や物理量の期待値を計算するには分子シミュレーションを実行して状態密度を推定する必要がある.

分子シミュレーションを実行して, h 個のサンプルを得たとする. λ に関する確率分布は, 分子シミュレーションによって得られるパラメータ λ のヒストグラム $H(\lambda)$ を用いて,

$$P(\lambda) = H(\lambda)/h \quad (3)$$

と推定することができる. したがって, 式 (1) より, 状態密度は

$$n(\lambda) = Z \frac{P(\lambda)}{W(\lambda)} = Z \frac{H(\lambda)}{hW(\lambda)} \quad (4)$$

と推定できる. また, 物理量 A の期待値は, 式 (2) より

$$\langle A \rangle = \frac{1}{Z} \int d\lambda A(\lambda)n(\lambda)W(\lambda) \quad (5)$$

$$= \frac{1}{Z} \int d\lambda A(\lambda)Z \frac{P(\lambda)}{W(\lambda)}W(\lambda) \quad (6)$$

$$= \int d\lambda A(\lambda)H(\lambda)/h \quad (7)$$

と計算できる.

続いて, 重み $W(\lambda)$ に基づいた分子シミュレーションのトラジェクトリーを用いて, 重み $W'(\lambda)$ に基づいた統計アンサンブルの確率分布と物理量の期待値を推定することを考える. この時, 確率分布と物理量の平均値は,

$$P(\lambda) = \frac{n(\lambda)W'(\lambda)}{\int d\lambda n(\lambda)W'(\lambda)} = \frac{1}{Z} n(\lambda)W'(\lambda) \quad (8)$$

$$\langle A \rangle = \frac{\int d\lambda A(\lambda)n(\lambda)W'(\lambda)}{\int d\lambda n(\lambda)W'(\lambda)} = \frac{1}{Z} \int d\lambda A(\lambda)n(\lambda)W'(\lambda) \quad (9)$$

と表される. 重み $W(\lambda)$ に基づいた分子シミュレーションで得られた状態密度 (4) を代入すると, 確率分布は,

$$P(\lambda) = \frac{n(\lambda)W'(\lambda)}{\int d\lambda n(\lambda)W'(\lambda)} = \frac{Z \frac{P(\lambda)}{W(\lambda)} W'(\lambda)}{\int d\lambda Z \frac{P(\lambda)}{W(\lambda)} W'(\lambda)} \quad (10)$$

$$= \frac{\{H(\lambda)/W(\lambda)\} W'(\lambda)}{\int d\lambda \{H(\lambda)/W(\lambda)\} W'(\lambda)} \quad (11)$$

と推定することができる. 同様にして, 物理量の期待値は

$$\langle A \rangle = \frac{\int d\lambda A(\lambda) \{H(\lambda)/W(\lambda)\} W'(\lambda)}{\int d\lambda \{H(\lambda)/W(\lambda)\} W'(\lambda)} \quad (12)$$

と推定できる.

1.2 カノニカルシミュレーションの場合

温度 T におけるカノニカルシミュレーションを実行し, h 個のサンプルを得たとする. この時, ポテンシャルエネルギー U の確率分布は,

$$P_{NVT}(U; T) = \frac{n(U)e^{-\beta U}}{\int dU n(U)e^{-\beta U}} = \frac{n(U)e^{-\beta U}}{Z(T)} \quad (13)$$

であり, 物理量の期待値は

$$\langle A \rangle_T = \frac{\int dU A(U)n(U)e^{-\beta U}}{Z(T)} \quad (14)$$

である. ここで, $\beta = 1/k_B T$, k_B はボルツマン定数, $n(U)$ は状態密度, $Z(T)$ は分配関数である. 式 (13) より, 状態密度は

$$n(U) = Z(T)P_{NVT}(U; T)/e^{-\beta U} \quad (15)$$

とかける. 通常, 状態密度は事前に分からないことが多く, シミュレーションを実行することで, はじめて系の状態密度を推定することができる. 確率分布は, シミュレーションで得られたポテンシャル U のヒストグラム $H(U)$ に比例し,

$$P_{NVT}(U; T) = H(U)/h \quad (16)$$

と書くことができる. したがって, 状態密度は

$$n(U) = \frac{Z(T)H(U)}{h e^{-\beta U}} \quad (17)$$

と計算することができる. 一方, 温度 $T' = 1/k_B \beta'$ における確率分布は

$$P_{NVT}(U; T') = \frac{n(U)e^{-\beta' U}}{\int dU n(U)e^{-\beta' U}} \quad (18)$$

とかける。ここに温度 T のカノニカルシミュレーションで得られた状態密度の式 (17) を代入すると、

$$P_{NVT}(U; T') = \frac{\frac{Z(T)H(U)}{he^{-\beta U}} e^{-\beta' U}}{\int dU \frac{Z(T)H(U)}{he^{-\beta U}} e^{-\beta' U}} \quad (19)$$

$$= \frac{\{H(U)/e^{-\beta U}\} e^{-\beta' U}}{\int dU \{H(U)/e^{-\beta U}\} e^{-\beta' U}} \quad (20)$$

$$= \frac{H(U)e^{-U(\beta' - \beta)}}{\int dU H(U)e^{-U(\beta' - \beta)}} \quad (21)$$

を得る。また、物理量の期待値は

$$\langle A \rangle_{T'} = \frac{\int dU A(U)H(U)e^{-U(\beta' - \beta)}}{\int dU H(U)e^{-U(\beta' - \beta)}} \quad (22)$$

とかける。このように、温度 T のサンプルからの温度 T' における確率分布を推定式を得ることができる。一般的には、 T' が T 近傍の値の時は比較的精度良く推定することができ、 T と T' の値が離れるにつれて、推定された物理量の精度は悪くなってしまう。

2 多ヒストグラム再重法 (WHAM: Weighted Histogram Analysis Method)

2.1 WHAM 方程式の導出

WHAM を用いることで、 M 個のシミュレーションで得られたサンプルから、任意のパラメータ設定における確率分布 P についての最適な推定方法を得ることができる。ここでは、カノニカルシミュレーションの場合を考えていく。 m 番目のシミュレーションは温度 $T_m = 1/(k_B\beta_m)$ で実行され、そこで得られたサンプルの数を N_m と表すことにする。

2.1.1 相関のある時系列データの統計的不確かさの評価法

ここでは、相関のある時系列データの統計的不確かさについて簡単にまとめる [3, 4]。定常で時間可逆な確率過程からサンプルされた観測量 A について、データ間で相関のあるような連続した観測量の時系列 $\{A_i\}_{i=1}^N$ を考える。期待値 $\langle A \rangle$ は単純な算術平均で推定することができる：

$$\hat{A} = \bar{A} = \frac{1}{N} \sum_{i=1}^N A_i \quad (23)$$

上付きの $\hat{}$ で推定値であることを表す。また、上付きの $\bar{}$ は算術平均であることを表す。ここでは $\langle A \rangle$ と \bar{A} を区別して考える。 $\langle A \rangle$ は常数であり、何かある値に求められるようなものである。一方、 \bar{A} は期待値 $\langle A \rangle$ の推定値であるため、取りうる値は期待値周辺を揺らぐ。

ある物理量 A の推定値 \hat{A} について、統計的不確かさについて考える。統計的不確かさは

$$\begin{aligned} \delta^2 \hat{A} &\equiv \left\langle \left(\hat{A} - \langle \hat{A} \rangle \right)^2 \right\rangle = \langle \hat{A}^2 \rangle - \langle \hat{A} \rangle^2 \\ &= \frac{1}{N^2} \sum_{i,j=1}^n \langle A_i A_j \rangle - \frac{1}{N^2} \sum_{i,j=1}^n \langle A_i \rangle \langle A_j \rangle \end{aligned} \quad (24)$$

と計算される。対角部分と非対角部分に分割すると、

$$\delta^2 \hat{A} = \frac{1}{N^2} \sum_{i=1}^n (\langle A_i^2 \rangle - \langle A_i \rangle^2) + \frac{1}{N^2} \sum_{i \neq j=1}^n (\langle A_i A_j \rangle - \langle A_i \rangle \langle A_j \rangle) \quad (25)$$

となる。右辺第一項は観測量の分散に $1/N$ を乗じたものであり、右辺第二項は観測量の相関である。第二項について i と j の対称性を考えると、和は $\sum_{i \neq j} = 2 \sum_{i=1}^N \sum_{j=i+1}^N$ とすることができる。さらに $k = |i - j|$ を定めると、 i と j に関する和は k についての和に置き直すことができる：

$$\sum_{i \neq j=1}^N (\langle A_i A_j \rangle - \langle A_i \rangle \langle A_j \rangle) = 2 \sum_{i=1}^N \sum_{j=i+1}^N (\langle A_i A_j \rangle - \langle A_i \rangle \langle A_j \rangle) \quad (26)$$

$$= 2 \sum_{k=1}^{N-1} (\langle A_i A_{i+k} \rangle - \langle A_i \rangle \langle A_{i+k} \rangle) (N - k) \quad (27)$$

ここで $(N - k)$ は、ある k について $i < j$ の時、 $|i - j| = k$ となるような整数の組の数に由来する。以上より、統計的不確かさは

$$\delta^2 \hat{A} = \frac{1}{N} (\langle A_i^2 \rangle - \langle A_i \rangle^2) + \frac{2}{N} \sum_{k=1}^{N-1} \frac{N - k}{N} (\langle A_i A_{i+k} \rangle - \langle A_i \rangle \langle A_{i+k} \rangle) \quad (28)$$

$$\equiv \frac{\sigma_A^2}{N} (1 + 2\tau) \quad (29)$$

$$\equiv \frac{\sigma_A^2}{N/g} \quad (30)$$

と計算される。ここで、 σ_A^2 は分散、 g は統計的非効率性因子、 τ は自己相関時間であり、それぞれ

$$\sigma_A^2 \equiv \langle A_i^2 \rangle - \langle A_i \rangle^2 \quad (31)$$

$$\tau \equiv \sum_{k=1}^{N-1} \left(1 - \frac{k}{N}\right) C_k \quad (32)$$

$$g \equiv 1 + 2\tau \quad (33)$$

と定義した。また C_k は

$$C_k \equiv \frac{\langle A_i A_{i+k} \rangle - \langle A_i \rangle^2}{\langle A_i^2 \rangle - \langle A_i \rangle^2} \quad (34)$$

で定義される、規格化された自己相関関数である。統計的非効率性因子 $g = 1 + 2\tau$ は 1 以上の値をもつ。 N/g は時系列データに含まれる有効サンプル数（非相関データの数）を与える。 g の値はシミュレーションにおいてサンプルを取得する時間ステップ間隔に依存する。より長い時間ステップ間隔でサンプリングすれば、 g の値は 1 に近づく。

2.1.2 ヒストグラムの数学的定式化

指示関数 $\chi_l(A)$ を、観測量 A が中心値 A_l 、幅 ΔA のビンに含まれるかどうかを判定する関数として、次のように定義する：

$$\chi_l(A) = \begin{cases} 1 & \text{if } A \in [A_l - \frac{\Delta A}{2}, A_l + \frac{\Delta A}{2}) \\ 0 & \text{otherwise} \end{cases} \quad (35)$$

ここで、 l は A を離散化した時のビン番号を表す。また、 m 番目のシミュレーションの n 番目の観測量を A_{mn} とすると、指示関数の時系列は $\{\chi_l(A_{mn})\}_{n=1}^{N_m}$ と表せる。 N_m は m 番目のシミュレーションのサンプル数である。 m 番目のシミュレーションに対する観測量 A_k に関するヒストグラム H_m は

$$H_m(A_l) = \sum_{n=1}^{N_m} \chi_l(A_{mn}) \quad (36)$$

と計算され、全てのシミュレーションに渡って観測量をカウントしたヒストグラム H は

$$H(A_l) = \sum_{m=1}^M \sum_{n=1}^{N_m} \chi_l(A_{mn}) \quad (37)$$

と計算することができる。

2.1.3 ヒストグラムの統計的不確かさ

m 番目のシミュレーションのヒストグラムは、

$$H_m(A_l) = \sum_{n=1}^{N_m} \chi_l(A_{mn}) = N_m \frac{1}{N_m} \sum_{n=1}^{N_m} \chi_l(A_{mn}) = N_m \hat{\chi}_{lm} \quad (38)$$

と表すことができる。統計的不確かさの式 (30) より、

$$\delta^2 H_m(A_l) = \delta^2 \{N_m \hat{\chi}_{lm}\} = N_m^2 \delta^2 \hat{\chi}_{lm} = N_m^2 \frac{\sigma_\chi}{N_m/g} \quad (39)$$

と書き下せる。これを具体的に計算すると、

$$\delta^2 H_m(A_l) = \frac{g_m}{N_m} N_m^2 (\langle \chi_{lm}^2 \rangle - \langle \chi_{lm} \rangle^2) \quad (40)$$

$$= \frac{g_m}{N_m} N_m^2 (\langle \chi_{lm} \rangle - \langle \chi_{lm} \rangle^2) \quad (41)$$

$$= g_m N_m \langle \chi_{lm} \rangle (1 - \langle \chi_{lm} \rangle) \quad (42)$$

$$= g_m \langle H_m(A_l) \rangle \left(1 - \frac{\langle H_m(A_l) \rangle}{N_m}\right) \quad (43)$$

と計算される。ここで、 χ_{lm} は指示関数のため $[\chi_l(A_m)]^2 = \chi_l(A_m)$ であることを用いた。ヒストグラムの分布が疎であり、十分な数のピンを用いている場合、 $\langle H_m(A_l) \rangle / N_m \ll 1$ となるので、

$$\delta^2 H_m(A_l) \simeq g_m \langle H_m(A_l) \rangle \quad (44)$$

と近似することができる。

2.1.4 各シミュレーションの状態密度の推定

ここでは、 m 番目のシミュレーションに対する状態密度 $\Omega_m(U)$ の推定方法を考える。そのためには、確率密度分布を数学的に表現する方法が必要である。J. D. Chodera ら [4] が指摘しているように、確率密度分布の推定にノンパラメトリックを用いることもできるが、ここではより簡便な、Kumar[5] らや Ferrenberg と Swendsen[6] が提案したヒストグラムに基づいた確率分布の推定方法を説明する。

まず、カノニカルアンサンブルにおけるポテンシャルの確率分布は、

$$P_{NVT}(U; T_m) = \frac{\Omega_m(U) e^{-\beta_m U}}{\int dU \Omega_m(U) e^{-\beta_m U}} = \frac{1}{Z(\beta_m)} \Omega_m(U) e^{-\beta_m U} \quad (45)$$

$$= \Omega_m(U) \exp[f_m - \beta_m U] \quad (46)$$

と書くことができる。ここで、 $Z_m(\beta_m)$ は分配関数であり、ヘルムホルツの自由エネルギー F_m あるいは無次元化されたヘルムホルツの自由エネルギー f_m を用いて

$$Z(\beta_m) = e^{-\beta_m F_m} = e^{-f_m} \quad (47)$$

と書かれることを用いた。

続いて、シミュレーションで得られたサンプルから得られるヒストグラム H に基づいた確率分布の推定を考える。ヒストグラムを用いるとポテンシャルの確率分布は、

$$P_{NVT}(U_l; T_m) = \Omega_m(U) \exp[f_m - \beta_m U_l] \simeq \frac{1}{\Delta U} \frac{H_m(U_l)}{N_m} \quad (48)$$

のように見積もることができるため、状態密度をヒストグラムを用いて

$$\hat{\Omega}_m(U_l) = \frac{H_m(U_l)}{N_m \Delta U \exp[f_m - \beta_m U_l]} \quad (49)$$

と推定することができる。

2.1.5 状態密度の最良推定値 $\hat{\Omega}(U)$ の表現方法

状態密度 $\hat{\Omega}(U_l)$ の最良推定値を、 M 個のシミュレーションに対する状態密度の推定値 $\hat{\Omega}_m(U)$ ($m = 1, \dots, M$) の重み付きの和で表す：

$$\hat{\Omega}(U_l) = \sum_{m=1}^M w_m(U_l) \hat{\Omega}_m(U_l) \quad (50)$$

ここで、重み w_i は拘束条件

$$\sum_{m=1}^M w_m(U_l) = 1 \quad (51)$$

を満たす。WHAM では、状態密度 $\hat{\Omega}(U_l)$ の最良推定値を得るような重み w_i の集合は、状態密度の統計誤差 $\delta^2 \hat{\Omega}(U_l)$ を最小化することによって導出する。状態密度の統計誤差は具体的に、

$$\delta^2 \hat{\Omega}(U_l) = \left[\hat{\Omega}(U_l) - \langle \hat{\Omega}(U_l) \rangle \right]^2 \quad (52)$$

$$= \left[\sum_{m=1}^M w_m(U_l) \hat{\Omega}_m(U_l) - \left\langle \sum_{m=1}^M w_m(U_l) \hat{\Omega}_m(U_l) \right\rangle \right]^2 \quad (53)$$

$$= \sum_{m=1}^M w_m^2(U_l) \left[\hat{\Omega}_m(U_l) - \langle \hat{\Omega}_m(U_l) \rangle \right]^2 \quad (54)$$

$$= \sum_{m=1}^M w_m^2(U_l) \delta^2 \hat{\Omega}_m(U_l) \quad (55)$$

と計算される。各シミュレーションの状態密度 (49) より

$$\delta^2 \hat{\Omega}_m(U_l) = \frac{\delta^2 H_m(U_l)}{\{N_m \Delta U \exp[f_m - \beta_m U_l]\}^2} \quad (56)$$

であるので、

$$\delta^2 \hat{\Omega}(U_l) = \sum_{m=1}^M w_m^2(U_l) \frac{\delta^2 H_m(U_l)}{\{N_m \Delta U \exp[f_m - \beta_m U_l]\}^2} \quad (57)$$

と表すことができる。このように、 $\delta^2 \hat{\Omega}(U_l)$ は、 M 個の見積もられた状態密度 $\hat{\Omega}_1(U_l), \dots, \hat{\Omega}_M(U_l)$ の統計誤差 $\delta^2 \hat{\Omega}_1(U_l), \dots, \delta^2 \hat{\Omega}_M(U_l)$ に由来するので、結局はヒストグラムの統計誤差 $\delta^2 H_1(U_l), \dots, \delta^2 H_M(U_l)$ に依存する形で書かれる。式 (44) を用いると、ヒストグラムの統計誤差をヒストグラムの期待値 $\langle H_m(U_l) \rangle$ で近似できる。ヒストグラムの期待値は、得られるべき状態密度の最良推定値で置き直すと、

$$\langle H_m(U_l) \rangle = N_m \Delta U P_{NVT}(U_l; T_m) \quad (58)$$

$$\simeq N_m \Delta U \Omega(\hat{U}_l) \exp[f_m - \beta_m U_l] \quad (59)$$

となるので、式 (44) を用いてヒストグラムの統計誤差を

$$\delta^2 H_m(U_l) \simeq g_m \langle H_m(U_l) \rangle \simeq g_m N_m \Delta U \Omega(\hat{U}_l) \exp[f_m - \beta_m U_l] \quad (60)$$

と計算することができる。これを式 (56) に代入することで、各シミュレーションに対する状態密度の推定値の統計誤差 $\hat{\Omega}_m$ を

$$\delta^2 \Omega_m(U_l) = \frac{\delta^2 H_m(U_l)}{\{N_m \Delta U \exp[f_m - \beta_m U_l]\}^2} \quad (61)$$

$$= \frac{\Omega(\hat{U}_l)}{g_m^{-1} N_m \Delta U \exp[f_m - \beta_m U_l]} \quad (62)$$

と推定することができる。

2.1.6 重み w_i の決定: Lagrange の未定乗数法

状態密度の統計誤差 $\delta^2 \hat{\Omega}(U_l)$ が拘束条件 (51) の下で最小になるような重み w_i の集合を得るために、Lagrange の未定乗数法を用いる。ラグランジュ関数は

$$\mathcal{L} = \delta^2 \hat{\Omega}(U_l) - \alpha \left[\sum_{m=1}^M w_m(U_l) - 1 \right] = \left[\sum_{m=1}^M w_m^2(U_l) \delta^2 \hat{\Omega}_m(U_l) \right] - \alpha \left[\sum_{m=1}^M w_m(U_l) - 1 \right] \quad (63)$$

とかける。極値をとる条件 $\partial \mathcal{L} / \partial \alpha = 0$, $\partial \mathcal{L} / \partial w_m = 0$ より、

$$\frac{\partial \mathcal{L}}{\partial \alpha} = \sum_{m=1}^M w_m(U_l) - 1 = 0 \quad (64)$$

$$\frac{\partial \mathcal{L}}{\partial w_m} = 2w_m(U_l) \delta^2 \hat{\Omega}_m(U_l) - \alpha = 0 \quad (65)$$

を得る。式 (65) より、

$$\alpha = 2w_1(U_l) \delta^2 \hat{\Omega}_1(U_l) = 2w_2(U_l) \delta^2 \hat{\Omega}_2(U_l) = \dots = 2w_M(U_l) \delta^2 \hat{\Omega}_M(U_l) \quad (66)$$

であるので、これを拘束条件 (51) に代入すると

$$w_1(U_l) + \frac{\delta^2 \hat{\Omega}_1(U_l)}{\delta^2 \hat{\Omega}_2(U_l)} w_1(U_l) + \frac{\delta^2 \hat{\Omega}_1(U_l)}{\delta^2 \hat{\Omega}_3(U_l)} w_1(U_l) + \dots + \frac{\delta^2 \hat{\Omega}_1(U_l)}{\delta^2 \hat{\Omega}_M(U_l)} w_1(U_l) = 1 \quad (67)$$

を得る。これより、

$$w_1(U_l) = \frac{1}{\left\{ \frac{1}{\delta^2 \hat{\Omega}_1(U_l)} + \frac{1}{\delta^2 \hat{\Omega}_2(U_l)} + \dots + \frac{1}{\delta^2 \hat{\Omega}_M(U_l)} \right\} \delta^2 \hat{\Omega}_1(U_l)} = \frac{\frac{1}{\delta^2 \hat{\Omega}_1(U_l)}}{\left\{ \sum_{m=1}^M \frac{1}{\delta^2 \hat{\Omega}_m(U_l)} \right\}} \quad (68)$$

と計算される。その他の重みも同様に得ることができる:

$$w_i(U_l) = \frac{\frac{1}{\delta^2 \hat{\Omega}_i(U_l)}}{\left\{ \sum_{m=1}^M \frac{1}{\delta^2 \hat{\Omega}_m(U_l)} \right\}} \quad (69)$$

2.1.7 状態密度 $\Omega(U)$ の推定

得られた重み (69) と各シミュレーションに対する状態密度の推定値の統計誤差 (62) を用いると、状態密度の推定値 (50) は

$$\Omega(\hat{U}_l) = \sum_{m=1}^M w_m(U_l) \Omega_m(\hat{U}_l) \quad (70)$$

$$= \sum_{m=1}^M \frac{[\delta^2 \hat{\Omega}_m(U_l)]^{-1} \Omega_m(\hat{U}_l)}{[\sum_{i=1}^M \delta^2 \hat{\Omega}_i(U_l)]^{-1}} \quad (71)$$

$$= \sum_{m=1}^M \frac{\frac{g_m^{-1} N_m \Delta U \exp[f_m - \beta_m U_l]}{\hat{\Omega}(U_l)} \frac{H_m(U_l)}{N_m \Delta U \exp[f_m - \beta_m U_l]}}{\left[\sum_{i=1}^M \frac{g_i^{-1} N_i \Delta U \exp[f_i - \beta_i U_l]}{\hat{\Omega}(U_l)} \right]} \quad (72)$$

$$= \sum_{m=1}^M \frac{g_m^{-1} H_m(U_l)}{\sum_{i=1}^M g_i^{-1} N_i \Delta U \exp[f_i - \beta_i U_l]} \quad (73)$$

と計算できる。

2.1.8 確率分布と物理量の平均値

以上より、確率分布は

$$P(U_l; T) = \Omega(\hat{U}_l) e^{-\beta U_l} = \sum_{m=1}^M \frac{g_m^{-1} H_m(U_l) e^{-\beta U_l}}{\sum_{i=1}^M g_i^{-1} N_i \Delta U \exp[f_i - \beta_i U_l]} \quad (74)$$

と計算される。ここで、各シミュレーションに対する無次元化されたヘルムホルツの自由エネルギーは

$$e^{-f_m} = \sum_l \hat{\Omega}(U_l) e^{-\beta U_l} = \sum_l P(U_l; T) \quad (75)$$

である。また、物理量の平均値は

$$\langle A \rangle_T = \sum_l \frac{A(U_l) P(U_l; T)}{P(U_l; T)} \quad (76)$$

で計算することができる。

2.2 WHAM 法のまとめ: カノニカルアンサンブルの場合

カノニカルアンサンブルにおける物理量 $A(U)$ の平均値は、

$$\langle T \rangle_T = \frac{\int dU A(U) \omega(U) e^{-\beta U}}{\int dU \omega(U) e^{-\beta U}} \quad (77)$$

と計算できる。WHAM を用いることで、 M 個のシミュレーションで得られたサンプルを用いて系の状態密度を見積もることができる。ここで k 番目のシミュレーションは温度 $\beta_k = 1/k_B T_k$ で実行されたカノニカルシミュレーションとする。また、 k 番目のシミュレーションで得られたサンプル数を h_k とする。状態密度 $\Omega(U)$ は、無次元化されたヘルムホルツの自由エネルギー f_i を用いて、

$$\Omega(U; T) = \frac{\sum_{m=1}^M g_m H_m(U)}{\sum_{m=1}^M g_m N_m e^{f_m - \beta_m U}} \quad (78)$$

$$e^{-f_m} = \sum_U \Omega(U; T_m) e^{-\beta_m U} \quad (79)$$

である。ここで、 $H_i(U)$ は i 番目のシミュレーションによって得られた、ポテンシャルエネルギー U に関するヒストグラム、 $f_i = \beta_i F_i$ は無次元化されたヘルムホルツの自由エネルギーである。式 (78) と式 (79) を自己無撞着に解くことで、系の状態密度を推定することができる。 f_i の値のセットの初期値は任意に取ることができるので、全ての f_i に対して 0 としても大丈夫である。

3 多状態ベネット受容比法 (MBAR: Multistate Bennett Acceptance Ratio Estimator)

3.1 熱平衡状態の表式

K 個の同じ熱力学平衡状態 (例えば、 NVT , NPT , μNT アンサンブルなど) から N_i 個の相関のないサンプルを得たとする。各サンプルに対する状態は、用いたアンサンブルに応じて、逆温度 β 、ポテンシャルエネルギー U 、圧力 p 、化学ポテンシャル μ などにより特徴づけられる。ここで、熱平衡状態 i に対する無次元ポテンシャル関数 $u_i(\mathbf{x})$ を

$$u_i(\mathbf{x}) = \beta_i [U_i(\mathbf{x}) + p_i V(\mathbf{x}) + \mu_i^t n(\mathbf{x})] \quad (80)$$

と定義する。ここで、 $\mathbf{x} \in \Gamma$ は位相空間 Γ 内の系の構造を表す。また、 $V(\mathbf{x})$ は体積 (定圧アンサンブルの場合)、 $n(\mathbf{x})$ は M 成分系における各成分の分子の数 (グランドカノニカルアンサンブルの場合) を表す。アンブレラサンプリングなどバイアスポテンシャルを課している場合、 U の中にはそのようなバイアスポテンシャルも含まれることに注意する。

熱平衡状態 i から得られたサンプル $\{\mathbf{x}_{in}\}_{n=1}^{N_i}$ は、確率分布 p_i

$$p_i(\mathbf{x}) = \frac{q_i(\mathbf{x})}{\int_{\Gamma} d\mathbf{x} q_i(\mathbf{x})} = \frac{q_i(\mathbf{x})}{c_i} \quad (81)$$

にしたがっている。ここで、 q_i は非負の規格化されていない密度関数であり、 c_i は規格化定数 (統計力学では分配関数) である。規格化定数が一般的に事前に分かっているという状況は多くない。温度や圧力制御されたモンテカルロ法や分子動力学法で得られたサンプル、あるいは実験で得られたデータの場合、関数 q_i はボルツマン因子で表される:

$$q_i(\mathbf{x}) = e^{-u_i(\mathbf{x})} \quad (82)$$

ただし、マルチカノニカルシミュレーションなどの非ボルツマン因子に基づいたシミュレーションでは、このように表すことができないアンサンブルがあることに注意する必要がある。

3.2 平衡状態間の自由エネルギー差と物理量の期待値の表式

無次元化された自由エネルギー差は、

$$\Delta f_{ij} \equiv f_j - f_i = -\ln \frac{c_j}{c_i} = -\ln \frac{\int_{\Gamma} d\mathbf{x} q_j(\mathbf{x})}{\int_{\Gamma} d\mathbf{x} q_i(\mathbf{x})} \quad (83)$$

である。ただし、 f_i は次元を持つ自由エネルギー F_i と $f_i = \beta_i F_i$ の関係で結ばれている。

平衡状態 i における物理量 A の期待値は、

$$\langle A \rangle_i \equiv \int_{\Gamma} d\mathbf{x} p_i(\mathbf{x}) A(\mathbf{x}) = \frac{\int_{\Gamma} d\mathbf{x} A(\mathbf{x}) q_i(\mathbf{x})}{\int_{\Gamma} d\mathbf{x} q_i(\mathbf{x})} \quad (84)$$

新たに密度関数 $q(\mathbf{x}) = A(\mathbf{x}) q_i(\mathbf{x})$ を定義すれば、物理量の期待値は規格化定数の比で計算することができる。この時、 $q(\mathbf{x})$ はもはや非負である必要はない。

3.3 拡張ブリッジサンプリング推定法 (Extended Bridge Sampling Estimators)

規格化定数 c_i の比を推定する方法を構築する. 恒等式 $c_i \langle \alpha_{ij} q_j \rangle_i = c_j \langle \alpha_{ij} q_i \rangle_j$ が成立することは

$$c_i \langle \alpha_{ij} q_j \rangle_i = \left[\int_{\Gamma} d\mathbf{x} q_i(\mathbf{x}) \right] \frac{\int_{\Gamma} d\mathbf{x} q_i(\mathbf{x}) \alpha_{ij}(\mathbf{x}) q_j}{\int_{\Gamma} d\mathbf{x} q_i(\mathbf{x})} \quad (85)$$

$$= \int_{\Gamma} d\mathbf{x} q_i(\mathbf{x}) \alpha_{ij}(\mathbf{x}) q_j \quad (86)$$

$$= \left[\int_{\Gamma} d\mathbf{x} q_j(\mathbf{x}) \right] \frac{\int_{\Gamma} d\mathbf{x} q_j(\mathbf{x}) \alpha_{ij}(\mathbf{x}) q_i}{\int_{\Gamma} d\mathbf{x} q_j(\mathbf{x})} \quad (87)$$

$$= c_j \langle \alpha_{ij} q_i \rangle_j \quad (88)$$

から分かる. $\alpha_{ij}(\mathbf{x})$ はゼロでない c_i について任意に選べる関数である.

この恒等式について両辺共に j について和をとる. 期待値は $\langle g \rangle_i$ は, 経験的に, $\sum_{n=1}^{N_i} g(\mathbf{x}_{in})$ のように算術平均で推定することができることを用いれば, $i = 1, \dots, K$ に対して K 個の方程式を得る:

$$\sum_{j=1}^K \frac{\hat{c}_i}{N_i} \sum_{n=1}^{N_i} \alpha_{ij} q_j(\mathbf{x}_{in}) = \sum_{j=1}^K \frac{\hat{c}_j}{N_j} \sum_{n=1}^{N_j} \alpha_{ij} q_i(\mathbf{x}_{jn}) \quad (89)$$

ここで, 上付のハット[^]は推定値であることを示している. $i = 1, \dots, K$ について, 全ての \hat{c}_i に関する方程式の組みを解くことで, サンプルされたデータから c_i の推定値 \hat{c}_i を得ることができる.

$\alpha_{ij}(\mathbf{x})$ には様々な選び方が可能であり, その中にはリウエーティングで一般的に使用されるものもある. MBAR 法では, 可能な $\alpha_{ij}(\mathbf{x})$ の選び方の中でも, 最も小さい分散を持つという意味において最適な $\alpha_{ij}(\mathbf{x})$ を用いる. 具体的には, 以下のような $\alpha_{ij}(\mathbf{x})$ を採用する:

$$\alpha_{ij}(\mathbf{x}) = \frac{N_j \hat{c}_j^{-1}}{\sum_{k=1}^K N_k \hat{c}_k^{-1} q_k(\mathbf{x})} \quad (90)$$

この推定式は, 漸近不偏量であり, 唯一の解を持つことが保証されている. 式 (89) と式 (90) からでは, \hat{c}_i の組について閉じた方程式を導けないため, 直接解くことができない. 代わりに, 自己無撞着にあるいは Newton-Raphson 法を用いて, 数値的にこの非線形方程式を解くことができる.

サンプルの数が大きい範囲では, 一般的には比 \hat{c}_i/\hat{c}_j の誤差は分散される. 漸近的な共分散行列 $\Theta = \text{cov}(\ln c_i, \ln c_j) \equiv \text{cov}(\theta_i, \theta_j)$ は次のように推定される:

$$\hat{\Theta} = \mathbf{W}^T (\mathbf{I}_N - \mathbf{W} \mathbf{N} \mathbf{W}^T)^+ \mathbf{W} \quad (91)$$

ここで, \mathbf{I}_N は $N \times N$ の単位行列, $N = \sum_{i=1}^M N_i$ は全サンプル数, $\mathbf{N} = \text{diag}(N_1, N_2, \dots, N_M)$ は各シミュレーションのサンプル数を成分に持つベクトル, 上付の $+$ は Moore-Penrose のような一般化逆行列を表している. また, \mathbf{W} は $N \times K$ の重み行列で,

$$W_{ni} = \hat{c}_i^{-1} \frac{q_i(\mathbf{x}_n)}{\sum_{k=1}^M N_k \hat{c}_k^{-1} q_k(\mathbf{x}_n)} \quad (92)$$

と計算される. ここで, サンプルは 1 つのラベル $n = 1, \dots, N$ でラベル付けしている. したがって, サンプル

\mathbf{x}_n が, どの分布 $p(\mathbf{x})$ からサンプルされているかは, もはや関係なくなっている. また, 上記の定義において,

$$\sum_{n=1}^N W_{ni} = 1, \quad (i = 1, \dots, K) \quad (93)$$

$$\sum_{n=1}^K N_i W_{ni} = 1, \quad (n = 1, \dots, N) \quad (94)$$

を満たす. $\hat{\Theta}$ を見積もるとき, $N \times N$ の擬逆行列を見積もる計算コストは, $K \times K$ 行列の固有値を見積もる計算コストまで減らすことが可能である. 多くの場合で, 共分散行列は $K \times K$ 行列の操作で見積もられる. 規格化定数に対数をとった θ_i に関する任意の関数 $\phi(\theta_1, \dots, \theta_K)$ と $\psi(\theta_1, \dots, \theta_K)$ に関する推定値の共分散は $\hat{\Theta}$ から見積もられる:

$$\text{cov}(\hat{\phi}, \hat{\psi}) = \sum_{i,j=1}^K \frac{\partial \phi}{\partial \theta_i} \hat{\theta}_{ij} \frac{\partial \psi}{\partial \theta_j} \quad (95)$$

3.4 無次元化された自由エネルギーの推定

構造がボルツマン統計, すなわち $q_i \equiv \exp[-u_i(\mathbf{x})]$ からサンプリングされた時, 式 (89) と式 (90) から, 以下のような無次元自由エネルギーの推定式が得られる:

$$\hat{f}_i = -\ln \sum_{j=1}^K \sum_{n=1}^{N_j} \frac{\exp[-u_i(\mathbf{x}_{jn})]}{\sum_{k=1}^K N_k \exp[\hat{f}_k - u_k(\mathbf{x}_{jn})]} \quad (96)$$

この推定式は \hat{f}_i に対して自己無撞着に解くことができる.

推定された自由エネルギー差の不確かさは式 (91) と式 (92) より

$$\delta^2 \Delta \hat{f}_{ij} \equiv \text{cov} \left(-\ln \frac{\hat{c}_j}{\hat{c}_i}, -\ln \frac{\hat{c}_j}{\hat{c}_i} \right) \quad (97)$$

$$= \hat{\Theta}_{ii} - 2\hat{\Theta}_{ij} + \hat{\Theta}_{ji} \quad (98)$$

と計算される.

3.4.1 無次元化された自由エネルギーの推定式の導出

式 (89) と式 (90) から, 無次元自由エネルギーの推定式 (96) を導出する. 式 (89) と式 (90) をもう一度書いておくと, それぞれ

$$\sum_{j=1}^K \frac{\hat{c}_i}{N_i} \sum_{n=1}^{N_i} \alpha_{ij} q_j(\mathbf{x}_{in}) = \sum_{j=1}^K \frac{\hat{c}_j}{N_j} \sum_{n=1}^{N_j} \alpha_{ij} q_i(\mathbf{x}_{jn})$$

$$\alpha_{ij}(\mathbf{x}) = \frac{N_j \hat{c}_j^{-1}}{\sum_{k=1}^K N_k \hat{c}_k^{-1} q_k(\mathbf{x})}$$

であった. ここで,

$$q_i(\mathbf{x}) = e^{-u_i(\mathbf{x})} \quad (99)$$

$$c_i = e^{-f_i}$$

である。式 (90) を式 (89) に代入することで、

$$\begin{aligned} \sum_{j=1}^K \frac{\hat{c}_i}{N_i} \sum_{n=1}^{N_i} \frac{N_j \hat{c}_j^{-1}}{\sum_{k=1}^K N_k \hat{c}_k^{-1} q_k(\mathbf{x}_{in})} q_j(\mathbf{x}_{in}) &= \sum_{j=1}^K \frac{\hat{c}_j}{N_j} \sum_{n=1}^{N_j} \frac{N_j \hat{c}_j^{-1}}{\sum_{k=1}^K N_k \hat{c}_k^{-1} q_k(\mathbf{x}_{jn})} q_i(\mathbf{x}_{jn}) \\ &= \sum_{j=1}^K \sum_{n=1}^{N_j} \frac{q_i(\mathbf{x}_{jn})}{\sum_{k=1}^K N_k \hat{c}_k^{-1} q_k(\mathbf{x}_{jn})} \end{aligned} \quad (100)$$

さらに、左辺について、

$$\begin{aligned} \sum_{j=1}^K \frac{\hat{c}_i}{N_i} \sum_{n=1}^{N_i} \frac{N_j \hat{c}_j^{-1}}{\sum_{k=1}^K N_k \hat{c}_k^{-1} q_k(\mathbf{x}_{in})} q_j(\mathbf{x}_{in}) &= \frac{\hat{c}_i}{N_i} \sum_{j=1}^K \sum_{n=1}^{N_i} \frac{N_j \hat{c}_j^{-1}}{\sum_{k=1}^K N_k \hat{c}_k^{-1} q_k(\mathbf{x}_{in})} q_j(\mathbf{x}_{in}) \\ &= \frac{\hat{c}_i}{N_i} \sum_{j=1}^K \left[\frac{N_j \hat{c}_j^{-1} q_j(\mathbf{x}_{i1})}{\sum_{k=1}^K N_k \hat{c}_k^{-1} q_k(\mathbf{x}_{i1})} + \dots + \frac{N_j \hat{c}_j^{-1} q_j(\mathbf{x}_{iN_i})}{\sum_{k=1}^K N_k \hat{c}_k^{-1} q_k(\mathbf{x}_{iN_i})} \right] \\ &= \frac{\hat{c}_i}{N_i} \left[\frac{\sum_{j=1}^K N_j \hat{c}_j^{-1} q_j(\mathbf{x}_{i1})}{\sum_{k=1}^K N_k \hat{c}_k^{-1} q_k(\mathbf{x}_{i1})} + \dots + \frac{\sum_{j=1}^K N_j \hat{c}_j^{-1} q_j(\mathbf{x}_{iN_i})}{\sum_{k=1}^K N_k \hat{c}_k^{-1} q_k(\mathbf{x}_{iN_i})} \right] \\ &= \frac{\hat{c}_i}{N_i} N_i \\ &= \hat{c}_i \end{aligned}$$

と計算されるので、

$$\hat{c}_i = \sum_{j=1}^K \sum_{n=1}^{N_j} \frac{q_i(\mathbf{x}_{jn})}{\sum_{k=1}^K N_k \hat{c}_k^{-1} q_k(\mathbf{x}_{jn})} \quad (101)$$

を得る。 $q_i = e^{-u_i(\mathbf{x})}$, $c_i = e^{-f_i}$ を用いると

$$\hat{f}_i = -\ln \sum_{j=1}^K \sum_{n=1}^{N_j} \frac{\exp[-u_i(\mathbf{x}_{jn})]}{\sum_{k=1}^K N_k \exp[\hat{f}_k - u_k(\mathbf{x}_{jn})]} \quad (102)$$

が導出される。

3.4.2 無次元化された自由エネルギーの効率的な解法

非線形方程式である式 (96) を解くために、ここでは主に二つの方法を見ていく。一つは、自己無撞着 (self-consistent iteration) に解くことであり、もう一つはニュートンラフソン法を用いることである。自己無撞着法を用いた場合、自由エネルギーの収束が遅いというデメリットがあるが、安定して計算をすることができるため、確実に解を求めることができる。一方、ニュートンラフソン法を用いた場合、自由エネルギーが早く収束するメリットがあるが、初期値が適切でない場合に計算が破綻してしまうというデメリットがある。そこで、最初は自己無撞着に解いて、解がある程度収束したらニュートンラフソン法に切り替えると言った使い方も可能である。

自己無撞着法 非線形方程式 (96) を自己無撞着に解くには、最後のイテレーションで得た自由エネルギーの推定値 $\{\hat{f}_i^{(n)}\}_{i=1}^K$ を次のイテレーションでの推定 $\{\hat{f}_i^{(n+1)}\}_{i=1}^K$ に使用する。すなわち、

$$\hat{f}_i^{(n+1)} = -\ln \sum_{j=1}^K \sum_{n=1}^{N_j} \frac{\exp[-u_i(\mathbf{x}_{jn})]}{\sum_{k=1}^K N_k \exp[\hat{f}_k^{(n)} - u_k(\mathbf{x}_{jn})]} \quad (103)$$

を計算する。この式は、初期推定値 $\hat{f}_i^{(0)}$ によらず収束することが保証されている。最も簡単な初期推定値として $\hat{f}_i^{(0)} = 0$ と設定してもよいし、あるいは、より早い収束を期待して

$$f_k^{(0)} = \frac{1}{N_k} \sum_{n=1}^{N_k} \ln q(\mathbf{x}_{kn}) \quad (104)$$

としても良い。ここで、サンプルがボルツマン因子に従う場合、 $q_k(\mathbf{x}) = \exp[-u_k(\mathbf{x})]$ である。

イテレーション中に推定値の変化の大きさがコントロールできなくなることを防ぎ、かつ一意的な解を得るため、 $f_1 = 0$ と拘束する。つまり、イテレーションで推定値を更新したら、その都度 f_i から f_1 を引く：

$$\hat{f}_i^{(n+1)} \leftarrow \hat{f}_i^{(n+1)} - \hat{f}_1^{(n+1)} \quad \text{for all } i \quad (105)$$

計算終了判定は次のように決める：

$$\max_{i=2,\dots,K} \left[\frac{|\hat{f}_i^{(n+1)} - \hat{f}_i^{(n)}|}{|\hat{f}_i^{(n)}|} \right] < \epsilon \quad (106)$$

ここで ϵ は例えば 10^{-7} など十分小さい値を選択する。収束の速さは、注目する系やサンプル数などによって異なるので、可能な限りモニターすると良い。

式 (104) のような指数関数 e^a の和の計算では、指数関数が大きい値を持つために数値オーバーフローが起こりやすい。オーバーフローを避けるには、指数関数の和について、最も大きい値を持つ指数関数項 $a_m \equiv \max_n a_n$ で規格化した後で和を計算すると良い。すなわち

$$\sum_{n=1}^N \exp[a_n] = \exp[a_m] \sum_{n=1}^N \frac{\exp[a_n]}{\exp[a_m]} \quad (107)$$

と計算する。あるいは、次のように表すこともできる：

$$\ln \sum_{n=1}^N \exp[a_n] = a_m + \ln \sum_{n=1}^N \exp[a_n - a_m] \quad (108)$$

3.5 物理量の期待値の推定

構造 \mathbf{x} のみに依存する物理量 $A(\mathbf{x})$ の平衡分布における期待値は、式 (84) で与えられる。平衡分布における期待値は、特徴付けられる 2 つの追加した状態 A と状態 a に対する規格化定数の比 c_A/c_a として計算することができる。

$$q_A(\mathbf{x}) = A(\mathbf{x})q(\mathbf{x}) \quad (109)$$

$$q_a(\mathbf{x}) = q(\mathbf{x}) \quad (110)$$

ボルツマン因子に比例した期待値を要求する場合、 $q(\mathbf{x}) \equiv \exp[-u(\mathbf{x})]$ となる。 $q_A(\mathbf{x})$ はもはや正確には非負ではないが、サンプル数が $N_A = N_a = 0$ なので期待値 $\langle A \rangle$ を求めるのに拡張ブリッジサンプリング推定法の式 (89) を使うことができる。同様に、重み行列 (92) についても、 $q_A(\mathbf{x})$, $q_a(\mathbf{x})$ に対応する列 W_{nA} , W_{na} を拡張すると、

$$W_{nA} = \hat{c}_A^{-1} \frac{A(\mathbf{x}_n) \exp[-u(\mathbf{x}_n)]}{\sum_{k=1}^K N_k \exp[\hat{f}_k - u_k(\mathbf{x}_n)]} \quad (111)$$

$$W_{na} = \hat{c}_a^{-1} \frac{\exp[-u(\mathbf{x}_n)]}{\sum_{k=1}^K N_k \exp[\hat{f}_k - u_k(\mathbf{x}_n)]} \quad (112)$$

と計算される。ここで、規格化定数の推定値 \hat{c}_A , \hat{c}_a は自己無撞着推定方程式で定義される:

$$\hat{c}_A = \sum_{n=1}^N \frac{A(\mathbf{x}_n) \exp[-u(\mathbf{x}_n)]}{\sum_{k=1}^K N_k \exp[\hat{f}_k - u_k(\mathbf{x}_n)]} \quad (113)$$

$$\hat{c}_a = \sum_{n=1}^N \frac{\exp[-u(\mathbf{x}_n)]}{\sum_{k=1}^K N_k \exp[\hat{f}_k - u_k(\mathbf{x}_n)]} \quad (114)$$

ある平衡状態 α における物理量の期待値の推定値は,

$$\hat{A}_\alpha = \frac{\hat{c}_A}{\hat{c}_a} = \sum_{n=1}^N W_{na} A(\mathbf{x}_n) \quad (115)$$

と計算される。具体的には,

$$\hat{A}_\alpha = \sum_{n=1}^N W_{na} A(\mathbf{x}_n) \quad (116)$$

$$= \sum_{n=1}^N \hat{c}_a^{-1} \frac{\exp[-u_\alpha(\mathbf{x}_n)]}{\sum_{k=1}^K N_k \exp[\hat{f}_k - u_k(\mathbf{x}_n)]} A(\mathbf{x}_n) \quad (117)$$

$$= \sum_{n=1}^N \frac{A(\mathbf{x}_n) \exp[\hat{f}_\alpha - u_\alpha(\mathbf{x}_n)]}{\sum_{k=1}^K N_k \exp[\hat{f}_k - u_k(\mathbf{x}_n)]} \quad (118)$$

$$= \sum_{j=1}^K \sum_{n=1}^{N_j} \frac{A(\mathbf{x}_{jn}) \exp[\hat{f}_\alpha - u_\alpha(\mathbf{x}_{jn})]}{\sum_{k=1}^K N_k \exp[\hat{f}_k - u_k(\mathbf{x}_{jn})]} \quad (119)$$

とである。途中, $\hat{c}_a^{-1} = e^{\hat{f}}$ であることを用いた。また, 重み因子を

$$w(\mathbf{x}_{jn}) = \frac{\exp[-u(\mathbf{x}_{jn})]}{\sum_{k=1}^K N_k \exp[\hat{f}_k - u_k(\mathbf{x}_{jn})]} \quad (120)$$

と定義すれば, 期待値は

$$\hat{A}_\alpha = \frac{\sum_{j=1}^K \sum_{n=1}^{N_j} A(\mathbf{x}_{jn}) w(\mathbf{x}_{jn})}{\sum_{j=1}^K \sum_{n=1}^{N_j} w(\mathbf{x}_{jn})} \quad (121)$$

とかける。不確かさは, 拡張した重み行列 \mathbf{W} から計算される共分散行列 $\hat{\Theta}$ を用いて

$$\delta^2 \hat{A} \equiv \text{cov} \left(\frac{\hat{c}_A}{\hat{c}_a}, \frac{\hat{c}_A}{\hat{c}_a} \right) = \hat{A}^2 \left(\hat{\Theta}_{AA} + \hat{\Theta}_{aa} - 2\hat{\Theta}_{Aa} \right) \quad (122)$$

と計算される。

3.6 平均力ポテンシャル (PMF: Potential of Mean Force)

描きたい反応座標を適当にビンに分ける。ビン i に系が見つかる確率は,

$$p_i = \langle \chi_i(\mathbf{x}_n) \rangle = \sum_{n=1}^N W_{na} \chi_i(\mathbf{x}_n) \quad (123)$$

と計算される。ここで、 $\chi_i(x_n)$ は指示関数 (indicator function) であり、サンプルがビン i にいる時に 1 をとり、それ以外の時は 0 の値となる。したがって、自由エネルギーの推定値は、

$$F_i = -k_B T \ln(p_i/w_i) \quad (124)$$

と計算される。ただし、 w_i はビン幅の相対値である。ビン幅が全て等しい時は 1 とすることができる。

4 リウエイティング Tips

指数関数の足し算や引き算は計算機のオーバーフローが発生しやすい。これを防ぐために、対数関数の中で足し算・引き算を行うアルゴリズムを述べる。[7]

$A > 0, B > 0$ に対して $C = A + B$ を計算するとき、 $\ln A$ や $\ln B$ から $\ln C = \ln(A + B)$ を計算する:

$$\begin{aligned} \ln C &= \ln \left[\max(A, B) \left\{ 1 + \frac{\min(A, B)}{\max(A, B)} \right\} \right] \\ &= \max(\ln A, \ln B) + \ln [1 + \exp \{ \min(\ln A, \ln B) - \max(\ln A, \ln B) \}] \end{aligned} \quad (125)$$

一方、引き算をしたい時は、 $A > 0, B > 0$ に対して $|C| = |A - B|$ の計算を考える。すなわち、 $\ln A$ や $\ln B$ から $\ln |C| = \ln(|A - B|)$ を計算する:

$$\begin{aligned} \ln |C| &= \ln \left[\max(A, B) \left\{ 1 + \frac{\min(A, B)}{\max(A, B)} \right\} \right] \\ &= \max(\ln A, \ln B) + \ln [1 - \exp \{ \min(\ln A, \ln B) - \max(\ln A, \ln B) \}] \end{aligned} \quad (126)$$

参考文献

- [1] Alan M Ferrenberg and Robert H Swendsen. New Monte Carlo technique for studying phase transitions. *Physical Review Letters*, Vol. 61, No. 23, pp. 2635–2638, dec 1988.
- [2] Alan M. Ferrenberg and Robert H. Swendsen. Erratum: New Monte carlo technique for studying phase transitions (Physical Review Letters (1989) 63, 15 (1658)). *Physical Review Letters*, Vol. 63, No. 15, p. 1658, 1989.
- [3] Wolfhard Janke. Statistical Analysis of Simulations: Data Correlations and Error Estimation. *Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms*, Vol. 10, pp. 423–445, 2002.
- [4] John D Chodera, William C Swope, Jed W Pitara, Chaok Seok, and Ken A Dill. Use of the Weighted Histogram Analysis Method for the Analysis of Simulated and Parallel Tempering Simulations. *Journal of Chemical Theory and Computation*, Vol. 3, No. 1, pp. 26–41, jan 2007.
- [5] Shankar Kumar, John M. Rosenberg, Djamal Bouzida, Robert H. Swendsen, and Peter A. Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry*, Vol. 13, No. 8, pp. 1011–1021, oct 1992.
- [6] Alan M. Ferrenberg and Robert H. Swendsen. Optimized Monte Carlo data analysis. *Physical Review Letters*, Vol. 63, No. 12, pp. 1195–1198, sep 1989.
- [7] Bernd A Berg. Multicanonical simulations step by step. *Computer Physics Communications*, Vol. 153, No. 3, pp. 397–406, jul 2003.