Mbar*

by Weiwei Zhang, wwzhangcn@foxmail.com
** The Mbar section is derived following the procedure in document (https://github.com/samuelymei/freeenergy)

If there are R different simulation with each performed at different termperature($\beta_1 \neq \beta_2 \neq .. \neq \beta_R, \beta_k = (k_B T_k)^{-1}$), and the number of snapshot at temperature T_k is N_k. Therefore, the probability(P_m) of finding snapshot(x) from all trajectory is:

$$P_m(x) = \frac{1}{N}\sum_{k=1}^{R} N_k p_k(x) \tag{1}$$

where $N = \sum_k N_k$, p$_k(x) = c_k^{-1} q_k(x)$ is the probability of locating at x for each snapshot in trajectory k, and $c_k = \int q_k(x)dx$

For any varibale $\hat{O}$, the expectation at Temperature T_i is $< O >_i$, which is difined by:

$$< O >_i = \int O(R)p_i(R)dR = \int O(R)\frac{p_i(R)}{P_m(R)}P_m(R)dR \approx \frac{1}{N}\sum_{n=1}^{N}\hat{O}(R_n)\frac{p_i(R_n)}{P_m(R_n)} = \sum_{n=1}^{N}\hat{O}(R_n)\frac{p_i(R_n)}{NP_m(R_n)} \tag{2}$$

$$= \sum_{n=1}^{N}\hat{O}(R_n)c_i^{-1}\frac{q_i(R_n)}{\sum_{k=1}^{R} N_k p_k(R_n)} \tag{3}$$

Therefore, the weight assigned to each snapshot is:

$$w_i(R_n) = c_i^{-1}\frac{q_i(R_n)}{\sum_{k=1}^{R} N_k p_k(R_n)} \tag{4}$$

Let $\hat{O}(R_n) = 1$, based on equation (3), we find:

$$c_i = \sum_{n=1}^{N}\frac{q_i(R_n)}{\sum_{k=1}^{R} N_k c_k^{-1} q_k(R_n)} \tag{5}$$

If q_i follows the Boltzmann statistics, $q_i(x) = exp(-\beta_i U_i(x))$, and $f_i = -\beta_i^{-1} lnc_i$, thus:

$$f_i = -\beta_i^{-1} ln\sum_{n=1}^{N}\frac{exp[-\beta_i U_i(R_n)]}{\sum_{k=1}^{R} N_k exp[\beta_k f_k - \beta_k U_k(R_n)]} \tag{6}$$

Therefore, normalized weight is :

$$w_i(R_n) = \frac{e^{\beta_i f_i - \beta_i U_i(R_n)}}{\sum_{k=1}^{R} N_k e^{\beta_k f_k - \beta_k U_k(R_n)}} \tag{7}$$

And, unnormalized weight is:

$$w2_i(R_n) = \frac{e^{-\beta_i U_i(R_n)}}{\sum_{k=1}^{R} N_k e^{\beta_k f_k - \beta_k U_k(R_n)}} \tag{8}$$

Where $e^{\beta_k f_k} = c_k$ is the normalized factor at temperature T_k.
Specially, if all simulations were performed at constant temperature(T) and the hamiltonians differ only by bias($V_i(R_n)$, constant temperature umbrella sampling), where $U_i(R_n) = U_0(R_n) + V_i(R_n)$ and U_0 is the system's original potential energy, it can be reduced to:

$$w_i(R_n) = \frac{e^{\beta f_i - \beta V_i(R_n)}}{\sum_{k=1}^{R} N_k e^{\beta f_k - \beta V_k(R_n)}} \tag{9}$$

$$w2_i(R_n) = \frac{e^{-\beta V_i(R_n)}}{\sum_{k=1}^{R} N_k e^{\beta f_k - \beta V_k(R_n)}} \tag{10}$$

Relationship with WHAM

If we discretize all samples into two dimensions: reaction coordinate($\hat{\xi}(R_n)$) and system original potential($\hat{U}_m^0(R_n)$), according to equation (7), we can get the probability(P(T_i, U_m^0, ξ)) of state with reaction coordinate equal to ξ and original potential energy equal to U_m^0 at temperature T_i, where the total potential energy $U_i(R_n) = \hat{U}_m^0(R_n) + V_i(\hat{\xi}(R_n)) = U_i(m, \xi)$, and express it by the following:

$$P(T_i, U_i(m,\xi)) = P(T_i, U_m^0, \xi) = P(T_i, m, \xi) = \sum_{n=1}^{N} w_i(R_n)\delta(\hat{U}_m^0(R_n) - U_m^0)\delta(\hat{\xi}(R_n) - \xi) \tag{11}$$

$$= \left(\frac{H_m(\xi)}{\sum_{k=1}^{K} N_k e^{\beta_k f_k - \beta_k U_k(m,\xi)}} \right) e^{\beta_i f_i - \beta_i U_i(m,\xi)} \tag{12}$$

$$= \left(\frac{\sum_{k=1}^{K} H_{mk}(\xi)}{\sum_{k=1}^{K} N_k e^{\beta_k f_k - \beta_k U_k(m,\xi)}} \right) e^{\beta_i f_i - \beta_i U_i(m,\xi)} = \left(\frac{\sum_{k=1}^{K} H_{mk}(\xi)}{\sum_{k=1}^{K} N_k e^{\beta_k f_k - \beta_k (U_m^0 + V_k(\xi))}} \right) e^{\beta_i f_i - \beta_i (U_m^0 + V_i(\xi))} \tag{13}$$

Therefore, the **unbiased** probability $P_0(T, m, \xi)$ at temperature T is:

$$P_0(T, m, \xi) = \left(\frac{\sum_{k=1}^{K} H_{mk}(\xi)}{\sum_{k=1}^{K} N_k e^{-\beta_k U_m^0} e^{\beta_k f_k - \beta_k V_k(\xi)}} \right) e^{\beta f_0 - \beta U_m^0} \tag{14}$$

where $H_m(\xi), H_{mk}(\xi)$ are the histogram of ξ from all trajectories and trajectory k respectively, where the system original potential energy($\hat{U}_m^0(R_n)$) is equal to U_m^0, and $e^{\beta f_0}$ is defined based on the equation (6):

$$e^{-\beta f_0} = \sum_{n=1}^{N} \frac{exp[-\beta U_m^0(R_n)]}{\sum_{k=1}^{R} N_k exp[\beta_k f_k - \beta_k (U_m^0(R_n) + V_k(\hat{\xi}(R_n))]} \tag{15}$$

Constant temperature umbrella sampling

Based on (14), when all the simulations were performed at the constant temperature($\beta_0 = \beta_1 = .. = \beta_k = \beta$, umbrella sampling), the **unbiased** probability($P_0(T, U_m, \xi), P_0(T, \xi)$) can be expressed by:

$$P_0(T, m, \xi) = \left(\frac{\sum_{k=1}^{K} H_{mk}(\xi)}{\sum_{k=1}^{K} N_k e^{-\beta_k U_m^0} e^{\beta_k f_k - \beta_k V_k(\xi)}} \right) e^{\beta f_0 - \beta U_m^0} = \frac{\sum_{k=1}^{K} H_{mk}(\xi)}{\sum_{k=1}^{K} N_k e^{\beta_k f_k - \beta_k V_k(\xi)}} e^{\beta f_0} (\beta_1 = ... = \beta_k = \beta) \tag{16}$$

$$P_0(T, \xi) = \sum_{m=1}^{M} P_0(T, m, \xi) = \frac{\sum_{m}^{M} \sum_{k=1}^{K} H_{mk}(\xi)}{\sum_{k=1}^{K} N_k e^{\beta_k f_k - \beta_k V_k(\xi)}} e^{\beta f_0} = \frac{\sum_{k=1}^{K} \overbrace{H_k(\xi)}}{\sum_{k=1}^{K} N_k e^{\beta_k f_k - \beta_k V_k(\xi)}} e^{\beta f_0} \tag{17}$$

where $\overbrace{H_k(\xi)} = \sum_{m}^{M} H_{mk}(\xi)$

Based on the equation (6), we can find:

$$e^{-\beta_i f_i(\xi)} = \sum_{m=1}^{M}\sum_{n=1}^{N} \frac{exp[-\beta_i(U_m^0(R_n) + V_i(\xi(\hat{R}_n)))]}{\sum_{k=1}^{R} N_k exp[\beta_k f_k - \beta_k(U_m^0(R_n) + V_k(\hat{\xi}(R_n))]}\delta(\hat{\xi}(R_n)-\xi)\delta(\hat{U}_m^0(R_n)-U_m^0)$$

$$= \frac{\sum_{k=1}^{K}\overbrace{H_k(\xi)}}{\sum_{k=1}^{K} N_k e^{\beta_k f_k - \beta_k V_k(\xi)}} e^{-\beta_i V_i(\xi)} (\beta_1 = ... = \beta_k = \beta) \tag{18}$$

$$e^{-\beta f_0(\xi)} = \sum_{m=1}^{M}\sum_{n=1}^{N} \frac{exp[-\beta U_m^0(R_n)]}{\sum_{k=1}^{R} N_k exp[\beta_k f_k - \beta_k(U_m^0(R_n) + V_k(\hat{\xi}(R_n))]}\delta(\hat{\xi}(R_n)-\xi)\delta(\hat{U}_m^0(R_n)-U_m^0)$$

$$= \frac{\sum_{k=1}^{K}\overbrace{H_k(\xi)}}{\sum_{k=1}^{R} N_k e^{\beta_k f_k - \beta_k V_k(\xi)}} (\beta_1 = ... = \beta_k = \beta) \tag{19}$$

$$e^{-\beta_i f_i} = \sum_{\xi} e^{-\beta_i f_i(\xi)} \tag{20}$$

$$e^{-\beta f_0} = \sum_{\xi} e^{-\beta f_0(\xi)} \tag{21}$$

Based on equation (17), the weight assigned to each snapshot is:

$$w_0(\hat{\xi}(R_n)) = \frac{e^{\beta f_0}}{\sum_{k=1}^{K} N_k e^{\beta_k f_k - \beta_k V_k(\xi)}} \tag{22}$$

Thus, For constant temperature simulation($\beta_1 = ... = \beta_k = \beta$),

$$e^{-\beta_k f_k} e^{\beta f_0} = \sum_{\xi} e^{-\beta f_k(\xi)} e^{\beta f_0} = \sum_{\xi} \frac{\sum_{k=1}^{K}\overbrace{H_k(\xi)}}{\sum_{k=1}^{K} N_k e^{\beta f_k - \beta V_k(\xi)}} e^{-\beta V_k(\xi)} e^{\beta f_0}$$

$$= \sum_{\xi} \left(\frac{\sum_{k=1}^{K} \overbrace{H_k(\xi)}}{\sum_{k=1}^{K} N_k e^{\beta f_k - \beta V_k(\xi)}} e^{\beta f_0} \right) e^{-\beta V_k(\xi)} = \sum_{\xi} P_0(T,\xi) e^{-\beta V_k(\xi)} =< e^{-\beta V_k(\xi)} >_0 \tag{23}$$

which is the same as equation(9) in paper "Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations" , Computer Physics Communications 135 (2001) 40–57.
As $e^{\beta f_0}$(constant C in equation(9) in the above suggested paper) is the normalized factor, if we assume $P_0(T_m, \xi)$ is already normalized with $w2_i(R_n) = \sum_1^N \frac{-\beta_i U_i^0(R_n)}{\sum_{k=1}^R N_k e^{\beta_k f_k - \beta_k U_k(R_n)}} = 1((8))$, we can find:
$e^{-\beta_k f_k} =< e^{-\beta V_k(\xi)} >_0$, which is widely used in caculating WHAM.

Different temperature umbrella sampling

Especially, when the umbrella simulations were performed at different temperature($\beta_1 \neq \beta_2 ... \neq \beta_k, V_k(\xi) \neq 0$), we can obtain that:

$$P_0(T,\xi) = \sum_{m=1}^{M} P_0(T,m,\xi) = \sum_{m}^{M} \left(\frac{\sum_{k=1}^{K} H_{mk}(\xi)}{\sum_{k=1}^{K} N_k e^{-\beta_k U_m^0} e^{\beta_k f_k - \beta_k V_k(\xi)}} \right) e^{\beta f_0 - \beta U_m^0} \tag{24}$$

Similarly, we find:

$$e^{-\beta_i f_i(\xi)} = \sum_{m}^{M} \sum_{n=1}^{N} \frac{exp[-\beta_i (U_m^0(R_n) + V_i(\xi(\hat{R}_n)))]}{\sum_{k=1}^R N_k exp[\beta_k f_k - \beta_k (U_m^0(R_n) + V_k(\hat{\xi}(R_n))]} \delta(\hat{\xi}(R_n) - \xi) \delta(\hat{U}_m^0(R_n) - U_m^0)$$

$$= \sum_{m}^{M} \frac{\sum_{k=1}^{K} H_{mk}(\xi)}{\sum_{k=1}^{K} N_k e^{\beta_k f_k - \beta_k (U_m^0(R_n) + V_k(\xi))}} e^{-\beta_i (U_m^0(R_n) + V_i(\xi))} \tag{25}$$

$$e^{-\beta f_0(\xi)} = \sum_{m}^{M}\sum_{n=1}^{N}\frac{exp[-\beta U_m^0(R_n)]}{\sum_{k=1}^{R} N_k exp[\beta_k f_k - \beta_k(U_m^0(R_n) + V_k(\hat{\xi}(R_n))]}\delta(\hat{\xi}(R_n) - \xi)\delta(\hat{U}_m^0(R_n) - U_m^0)$$

$$= \sum_{m}^{M}\frac{\sum_{k=1}^{K} H_{mk}(\xi)}{\sum_{k=1}^{R} N_k e^{\beta_k f_k - \beta_k(U_m^0(R_n)+V_k(\xi))}} e^{-\beta U_m^0(R_n)} \tag{26}$$

$$e^{-\beta_i f_i} = \sum_{\xi} e^{-\beta_i f_i(\xi)} \tag{27}$$

$$e^{-\beta f_0} = \sum_{\xi} e^{-\beta f_0(\xi)} \tag{28}$$

Similarly, we can also get the weight assigned to snapshot at temperature T($\beta = k_B T$) with **unbiased** potential according to equation (24):

$$w_0(\hat{\xi}(R_n)) = \sum_{m}^{M}\frac{e^{\beta f_0 - \beta U_m^0(R_n)}}{\sum_{k=1}^{K} N_k e^{-\beta_k U_m^0} e^{\beta_k f_k - \beta_k V_k(\xi)}} \tag{29}$$

$$e^{-\beta_i f_i} e^{\beta f_0} = \sum_{\xi} e^{-\beta_i f_i(\xi)} e^{\beta f_0} = \sum_{\xi}\sum_{m}^{M}\frac{\sum_{k=1}^{K} H_{mk}(\xi)}{\sum_{k=1}^{K} N_k e^{\beta_k f_k - \beta_k(U_m^0(R_n)+V_k(\xi))}} e^{-\beta_i(U_m^0(R_n)+V_i(\xi))} e^{\beta f_0}$$

$$=\sum_{\xi}\left(\sum_{m}^{M}\frac{\sum_{k=1}^{K} H_{mk}(\xi)}{\sum_{k=1}^{K} N_k e^{\beta_k f_k - \beta_k(U_m^0(R_n)+V_k(\xi))}} e^{\beta f_0 - \beta_i U_m^0(R_n)}\right) e^{-\beta_i V_i(\xi)} (30)$$

If $\beta_i = \beta$, we can get:

$$e^{-\beta_i f_i} e^{\beta f_0} = \sum_{\xi} P_0(T, \xi) e^{-\beta V_k(\xi)} =< e^{-\beta V_k(\xi)} >_0 \tag{31}$$

Similarly, if we assume that $e^{\beta f_0} = 1$, we also get that: $e^{\beta_i f_i} =< e^{-\beta V_k(\xi)} >_0$

Parallel tempering simulation

While all the simulations were performed at different temperature without bias($\beta_1 \neq \beta_2 ... \neq \beta_k, V_k(\xi) = 0$, parallel tempering simulation(replica-exchange)), we can get $P(T_i, U_m, \xi), P(T_i, m)$ at temperature T_i:

$$P(T_i, m) = \sum_{\xi} P(T_i, U_m, \xi) = \left(\frac{\sum_{\xi} \sum_{k=1}^{K} H_{mk}(\xi)}{\sum_{k=1}^{K} N_k e^{-\beta_k U_m^0} e^{\beta_k f_k}} \right) e^{\beta_i f_i - \beta_i U_m^0} = \left(\frac{\sum_{k=1}^{K} \widehat{H_{mk}}}{\sum_{k=1}^{K} N_k e^{\beta_k f_k - \beta_k U_m^0}} \right) e^{\beta_i f_i - \beta_i U_m^0} \tag{32}$$

Therefore, if $e^{\beta_i f_i} = 1$

$$P(T, E) = \left(\frac{\sum_{k=1}^{K} \widehat{H_{mk}}}{\sum_{k=1}^{K} N_k e^{\beta_k f_k - \beta_k E}} \right) e^{-\beta E} \tag{33}$$

where $\widehat{H_{mk}} = \sum_{\xi} H_{mk}(\xi)$, which is independent on the ξ. Calculating the expectation of variable $\hat{A}$,

$$< \hat{A} >_{T_i} = \frac{\sum_{m=1}^{M} \sum_{\xi} P(T_i, U_m, \xi) \hat{A}}{\sum_{m=1}^{M} \sum_{\xi} P(T_i, U_m, \xi)} = \frac{\sum_{m=1}^{M} P(T_i, U_m) \hat{A}}{\sum_{m=1}^{M} P(T_i, U_m)} \tag{34}$$

or

$$< \hat{A} >_{T} = \frac{\sum_{E} P(T, E) \hat{A}}{\sum_{E} P(T, E)} \tag{35}$$