The Weighted Histogram Analysis Method (WHAM)

Michael Andrec

January 21, 2010

1 Introduction

There are situations in which it is necessary to combine samples obtained under different simulation conditions to determine properties (e.g. averages or PMFs) of the system under a single condition (which may not even be one of the conditions that were actually simulated). One example is the determination of a PMF to obtain accurate measures of a free energy barrier. If the barrier is substantially higher than the minima, then a standard canonical simulation will provide little sampling at the barrier, and thus may not be able to accurately estimate the barrier height. An alternative strategy is to use umbrella sampling with a series of biasing potentials to confine the system to small regions of the PMF. Together, these provide adequate sampling of the entire reduced coordinate of interest (see, for example, references 1,2). Since we know the biasing potential, we can "unbias" the samples of each simulation. Suppose that the underlying potential of interest is $U(\mathbf{p})$ and we apply a biasing potential $V_i(x)$ to the i-th simulation, where $x(\mathbf{p})$ is the reduced coordinate with respect to which we are interested in obtaining the PMF, i.e. we obtain a canonical sample distributed according to $\exp[-\beta(U(\mathbf{p})+V_i(x(\mathbf{p}))]$. To unbias the estimated population $P(x = x_j)$ we can multiply the estimated probability by $\exp[\beta V_i(x_j)]$. To obtain the full unbiased distribution, we need to combine the data from all of the simulations. If the sampling were extremely thorough, then

the unbiased distributions for different simulations i would be the same, and any linear combination would give the same result. However, this is a very unrealistic limit, since if this were the case, then there would have been no reason to use a biasing potential in the first place. For any realistic situation, the optimal weights will not be equal, but should be chosen so that the simulations which have a higher uncertainty in their estimates of the unbiased probabilities are weighted less.

Another example of a situation where reweighting is necessary is when simulations are performed at multiple temperatures, but where the property of interest is only needed at one temperature. This arises in replica exchange simulations, for example. Simulations at higher temperature than the target temperature allow better sampling of regions that are high in potential energy, while simulations at lower temperature provide better sampling of regions that have high free energy due to low entropy.[3]

The general problem which will be studied in this chapter is formulated as follows. Suppose that we have performed S simulations and have generated N_i samples from the i-th simulation. We then discretize the samples into M bins to determine a histogram with respect to the biasing coordinate. If the simulations were performed using a biasing potential with respect to a reduced coordinate x, then the histogram must be with respect to x. If the simulations were temperature-biased, then the histogram must be with respect to the potential energy $E(\mathbf{p})$. If both temperature and coordinate biasing were used, then a two-dimensional histogram with respect to both x and E must be constructed. Let p_{ij} be the estimate of the (biased) probability in the j-th bin in the i-th simulation. We assume that p_{ij} is related to p_j°, the unbiased probability of bin j, via

$$p_{ij} = f_i c_{ij} p_j^\circ, \tag{1}$$

where c_{ij} is the biasing factor and f_i is a normalizing constant chosen such that $\sum_j p_{ij} = 1$, i.e. $f_i^{-1} = \sum_j c_{ij} p_j^\circ$. The biasing factor must be chosen appropriate to the simulation performed, e.g. $c_{ij} = \exp[-(\beta_i - \beta_0)E_j]$ for temperature biasing (where β_i and β_0 are the inverse temperatures of

simulation i and the target temperature, respectively). For coordinate biasing, $c_{ij} = \exp[-\beta V_i(x_j)]$. In addition, the normalizing constants f_i have physical meaning, as $\beta_i^{-1} \log f_i$ is the Helmholtz free energy of the i-th simulation.[4] We wish to find an optimal estimate of p_j°.

We will show below that an optimal estimate of p_j° is given by

$$p_j^\circ = \frac{\sum_{i=1}^{S} n_{ij}}{\sum_{i=1}^{S} N_i f_i c_{ij}} \tag{2}$$

where n_{ij} is the number of counts in histogram bin j for simulation i, and N_i is the total number of samples generated by the i-th simulation. As stated above, the unbiased probabilities are also constrained by the normalization condition

$$f_i^{-1} = \sum_{j=1}^{M} c_{ij} p_j^\circ. \tag{3}$$

Equations 2 and 3 are collectively known as the Weighted Histogram Analysis Method (WHAM) equations, which were first proposed by Ferrenberg and Swendsen in the context of the statistical physics of Ising models[5,6] and were introduced into the molecular simulation literature by Kumar, et al.[4] The WHAM equations are a system of $M + S$ nonlinear equations that must be solved self-consistently for p_j° ($j = 1, \ldots, M$) and f_i ($i = 1, \ldots, S$). In practice, this is typically done in an iterative fashion by using an arbitrary set of starting values for f_i (often $f_1 = f_2 = \cdots = f_S = 1$) to calculate p_j° using Equation 2, then using those p_j° values to update f_i, and repeating the process until convergence is achieved. In principle, any numerical method for solving systems of nonlinear equations[7] could be used. It is interesting to note that Equation 2 depends on the bin counts n_{ij} only via their sum over all simulations, i.e. we only need to know the total number of counts in a given bin from all simulations and not what proportion of the counts came from which simulation.

2 Two derivations of the WHAM equations

In this section we present two alternative derivations of Equation 2. The first derivation (section 2.1) roughly follows that of Ferrenberg and Swendsen[5] and Kumar, et al.,[4] and is based on a minimization of the variance of the estimated unbiased probability. The second (section 2.2) is based on the work of Bartels and Karplus,[8] which uses a maximum likelihood approach. It is somewhat surprising that although the two derivations appear to be based on very different assumptions and approximations, they end up giving the same final result. This would seem to imply that the two approaches are in some way equivalent. The connection may be related to the fact that maximal likelihood estimators in the "asymptotic limit" of large sample sizes have the smallest possible variance in the estimated parameters achievable by any unbiased estimator,[9] however a full examination of this idea is beyond the scope of this course.

2.1 The "traditional" derivation

Let Ω_{ij} be the best estimate of the unbiased probability of the j-th bin using the i-th simulation, i.e.

$$\Omega_{ij} = \frac{n_{ij}}{N_i c_{ij} f_i} \tag{4}$$

(which is obtained by rearranging the expression analogous to Equation 1 but with Ω_{ij} substituting for p_j°). We will estimate p_j° as a weighted sum of Ω_{ij} over all simulations i:

$$p_j^\circ = \sum_{i=1}^{S} w_i \Omega_{ij}. \tag{5}$$

We wish to find the values of w_i (subject to the constraint $\sum_{i=1}^{S} w_i = 1$) which minimize the expected variance in p_j°:

$$\begin{aligned}
\mathrm{var}(p_j^\circ) &= \langle (p_j^\circ - \langle p_j^\circ \rangle)^2 \rangle \\
&= \left\langle \left(\sum_{i=1}^{S} w_i \Omega_{ij} - \left\langle \sum_{i=1}^{S} w_i \Omega_{ij} \right\rangle \right)^2 \right\rangle \\
&= \left\langle \left(\sum_{i=1}^{S} w_i (\Omega_{ij} - \langle \Omega_{ij} \rangle) \right)^2 \right\rangle .
\end{aligned}$$

Let $\delta\Omega_{ij} \equiv \Omega_{ij} - \langle \Omega_{ij} \rangle$. Then

$$\begin{aligned}
\mathrm{var}(p_j^\circ) &= \left\langle \left(\sum_{i=1}^{S} w_i \delta\Omega_{ij} \right)^2 \right\rangle \\
&= \left\langle \sum_{i=1}^{S} w_i^2 (\delta\Omega_{ij})^2 + \sum_{k \neq l=1}^{S} w_k w_l \delta\Omega_{kj} \delta\Omega_{lj} \right\rangle \\
&= \sum_{i=1}^{S} w_i^2 \langle (\delta\Omega_{ij})^2 \rangle + \sum_{k \neq l=1}^{S} w_k w_l \langle \delta\Omega_{kj} \delta\Omega_{lj} \rangle .
\end{aligned}$$

But $\langle (\delta\Omega_{ij})^2 \rangle = \mathrm{var}(\Omega_{ij})$, and if we assume that simulations k and l ($k \neq l$) are uncorrelated, then $\langle \delta\Omega_{kj} \delta\Omega_{lj} \rangle = 0$. Therefore

$$\mathrm{var}(p_j^\circ) = \sum_{i=1}^{S} w_i^2 \mathrm{var}(\Omega_{ij}). \tag{6}$$

Next, we express the variance of Ω_{ij} in terms of the variance of n_{ij}. Making use of the fact that $\mathrm{var}(ax) = \langle a^2 x^2 \rangle - \langle ax \rangle^2 = a^2 \langle x^2 \rangle - (a \langle x \rangle)^2 = a^2 \mathrm{var}(x)$, it follows from Equation 4 that

$$\mathrm{var}(\Omega_{ij}) = \frac{\mathrm{var}(n_{ij})}{N_i^2 c_{ij}^2 f_i^2}$$

and, therefore, that

$$\mathrm{var}(p_j^\circ) = \sum_{i=1}^{S} \frac{w_i^2 \mathrm{var}(n_{ij})}{N_i^2 c_{ij}^2 f_i^2}. \tag{7}$$

We now need to determine the variance of n_{ij}. In general, if there are N independent samples in a simulation, then the probability of having n counts in a histogram bin is given by the binomial distribution

$$P(n) = \binom{N}{n} p^n (1-p)^{N-n}, \tag{8}$$

where p is the probability of the bin. The mean and variance of the binomial distribution are np and variance $np(1-p)$, respectively. In the limit of large N and small p, we can approximate the binomial distribution by the Poisson distribution

$$P(n) = \exp(-Np)\frac{(Np)^n}{n!},$$

which has mean and variance both equal to np (ref. 10 p. 32). The probability of bin j in simulation i is $f_i c_{ij} p_j^\circ$, and therefore (in limit of the Poisson approximation) the variance of n_{ij} is equal to $N_i f_i c_{ij} p_j^\circ$. Inserting this into Equation 7 we find that

$$\mathrm{var}(p_j^\circ) = \sum_{i=1}^{S} \frac{w_i^2 p_j^\circ}{N_i c_{ij} f_i}. \tag{9}$$

We now minimize Equation 9 with respect to w_i subject to the constraint that $\sum_{i=1}^{S} w_i = 1$. Specifically, we minimize

$$Q = \sum_{i=1}^{S} \frac{w_i^2 p_j^\circ}{N_i c_{ij} f_i} + \lambda \sum_{i=1}^{S} w_i,$$

where λ is a Lagrange multiplier. Taking derivatives

$$\frac{\partial Q}{\partial w_k} = \frac{2 w_k p_j^\circ}{N_k c_{kj} f_k} + \lambda$$

and setting them to zero, we find that

$$w_k = \frac{-N_k c_{kj}}{2 p_j^\circ}\lambda.$$

Applying the constraint

$$\sum_{k=1}^{S} w_k = \frac{-\lambda}{2 p_j^\circ} \sum_{k=1}^{S} N_k c_{kj} = 1,$$

we obtain

$$\lambda = \frac{-2 p_j^\circ}{\sum_{i=1}^{S} N_i c_{ij}},$$

which finally gives the optimal w_i:

$$w_i = \frac{N_i c_{ij}}{\sum_{k=0}^{S} N_k c_{ik}}.$$

Substituting these weights into Equation 5, we obtain equation 2.

2.2 The maximum likelihood derivation

We begin by noting that the probability of observing a histogram n_{ij} $(j = 1 \ldots M)$ is given by the multinomial distribution

$$P(n_{i1}, \ldots, n_{iM}) = \frac{N_i!}{\prod_{k=1}^{M} n_{ik}!} \prod_{j=1}^{M} (p_{ij})^{n_{ij}}, \tag{10}$$

which can be thought of as the multivariate generalization of the binomial distribution of Equation 8. If the S simulations are uncorrelated, then the total probability of observing all S histograms is the product of multinomials

$$\begin{aligned} P(n_{i1}, \ldots, n_{iM}, \ldots, n_{S1}, \ldots, n_{SM}) &= \prod_{i=1}^{S} \left[\frac{N_i!}{\prod_{k=1}^{M} n_{ik}!} \prod_{j=1}^{M} (p_{ij})^{n_{ij}} \right] \\ &= \prod_{i=1}^{S} \left[\frac{N_i!}{\prod_{k=1}^{M} n_{ik}!} \prod_{j=1}^{M} (f_i c_{ij} p_j^{\circ})^{n_{ij}} \right]. \end{aligned} \tag{11}$$

We will maximize Equation 11 with respect to f_i $(i = 1, \ldots, S)$ and p_j° $(j = 1, \ldots, M)$ subject to the S constraints $\sum_{j=1}^{m} f_i c_{ij} p_j^{\circ} = 1$. To facilitate this, we take the logarithm of Equation 11 and drop all terms that are independent of f_i and p_j° to obtain the log-likelihood function

$$L = \sum_{i=1}^{S} \sum_{j=1}^{M} n_{ij} \ln(f_i c_{ij} p_j^{\circ}).$$

Since there are S constraints, there will be S Lagrange multipliers λ_i, giving

$$F = \sum_{i=1}^{S} \sum_{j=1}^{M} \left[n_{ij} \ln(f_i c_{ij} p_j^{\circ}) + \lambda_i f_i c_{ij} p_j^{\circ} \right].$$

Taking derivatives with respect to f_i and p_j° we obtain

$$\begin{aligned} \frac{\partial F}{\partial f_k} &= \sum_{j=1}^{M} \left[\frac{n_{kj}}{f_k} + \lambda_k c_{kj} p_j^{\circ} \right] \\ &= \frac{N_k}{f_k} + \lambda_k \sum_{j=1}^{M} c_{kj} p_j^{\circ} \end{aligned} \tag{12}$$

and

$$\begin{aligned}\frac{\partial F}{\partial p_k^\circ} &= \sum_{i=1}^{S}\left[\frac{n_{kj}}{p_k^\circ} + \lambda_i f_i c_{ik}\right] \\ &= \frac{\sum_{i=1}^{S} n_{ik}}{p_k^\circ} + \sum_{i=1}^{S}\lambda_i f_i c_{ik}. \end{aligned} \tag{13}$$

Setting Equation 12 equal to zero and solving for λ_k, we find that

$$\lambda_k = \frac{-N_k}{\sum_{j=1}^{M} f_k c_{kj} p_j^\circ} = -N_k, \tag{14}$$

where we have made use of the constraint $\sum_{j=1}^{m} f_i c_{ij} p_j^\circ = 1$. Substituting Equation 14 into Equation 13 and setting it equal to zero, we find that

$$\frac{\sum_{i=1}^{S} n_{ik}}{p_k^\circ} = \sum_{i=1}^{S} N_i f_i c_{ik},$$

which when solved for p_k° gives Equation 2.

2.3 Independent estimation of the normalizing constants f_i

In some cases, if particularly accurate estimates of f_i are needed, or if the unbiased probabilities are not required (i.e. if we are only interested in relative free energies), then the f_i values can be estimated directly without recourse to Equation 2. This approach does not require one to create explicit histograms, which is useful if there is concern about the appropriate choice of histogram bin widths for the finite sample sizes. To obtain the f_i values, one can iteratively solve

$$f_i^{-1} = \sum_{j=1}^{S}\sum_{k=1}^{N_j} \frac{\exp(-\beta_i V_{ij})\exp[-(\beta_i - \beta_0)E_{ij}]}{\sum_{l=1}^{S} N_l f_l \exp(-\beta_l V_{lk})\exp[-(\beta_l - \beta_0)E_{lj}]} \tag{15}$$

for the m values of f_i,[4] where V_{ij} and E_{ij} are the values of the biasing potential and unbiased potential energy, respectively, for the j-th sample in the i-th simulation. Although Kumar, et al. do not provide a derivation for Equation 15, the proof is relatively straightforward. One begins by substituting Equation 2 into Equation 3 to obtain

$$f_i^{-1} = \sum_{j=1}^{M} \frac{c_{ij}\sum_{k=1}^{S} n_{kj}}{\sum_{l=1}^{S} N_l f_l c_{lj}}. \tag{16}$$

Next, we imagine that we make the histogram bin sizes increasingly small, so that we reach the limit where each bin has exactly one sample (from all simulations) or none at all, i.e. $\sum_{k=1}^{S} n_{kj}$ is either 0 or 1 for all bins j. Then we can rewrite Equation 16 as

$$f_i^{-1} = \sum_{j=1}^{M} \sum_{k} \frac{c_{ik}}{\sum_{l=1}^{S} N_l f_l c_{lk}},$$

where the sum over k is over all occupied bins. However, the number of occupied bins is simply N_j, and we can replace the biasing factor c_{ik} associated with bin k with the actual values of the bias and potential energy of the k-th sample, which leads to Equation 15.

3 The Effect of Correlations

In several places in the above derivations, assumptions were made about the histogram counts being uncorrelated. In fact, two distinct assumptions were made, namely, that the counts contributing to a given histogram are independent (which is implicit in Equations 8 and 10), and that the S histograms are independent of each other (which is implicit in Equation 11 as well as the assumption that $\langle \delta\Omega_{kj} \delta\Omega_{lj} \rangle = 0$ which was used to arrive at Equation 6). We will now consider these two assumptions individually.

The assumption of independence of histogram counts is clearly not valid for data generated using MC or MD simulations, as samples near each other in time or sequence will be highly correlated with each other. However, it is possible to correct for this effect by reducing the effective sample size, as we will now demonstrate using an argument first suggested by Müller-Krumbhaar and Binder[11] and recently more clearly formulated by Chodera, et al.[12] This strategy is most easily understood in the context of the "traditional" WHAM equation derivation of section 2.1, where the problem is calculating the variance of n_{ij} in the presence of correlations.

Before tackling this problem, let us first consider the more general problem of calculating the

standard error of the mean of a time-ordered sequence of random variables $\{x_1, x_2, \ldots, x_n\}$ which will represent some property of the sequence of samples from a single MD or MC simulation. The mean value of x is simply $\hat{x} = n^{-1}\sum_{i=1}^{n} x_i$, and this is unaffected by the presence of correlations. We are interested in the variance of $\hat{x}$, i.e. the standard error of the mean. If the samples were independent, then $\mathrm{var}(\hat{x}) = \mathrm{var}(x)/n$. The presence of correlations will increase this variance, as we will now calculate. Starting from the definition of the variance, we have

$$\begin{aligned}
\mathrm{var}(\hat{x}) &= \langle \hat{x}^2 \rangle - \langle \hat{x} \rangle^2 \\
&= n^{-2}\left[\left\langle \left(\sum_{i=1}^{n} x_i \right)^2 \right\rangle - \left\langle \sum_{i=1}^{n} x_i \right\rangle^2 \right] \\
&= n^{-2}\left[\sum_{i=1}^{n}\sum_{j=1}^{n} \langle x_i x_j \rangle - \sum_{i=1}^{n}\sum_{j=1}^{n} \langle x_i \rangle \langle x_j \rangle \right] \\
&= n^{-2}\left[\sum_{i=1}^{n} \langle x_i^2 \rangle + \sum_{i \neq j=1}^{n} \langle x_i x_j \rangle - \sum_{i=1}^{n} \langle x_i \rangle^2 - \sum_{i \neq j=1}^{n} \langle x_i \rangle \langle x_j \rangle \right] \\
&= n^{-1}\mathrm{var}(x) + n^{-2} \sum_{i \neq j=1}^{n} (\langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle),
\end{aligned}$$

where in the last step we have assumed that the sequence $\{x_i\}$ is stationary, i.e. that $\langle x_j \rangle = \langle x_i \rangle$ and $\langle x_j^2 \rangle = \langle x_i^2 \rangle$ and for all i and j, which implies that $\sum_{i=1}^{n} \langle x_i \rangle^2 = n \langle x \rangle^2$ and $\sum_{i=1}^{n} \langle x_i^2 \rangle = n \langle x^2 \rangle$. If we further assume that sequence $\{x_i\}$ is time-reversible, then we can write

$$\begin{aligned}
\mathrm{var}(\hat{x}) &= n^{-1}\mathrm{var}(x) + 2n^{-2} \sum_{j>i} (\langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle) \\
&= n^{-1}\mathrm{var}(x) + 2n^{-2} \sum_{t=1}^{n-1} (n-t)(\langle x_i x_{i+t} \rangle - \langle x_i \rangle \langle x_{i+t} \rangle) \\
&= n^{-1}\mathrm{var}(x) + 2n^{-1} \sum_{t=1}^{n-1} (1 - t/n)(\langle x_i x_{i+t} \rangle - \langle x \rangle^2).
\end{aligned}$$

If we further define $g = 1 + 2\tau$, where the correlation time τ is given by

$$\tau = \sum_{t=1}^{n-1} (1 - \frac{t}{n}) \frac{\langle x_i x_{i+t} \rangle - \langle x \rangle^2}{\mathrm{var}(x)},$$

then we can write

$$\mathrm{var}(\hat{x}) = \frac{\mathrm{var}(x)}{n/g}, \tag{17}$$

from which we can see that the effect of correlations is to reduce the effective sample size by a factor of g.

Let us now return to the original problem of finding the variance of n_{ij} in the presence of correlations. To simplify the notation, let us assume that we are only interested in the counts in one bin of one simulation, which will allow us to dispense with subscripts. Define the indicator function

$$\psi_m = \begin{cases} 1 & \text{if the } m\text{-th sample of simulation } i \text{ is in bin } j \\ \\ 0 & \text{otherwise} \end{cases}.$$

Then

$$\begin{aligned} n_{ij} &= \sum_{m=1}^{N_i} \psi_m \\ &= N_i \left(\frac{1}{N_i} \sum_{m=1}^{N_i} \psi_m \right), \end{aligned}$$

i.e. n_{ij}/N_i is the "time average" of ψ_m. Using Equation 17, we obtain $\mathrm{var}(\hat{\psi_m}) = g\,\mathrm{var}(\psi_m)/N_i$. But $n_{ij} = N_i\hat{\psi_m}$, so

$$\begin{aligned} \mathrm{var}(n_{ij}) &= N_i^2 \mathrm{var}(\hat{\psi_m}) \\ &= gN_i(\langle\psi_m^2\rangle - \langle\psi_m\rangle^2). \end{aligned}$$

But ψ_m is a $(0,1)$ indicator function. Therefore, $\psi_m^2 = \psi_m$ and $\mathrm{var}(n_{ij}) = gN_i\langle\psi_m\rangle(1 - \langle\psi_m\rangle)$. However, $\langle\psi_m\rangle$ is the probability of landing in bin j and simulation i. This is the same as the variance of the binomial distribution 8, but with the sample size scaled by g, leading to the modified WHAM equation

$$p_j^\circ = \frac{\sum_{i=1}^{S} g_i^{-1} n_{ij}}{\sum_{i=1}^{S} g_i^{-1} N_i f_i c_{ij}}, \tag{18}$$

where g_i is the scaling factor for the i-th simulation. It should be noted that in general g will also depend on the bin index,[12] though this is often neglected. If the g factors are independent of simulation and bin, as Kumar et al.[4] have argued for simulations that do not involve phase transitions, then they cancel out, leaving the optimal estimate of p_j° unchanged from what would be obtained in the absence of correlations.

The second assumption, that the simulations are independent, may or may not be reasonable, depending on the simulations being analyzed. If separate simulations are performed using different biasing potentials, then this assumption is obviously reasonable. In the case of temperature biasing in replica exchange simulations, this may not be reasonable. This is because samples remain conformationally and energetically correlated even though temperatures have been exchanged. Therefore, two highly correlated samples could contribute to the same bin in two different histograms. The proper treatment of this correlation is non-trivial, though at least one attempt have been made in the literature.[12]

4 Estimation of errors in the unbiased probabilities

The maximum likelihood formulation of the WHAM equations lend themselves naturally to a Bayesian interpretation. The product of multinomials (Equation 11) can be viewed not only as a likelihood function, but also as the Bayesian posterior probability distribution of p_j° under the assumption of a uniform prior:[3]

$$P(p_1^\circ, \ldots, p_M^\circ | \text{bin counts}) \propto \prod_{i=1}^{S} \prod_{j=1}^{M} (f_i c_{ij} p_j^\circ)^{n_{ij}}. \tag{19}$$

We can rearrange the factors in Equation 19 to rewrite the posterior probability as

$$\begin{aligned} P(p_1^\circ, \ldots, p_M^\circ | \text{bin counts}) &\propto \prod_{i=1}^{S} f_i^{N_i} \prod_{j=1}^{M} c_{ij}^{n_{ij}} p_j^{\circ n_{ij}} \\ &= \left(\prod_{i=1}^{S} f_i^{N_i} \right) \left(\prod_{i=1}^{S} \prod_{j=1}^{M} c_{ij}^{n_{ij}} \right) \left(\prod_{i=1}^{S} \prod_{j=1}^{M} p_j^{\circ n_{ij}} \right) \\ &\propto \left[\prod_{i=1}^{S} \left(\sum_{k=1}^{M} c_{ik} p_k^\circ \right)^{-N_i} \right] \left(\prod_{j=1}^{M} p_j^{\circ \nu_j} \right), \end{aligned} \tag{20}$$

where $\nu_j = \sum_{i=1}^{S} n_{ij}$ and where we have absorbed the factors independent of p_j° into the proportionality constant. The second factor in Equation 20 is a Dirichlet distribution, for which standard random number generators are available. We can use draws from such a generator as MC sampling proposals to be accepted or rejected using a Metropolis-type rule in order to sample from the posterior.[3] Once such a sample has been generated, any moments of interest can be easily obtained. Correlations can be accounted for by reducing n_{ij} by g as described above. This will increase the width of the posterior, reflecting the effective loss of information due to correlations.

References

1. Benoît Roux. The calculation of the potential of mean force using computer simulations. *Comp. Phys. Comm.*, 91:275–282, 1995.

2. Dmitriy S. Chekmarev, Tateki Ishida, and Ronald M. Levy. Long-time conformational transitions of alanine dipeptide in aqueous solution: Continuous and discrete-state kinetic models. *J. Phys. Chem. B*, 108:19487–19495, 2004.

3. E. Gallicchio, M. Andrec, A. K. Felts, and R. M. Levy. Temperature weighted histogram analysis method, replica exchange and transition paths. *J. Phys. Chem. B*, 109:6722–6731, 2005.

4. Shankar Kumar, Djamal Bouzida, Robert H. Swendsen, Peter A. Kollman, and John M. Rosenberg. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comp. Chem.*, 13:1011–1021, 1992.

5. Alan M. Ferrenberg and Robert H. Swendsen. Optimized Monte Carlo data analysis. *Phys. Rev. Lett.*, 63:1195–1198, 1989.

6. Alan M. Ferrenberg and Robert H. Swendsen. Optimized Monte Carlo data analysis. *Computers in Physics*, Sept/Oct:101–104, 1989.

7. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Cambridge, 1992.

8. Christian Bartels and Martin Karplus. Multidimensional adaptive umbrella sampling: Application to main chain and side chain peptide conformations. *J. Comp. Chem.*, 18:1450–1462, 1997.

9. J. C. Kiefer. *Introduction to Statistical Inference*. Springer-Verlag, New York, 1987.

10. S. M. Ross. *Introduction to Probability Models*. Harcourt Academic Press, San Diego, 7th ed. edition, 2000.

11. H. Müller-Krumbhaar and K. Binder. Dynamic properties of the Monte Carlo method in statistical physics. *J. Stat. Phys.*, 8:1–24, 1973.

12. John D. Chodera, William C. Swope, Jed W. Pitera, Chaok Seok, and Ken A. Dill. Use of the weighted histogram analysis method for the analysis of simulated and parallel tempering simulations. *J. Chem. Theory Comput.*, 3:26–41, 2007.