1 Multiple histogram method

We wish now to combine results obtained at different values of β. In the simplest setting, imagine that you have performed two runs at β_0 and β_1, that are reasonably close so that the energy distributions for the two values of β essentially overlap. Using the reweighting method discussed before we can:

1) use the data at β_1 to compute averages at β_0 and viceversa;

2) use the data at β_1 to compute averages at any β that is between β_0 and β_1 ($h(E,\beta)$ will be "in the middle", between $h(E,\beta_0)$ and $h(E,\beta_1)$, and hence it will overlap with both of them).

In this way, for any β between β_0 and β_1, and also for these two values, we have two different estimates, one obtained from the run at β_0, one for the run at β_1. The question now is: *How do we combine them, to obtain a more precise estimate?*

Another imteresting problem is the following. Suppose we need an estimate of a given observable A for *any* β in a relatively large interval $[\beta_a,\beta_b]$. How can we adress this problem? The idea here is to consider a set of β values $\beta_a=\beta_1<\beta_2<\ldots\beta_R=\beta_b$ (R is the number of β values) such that the energy distribution at β_i overlaps with the energy distributions at β_{i-1} and β_{i+1}. Then, one performs numerical simulations at all values β_i. Now, the question is: *How do we combine the numerical data, to obtain an estimate at a given value of β that belongs to the interval $[\beta_a,\beta_b]$?*

The most naive method would consist in performing a weighted average of the reweighted data. To explain the shortcomings of this approach, let us assume $R=2$ and, for instance, let us consider a value of β between β_1 and β_2 which is closer to β_1 than to β_2. A formally correct strategy to compute an average $\langle A\rangle_\beta$ could be the following. We first use the data at β_1 to obtain an estimate A_1 with error σ_1 and then the data at β_2 to obtain an estimate A_2 with error σ_2. Finally, one could combine the two estimates as

$$A_{12}=\frac{A_1\sigma_1^{-2}+A_2\sigma_2^{-2}}{\sigma_1^{-2}+\sigma_2^{-2}}\ .$$

However, since β is not close to β_2, A_2 has a somewhat large error; but, what is worse, also the error estimate σ_2 has a somewhat large error. Hence, σ_2 as estimated from the data could be largely underestimated. But, if σ_2 is largely underestimated, we would give too much weight to A_2, adding essentially noise and not signal to A_1. In these cases A_{12} would be a worst estimate than A_1.

We now discuss a more robust method, the multihistogram method of Ferrenberg and Swendsen. As before, we consider a system with discrete phase space. The basic quantity of the method is the energy histogram

$$h(E_0,\beta)=\langle\delta_{H(x),E_0}\rangle_\beta=\frac{1}{Z_\beta}\sum_E\rho(E)\delta_{E,E_0}e^{-\beta E}=\frac{1}{Z_\beta}\rho(E_0)e^{-\beta E_0}\ .$$

where $\rho(E)$ is the density of states (for a discrete system it represents the number of states with a given energy). It satisfies

$$\sum_E h(E,\beta)=1. \tag{1}$$

From the computation of $h(E,\beta)$ we can estimate $\rho(E)$ using

$$\rho(E)=Z_\beta e^{\beta E}h(E,\beta).$$

It is important to realize that $h(E,\beta) \sim 1/\sqrt{N}$. Indeed, $h(E,\beta)$ is roughly a Gaussian centered around $\langle E\rangle_\beta$ with a variance of order N. Because of the normalization condition (1), we have

$$h(E,\beta) = \frac{1}{\sqrt{\pi a N}} e^{-(E-\langle E\rangle_\beta)^2/(aN)}.$$

Let us now compute the variance associated to $h(E,\beta)$. We have

$$\mathrm{var}\left[h(E_0,\beta)\right] = \langle \delta^2_{H(x),E_0}\rangle_\beta - \langle \delta_{H(x),E_0}\rangle^2_\beta = h(E_0,\beta)[1-h(E_0,\beta)] \approx h(E_0,\beta)\ ,$$

where we have used the obvious property $\delta^2_{H(x),E_0} = \delta_{H(x),E_0}$ and, in the last step, that $h(E_0,\beta) \sim 1/\sqrt{N} \ll 1$.

We can now define the method. We have performed simulations at $\beta_1 < \beta_2 \ldots < \beta_R$. Suppose we have taken n_i *independent* measurements[1] at β_i and let us denote with $N_i(E)$ the number of measures with energy E at this value of the inverse temperature. The ratio $N_i(E)/n_i$ is an estimator of the histogram variable $h(E,\beta_i)$. Indeed, assume that $E_i^{(1)}$, ..., $E_i^{(n_i)}$ are the energy estimates obtained in the run at β_i. The estimator of $h(E,\beta_i)$ is the sample mean of $\delta_{E,H(x)}$, i.e.

$$\frac{1}{n_i}\sum_{\alpha=1}^{n_i} \delta_{E,E_i^{(\alpha)}}$$

Now, the δ term gives one only if $E = E_i^{(\alpha)}$, i.e., if the measurement at time α is E. Summing over α we obtain $N_i(E)$, the number of measures with energy E .

Using the data at β_i we can therefore estimate $\rho(E)$ as

$$\rho_i(E) \approx n_i^{-1} N_i(E) e^{\beta_i E} Z_i\ ,$$

where Z_i, the partition function at β_i, is however not known. Note that we have appended an index i to $\rho_i(E)$ only to specify that this is the estimate obtained from the data computed at β_i. Indeed, the density of states is a property of the state space, not of the probability distribution (therefore it does not depend on β_i).

The variance of the estimator of $\rho_i(E)$ can be computed if one assumes that the partition function Z_i is known. Indeed, with this assumption

$$\begin{aligned} \sigma_i^2(E) &= n_i^{-1}\mathrm{var}\left[\rho(E)\right] = n_i^{-1} e^{2\beta_i E} Z_i^2 \mathrm{var}\left[h(E,\beta_i)\right] = \\ &= n_i^{-1} e^{2\beta_i E} Z_i^2 h(E,\beta_i) = n_i^{-1} e^{\beta_i E} Z_i \rho(E)\ . \end{aligned} \tag{2}$$

If we would really want to compute the error, we should replace $\rho(E)$ in the last term with its estimator ρ_i. Since we know that this estimator may be very imprecise—it provides an accurate estimate of $\rho(E)$ only if E is a typical energy at inverse temperature β_i—we do not do it here. This is a crucial point in the method and it is the one that guarantees the robustness of the results. It is also important to stress that σ_i is more properly related to the error on $\rho_i(E)/Z_i$, the only quantity that we can compute by simulations. This is not relevant in practive as we will only use σ_i to write down a weighted average of the estimators $\rho_i(E)$.

[1] If measures are correlated with an autocorrelation time τ_i, then $n_i = \tilde{n}_i/(2\tau_i)$, where $\tilde{n}_i$ is the total number of measurements. Analogously $N_i(E)$ is the total number of times we obtained an energy E, divided by the autocorrelation time τ_i. In the following we always assume measures to be independent for simplicity.

A robust estimate of the density of states using all R datasets is given by a weighted average, where each estimate $\rho_i(E)$ enters with a weight proportional to $1/\sigma_i^2(E)$:

$$\begin{aligned}\rho(E) &= \frac{\sum_{i=1}^R \rho_i(E)/\sigma_i^2(E)}{\sum_{j=1}^R 1/\sigma_j^2(E)} = \frac{\sum_{i=1}^R (N_i/n_i)\; e^{\beta_i E} Z_i\, n_i e^{-\beta_i E}/(Z_i\rho(E))}{\sum_{j=1}^R n_j e^{-\beta_j E}/(Z_j\rho(E))} = \\ &= \frac{\sum_{i=1}^R N_i(E)}{\sum_{j=1}^R n_j\, e^{-\beta_j E}\, Z_j^{-1}}\;. \end{aligned} \tag{3}$$

Note that $\rho(E)$ in the right-hand side simplifies (it appears both in the numerator and in the denominator). At this point it is important to stress a very important difference between this method and the naive method presented at the beginning. Here, we do not use estimates of the errors (that may very imprecise). We use instead an expression that follows from the exact result for the variance of $h(E,\beta)$ and do not replace $\rho(E)$ in the errors with its estimate. This is what makes the final expression robust. Second, observe that for any given E, the only runs that contribute to the determination of $\rho(E)$ are those for which $N_i(E) \neq 0$. This means that we are using the data at β_i only where they are relevant.

Eq. (3) still depends on the unknown partition functions Z_i. They can be determined in a self-consistent way by noting that

$$Z_k = \sum_E \rho(E) e^{-\beta_k E} = \sum_E \frac{\sum_{i=1}^R N_i(E)}{\sum_{j=1}^R n_j\, e^{(\beta_k-\beta_j)E}\, Z_j^{-1}}\;. \tag{4}$$

which can be rewritten as

$$\sum_E \frac{\sum_{i=1}^R N_i(E)}{\sum_{j=1}^R n_j\, e^{(\beta_k-\beta_j)E}\, (Z_k/Z_j)} = 1\;. \tag{5}$$

The consistency condition gives us R equations (one for each value of k) for the partition function ratios Z_i/Z_j. However, there are only $R-1$ independent partition function ratios. For instance, all ratios can be expressed in terms of $z_j = Z_j/Z_1$, $(j = 2,\ldots,R)$ as

$$\frac{Z_j}{Z_k} = \frac{Z_j}{Z_1}\frac{Z_1}{Z_k} = \frac{z_j}{z_k}.$$

Apparently, the consistency system appears to be overdefined (there are more equations than unknowns). But this is not the case, since the R equations are linearly dependent. To clarify this point, let us define

$$f_k = \sum_E \frac{\sum_{i=1}^R N_i(E)}{\sum_{j=1}^R n_j\, e^{(\beta_k-\beta_j)E}\, (Z_k/Z_j)};$$

this is the left-hand side of equation (5). The consistency equations become $f_k = 1$. Now let us compute

$$\begin{aligned}\sum_{k=1}^R n_k f_k &= \sum_{k=1}^R n_k \sum_E \frac{\sum_{i=1}^R N_i(E)}{\sum_{j=1}^R n_j\, e^{(\beta_k-\beta_j)E}\, (Z_k/Z_j)} \\ &= \sum_E \sum_{k=1}^R n_k (e^{-\beta_k E}/Z_k) \frac{\sum_{i=1}^R N_i(E)}{\sum_{j=1}^R n_j\, e^{-\beta_j E}/Z_j} \\ &= \sum_E \sum_{i=1}^R N_i(E) \frac{\sum_{k=1}^R n_k e^{-\beta_k E}/Z_k}{\sum_{j=1}^R n_j\, e^{-\beta_j E}/Z_j} = \sum_E \sum_{i=1}^R N_i(E) = \sum_{i=1}^R \sum_E N_i(E) = \sum_{i=1}^R n_i\;. \end{aligned} \tag{6}$$

It is now easy to verify that the system of equations is not overdetermined. Suppose we use the equations $f_1 = 1$, ..., $f_{R-1} = 1$, to determine the ratios z_i. Let us now show that the equation $f_R = 1$ is automatically satisfied. Indeed, the previous relation gives

$$f_R = \frac{1}{n_R}\left(\sum_{i=1}^{R} n_i - \sum_{i=1}^{R-1} f_i n_i\right) = \frac{1}{n_R}\left(\sum_{i=1}^{R} n_i - \sum_{i=1}^{R-1} n_i\right) = \frac{1}{n_R} \times n_R = 1$$

Note that the consistency equations do not provide the partition functions but only the ratios of the partition function, so that we are only able to estimate $\rho(E)$ up to a constant that is independent of E. In practice we only compute the ratio $\rho(E)/Z_i$ for any i. However, this is enough to compute averages of functions of the energy since

$$\langle g(E)\rangle_{\beta_i} = \frac{1}{Z_i}\sum_x g(H(x))e^{\beta_i H(x)} = \sum_E g(E)e^{-\beta_i E}[\rho(E)/Z_i] \ .$$

or ratios of partition functions

$$\frac{Z_\beta}{Z_i} = \sum_E e^{-\beta E}[\rho(E)/Z_i] \ .$$

At a generic β we can use

$$\langle g(E)\rangle_\beta = \frac{\sum_E g(E)e^{-\beta E}[\rho(E)/Z_i]}{\sum_E e^{-\beta E}[\rho(E)/Z_i]}.$$

The procedure in practice works as follows. Using the standard reweighting method we compute the ratios Z_i/Z_{i-1} by means of a single simulation at β_i or β_{i-1}. Then, we obtain the solution of the R consistency equations by minimizing

$$\sum_{k=1}^{R}(f_k - 1)^2$$

starting the minimization from the estimates obtained using the reweighting method. Once the ratios of partition functions are known, we can compute any average of functions of the energy for any β in the relevant interval $[\beta_a, \beta_b]$.

The procedure we have discussed works for discrete systems. We will now rewrite the expressions in the different form, that also works for continuous systems, which avoids the use of $N_i(E)$, which is well defined only in systems with discrete energy values. Let $E_i^{(\alpha)}$ be the measurements at β_i (α runs from 1 to n_i). Let us consider for instance f_k that defines the consistency equations. We rewrite it as

$$f_k = \sum_E \sum_{i=1}^{R} N_i(E)s(E) \qquad s(E) = \left[\frac{1}{\sum_{j=1}^{R} n_j \, e^{(\beta_k - \beta_j)E} \, (Z_k/Z_j)}\right].$$

Now, we can equally write it as

$$f_k = \sum_{i=1}^{R}\sum_{\alpha=1}^{n_i} s[E_i^{(\alpha)}]. \tag{7}$$

since, for each i, we have

$$\sum_{\alpha=1}^{n_i} s[E_i^{(\alpha)}] = \sum_E N_i(E)s(E).$$

In the form (7) the expression is also valid for continuous systems. The same procedure can be applied to the computation of mean values: the sums over E are replaced by sums over all measures and $N_i(E)$ is simply replaced by 1.

The procedure we have presented can be generalized to allow us to compute averages of generic observables $A(x)$. In this case, the basic quantity is the joint histogram with respect to E and A

$$h(E_0, A_0, \beta) = \langle \delta_{E,E_0} \delta_{A,A_0} \rangle_\beta \ ,$$

its estimator $N_i(E_0, A_0)/n_i$, and the density of states $\rho(E_0, A_0)$ which counts the number of states such that $E = E_0$ and $A = A_0$. Repeating the same steps as before, we end up with

$$\rho(E, A) = \frac{\sum_{i=1}^{R} N_i(E, A)}{\sum_{j=1}^{R} n_j \, e^{-\beta_j E} \, Z_j^{-1}} \ .$$

Once $\rho(E, A)/Z_i$ is known, any average involving E and A can be directly computed.