1 Simulated tempering

As the umbrella sampling (US) method, also the simulated tempering (ST) method aims at sampling the configurations that are typical at a set of inverse temperatures $\beta_1 < \ldots < \beta_R$ and, indeed, it represents a stochastic version of the US method. In the ST case, one enlarges the configuration space by adding an index i which runs from 1 to R. Hence, a configuration in the ST simulation is a pair (x, i). Configurations are sampled with probability ($\alpha_i > 0$)

$$\Pi(x, i) = \frac{\alpha_i e^{-\beta_i H}}{Z_\pi} \ , \tag{1}$$

where Z_π is the same partition function we considered in the US case. Indeed, the condition

$$\sum_i \sum_x \Pi(x, i) = 1$$

implies

$$1 = \sum_{ix} \frac{\alpha_i e^{-\beta_i H}}{Z_\pi} = \sum_i \alpha_i \frac{Z_i}{Z_\pi} \Rightarrow Z_\pi = \sum_i \alpha_i Z_i.$$

The ST and the US method are essentially equivalent. The simulated tempering method is simply a stochastic version of the umbrella sampling method. For instance, consider a variable $A(x)$. If π is the US probability distribution, the umbrella sampling average of A is

$$\langle A \rangle_{US} = \sum_x \pi(x) A(x) = \frac{1}{Z_\pi} \sum_{xi} \alpha_i e^{-\beta_i H} A(x)$$

while the ST average is

$$\langle A \rangle_{ST} = \sum_{x,i} \Pi(x, i) A(x) = \frac{1}{Z_\pi} \sum_{xi} \alpha_i e^{-\beta_i H} A(x)$$

The two expressions are clearly the same.

The simulated tempering dynamics is usually thought as a dynamics in which the temperature of the systam varies. If the configuration is (x, i), one thinks of the configuration as the system x at inverse temperature β_i. Analogously, transitions $(x, i) \to (x, j)$, are interpreted as temperature changes, from β_i to β_j.

Since the US method and the ST method are equivalent, the ST method works only if the conditions we mentioned for the US method are satisfied:

- The temperatures should be finely spaced, so that typical configurations at inverse temperature β_i overlap with those at $\beta_{i\pm1}$.
- The coefficients α_i should satisfy

$$\frac{\alpha_i}{\alpha_j} \approx \frac{Z_j}{Z_i} = e^{\beta_i F(\beta_i) - \beta_j F(\beta_j)} \ . \tag{2}$$

  The algorithm is correct, though not optimal, for any choice of the α_i's,

Again, we should also specify how to obtain canonical averages from ST simulations. As we already proved that ST and US averages are the same, we can use

$$\langle A\rangle_\beta = \frac{\langle A\, e^{-\beta H}/f_\pi\rangle}{\langle e^{-\beta H}/f_\pi\rangle}\ , \tag{3}$$

where the mean values in the right-hand side are ST averages and

$$f_\pi(x) = \sum_i \alpha_i e^{-\beta_i H(x)}$$

so that $\sum_i \Pi(x,i) = f_\pi(x)/Z_\pi$.
However, in ST simulations one often uses a different formula to compute average values at the inverse temperatures β_i. Define the *indicator* function

$$I_i(x,j) = \delta_{ij}.$$

Then, we consider

$$\langle AI_i\rangle = \sum_{xj} \Pi(x,j)A(x)I_i(x,j) = \frac{1}{Z_\pi}\sum_{xj}\alpha_j e^{-\beta_j H}A(x)\delta_{ij} = \frac{1}{Z_\pi}\alpha_i\sum_x e^{-\beta_i H}A(x) = \frac{Z_i\alpha_i}{Z_\pi}\langle A\rangle_{\beta_i}$$

If we set $A(x)=1$, since $\langle A\rangle_{\beta_i}=1$, we have

$$\langle I_i\rangle = \frac{Z_i}{Z_\pi}\alpha_i \tag{4}$$

These two relations imply

$$\langle A\rangle_{\beta_i} = \frac{\langle AI_i\rangle}{\langle I_i\rangle}.$$

Let us understand the meaning of Eq. (4). If we work in the simulated tempering ensemble, configurations are (x,i). In the Monte Carlo simulation we update both x, the configuration in the state space, and the index, that is the inverse temperature. So, if we intend to measure averages of $A(x)$, we save, at constant intervals, both the index (the temperature) and the value of the observable. Therefore, we collect the measures (i_1,A_1), (i_2,A_2), …, (i_n,A_n), where A_k is the value of A computed on the configuration at the k-th measure, while i_k is the value of the index when we performed the k-th measure. Here n is the total number of measures. Now, let us estimate $\langle I_i\rangle$. As usual, the estimator is the sample mean:

$$\overline{I_i} = \frac{1}{n}\sum_{k=1}^{n}\delta(i,i_k)$$

Now the sum simply counts how many measures were performed with $i=i_k$, so that

$$\overline{I_i} = \frac{1}{n}\times(\text{number of measures performed at }\beta_i) = \frac{N_i}{n}.$$

Using Eq. (4), we see that the optimality condition $\alpha_i Z_i \approx \alpha_j Z_j$ can be rephrased in the requirement that the number of measures N_i performed at each β_i is essentially the same: during the simulation the system equally visits all temperatures.

In an analogous way we can understand the meaning of $\langle AI_i \rangle$. The estimator is

$$\overline{AI_i} = \frac{1}{n} \sum_{k=1}^{n} A_k \delta(i, i_k)$$

The sum in the right-hand side is the sum of the measures A_k that correspond to systems that are at temperature β_i (there are N_i measures at temperature β_i). We call this sum $S_i(A)$, so that

$$\overline{AI_i} = \frac{1}{n} S_i(A).$$

It follows

$$\langle A \rangle_{\beta_i} \approx \frac{\frac{1}{n} S_i(A)}{\frac{1}{n} N_i(A)} = \frac{S_i(A)}{N_i(A)}$$

This formula has a simple interpretation: the average at β_i is the usual sample average in which we only consider the data collected at the inverse temperature β_i. There is only one caveat: while the average is the usual one, one cannot use the usual formulae for the errors, since here the number of measures N_i is a stochastic variable. Errors should be computed using a robust method, like the jackknife method.

2 Implementation

In a simulated tempering simulation we must update both x and i. Let us see how to implement a Metropolis simulation.

Update of x. We keep i fixed and propose $x \to y$ with proposal matrix $P^{(0)}_{xy}$. We accept the move with probability $\min(1, R_{xy})$, where

$$R_{xy} = \frac{\Pi(y,i) P^{(0)}_{yx}}{\Pi(x,i) P^{(0)}_{xy}} = e^{-\beta_i (H(y) - H(x))} \frac{P^{(0)}_{yx}}{P^{(0)}_{xy}}.$$

The acceptance is the same we use in the case of canonical simulations, setting β equal to the *current* inverse temperature.

Update of i. We keep x fixed and change i. We typically only propose moves $i \to i \pm 1$. We have already discussed this algorithm in `Lesson 8: Metropolis. Examples (I)`. The proposal is

$$\begin{aligned}
&P^{(0)}_{i,i+1} = 1/2 \qquad 2 \le i \le R-1 \\
&P^{(0)}_{i,i-1} = 1/2 \qquad 2 \le i \le R-1 \\
&P^{(0)}_{1,2} = 1 \\
&P^{(0)}_{R,R-1} = 1
\end{aligned}$$

The acceptance is $\min(1, R_{i,i\pm1})$ with

$$R_{i,i\pm1} = \frac{P^{(0)}_{i\pm1,i} \Pi(x, i \pm 1)}{P^{(0)}_{i,i\pm1} \Pi(x,i)} = \frac{P^{(0)}_{i\pm1,i}}{P^{(0)}_{i,i\pm1}} \exp[-(\beta_{i\pm1} - \beta_i) H(x)].$$

Note that the ratio of the proposal matrices is 1, except when i or $i \pm 1$ correspond to the boundary values (see the discussion in Lesson 8, where a different, less efficient updating method is also discussed).

The full algorithm. The full algorithm is obtained by performing both updates of x and updates of i. The relative frequency of the two moves is a free parameter of the algorithm, that can be optimized to achieve the smallest errors.

It is usually a good idea to collect the transition frequencies $a_{i,i\pm1}$ that give the probability that a transition $i \to i \pm 1$ is performed. One would like the transition frequencies to vary between 0.2 and 0.5. If larger numbers are observed, one can decrease the number of temperatures; if smaller numbers are observed, it is possible that the parameters α_i are not optimized or that the β intervals are too wide, so that the energy distributions have little overlap. In this case it is better to review the parameters that have been chosen.