1. What is the purpose of MBAR?

(1) To estimate the difference in dimensionless free energies

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

(2) To estimate equilibrium expectations of some observable (energy, volume, pair distances, etc.) of the coordinates O(x).

$$\langle O \rangle_i = \int_\Gamma d\mathbf{x} p_i(\mathbf{x}) O(\mathbf{x}) = \frac{\int_\Gamma d\mathbf{x} q_i(\mathbf{x}) O(\mathbf{x})}{\int_\Gamma d\mathbf{x} q_i(\mathbf{x})}$$

2. What is the potential of mean force, and how does it differ from a free energy surface?

Kin-Yiu Wong, and Darrin M. York, Exact Relation between Potential of Mean Force and Free-Energy Profile, *Journal of chemical theory and computation*, 2012, 8, 3998-4003 (https://pubs.acs.org/doi/10.1021/ct300392f) (Quantum scenario)

3. What other methods/estimators have a similar purpose/use to MBAR?

(1) Compared with WHAM: WHAM reduces to MBAR as bin goes to zero width.

(2) Compared with BAR: MBAR reduces exactly to Bennett's acceptance ratio formula in the case of only two states.

4. What advantages does MBAR offer over other methods/estimators?

(1) Compared with WHAM

(A) WHAM rely on histograms of width sufficient to contain many samples, introducing a bias that can be substantial and often difficult to access.

(B) WHAM: difficult to compute error estimates in the free energies or ensemble averages obtained. (MBAR uses no histograms, estimating quantities using the samples directly, ad thus has no histogram bias.)

(2) Compared with BAR

(A) BAR does not have provably lowest variance for the free energy calculation between two states.

(B) It cannot be used to directly compute free energies using samples for multiple states all together (only in pairs), and cannot be used to compute ensemble averages in its standard formulation.

5. How does MBAR differ, in approach, from other popular estimators that incorporate samples from multiple equilibrium states?

6. Identify a current research endeavor in which you could apply MBAR, and outline the steps to achieve this.

7. Are there any practical considerations that one must make before using MBAR? (ie, What sampling level and/or number of equilibrium states are required to achieve convergence of the estimator?)

8. What specific equation(s) are solved during execution of MBAR? What are the known and unknown variables? Are the unknown variables coupled to or independent of one another?

9. Why does MBAR assume Boltzmann statistics? Can you think of a scenario where we would want to apply a non-Boltzmann distribution function?

10. Why do we use the Newton-Raphson solver with MBAR? What does this solver do?

Newton's method is guaranteed to converge under certain conditions. One popular set of such conditions is this: if a function has a root and has a non-zero derivative at that root, and it's continuously differentiable in some interval around that root, then there's some neighborhood of

the root so that if we pick our starting point in that region, the iterations will converge to the given root.

∗ When does it fail?

(1) Bad starting point that lies outside to range of guaranteed convergence

(2) The chosen starting point has a zero derivative.

(3) The root has a zero derivative → converge extremely slow

11. Is MBAR designed to be implemented with correlated time-series data (obtained from molecular dynamics), decorrelated data (ie, umbrella sampling), neither of these, or both of these?

12. What is the asymptotic covariance matrix? What method is applied to compute the covariance matrix in MBAR? What information can we obtain from the covariance matrix?

13. How can we prove that MBAR the lowest variance unbiased estimators of both free energies and ensemble averages? And how we can mathematically prove the relationship of MBAR with BAR and WHAM?

14. **What is extended bridge sampling estimators?**

Reweighting from the mixture distribution as the better way to describe the MBAR

Things I kind of get… Extracting partition functions from each state to create a partition function over the ensemble of sampled states (All simulations)… Reweight observables with data from all sampled partition functions?

When does free energy come into this?

I don't like the the whole 1 math thing used in to show how they system of MBAR equations…

$$c_i \langle \alpha_{ij} q_j \rangle_i = \int_{\mathbf{\Gamma}} d\boldsymbol{x}\, q_i(\boldsymbol{x})\, \alpha_{ij}(\boldsymbol{x})\, q_j(\boldsymbol{x}) = c_j \langle \alpha_{ij} q_i \rangle_j \quad (5)$$

What is a verbal explanation of this? The ratios of partition functions from each state is equal to the ratios of some function α_{ij} times the un-normalized probability function of state j averaged in state i? What is the alpha?

1. **How can we prove that MBAR the lowest variance unbiased estimators of both free energies and ensemble averages? And how we can mathematically prove the relationship of MBAR with BAR and WHAM?**