Equilibrium Free Energies from Nonequilibrium Measurements Using Maximum-Likelihood Methods

Michael R. Shirts
Department of Chemistry, Stanford University, Stanford, California 94305-5080, USA

Eric Bair and Giles Hooker
Department of Statistics, Stanford University, Stanford, California 94305-4065, USA

Vijay S. Pande
Department of Chemistry, Stanford University, Stanford, California 94305-5080, USA
(Received 27 June 2003; published 2 October 2003)

We present a maximum likelihood argument for the Bennett acceptance ratio method, and derive a simple formula for the variance of free energy estimates generated using this method. This derivation of the acceptance ratio method, using a form of logistic regression, a common statistical technique, allows us to shed additional light on the underlying physical and statistical properties of the method. For example, we demonstrate that the acceptance ratio method yields the lowest variance for *any* estimator of the free energy which is unbiased in the limit of large numbers of measurements.

DOI: 10.1103/PhysRevLett.91.140601 PACS numbers: 05.10.–a, 05.70.Ce

Introduction.—Finding the free energy difference between different states of a physical system is of great general interest in many scientific fields, including drug design [1], basic statistics [2], and even nonperturbative quantum chromodynamics [3]. It is of interest to the experimental community as well as the theoretical and computational communities [4]. Recently, there has been increased interest in determining the uncertainty and bias in any attempt to extract free energies from a suitable set of data [2,5–11].

We can separate the calculation of precise and accurate free energy differences into two nonoverlapping problems. First, we must generate a number, n, of statistically uncorrelated measurements of the system. Second, we must extract a free energy estimate from these n measurements, along with reliable estimates for the statistical bias and variances of our estimate. We will assume that we are already in possession of a set of n uncorrelated measurements of the proper observable for our method, and address only the statistical issues related to the extraction of free energy estimates from these measurements.

There are a variety of commonly used methods for finding the free energy of a physical change in the system. Many of these can be expressed finding the equilibrium free energy from nonequilibrium work distributions. Thermodynamic perturbation theory (TPT) or free energy perturbation (FEP) estimates free energy differences by exponentially averaging potential energy differences between a reference state sampled at equilibrium and a target state [12]. However, FEP can be seen as a special case of "fast growth" [5,6] nonequilibrium exponential work averaging, as the energy difference is the infinitely fast adiabatic work of transition between the two states [13,14]. "Slow growth" thermodynamic integration has been shown to have high intrinsic biases and is unreliable as originally implemented [15,16]. However, the "free energies" obtained from these simulations are actually measurements from a nonequilibrium work distribution and an ensemble of values can therefore also be used to obtain correct free energies [6]. The rest of this Letter will focus on the generalized nonequilibrium problem.

Assume there are two different equilibrium states defined on a phase space by energy functions $U_0(\vec{q})$ and $U_1(\vec{q})$. Let ΔF be the free energy between these states, defined as the log of the ratio of the partition functions associated with $U_0(\vec{q})$ and $U_1(\vec{q})$. We can associate a work with the process of changing energy functions from U_0 to U_1 or vice versa while the system is maintained in temperature equilibrium with the surroundings. By sampling initial conditions from equilibrium, we obtain a distribution in either direction of such work values. For infinitely fast switching, these distributions are simply of $\pm\Delta U = \pm(U_1 - U_0)$ canonically sampled from the initial state.

It has long been known that the exponential average of equilibrium energy differences between two states yields the free energy difference between the states [12]. More recently, Jarzynski demonstrated that distribution of nonequilibrium work values can yield an equilibrium free energy by taking the exponential average of the set of nonequilibrium work values [13]. However, the exponential average of a set of data $X = \{x_i, \ldots, x_n\}$, defined as $-(1/\beta)\ln\langle\exp(-\beta X)\rangle$ (where $\beta = 1/kT$), is a statistic that is both inherently noisy and biased, even if the spread of the data is only moderately larger than kT. The results of exponential averaging strongly depend on the behavior at the tails of the distribution, which, by definition, are not as well sampled as the rest of the distribution. Previous studies have explored and demonstrated the

 0031-9007/03/91(14)/140601(4)$20.00

poor behavior of exponential averaging for small sample sizes [7–10].

In an examination of free energy estimation between two states sampled at equilibrium, Bennett [17] demonstrated that it is possible to use the information contained in both the forward and reverse distributions of the potential energy difference together in a manner which was significantly better than estimates using only energy difference data in one direction. This derivation can trivially be generalized to the nonequilibrium work case, replacing ΔU with the nonequilibrium work [14]. Bennett, in the FEP case, and Crooks, in the general case, showed that the equation:

$$\exp(-\beta\Delta F) = \frac{\langle f(W)\rangle_F}{\langle f(-W)\exp(-\beta W)\rangle_R} \qquad (1)$$

is true for any function $f(W)$, where we define the measurement from the initial state to the final state as the "forward" direction (denoted by the subscript F) and the measurement from the final state to the initial state as the "reverse" direction (denoted by the subscript R).

Bennett then minimized the statistical variance with respect to this function $f(W)$ to find that $f(W) = [1 + n_F/n_R \exp(\beta(W - \Delta F)]^{-1}$ minimizes the variance in this free energy estimate, where n_F and n_R are the number of simulations in the forward and reverse direction, respectively. The free energy difference ΔF can easily be found by iterative methods [17]. Although this method is known and referenced in the literature, it is strangely seldom used in practice despite its theoretical advantages.

We have found that it is possible to derive this acceptance ratio method motivated by an entirely different reasoning—the maximum likelihood of the free energy difference given a set of work measurements. This derivation has theoretical advantages. Maximum likelihood estimators are particularly well-behaved estimators, as they can be shown under relatively weak conditions to be asymptotically efficient, meaning that no other asymptotically unbiased estimator can have lower variance [18]. The term "asymptotically unbiased" means that the estimate becomes unbiased as the number of measurements goes to infinity. The acceptance ratio method is therefore the *best* asymptotically unbiased estimate possible given a set of nonequilibrium work values in both directions.

Maximum likelihood estimator derivation.—We start from the fact that [14]:

$$\ln\left[\frac{P_F(W)}{P_R(-W)}\right] = \beta(W - \Delta F), \qquad (2)$$

where $P_F(W)$ and $P_R(W)$ are probability distributions for the work of nonequilibrium processes from the two states in opposing directions. In order to simplify the notation, and without loss of generality, we will replace the reverse work distribution with the equivalent distribution formed by substituting $-W$ for W.

Equation (2) can be recognized as a specific case of logistic regression, a common statistical technique usually used for epidemiological outcome prediction [19], and we will use several important results from this field in our derivation. Logistic regression models are solved by maximum likelihood methods [19,20], which we apply here. Our specific problem is simplified because the exact value for β is given as an input to the simulation; we need only estimate the free energy. One could in principle estimate both β and ΔF from the two distributions with logistic regression techniques, though it is unclear at this point what practical advantage it may have.

The ratio in Eq. (2) can then be written as $P(W|F)/P(W|R)$ where $P(W|F)$ is the conditional probability of a work value given that it is a forward measurement and $P(W|R)$ is the conditional probability of a work value given that it is a reverse measurement. We would like to compute the likelihood of a free energy estimate of a given number of work measurements which are specified as either forward or reverse. Although either $P(W|R)$ or $P(W|F)$ can be eliminated, we are left with one independent continuous free energy distribution, and writing either $P(W|F)$ or $P(W|R)$ in a closed form is system dependent. Although the problem can be solved by finding the likelihood over the this continuous distribution without the need for a closed form [21,22], it is possible to rewrite this problem in a much more tractable way.

Using the rules of conditional probabilities, and the fact that $P(F|W) + P(R|W) = 1$, we rewrite this probability distribution as follows:

$$\frac{P(W|F)}{P(W|R)} = \frac{P(F|W)P(R)}{P(R|W)P(F)} = \frac{P(F|W)}{1 - P(F|W)}\frac{P(R)}{P(F)}. \qquad (3)$$

We note that $P(R)/P(F) = n_R/n_F$, where n_F and n_R are the number of forward and reverse measurements, respectively. We define the constant $M = kT\ln(n_F/n_R)$ and rewrite Eq. (2) as

$$\ln\frac{P(F|W)}{1 - P(F|W)} = \beta(M + W - \Delta F). \qquad (4)$$

Given Eq. (4) and an estimate for ΔF, we can rewrite the probability of a single measurement $P(F|W_i)$ as

$$P(F|W_i) = \frac{1}{1 + \exp[-\beta(M + W_i - \Delta F)]}. \qquad (5)$$

Similarly,

$$P(R|W_i) = \frac{1}{1 + \exp[\beta(M + W_i - \Delta F)]}. \qquad (6)$$

We now have expressions for the probabilities $P(F|W_i)$ and $P(R|W_i)$ given a value of the free energy ΔF without introducing any additional parameters and eliminating the need to parametrize a continuous distribution. We can then estimate the free energy that would maximize the likelihood of having found the specified distribution of forward and reverse values of W.

Given a value for the free energy ΔF, the overall likelihood L of obtaining the given measurements can be expressed as a joint probability of obtaining the forward measurements of work at the specified work values times the joint probability of obtaining the reverse measurements at the specified work values:

$$L(\Delta F) = \prod_{i=1}^{n_F} P(F|W_i) \prod_{j=1}^{n_R} P(R|W_j). \tag{7}$$

The most likely value of ΔF is the value that maximizes the likelihood, but it is usually easier to solve the equivalent problem of maximizing the log likelihood. Taking first the log of the likelihood and then taking the derivative with respect to ΔF and setting it equal to zero we obtain:

$$\frac{\partial \ln L(\Delta F)}{\partial \Delta F} = \sum_{i=1}^{n_F} \frac{1}{1 + \exp[\beta(M + W_i - \Delta F)]} - \sum_{j=1}^{n_R} \frac{1}{1 + \exp[-\beta(M + W_j - \Delta F)]} = 0. \tag{8}$$

The left side of Eq. (8) is a strictly increasing function in ΔF, and has limits of $-n_R$ as $\Delta F \rightarrow -\infty$ and n_F as $\Delta F \rightarrow \infty$, so we are guaranteed that ΔF has one unique root. This value of the free energy difference ΔF is the maximum likelihood estimate (MLE) of the measured data. This is the likelihood with a fixed *probability* for forward and reverse measurements obtained from the ratio $P(F|W)/P(R|W)$, not a fixed *number* of each type of measurement, obtained from the ratio $P(W|F)/P(W|R)$. However, it has been shown that these two approaches to logistic regression yield the same values for the parameter or parameters being estimated [21,22].

Equation (8) is exactly equivalent to the Bennett acceptance ratio method, as can be seen by comparison to Eqs. 12(a) and 12(b) of Bennett's paper describing the method [17]. In other words, the Bennett acceptance ratio yields the free energy which, given a series of work measurements in the forward and reverse direction, maximizes the chance these work values would be observed.

As mentioned before, the MLE is an asymptotically efficient estimator if certain weak conditions are met. These conditions are that there exists a unique root in the likelihood equation for every n, and that the third derivative with respect to the parameter (here, the free energy ΔF) is bounded [18]. These conditions are satisfied in this case.

Variance estimates from maximum likelihood methods.—The variance of the MLE of a parameter θ asymptotically converges to $1/I(\theta)$ as $n \rightarrow \infty$, where $I(\theta)$ is the Fisher information of the joint distribution of the work values with respect to the parameter θ [18]. The Fisher information is defined as the negative of the expectation value of the second derivative of the log likelihood with respect to the parameter θ [18]. In our case, the parameter is ΔF, the free energy difference, and the estimate of the Fisher information given finite sampling is

$$I(\Delta F) = \frac{\beta^2}{2}\left(\sum_{i=n_{tot}} \frac{1}{1 + \cosh[\beta(M + W_i - \Delta F)]}\right), \tag{9}$$

where this sum goes over both the forward and reverse work measurements.

However, this leads to a variance estimate for a fixed *probability* of forward and reverse simulations, not a fixed *number* of simulations in the forward and reverse direction, including evaluations of free energy over the cases that n_F is not the specified number. We must include this restriction in order to obtain the correct variance for the maximum likelihood estimate for the fixed number case. Anderson showed that this difference can be compensated for by subtracting $1/n_F + 1/n_R$ from the variance [21], yielding:

$$\frac{1}{\beta^2 n_{\mathrm{tot}}}\left[\left\langle\frac{1}{2 + 2\cosh[\beta(M + W_i - \Delta F)]}\right\rangle^{-1} - \left(\frac{n_{\mathrm{tot}}}{n_F} + \frac{n_{\mathrm{tot}}}{n_R}\right)\right], \tag{10}$$

where the average in the above equations is over all work measurements, both forward and reverse. Anderson's derivation is rather technical, and we provide an alternate derivation as supplementary materials specific to this case [23]. Comparing the methods obtained by the maximum likelihood and Bennett's original variance calculation, we find that Eq. (10) can be identified with Eq. 10(b) in Ref. [17], by recognizing that ρ_1/ρ_0 is simply equal to the factor $\exp[\beta(W - \Delta F)]$.

Although we know that the MLE is asymptotically unbiased, we still need some sense of the bias for finite numbers of measurements. Typically, the bias of a MLE is proportional to n^{-1} [20]. Since the standard error is proportional to $n^{-1/2}$, in general the statistical bias will not be significant with the respect to the statistical uncertainty.

Discussion.—There are strong connections between the acceptance ratio method and exponential averaging. In the limit that $n_F \gg n_R$, Eq. (8) becomes

$$\begin{aligned} \sum_{i=n_F} \frac{n_R}{n_F} \exp[-\beta(W_i - \Delta F)] - \sum_{i=n_R} 1 = 0, \\ n_R\langle \exp[-\beta(W - \Delta F)]\rangle_F = n_R, \\ \langle \exp(-\beta W)\rangle_F = \exp(-\beta \Delta F). \end{aligned} \tag{11}$$

Assuming that $n_R \gg n_F$, we similarly find that $\langle \exp(\beta W)\rangle_R = \exp(\beta \Delta F)$, thus recovering the well-known fact that the free energy is the expectation value of the exponential average of either the forward or the reverse distribution. Thus, one can interpret the exponential average as the maximum likelihood estimator for the free energy difference in the limit that samples are drawn from only one distribution. No other asymptotically unbiased estimate that draws from only one distribution can have a lower variance.

However, in an analysis of the computational efficiency of the acceptance ratio method as the ratio n_F/n_R is varied, Bennett showed that the optimally efficient ratio was always relatively close to 1 [17]. The limits of $n_R \gg n_F$ and $n_F \gg n_R$ are significantly less efficient, sometimes by 3 or 4 orders of magnitude. Essentially, there is only one independent distribution, as the forward and reverse distributions are related by Eq. (2). Above ΔF, the forward distribution will be more statistically accurate in estimating this independent distribution, and below ΔF, the reverse distribution will be the most statistically accurate. It will therefore always be better to partition the available simulations or experiments between both directions rather than explore the distribution of work values in one direction for twice as long. The acceptance ratio method will therefore always be preferable to FEP. We note that if FEP is performed in both the forward and reverse directions, as is typically the case, then the exact same data can be instead used in the acceptance ratio method.

The fact that the MLE is the asymptotically unbiased estimator with the minimum variance does not necessarily guarantee that is the best estimator by all measures. It is sometimes possible to find an estimator that *is* asymptotically biased, but has significantly smaller variance, resulting in a smaller mean squared error for a finite number of measurements. Preliminary evidence we have obtained indicates, for example, that the probability distributions $P_R(W)$ and $P_F(W)$ can be smoothed by such methods as convolution with a kernel to obtain mean squared errors similar to or smaller than acceptance ratio estimates in some cases.

Conclusions.—We have demonstrated that the Bennett acceptance ratio method can be interpreted in terms of the maximum likelihood estimate of the free energy difference given a set of nonequilibrium work values in the forward and reverse directions. This extends Bennett's work by showing that the acceptance ratio method is the minimum variance asymptotically unbiased estimate. We have also derived an expression for the variance of the Bennett acceptance ratio method using the same methods. We note use of the Bennett acceptance ratio will always yield lower variance than exponential averaging. We also note that the conclusions presented in this Letter are valid for the general case of work distributions from an initial state to a final state, not only the specific case of energy differences from an equilibrium state to another state.

The authors thank Chris Jarzynski for many useful comments and extensive editorial suggestions, and Trevor Hastie and Rob Tibshirani for useful literature suggestions. M. R. S. thanks the John and Fannie Hertz Foundation for financial support.

[1] *Free Energy Calculations in Rational Drug Design*, edited by M. R. Reddy and M. D. Erion (Kluwer Academic, New York, 2001).
[2] A. Gelman and X. L. Meng, Stat. Sci. **13**, 163 (1998).
[3] T. Schafer and E. V. Shuryak, Rev. Mod. Phys. **70**, 323 (1998).
[4] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, and C. Bustamante, Science **296**, 1832 (2002).
[5] G. Hummer, Mol. Simul. **28**, 81 (2002).
[6] D. A. Hendrix and C. Jarzynski, J. Chem. Phys. **114**, 5974 (2001).
[7] D. A. Kofke and P. T. Cummings, Fluid Phase Equilib. **151**, 41 (1998).
[8] D. M. Zuckerman and T. B. Woolf, Phys. Rev. Lett. **89**, 180602 (2002); Chem. Phys. Lett. **351**, 445 (2002).
[9] N. D. Lu and D. A. Kofke, J. Chem. Phys. **114**, 7303 (2001); **115**, 6866 (2001); **111**, 4414 (1999).
[10] N. D. Lu, J. K. Singh, and D. A. Kofke, J. Chem. Phys. **118**, 2977 (2003).
[11] M. R. Shirts, J. W. Pitera, W. C. Swope, and V. S. Pande, J. Chem. Phys. (to be published).
[12] R. W. Zwanzig, J. Chem. Phys. **22**, 1420 (1954).
[13] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).
[14] G. E. Crooks, Phys. Rev. E **61**, 2361 (2000).
[15] T. P. Straatsma and J. A. McCammon, J. Chem. Phys. **95**, 1175 (1991).
[16] D. A. Pearlman and P. A. Kollman, J. Chem. Phys. **91**, 7831 (1989).
[17] C. H. Bennett, J. Comput. Phys. **22**, 245 (1976).
[18] E. L. Lehmann and G. Casella, *Theory of Point Estimation* (Springer-Verlag, New York, 1998).
[19] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer-Verlag, New York, 2001).
[20] J. A. Nelder and P. McCullagh, *Generalized Linear Models* (CRC Press, Boca Raton, Florida, 1989), 2nd ed.
[21] J. A. Anderson, Biometrika **59**, 19 (1972).
[22] R. L. Prentice and R. Pyke, Biometrika **66**, 403 (1979).
[23] See EPAPS Document No. E-PRLTAO-91-014337 for an alternative derivation for the correction term. A direct link to this document may be found in the online article's HTML reference section. The document may also be reached via the EPAPS homepage (http://www.aip.org/pubservers/epaps.html) or from ftp.aip.org in the directory /epaps/. See the EPAPS homepage for more information.