Optimal estimators and asymptotic variances for nonequilibrium path-ensemble averages

David D. L. Minh[1,a)] and John D. Chodera[2,b)]

[1]*Laboratory of Chemical Physics, NIDDK, National Institutes of Health, Bethesda, Maryland 20892, USA*
[2]*California Institute for Quantitative Biosciences (QB3), University of California, Berkeley, Berkeley, California 94720, USA*

Existing optimal estimators of nonequilibrium path-ensemble averages are shown to fall within the framework of extended bridge sampling. Using this framework, we derive a general minimal-variance estimator that can combine nonequilibrium trajectory data sampled from multiple path-ensembles to estimate arbitrary functions of nonequilibrium expectations. The framework is also applied to obtain asymptotic variance estimates, which are a useful measure of statistical uncertainty. In particular, we develop asymptotic variance estimates pertaining to Jarzynski's equality for free energies and the Hummer–Szabo expressions for the potential of mean force, calculated from uni- or bidirectional path samples. These estimators are demonstrated on a model single-molecule pulling experiment. In these simulations, the asymptotic variance expression is found to accurately characterize the confidence intervals around estimators when the bias is small. Hence, the confidence intervals are inaccurately described for unidirectional estimates with large bias, but for this model it largely reflects the true error in a bidirectional estimator derived by Minh and Adib. © *2009 American Institute of Physics*. [doi:10.1063/1.3242285]

I. INTRODUCTION

Path-ensemble averages play a central role in nonequilibrium statistical mechanics, akin to the role of configurational ensemble averages in equilibrium statistical mechanics. Expectations of various functionals over processes where a system is driven out of equilibrium by a time-dependent external potential have been shown to be related to equilibrium properties, including free energy differences[1,2] and thermodynamic expectations.[3,4] The latter relationship, between equilibrium and nonequilibrium expectations, has been applied to several specific cases, such as the potential of mean force (PMF) along the pulling coordinate[5–7] (or other observed coordinates[8]) in single-molecule pulling experiments, RNA folding free energies as a function of a control parameter,[9] the root mean square deviation from a reference structure,[10] the potential energy distribution[10] and average,[11] and the thermodynamic length.[12]

Compared to equilibrium sampling, nonequilibrium processes may be advantageous for traversing energetic barriers and accessing larger regions of phase space per unit time. This is useful, for example, in reducing the effects of experimental apparatus drift or increasing the sampling of barrier-crossing events. Thus, there has been interest in calculating equilibrium properties from nonequilibrium trajectories collected in simulations or laboratory experiments. Indeed, single-molecule pulling data have been used to experimentally verify relationships between equilibrium and nonequilibrium quantities.[13,14]

While many estimators for free energy differences[3,15–17] and equilibrium ensemble averages can be constructed from nonequilibrium relationships, they will differ in the efficiency with which they utilize finite data sets, leading to varying amounts of statistical bias and uncertainty. Characterization of this bias and uncertainty is helpful in comparing the quality of different estimators[18] and assessing the accuracy of a particular estimate. The statistical uncertainty of an estimator is usually quantified by its variance in the asymptotic, or large sample, limit, where estimates from independent repetitions of the experiment often approach a normal distribution about the true value due to the central limit theorem. It is an important goal to find an optimal estimator which minimizes this asymptotic variance.

Although numerical estimates of the asymptotic variance may be provided by bootstrapping (e.g., Ref. 19), closed-form expressions can provide computational advantages in the calculation of confidence intervals, allow comparison of asymptotic efficiency,[18,20] and facilitate the design of adaptive sampling strategies to target data collection in a manner that most rapidly reduces statistical error.[21–23] In the asymptotic limit, the statistical error in functions of the estimated parameters can be estimated by propagating this variance estimate via a first-order Taylor series expansion. While this procedure is relatively straightforward for simple estimators, it can be difficult for estimators that involve arbitrary functions (e.g., nonlinear or implicit equations) of nonequilibrium path-ensemble averages.

Fortunately, the extended bridge sampling (EBS) estimators,[20,24–26] a class of equations for estimating the ratios of normalizing constants, are known to have both minimal-variance forms and associated asymptotic variance

a)Electronic mail: daveminh@gmail.com. Present address: Biosciences Division, Argonne National Laboratory, 9700 S. Cass Avenue, Argonne, IL 60439 USA.
b)Electronic mail: jchodera@berkeley.edu.

expressions. Recently, Shirts and Chodera[27] applied the EBS formalism to generalize the Bennett acceptance ratio,[15] producing an optimal estimator combining data from multiple equilibrium states to compute free energy differences, thermodynamic expectations, and their associated uncertainties. Here, we apply the EBS formalism to estimators utilizing nonequilibrium trajectories. We first construct a general minimal-variance path-average estimator that can use samples collected from multiple nonequilibrium path ensembles. We then show that some existing path-average estimators using uni- and bidirectional data are special cases of this general estimator, proving their optimality. This also allows us to develop asymptotic variance expressions for estimators based on Jarzynski's equality[1,2] and the Hummer–Szabo expressions for the PMF.[5,6,28] We then demonstrate them on simulation data from a simple one-dimensional system and comment on their applicability.

II. EXTENDED BRIDGE SAMPLING

Suppose that we sample N_i paths (trajectories) from each of K path ensembles indexed by $i=1,2,\ldots,K$. The path-ensemble average of an arbitrary functional $\mathcal{F}[X]$ in path ensemble i is defined by

$$\langle \mathcal{F} \rangle_i \equiv \int dX\, \mathcal{F}[X]\, \rho_i[X], \tag{1}$$

where $\rho_i[X]$ is a probability density over trajectories,

$$\rho_i[X] = c_i^{-1} q_i[X]; \quad c_i = \int dX\, q_i[X], \tag{2}$$

with unnormalized density $q_i[X] > 0$ and the normalization constant c_i (a path partition function). The above integrals, in which dX is an infinitesimal path element, are taken over all possible paths, X. EBS estimators provide a way of estimating ratios of normalization constants c_i/c_j, which will prove useful in estimating free energies and thermodynamic expectations.

To construct these estimators, we first note the importance sampling identity,

$$c_i \langle \alpha_{ij} q_j \rangle_i = \left[\int dX\, q_i[X] \right] \frac{\int dX\, \alpha_{ij}[X]\, q_i[X]\, q_j[X]}{\int dX\, q_i[X]}$$
$$= \left[\int dX\, q_j[X] \right] \frac{\int dX\, \alpha_{ij}[X]\, q_i[X]\, q_j[X]}{\int dX\, q_j[X]}$$
$$= c_j \langle \alpha_{ij} q_i \rangle_j, \tag{3}$$

where j is another path-ensemble index, $\alpha_{ij}[X]$ is an arbitrary functional of X, and all normalization constants are nonzero.

Summing over the index j in Eq. (3) and using the sample mean, $N_i^{-1} \Sigma_{n=1}^{N_i} \mathcal{F}[X_{in}]$, as an estimator for $\langle \mathcal{F} \rangle_i$, we obtain a set of K estimating equations, indexed by $i=1,\ldots,K$,

$$\sum_{j=1}^{K} \frac{\hat{c}_i}{N_i} \sum_{n=1}^{N_i} \alpha_{ij}[X_{in}]\, q_j[X_{in}] = \sum_{j=1}^{K} \frac{\hat{c}_j}{N_i} \sum_{n=1}^{N_j} \alpha_{ij}[X_{jn}]\, q_i[X_{jn}], \tag{4}$$

whose solutions yield estimates $\hat{c}_i$ for the normalization constants c_i, up to an irrelevant scalar multiple. Each path, X_{in}, is indexed by the ensemble i from which it is sampled, and the sample number $n=1,2,\ldots,N_i$. This coupled set of nonlinear equations defines a *family* of estimators parametrized by the choice of $\alpha_{ij}[X]$, all of which are asymptotically consistent, but whose statistical efficiencies will vary.[20]

With the choice,

$$\alpha_{ij}[X] = \frac{N_j\, \hat{c}_j^{-1}}{\Sigma_{k=1}^{K} N_k\, \hat{c}_k^{-1} q_k[X]}, \tag{5}$$

Eq. (4) simplifies to the optimal EBS estimator,

$$\hat{c}_i = \sum_{j=1}^{K} \sum_{n=1}^{N_j} \left[\sum_{k=1}^{K} \frac{N_k\, q_k[X_{jn}]}{\hat{c}_k\, q_i[X_{jn}]} \right]^{-1}. \tag{6}$$

This choice for $\alpha_{ij}[X]$ is optimal in that the asymptotic variance of the ratios $\hat{c}_i/\hat{c}_j$ is minimal.[20,27] These equations may be solved by any appropriate algorithm, including a number of efficient and stable methods suggested by Shirts and Chodera.[27]

The asymptotic covariance of Eq. (6) is estimated by

$$\hat{\Theta} = M^T (I_N - MNM^T)^+ M, \tag{7}$$

where the elements of Θ are the covariances of the logarithms of the estimated normalization constants, $\Theta_{ij} \equiv \mathrm{cov}(\hat{\gamma}_i, \hat{\gamma}_j)$, and $\hat{\gamma}_i \equiv \ln \hat{c}_i$.[26] The superscript $(\cdots)^+$ denotes an appropriate generalized inverse, such as the Moore–Penrose pseudoinverse, I_N is the $N \times N$ identity matrix (where $N = \Sigma_{i=1}^{K} N_i$ is the total number of samples), $N \equiv \mathrm{diag}(N_1, N_2, \ldots, N_K)$ is the diagonal matrix of sample sizes, and M is the $N \times K$ weight matrix with elements,

$$M_{ni} = \hat{c}_i^{-1} \frac{q_i[X_n]}{\Sigma_{k=1}^{K} N_k\, \hat{c}_k^{-1} q_k[X_n]}. \tag{8}$$

In this matrix, the distribution from which samples are drawn from is irrelevant and X is only indexed by $n=1,\ldots,N$. We note that the sum over each column, $\Sigma_{n=1}^{N} M_{ni}$, is unity. (Efficient methods for computing $\hat{\Theta}$ are discussed in Appendix D of Ref. 27.)

For arbitrary functions of the logarithms of the normalization constants, $\phi(\hat{\gamma}_1, \ldots, \hat{\gamma}_K)$ and $\psi(\hat{\gamma}_1, \ldots, \hat{\gamma}_K)$, the asymptotic covariance $\mathrm{cov}(\hat{\phi}, \hat{\psi})$ can be estimated from $\hat{\Theta}$ according to

$$\mathrm{cov}(\hat{\phi}, \hat{\psi}) \approx \sum_{i,j=1}^{K} \frac{\partial \phi}{\partial \hat{\gamma}_i} \hat{\Theta}_{ij} \frac{\partial \psi}{\partial \hat{\gamma}_j}, \tag{9}$$

through first-order Taylor series expansion of ϕ and ψ.

III. GENERAL PATH-ENSEMBLE AVERAGES

Following previous work,[27,29] we estimate nonequilibrium expectations by defining additional path ensembles with "unnormalized densities,"

$$q_{\mathcal{F}_i}[X] = \mathcal{F}[X]\, q_i[X]; \quad c_{\mathcal{F}_i} = \int dX\, q_{\mathcal{F}_i}[X]. \tag{10}$$

Using Eqs. (1), (2), and (10), we can express nonequilibrium expectations as a ratio of the appropriate normalization constants, $\langle \mathcal{F} \rangle_i = c_{\mathcal{F}_i}/c_i$. Notably, this can be estimated *without* actually sampling path ensembles biased by some function of $\mathcal{F}[X]$ (although it is sometimes possible to do so in computer simulations[30,31] via transition path sampling[32,33]). If no paths are drawn from the path ensemble corresponding to $q_{\mathcal{F}_i}[X]$, then $N_{\mathcal{F}_i}=0$ and it is no longer required that $q_{\mathcal{F}_i}[X]>0$.[20,27]

For each defined path ensemble, the weight matrix $\boldsymbol{M}$ is augmented by one column with elements,

$$M_{n\mathcal{F}_i} = \hat{c}_{\mathcal{F}_i}^{-1} \frac{\mathcal{F}[X_n]\, q_i[X_n]}{\sum_{k=1}^{K} N_k\, \hat{c}_k^{-1} q_k[X_n]}. \tag{11}$$

The estimator for the path-ensemble average, $\overline{\mathcal{F}}_i \approx \langle \mathcal{F} \rangle_i$, can be expressed in terms of weight matrix elements,

$$\overline{\mathcal{F}}_i = \sum_{n=1}^{N} M_{ni} \mathcal{F}[X_n], \tag{12}$$

and its uncertainty estimated by

$$\sigma^2(\overline{\mathcal{F}}_i) \approx \overline{\mathcal{F}}_i^2 (\hat{\Theta}_{\mathcal{F}_i\mathcal{F}_i} - 2\hat{\Theta}_{\mathcal{F}_i i} + \hat{\Theta}_{ii}). \tag{13}$$

IV. EXPERIMENTALLY RELEVANT PATH ENSEMBLES

The above formalism is fully general, and may be applied to *any* situation where the ratio $q_i[X]/q_j[X]$ can be computed. For arbitrary path ensembles, unfortunately, calculating this ratio is only possible in computer simulations unless certain assumptions are made about the dynamics.[34] In a few special path ensembles, however, we can use the Crooks fluctuation theorem[35,36] to estimate this ratio, allowing us to apply the EBS estimator to laboratory experiments. We examine these here.

First, consider a *forward process*, in which a system, initially in equilibrium, is propagated under some time-dependent dynamics for a time τ, which may cause it to be driven out of equilibrium. The time dependence of the evolution law (e.g., Hamiltonian dynamics in a time-dependent potential) is the same for all paths sampled from this ensemble.

For a sample of paths only drawn from this ensemble, the optimal EBS estimator of $\langle \mathcal{F} \rangle_f$ reduces to the sample mean estimator, which we call the *unidirectional* path-ensemble average estimator,

$$\overline{\mathcal{F}}_f = \frac{1}{N_f} \sum_{n=1}^{N_f} \mathcal{F}[X_{fn}], \tag{14}$$

and the associated asymptotic variance from Eq. (9) reduces to the variance of the sample mean (see Appendix A)

$$\sigma^2(\overline{\mathcal{F}}_f) \approx \frac{1}{N_f} \left[\frac{1}{N_f} \sum_{n=1}^{N_f} (\mathcal{F}[X_{fn}] - \overline{\mathcal{F}}_f)^2 \right]. \tag{15}$$

The forward process has a unique counterpart known as the *reverse process*. Here, the system moves via the opposite protocol in thermodynamic state space; after initial configurations are drawn from the final thermodynamic state of the forward path ensemble, they are driven toward the initial state. If the dynamical law satisfies detailed balance when the control parameters are held constant at each fixed time t, the path probabilities in the conjugate forward and reverse path ensembles are related according to the Crooks fluctuation theorem,[35,36]

$$\frac{\rho_f[X]}{\rho_r[\tilde{X}]} = \frac{q_f[X]\, c_r}{q_r[\tilde{X}]\, c_f} = e^{w_\tau[X] - \Delta f_\tau} \equiv e^{\Omega[X]}, \tag{16}$$

in which $\tilde{X}$ is the time reversal, or *conjugate twin*,[37] of X, $\Delta f_t = -\ln(c_t/c_0)$ is the dimensionless free energy difference between thermodynamic states at times 0 and t (with τ being the fixed total trajectory length), and $w_t[X]$ is the appropriate dimensionless work. With Hamiltonian dynamics, for example, this work is $w_t[X] = \beta \int_0^t dt'(\partial H/\partial t')$. For convenience, we define the total *dissipative work* as $\Omega[X] \equiv w_\tau[X] - \Delta f_\tau$.

We will refer to data sets which only include realizations from the forward path ensemble as "unidirectional," and those with paths from both path ensembles as "bidirectional." Notably, sampling paths from these conjugate ensembles and calculating the associated work $w_t[X]$ are possible in single-molecule pulling experiments as well as computer simulations (cf. Refs. 6 and 14). To combine bidirectional data to estimate $\langle \mathcal{F} \rangle_f$, we apply the Crooks fluctuation theorem[35,36] to Eq. (6) and divide by $\hat{c}_f$, leading to

$$\overline{\mathcal{F}}_f = \sum_{n=1}^{N_f} \frac{\mathcal{F}[X_{fn}]}{N_f + N_r\, e^{-\hat{\Omega}[X_{fn}]}} + \sum_{n=1}^{N_r} \frac{\mathcal{F}[X_{rn}]}{N_f + N_r\, e^{-\hat{\Omega}[X_{rn}]}}, \tag{17}$$

which is bidirectional path-average estimator of Minh and Adib,[28] derived here by a different route which demonstrates its optimality. (The asymptotic variance estimator for this equation is written in a closed form in Appendix B.) In these bidirectional expressions, samples drawn from the reverse path ensemble are time reversed to obtain the paths X_{rn}. The dissipated work estimate, $\hat{\Omega}[X] \equiv w_\tau[X] - \Delta \hat{f}_\tau$, requires an estimate of Δf_τ. A method for obtaining this estimate will be described next.

V. FREE ENERGY

Jarzynski's equality,[1,2]

$$e^{-\Delta f_t} = \langle e^{-w_t} \rangle_f, \tag{18}$$

relates nonequilibrium work and free energy differences. To facilitate the use of EBS in Jarzynski's equality, we define a

path ensemble by choosing $\mathcal{F}[X] = e^{-w_t[X]}$ in Eq. (10), leading to

$$q_{w_t}[X] = e^{-w_t[X]} q_f[X]; \quad c_{w_t} = \int dX\, e^{-w_t[X]} q_f[X]. \quad (19)$$

When only unidirectional data is available, the optimal EBS estimator for Jarzynski's equality is

$$e^{-\Delta \hat{f}_t} = \frac{1}{N_f} \sum_{n=1}^{N_f} e^{-w_t[X_{fn}]} \quad (20)$$

and its asymptotic variance is straightforwardly given by error propagation.[38] Estimators[30,31,39] and asymptotic variances[39,40] have also been developed for unidirectional *importance sampling* forms of the equality.

When bidirectional data is available, the same choice of $\mathcal{F}[X]$ in Eq. (17) gives the estimator

$$e^{-\Delta \hat{f}_t} = \sum_{n=1}^{N_f} \frac{e^{-w_t[X_{fn}]}}{N_f + N_r\, e^{-\hat{\Omega}[X_{fn}]}} + \sum_{n=1}^{N_r} \frac{e^{-w_t[X_{rn}]}}{N_f + N_r\, e^{-\hat{\Omega}[X_{rn}]}}. \quad (21)$$

In this equation, choosing $t = 0$ or $t = \tau$ leads to an implicit function mathematically equivalent to the Bennett acceptance ratio method,[3,15] as previously explained.[28,41] The asymptotic variance of $\Delta \hat{f}_t$ is calculated by augmenting the matrices M and $\hat{\Theta}$ and using $\phi = \psi = \Delta f_t = -\ln(c_{w_t}/c_f)$ in Eq. (9), such that

$$\sigma^2(\Delta \hat{f}_t) = \hat{\Theta}_{w_t w_t} - 2\hat{\Theta}_{w_t f} + \hat{\Theta}_{ff}. \quad (22)$$

VI. POTENTIAL OF MEAN FORCE

Building on Jarzynski's equality, Hummer and Szabo developed expressions for the PMF,[5,6] the free energy as a function of a *order parameter* rather than a thermodynamic state, which may be used to interpret single-molecule pulling experiments. In these experiments, a molecule is mechanically stretched by a force-transducing apparatus, such as a laser optical trap or atomic force microscope tip (cf. Ref. 6). The Hamiltonian governing the time evolution in these experiments, $H(x;t) = H_0(x) + V(z(x);t)$, is assumed to contain both a term corresponding to the unperturbed system, $H_0(x)$, and a time-dependent (typically harmonic) external bias potential imposed by the apparatus, $V(z;t)$, which acts along a pulling coordinate, $z(x)$. As the coordinate $z_t \equiv z(x(t))$ is observed at fixed intervals Δt over the course of the experiment, we will henceforth use $t = 0, 1, \ldots, T$ as an integer time index. We calculate the work with a discrete sum as $w_t = \sum_{n=1}^{t} [V_n(z_n) - V_{n-1}(z_n)]$, where $V_n(z) \equiv V(z; n\Delta t)$.

While the expressions in Sec. V provide an estimate of relative free energies of the equilibrium thermodynamic states defined by $H(x;t)$, they are not immediately useful as an estimate for the PMF along z.[5,6,42] By applying the non-equilibrium estimator for thermodynamic expectations,[3,4] it was shown that the PMF in the absence of an external potential is given by[5,6]

$$e^{-g_0(z)} = \langle \delta(z - z_t) e^{-w_t} \rangle_f\, e^{V(z_t;t)}, \quad (23)$$

where the dimensionless PMF, $g_0(z)$, is defined in relation to the normalized density as $g_0(z) = -\ln p_0(z) - \delta g$. In this equation, δg is a time-independent constant, $e^{-\delta g} = \int dx\, e^{-H(x;0)} / \int dx\, e^{-H_0(x)}$.[6]

This theorem [Eq. (23)] can be used to develop estimators for the PMF by replacing the delta function using a kernel function of finite width, such as

$$h(z - z_t) = \begin{cases} \dfrac{1}{\Delta z}, & \text{if } |z - z_t| < \dfrac{\Delta z}{2}, \\ 0, & \text{else.} \end{cases} \quad (24)$$

The width Δz must be small so that $e^{V(z;t)}$ does not vary substantially across it.

As this theorem is valid at all times, it is possible to obtain an asymptotically unbiased density estimate $\hat{p}_t$ from each time slice. It is far more efficient, however, to estimate the PMF using *all* recorded time slices. While any linear combination of time slices will lead to a valid estimate, certain choices will be more statistically efficient (leading to lower variance) than others. One way to combine time slices is to use the asymptotic covariance matrix in the method of control variates,[20] leading to a generalized least-squares optimal estimate of the PMF. Unfortunately, we empirically found this approach to be numerically unstable. A more numerically stable approach, which was proposed by Hummer and Szabo,[5,6] is based on the weighted histogram analysis method,[43,44] which was generalized to,[45–47]

$$\hat{p}_0(z) = \frac{\sum_t \mu_t(z) \hat{p}_t(z)}{\sum_t \mu_t(z)}; \quad \mu_t(z) \equiv e^{-V(z;t) + \Delta \hat{f}_t}. \quad (25)$$

While this weighting scheme is optimal, in a minimal-variance sense, for independent samples from multiple *equilibrium* distributions, these assumptions do not hold for time slices from nonequilibrium trajectories. However, Oberhofer and Dellago did not observe substantial improvement in PMF estimates when using other time-slice weighting schemes.[48]

By defining the path ensemble,

$$q_{z_t}[X] = \delta(z - z_t) e^{-w_t[X]} q_f[X]; \quad c_{z_t} = \int dX\, q_{z_t}[X], \quad (26)$$

and making use of Jarzynski's equality [Eq. (18)] for $e^{-\Delta \hat{f}_t}$, we can write Hummer and Szabo's PMF estimator as

$$e^{-\hat{g}_0(z)} = \frac{\sum_t (\hat{c}_{z_t}/\hat{c}_{w_t})}{\sum_t e^{-V(z;t)}(\hat{c}_f/\hat{c}_{w_t})}, \quad (27)$$

which can be readily analyzed in terms of EBS. While Hummer and Szabo proposed using the unidirectional path average estimator [Eq. (14)] to estimate the expectations in Eq. (27), Minh and Adib later applied a bidirectional estimator [Eq. (17)], leading to significantly improved statistical properties.[28]

The asymptotic variance of these estimators can be determined by choosing $\phi = \psi = p_0(z)$ in Eq. (9). For the bidi-
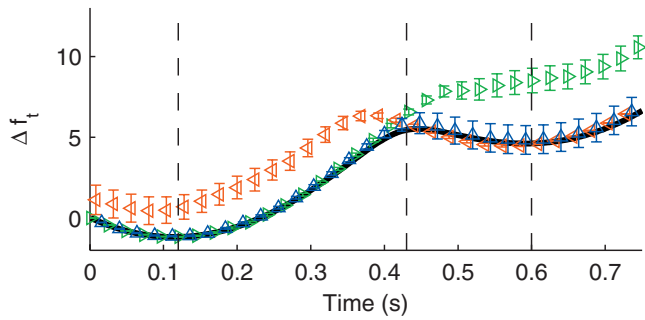
FIG. 1. Comparison of estimators for Δf_t. This figure is similar to Fig. 1 of Ref. [28], except that error bars are now included and the sample size is halved. The unidirectional estimator [Eq. (20)] is applied to 250 forward (rightward triangles) or reverse (leftward triangles, time reversed) sampled paths, and the bidirectional estimator [Eq. (21)] to 125 paths in each direction (upward triangles). The exact Δf_t is shown as a solid line. Error bars (sometimes smaller than the markers) denote one standard deviation of $\Delta \hat{f}_t$, estimated using the expressions presented here. The vertical dashed lines are at the times considered in Fig. 4.

FIG. 2. Comparison of PMF estimators. This figure is similar to Fig. 2 of Ref. [28], except that error bars are now included and the sample size is halved. In the left panel, the unidirectional Hummer and Szabo estimator is applied to (a) 250 forward (rightward triangles) or 250 reverse (leftward triangles) sampled paths. In the right panel, the bidirectional estimator is applied to 125 sampled paths in each direction (upward triangles). The exact PMF is shown as a solid line in both panels. Error bars (sometimes smaller than the markers) denote one standard deviation of $\Delta \hat{g}_0(z)$, estimated using the expressions presented here. The vertical dashed lines are at the positions considered in Fig. 4.

rectional estimator, the matrices M and $\hat{\Theta}$ will contain one column each for the f and r path ensembles, and $T+1$ columns each for the path ensembles associated with $\{w_t\}_{t=0}^T$ and $\{z_t\}_{t=0}^T$. The relevant partial derivatives are

$$\frac{\partial p_0(z)}{\partial \gamma_f} = -p_0(z), \tag{28}$$

$$\frac{\partial p_0(z)}{\partial \gamma_{w_t}} = -\frac{1}{\mathcal{D}}\frac{c_{z_t}}{c_{w_t}} + \frac{\mathcal{N}}{\mathcal{D}^2}\left(e^{-V(z;t)}\frac{c_f}{c_{w_t}}\right), \tag{29}$$

$$\frac{\partial p_0(z)}{\partial \gamma_{z_t}} = \frac{1}{\mathcal{D}}\frac{c_{z_t}}{c_{w_t}}, \tag{30}$$

where $\gamma_i = \ln c_i$, $\mathcal{N} = \Sigma_t(c_{z_t}/c_{w_t})$ is the numerator of Eq. (27), and $\mathcal{D} = \Sigma_t e^{-V(z;t)}(c_f/c_{w_t})$ is its denominator. These lead to an estimate for $\sigma^2(\hat{p}_0(z))$. Finally, the asymptotic variance in the PMF is given by the error propagation formula, $\sigma^2(\hat{g}_0(z)) \approx \sigma^2(\hat{p}_0(z))/\hat{p}_0(z)^2$.

VII. ILLUSTRATIVE EXAMPLE

We demonstrate these results with Brownian dynamics simulations on a one-dimensional potential with $U_0(z) = (5z^3 - 10z + 3)z$, which were run as previously described.[28] A time-dependent external perturbation, $V(z;t) = k_s(z - \bar{z}(t))^2/2$, with $k_s = 15$ is applied, such that the total potential is $U(z;t) = U_0(z) + V(z;t)$. After 100 steps of equilibration at the initial $\bar{z}(t)$, $\bar{z}(t)$ is linearly moved over 750 steps from -1.5 to 1.5 in forward processes and 1.5 to -1.5 in the reverse. The position at each time step is calculated using the equation $z_t = z_{t-1} - (dU(x_{t-1})/dx)D\Delta t + (2D\Delta t)^{1/2}R_t$, where the diffusion coefficient is $D = 1$, the time step is $\Delta t = 0.001$, and $R_t \sim N(0,1)$ is a random number from the standard normal distribution.

As previously noted,[28,49–51] unidirectional sampling leads to significant apparent bias in estimates of Δf_t (Fig. 1). In addition to the increased bias as the system is driven further from equilibrium, we further observe that the estimated variance also increases. Bidirectional sampling, on the other
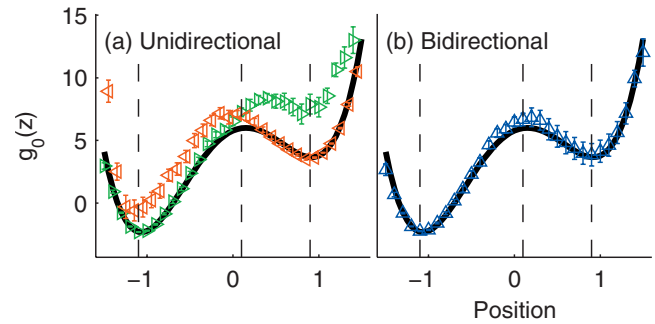
hand, leads to a significant reduction in bias and variance,[28] such that free energy estimate is within error bars of the actual free energy. Because $\Delta \hat{f}_t$ represents the estimated free energy difference with respect to t, the estimated $\sigma^2(\Delta \hat{F}_t)$ increases with t, becoming equal to the well-known Bennett acceptance ratio asymptotic variance estimate[15,41] when $t = \tau$.

Similar trends are observed with the Hummer–Szabo PMF estimates (Fig. 2). For unidirectional sampling, the finite-sampling bias and estimated variance increases when the PMF is far from the region sampled by the initial state. With bidirectional sampling, the bias is significantly reduced; the PMF estimate is largely within error bars of the actual PMF.

To analyze these trends more quantitatively, we repeated the experiment 1000 times. For both Δf_t and $g_0(z)$, we calculated the bias as $\bar{B}(\bar{\mathcal{F}}_f) = (1/S)\Sigma_{s=1}^S(\bar{\mathcal{F}}_{f,s} - \langle \mathcal{F} \rangle)$ and the standard deviation as $\bar{\sigma}(\bar{\mathcal{F}}_f) = \sqrt{(1/S)\Sigma_{s=1}^S(\bar{\mathcal{F}}_{f,s} - \langle \mathcal{F} \rangle)^2}$, where $S = 1000$ is the number of replicates. The results from these more extensive simulations support our described trends (Fig. 3). For unidirectional sampling, the bias in both Δf_t and $g_0(x)$ appear to significantly increase around the barrier
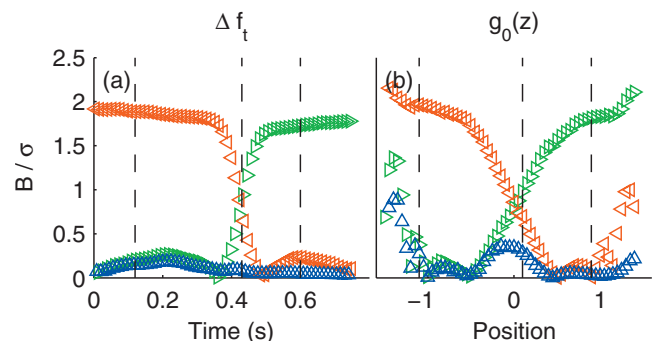

FIG. 3. Ratio of estimator bias to standard deviation. This ratio is calculated for the (a) free energy and (b) PMF, using 1000 independent estimates. Each estimate is obtained and the type of path sample is indicated as in Figs. 1 and 2. The vertical dashed lines are at the times/positions considered in Fig. 4.
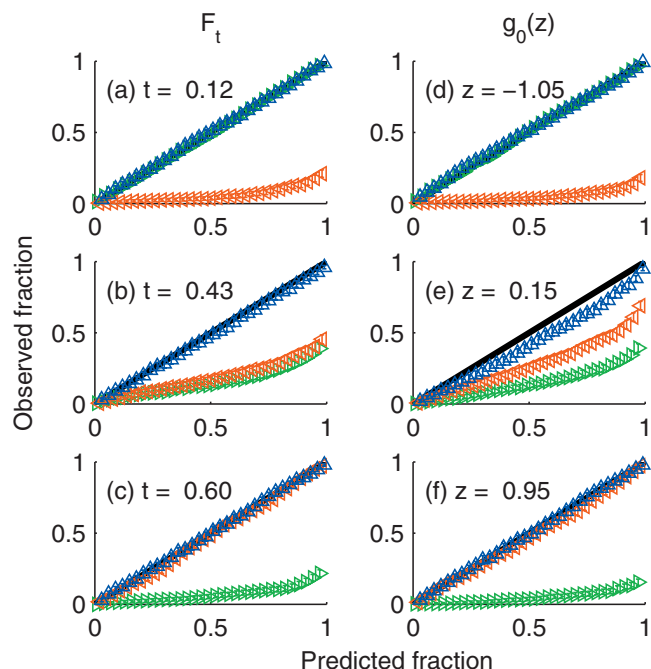
FIG. 4. Validation of asymptotic variance estimators. Predicted vs observed fraction of 1000 independent estimates that are within an interval of the true value for [(a)–(c)] Δf_t and [(d)–(f)] $g_0(z)$ at the indicated times or positions. Each estimate is obtained and the type of path sample is indicated as in Figs. 1 and 2. Error bars on these fractions are 95% confidence intervals calculated using a Bayesian scheme described in Appendix B of Chodera et al. (Ref. 62) except that, for numerical reasons, the confidence interval was estimated from the variance of the beta distribution assuming approximate normality, rather than from the inverse beta cumulative distribution function.

crossing. In the bidirectional free energy estimate, however, the bias is small relative to the variance at all times. Notably, in the bidirectional PMF estimate, there is a small spike in the bias near the barrier, potentially due to reduced sampling in the region.

While in the large sample limit, the bias in the unidirectional estimate is expected to be small compared to the variance,[50] our distribution of unidirectional $e^{-\Delta f_t}$ estimates is significantly skewed and does not resemble a Gaussian distribution expected by the central limit theorem (data not shown). Hence, the asymptotic limit has not been reached and the large relative bias is caused by insufficient sampling of rare events with low work values that dominate the exponential average.[37] Larger sample sizes would be necessary for the distribution of estimates to be normally distributed and for the error to be dominated by the variance (which we estimate here) rather than the bias.

The accuracy of variance estimates may be assessed by comparing predicted and observed confidence intervals. If the estimates are indeed normally distributed about the true value, about 68% of estimates from many independent replicates of the experiment should be within one standard deviation of the true value, 95% within two, and so forth. Figure 4 compares confidence intervals predicted using the described asymptotic variance estimators and the actual fraction of estimates within the interval.

We observe that the accuracy of our asymptotic variance estimate in characterizing the confidence interval largely depends on the presence of bias. In the bidirectional $\Delta \hat{f}_t$ esti-

mate, where there is little bias, the asymptotic variance estimate works very well. For the unidirectional $\Delta \hat{f}_t$ estimates, it works well near the initial state but underestimates the error as the system is driven further away from equilibrium, concurring with the bias trend. In the bidirectional PMF estimate, the asymptotic variance estimate accurately describes the confidence interval except near the barrier, where it slightly underestimates the uncertainty, probably due to the small spike in bias.

In the regime where the bias is much smaller than the variance, $\bar{B} \ll \bar{\sigma}$, the asymptotic variance estimate provides a good estimate of the actual statistical error in the estimate. This also permits us to model the posterior distribution of quantity being estimated as a multivariate normal distribution with mean $\overline{\mathcal{F}}$ and covariance $\hat{\Theta}$. Doing so provides a route to combining estimates from independent data sets collected from different path ensembles—such as different pulling speeds or from equilibrium and nonequilibrium path ensembles—without knowledge of path probability ratios. This is achieved by maximizing the product of these posterior distributions in a manner similar to the Bayesian approach for estimating Δf_τ described in Ref. 17.

VIII. DISCUSSION

In addition to the described experimentally relevant applications, we anticipate the use of our estimator in conjunction with a number of computational techniques. Recent work suggested[30,31,40,52,53] that the convergence of Jarzynski's equality may be improved by sampling from an alternate path ensemble, biased according to the work. Unfortunately, the optimal bias for this path ensemble includes the free energy itself,[53] limiting its practicality. However, an iterative procedure in which progressively improving free energy estimates are used in successive biased path ensembles is feasible, and our estimator provides a way to combine all of the data. Free energy estimates may also be improved by optimizing the switching protocol in thermodynamic state space to minimize the dissipated work.[54–61] When the optimal protocol is known,[59–61] the bidirectional form of our estimator is likely to improve free energy estimates over a unidirectional procedure. In the general case, protocol optimization may require an iterative procedure for which our estimator can also be used to combine the data.

Lastly, while we emphasized the use of EBS in *nonequilibrium* path-ensemble averages, the formalism described in Secs. II and III is equally applicable to other path ensembles. Given the explosion in popularity of transition path sampling,[32,33] in which novel path ensembles can be designed and sampled from, our estimator should find use in some of its applications.

ACKNOWLEDGMENTS

APPENDIX A: CLOSED-FORM EXPRESSION FOR THE ASYMPTOTIC VARIANCE, GIVEN UNIDIRECTIONAL DATA

In this appendix, we show that given unidirectional data, the optimal EBS estimate is the sample mean and its variance simplifies to the variance of a sample mean. First, consider the application of the optimal EBS estimator, Eq. (6), to estimating a nonequilibrium path-ensemble average from a unidirectional data set,

$$\hat{c}_{\mathcal{F}_f} = \sum_{n=1}^{N_f} \left[\frac{N_f \, q_f[X_{fn}]}{\hat{c}_f \, q_{\mathcal{F}_f}[X_{fn}]} \right]^{-1}, \tag{A1}$$

$$= \sum_{n=1}^{N_f} \frac{\mathcal{F}[X_{fn}]\hat{c}_f}{N_f}. \tag{A2}$$

Dividing both sides by $\hat{c}_f$, we obtain the sample mean estimator,

$$\overline{\mathcal{F}}_f = \frac{1}{N_f} \sum_{n=1}^{N_f} \mathcal{F}[X_{fn}]. \tag{A3}$$

We shall now simplify the asymptotic variance estimate by closely following the procedure of Shirts and Chodera.[27] When M has full column rank, $\hat{\Theta}$ can be written as (Eq. D7 of Ref. 27)

$$\hat{\Theta} = [(M^{\mathrm{T}}M)^{-1} - N + b\mathbf{1}_K\mathbf{1}_K^{\mathrm{T}}]^{-1}, \tag{A4}$$

where b is an arbitrary multiplicative factor and $\mathbf{1}_K$ is a $1 \times K$ matrix of ones.

The weight matrix M consists of two columns,

$$M_{nf} = \frac{\hat{c}_f^{-1} q_f[X_{fn}]}{N_f \, \hat{c}_f^{-1} q_f[X_{fn}]} = \frac{1}{N_f}, \tag{A5}$$

$$M_{n\mathcal{F}_f} = \frac{\hat{c}_{\mathcal{F}_f}^{-1} q_{\mathcal{F}_f}[X_{fn}]}{N_f \, \hat{c}_f^{-1} q_f[X_{fn}]} = \frac{\mathcal{F}[X_{fn}]}{N_f \overline{\mathcal{F}}_f}, \tag{A6}$$

obtained by applying Eqs. (8) and (11). This leads to,

$$M^{\mathrm{T}}M = \begin{bmatrix} N_f^{-1} & N_f^{-1} \\ N_f^{-1} & \sum_{n=1}^{N_f} M_{n\mathcal{F}_f}^2 \end{bmatrix} \equiv \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}. \tag{A7}$$

The matrix $M^{\mathrm{T}}M$ has the determinant

$$D = \frac{1}{N_f} \sum_{n=1}^{N_f} M_{n\mathcal{F}_f}^2 - \frac{1}{N_f^2}, \tag{A8}$$

which allows us to write the inverse covariance matrix as

$$\hat{\Theta}^{-1} = \begin{bmatrix} \dfrac{a_{22}}{D} - N_f + b & -\dfrac{a_{21}}{D} + b \\[3mm] -\dfrac{a_{12}}{D} + b & \dfrac{a_{11}}{D} + b \end{bmatrix}. \tag{A9}$$

By applying the same steps as Appendix E of Shirts and Chodera,[27] we then obtain the determinant

$$|\hat{\Theta}^{-1}| = \frac{4ab}{D}, \tag{A10}$$

where $a = a_{12} = a_{21}$. We then obtain the asymptotic covariance estimate

$$\hat{\Theta} = \frac{D}{4ab} \begin{bmatrix} \dfrac{a_{11}}{D} + b & \dfrac{a}{D} - b \\[3mm] \dfrac{a}{D} - b & \dfrac{a_{22}}{D} - N_f + b \end{bmatrix}. \tag{A11}$$

To estimate the variance, we apply Eq. (13), leading to

$$\sigma^2(\overline{\mathcal{F}}_f^2) \approx \overline{\mathcal{F}}_f^2(\Theta_{\mathcal{F}_f\mathcal{F}_f} - 2\Theta_{\mathcal{F}_f f} + \Theta_{ff}), \tag{A12}$$

$$= \overline{\mathcal{F}}_f^2 \left(\sum_{n=1}^{N_f} M_{n\mathcal{F}_f}^2 - \frac{1}{N_f} \right), \tag{A13}$$

$$= \sum_{n=1}^{N_f} \frac{\mathcal{F}[X_{fn}]^2}{N_f^2} - \frac{\overline{\mathcal{F}}_f^2}{N_f}, \tag{A14}$$

$$= \frac{1}{N_f} \left[\frac{1}{N_f} \sum_{n=1}^{N_f} \mathcal{F}[X_{fn}]^2 - \left(\frac{1}{N_f} \sum_{n=1}^{N_f} \mathcal{F}[X_{fn}] \right)^2 \right], \tag{A15}$$

$$= \frac{1}{N_f} \left[\frac{1}{N_f} \sum_{n=1}^{N_f} (\mathcal{F}[X_{fn}] - \overline{\mathcal{F}}_f)^2 \right], \tag{A16}$$

which is the variance of a sample mean estimate.

APPENDIX B: CLOSED-FORM EXPRESSION FOR THE ASYMPTOTIC VARIANCE, GIVEN BIDIRECTIONAL DATA

In this appendix, we obtain a closed-form expression for the asymptotic variance of the optimal EBS estimate for $\overline{\mathcal{F}}_f$, given bidirectional data. We will follow a similar procedure as in Appendix A. For the bidirectional case, the weight matrix M consists of three columns, $M = [m_f m_r m_{\mathcal{F}_f}]$, where m_i is a column matrix of weights from Eqs. (8) and (11) corresponding to path ensemble i. The elements of M are

$$M_{nf} = \frac{\hat{c}_f^{-1} q_f[X_n]}{N_f \, \hat{c}_f^{-1} q_f[X_n] + N_r \, \hat{c}_r^{-1} q_r[\widetilde{X}_n]}$$

$$= \frac{1}{N_f + N_r \, e^{-\hat{\Omega}[X_n]}} = N_f^{-1} \epsilon(L_n), \tag{B1}$$

$$M_{nr} = \frac{\hat{c}_r^{-1} q_r[\widetilde{X}_{fn}]}{N_f \, \hat{c}_f^{-1} q_f[X_n] + N_r \, \hat{c}_r^{-1} q_r[\widetilde{X}_n]}$$

$$= \frac{1}{N_f \, e^{\hat{\Omega}[X_n]} + N_r}$$

$$= N_r^{-1} \epsilon(-L_n), \tag{B2}$$

$$M_{n\mathcal{F}_f} = \left(\frac{\mathcal{F}[X]}{\overline{\mathcal{F}}_f}\right) \frac{1}{N_f + N_r \, e^{-\hat{\Omega}[X_n]}} = \left(\frac{\mathcal{F}[X]}{\overline{\mathcal{F}}_f}\right) N_f^{-1} \epsilon(L_n), \tag{B3}$$

where ϵ is defined as the Fermi function, $\epsilon(L_n) = 1/1 + e^{-L_n}$, and we define $L_n = W[X_n] - \Delta\hat{f}_t + \ln(N_f/N_r)$. This allows us to write $\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M}$ as

$$\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M} = \sum_{n=1}^{N} \begin{bmatrix} \frac{1}{N_f^2}\epsilon(L_n)^2 & \frac{1}{N_f N_r}\epsilon(L_n)\epsilon(-L_n) & \frac{1}{N_f^2}\left(\frac{\mathcal{F}[X]}{\overline{\mathcal{F}}_f}\right)\epsilon(L_n)^2 \\ \frac{1}{N_f N_r}\epsilon(L_n)\epsilon(-L_n) & \frac{1}{N_r^2}\epsilon(-L_n)^2 & \frac{1}{N_f N_r}\left(\frac{\mathcal{F}[X]}{\overline{\mathcal{F}}_f}\right)\epsilon(L_n)\epsilon(-L_n) \\ \frac{1}{N_f^2}\left(\frac{\mathcal{F}[X]}{\overline{\mathcal{F}}_f}\right)\epsilon(L_n)^2 & \frac{1}{N_f N_r}\left(\frac{\mathcal{F}[X]}{\overline{\mathcal{F}}_f}\right)\epsilon(L_n)\epsilon(-L_n) & \frac{1}{N_f^2}\left(\frac{\mathcal{F}[X]}{\overline{\mathcal{F}}_f}\right)^2\epsilon(L_n)^2 \end{bmatrix} \equiv \begin{bmatrix} a_{ff} & a_{fr} & a_{f\mathcal{F}_f} \\ a_{fr} & a_{rr} & a_{r\mathcal{F}_f} \\ a_{\mathcal{F}_f\mathcal{F}} & a_{r\mathcal{F}_f} & a_{\mathcal{F}_f\mathcal{F}_f} \end{bmatrix}. \tag{B4}$$

Using the determinant,

$$D = -a_{\mathcal{F}_f\mathcal{F}_f}a_{fr}^2 + 2a_{f\mathcal{F}_f}a_{fr}a_{r\mathcal{F}_f} - a_{ff}a_{r\mathcal{F}_f}^2 - a_{f\mathcal{F}_f}^2 a_{rr} + a_{\mathcal{F}_f\mathcal{F}_f}a_{ff}a_{rr}, \tag{B5}$$

we write the inverse covariance matrix estimator as

$$\hat{\boldsymbol{\Theta}}^{-1} = \begin{bmatrix} \frac{-a_{r\mathcal{F}_f}^2 + a_{\mathcal{F}_f\mathcal{F}_f}a_{rr}}{D} - N_f + b & \frac{-a_{\mathcal{F}_f\mathcal{F}_f}a_{fr} + a_{f\mathcal{F}_f}a_{r\mathcal{F}_f}}{D} + b & \frac{a_{fr}a_{r\mathcal{F}_f} - a_{f\mathcal{F}_f}a_{rr}}{D} + b \\ \frac{-a_{\mathcal{F}_f\mathcal{F}_f}a_{fr} + a_{f\mathcal{F}_f}a_{r\mathcal{F}_f}}{D} + b & \frac{-a_{f\mathcal{F}_f}^2 + a_{\mathcal{F}_f\mathcal{F}_f}a_{ff}}{D} - N_r + b & \frac{a_{f\mathcal{F}_f}a_{fr} - a_{ff}a_{r\mathcal{F}_f}}{D} + b \\ \frac{a_{fr}a_{r\mathcal{F}_f} - a_{f\mathcal{F}_f}a_{rr}}{D} + b & \frac{a_{f\mathcal{F}_f}a_{fr} - a_{ff}a_{r\mathcal{F}_f}}{D} + b & \frac{-a_{fr}^2 + a_{ff}a_{rr}}{D} + b \end{bmatrix}. \tag{B6}$$

By applying the same steps as Appendix E of Shirts and Chodera,[27] we obtain the determinant

$$|\hat{\boldsymbol{\Theta}}^{-1}| = \frac{9b(a_{fr}^2 N_f + a_{fr} a_{rr} N_f)}{D}. \tag{B7}$$

Applying Eq. (13) to $\hat{\boldsymbol{\Theta}}$ and simplifying, it can be shown that the variance estimate is

$$\sigma^2(\overline{\mathcal{F}}_f^2) \approx \overline{\mathcal{F}}_f^2(\Theta_{\mathcal{F}_f\mathcal{F}_f} - 2\Theta_{\mathcal{F}_f f} + \Theta_{ff}), \tag{B8}$$

$$= \frac{a_{\mathcal{F}_f\mathcal{F}_f}a_{fr} - a_{f\mathcal{F}_f}a_{fr} - a_{f\mathcal{F}_f}a_{r\mathcal{F}_f} + a_{ff}\,a_{r\mathcal{F}_f}}{a_{fr}(a_{fr}\,N_f + a_{rr}\,N_r)}. \tag{B9}$$

[1] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).
[2] C. Jarzynski, Phys. Rev. E **56**, 5018 (1997).
[3] G. E. Crooks, Phys. Rev. E **61**, 2361 (2000).
[4] R. M. Neal, Stat. Comput. **11**, 125 (2001).
[5] G. Hummer and A. Szabo, Proc. Natl. Acad. Sci. U.S.A. **98**, 3658 (2001).
[6] G. Hummer and A. Szabo, Acc. Chem. Res. **38**, 504 (2005).
[7] D. D. L. Minh, Phys. Rev. E **74**, 061120 (2006).
[8] D. D. L. Minh, J. Phys. Chem. B **111**, 4137 (2007).
[9] I. Junier, A. Mossa, M. Manosas, and F. Ritort, Phys. Rev. Lett. **102**, 070602 (2009).
[10] E. Lyman and D. M. Zuckerman, J. Chem. Phys. **127**, 065101 (2007).
[11] J. Nummela, F. Yassin, and I. Andricioaei, J. Chem. Phys. **128**, 024104 (2008).
[12] E. H. Feng and G. E. Crooks, Phys. Rev. E **79**, 012104 (2009).
[13] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, Jr., and C. Bustamante, Science **296**, 1832 (2002).
[14] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco, and C. Bustamante, Nature (London) **437**, 231 (2005).
[15] C. H. Bennett, J. Comput. Phys. **22**, 245 (1976).
[16] P. Maragakis, M. Spichty, and M. Karplus, Phys. Rev. Lett. **96**, 100602 (2006).
[17] P. Maragakis, F. Ritort, C. Bustamante, M. Karplus, and G. E. Crooks, J. Chem. Phys. **129**, 024102 (2008).
[18] M. R. Shirts and V. S. Pande, J. Chem. Phys. **122**, 144107 (2005).
[19] C. P. Calderon, L. Janosi, and I. Kosztin, J. Chem. Phys. **130**, 144908 (2009).
[20] Z. Tan, J. Am. Stat. Assoc. **99**, 1027 (2004).
[21] N. Singhal and V. S. Pande, J. Chem. Phys. **123**, 204909 (2005).
[22] N. S. Hinrichs and V. S. Pande, J. Chem. Phys. **126**, 244101 (2007).
[23] A. M. Hahn and H. Then, e-print arXiv:0904.0625v2.
[24] Y. Vardi, Ann. Stat. **13**, 178 (1985).
[25] R. D. Gill, Y. Vardi, and J. A. Wellner, Ann. Stat. **16**, 1069 (1988).
[26] A. Kong, P. McCullagh, X.-L. Meng, D. Nicolae, and Z. Tan, J. R. Stat. Soc. Ser. B (Stat. Methodol.) **65**, 585 (2003).
[27] M. R. Shirts and J. D. Chodera, J. Chem. Phys. **129**, 124105 (2008).
[28] D. D. L. Minh and A. B. Adib, Phys. Rev. Lett. **100**, 180602 (2008).
[29] H. Doss makes this suggestion in the conference discussion of Ref. 26.
[30] S. Sun, J. Chem. Phys. **118**, 5769 (2003).
[31] F. M. Ytreberg and D. M. Zuckerman, J. Chem. Phys. **120**, 10876 (2004).

[32] L. Pratt, J. Chem. Phys. **85**, 5045 (1986).

[33] C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, J. Chem. Phys. **108**, 1964 (1998).

[34] J. Nummela and I. Andricioaei, Biophys. J. **93**, 3373 (2007).

[35] G. E. Crooks, J. Stat. Phys. **90**, 1481 (1998).

[36] G. E. Crooks, Phys. Rev. E **60**, 2721 (1999).

[37] C. Jarzynski, Phys. Rev. E **73**, 046105 (2006).

[38] L. Lu and T. B. Woolf, in *Free Energy Calculations*, edited by C. Chipot and A. Pohorille (Springer, Berlin, 2007), Vol. 86.

[39] D. D. L. Minh, J. Chem. Phys. **130**, 204102 (2009).

[40] H. Oberhofer, C. Dellago, and P. Geissler, J. Phys. Chem. B **109**, 6902 (2005).

[41] M. R. Shirts, E. Bair, G. Hooker, and V. S. Pande, Phys. Rev. Lett. **91**, 140601 (2003).

[42] D. D. L. Minh and J. A. McCammon, J. Phys. Chem. B **112**, 5892 (2008).

[43] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989).

[44] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg, J. Comput. Chem. **13**, 1011 (1992).

[45] E. Veach and L. J. Guibas, *Optimally combining sampling techniques for Monte Carlo rendering*, SIGGRAPH '95: Proceedings of the 22nd annual conference on computer graphics and interactive techniques (ACM, New York, 1995), pp. 419–428.

[46] E. Veach, Ph.D. dissertation, Stanford University, 1997.

[47] M. Souaille and B. Roux, Comput. Phys. Commun. **135**, 40 (2001).

[48] H. Oberhofer and C. Dellago, J. Comput. Chem. **30**, 1726 (2009).

[49] D. M. Zuckerman and T. B. Woolf, Phys. Rev. Lett. **89**, 180602 (2002).

[50] J. Gore, F. Ritort, and C. Bustamante, Proc. Natl. Acad. Sci. U.S.A. **100**, 12564 (2003).

[51] D. M. Zuckerman and T. B. Woolf, J. Stat. Phys. **114**, 1303 (2004).

[52] E. Atilgan and S. X. Sun, J. Chem. Phys. **121**, 10392 (2004).

[53] H. Oberhofer and C. Dellago, Comput. Phys. Commun. **179**, 41 (2008).

[54] A. E. Mark, W. F. van Gunsteren, and H. J. C. Berendsen, J. Chem. Phys. **94**, 3808 (1991).

[55] W. P. Reinhardt and J. E. Hunter, J. Chem. Phys. **97**, 1599 (1992).

[56] J. E. Hunter, W. P. Reinhardt, and T. F. Davis, J. Chem. Phys. **99**, 6856 (1993).

[57] J. C. Schon, J. Chem. Phys. **105**, 10072 (1996).

[58] C. Jarque and B. Tidor, J. Phys. Chem. B **101**, 9402 (1997).

[59] T. Schmiedl and U. Seifert, Phys. Rev. Lett. **98**, 108301 (2007).

[60] H. Then and A. Engel, Phys. Rev. E **77**, 041105 (2008).

[61] A. Gomez-Marin, T. Schmiedl, and U. Seifert, J. Chem. Phys. **129**, 024114 (2008).

[62] J. D. Chodera, W. C. Swope, J. W. Pitera, C. Seok, and K. A. Dill, J. Chem. Theory Comput. **3**, 26 (2007).