Characteristic of Bennett's acceptance ratio method

Aljoscha M. Hahn* and Holger Then
Institut für Physik, Carl von Ossietzky Universität, 26111 Oldenburg, Germany
(Received 12 June 2009; published 10 September 2009)

A powerful and well-established tool for free-energy estimation is Bennett's acceptance ratio method. Central properties of this estimator, which employs samples of work values of a forward and its time-reversed process, are known: for given sets of measured work values, it results in the best estimate of the free-energy difference in the large sample limit. Here we state and prove a further characteristic of the acceptance ratio method: the convexity of its mean-square error. As a two-sided estimator, it depends on the ratio of the numbers of forward and reverse work values used. Convexity of its mean-square error immediately implies that there exists a unique optimal ratio for which the error becomes minimal. Further, it yields insight into the relation of the acceptance ratio method and estimators based on the Jarzynski equation. As an application, we study the performance of a dynamic strategy of sampling forward and reverse work values.

DOI: 10.1103/PhysRevE.80.031111 PACS number(s): 02.50.Fz, 05.40.−a, 05.70.Ln

I. INTRODUCTION

A quantity of central interest in thermodynamics and statistical physics is the (Helmholtz) free energy, as it determines the equilibrium properties of the system under consideration. In practical applications, e.g., drug design, molecular association, thermodynamic stability, and binding affinity, it is usually sufficient to know free-energy differences. As recent progress in statistical physics has shown, free-energy differences, which refer to equilibrium, can be determined via nonequilibrium processes [1,2].

Typically, free-energy differences are beyond the scope of analytic computations and one needs to measure them experimentally or compute them numerically. Highly efficient methods have been developed in order to estimate free-energy differences precisely, including thermodynamic integration [3,4], free-energy perturbation [5], umbrella sampling [6–8], adiabatic switching [9], dynamic methods [10–12], asymptotics of work distributions [13], optimal protocols [14], targeted, and escorted free-energy perturbation [15–19].

A powerful [20–22] and frequently [23–25] used method for free-energy determination is two-sided estimation, i.e., Bennett's acceptance ratio method [26], which employs a sample of work values of a driven nonequilibrium process together with a sample of work values of the time-reversed process [27].

The performance of two-sided free-energy estimation depends on the ratio

$$r = \frac{n_1}{n_0} \tag{1}$$

of the number of forward and reverse work values used. Think of an experimenter who wishes to estimate the free-energy difference with Bennett's acceptance ratio method and has the possibility to generate forward as well as reverse work values. The capabilities of the experiment give rise to an obvious question: if the total amount of draws is intended to be $N=n_0+n_1$, which is the optimal choice of partitioning N into the numbers n_0 of forward and n_1 of reverse work values or, equivalently, what is the optimal choice r_o of the ratio r? The problem is to determine the value of r that minimizes the (asymptotic) mean-square error of Bennett's estimator when $N=n_0+n_1$ is held constant.

While known since Bennett [26], the optimal ratio is underutilized in the literature. Bennett himself proposed to use a suboptimal equal-time strategy, instead, because his estimator for the optimal ratio converges too slowly in order to be practicable. Even questions as fundamental as the existence and uniqueness are unanswered in the literature. Moreover, it is not always clear *a priori* whether two-sided free-energy estimation is better than one-sided exponential work averaging. For instance, Shirts and Pande presented a physical example where it is optimal to draw work values from only one direction [28].

The paper is organized as follows. In Secs. II and III we rederive two-sided free-energy estimation and the optimal ratio. We also remind that two-sided estimation comprises one-sided exponential work averaging as limiting cases for $\ln r \to \pm\infty$, a result that is also true for the mean-square errors of the corresponding estimators.

The central result is stated in Sec. IV: the asymptotic mean-square error of two-sided estimation is convex in the fraction $\frac{n_0}{N}$ of forward work values used. This fundamental characteristic immediately implies that the optimal ratio r_o exists and is unique. Moreover, it explains the generic superiority of two-sided estimation if compared with one sided, as found in many applications.

To overcome the slow convergence of Bennett's estimator of the optimal ratio, which is based on estimating second moments, in Sec. V we transform the problem into another form such that the corresponding estimator is entirely based on first moments, which enhances the convergence enormously.

As an application, in Sec. VII we present a dynamic strategy of sampling forward and reverse work values that maximizes the efficiency of two-sided free-energy estimation.

*Present address: Technische Universität Berlin, Institut für Theoretische Physik, 10623 Berlin, Germany.

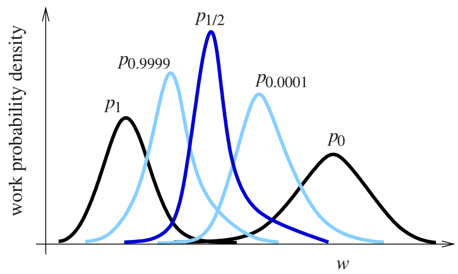

FIG. 1. (Color online) The overlap density $p_\alpha(w)$ bridges the densities $p_0(w)$ and $p_1(w)$ of forward and reverse work values, respectively. α is the fraction $\frac{n_0}{n_0+n_1}$ of forward work values, here schematically shown for $\alpha=0.0001$, $\alpha=0.5$, and $\alpha=0.9999$. The accuracy of two-sided free-energy estimates depends on how good $p_\alpha(w)$ is sampled when drawing from $p_0(w)$ and $p_1(w)$.

II. TWO-SIDED FREE-ENERGY ESTIMATION

Given a pair of samples of n_0 forward and n_1 reverse work values drawn from the probability densities $p_0(w)$ and $p_1(w)$ of forward and reverse work values and provided the latter are related to each other via the fluctuation theorem [2],

$$\frac{p_0(w)}{p_1(w)} = e^{w-\Delta f}, \tag{2}$$

Bennett's acceptance ratio method [20,26,27,29] is known to give the optimal estimate of the free-energy difference Δf in the limit of large sample sizes. Throughout the paper, $\Delta f = \Delta F/kT$ and $w=W/kT$ are understood to be measured in units of the thermal energy kT. The normalized probability densities $p_0(w)$ and $p_1(w)$ are assumed to have the same support Ω, and we choose the following sign convention: $p_0(w) := p_{\text{forward}}(+w)$ and $p_1(w) := p_{\text{reverse}}(-w)$.

Now define a normalized density $p_\alpha(w)$ with

$$p_\alpha(w) = \frac{1}{U_\alpha}\frac{p_0(w)p_1(w)}{\alpha p_0(w) + \beta p_1(w)}, \tag{3}$$

$w \in \Omega$, where $\alpha \in [0,1]$ is a real number and

$$\alpha + \beta = 1. \tag{4}$$

The normalization constant U_α is given by

$$U_\alpha = \int_\Omega \frac{p_0 p_1}{\alpha p_0 + \beta p_1} dw. \tag{5}$$

The density $p_\alpha(w)$ is a normalized harmonic mean of p_0 and p_1, $\frac{p_0 p_1}{\alpha p_0 + \beta p_1} = [\alpha \frac{1}{p_1} + \beta \frac{1}{p_0}]^{-1}$, and thus bridges between p_0 and p_1 (see Fig. 1). In the limit $\alpha \to 0$, $p_\alpha(w)$ converges to the forward work density $p_0(w)$ and, conversely, for $\alpha \to 1$ it converges to the reverse density $p_1(w)$. As a consequence of the inequality of the harmonic and arithmetic mean $[\alpha \frac{1}{p_1} + \beta \frac{1}{p_0}]^{-1} \leq \alpha p_1 + \beta p_0$, U_α is bounded from above by unity,

$$U_\alpha \leq 1, \tag{6}$$

$\forall \alpha \in [0,1]$. Except for $\alpha=0$ and $\alpha=1$, the equality holds if and only if $p_0 \equiv p_1$. Using the fluctuation theorem (2), U_α can be written as an average in p_0 and p_1,

$$U_\alpha = \left\langle \frac{1}{\alpha + \beta e^{-w+\Delta f}} \right\rangle_1 = \left\langle \frac{1}{\beta + \alpha e^{w-\Delta f}} \right\rangle_0, \tag{7}$$

where the angular brackets with subscript $\gamma \in [0,1]$ denote an ensemble average with respect to p_γ, i.e.,

$$\langle g \rangle_\gamma = \int_\Omega g(w) p_\gamma(w) dw, \tag{8}$$

for an arbitrary function $g(w)$.

In setting $\alpha=1$, Eq. (7) reduces to the nonequilibrium work relation [1],

$$1 = \langle e^{-w+\Delta f} \rangle_0, \tag{9}$$

in the *forward* direction and, conversely, with $\alpha=0$ we obtain the nonequilibrium work relation in the *reverse* direction,

$$1 = \langle e^{w-\Delta f} \rangle_1. \tag{10}$$

The last two relations can, of course, be obtained more directly from the fluctuation theorem (2). An important application of these relations is the *one-sided* free-energy estimation. Given a sample $\{w_1^0 \dots w_N^0\}$ of N forward work values drawn from p_0, Eq. (9) is commonly used to define the *forward* estimate $\widehat{\Delta f}_0$ of Δf with

$$\widehat{\Delta f}_0 = -\ln \frac{1}{N}\sum_{k=1}^{N} e^{-w_k^0}. \tag{11}$$

Conversely, given a sample $\{w_1^1 \dots w_N^1\}$ of N reverse work values drawn from p_1, Eq. (10) suggests the definition of the *reverse* estimate $\widehat{\Delta f}_1$ of Δf,

$$\widehat{\Delta f}_1 = \ln \frac{1}{N}\sum_{l=1}^{N} e^{w_l^1}. \tag{12}$$

If we have drawn both, a sample of n_0 forward *and* a sample of n_1 reverse work values then Eq. (7) can serve us to define a two-sided estimate $\widehat{\Delta f}$ of Δf by replacing the ensemble averages with sample averages,

$$\frac{1}{n_1}\sum_{l=1}^{n_1} \frac{1}{\alpha + \beta e^{-w_l^1 + \widehat{\Delta f}}} = \frac{1}{n_0}\sum_{k=1}^{n_0} \frac{1}{\beta + \alpha e^{w_k^0 - \widehat{\Delta f}}}. \tag{13}$$

$\widehat{\Delta f}$ is understood to be the unique root of Eq. (13), which exists for any $\alpha \in [0,1]$. Different values of α result in different estimates for Δf. Choosing

$$\alpha = \frac{n_0}{N}, \quad \beta = \frac{n_1}{N}, \tag{14}$$

$N = n_0 + n_1$, the estimate (13) coincides with Bennett's optimal estimate, which defines the two-sided estimate with a least asymptotic mean-square error for a given value $\alpha = \frac{n_0}{N}$ or, equivalently, *for a given ratio* $r = \frac{\beta}{\alpha} = \frac{n_1}{n_0}$ [20,26]. We denote the optimal two-sided estimate, i.e., the solution of Eq. (13) under the constraint (14), by $\widehat{\Delta f}_{1-\alpha}$ and simply refer to it as the two-sided estimate. Note that the optimal estimator can be written in the familiar form

$$\sum_{l=1}^{n_1} \frac{1}{1+e^{-w_l^1+\widehat{\Delta f}+\ln n_1/n_0}} = \sum_{k=1}^{n_0} \frac{1}{1+e^{w_k^0-\widehat{\Delta f}-\ln n_1/n_0}}. \qquad (15)$$

In the limit $\alpha=\frac{n_0}{N}\to 1$, the two-sided estimate reduces to the one-sided forward estimate (11), $\widehat{\Delta f}_{1-\alpha} \overset{\alpha\to 1}{\longrightarrow} \widehat{\Delta f}_0$, and, conversely, $\widehat{\Delta f}_{1-\alpha} \overset{\alpha\to 0}{\longrightarrow} \widehat{\Delta f}_1$. Thus, the one-sided estimates are the optimal estimates if we have given draws from only one of the densities p_0 or p_1.

A characteristic quantity to express the performance of the estimate $\widehat{\Delta f}_{1-\alpha}$ is the mean-square error,

$$\langle(\widehat{\Delta f}_{1-\alpha}-\Delta f)^2\rangle, \qquad (16)$$

which depends on the total sample size $N=n_0+n_1$ and the fraction $\alpha=\frac{n_0}{N}$. Here, the average is understood to be an ensemble average in the value distribution of the estimate $\widehat{\Delta f}_{1-\alpha}$ for fixed N and α. In the limit of large n_0 and n_1, the asymptotic mean-square error X (which then equals the variance) can be written as [20,26]

$$X(N,\alpha) = \frac{1}{N}\frac{1}{\alpha\beta}\left(\frac{1}{U_\alpha}-1\right). \qquad (17)$$

Provided the right-hand side of Eq. (17) exists, which is guaranteed for any $\alpha\in(0,1)$, the N dependence of X is simply given by the usual $\frac{1}{N}$ factor, whereas the α dependence is determined by the function U_α given in Eq. (5). Note that if a two-sided estimate $\widehat{\Delta f}_{1-\alpha}$ is calculated then essentially the normalizing constant U_α is estimated from two sides 0 and 1 [cf. Eqs. (7) and (13)]. With an estimate $\widehat{\Delta f}_{1-\alpha}$, we therefore always have an estimate of the mean-square error at hand. However, the reliability of the latter naturally depends on the degree of convergence of the estimate $\widehat{\Delta f}_{1-\alpha}$. The convergence of the two-sided estimate can be checked with the convergence measure introduced in Ref. [19].

In the limits $\alpha=\frac{n_0}{N}\to 1$ and $\alpha\to 0$, respectively, the asymptotic mean-square error X of the two-sided estimator converges to the asymptotic mean-square error of the appropriate one-sided estimator [30],

$$\lim_{\alpha\to 1} X(N,\alpha) = \frac{1}{N}\mathrm{Var}_0\left(\frac{p_1}{p_0}\right) = \frac{1}{N}\mathrm{Var}_0(e^{-w+\Delta f}), \qquad (18)$$

and

$$\lim_{\alpha\to 0} X(N,\alpha) = \frac{1}{N}\mathrm{Var}_1\left(\frac{p_0}{p_1}\right) = \frac{1}{N}\mathrm{Var}_1(e^{w-\Delta f}), \qquad (19)$$

where Var_γ denotes the variance operator with respect to the density p_γ, i.e.,

$$\mathrm{Var}_\gamma(g) = \langle(g-\langle g\rangle_\gamma)^2\rangle_\gamma, \qquad (20)$$

for an arbitrary function $g(w)$ and $\gamma\in[0,1]$.

III. THE OPTIMAL RATIO

Now we focus on the question raised in the introduction: which value α_o of α in the range $[0,1]$ minimizes the mean-square error (17) when the total sample size $N=n_0+n_1$ is held fixed?

Let M be the rescaled asymptotic mean-square error given by

$$M(\alpha) = NX(N,\alpha), \qquad (21)$$

which is a function of α only. Assuming $\alpha_o\in(0,1)$, a necessary condition for a minimum of M is that the derivative $M'(\alpha)=\frac{dM}{d\alpha}$ of M vanishes at α_o. Before calculating M' explicitly, it is beneficial to rewrite M by using the identity

$$U_\alpha = \int_\Omega \frac{p_0p_1(\alpha p_0+\beta p_1)}{(\alpha p_0+\beta p_1)^2}dw = \alpha\left\langle\frac{p_0^2}{(\alpha p_0+\beta p_1)^2}\right\rangle_1 + \beta\left\langle\frac{p_1^2}{(\alpha p_0+\beta p_1)^2}\right\rangle_0. \qquad (22)$$

Subtracting $(\alpha+\beta)U_\alpha^2=U_\alpha^2$ from Eq. (22) and recalling the definition (3) of p_α, one obtains

$$U_\alpha(1-U_\alpha) = [\alpha\theta_1(\alpha)+\beta\theta_0(\alpha)]U_\alpha^2, \qquad (23)$$

where the functions θ_i are defined as

$$\theta_1(\alpha) = \mathrm{Var}_1\left(\frac{p_\alpha}{p_1}\right) = \frac{1}{U_\alpha^2}\mathrm{Var}_1\left(\frac{1}{\alpha+\beta e^{-w+\Delta f}}\right),$$

$$\theta_0(\alpha) = \mathrm{Var}_0\left(\frac{p_\alpha}{p_0}\right) = \frac{1}{U_\alpha^2}\mathrm{Var}_0\left(\frac{1}{\beta+\alpha e^{w-\Delta f}}\right). \qquad (24)$$

θ_0 and θ_1 describe the relative fluctuations of the quantities that are averaged in the two-sided estimation of Δf [cf. Eq. (13)].

With the use of formula (23), M can be written as

$$M(\alpha) = \frac{\theta_0(\alpha)}{\alpha} + \frac{\theta_1(\alpha)}{\beta}, \qquad (25)$$

and the derivative yields

$$M'(\alpha) = \frac{\theta_1(\alpha)}{\beta^2} - \frac{\theta_0(\alpha)}{\alpha^2} + \frac{\beta\theta_0'(\alpha)+\alpha\theta_1'(\alpha)}{\alpha\beta}. \qquad (26)$$

The derivatives of the θ functions involve the first two derivatives of U_α, which will thus be computed first,

$$U_\alpha' := \frac{d}{d\alpha}U_\alpha = \int_\Omega \frac{p_0p_1(p_1-p_0)}{(\alpha p_0+\beta p_1)^2}dw, \qquad (27)$$

and

$$U_\alpha'' := \frac{d^2}{d\alpha^2}U_\alpha = 2\int_\Omega \frac{p_0p_1(p_1-p_0)^2}{(\alpha p_0+\beta p_1)^3}dw. \qquad (28)$$

From this equation, it is clear that U_α is convex in α, $U_\alpha''\geq 0$, with a unique minimum in (0,1) (as $U_0=U_1=1$). We can rewrite the θ functions with U_α and U_α' as follows:

$$\theta_1(\alpha) = \frac{U_\alpha-\beta U_\alpha'}{U_\alpha^2} - 1,$$

$$\theta_0(\alpha) = \frac{U_\alpha + \alpha U'_\alpha}{U_\alpha^2} - 1. \tag{29}$$

Differentiating these expressions gives

$$\theta'_1(\alpha) = -\frac{\beta}{U_\alpha^3}(U''_\alpha U_\alpha - 2U'^2_\alpha),$$

$$\theta'_0(\alpha) = \frac{\alpha}{U_\alpha^3}(U''_\alpha U_\alpha - 2U'^2_\alpha). \tag{30}$$

θ_0 and θ_1 are monotonically increasing and decreasing, respectively. This immediately follows from writing the term occurring in the brackets of Eqs. (30) as a variance in the density p_α,

$$U''_\alpha U_\alpha - 2U'^2_\alpha = 2\,\mathrm{Var}_\alpha\left(\frac{p_1 - p_0}{\alpha p_0 + \beta p_1}\right)U_\alpha^2, \tag{31}$$

which is thus positive.

As a consequence of Eq. (30), the relation

$$\beta\theta'_0(\alpha) + \alpha\theta'_1(\alpha) = 0 \quad \forall\, \alpha \in [0,1] \tag{32}$$

holds and M' reduces to

$$M'(\alpha) = \frac{\theta_1(\alpha)}{\beta^2} - \frac{\theta_0(\alpha)}{\alpha^2}. \tag{33}$$

The derivatives of the θ functions do not contribute to M' due to the fact that the specific form of the two-sided estimator (13) originates from minimizing the asymptotic mean-square error (cf. [26]). The necessary condition for a local minimum of M at α_o, $M'(\alpha_o)=0$, now reads as

$$\frac{\beta_o^2}{\alpha_o^2} = \frac{\theta_1(\alpha_o)}{\theta_0(\alpha_o)}, \tag{34}$$

where $\beta_o = 1-\alpha_o$ is introduced. Using Eqs. (24) and (2), the condition (34) results in

$$\mathrm{Var}_1\left(\frac{1}{1+e^{-w+\Delta f+\ln r_o}}\right) = \mathrm{Var}_0\left(\frac{1}{1+e^{w-\Delta f-\ln r_o}}\right). \tag{35}$$

This means, the optimal ratio r_o is such that the variances of the random functions, which are averaged in the two-sided estimation (15), are equal. However, the existence of a solution of $M'(\alpha)=0$ is not guaranteed in general.

Writing Eq. (35) in the form

$$\mathrm{Var}_1\left(\frac{p_1 - p_0}{\alpha p_0 + \beta p_1}\right) = \mathrm{Var}_0\left(\frac{p_1 - p_0}{\alpha p_0 + \beta p_1}\right) \tag{36}$$

prevents the equation from becoming a tautology.

IV. CONVEXITY OF THE MEAN-SQUARE ERROR

Theorem. The asymptotic mean-square error $M(\alpha)$ is convex in α.

In order to prove the convexity, we introduce the operator $\Gamma_\alpha(f)$, which is defined for an arbitrary function $f(w)$ by

$$\Gamma_\alpha(f) = \beta\,\mathrm{Var}_0(f) + \alpha\,\mathrm{Var}_1(f) - U_\alpha\,\mathrm{Var}_\alpha(f). \tag{37}$$

Lemma. Γ_α is positive semidefinite, i.e.,

$$\Gamma_\alpha(f) \geq 0 \quad \forall\, f(w). \tag{38}$$

For $\alpha \in (0,1)$ and $f(w) \neq \mathrm{const}$, the equality holds if and only if $p_0 \equiv p_1$.

Proof of the Lemma. Let $\delta f_\gamma = f(w) - \langle f\rangle_\gamma$, $\gamma \in [0,1]$. Then

$$\begin{aligned}\Gamma_\alpha(f) &= \int_\Omega \left(\beta\delta f_0^2 p_0 + \alpha\delta f_1^2 p_1 - \delta f_\alpha^2 \frac{p_0 p_1}{\alpha p_0 + \beta p_1}\right)dw \\ &= \int_\Omega \frac{(\beta\delta f_0^2 p_0 + \alpha\delta f_1^2 p_1)(\alpha p_0 + \beta p_1) - \delta f_\alpha^2 p_0 p_1}{\alpha p_0 + \beta p_1}dw \\ &= \alpha\beta\int_\Omega \frac{(\delta f_1 p_1 - \delta f_0 p_0)^2}{\alpha p_0 + \beta p_1}dw + U_\alpha(\beta\langle f\rangle_0 + \alpha\langle f\rangle_1 \\ &\quad - \langle f\rangle_\alpha)^2,\end{aligned} \tag{39}$$

which is clearly positive. Provided $f \neq \mathrm{const}$ and $\alpha \neq 0,1$, the integrand in the last line is zero $\forall w$ if and only if $p_0 \equiv p_1$. This completes the proof of the lemma. □

Proof of the Theorem. Consulting Eqs. (33) and (32), the second derivative of M reads as

$$M''(\alpha) = 2\left(\frac{\theta_1(\alpha)}{\beta^3} + \frac{\theta_0(\alpha)}{\alpha^3}\right) - \frac{1}{\alpha^2\beta}\theta'_0(\alpha). \tag{40}$$

Expressing $p_0 = p - \beta d$ and $p_1 = p + \alpha d$ in center and relative "coordinates" $p = \alpha p_0 + \beta p_1$ and $d = p_1 - p_0$, respectively, gives

$$\theta_1(\alpha) = \frac{1}{U_\alpha^2}\mathrm{Var}_1\left(\frac{p_0}{p}\right) = \frac{\beta^2}{U_\alpha^2}\mathrm{Var}_1\left(\frac{d}{p}\right),$$

$$\theta_0(\alpha) = \frac{1}{U_\alpha^2}\mathrm{Var}_0\left(\frac{p_1}{p}\right) = \frac{\alpha^2}{U_\alpha^2}\mathrm{Var}_0\left(\frac{d}{p}\right),$$

$$\theta'_0(\alpha) = \frac{2\alpha}{U_\alpha}\mathrm{Var}_\alpha\left(\frac{d}{p}\right). \tag{41}$$

Therefore, $\frac{1}{2}\alpha\beta U_\alpha^2 M'' = \Gamma_\alpha(\frac{d}{p})$, which is positive according to the lemma. □

The convexity of the mean-square error is a fundamental characteristic of Bennett's acceptance ratio method. This characteristic allows us to state a simple criterion for the existence of a *local* minimum of the mean-square error in terms of its derivatives at the boundaries. Namely, if

$$M'(0) = \mathrm{Var}_1(e^{w-\Delta f}) - \mathrm{Var}_0(e^{w-\Delta f}) \tag{42}$$

is negative and

$$M'(1) = \mathrm{Var}_1(e^{-w+\Delta f}) - \mathrm{Var}_0(e^{-w+\Delta f}) \tag{43}$$

is positive there exists a local minimum of $M(\alpha)$ for $\alpha \in (0,1)$. Otherwise, no local minimum exists and the global minimum is found on the boundaries of α: if $M'(0) > 0$, the global minimum is found for $\alpha=0$; thus, it is optimal to measure work values in the reverse direction only and to use the one-sided reverse estimator (12). Else, if $M'(1) < 0$, the global minimum is found for $\alpha=1$, implying the one-sided forward estimator (11) to be optimal.

In addition, the convexity of the mean-square error proves the existence and uniqueness of the optimal ratio since a convex function has a global minimum on a closed interval.

Corollary. If a solution of $M'(\alpha)=0$ exists, it is unique and $M(\alpha)$ attains its global minimum ($\alpha\in[0,1]$) there.

V. ESTIMATING THE OPTIMAL RATIO WITH FIRST MOMENTS

In situations of practical interest, the optimal ratio is not available *a priori*. Thus, we are going to estimate the optimal ratio. There exist estimators of the optimal ratio since Bennett. In addition, we have just proven that the optimal ratio exists and is unique. However, there is still one obstacle to overcome. Yet, all expressions for estimating the optimal ratio are based on second moments [see, e.g., Eq. (35)]. Due to convergence issues, it is not practicable to base any estimator on expressions that involve second moments. The estimator would converge far too slowly. For this reason, we transform the problem into a form that employs first moments only.

Assume we have given n_0 and n_1 work values in forward and reverse direction, respectively, and want to estimate U_a, with $0\leq a\leq 1$. According to Eq. (7), we can estimate the overlap measure U_a by using draws from the forward direction,

$$\hat{U}_a^{(0)}=\frac{1}{n_0}\sum_{k=1}^{n_0}\frac{1}{b+ae^{w_k^0-\widehat{\Delta f}}},\tag{44}$$

where b equals $1-a$ and for $\widehat{\Delta f}$ the best available estimate of Δf is inserted, i.e., the two-sided estimate based on the n_0+n_1 work values. Similarly, we can estimate the overlap measure by using draws from the reverse direction,

$$\hat{U}_a^{(1)}=\frac{1}{n_1}\sum_{l=1}^{n_1}\frac{1}{a+be^{-w_l^1+\widehat{\Delta f}}}.\tag{45}$$

Since in general draws from both directions are available, it is reasonable to take an arithmetic mean of both estimates

$$\hat{U}_a=a\hat{U}_a^{(1)}+b\hat{U}_a^{(0)},\tag{46}$$

where the weighting is chosen such that the better estimate $\hat{U}_a^{(0)}$ or $\hat{U}_a^{(1)}$ contributes stronger: with increasing a the estimate $\hat{U}_a^{(1)}$ becomes more reliable, as U_a is the normalizing constant of the bridging density p_a [Eq. (3)] and $p_a\overset{a\to 1}{\longrightarrow}p_1$, and conversely for decreasing a.

From the estimate of the overlap measure, we can estimate the rescaled mean-square error by

$$\hat{M}(a)=\frac{1}{ab}\left(\frac{1}{\hat{U}_a}-1\right)\tag{47}$$

for all $a\in(0,1)$, a result that is entirely based on first moments. The infimum of $\hat{M}(a)$ finally results in an estimate $\hat{\alpha}_o$ of the optimal choice α_o of $\frac{n_0}{N}$,

$$\hat{\alpha}_o:\Leftrightarrow\hat{M}(\hat{\alpha}_o)=\inf_a\hat{M}(a).\tag{48}$$

When searching for the infimum, we also take

$$\hat{M}(0)=\frac{1}{n_0}\sum_{k=1}^{n_0}e^{w_k^{(0)}-\widehat{\Delta f}}-\frac{1}{n_1}\sum_{l=1}^{n_1}e^{w_l^{(1)}-\widehat{\Delta f}},$$

$$\hat{M}(1)=\frac{1}{n_1}\sum_{l=1}^{n_1}e^{-w_l^{(1)}+\widehat{\Delta f}}-\frac{1}{n_0}\sum_{k=1}^{n_0}e^{-w_k^{(0)}+\widehat{\Delta f}},\tag{49}$$

into account which follow from a series expansion of Eq. (47) in a at $a=0$ and $a=1$, respectively.

VI. INCORPORATING COSTS

The costs of measuring a work value in forward direction may differ from the costs of measuring a work value in reverse direction. The influence of costs on the optimal ratio of sample sizes is investigated here.

Different costs can be due to a direction dependent effort of experimental or computational measurement of work (unfolding a RNA may be much easier than folding it). We assume the work values to be uncorrelated, which is essential for the validity of the theory presented in this paper. Thus, a source of nonequal costs, which arises especially when work values are obtained via computer simulations, is the difference in the strength of correlations of consecutive Monte Carlo steps in forward and reverse direction. To achieve uncorrelated draws, the "correlation lengths" or "correlation times" have to be determined within the simulation too. However, this is advisable in any case of two-sided estimation, independent of the sampling strategy.

Let c_0 and c_1 be the costs of drawing a single forward and reverse work value, respectively. Our goal is to minimize the mean-square error $X=\frac{1}{N}M$ while keeping the total costs $c=n_0c_0+n_1c_1$ constant. Keeping c constant results in

$$N(c,\alpha)=\frac{c}{\alpha c_0+\beta c_1},\tag{50}$$

which in turn yields

$$X(c,\alpha)=\frac{1}{N(c,\alpha)}M(\alpha).\tag{51}$$

If a local minimum exists, it results from $\frac{\partial}{\partial\alpha}X(c,\alpha)=0$, which leads to

$$\frac{\beta_o^2}{\alpha_o^2}=\frac{c_0\theta_1(\alpha_o)}{c_1\theta_0(\alpha_o)},\tag{52}$$

a result Bennett was already aware of [26]. However, based on second moments, it was not possible to estimate the optimal ratio r_o accurately and reliably. Hence, Bennett proposed to use a suboptimal *equal-time strategy* or *equal cost strategy*, which spends an equal amount of expenses to both directions, i.e., $n_0c_0=n_1c_1=\frac{c}{2}$ or

$$\frac{\beta_{ec}}{\alpha_{ec}} = \frac{c_0}{c_1}, \tag{53}$$

where $\alpha_{ec}=1-\beta_{ec}$ is the equal cost choice for $\alpha=\frac{n_0}{N}$. This choice is motivated by the following result:

$$X(c,\alpha) \geq \frac{1}{2}X(c,\alpha_{ec}) \quad \forall\ \alpha \in [0,1], \tag{54}$$

which states that the asymptotic mean-square error of the equal cost strategy is at most suboptimal by a factor of 2 [26]. Note, however, that the equal cost strategy can be far more suboptimal if the asymptotic limit of large sample sizes is not reached.

Since we can base the estimator for the optimal ratio r_o on first moments (see Sec. V), we propose a *dynamic strategy* that performs better than the equal cost strategy. The infimum of

$$\hat{X}(c,a) = \frac{ac_0 + bc_1}{c}\hat{M}(a) \tag{55}$$

results in the estimate $\hat{\alpha}_o$ of the optimal choice α_o of $\frac{n_0}{N}$,

$$\hat{\alpha}_o : \Leftrightarrow \hat{X}(c,\hat{\alpha}_o) = \inf_a \hat{X}(c,a). \tag{56}$$

We remark that opposed to $M(\alpha)$, $X(c,\alpha)$ is not necessarily convex. However, a global minimum clearly exists and can be estimated.

VII. A DYNAMIC SAMPLING STRATEGY

Suppose we want to estimate the free-energy difference with the acceptance ratio method but have a limit on the total amount of expenses c that can be spend for measurements of work. In order to maximize the efficiency, the measurements are to be performed such that $\frac{n_0}{N}$ finally equals the optimal fraction α_o of forward measurements.

The dynamic strategy is as follows:

(1) In absence of preknowledge on α_o, we start with Bennett's equal cost strategy (53) as an initial guess of α_o.

(2) After drawing a small number of work values, we make preliminary estimates of the free-energy difference, the mean-square error, and the optimal fraction α_o.

(3) Depending on whether the estimated rescaled mean-square error $\hat{M}(a)$ is convex, which is a necessary condition for convergence, our algorithm updates the estimate $\hat{\alpha}_o$ of α_o.

(4) Further work values are drawn such that $\frac{n_0}{N}$ dynamically follows $\hat{\alpha}_o$, while $\hat{\alpha}_o$ is updated repeatedly.

There is no need to update $\hat{\alpha}_o$ after each individual draw. Splitting the total costs into a sequence $0<c^{(1)}<\ldots<c^{(p)}=c$, not necessarily equidistant, we can predefine when and how often an update in $\hat{\alpha}_o$ is made. Namely, this is done whenever the actually spent costs reach the next value $c^{(\nu)}$ of the sequence.

The dynamic strategy can be cast into an algorithm.

Algorithm. Set the initial values $n_0^{(0)}=n_1^{(0)}=0$, $\hat{\alpha}_o^{(1)}=\alpha_{ec}$. In the νth step of the iteration $\nu=1,\ldots,p$ determine

$$n_0^{(\nu)} = \lfloor \hat{\alpha}_o^{(\nu)} N^{(\nu)} \rfloor,$$

$$n_1^{(\nu)} = \lfloor \hat{\beta}_o^{(\nu)} N^{(\nu)} \rfloor, \tag{57}$$

with

$$N^{(\nu)} = \frac{c^{(\nu)}}{\hat{\alpha}_o^{(\nu)} c_0 + \hat{\beta}_o^{(\nu)} c_1}, \tag{58}$$

where $\lfloor\ \rfloor$ means rounding to the next lower integer. Then, $\Delta n_0^{(\nu)}=n_0^{(\nu)}-n_0^{(\nu-1)}$ additional forward and $\Delta n_1^{(\nu)}=n_1^{(\nu)}-n_1^{(\nu-1)}$ additional reverse work values are drawn. Using the entire present samples, an estimate $\widehat{\Delta f}^{(\nu)}$ of Δf is calculated according to Eq. (13). With the free-energy estimate at hand, $\hat{M}^{(\nu)}(a)$ is calculated for all values of $a \in [0,1]$ via Eqs. (44)–(47) and (49) discretized, say in steps $\Delta a=0.01$. If $\hat{M}^{(\nu)}(a)$ is convex, we update the recent estimate $\hat{\alpha}_o^{(\nu)}$ of α_o to $\hat{\alpha}_o^{(\nu+1)}$ via Eqs. (55) and (56). Otherwise, if $\hat{M}^{(\nu)}(a)$ is not convex, the corresponding estimate of α_o is not yet reliable and we keep the recent value, $\hat{\alpha}_o^{(\nu+1)}=\hat{\alpha}_o^{(\nu)}$. Increasing ν by one, we iteratively continue with Eq. (57) until we finally obtain $\widehat{\Delta f}^{(p)}$, which is the optimal estimate of the free-energy difference after having spend all costs c.

Note that an update in $\hat{\alpha}_o^{(\nu)}$ may result in negative values of $\Delta n_0^{(\nu)}$ or $\Delta n_1^{(\nu)}$. Should $\Delta n_0^{(\nu)}$ happen to be negative, we set $n_0^{(\nu)}=n_0^{(\nu-1)}$ and

$$n_1^{(\nu)} = \left\lfloor \frac{c^{(\nu)} - c_0 n_0^{(\nu-1)}}{c_1} \right\rfloor. \tag{59}$$

We proceed analogously, if $\Delta n_1^{(\nu)}$ happens to be negative.

The optimal fraction α_o depends on the cost ratio c_1/c_0, i.e., the algorithm needs to know the costs c_0 and c_1. However, the costs are not always known in advance and may also vary over time. Think of a long-time experiment which is subject to currency changes, inflation, terms of trade, innovations, and so on. Of advantage is that the dynamic sampling strategy is capable of incorporating varying costs. In each iteration step of the algorithm, one just inserts the actual costs. If desired, the breakpoints $c^{(\nu)}$ may also be adapted to the actual costs. Should the costs initially be unknown (e.g., the "correlation length" of a Monte Carlo simulation needs to be determined within the simulation first) one may use any reasonable guess until the costs are known.

VIII. EXAMPLE

For illustration of results, we choose exponential work distributions

$$p_i(w) = \frac{1}{\mu_i} e^{-w/\mu_i}, \quad w \in \Omega = \mathbb{R}^+, \tag{60}$$

$\mu_i>0$, $i=0,1$. According to the fluctuation theorem (2), we have $\mu_1=\frac{\mu_0}{1+\mu_0}$ and $\Delta f=\ln(1+\mu_0)$.

Exponential work densities arise in a natural way in the context of a two-dimensional harmonic oscillator with Boltzmann distribution $\rho(x,y)=e^{-(1/2)\omega^2(x^2+y^2)}/Z$, where Z

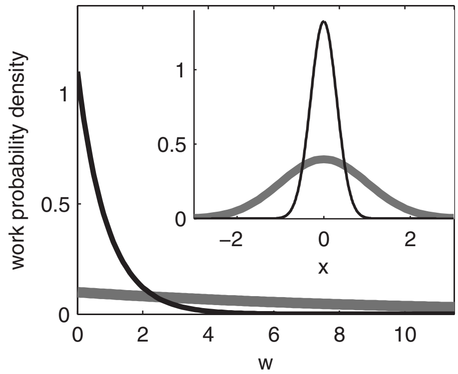

FIG. 2. The main figure displays the exponential work densities p_0 (thick line) and p_1 (thin line) for the choice of $\mu_0=10$ and, according to the fluctuation theorem, $\mu_1=10/11$. The inset displays the corresponding Boltzmann distributions $\rho_0(x,y)$ (thick) and $\rho_1(x,y)$ (thin) both for $y=0$. Here, ω_0 is set equal to 1 arbitrarily, hence, $\omega_1^2=(1+\mu_0)\omega_0^2=11$. The free-energy difference is $\Delta f=\ln(1+\mu_0)=\ln(\omega_1^2/\omega_0^2)\approx 2.38$.

$=2\pi/\omega^2$ is a normalizing constant (partition function) and $(x,y)\in\mathbb{R}^2$ [28]. Drawing a point (x,y) from the initial density $\rho=\rho_0$ defined by setting $\omega=\omega_0$, and switching the frequency to $\omega_1>\omega_0$ instantaneously amounts in the work $\frac{1}{2}(\omega_1^2-\omega_0^2)(x^2+y^2)$. The probability density of observing a specific work value w is given by the exponential density p_0 with $\mu_0=\frac{\omega_1^2-\omega_0^2}{\omega_0^2}$. Switching the frequency in the reverse direction $\omega_1\to\omega_0$, with the point (x,y) drawn from $\rho=\rho_1$ with $\omega=\omega_1$, the density of work (with interchanged sign) is given by p_1 with $\mu_1=\frac{\omega_1^2-\omega_0^2}{\omega_1^2}=\frac{\mu_0}{1+\mu_0}$. The free-energy difference of the states characterized by ρ_0 and ρ_1 is the log ratio of their normalizing constants $\Delta f=-\ln\frac{Z_1}{Z_0}=\ln(1+\mu_0)$. A plot of the work densities for $\mu_0=10$ is enclosed in Fig. 2.

Now, with regard to free-energy estimation, is it better to use one- or two-sided estimators? In other words, we want to know whether the global minimum of $M(\alpha)$ is on the boundaries $\{0,1\}$ of α or not. By the convexity of M, the answer is determined by the signs of the derivatives $M'(0)$ and $M'(1)$ at the boundaries. The asymptotic mean-square errors (18) and (19) of the one-sided estimators are calculated to be

$$M(1)=\mathrm{Var}_0(e^{-w+\Delta f})=\frac{\mu_0^2}{1+2\mu_0}, \tag{61}$$

for the forward direction and

$$M(0)=\mathrm{Var}_1(e^{w-\Delta f})=\frac{\mu_0^2}{1-\mu_0^2},\quad \mu_0<1, \tag{62}$$

for the reverse direction. For $\mu_0\geq 1$, the variance of the reverse estimator diverges. Note that $M(0)>M(1)$ holds for all $\mu_0>0$, i.e., forward estimation of Δf is always superior if compared to reverse estimation. Furthermore, a straightforward calculation gives

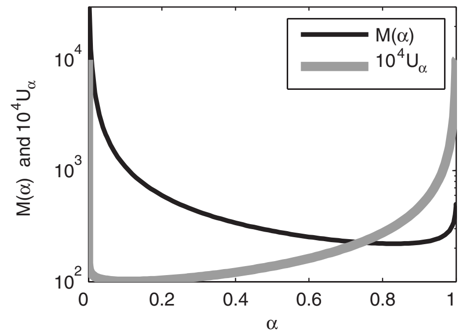

FIG. 3. The overlap function U_α and the rescaled asymptotic mean-square error M for $\mu_0=1000$. Note that $M(\alpha)$ diverges for $\alpha\to 0$.

$$M'(1)=\frac{\mu_0^3(\mu_0+\xi_-)(\mu_0-\xi_+)}{(1+2\mu_0)^2(1+3\mu_0)}, \tag{63}$$

where $\xi_\pm=\frac{1}{2}(\sqrt{17}\pm 3)$, and

$$M'(0)=-\frac{\mu_0^3(2+(1-2\mu_0)\mu_0)}{(1-\mu_0^2)^2(1-2\mu_0)},\quad \mu_0<\frac{1}{2}, \tag{64}$$

and $M'(0)=-\infty$ for $\mu_0\geq\frac{1}{2}$. Thus, for the range $\mu_0\in(0,\xi_+)$ we have $M'(0)<0$ as well as $M'(1)<0$ and therefore $\alpha_o=1$, i.e., the forward estimator is superior to any two-sided estimator (13) in this range. For $\mu_0\in(\xi_+,\infty)$, we have $M'(0)<0$ and $M'(1)>0$, specifying that $\alpha_o\in(0,1)$, i.e., two-sided estimation with an appropriate choice of α is optimal.

Numerical calculation of the function U_α and subsequent evaluation of $M(\alpha)$ allows to find the "exact" optimal fraction α_o. Examples for U_α and M are plotted in Fig. 3.

The behavior of α_o as a function of μ_0 is quite interesting (see Fig. 4). We can interpret this behavior in terms of the Boltzmann distributions as follows. Without loss of generality, assume $\omega_0=1$ is fixed. Increasing μ_0 then means increasing ω_1. The density ρ_1 is fully nested in ρ_0 (cf. the inset of Fig. 2) (remember that $\omega_1>\omega_0$) and converges to a delta peak at the origin with increasing ω_1. This means that by

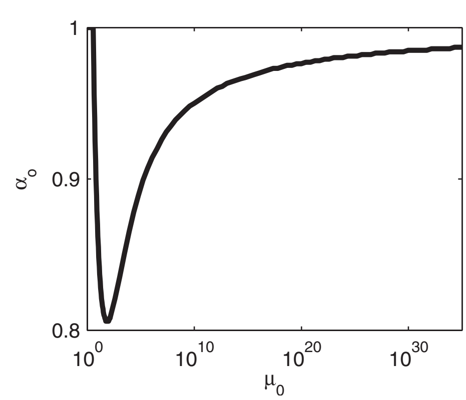

FIG. 4. The optimal fraction $\alpha_o=\frac{n_0}{N}$ of forward work values for the two-sided estimation in dependence of the average forward work μ_0. For $\mu_0\leq\xi_+\approx 3.56$, the one-sided forward estimator is optimal, i.e., $\alpha_o=1$.

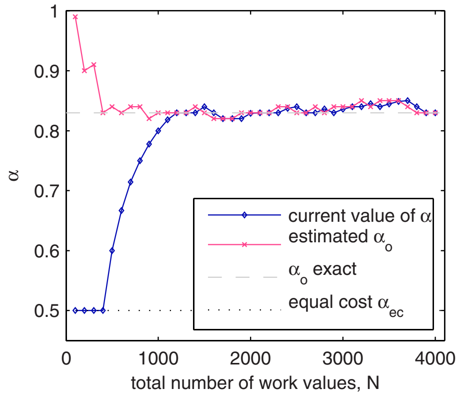

FIG. 5. (Color online) Example of a single run using the dynamic strategy: the optimal fraction α_o of forward measurements for the two-sided free-energy estimation is estimated at predetermined values of total sample sizes $N=n_0+n_1$ of forward and reverse work values. Subsequently, taking into account the current actual fraction $\alpha=\frac{n_0}{N}$, additional work values are drawn such that we come closer to the estimated $\hat{\alpha}_o$.

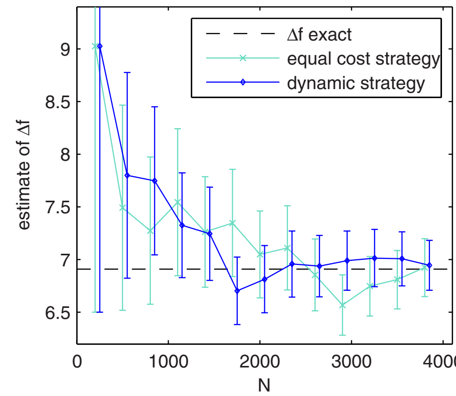

FIG. 7. (Color online) Comparison of a single run of free-energy estimation using the equal cost strategy versus a single run using the dynamic strategy. The error bars are the square roots of the estimated mean-square error X.

sampling from ρ_0 we can obtain information about the full density ρ_1 quite easily, whereas sampling from ρ_1 provides only poor information about ρ_0. This explains why $\alpha_o=1$ holds for small values of μ_0. However, with increasing ω_1 the density ρ_1 becomes so narrow that it becomes difficult to obtain draws from ρ_0 that fall into the main part of ρ_1. Therefore, it is better to add some information from ρ_1, hence, α_o decreases. Increasing ω_1 further, the relative number of draws needed from ρ_1 will decrease, as the density converges toward the delta distribution. Finally, it will become sufficient to make only *one* draw from ρ_1 in order to obtain the full information available. Therefore, α_o converges toward 1 in the limit $\mu_0 \to \infty$.

In the following, the dynamic strategy proposed in Sec. VII is applied. We choose $\mu_0=1000$ and $c_0=c_1$. The equal cost strategy draws according to $\alpha_{ec}=0.5$, which is used as initial value in the dynamic strategy. The results of a single run are presented in Figs. 5–7. Starting with $N=100$, the estimate of α_o is updated in steps of $\Delta N=100$. The actual forward fractions α together with the estimated values of the optimal fraction α_o are shown in Fig. 5. The first three estimates of α_o are rejected because the estimated function $\hat{M}(\alpha)$ is not yet convex. Therefore, α remains unchanged at the beginning. Afterward, α follows the estimates of α_o and starts to fluctuate about the exact value of α_o. Some estimates of the function M corresponding to this run are depicted in Fig. 6. For these estimates, α is discretized in steps $\Delta\alpha=0.01$. Remarkably, the estimates of α_o that result from these curves are quite accurate even for relatively small N. Finally, Fig. 7 shows the free-energy estimates of the run (not for all values of N) compared with those of a single run where the equal cost strategy is used. We find some increase in accuracy when using the dynamic strategy.

In combination with a good *a priori* choice of the initial value of α, the use of the dynamic strategy enables a superior convergence and precision of free-energy estimation (see Figs. 8 and 9). Due to insight into some particular system under consideration, it is not unusual that one has *a priori*

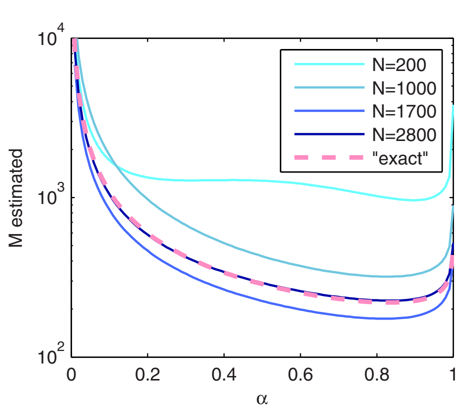

FIG. 6. (Color online) Displayed are estimated mean-square errors $\hat{M}$ in dependence of α for different sample sizes. The global minimum of the estimated function $\hat{M}$ determines the estimate of the optimal fraction α_o of forward work measurements.

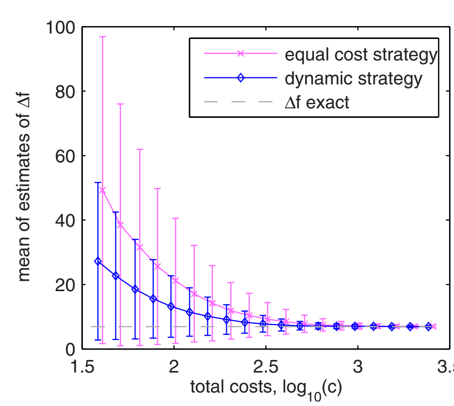

FIG. 8. (Color online) Averaged estimates from 10 000 independent runs with dynamic strategy versus 10 000 runs with equal cost strategy in dependence of the total cost $c=n_0c_0+n_1c_1$ spend. The cost ratio is $c_1/c_0=0.01$, $c_0+c_1=2$, and $\mu_0=1000$. The error bars represent one standard deviation. Here, the initial value of α in the dynamic strategy is 0.5, while the equal cost strategy draws with $\alpha_{ec}\approx 0.01$. We note that $\alpha_o \approx 0.08$.

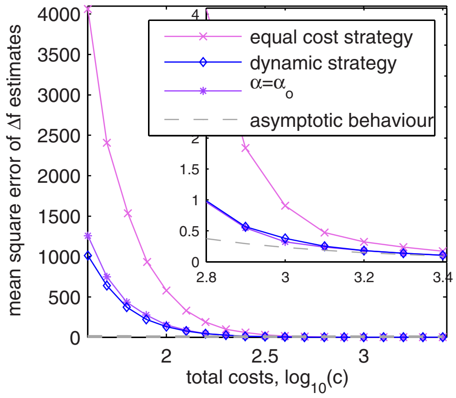

FIG. 9. (Color online) Displayed are mean-square errors of free-energy estimates using the same data as in Fig. 8. In addition, the mean-square errors of estimates with constant $\alpha=\alpha_o$ are included, as well as the asymptotic behavior (51). The inset shows that the mean-square error of the dynamic strategy approaches the asymptotic optimum, whereas the equal cost strategy is suboptimal. Note that for small sample sizes, the asymptotic behavior does not represent the actual mean-square error.

knowledge which results in a better guess for the initial choice of α in the dynamic strategy than starting with $\alpha=\alpha_{ec}$. For instance, a good initial choice is known when estimating the chemical potential via Widom's particle insertion and deletion [31]. Namely, it is *a priori* clear that inserting particles yields much more information than deleting particles since the phase space which is accessible to particles in the "deletion system" is effectively contained in the phase space accessible to the particles in the "insertion system" (cf., e.g., [19]). A good *a priori* initial choice for α may be $\alpha=0.9$ with which the dynamic strategy outperforms any other strategy that the authors are aware of.

Once reaching the limit of large sample sizes, the dynamic strategy is insensitive to the initial choice of α since the strategy is robust and finds the optimal fraction α_o of forward measurements itself.

IX. CONCLUSION

Two-sided free-energy estimation, i.e., the acceptance ratio method [26], employs samples of n_0 forward and n_1 reverse work measurements in the determination of free-energy differences in a statistically optimal manner. However, its statistical properties depend strongly on the ratio $\frac{n_1}{n_0}$ of work values used. As a central result, we have proven the convexity of the asymptotic mean-square error of two-sided free-energy estimation as a function of the fraction $\alpha=\frac{n_0}{N}$ of forward work values used. From here follows immediately the existence and uniqueness of the optimal fraction α_o, which minimizes the asymptotic mean-square error. This is of particular interest if we can control the value of α, i.e., can make additional measurements of work in either direction. Drawing such that we finally reach $\frac{n_0}{N}=\alpha_o$, the efficiency of two-sided estimation can be enhanced considerably. Consequently, we have developed a dynamic sampling strategy which iteratively estimates α_o and makes additional draws or measurements of work. Thereby, the convexity of the mean-square error enters as a key criterion for the reliability of the estimates. For a simple example, which allows to compare with analytic calculations, the dynamic strategy has shown to work perfectly.

In the asymptotic limit of large sample sizes, the dynamic strategy is optimal and outperforms any other strategy. Nevertheless, in this limit it has to compete with the near optimal equal cost strategy of Bennett, which also performs very good. It is worth mentioning that even if the latter comes close to the performance of ours, it is worthwhile the effort of using the dynamic strategy since the underlying algorithm can be easily implemented and does cost quite anything if compared to the effort required for drawing additional work values.

Most important for experimental and numerical estimation of free-energy differences is the range of small and moderate sample sizes. For this relevant range, it is found that the dynamic strategy performs very good too. It converges significantly better than the equal cost strategy. In particular, for small and moderate sample sizes it can improve the accuracy of free-energy estimates by half an order of magnitude.

We close our considerations by mentioning that the two-sided estimator is typically far superior with respect to one-sided estimators: assume the support of p_0 and p_1 is symmetric about Δf [32]; then, if the densities are symmetric to each other, $p_0(\Delta f+w)=p_1(\Delta f-w)$, the optimal fraction of forward draws is $\frac{n_0}{N}=\frac{1}{2}$ by symmetry. Therefore, if the symmetry is violated not too strongly, the optimum will remain near 0.5. Continuous deformations of the densities change the optimal fraction α_o continuously. Thus, α_o does not reach 0 and 1, respectively, for some certain strength of asymmetry. It is exceptionally hard to violate the symmetry such that α_o hits the boundary 0 or 1. In consequence, in almost all situations, the two-sided estimator is superior.

ACKNOWLEDGMENTS

We thank Andreas Engel for a critical reading of the manuscript.

[1] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).
[2] G. E. Crooks, Phys. Rev. E **60**, 2721 (1999).
[3] J. G. Kirkwood, J. Chem. Phys. **3**, 300 (1935).
[4] A. Gelman and X.-L. Meng, Stat. Sci. **13**, 163 (1998).
[5] R. W. Zwanzig, J. Chem. Phys. **22**, 1420 (1954).
[6] G. M. Torrie and J. P. Valleau, J. Comput. Phys. **23**, 187 (1977).
[7] M.-H. Chen and Q.-M. Shao, Ann. Stat. **25**, 1563 (1997).

[8] H. Oberhofer and C. Dellago, Comput. Phys. Commun. **179**, 41 (2008).
[9] M. Watanabe and W. P. Reinhardt, Phys. Rev. Lett. **65**, 3301 (1990).
[10] S. X. Sun, J. Chem. Phys. **118**, 5769 (2003).
[11] F. M. Ytreberg and D. M. Zuckerman, J. Chem. Phys. **120**, 10876 (2004).
[12] C. Jarzynski, Phys. Rev. E **73**, 046105 (2006).
[13] A. Engel, Phys. Rev. E **80**, 021120 (2009).
[14] H. Then and A. Engel, Phys. Rev. E **77**, 041105 (2008).
[15] X.-L. Meng and S. Schilling, J. Comput. Graph. Stat. **11**, 552 (2002).
[16] C. Jarzynski, Phys. Rev. E **65**, 046122 (2002).
[17] H. Oberhofer, C. Dellago, and S. Boresch, Phys. Rev. E **75**, 061106 (2007).
[18] S. Vaikuntanathan and C. Jarzynski, Phys. Rev. Lett. **100**, 190601 (2008).
[19] A. M. Hahn and H. Then, Phys. Rev. E **79**, 011113 (2009).
[20] X.-L. Meng and W. H. Wong, Stat. Sin. **6**, 831 (1996).
[21] A. Kong, P. McCullagh, X.-L. Meng, D. Nicolae, and Z. Tan, J. R. Stat. Soc. Ser. B (Stat. Methodol.) **65**, 585 (2003).
[22] M. R. Shirts and J. D. Chodera, J. Chem. Phys. **129**, 124105 (2008).
[23] D. M. Ceperley, Rev. Mod. Phys. **67**, 279 (1995).
[24] D. Frenkel and B. Smit, *Understanding Molecular Simulation*, 2nd ed. (Academic Press, London, 2002).
[25] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco, Jr., and C. Bustamante, Nature (London) **437**, 231 (2005).
[26] C. H. Bennett, J. Comput. Phys. **22**, 245 (1976).
[27] G. E. Crooks, Phys. Rev. E **61**, 2361 (2000).
[28] M. R. Shirts and V. S. Pande, J. Chem. Phys. **122**, 144107 (2005).
[29] M. R. Shirts, E. Bair, G. Hooker, and V. S. Pande, Phys. Rev. Lett. **91**, 140601 (2003).
[30] J. Gore, F. Ritort, and C. Bustamante, Proc. Natl. Acad. Sci. U.S.A. **100**, 12564 (2003).
[31] B. Widom, J. Chem. Phys. **39**, 2808 (1963).
[32] Which is not the case for the densities studied in Sec. VIII.