On the Accurate Estimation of Free Energies Using the Jarzynski Equality

Mehrnoosh Arrar,[†[a]] Fernando Martín Boubeta,[†[a]] Maria Eugenia Szretter,[b] Mariela Sued,[c] Leonardo Boechi,*[c] and Daniela Rodriguez*[c]

The Jarzynski equality is one of the most widely celebrated and scrutinized nonequilibrium work theorems, relating free energy to the external work performed in nonequilibrium transitions. In practice, the required ensemble average of the Boltzmann weights of infinite nonequilibrium transitions is estimated as a finite sample average, resulting in the so-called Jarzynski estimator, $\Delta\hat{F}_J$. Alternatively, the second-order approximation of the Jarzynski equality, though seldom invoked, is exact for Gaussian distributions and gives rise to the Fluctuation-Dissipation estimator $\Delta\hat{F}_{FD}$. Here we derive the parametric maximum-likelihood estimator (MLE) of the free energy $\Delta\hat{F}_{ML}$ considering unidirectional work distributions belonging to Gaussian or Gamma families, and compare this estimator to $\Delta\hat{F}_J$. We further consider bidirectional work distributions belonging to the same families, and compare the corresponding bidirectional $\Delta\hat{F}_{ML*}$ to the Bennett acceptance ratio ($\Delta\hat{F}_{BAR}$) estimator. We show that, for Gaussian unidirectional work distributions, $\Delta\hat{F}_{FD}$ is in fact the parametric MLE of the free energy, and as such, the most efficient estimator for this statistical family. We observe that $\Delta\hat{F}_{ML}$ and $\Delta\hat{F}_{ML*}$ perform better than $\Delta\hat{F}_J$ and $\Delta\hat{F}_{BAR}$, for unidirectional and bidirectional distributions, respectively. These results illustrate that the characterization of the underlying work distribution permits an optimal use of the Jarzynski equality. © 2018 Wiley Periodicals, Inc.

DOI:10.1002/jcc.25754

Introduction

The nonequilibrium work theorem popularly known as the "Jarzynski equality" was first communicated to the scientific community in 1997 and represents a significant turning point for both experimental and computational biophysics.[1]

The Jarzynski equality (eq. 1) relates the equilibrium free energy change ΔF of a given process to the ensemble average of the Boltzmann weights of the external work W performed in infinite repetitions of the process of interest carried out far from equilibrium:

$$\Delta F = -\ln\langle e^{-W}\rangle, \tag{1}$$

where both W and F are in units of k_BT, with k_B being the Boltzmann constant and T being the temperature. In the limit of instantaneous processes, eq. 1 reduces to the Zwanzig relationship.[2] The Jarzynski equality is particularly attractive because of its generality; that is, the equality holds regardless of the rate at which the nonequilibrium processes are carried out, and the nature of the process itself. Interest in the Jarzynski equality spread rapidly among the scientific community, not only among theoretical physicists but also among experimental biophysicists, who were eager to test and exploit its full potential, in particular in single-molecule pulling experiments.[3–11]

In practice, the ensemble average indicated by the brackets $\langle\cdot\rangle$ in eq. 1 is estimated using the average of a finite sample of $W_1, \ldots, W_N$ values, resulting in the so-called Jarzynski free energy estimator $\Delta\hat{F}_J$, where the hat notation indicates an estimator:

$$\Delta\hat{F}_J = -\ln\left[\frac{1}{N}\sum_{i=1}^{N} e^{-W_i}\right]. \tag{2}$$

In many cases, the nonequilibrium work distributions are reported to be Gaussian, or mixtures of Gaussians.[12,13] More specifically, nonequilibirum work distributions generated from steered Molecular Dynamics (sMD) simulations in which a sufficiently high force constant is used, and the pulling velocity is slow enough, compared to the dynamic relaxation of the system, such that the perturbation is considered to be in the "near-equilibrium" or "low-dissipation" regime, the resulting work distribution is expected to be Gaussian.[12,14,15] For a

[a] M. Arrar, F. Martín Boubeta
Instituto de Química-Física de los Materiales, Medio Ambiente y Energía, CONICET-Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina, Departamento de Química Inorgánica, Analítica y Química Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina
[b] M. Eugenia Szretter
Departamento de Matemática, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina, Instituto de Cálculo, CONICET-Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina
[c] M. Sued, L. Boechi, D. Rodriguez
Instituto de Cálculo, CONICET-Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina
E-mail: lboechi@ic.fcen.uba.ar or drodrig@dm.uba.ar
[†] These authors contributed equally to this work
Contract Grant sponsor: CONICET; Contract Grant sponsor: University of Buenos Aires

Gaussian distribution with mean μ and variance σ^2, the right-hand side of the Jarzynski equality (eq. 1) is equal to its second-order Taylor expansion ($\Delta F = \mu - \frac{\sigma^2}{2}$), an expression that has been independently derived from the Fluctuation-Dissipation (FD) theorem and gives rise to the estimator $\Delta\hat{F}_{FD}$[16–18]:

$$\Delta\hat{F}_{FD} = \overline{W}_N - \frac{\hat{\sigma}_W^2}{2}, \tag{3}$$

where $\overline{W}_N$ and $\hat{\sigma}_W^2$ are the work sample average and variance, respectively:

$$\overline{W}_N = \frac{1}{N}\sum_{i=1}^{N} W_i \quad \text{and} \quad \hat{\sigma}_W^2 = \frac{1}{N}\sum_{i=1}^{N}\left(W_i - \overline{W}\right)^2. \tag{4}$$

The exponential nature of the Boltzmann weights makes the average in $\Delta\hat{F}_J$ notoriously difficult to converge because it is dominated by typically poorly sampled negative work values, a problem that is exacerbated by large dispersion in the work values.[19–23] Accordingly, many have developed methodologies that limit the dispersion of the nonequilibrium work values in the first place.[8,24–29] Others have proposed corrections or block-averaging, combinatorial averaging, and extrapolation analysis protocols that have proved useful in improving the exponential average in the finite sampling limit.[30–34]

When possible, independent pulling experiments performed in both forward and reverse directions is not only a convenient way to check the convergence of a free energy estimate, but also a source of additional information that can be leveraged to improve the free energy estimate. Specifically, we refer to the forward work distribution as that resulting from multiple independent pulling experiments starting from equilibrium structures of state A with the reaction coordinate restrained to its initial value, and ending with the value of the reaction coordinate corresponding to state B. The reverse work ensemble is defined as just the opposite; beginning with equilibrium structures of state B with the reaction coordinate restrained to an initial value, and ending with the value of the reaction coordinate corresponding to state A. We use the term bidirectional pulling experiment to refer to this pair of forward and reverse pulling experiments. The distributions of the forward and reverse work ensembles are directly related to the free energy, according to the Crooks Fluctuation Theorem (CFT)[35]:

$$f_F(w) = f_R(-w)\exp(w - \Delta F), \tag{5}$$

where $f_F(\cdot)$ and $f_R(\cdot)$ are the probability density functions (pdfs) of W in the forward and reverse directions, respectively, and ΔF is the free energy change in the forward direction. As the CFT relates the forward work distribution to that of the sign-changed works from the reverse process, we will let $f_{-R}(w) = f_R(-w)$ to simplify the notation. According to the CFT, the pdfs f_F and f_{-R} intersect at $w = \Delta F$. Using logistic regression and the CFT, the Bennett Acceptance Ratio ($\Delta\hat{F}_{\mathrm{BAR}}$) has been shown to be a nonparametric maximum-likelihood estimator (MLE) of the free energy difference, leveraging forward and reverse work ensembles.[36,37] As the directionality is arbitrarily assigned and is only relevant when both directions are considered, from here on out, when referring to a unidirectional work distribution we will simply omit the subscript *F* or *R*.

Often it is of interest not only to calculate a two-state free energy difference, but a free energy profile, or change along a reaction coordinate, in which case the Jarzynski equality can be applied along the reaction coordinate. Specifically for the reconstruction of free energy profiles, although, Hummer and Szabo have proposed an efficient free energy estimator akin to the weighted histogram approach.[38] In the limit of two states, the Hummer and Szabo estimator reduces to the $\Delta\hat{F}_{\mathrm{BAR}}$ estimator. Minh and colleagues have further generalized the Hummer and Szabo estimator to include bidirectional pulling experiments,[39,40] and recently Nicolini et al. have proposed a similar estimator for free energy profiles using bidirectional work trajectories that was shown to be advantageous particularly when the dispersion in the work distribution is large.[41] These estimators have been compared computationally using either toy examples with analytical solutions or simple test systems.[42–44] A general conclusion that can be drawn from these studies, however, is that the advantage of $\Delta\hat{F}_{\mathrm{BAR}}$ or Hummer and Szabo estimators in practice depends on the specific system and transformation.

In this work, we will focus on the estimation of two-state free energy differences, although the results should be generalizable to the reconstruction of free energy profiles. Rather than considering toy problems, we consider representative pdfs of nonequilibrium work values from Gaussian and Gamma families. For each family we derive the parametric MLE of the free energy using either unidirectional ($\Delta\hat{F}_{ML}$) or bidirectional ($\Delta\hat{F}_{ML*}$) work distributions, and numerically compare these estimators to $\Delta\hat{F}_J$ and $\Delta\hat{F}_{\mathrm{BAR}}$ for unidirectional and bidirectional work distributions, respectively.

Briefly, we will describe the motivation for choosing these two statistical families. First, the Gaussian distribution has been analytically derived as the expected distribution for near-equilibrium pulling conditions; namely, when the dynamics of the reaction coordinate can be approximated by a particle diffusing along a moving harmonic potential.[15,45] Park et al. have illustrated that, in the specific context of SMD simulations, a sufficiently stiff spring along with sufficiently fast relaxations of the system result in such Gaussian distributions.[12] Indeed, many pulling experiments, both computational and experimental, reportedly result in Gaussian work distributions.[46–50] Nevertheless, there are also many real-world examples of non-Gaussian work distributions.[51,52] In particular, Kofke and coworkers have made substantial advances in characterizing perturbation energy (or instantaneous-switching work) distributions from Free Energy Perturbation (FEP) calculations, and have highlighted the marked asymmetry between insertion/desertion FEP calculations, making these pdfs strictly non-Gaussian.[53–56] In fact, the typical histograms of perturbation energies from these calculations are well-described by Gamma distributions.[53,54] Moreover, for several analytically solvable test cases, such as certain cases of the multiharmonic model,[56]

nested harmonic potentials,[42] and the adiabatic compression/expansion of a dilute gas,[57] the derived work distributions belong to the Gamma family.

Our intention is to clarify that the Jarzynski equality needs not be exclusively linked to the $\Delta\hat{F}_J$ estimator, which, though frequently invoked, can perform quite poorly for certain work distributions. We conclude that the MLE for a particular parametric family of the underlying work distribution always leads to an optimal use of the Jarzynski equality. Although we show that $\Delta\hat{F}_{\mathrm{BAR}}$ is a powerful nonparametric estimator, we note that, when possible, correctly identifying the statistical family and using the corresponding $\Delta\hat{F}_{ML*}$ is more advantageous.

Methodology

Maximum-likelihood approach

Briefly, we will summarize the method of deriving MLEs as used here, more details can be found elsewhere.[58]

Unidirectional case. Based on a sample $W_1, \ldots, W_N$ from a unidirectional work distribution characterized by the pdf $f(\cdot, \theta_1, \theta_2)$, where θ_1 and θ_2 are the parameters indexing the statistical family (e.g., μ and σ^2 for the Gaussian case). The MLEs ($\hat{\theta}_1$ and $\hat{\theta}_2$) of these parameters maximize the likelihood function, which is

$$L(\theta_1,\theta_2)=\prod_{i=1}^{N} f(W_i;\theta_1,\theta_2). \tag{6}$$

The maximizers of $L(\theta_1, \theta_2)$ are critical points, and are also critical points of its natural logarithm (the log-likelihood function), which turns out to be easier to work with. We take the partial derivative of $\ln(L(\theta_1, \theta_2))$ with respect to θ_1 and θ_2, such that the maximizers of the log-likelihood $\hat{\theta}_1$ and $\hat{\theta}_2$ satisfy:

$$0=\sum_{i=1}^{N}\frac{\partial}{\partial\theta_1}\ln f\left(W_i;\hat{\theta}_1,\hat{\theta}_2\right); \quad \text{and} \quad 0=\sum_{i=1}^{N}\frac{\partial}{\partial\theta_2}\ln f\left(W_i;\hat{\theta}_1,\hat{\theta}_2\right) \tag{7}$$

It is a simple exercise to show that the maximizers of the log-likelihood function in the unidirectional Gaussian case are $\overline{W}_N$ for μ and $\hat{\sigma}^2_W$ for σ^2 (eq. 4).

Bidirectional case. To leverage the data from both work distributions in the estimation of the necessary parameters, we first recognize that we can describe the reverse distribution in terms of the parameters of the forward distribution, θ_1 and θ_2, and express the joint-likelihood function:

$$L(\theta_1,\theta_2)=\prod_{i=1}^{n_F} f_F(W_i,\theta_1,\theta_2)\prod_{j=1}^{n_R} f_R\left(-Z_j,\theta_1,\theta_2\right), \tag{8}$$

where $-Z_1,\ldots,-Z_{n_R}$ are the work values with signs changed from the reverse distribution. The same steps as described in unidirectional case can be taken to find the maximizers of this joint-likelihood function (eq. 12).

Numerical simulations

For a given statistical family, fixing the corresponding parameters, and a sample size N, the following steps were carried out:

1. A sample of works $W_1,\ldots,W_{n_F}$ was generated from $f_F(w)$, and, for bidirectional cases, a corresponding sample $Z_1,\ldots,Z_{n_R}$ from $f_R(w)$ was generated. In unidirectional cases, $n_F = N$, whereas in bidirectional cases $n_F = n_R = N/2$.
2. The free energy ΔF was estimated based on the generated sample, according to $\Delta\hat{F}_J$, $\Delta\hat{F}_{ML}$, and, for bidirectional cases, $\Delta\hat{F}_{\mathrm{BAR}}$ and $\Delta\hat{F}_{ML*}$.
3. For each of the estimators the mean squared error (MSE) was computed from 10,000 repetitions of steps 1 and 2, giving rise to 10,000 estimated free energy values $\Delta\hat{F}_{\ominus}^{(1)}\ldots\Delta\hat{F}_{\ominus}^{(10{,}000)}$, each based on a different sample of work distribution of size N:

$$MSE\left(\Delta\hat{F}_{\ominus}\right)=\frac{1}{10{,}000}\sum_{j=1}^{j=10{,}000}\left(\Delta\hat{F}_{\ominus}^{(j)}-\Delta F\right)^2, \tag{9}$$

where $\ominus$ corresponds to ML, J, ML*, or BAR.

For each of the estimators we also constructed a density estimator based on the same 10,000 estimated free energy values $\Delta\hat{F}_{\Theta}^{(1)},\ldots,\Delta\hat{F}_{\Theta}^{(10{,}000)}$, using a Gaussian kernel and Silverman's rule[59] to determine the bandwidth.

For $\Delta\hat{F}_{\mathrm{BAR}}$ and $\Delta\hat{F}_{ML*}$ the reverse distributions $f_{-R}(w)$ were derived in terms of the parameters of the forward distribution $f_F(w)$, according to the CFT (eq. 5), as described in the Results section (details provided in Supporting Information). We considered sample sizes N of 50, 100, 500, 1000, 5000, 10,000, and 500,000. All simulations were run using python, the scipy.stats module was used to describe and calculate parameters for both statistical families,[60] and the pymbar script was used to calculate the $\Delta\hat{F}_{\mathrm{BAR}}$ estimates.[61]

Results

It will be worthwhile to begin by recalling that the ensemble average is an expected value and can be calculated as:

$$\left\langle e^{-W}\right\rangle=\int e^{-w}f(w)dw, \tag{10}$$

where $f(w)$ is, again, the pdf of the work distribution.

If a parametric family is postulated for $f(w)$, the ensemble average can be characterized in terms of the parameters indexing the family, which can in turn be estimated invoking the maximum-likelihood principle (see Methods).

In the case of the Gaussian distribution the ensemble average can be expressed *exactly* as a function of the work distribution's mean μ and variance σ^2. By substituting the pdf (column 2 of Table 1) into eq. 10, we arrive at the following identity:

Table 1. Summary of statistical families and exact expressions for ΔF.

Distribution	Probability density function	ΔF
Normal $N(\mu, \sigma^2)$	$\frac{1}{\sqrt{2\pi}\sigma} e^{-(w-\mu)^2/2\sigma^2}$	$\mu - \frac{\sigma^2}{2}$
Gamma $\Gamma(\alpha, \lambda)$	$\frac{\lambda^\alpha}{\Gamma(\alpha)} w^{\alpha-1} e^{-\lambda w} \quad w > 0$[a]	$\alpha \ln\left(\frac{\lambda+1}{\lambda}\right)$

[a] $\Gamma(\cdot)$ is the Gamma function.

$$\left\langle e^{-W} \right\rangle = \exp\left(-\mu + \frac{\sigma^2}{2}\right). \quad (11)$$

Substituting eq. 11 into the Jarzynski equality (eq. 1), we obtain an exact expression for ΔF in terms of the parameters μ and σ (Table 1). The MLEs of μ and σ^2, $\hat{\mu}$ and $\hat{\sigma}^2$, respectively, are precisely $\overline{W}_N$ and $\hat{\sigma}^2_W$, defined in eq. 3. By replacing these MLEs in the expression for ΔF in Table 1, we obtain the parametric MLE of ΔF, which coincides with the familiar estimator $\Delta\hat{F}_{FD}$ (eq. 3) resulting from the FD nonequilibrium work theorem.[18] In the limit of large N, the two estimators $\Delta\hat{F}_{FD}$ and $\Delta\hat{F}_J$ of course converge to the true value of ΔF. The $\Delta\hat{F}_J$ estimator requires no knowledge of the underlying work distribution; however, if such information is available, the MLE ($\Delta\hat{F}_{FD}$ for a Gaussian distribution) should be used as it is the most efficient possible estimator.[58,62,63]* In the sections that follow, we first compare $\Delta\hat{F}_{ML}$ and $\Delta\hat{F}_J$ for unidirectional Gaussian and Gamma families; we then derive the bidirectional MLEs $\Delta\hat{F}_{ML*}$ for both families, and compare $\Delta\hat{F}_{ML*}$ to $\Delta\hat{F}_{\mathrm{BAR}}$.

Numerical simulations

Here, we consider numerically generated work samples from Gaussian and Gamma distributions of varying spread.

As mentioned earlier, the right-hand side of the Jarzynski equality (eq. 1) can be solved exactly in terms of the parameters indexing each statistical family, by substituting the pdf (second column of Table 1) into eq. 10. Replacing each of these parameters with its MLE, calculated from either unidirectional or bidirectional work distributions, yields the corresponding parametric MLE of ΔF. The parameters and exact expressions for ΔF are summarized in Table 1, columns 1 and 3, respectively, for Gaussian and Gamma distributions.

Unidirectional work distributions. For each family, we considered work distributions of varying spread by fixing the corresponding parameters, and for each scenario generated 10,000 random samples of increasing size N, ranging from N = 50 to N = 500,000, which correspond to independent unidirectional nonequilibrium work values. For each of these samples, we estimated the free energy using $\Delta\hat{F}_J$ and the corresponding MLE, $\Delta\hat{F}_{ML}$, which, for a Gaussian distribution is $\Delta\hat{F}_{FD}$. From these 10,000 samples, we calculated the MSE (eq. 9 in Methods) and assessed its behavior with increasing sample sizes.

We considered various Gaussian distributions that correspond to a wide range of associated ΔF values (−70, 0, or 70 k_BT) (Supporting Information Fig. S1).† The convergence of the free energy estimate, however, depends exclusively on the standard deviation σ of the work distribution, and not on its mean μ (Fig. 1). Even sample sizes as large as N = 500,000 are not large enough to converge the estimate obtained with $\Delta\hat{F}_J$, for work distributions with σ above 4 k_BT, a magnitude that is easily observed in driven nonequilibrium processes, particularly those involving biomolecular systems with many degrees of freedom.[48,64,65]

As anticipated, the MLE for Gaussian work distributions $\Delta\hat{F}_{FD}$ performs well, regardless of the spread of the distribution, and even for sample sizes as small as N = 50.

To further understand how the performance of the estimators are affected by the nature of the underlying work distribution, we also considered distributions from the Gamma family (Supporting Information Fig. S2), which has been identified to be relevant in certain driven nonequilibrium processes.[42,52,57] Depending on the shape and rate parameters (α and λ, respectively), a Gamma distribution can resemble a Gaussian distribution, however the Gamma distribution is asymmetric and restricted to strictly positive values of W. In this case, the convergence of $\Delta\hat{F}_J$ depends on both α and λ of the distribution and performance worsens particularly as the rate parameter λ decreases, and as the shape parameter α increases (toward the left and toward the bottom of Fig. 2). We note that even without negative work values, in this regime, in which the work distributions are extremely disperse, we again see the poor performance of $\Delta\hat{F}_J$.

Bidirectional work distributions. To consider bidirectional work distributions, we first used the CFT (eq. 5) to describe f_{-R} in terms of the parameters of f_F and then generated samples of size n_F and n_R from both forward and reverse distributions, with $n_F = n_R = N/2$. For a Gaussian family, a forward distribution $f_F(w)$ with mean μ and variance σ^2 will have a corresponding reverse distribution $f_{-R}(w)$ that is also Gaussian with mean $\mu_R = \mu - \sigma^2$ and variance $\sigma^2_R = \sigma^2$. In this way, the parameters μ and σ^2 sufficiently describe the set of forward and reverse distributions. Similarly, a Gamma distribution $f_F(w)$ with shape and rate parameters α and λ, respectively, has a corresponding $f_{-R}(w)$ that is also Gamma with shape and rate parameters $\alpha_R = \alpha$ and $\lambda_R = \lambda + 1$ (detailed derivation in Supporting Information).

In the case in which $n_F = n_R$ the estimator for μ that maximizes the joint-likelihood expression for the Gaussian case (see Methods) is:

$$\hat{\mu}_{ML*} = \frac{\overline{W}_{n_F} + \overline{Z}_{n_R}}{2} + \frac{\hat{\sigma}^2_{ML*}}{2}, \quad (12)$$

where $\overline{W}_{n_F}$ and $\overline{Z}_{n_R}$ are the sample means of the forward and reverse work distributions, as defined earlier in eq. 4. Upon substitution into the exact expression for ΔF from column 3 of Table 1, we observe that $\hat{\sigma}^2_{ML*}$ cancels out and the bidirectional MLE $\Delta\hat{F}_{ML*}$ for Gaussian distributions reduces nicely to

*For a Gaussian distribution, $\Delta\hat{F}_{ML} = \Delta\hat{F}_{FD}$.

†1 k_{BT} at 300 K is approximately 0.56 kcal/mol.

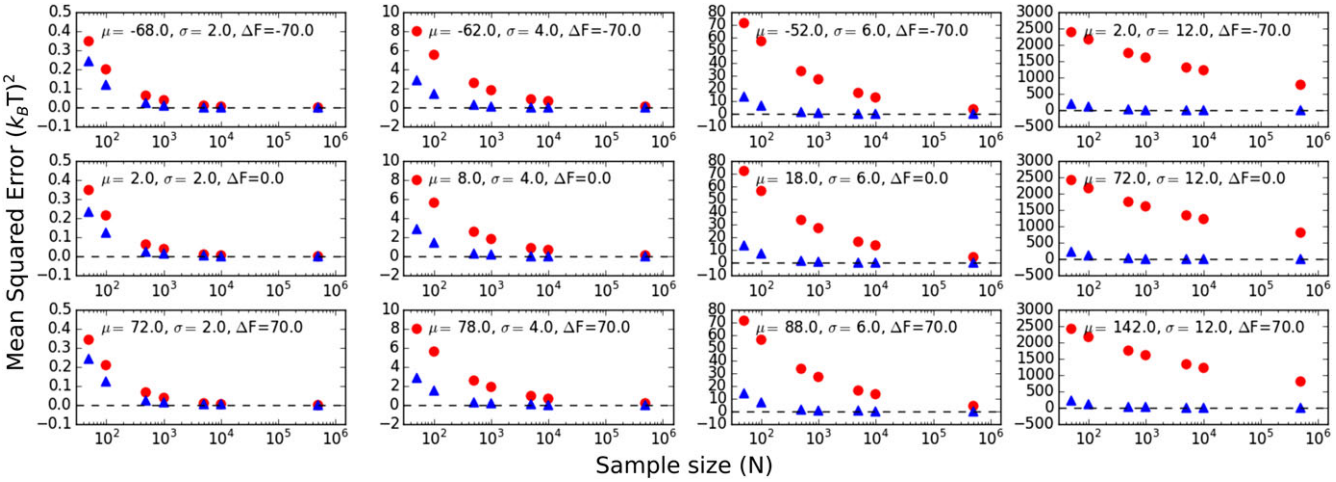

Figure 1. Comparison of mean squared error for $\Delta\hat{F}_J$ (red circles) and $\Delta\hat{F}_{ML}$ (blue triangles) for unidirectional Gaussian distributions. For each estimator the mean squared error (units: $(k_BT)^2$) from 10,000 repetitions is shown as a function of sample size *N*. The grid is organized such that σ increases from left to right, μ increases from top to bottom. Note that subplots have different scales on the y-axes. Arrar M, Boubeta FM, Szretter ME, Sued M, Boechi L, Rodriguez D J. Comput. Chem. [Color figure can be viewed at wileyonlinelibrary.com]

$$\Delta\hat{F}_{ML*} = \hat{\mu}_{ML*} - \frac{\hat{\sigma}^2_{ML*}}{2} = \frac{\overline{W}_{n_F} + \overline{Z}_{n_R}}{2}. \quad (13)$$

The same procedure was applied to the Gamma family, resulting in analogous bidirectional MLEs for α and λ, the latter of which must be determined numerically. Replacing these estimators in the exact formula for ΔF in column 3 of Table 1 for Gamma distributions, we obtain the corresponding bidirectional MLE of ΔF.

The $\Delta\hat{F}_{\mathrm{BAR}}$ estimator, though requiring no knowledge of the underlying distribution, performs almost as well as the parametric bidirectional MLE for Gaussian distributions. For the more disperse distributions, however, in which there is very little overlap between $f_{-R}(w)$ and $f_F(w)$ (far right column of Fig. 3), there is a substantial advantage in using $\Delta\hat{F}_{ML*}$, which we emphasize is simply the average of the average work obtained in the forward and sign-changed reverse distributions. For the Gamma distributions we find that, for the most part, $\Delta\hat{F}_{\mathrm{BAR}}$ and $\Delta\hat{F}_{ML*}$ perform comparably. For distributions with low rate parameters ($\lambda = 0.1(k_BT)^{-1}$), there is a tremendous advantage in using the parametric MLE $\Delta\hat{F}_{ML*}$, as the MSE can reach values as high as 700 $(k_BT)^2$ for small ($n_F = n_R = 50$) sample sizes (Fig. 4); as mentioned earlier, this is a regime in which $f_F(w)$ is very disperse and there is poor overlap between $f_F(w)$ and $f_{-R}(w)$ (Supporting Information Fig. S2). As anticipated, both free energy estimators that utilize bidirectional work distributions perform better than the unidirectional $\Delta\hat{F}_{ML}$ and $\Delta\hat{F}_J$ estimators, as we highlight the substantial difference in scale between either Figures 1 and 3 or Figures 2 and 4.

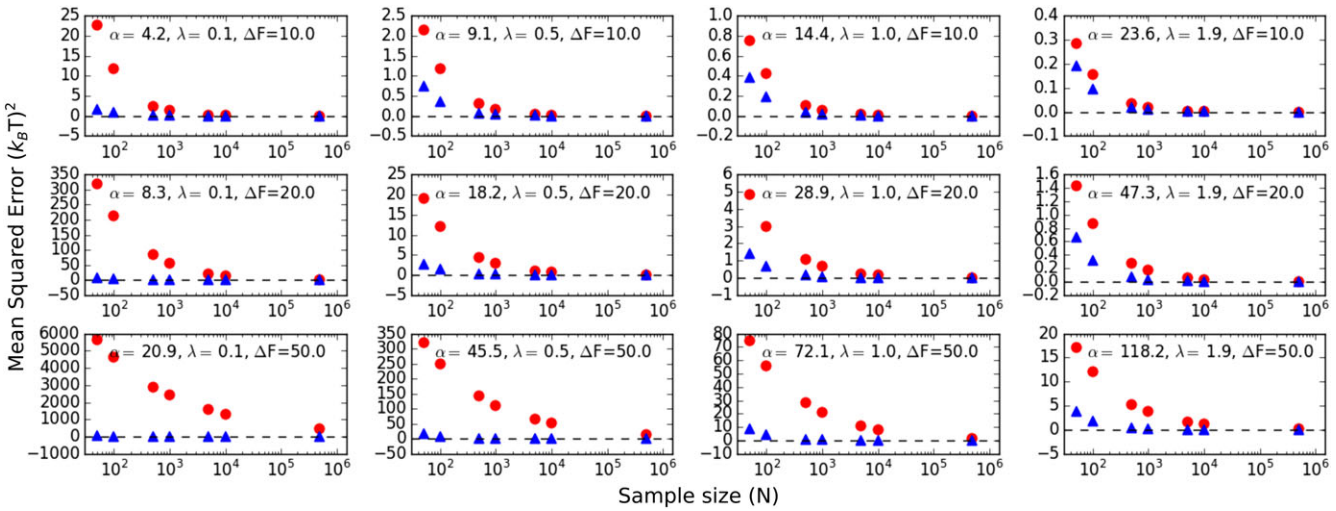

Figure 2. Comparison of mean squared error for $\Delta\hat{F}_J$ (red circles) and $\Delta\hat{F}_{ML}$ (blue triangles) for unidirectional Gamma distributions. For each estimator the mean squared error (units: $(k_BT)^2$) from 10,000 repetitions is shown as a function of sample size *N*. Note that subplots have different y-axes. The grid is organized such that λ increases from left to right, α increase from top to bottom. Arrar M, Boubeta FM, Szretter ME, Sued M, Boechi L, Rodriguez D J. Comput. Chem. [Color figure can be viewed at wileyonlinelibrary.com]

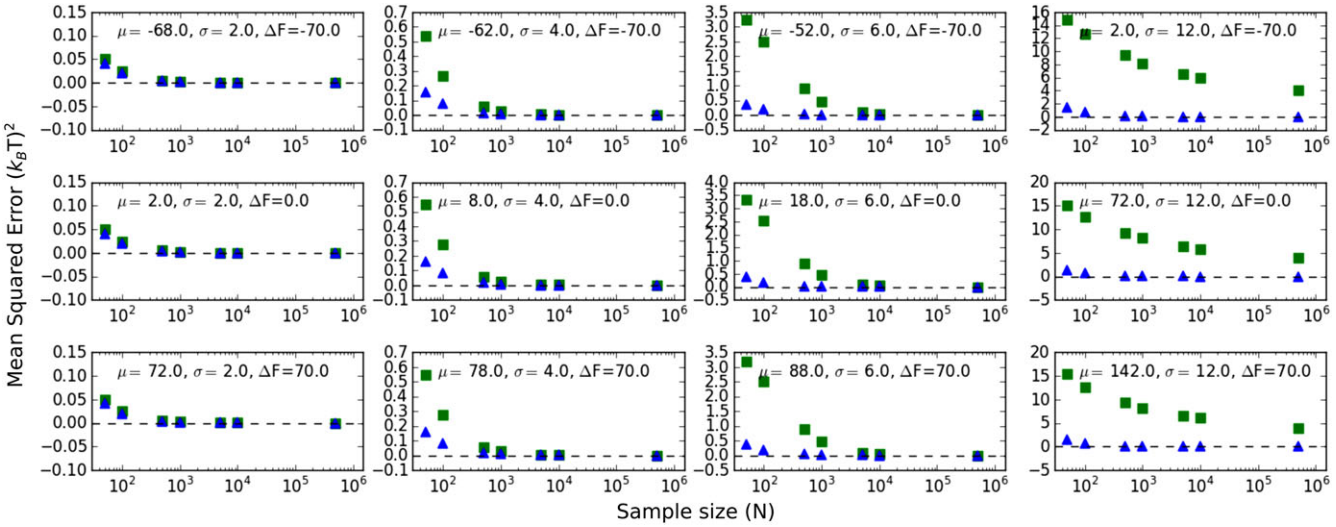

Figure 3. Comparison of mean squared error for $\Delta\hat{F}_{\mathrm{BAR}}$ (green squares) and $\Delta\hat{F}_{ML*}$ (blue triangles) for bidirectional Gaussian distributions. For each estimator the mean squared error (units: $(k_BT)^2$) from 10,000 repetitions is shown as a function of sample size N, with $n_R = n_F = N/2$. Parameters for the forward distribution are shown in each subplot. Arrar M, Boubeta FM, Szretter ME, Sued M, Boechi L, Rodriguez D J. Comput. Chem. [Color figure can be viewed at wileyonlinelibrary.com]

Variance of free energy estimates

Up until now, we have used the MSE as a criterion to compare free energy estimators that do ($\Delta\hat{F}_{ML}$ and $\Delta\hat{F}_{ML*}$) or do not ($\Delta\hat{F}_J$ and $\Delta\hat{F}_{\mathrm{BAR}}$) require assigning a statistical family to the underlying work distributions. As each MSE is calculated with *Nrep* = 10,000 repeated ΔF estimates, we can go one step further in this analysis and compare the empirical density that describes these *Nrep* free energy estimates for each estimator and sample size *N*. In what follows, we will focus on the unidirectional and bidirectional estimators in the Gaussian case, however the analogous analysis for the Gamma case is included as Supporting Information. Because we have already observed that the behavior of the estimators only depends on σ, and not on μ, we will arbitrarily let $\Delta F = 0$.

In Figure 5, we observe that the empirical density of $\Delta\hat{F}_{ML}$ is always centered around the true value of ΔF, whereas the empirical density of $\Delta\hat{F}_J$ clearly indicates a bias, specifically to the right of the true ΔF value, as anticipated by the Jensen inequality.[66,67] Although this bias decreases with increasing sample size, the distribution of free energies obtained with $\Delta\hat{F}_J$ is still biased. In fact, for the work distribution with $\sigma = 12k_BT$, the distribution of $\Delta\hat{F}_J$ still does not include the true value of ΔF, even with a sample size of $N = 500{,}000$.

The bias associated with $\Delta\hat{F}_J$ cannot be predicted without first assigning a statistical model. This is an important and unresolved issue because, as can be seen in Figure 5, the bias can constitute a substantial part of the error associated with the free energy estimate. Nevertheless, we can use error propagation (see Supporting

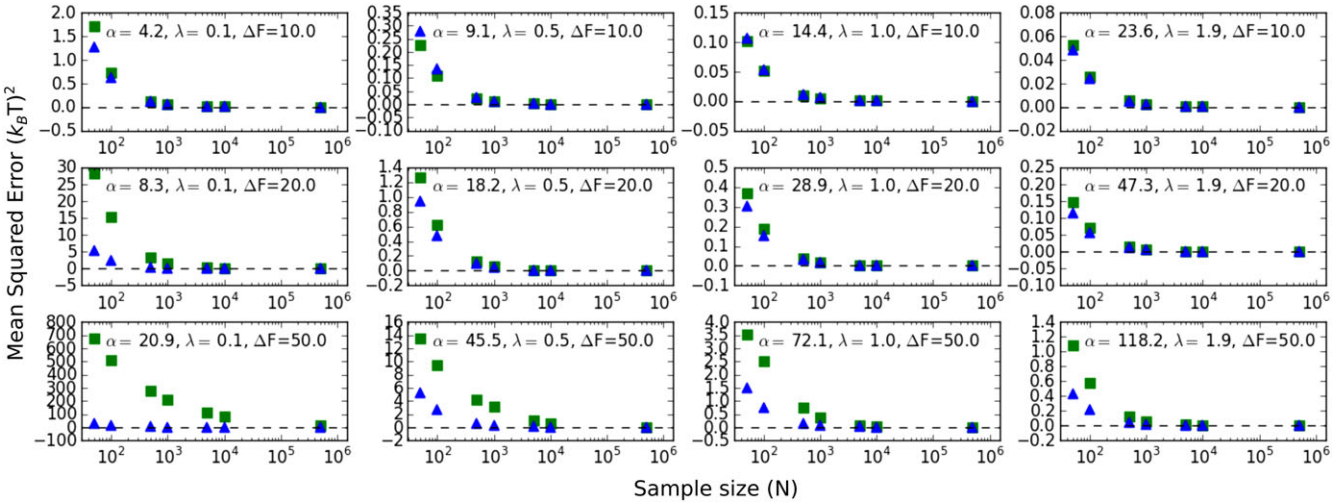

Figure 4. Comparison of mean squared error for $\Delta\hat{F}_{\mathrm{BAR}}$ (green squares) and $\Delta\hat{F}_{ML*}$ (blue triangles) for bidirectional Gamma distributions. For each estimator the mean squared error (units: $(k_BT)^2$) from 10,000 repetitions is shown as a function of sample size *N*. Arrar M, Boubeta FM, Szretter ME, Sued M, Boechi L, Rodriguez D J. Comput. Chem. [Color figure can be viewed at wileyonlinelibrary.com]

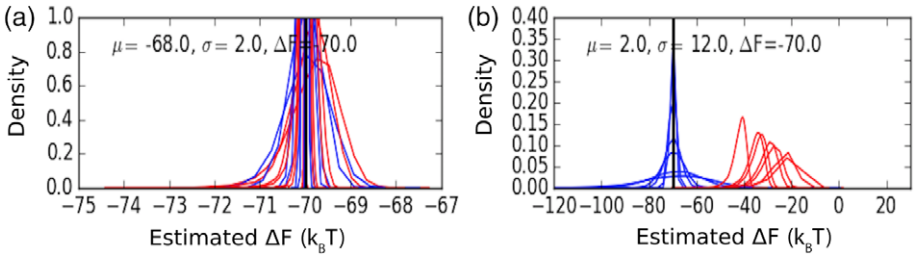

Figure 5. Comparison of empirical densities for of free energy estimates from unidirectional Gaussian work distributions, with $\Delta\hat{F}_J$ in red series and $\Delta\hat{F}_{ML}$ in blue series. For each estimator the empirical density was calculated from 10,000 repetitions and plotted for increasing sample sizes N, showing how the density either approaches the true value of 0 k_BT or becomes more concentrated around it, with larger sample sizes. Arrar M, Boubeta FM, Szretter ME, Sued M, Boechi L, Rodriguez D J. Comput. Chem. [Color figure can be viewed at wileyonlinelibrary.com]

Information) to approximate the variance associated with $\Delta\hat{F}_J$, without assuming a statistical family:

$$\mathrm{var}\left(\Delta\hat{F}_J\right) \simeq \frac{\mathrm{var}[e^{-W}]}{N\langle e^{-W}\rangle^2}. \quad (14)$$

And the asymptotic variance of $\Delta\hat{F}_{ML}$ for the unidirectional Gaussian case is:

$$\mathrm{var}\left(\Delta\hat{F}_{ML}\right) = \frac{\sigma^2}{N} + \frac{\sigma^4(N-1)}{2N^2}. \quad (15)$$

Now, assuming a Gaussian distribution, we can compare var$(\Delta\hat{F}_J)$ (eq. 14) to var$(\Delta\hat{F}_{ML})$ (eq. 15) by expressing $\langle e^{-W}\rangle$ (eq. 11) and var$[e^{-W}]$ in var$(\Delta\hat{F}_J)$ in terms of μ and σ, such that var$(\Delta\hat{F}_J)$ reduces to:

$$\mathrm{var}\left(\Delta\hat{F}_J\right) \simeq \frac{1}{N}\left(e^{\sigma^2}-1\right) = \frac{1}{N}\left(\sigma^2 + \frac{\sigma^4}{2} + R_3\left(\sigma^2\right)\right), \quad (16)$$

where we have rewritten the exponential term as a second-order Taylor expansion, plus the higher order terms $R_3(\sigma^2)$. By further regrouping terms in var$(\Delta\hat{F}_{ML})$ (eq. 15), we can rewrite the expression as $\mathrm{var}(\Delta\hat{F}_{ML}) = \frac{1}{N}\left(\sigma^2 + \frac{\sigma^4}{2}\right) - \frac{\sigma^4}{2N^2}$, making it easier to observe that $\mathrm{var}\left(\Delta\hat{F}_{ML}\right) < \mathrm{var}\left(\Delta\hat{F}_J\right)$. Thus, even in the small-σ regime of Gaussian distributions, in which the bias of $\Delta\hat{F}_J$ appears to be small (Fig. 5), the free energy estimate of $\Delta\hat{F}_{ML}$ is still preferable due to its smaller variance.

For the case of bidirectional Gaussian distributions (Supporting Information Fig. S4), the empirical densities for both $\Delta\hat{F}_{BAR}$ and $\Delta\hat{F}_{ML*}$ show that both estimators are always centered around the true value of ΔF, however the variance of the empirical density of $\Delta\hat{F}_{ML*}$ is smaller than that of $\Delta\hat{F}_{BAR}$. An approximation for the variance of $\Delta\hat{F}_{BAR}$ has been derived in Ref 48 using propagation of errors. The asymptotic variance of $\Delta\hat{F}_{ML*}$ is simply:

$$\mathrm{var}\left(\Delta\hat{F}_{ML*}\right) = \frac{\sigma^2}{N}, \quad (17)$$

and, based on a comparison of var$(\Delta\hat{F}_{ML*})$ (eq. 17) to var$(\Delta\hat{F}_{ML})$ (eq. 15), we note that for Gaussian distributions it is more advantageous to consider $N/2$ work samples in forward and reverse directions, rather than N works in a single direction.

Discussion

We have evaluated the widely celebrated and scrutinized Jarzynski equality, placing special emphasis on the choice of the procedure used to estimate the ensemble average of Boltzmann weights. Although the conventionally invoked Jarzynski estimator requires no characterization of the underlying work distribution, this robustness comes with a loss in statistical efficiency.

One of the most notable validations of the Jarzynski equality, published in 2002 in *Science*, has since attracted a great deal of interest, with over 900 citations to date.[46] In that seminal article, the authors validated the Jarzynski equality with single-molecule pulling experiments, and further assessed that the $\Delta\hat{F}_J$ estimator was more accurate, in certain regimes, than the $\Delta\hat{F}_{FD}$ estimator. This finding, supported by others,[68,69] sparked a surge of applications that invoke the Jarzynski equality specifically using the $\Delta\hat{F}_J$ estimator.[3–11,70] Here, we have demonstrated, however, that if a work distribution is actually Gaussian, as was the case in Ref [46], $\Delta\hat{F}_{FD}$ is guaranteed to be more accurate than $\Delta\hat{F}_J$.

We have shown that the penalty for excluding a known characterization of the underlying work distribution can be tremendous as even a sample size as large as N = 500,000 in some cases is insufficient for $\Delta\hat{F}_J$ to converge. The crux of the matter is that the bias associated with the Jarzynski estimator is unknown, and although expressions for the bias have been proposed,[15,71,72] these estimates require first an assumption of a statistical family for the work distribution. Once a statistical family is assumed, however, the corresponding $\Delta\hat{F}_{ML}$ is the preferred estimator, with lower variance than $\Delta\hat{F}_J$, in the first place.

The problematic convergence of $\Delta\hat{F}_J$ has been identified by others[19,21,22,36,38,40,61,73,74]; indeed its functional form alone makes it a biased estimator that, on average, will over-estimate ΔF for any finite number of work values, according to Jensen's inequality.[66,67] One key pitfall in the efficiency of the Jarzynski estimator, and in any estimator that requires the ensemble average of Boltzmann weights (e.g., in Umbrella Sampling or FEP), lies in the presence of improbable low work values in the left tail of the underlying work distribution. This is, in part, why $\Delta\hat{F}_J$ performs considerably well for work distributions from the Gamma family, in which work values are strictly nonnegative.

We emphasize, however, that even without negative work values, $\Delta\hat{F}_J$ can be grossly inaccurate when the dispersion in the nonequilibrium work values is large. The bias inherent in $\Delta\hat{F}_J$ has been addressed in greater detail elsewhere.[56]

Through an analysis of the phase-space overlap and entropy difference between initial and final states, Kofke and coworkers have proposed a useful heuristic, the so-called pi criteria, as a fail-safe indicator of bias in free energies estimated with the $\Delta\hat{F}_J$ estimator.[55,75] This heuristic does not require explicit assumption of a particular pdf, rather it holds when the phase space of the final state is a subset of the initial state. The π criteria reliably takes on negative values when free energy estimates are biased, and changes sign when bias becomes negligible. For the Gaussian and Gamma unidirectional distributions considered here, we confirmed that the π criteria takes on negative values when the MSE was larger than 1 $(k_BT)^2$ (data not shown), making this heuristic a good general complement to exponential averaging estimators.

Once a statistical family is correctly assigned, its corresponding expression for ΔF (column 3 of Table 1) as a function of the parameters indexing the statistical family gives rise to a more efficient estimator than $\Delta\hat{F}_J$. In fact, these parameters could be estimated by any approach, resulting in other estimators for the free energy. In this article, we have specifically used MLEs as they are the most efficient ones when the statistical family is known,[58] but it is important to emphasize that leveraging the pdf of the work values in the first place generally leads to a preferential use of the Jarzynski equality, regardless of the efficiency of the estimators of the statistical parameters themselves.

Even though several analytically solvable test cases result in Gaussian or Gamma work distributions, and examples of these work distributions can be found in real-world applications, future work is needed to address the conditions under which these distributions can be anticipated beforehand. Therefore, an important concern in using parametric MLEs is how much deviations from the assigned statistical family impact the estimated ΔF values. Strictly speaking, this question cannot be answered without knowing the true underlying distribution, but to gauge the effect of incorrectly assuming a Gaussian distribution, we have evaluated the performance of the parametric MLE corresponding to unidirectional Gaussian work distributions (i.e., the second-order cumulant expansion of the Jarzynski equality) on non-Gaussian distributions, either Skew-Normal or Gamma. From this analysis (Supporting Information Fig. S7), we observe that for left-skewed work distributions (i.e., longer left tail than right tail), both $\Delta\hat{F}_J$ and Gaussian $\Delta\hat{F}_{ML}$ overestimate the true ΔF value, and this effect is more pronounced with increased skewness. For right-skewed distributions, however, the incorrectly assigned $\Delta\hat{F}_{ML}$ estimator consistently underestimates the true ΔF value, to an extent that depends on the magnitude of the skewness of the underlying distribution. We note that Gamma distributions are included among right-skewed distributions. Although this point merits further analysis, a practical approach may leverage this observation along with the fact that $\Delta\hat{F}_J$ consistently overestimates the true ΔF value, and use the incorrectly assigned Gaussian $\Delta\hat{F}_{ML}$ and $\Delta\hat{F}_J$ as lower and upper limits, respectively, for the true ΔF value, for any arbitrary right-skewed work distribution.

When possible, independent pulling simulations can be performed in both forward and reverse directions. Using logistic regression, Shirts et al. have arrived at the $\Delta\hat{F}_{BAR}$ estimator[37] as a nonparametric maximum-likelihood free energy estimator, leveraging bidirectional work distributions. Here, we have considered Gaussian and Gamma statistical families, and, using the CFT have confirmed that, in both cases, the forward and reverse distributions belong to the same family, such that the parameters of the distribution in one direction are sufficient to describe those of the opposite direction. With this information, we derived MLEs for the necessary parameters (e.g., $\hat{\mu}_{ML*}$ and $\hat{\sigma}^2_{ML*}$ for the Gaussian distribution). Interestingly, for the Gaussian case, and when the same number of work values are considered in either direction, we found that the free energy estimator reduces to the average of the works obtained from the forward and reverse (with signs changed for the latter) directions, and this simple average performs even better than the $\Delta\hat{F}_{BAR}$ estimator for Gaussian distributions.

For the case of Gamma distributions the $\Delta\hat{F}_{BAR}$ estimator can be particularly slow to converge, as evidenced by MSEs as high as 700 $(k_BT)^2$ for sample sizes of $N = 100$. This finding is consistent with that of another comparison of free energy estimators,[42] in which $\Delta\hat{F}_{BAR}$ was found to perform poorly in an analytically solvable test case with nested harmonic potentials; we point out that the resulting work distribution in that case is Gamma as well, with $\alpha = 1$.

Together these results highlight regimes of two statistical families in which the parametric MLE of ΔF is particularly more advantageous than $\Delta\hat{F}_J$ or $\Delta\hat{F}_{BAR}$, for unidirectional or bidirectional work distributions, respectively. Furthermore, a comparison of the MSE and the asymptotic variance of the MLEs in the unidirectional and bidirectional cases, we can highlight that considering $N/2$ forward and $N/2$ reverse works is preferable over N unidirectional work families.

The appropriate designation of the statistical family of the work distribution is of utmost importance when translating the results of these numerical simulations to any practical analysis of pulling experiments. With small sample sizes the statistical family may not be clearly distinguishable based on the sample itself. Depending on the way in which the nonequilibrium process is driven, however, the statistical family of the underlying work distribution may be anticipated beforehand. This theoretical anticipation, regardless of evidence from the sampled work distribution itself warrants the use of the corresponding parametric MLE. In this sense, computational single-molecule applications can be more straightforward than analogous force spectroscopic experiments, as the work distribution is generally expected to be Gaussian (or a mixture of Gaussians[13]) as long as a sufficiently stiff spring is used and the relaxations of the system with respect to the changing reaction coordinate are sufficiently fast, which corresponds to being in the low-dissipation regime.[12,45] When the statistical family is incorrectly assigned, however, the resulting free energy estimates can be less accurate than the biased free energy estimate obtained with $\Delta\hat{F}_J$, and thus extreme caution must be exercised in choosing the appropriate $\Delta\hat{F}_{ML}$.

These results highlight the tremendous advantage of theoretically anticipating the nature of the underlying work distributions in estimating free energies as defined by the Jarzynski equality.

Acknowledgments

The authors acknowledge support from the University of Buenos Aires and CONICET.

Keywords: free energy · Jarzynski · steered molecular dynamics · maximum-likelihood

How to cite this article: M. Arrar, F. M. Boubeta, M. E. Szretter, M. Sued, L. Boechi, D. Rodriguez. *J. Comput. Chem.* **2018**, 9999, 1–9. DOI: 10.1002/jcc.25754

Additional Supporting Information may be found in the online version of this article.

[1] C. Jarzynski, *Phys. Rev. Lett.* **1997a**, *78*, 2690.
[2] R. W. Zwanzig, *J. Chem. Phys.* **1955**, *23*, 1915.
[3] A. R. Bizzarri, S. Cannistraro, *Phys. Chem. Chem. Phys.* **2011**, *13*, 2738. https://doi.org/10.1039/C0CP01474E.
[4] M. Orzechowski, P. Cieplak, *J. Phys. Condens. Matter* **2005**, *17*, S1627. http://stacks.iop.org/0953-8984/17/i=18/a=018.
[5] H. Oberhofer, C. Dellago, S. Boresch, *Phys. Rev. E* **2007**, *75*, 061106. http://link.aps.org/doi/10.1103/PhysRevE.75.061106.
[6] W. Liu, V. Montana, V. Parpura, and U. Mohideen, *Biophys. J. 95*, 419 (**2008**), ISSN 0006-3495, URL http://www.sciencedirect.com/science/article/pii/S000634950870314X.
[7] C. Hyeon, D. Thirumalai, *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6789. http://www.pnas.org/content/102/19/6789.full.pdf, http://www.pnas.org/content/102/19/6789.abstract.
[8] G. Ozer, E. F. Valeev, S. Quirk, R. Hernandez, *J. Chem. Theory Comput.* **2010**, *6*, 3026. https://doi.org/10.1021/ct100320g, https://doi.org/10.1021/ct100320g.
[9] F. P. Nicoletti, A. Comandini, A. Bonamore, L. Boechi, F. M. Boubeta, A. Feis, G. Smulevich, A. Boffi, *Biochemistry* **2010**, *49*, 2269.
[10] L. Boechi, M. Arrar, M. A. Martí, J. S. Olson, A. E. Roitberg, D. A. Estrin, *J. Biol. Chem.* **2013**, *288*, 6754.
[11] P. Arroyo Manez, D. E. Bikiel, L. Boechi, L. Capece, S. Di Lella, D. A. Estrin, M. A. Marti, D. M. Moreno, A. D. Nadra, A. A. Petruk, *Biochim. Biophys. Acta* **2011**, *1814*, 1054.
[12] S. Park, K. Schulten, *J. Chem. Phys.* **2004**, *120*, 5946.
[13] P. Procacci and S. Marsili, *Chem. Phys. 375*, 8 (**2010**), ISSN 0301-0104, URL http://www.sciencedirect.com/science/article/pii/S0301010410003216.
[14] C. Jarzynski, *Annu. Rev. Condens. Matter Phys.* **2011**, *2*, 329.
[15] J. Gore, F. Ritort, C. Bustamante, *Proc. Natl. Acad. Sci.* **2003**, *100*, 12564.
[16] J. Hermans, *J. Phys. Chem.* **1991**, *95*, 9029.
[17] R. H. Wood, W. C. Muhlbauer, P. T. Thompson, *J. Phys. Chem.* **1991**, *95*, 6670.
[18] H. B. Callen, T. Welton, *Phys. Rev.* **1951**, *83*, 34.
[19] C. Jarzynski, *Phys. Rev. E* **2006**, *73*, 046105.
[20] N. Yunger Halpern, C. Jarzynski, *Phys. Rev. E* **2016**, *93*, 052144.
[21] A. Pohorille, C. Jarzynski, C. Chipot, *J. Phys. Chem. B* **2010**, *114*, 10235.
[22] G. Hummer, Free Energy Calculations, Springer: New York, **2007**, p. 171.
[23] G. Hummer, *J. Chem. Phys.* **2001**, *114*, 7330.
[24] C. L. Ramírez, A. Zeida, G. E. Jara, A. E. Roitberg, M. A. Marti, *J. Chem. Theory Comput.* **2014**, *10*, 4609.
[25] S. Vaikuntanathan, C. Jarzynski, *Phys. Rev. Lett.* **2008**, *100*, 190601.
[26] T. Schmiedl, U. Seifert, *Phys. Rev. Lett.* **2007**, *98*, 108301.
[27] R. Chelli, *J. Chem. Theory Comput.* **2012**, *8*, 4040.
[28] P. Nicolini, D. Frezzato, R. Chelli, *J. Chem. Theory Comput.* **2011**, *7*, 582.
[29] M. Zerbetto, A. Piserchia, D. Frezzato, *J. Comput. Chem.* **2014**, *35*, 1865.
[30] D. M. Zuckerman and T. B. Woolf, arXiv preprint Phys. Ther./0107066 (**2001**), related:9VAObcuyCwwJ.
[31] F. M. Ytreberg, D. M. Zuckerman, *J. Comput. Chem.* **2004**, *25*, 1749.
[32] D. M. Zuckerman, T. B. Woolf, *Chem. Phys. Lett.* **2002**, *351*, 445.
[33] D. Bucher, R. C. Walker, J. A. McCammon, *J. Chem. Theory Comput.* **2014**, *10*, 2677.
[34] I. Echeverria, L. M. Amzel, *J. Phys. Chem. B* **2012**, *116*, 10986.
[35] G. E. Crooks, *Phys. Rev. E* **2000**, *61*, 2361.
[36] M. R. Shirts, E. Bair, G. Hooker, V. S. Pande, *Phys. Rev. Lett.* **2003**, *91*, 140601.
[37] C. H. Bennett, *J. Comput. Phys.* **1976**, *22*, 245.
[38] G. Hummer, A. Szabo, *Proc. Natl. Acad. Sci.* **2001**, *98*, 3658.
[39] D. D. L. Minh, A. B. Adib, *Phys. Rev. Lett.* **2008**, *100*, 180602. http://link.aps.org/doi/10.1103/PhysRevLett.100.180602.
[40] D. D. L. Minh, J. D. Chodera, *J. Chem. Phys.* **2009**, *131*, 134110.
[41] P. Nicolini, P. Procacci, R. Chelli, *J. Phys. Chem. B* **2010**, *114*, 9546.
[42] M. R. Shirts, V. S. Pande, *J. Chem. Phys.* **2005a**, *122*, 144107.
[43] D. D. Minh, J. A. McCammon, *J. Phys. Chem. B* **2008**, *112*, 5892.
[44] D. M. Ferguson, *J. Chem. Phys.* **1993**, *99*, 10086.
[45] O. Mazonka and C. Jarzynski, arXiv cond-mat/9912121 (**1999**).
[46] J. Liphardt, S. Dumont, S. B. Smith, I. J. Tinoco, C. Bustamante, *Science* **2002**, *296*, 1832.
[47] T. Giorgino, G. De Fabritiis, *J. Chem. Theory Comput.* **2011**, *7*, 1943.
[48] F. M. Boubeta, S. E. Bari, D. A. Estrin, L. Boechi, *J. Phys. Chem. B* **2016**, *120*, 9642.
[49] C. Chipot, J. Hénin, *J. Chem. Phys.* **2005**, *123*, 244906.
[50] H. Xiong, A. Crespo, M. Marti, D. Estrin, A. E. Roitberg, *Theor. Chem. Accounts* **2006**, *116*, 338.
[51] S. Boresch, H. L. Woodcock, *Mol. Phys.* **2017**, *115*, 1200.
[52] M. Nategholeslam, C. G. Gray, B. Tomberli, *J. Phys. Chem. B* **2017**, *121*, 391. https://doi.org/10.1021/acs.jpcb.6b08701, https://doi.org/10.1021/acs.jpcb.6b08701.
[53] N. Lu, D. A. Kofke, *J. Chem. Phys.* **2001a**, *114*, 7303.
[54] N. Lu, D. A. Kofke, *J. Chem. Phys.* **2001b**, *115*, 6866.
[55] D. Wu, D. A. Kofke, *J. Chem. Phys.* **2005a**, *123*, 054103.
[56] D. Wu, D. A. Kofke, *Phys. Rev. E* **2004a**, *70*, 066702.
[57] G. E. Crooks, C. Jarzynski, *Phys. Rev. E* **2007**, *75*, 021116. http://link.aps.org/doi/10.1103/PhysRevE.75.021116.
[58] L. Wasserman, All of statistics: a concise course in statistical inference, Springer Science & Business Media: New York, **2013**.
[59] D. W. Scott, Multivariate density estimation: theory, practice, and visualization, John Wiley & Sons: New Jersey, **2015**.
[60] P. P. Eric Jones, Travis Oliphant, *SciPy: Open source scientific tools for Python* (**2001**), URL http://www.scipy.org/.
[61] M. R. Shirts, J. D. Chodera, *J. Chem. Phys.* **2008**, *129*, 124105.
[62] E. L. Lehmann, G. Casella, Theory of point estimation, Springer Science & Business Media: New York, **2006**.
[63] G. Casella, R. L. Berger, Statistical inference, Vol. *2*, Duxbury Pacific Grove, CA, **2002**.
[64] D. Zhang, J. Gullingsrud, J. A. McCammon, *J. Am. Chem. Soc.* **2006**, *128*, 3019.
[65] K. Hayashi, H. Ueno, R. Iino, H. Noji, *Phys. Rev. Lett.* **2010**, *104*, 218103.
[66] J. L. W. V. Jensen, *Acta Math.* **1906**, *30*, 175.
[67] C. Jarzynski, *Phys. Rev. E* **1997b**, *56*, 5018.
[68] F. Douarche, S. Ciliberto, A. Petrosyan, I. Rabbiosi, *Europhy. Lett.* **2005**, *70*, 593.
[69] N. C. Harris, Y. Song, C.-H. Kiang, *Phys. Rev. Lett.* **2007**, *99*, 068101. http://link.aps.org/doi/10.1103/PhysRevLett.99.068101.
[70] F. Forti, L. Boechi, D. A. Estrin, and M. A. Marti, *J. Comput. Chem. 32*, 2219 (**2011**), ISSN 1096-987X, URL https://doi.org/10.1002/jcc.21805, 2231.
[71] M. Palassini, F. Ritort, *Phys. Rev. Lett.* **2011**, *107*, 060601. http://link.aps.org/doi/10.1103/PhysRevLett.107.060601.
[72] D. Wu, D. A. Kofke, *J. Chem. Phys.* **2004b**, *121*, 8742.
[73] M. R. Shirts, V. S. Pande, *J. Chem. Phys.* **2005b**, *122*, 144107.
[74] D. M. Zuckerman, T. B. Woolf, *Phys. Rev. Lett.* **2002**, *89*, 180602.
[75] D. Wu, D. A. Kofke, *J. Chem. Phys.* **2005b**, *123*, 084109.

Received: 28 March 2018
Revised: 18 June 2018
Accepted: 3 September 2018