Theory of a Systematic Computational Error in Free Energy Differences

Daniel M. Zuckerman[1,*] and Thomas B. Woolf[2,†]

[1]Department of Physiology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205

[2]Department of Physiology and Department of Biophysics, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205

(Received 12 March 2002; published 15 October 2002)

Systematic inaccuracy is inherent in any computational estimate of a nonlinear average, due to the availability of only a finite number of data values, N. Free energy differences ΔF between two states or systems are critically important examples of such averages. Previous work has demonstrated, empirically, that the "finite-sampling error" can be very large—many times k_BT—in ΔF estimates for simple molecular systems. Here we present a theoretical description of the inaccuracy, including the exact solution of a sample problem, the precise asymptotic behavior in terms of $1/N$ for large N, the identification of a universal law, and numerical illustrations. The theory relies on corrections to the central and other limit theorems.

DOI: 10.1103/PhysRevLett.89.180602 PACS numbers: 05.70.Ce, 02.70.-c, 05.40.Fb

Recent interest in free energy difference (ΔF) calculations [1–9] stems from their tremendous range of applications to physical, chemical, and biological systems. Examples include computations relating to crystalline lattices [3,4], the behavior of magnetic models [4,10], and biomolecular binding events—of ligands to both DNA and proteins (e.g., [1,11]). Computations of ΔF, moreover, are formally equivalent to calculating the temperature dependence $F(T)$ [4]. Most recently, it has been pointed out that ΔF calculations are required to convert *experimental* data from nonequilibrium single-molecule pulling measurements to free energy vs extension profiles [5,8]; see also [7].

For more than ten years, however, it has been appreciated that computational estimates of ΔF are inherently subject to "finite-sampling error" [12], that is, to bias whenever the computation is of finite length. Because these inaccuracies can be many times k_BT [9], especially in the important context of large biomolecular calculations, there is a strong motivation to understand and overcome these errors. Although finite-sampling errors accompanying susceptibility computations have been understood on the basis of elementary statistical principles [13], the errors in *nonlinear averages* like ΔF have remained without an explicit theoretical basis. This Letter bridges that gap.

Since the work of Kirkwood [14], it has been appreciated that the free energy difference, $\Delta F \equiv \Delta F_{0\to1}$, of switching from a Hamiltonian $\mathcal{H}_0$ to $\mathcal{H}_1$ is given by a nonlinear average,

$$\Delta F = -k_BT \log[\langle \exp(-W_{0\to1}/k_BT)\rangle_0], \tag{1}$$

where k_BT is the thermal unit of energy at temperature T and $W_{0\to1}$ is the work required to switch the system from $\mathcal{H}_0$ to $\mathcal{H}_1$. The angle brackets indicate an average over switches starting from configurations drawn from the equilibrium distribution governed by $\mathcal{H}_0$. In instantaneous switching the work is defined by $W_{0\to1} = \mathcal{H}_1(\mathbf{x}) - \mathcal{H}_0(\mathbf{x})$ for a start (and end) configuration $\mathbf{x}$; however, Jarzynski noted that gradual switches requiring a "trajectory"-based work definition may also be used [2].

Whenever a convex, nonlinear average such as (1) is estimated computationally, that result will *always* be systematically biased [15] because one has only a finite amount of data, say, N work values. The bias results from incomplete sampling of the smallest (or most negative) $W_{0\to1}$ values: these values dominate the average (1) and cannot be sampled perfectly for finite N, regardless of the $W_{0\to1}$ distribution. Thus, a running estimate of ΔF will typically decline as data are gathered. Such considerations led Wood *et al.* [12] to consider the block-averaged n-data-point estimate of the free energy based on $N = mn$ total work values $\{W^{(k)}\}$, namely,

$$\Delta F_n = \lim_{m\to\infty} \frac{1}{m}\sum_{j=1}^{m} -k_BT \times \log\left[\frac{1}{n}\sum_{k=(j-1)n+1}^{jn} \exp(-W^{(k)}/k_BT)\right]. \tag{2}$$

In the limit, ΔF_n is well defined; it represents the expected value of

$$\mathcal{F}_n = -k_BT \log[(e^{-W_1/k_BT} + \cdots + e^{-W_n/k_BT})/n] \tag{3}$$

(see Fig. 1). Wood *et al.* estimated the lowest order correction to $\Delta F \equiv \Delta F_\infty$ as $\sigma_w^2/2nk_BT$, where σ_w^2 is the variance in the distribution of work values, W [12].

More recently, Zuckerman and Woolf [9] studied ΔF_n. They suggested a means by which a range of ΔF_n values for $n < N$ could be used to extrapolate to the true, infinite-data answer, ΔF. The authors also observed that the free energy appears to be bound according to

$$\Delta F \le \Delta F_n, \qquad \text{any } n, \tag{4}$$

which substantially extends [16] the previous bound, $\Delta F \le \langle W\rangle \equiv \Delta F_1$ [10]. With a lower bound from the

 0031-9007/02/89(18)/180602(4)$20.00

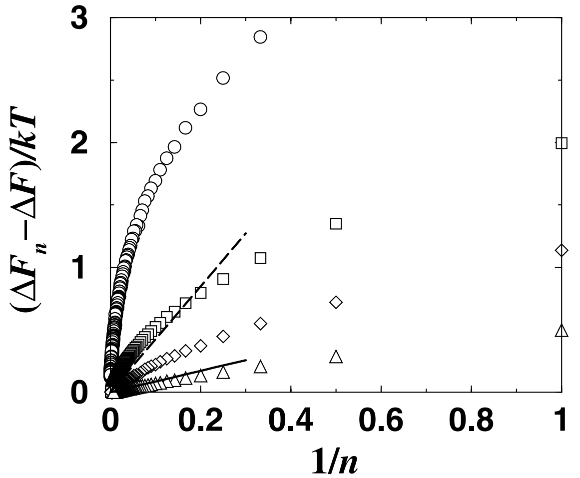

FIG. 1. Finite-sampling error for Gaussian-distributed work values. The expected value of the dimensionless finite-sampling inaccuracy, $(\Delta F_n - \Delta F)/k_BT$, for n data points is plotted as a function of $1/n$. From top to bottom, the data sets represent numerical values of the error for Gaussian distributions of work values with standard deviations, σ_w/k_BT, of 3, 2, 1.5, and 1. The lines (dashed for $\sigma_w/k_BT = 1.5$, solid for $\sigma_w/k_BT = 1$) depict the asymptotic linear behavior for the smallest widths.

reverse process ($1 \to 0$), one can obtain practical brackets on ΔF (e.g., [17]). Finally, Zuckerman and Woolf noted that the leading behavior of ΔF_n appeared to be *not always linear* in $1/n$ but, rather, seemed to behave as $(1/n)^{\tau_1}$ for $\tau_1 \leq 1$.

This Letter presents the theory—apparently for the first time—describing the finite-sampling inaccuracy for ΔF estimates, and extends previous empirical work [9,12]. The inaccuracy occurs regardless of the quality of the conformational sampling (which is assumed to be perfect in the present discussion). Our report includes (i) the formal analytic expression for the expected value of the error from N work values, $\Delta F_N - \Delta F$; (ii) an exact solution, for all N, of this expected value when the Boltzmann factor of the work value $z \equiv e^{-W/k_BT}$ follows a gamma distribution; (iii) exact asymptotic expressions for ΔF_n and the variance of $\mathcal{F}_n$ as $n \to \infty$ for arbitrary W distributions, including nonanalytic behavior in the case when the variance and higher moments of z diverge; and (iv) discussion and numerical illustrations based on Gaussian distributions of W, plus corrections expected from skewed Gaussian distributions. The present discussion makes use of mathematical results regarding the convergence—to "stable" limiting distributions [18–20], also known as Lévy processes (e.g., [21])—of the distributions of sums of variables. The results are expected to have practical application in the extrapolation process outlined in [9].

The theory begins with continuum expressions simplified by the definitions $w \equiv W/k_BT$, $f \equiv \Delta F/k_BT$, and $f_n \equiv \Delta F_n/k_BT$. First, in terms of the probability density ρ_w of work values, which is normalized by $\int dw \rho_w(w) = 1$, the free energy is given by the continuum analog of (1),

$$f = \Delta F/k_BT = -\log\left[\int dw \rho_w(w) e^{-w}\right]. \tag{5}$$

The finite-data average free energy, following (2), must apply the logarithm *before* the average of the n Boltzmann factors, and one has

$$f_n = -\int \prod_{i=1}^{n} [dw_i \rho_w(w_i)] \log\left[\frac{1}{n}\sum_{i=1}^{n} e^{-w_i}\right]. \tag{6}$$

To consider the $n \to \infty$ asymptotics, we introduce a change of variables, motivated by the central and related limit theorems [18,20] for the sum of the e^{-w} variables. We define

$$y = (e^{-w_1} + \cdots + e^{-w_n} - ne^{-f}) / b_1 n^{1/\alpha}, \tag{7}$$

where b_1 is a constant and $\alpha \leq 2$ is an exponent characterizing the distribution of the variable e^{-w}. The finite-data free energy difference can now be written

$$f_n = -\int_{-cn^a}^{\infty} dy \rho_n(y) \log\left(e^{-f} + \frac{b_1}{n^a} y\right), \tag{8}$$

where $c = \exp(-f)/b_1$, $a \equiv (\alpha - 1)/\alpha < 1/2$, and ρ_n is the probability density of the variable y. The requirement that ΔF be finite in (5) further implies $\alpha > 1$, $a > 0$.

To continue, we must call upon some mathematical results regarding the approach, with increasing n, to general stable limit distributions (of which the Gaussian, for $\alpha = 2$, is the best known [18,20]). More precisely, the sum of *any* set of random variables, suitably normalized as in (7), has a distribution with zero mean which may be expressed as a stable distribution function multiplied by a large-n asymptotic expansion [18,22].

To illustrate the case of a Gaussian limit ($\alpha = 2$), assume the variable e^{-w} possesses finite "Boltzmann moments" (a mean $\hat{\mu} = e^{-f}$, variance $\hat{\sigma}^2$, and third moment $\hat{\mu}_3$) not to be confused with the moments of the distribution of w. The finite-n corrections to the central limit theorem indicate that the variable $y = (\sum^n e^{-w_i} - n\hat{\mu})/\sqrt{n}\hat{\sigma}$ [cf. (7)] is distributed according to [18]

$$\rho_n(y) = \rho_G(u; 1)[1 + \nu_1(y)/\sqrt{n} + \nu_2(y)/n + \cdots], \tag{9}$$

for large n, where the remaining terms are higher integer powers of $1/\sqrt{n}$ and the Gaussian density is $\rho_G(y; \sigma) = \exp(-y^2/2\sigma^2)/\sqrt{2\pi}\sigma$. The ν_i depend on the original distribution of e^{-w}; for instance, $\nu_1(y) = (\hat{\mu}_3/6\hat{\sigma}^3)(y^3 - 3y)$ [18]. Moreover, the ν functions are odd or even according to whether i is odd or even, in this $\alpha = 2$ case. See [16] for details.

One arrives at the explicit form of the finite-data-corrected free energy for the case of finite $\hat{\sigma}^2$ and $\hat{\mu}_3$ by substituting (9) into (8), along with an expansion of the logarithm about $y = 0$. (More careful consideration of series convergence for large y yields the same final result for f_n [16].) One finds an expansion consisting *solely of integer powers of* $1/n$, namely,

$$f_n = f + \varphi_1/n + \varphi_2/n^2 + \cdots, \tag{10}$$

with $\varphi_1 = \hat{\sigma}^2/2\hat{\mu}^2$ and $\varphi_2 = -(4\hat{\mu}\hat{\mu}_3 - 9\hat{\sigma}^4)/12\hat{\mu}^4$. To compare this with the finding of Wood *et al.* for $f_n - f$, one can consider a Gaussian distribution of $W = k_BTw$ with variance σ_w^2: expanding the resulting Boltzmann moments for small σ_w yields $\varphi_1 = \{\exp[(\sigma_w/k_BT)^2] - 1\}/2 \approx (\sigma_w/k_BT)^2/2$, which yields precisely the first-order prediction of Wood *et al.* [12].

Figure 1 illustrates the behavior of the finite-data free energy for a Gaussian distribution of work values, based on numerical block averages (2) and the asymptotic behavior given in (10). Although the leading term in $f_n - f$ is linear in $1/n$, the leading coefficients are *exponential* in the *square* of the distribution's width. The asymptotic expressions (10) thus represent viable approximations for only a very small window about $1/n = 0$ for large widths. Figure 1 shows that such behavior is easily mistaken for nonanalytic (e.g., power-law) behavior.

An exactly solvable case occurs when the Boltzmann factor $e^{-w} \equiv z$ is distributed according to a gamma distribution, namely,

$$\rho_\Gamma(z; b, q) = (z/b)^{q-1}\exp(-z/b)/b\Gamma(q). \tag{11}$$

Because this density is "infinitely divisible" [18] the required sums (3) also follow gamma distributions, and after performing the integration in (8) one finds

$$f_n(n; b, q) = \log(n/b) - \psi(nq), \tag{12}$$

where the digamma function is defined by $\psi(x) = (d/dx)\Gamma(x)$. The exact solution is illustrated in Fig. 2 for $b = 10$, $q = 2$.

When asymmetry is added to a Gaussian distribution via the first Edgeworth correction [see (9) and, e.g., [18]], one finds that the exponential dependence of the φ_i on σ_w is corrected only linearly by the now nonzero third moment of the W distribution.

Fundamentally different behavior occurs when the variable $e^{-w} \equiv z$ in (7) possesses a long-tailed distribution ρ_z: the limiting distribution is not a Gaussian and the results (9) and (10) no longer hold. In particular, if one of the tails of $\rho_z(z)$ decays as $z^{-(1+\alpha)}$ with $\alpha < 2$(implying an infinite Boltzmann variance, $\hat{\sigma}^2$), then the distribution of the variable y in (7) approaches a non-Gaussian stable law for large n [20]. Note that such power-law behavior in z corresponds to *simple exponential decay in the work distribution.* Further, because the mean of e^{-w} must be finite for ΔF to exist [recall (5)], we also have $\alpha > 1$. Unfortunately, no explicit forms for stable distributions are known in the range $1 < \alpha < 2$ [20].

A long-tailed z distribution $\rho_z \equiv \rho_1$ also alters the *form* of the asymptotic expansion of the sum-variable y distribution and, hence, the expansion of f_n. Instead of (9), we now have [22]

$$\rho_n(y) = \rho_\alpha(y)\left[1 + \sum\nolimits^* \nu_{uv}(y)/n^{\theta(u,v)}\right], \tag{13}$$

where ρ_α is the appropriate stable probability density with exponent α. The functions ν_{uv}, which are not available analytically, depend on the original distribution of e^{-w} and partial derivatives of the stable distribution. The exponents are given by $\theta(u, v) = (u + \alpha v)/\alpha$, and the summation $\sum^*$ includes $u \geq 0$ and $v \geq -\lceil u/2 \rceil$, where $\lceil x \rceil$ denotes the integer part of x. Note that we have omitted an asymmetry parameter, $\beta = 1$, of the stable laws [20] which does not affect the form of the expansions.

Development of the expansion of f_n in the case of diverging Boltzmann moments will only be sketched here. The basic strategy is to ensure that coefficients of the powers of $1/n$ are rendered in terms of convergent

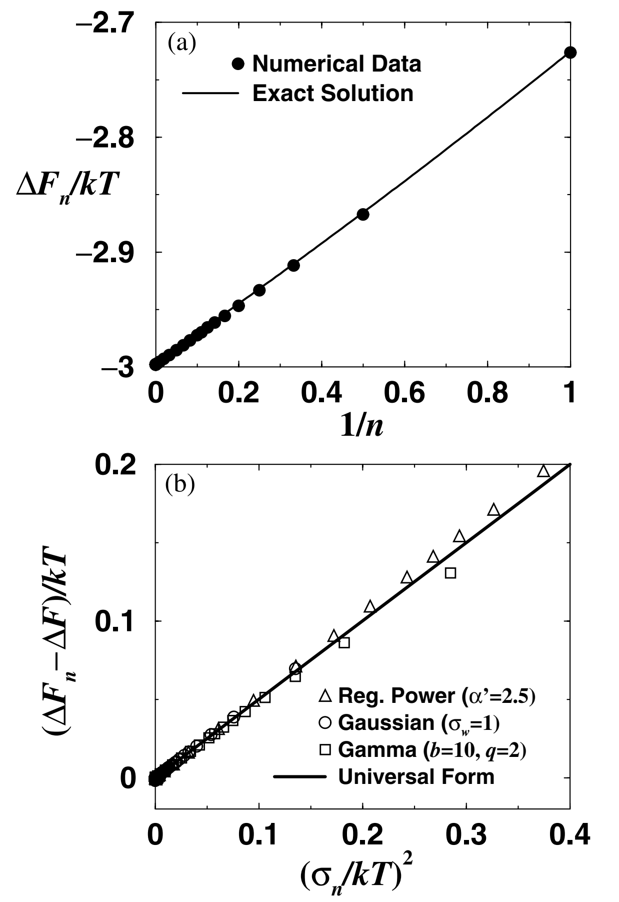

FIG. 2. An exact solution and a universal law. The top panel illustrates the exact solution (12) for the analytic form of $\Delta F_n/k_BT$ when the work Boltzmann factor e^{-W/k_BT} is distributed according to a gamma distribution (11). The bottom plot illustrates the *universal* asymptotic behavior (from several unrelated distributions of work values, W) of the finite-data free energy difference as a function of its fluctuations, σ_n^2; see (17) and text.

integrals. One finds

$$f_n - f \approx \varphi_{\alpha-1}/n^{\alpha-1}, \tag{14}$$

for $n \to \infty$, where $\varphi_{\alpha-1}$ depends on α and on the distribution ρ_1 in a nontrivial way [16].

The fluctuations in the finite-data free energy, $f_n = \Delta F_n/k_BT$, as measured by the variance σ_n of $\mathcal{F}_n$, are of considerable interest because of their potential to provide parameter-free extrapolative estimates of $f_\infty = \Delta F/k_BT$ [9]. Formally, the variance is given by

$$\left(\frac{\sigma_n}{k_BT}\right)^2 = \int_{-cn^a}^{\infty} dy \rho_n(y)[\log(1 + y/cn^a)]^2 - (f_n - f)^2. \tag{15}$$

Techniques analogous to those used above yield asymptotic expansions for the fluctuations. In the case of finite Boltzmann moments, one finds

$$(\sigma_n/k_BT)^2 \approx (\hat{\sigma}/\hat{\mu})^2/n + O(n^{-2}), \tag{16}$$

where the unsubscripted moments refer to the density ρ_z.

Remarkably, comparison with φ_1 for (10) shows that

$$f_n - f = (\sigma_n/k_BT)^2/2 + O(n^{-2}) \tag{17}$$

exactly, as $n \to \infty$, and *independent of any parameters of the distribution*. This *universal* law, valid for the case when the second Boltzmann moment is finite, is illustrated in Fig. 2. In the figure, the "regulated power-law" distribution is defined by $\rho_{rp}(z) = \alpha'/(1+z)^{\alpha'+1}$, and we set $\alpha' = 2.5$. Note that the relation between ΔF_n and its fluctuation seems to parallel that described by Meirovitch using qualitatively different estimates for ΔF [23]. For diverging Boltzmann moments, see [16].

In conclusion, we have presented the first general statistical theory describing the systematic inaccuracy inherent in free-energy-difference estimates based on a finite amount of data (N work values). Our focus has been on the large-N asymptotic behavior, and we have not considered the *independent* problem of conformational sampling in this work. Two cases were formally identified, distinguished by whether the second moment of the distribution of *Boltzmann factors* of the required work values is finite. The asymptotic behavior was discussed for both cases, and, for the finite-second-Boltzmann-moment case, an exact solution and a universal law were presented. Important future work includes the application of the theory outlined here to more robust techniques for extrapolating to $\Delta F \equiv \Delta F_\infty$ from a sequence of ΔF_n values, building on the initial implementation in [9] which suggested that dramatic increases in computational efficiency may be possible. The analysis and extrapolation of ΔF_n data in the case of temperature variation [4] should also prove fruitful.

Finally, as will be emphasized in forthcoming work [16], the analysis presented here applies to a broad class of nonlinear computations beyond that for the free energy.

The authors have benefited greatly from discussions with Michael E. Fisher, Gerhard Hummer, Chris Jarzynski, Hirsh Nanda, Lawrence Pratt, Attila Szabo, Art Voter, and David Zuckerman. We gratefully acknowledge funding provided by the NIH (under Grants No. GM54782 and No. F32GM20394).

*Email address: dzuckerman@ceoh.pitt.edu
Present address: Center for Computational Biology and Bioinformatics, University of Pittsburgh, Pittsburgh, PA 15213.
†Email address: woolf@groucho.med.jhmi.edu

[1] S. B. Singh, Ajay, D. E. Wemmer, and P. A. Kollman, Proc. Natl. Acad. Sci. U.S.A. **91**, 7673 (1994).
[2] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997), Phys. Rev. E **56**, 5018 (1997).
[3] A. D. Bruce, N. B. Wilding, and G. J. Ackland, Phys. Rev. Lett. **79**, 3002 (1997).
[4] M. de Koning, A. Antonelli, and S. Yip, Phys. Rev. Lett. **83**, 3973 (1999).
[5] G. Hummer and A. Szabo, Proc. Natl. Acad. Sci. U.S.A. **98**, 3658 (2001).
[6] C. Jarzynski, Proc. Natl. Acad. Sci. U.S.A. **98**, 3636 (2001).
[7] B. Isralewitz, M. Gao, and K. Schulten, Curr. Opin. Struct. Biol. **11**, 224 (2001).
[8] J. Liphardt *et al.*, Science **296**, 1832 (2002).
[9] D. M. Zuckerman and T. B. Woolf, Chem. Phys. Lett. **351**, 445 (2002).
[10] W. P. Reinhardt and J. E. Hunter, J. Chem. Phys. **97**, 1599 (1992).
[11] P. A. Kollman, Chem. Rev. **93**, 2395 (1993).
[12] R. H. Wood, W. C. F. Mühlbauer, and P. T. Thompson, J. Phys. Chem. **95**, 6670 (1991).
[13] A. M. Ferrenberg, D. P. Landau, and K. Binder, J. Stat. Phys. **63**, 867 (1991).
[14] J. G. Kirkwood, J. Chem. Phys. **3**, 300 (1935).
[15] A. D. Stone and J. D. Joannopoulos, Phys. Rev. E **25**, 2400 (1982).
[16] D. M. Zuckerman and T. B. Woolf, physics/0208015 (unpublished).
[17] C. Jarque and B. Tidor, J. Phys. Chem. B **101**, 9402 (1997).
[18] W. Feller, *An Introduction to Probability Theory and Its Applications* (Wiley, New York, 1971), Vol. 2.
[19] V. M. Zolotarev, *One-Dimensional Stable Distributions* (American Mathematical Society, Providence, 1986).
[20] V. V. Uchaikin and V. M. Zolotarev, *Chance and Stability: Stable Distributions and Their Applications* (VSP, Utrecht, 1999).
[21] M. F. Shlesinger, G. M. Zaslavsky, and U. Frisch, *Lévy Flights and Related Topics in Physics* (Springer, Berlin, 1995).
[22] G. Christoph and W. Wolf, *Convergence Theorems with a Stable Limit Law* (Akadmie Verlag, Berlin, 1992).
[23] H. Meirovitch, J. Chem. Phys. **111**, 7215 (1999).