Measures of trajectory ensemble disparity in nonequilibrium statistical dynamics

Gavin E. Crooks and David A. Sivak

Physical Biosciences Division Lawrence Berkeley National Laboratory One Cyclotron Road Berkeley, CA 94720

Abstract.   Many interesting divergence measures between conjugate ensembles of nonequilibrium trajectories can be experimentally determined from the work distribution of the process.   Herein, we review the statistical and physical significance of several of these measures, in particular the relative entropy (dissipation), Jeffreys divergence (hysteresis), Jensen-Shannon divergence (time-asymmetry), Chernoff divergence (work cumulant generating function), and Rényi divergence. ‡

1. Csiszár f-divergences and work measurements

Many interesting divergence measures between probability distributions can be written as (or related to) an f-divergence (also know as Csiszár, Csiszár-Morimoto or Ali-Silvey divergence) [1–6],

$$C_f(p;q) \equiv \sum_i p_i \; f\!\left(\frac{q_i}{p_i}\right) , \tag{1}$$

where f is a function that takes the ratio of the probabilities as an argument. We will discuss several examples, including the relative entropy, Eq. (6), Jeffreys divergence (9), Jensen-Shannon divergence (11), Chernoff divergence (15), Bhattacharyya distance (24), Rényi divergence (25) and Tsallis divergence (26). Typically, the function f is convex and $f(1) = 0$, which implies that $C_f(p;q) \geq 0$ from an application of Jensen's inequality. Note that the first argument to the f-divergence appears in the denominator of the ratio and is the distribution to be averaged over. The opposite convention is also common.

   Several of these divergence measures have interesting physical interpretations in equilibrium statistical mechanics. For instance, the relative entropy between the nonequilibrium ensemble and the canonical ensemble of the same system relaxed to thermal equilibrium is equal to the free energy difference [7]; the Rényi entropy (the Rényi divergence with a uniform reference distribution) is proportional to the free energy change due to a jump in temperature [8, 9]; and the Tsallis entropy [10] has been advanced as an extension of Boltzmann-Gibbs-Shannon entropy to non-extensive systems.

   In this paper, we wish to consider the physical significance of f-divergences applied to nonequilibrium dynamics. Consider a physical system driven from thermal equilibrium by an external perturbation. For such an experimental protocol Λ denotes the time course of a set of controllable parameters $\lambda(t)$ for $t \in [a, b]$ which describe how the system is driven from the initial equilibrium at $\lambda(a)$. In the conjugate, time-reversed protocol $\tilde{\Lambda}$ the system begins in thermal equilibrium at $\lambda(b)$ and the controllable parameters retrace the same series of changes, in reverse, back to $\lambda(a)$. As a consequence of the time-reversal symmetry of the underlying dynamics, the ratio of the probability of a trajectory during the forward protocol $P[z|\Lambda]$ and the probability of its conjugate trajectory during the reverse protocol $P[\tilde{z}|\tilde{\Lambda}]$ is [11–15]

$$\frac{P[z|\Lambda]}{P[\tilde{z}|\tilde{\Lambda}]} = e^{\beta W[z,\Lambda] - \beta \Delta F_\Lambda} , \tag{2}$$

where z represents a trajectory through phase space, z̃ represents the conjugate, time-reversed trajectory, $\beta = 1/k_\mathrm{B}T$, T is the temperature of the environment in natural units (k_B is the Boltzmann constant), $W[z, \Lambda]$ is the work performed on the system during the forward protocol Λ [11, 16, 17] and $\Delta F_\Lambda = F_{\lambda(b)} - F_{\lambda(a)}$ is the difference in free energy between the initial equilibrium and final equilibrium ensembles. We drop the explicit dependence of work and free energy change on the protocol and trajectory when these dependences are clear from the context. Note that both work and free energy change are odd under time-reversal, $W[z, \Lambda] = -W[\tilde{z}, \tilde{\Lambda}]$ and $\Delta F_\Lambda = -\Delta F_{\tilde{\Lambda}}$.

   For nonequilibrium systems, the key feature of the f-divergence is that, apart from the overall average, the two probability distributions enter only as a ratio. As a consequence, we can always relate the divergence between conjugate pairs of trajectory ensembles to averages of the work measured under a protocol Λ,

$$C_f\big(P[z|\Lambda] \; ; \; P[\tilde{z}|\tilde{\Lambda}]\big) = \big\langle f(e^{-\beta W + \beta \Delta F}) \big\rangle_\Lambda . \tag{3}$$

Hence, if we can measure the work histogram for a process and its reversal, for example in a computer simulation or single-molecule unfolding/refolding experiment [18–20], then we can also measure the Csiszár f-divergences for that system.

Many interesting instances of the f-divergence are symmetric under interchange of the two distributions,

$$JC_f(p;q) \equiv \sum_i p_i f\left(\frac{q_i}{p_i}\right) + \sum_i q_i f\left(\frac{p_i}{q_i}\right) \tag{4}$$

$$= C_g(p;q) , \quad g(x) = f(x) + xf(x^{-1}) .$$

In these cases, the conjugate trajectory f-divergences can be evaluated from averages over each conjugate protocol,

$$JC_f\big(P[\mathrm{z}|\Lambda] \; ; \; P[\tilde{\mathrm{z}}|\tilde{\Lambda}]\big) = \big\langle f(e^{-\beta W + \beta \Delta F})\big\rangle_\Lambda + \big\langle f(e^{-\beta W + \beta \Delta F})\big\rangle_{\tilde{\Lambda}} . \tag{5}$$

In the following discussion, we examine various instances of Csiszár f-divergence that have interesting statistical and physical interpretations when applied to conjugate ensembles of nonequilibrium trajectories. In particular, we review the physical significance of the relative entropy (dissipation), Jeffreys divergence (hysteresis), and Jensen-Shannon divergence (time-asymmetry), and investigate the application of the Chernoff divergence (work cumulant generating function).

2. Relative entropy and dissipation

The relative entropy (or Kullback-Leibler divergence) is defined as [21, 22]

$$D(p\|q) \equiv \sum_i p_i \ln \frac{p_i}{q_i} \tag{6}$$

$$= C_f(p;q), \quad f(x) = -\ln x .$$

Roughly speaking, the relative entropy measures the difference between two distributions, although it is not a metric since it is not symmetric [$D(p\|q) \neq D(q\|p)$ in general], nor does it obey the triangle inequality. Recall that a metric provides a measure of 'distance' between points: it is a real function $g(a, b)$ such that (1) distances are non-negative, $g(a, b) \geq 0$ with equality if, and only if, $a = b$, (2) distances are symmetric, $g(a, b) = g(b, a)$, and (3) it is generally shorter to go directly from point a to c than to go by way of b, $g(a, b) + g(b, c) \geq g(a, c)$ (the triangle inequality). One interpretation of relative entropy is that it represents an encoding cost [22]: if we encode messages using an optimal code for a probability distribution q_i of messages i, but the messages actually arrive with probabilities p_i, then each message requires, on average, an additional $D(p\|q)$ nats (1 nat = $\ln 2$ bits) to encode compared to the optimal encoding.

For conjugate trajectory ensembles the relative entropy is equal to [14, 15, 23–28]

$$D\big(P[\mathrm{z}|\Lambda] \; \big\| \; P[\tilde{\mathrm{z}}|\tilde{\Lambda}]\big) = \beta\langle W\rangle_\Lambda - \beta \Delta F_\Lambda . \tag{7}$$

Here, $\langle W\rangle_\Lambda$ is the average work of the protocol and ΔF is the free energy change. The free energy change is equal to the reversible work, the amount of work performed on the system during a thermodynamically reversible transformation. The difference $\langle W\rangle_\Lambda - \Delta F$ is the average excess work performed on the system by an irreversible protocol over and above the work performed during a reversible transformation. As

the system relaxes back to equilibrium, any excess energy is dissipated as heat, raising the entropy of the environment. Consequently,

$$\begin{aligned}
\beta\langle W\rangle_\Lambda - \beta\Delta F_\Lambda &= \beta\langle W\rangle_\Lambda - \beta\Delta\langle E\rangle_\Lambda + \Delta S_\Lambda^{\text{system}} \\
&= -\beta\langle Q\rangle_\Lambda + \Delta S_\Lambda^{\text{system}} \\
&= \Delta S_\Lambda^{\text{environment}} + \Delta S_\Lambda^{\text{system}} \\
&= \Delta S_\Lambda^{\text{total}} \ .
\end{aligned} \tag{8}$$

Here, Q is the heat and $\Delta E = Q + W$. Thus, the relative entropy of conjugate trajectory ensembles is equal to the total dissipation, the total increase in entropy due to the protocol, once the system has relaxed back to equilibrium. For conjugate trajectory ensembles, the relative entropy also represents a thermodynamic cost, since the dissipation is the average work irretrievably lost into the environment during execution of the protocol.

3. Jeffreys divergence and hysteresis

The Jeffreys divergence (J-divergence or symmetrized Kullback-Leibler divergence) [21, 29],

$$\begin{aligned}
\text{Jeffreys}(p; q) &\equiv D(p\|q) + D(q\|p) \tag{9} \\
&= \sum_i p_i \ln \frac{p_i}{q_i} + \sum_i q_i \ln \frac{q_i}{p_i} \\
&= C_f(p; q), \quad f(x) = (x-1)\ln x \ ,
\end{aligned}$$

is a symmetrized relative entropy, Eq. (6). It is symmetric and non-negative, but not a metric since it does not obey the triangle inequality. Note that some authors define Jeffreys divergence as the symmetrized average of the relative entropy, half the value defined by Jeffreys.

For conjugate trajectory ensembles, the Jeffreys divergence is twice the hysteresis, the average dissipation of the forward and reverse protocols [26],

$$\begin{aligned}
\text{Jeffreys}\left(P[\text{z}|\Lambda] \ ; \ P[\tilde{\text{z}}|\tilde{\Lambda}]\right) &\tag{10} \\
&= \beta\langle W\rangle_\Lambda - \beta\Delta F_\Lambda + \beta\langle W\rangle_{\tilde{\Lambda}} - \beta\Delta F_{\tilde{\Lambda}} \\
&= \beta\langle W\rangle_\Lambda + \beta\langle W\rangle_{\tilde{\Lambda}} \\
&= 2 \times \text{hysteresis} \ .
\end{aligned}$$

4. Jensen-Shannon divergence and time-asymmetry

The Jensen-Shannon divergence is defined as [30]

$$\begin{aligned}
\text{JS}(p; q) &\equiv \frac{1}{2}D\left(p \ \| \ \frac{1}{2}(p+q)\right) + \frac{1}{2}D\left(q \ \| \ \frac{1}{2}(p+q)\right) \tag{11} \\
&= \frac{1}{2}\sum_i p_i \ln \frac{p_i}{\frac{1}{2}(p_i+q_i)} + \frac{1}{2}\sum_i q_i \ln \frac{q_i}{\frac{1}{2}(p_i+q_i)} \\
&= C_f(p; q), \quad f(x) = \frac{1}{2}\ln\frac{2}{1+x} + \frac{1}{2}x\ln\frac{2x}{1+x} \ .
\end{aligned}$$

Each of the two summands is the relative entropy between one of the distributions and the mean of the two distributions.

One interpretation of the Jensen-Shannon divergence is in terms of a Bayesian inference problem [31]: given a sample taken from one of two probability distributions, p or q, the Jensen-Shannon divergence is the average information the sample provides about the identity of the distribution. The divergence is equal to zero only if the two distributions are identical, and therefore indistinguishable, and reaches its maximum value of $\ln 2$ nats (i.e. 1 bit) if the two distributions do not overlap and therefore are perfectly distinguishable from a single sample.

For conjugate trajectory ensembles, the Jensen-Shannon divergence measures the distinguishability of conjugate forward and reverse experiments. In other words, given a microscopic trajectory z, can we tell if it was generated by the protocol Λ, or if it is the time-reversal of a trajectory generated by the reverse protocol $\tilde{\Lambda}$? The Jensen-Shannon divergence measures the average gain in information about the orientation of time's arrow obtained from one realization of the experiment, essentially the time-asymmetry A of a driven system [26],

$$\mathrm{JS}\left(P[z|\Lambda]\ ;\ P[\tilde{z}|\tilde{\Lambda}]\right) = A[\Lambda; \tilde{\Lambda}]\ . \tag{12}$$

As the square root of the Jensen-Shannon divergence is a metric between probability distributions [31, 32], so too the square root of the time-asymmetry is a metric between the forward and reverse protocols (i.e. between conjugate ensembles of trajectories), the length of time's arrow [26].

The time-asymmetry A equals a non-linear average of the work,

$$A[\Lambda; \tilde{\Lambda}] = \frac{1}{2}\left\langle \ln \frac{2}{1 + \exp(-\beta W[z|\Lambda] + \beta \Delta F)}\right\rangle_{\Lambda} \tag{13}$$
$$+ \frac{1}{2}\left\langle \ln \frac{2}{1 + \exp(-\beta W[\tilde{z}|\tilde{\Lambda}] - \beta \Delta F)}\right\rangle_{\tilde{\Lambda}}\ .$$

Note that the work distribution is a sufficient statistic for the time-asymmetry: all other details of the microscopic trajectories are superfluous.

The Jensen-Shannon and Jeffreys divergences are related by the inequalities [5, 26],

$$\mathrm{JS}(p;q) \leq \frac{1}{8}\, \mathrm{Jeffreys}(p;q) \tag{14}$$

$$\mathrm{JS}(p;q) \leq \ln \frac{2}{1 + \exp\left(-\frac{1}{2}\, \mathrm{Jeffreys}(p;q)\right)}\ .$$

These inequalities imply that a given time-asymmetry requires a certain minimum hysteresis.

5. Chernoff divergence and work cumulant generating functions

The Chernoff divergence of order α is defined as [33, 34]

$$\mathrm{Chernoff}_{\alpha}(p;q) \equiv -\ln \sum_{i} p_i \left(\frac{p_i}{q_i}\right)^{\alpha-1} \tag{15}$$
$$= -\ln\left[C_f(p;q) + 1\right], \quad f(x) = x^{1-\alpha} - 1\ .$$

The Chernoff divergence is zero for $\alpha = 1$ and $\alpha = 0$, and reaches a maximum, the Chernoff information [22, 33], for some intermediate value of alpha. The Chernoff

divergence is well defined for $\alpha > 1$ if $q_i > 0$ whenever $p_i > 0$, and for $\alpha < 0$ if $p_i > 0$ whenever $q_i > 0$, and thus defined for all α if the distributions have the same support.

The Chernoff divergence of order α is related to the Chernoff divergence of order $1 - \alpha$ with the distributions interchanged [34],

$$\text{Chernoff}_\alpha(p; q) = \text{Chernoff}_{1-\alpha}(q; p) . \tag{16}$$

This relation always holds for $\alpha \in [0, 1]$, and for all α when the distributions have the same support.

For conjugate trajectory ensembles, the Chernoff divergence of order $1 - \alpha$ is proportional to the cumulant generating function for the excess work,

$$\text{Chernoff}_{1-\alpha}\Big(P[z|\Lambda]; P[\tilde{z}|\tilde{\Lambda}]\Big) = -\ln\left\langle e^{-\alpha(\beta W - \beta \Delta F)}\right\rangle_\Lambda . \tag{17}$$

Recall that a cumulant generating function has the form

$$\ln\left\langle e^{tz}\right\rangle = \sum_{n=1}^{\infty} \frac{t^n}{n!} \kappa_n[z] , \tag{18}$$

and that the kth-derivative of the cumulant generating function evaluated at zero is the kth cumulant,

$$\frac{d^k}{dt^k} \ln\left\langle e^{tz}\right\rangle \bigg|_{t=0} = \kappa_n[z] . \tag{19}$$

Cumulants of order n are functions of nth order and lower moments. The first cumulant κ_1 is the mean, the second κ_2 is the variance, and the third κ_3 is equal to the third central moment.

The symmetry of Chernoff divergence under interchange of the distributions, Eq. (16), implies that the excess work cumulant generating functions for the forward and reverse protocols are related [35, 36],

$$\ln\left\langle e^{-\alpha(\beta W - \beta \Delta F)}\right\rangle_\Lambda = \ln\left\langle e^{-(1-\alpha)(\beta W - \beta \Delta F)}\right\rangle_{\tilde{\Lambda}} , \tag{20}$$

or equivalently

$$\ln\left\langle e^{-\alpha \beta W}\right\rangle_\Lambda = \ln\left\langle e^{-(1-\alpha)\beta W}\right\rangle_{\tilde{\Lambda}} - \beta \Delta F_\Lambda . \tag{21}$$

Note that an additional minus sign enters into the right hand expression because both the work and free energy change are odd under time-reversal, $W[z, \Lambda] - \Delta F_\Lambda = -W[\tilde{z}, \tilde{\Lambda}] + \Delta F_{\tilde{\Lambda}}$. If we set $\alpha = 1$ or 0 we recover the Jarzynski identity, $\ln\langle \exp\{-\beta W\}\rangle_\Lambda = -\beta \Delta F_\Lambda$ [16].

This symmetry between cumulant generating functions implies that the work cumulants under a given protocol are related to the work cumulants of the conjugate protocol [36],

$$\kappa_k\big[\beta W | \Lambda\big] = \sum_{n=k}^{\infty} \frac{(-1)^n}{(n-k)!} \kappa_n\big[\beta W | \tilde{\Lambda}\big] . \tag{22}$$

From this relation Hummer and Szabo [36, 37] derive optimal estimators of free energy, given only the first m work cumulants. In a parallel development, this symmetry is also exploited in the large deviation approach to steady state fluctuation theorems [35].

The case $\alpha = \frac{1}{2}$ is related to the Bhattacharyya distance, another measure of probability distribution overlap [38],

$$\text{Bhattacharyya}(p; q) = -\ln \sum_i \sqrt{p_i q_i} \tag{23}$$

$$= \text{Chernoff}_{\frac{1}{2}}(p; q) .$$

The Bhattacharyya distance is invariant to interchange of p and q. For conjugate trajectory ensembles, this symmetry implies a relation between exponential averages of work,

$$\ln \left\langle e^{-\frac{1}{2}(\beta W - \beta \Delta F)} \right\rangle_\Lambda = \ln \left\langle e^{-\frac{1}{2}(\beta W - \beta \Delta F)} \right\rangle_{\tilde{\Lambda}} . \tag{24}$$

This relation has no immediately obvious utility, other than being attractively symmetric.

6. Rényi divergence and the maximum work

The Chernoff divergence is closely related to the Rényi, Tsallis and Cressie-Read divergences. The Rényi divergence (or relative Rényi entropy) of order α is a one-parameter generalization of the relative entropy [39],

$$\text{Rényi}_\alpha(p; q) \equiv \frac{1}{\alpha - 1} \ln \sum_i p_i \left(\frac{p_i}{q_i} \right)^{\alpha - 1} \tag{25}$$

$$= \frac{1}{1 - \alpha} \text{Chernoff}_\alpha(p; q)$$

$$= \frac{1}{\alpha - 1} \ln \left[C_f(p; q) + 1 \right], \quad f(x) = x^{1-\alpha} - 1 .$$

Higher values of α give a Rényi divergence dominated by the greatest ratio between the two distributions, whereas as α approaches zero the Rényi entropy weighs all possibilities more equally, regardless of their dissimilarities. We recover the relative entropy in the limit of $\alpha \to 1$.

Other closely related divergences include the relative Tsallis entropy [10],

$$\text{Tsallis}_\alpha(p; q) \equiv \frac{1}{\alpha - 1} \sum_i p_i \left[\left(\frac{p_i}{q_i} \right)^{\alpha - 1} - 1 \right] \tag{26}$$

$$= \frac{1}{\alpha - 1} \left[e^{(\alpha - 1) \, \text{Rényi}_\alpha(p; q)} - 1 \right]$$

$$= C_f(p; q), \quad f(x) = \frac{x^{1-\alpha} - 1}{\alpha - 1} ,$$

and the equivalent, but less well known, Cressie-Read divergence [40],

$$\text{CressieRead}_\alpha(p; q) \equiv \frac{1}{\alpha(\alpha + 1)} \sum_i p_i \left[\left(\frac{p_i}{q_i} \right)^\alpha - 1 \right] \tag{27}$$

$$= \frac{1}{\alpha(\alpha + 1)} \left[e^{\alpha \, \text{Rényi}_{\alpha+1}(p; q)} - 1 \right]$$

$$= \frac{1}{\alpha + 1} \text{Tsallis}_{\alpha+1}(p; q)$$

$$= C_f(p; q), \quad f(x) = \frac{x^{-\alpha} - 1}{\alpha(\alpha + 1)} .$$

Interesting special cases of the Rényi divergence occur for $\alpha = 0, \frac{1}{2}, 1$ and ∞. As previously mentioned, $\alpha = 1$ gives the relative entropy (6), and $\alpha = \frac{1}{2}$ gives the Bhattacharyya distance (24). In the limit $\alpha \to 0$, the Rényi divergence slides to the

negative log probability under q that p is non-zero,

$$\lim_{\alpha \to 0} \text{Rényi}_\alpha(p; q) = -\ln \sum_i \lim_{\alpha \to 0} p_i^\alpha q_i^{1-\alpha} \tag{28}$$

$$= -\ln \sum_i q_i[p_i > 0] .$$

Here we have used the Iverson bracket, which evaluates to one if the condition inside the bracket is true, and zero otherwise. For conjugate trajectory ensembles, the two distributions have the same support and thus in the limit $\alpha \to 0$ the Rényi divergence is zero.

In the limit $\alpha \to \infty$ the Rényi divergence picks out the maximum divergence, the log maximum ratio of the probabilities,

$$\text{Rényi}_\infty(p; q) = \ln \max_i \frac{p_i}{q_i} . \tag{29}$$

For conjugate trajectory ensembles, the maximum divergence reflects the maximum possible work during a single realization of the protocol,

$$\text{Rényi}_\infty \left(P[\text{z}|\Lambda]; P[\tilde{\text{z}}|\tilde{\Lambda}] \right) = \max_{\text{z}} \left(\beta W[\text{z}, \Lambda] - \beta \Delta F_\Lambda \right) . \tag{30}$$

In a thermodynamic context, the maximum work represents the worst case scenario under a particular protocol, the single realization that creates the largest increase in entropy.

7. Illustrative analytic example

To make these relations more concrete, we consider a system consisting of a micron-sized bead suspended in water at inverse temperature β by an initially stationary optical laser trap, with spring constant k. The trap is then translated at a constant velocity v, dragging the bead through the fluid with friction coefficient γ for a time t. This system can be modeled by a single particle undergoing overdamped Langevin dynamics in a one dimensional moving harmonic potential, and the pertinent properties of the model have been analyzed [27, 41]. The free energy change is zero and the work distribution is Gaussian with mean

$$W_0 \equiv \gamma v^2 \left[t - (1 - e^{-kt/\gamma}) \gamma/k \right] , \tag{31}$$

and variance $2W_0$.

The relative entropy, Jeffreys divergence and Chernoff divergence for this model are all simple functions of βW_0:

$$D\left(P[\text{z}|\Lambda] \parallel P[\tilde{\text{z}}|\tilde{\Lambda}] \right) \qquad = \beta W_0 \tag{32}$$

$$\text{Jeffreys} \left(P[\text{z}|\Lambda] \; ; \; P[\tilde{\text{z}}|\tilde{\Lambda}] \right) \quad = 2\beta W_0$$

$$\text{Chernoff}_{1-\alpha} \left(P[\text{z}|\Lambda]; P[\tilde{\text{z}}|\tilde{\Lambda}] \right) = \alpha(\alpha - 1)\beta W_0 .$$

The Jensen-Shannon divergence does not appear to have a simple closed-form solution for this system.

8. Epilogue

As we have seen, a number of f-divergences have both interesting statistical and physical interpretations for conjugate ensembles of nonequilibrium trajectories, and can be measured in computer simulation and real world experiments. This allows us to exploit the statistical machinery of information theory to gain physical insights into the behavior of nonequilibrium systems. It is an open problem whether any other f-divergences (such as the arithmetic-geometric mean divergence, the triangular discrimination, the symmetric chi-square divergence, the variational distance, and the Pearson divergence [1, 2, 5]) have interesting physical applications.

References

[1] Ali S M and Silvey S D 1966 *J. Roy. Stat. Soc. B* **28** 131–142

[2] Csiszár I 1967 *Studia Sci. Math. Hungar.* **2** 299–318

[3] Morimoto T 1963 *J. Phys. Soc. Jpn* **12** 328–331

[4] Johnson D H and Sinanović S 2001 Symmetrizing the Kullback-Leibler distance Tech. rep. Rice Univ., Houston, TX

[5] Taneja I J 2005 *Aust. J. Math. Anal. Appl.* **2** 8

[6] Gorban A N, Gorban P A and Judge G 2010 *Entropy* **12** 1145–1193

[7] Qian H 2001 *Phys. Rev. E* **63** 042103

[8] Beck C and Schlögl F 1995 *Thermodynamics of chaotic systems* (Cambridge University Press)

[9] Baez J C 2011 Rényi entropy and free energy arXiv:1102.2098

[10] Tsallis C 1988 *J. Stat. Phys.* **52** 479–487

[11] Crooks G E 1998 *J. Stat. Phys.* **90** 1481–1487

[12] Maes C and Netŏcný K 2003 *J. Stat. Phys.* **110** 269–310

[13] Seifert U 2005 *Phys. Rev. Lett.* **95** 040602

[14] Kawai R, Parrondo J M R and Van den Broeck C 2007 *Phys. Rev. Lett.* **98** 080602

[15] Andrieux D, Gaspard P, Ciliberto S, Garnier N, Joubaud S and Petrosyan A 2008 *J. Stat. Mech.: Theor. Exp.* P01002

[16] Jarzynski C 1997 *Phys. Rev. Lett.* **78** 2690–2693

[17] Peliti L 2008 *J. Stat. Mech.: Theor. Exp.* P05002

[18] Liphardt J, Dumont S, Smith S B, Tinoco Jr I and Bustamante C 2002 *Science* **296** 1832–1835

[19] Collin D, Ritort F, Jarzynski C, Smith S B, Tinoco Jr I and Bustamante C 2005 *Nature* **437** 231–234

[20] Ritort F 2006 *J. Phys.: Condens. Matter* **18** R531–R583

[21] Kullback S and Leibler R A 1951 *Ann. Math. Statist.* **22** 79–86

[22] Cover T M and Thomas J A 2006 *Elements of Information Theory* 2nd ed (New York: Wiley)

[23] Gaspard P 2004 *J. Chem. Phys.* **120** 8898–8905

[24] Jarzynski C 2006 *Phys. Rev. E* **73** 046105

[25] Gomez-Marin A, Parrondo J M R and Van den Broeck C 2008 *Europhys. Lett.* **82** 50002

[26] Feng E H and Crooks G E 2008 *Phys. Rev. Lett.* **101** 090602

[27] Horowitz J and Jarzynski C 2009 *Phys. Rev. E* **79** 021106

[28] Parrondo J M R, Van den Broeck C and Kawai R 2009 *New J. Phys.* **11** 073008

[29] Jeffreys H 1948 *Theory of probability* 2nd ed (Oxford: Clarendon Press)

[30] Lin J 1991 *IEEE Trans. Inf. Theory* **37** 145–151

[31] Endres D M and Schindelin J E 2003 *IEEE Trans. Inf. Theory* **49** 1858–1860

[32] Österreicher F and Vajda I 2003 *Ann. Inst. Statist. Math.* **55** 639–653

[33] Chernoff H 1952 *Ann. Math. Statist.* **23** 493–507

[34] Kakizawa Y, Shumway R H and Taniguchi M 1998 *J. Amer. Statist. Assoc.* **93** 328–340

[35] Lebowitz J L and Spohn H 1999 *J. Stat. Phys.* **95** 333–365

[36] Hummer G 2001 *J. Chem. Phys.* **114** 7330–7337

[37] Hummer G and Szabo A 1996 *J. Chem. Phys.* **105** 2004–2010

[38] Bhattacharyya A 1943 *Bull. Calcutta Math. Soc.* **35** 99–109

[39] Rényi A 1960 On measures of entropy and information *Proceedings of the 4th Berkeley symposium on mathematics, statistics and probability* pp 547–561

[40] Cressie N and Read T R C 1984 *J. Roy. Statist. Soc. B* **46** 440–464

[41] Mazonka O and Jarzynski C 1999 Exactly solvable model illustrating far-from-equilibrium predictions cond-mat/991212