On the fluctuation relation

The 1990s have witnessed remarkable progress in the domain of non-equilibrium statistical mechanics. One of the most celebrated results is the so-called Jarzynski relation, that will be derived below. Some non-equilibrium averages can consequently be now expressed in terms of equalities, that have been verified experimentally in a wealth of small systems where fluctuations are important. The approach presented here complements the derivations provided in class for stochastic processes: here, we start from a first principle standpoint, with deterministic systems obeying Hamiltonian dynamics.

◇

We consider a classical Hamiltonian system, initially in thermal equilibrium at temperature T. One of our goals here is to show that the free energy difference between two states of the system can be obtained from *non equilibrium averages*, taken over a large number of realizations of the same experiment, where an external parameter λ is modified (λ can be, for instance, the system volume, or any other quantity that can be controlled by the experimentalist). Between the initial time $t = 0$ and the final time t_f, λ changes from $\lambda = 0$ (the system is then in the macroscopic state A_0) to $\lambda = 1$ (the system is then in the macroscopic state A_1), following a temporal evolution $\lambda(t)$ that is given once and for all. During that operation, the work received by the system reads

$$W = \mathcal{H}_1(\Gamma(t_f)) - \mathcal{H}_0(\Gamma(0)) \tag{1}$$

where $\Gamma(t)$ denotes the point in phase space where the system sits at time t, and $\mathcal{H}_\lambda$ is the Hamiltonian, parameterized by λ. The partition function associated to $\mathcal{H}_\lambda$ at temperature T is denoted Z_λ. From a realization of the experiment to the next (in all realizations, the system is brought from state A_0 to state A_1), the work W fluctuates, e.g. because initial conditions correspond to different starting points $\Gamma(0)$ in phase space. We denote $\overline{(\ldots)}$ average values taken over all possible realizations of the experimental protocol, that is encoded in the function $\lambda(t)$ (it is not necessary to specify its time dependence). In the following, we suppose that at $t = 0$, the system is decoupled from the thermostat and subsequently evolves following Hamilton equations of motion.

1) Jarzynski equality

a) We have

$$\overline{e^{-\beta W}} = \int e^{-\beta W} \rho(\Gamma(0))\, d\Gamma(0) \tag{2}$$

where $\beta^{-1} = kT$ and the integral runs over all phase space. What is the expression of the probability density $\rho(\Gamma(0))$?

b) In expression (2), the work W only depends on $\Gamma(0)$. Using definition (1), and invoking some properties of Hamiltonian systems, show that

$$\boxed{\overline{e^{-\beta W}} = e^{-\beta \Delta F}}. \tag{3}$$

where ΔF is the free energy difference $F(A_1) - F(A_0)$. It is important to note that when the protocol stops at $t = t_f$, the system is generically out of equilibrium. Yet, the right hand side of the previous equality involves an equilibrium object only. The central identity (3) bears the name of Jarzynski, and we shall admit that it remains valid when the system remains in contact with the thermostat during the experiment.

c) Obtain from (3) an inequality between $\overline{W}$ and ΔF. It should be familiar...

2) Limiting cases

a) In the reversible limit (where $t_f \to \infty$ and the system can be considered at thermal equilibrium at all times), what is the expression of $\overline{W}$? How can this result be compatible with equality (3) ?

b) We would like to explicitly check the above result. First express W as an integral of $\partial \mathcal{H}_\lambda / \partial \lambda$ and show that

$$\frac{\partial \log Z_\lambda}{\partial \lambda} = -\beta \left\langle \frac{\partial \mathcal{H}_\lambda}{\partial \lambda} \right\rangle_\lambda , \tag{4}$$

where the meaning of the average $\langle \ldots \rangle_\lambda$ should be specified. Conclude.

c) Conversely, in the limit where the transformation is instantaneous ($t_f \to 0^+$), (3) has a particular meaning. Under such circumstances, we indeed have

$$W = \mathcal{H}_1(\Gamma(0)) - \mathcal{H}_0(\Gamma(0)) \tag{5}$$

and it can be noted that the definition (2) yields $\overline{e^{-\beta W}} = \left\langle e^{-\beta W} \right\rangle_0$. Show that Jarzynski's equality is recovered.

3) Crooks relation

We denote $p(W)$ the probability density function of W, obtained from the ensemble of realizations of the experimental protocol. We then have

$$\overline{e^{-\beta W}} = \int e^{-\beta W} p(W)\, dW. \tag{6}$$

It is possible to derive a more general relation than Jarzynski's. It relates the probability density $p_F(W)$ holding for the "direct" experiment, to $p_B(W)$ characteristic of the "inverse" process. F – not be confused with the free energy – stands here for "forward", and B stands for "backward", in which case the system goes from state A_1 to A_0, where λ changes from 1 (at $t = 0$) to 0 (at $t = t_f$) following the law $\lambda(t_f - t)$.

a) On general grounds, it is possible to express the probability density function of a given random variable from an appropriate average of a Dirac distribution. Specifically, this gives here :

$$p_F(W) = \frac{1}{Z_0} \int e^{-\beta \mathcal{H}_0(\Gamma(0))}\, \delta\left[W - \mathcal{H}_1(\Gamma(t_f)) + \mathcal{H}_0(\Gamma(0))\right] d\Gamma(0). \tag{7}$$

In such conditions, how is $p_B(W)$ defined ?

b) Show that

$$\boxed{p_F(W)\, e^{-\beta W} = e^{-\beta \Delta F}\, p_B(-W)}. \tag{8}$$

We again assume here that during the transformation, the system is decoupled from the thermostat, and we admit that the results obtained still hold when the system remains in thermal contact with the heat bath. Eq. (8) is our second key result. It actually subsumes previous considerations, as we now prove.

c) Show that Jarzynski equality (3) can be readily recovered from (8).

d) Discuss briefly the reversible limit.

4) The case of Gaussian fluctuations - Fluctuation/dissipation relation

We are interested here only in the direct process ("forward"). In the limit where the transformation is sufficiently slow (without being necessarily reversible), $p(W)$ takes a Gaussian form. We denote $\overline{W}$ its mean, and σ the corresponding standard deviation.

a) Show that equation (3) implies that

$$\overline{W} = \Delta F + \beta\, \sigma^2 /2 \tag{9}$$

b) In what sense can the previous relation be coined "fluctuation-dissipation" ? Briefly discuss the reversible limiting case.

5) . . . where one measures the length of time's arrow . . .

Reminder : the Kullback-Leibler distance between two discrete probability distributions $\{p_i\}$ and $\{q_i\}$ reads :

$$D(p||q) = \sum_i p_i \log\left(\frac{p_i}{q_i}\right). \tag{10}$$

Generalize the above definition in the continuous case, and give the expression of the distance between the distributions $p_F(W)$ and $p_B(-W)$. Establish then a connection between the irreversibility on the one hand, and the distinguishability of forward and reverse protocols (as measured by their distances) on the other hand.

6) Application to single molecule experiments

Figure 1 represents work distributions as measured in experiments where an RNA strand in a folded configuration (forming an hairpin) is unfolded by a mechanical force applied on both ends (denoted 3' and 5' on figure 1). The reverse process (folding) is spontaneous (with mostly negative values of W). The results shown have been obtained using two different experimental protocols : one is fast (data at 20 $pN\, s^{-1}$, corresponding to the circles), and the other is slow (data at 7 $pN\, s^{-1}$, corresponding to the triangles).

a) Invoking Crooks relation, the free energy difference for unfolding can be directly read on figure 1. What is this difference ?

b) How can the offset between folding and unfolding curves be explained ? Are the data shown in figure 1 compatible with the body of results obtained earlier ?

References :

- *Nonequilibrium equality for free energy differences*, C. Jarzynski, Physical Review Letters **78**, 2690 (1997).
- *Entropy production fluctuation theorem and the non-equilibrium work relation for free energy differences*, G.E. Crooks, Physical Review E **60**, 2721 (1999).

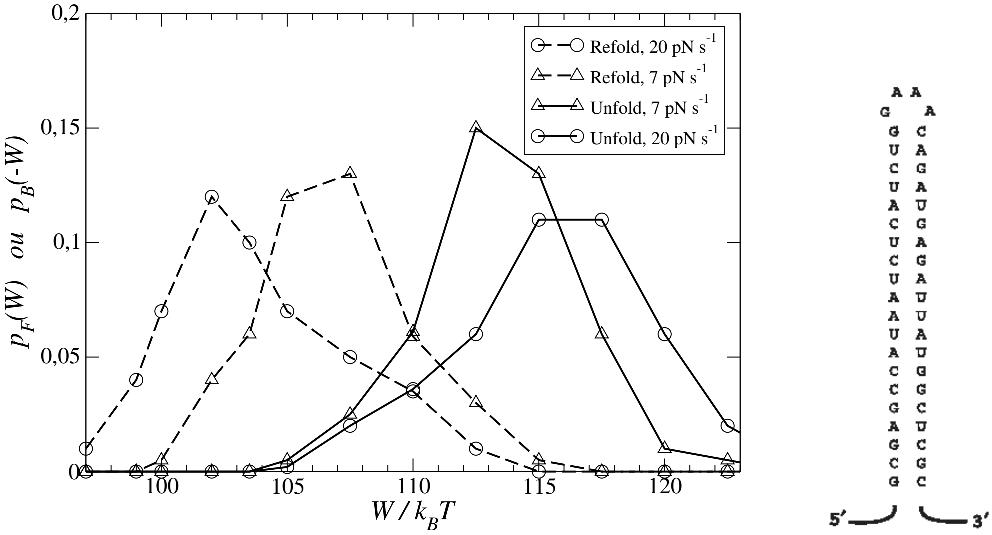

Figure 1: Left : plots of probability densities $p_F(W)$ (continuous curves) and $p_B(-W)$ (dashed curves) in an experiment of unfolding/refolding of an RNA strand. Right : the RNA strand in its hairpin configuration. Doctored from Collin *et al.*, Nature **437**, 231 (2005).