

Biophysics of Cells and Single Molecules: Problem set for weeks 36-37, 2023

Tomás Ricardo Basile Álvarez
316617194

September 29, 2023

Exercises week 1

Exercise 1

Consider a master with four states, $x \in \{1, 2, 3, 4\}$, and with rates $W_{xx'}$ ($\forall x \neq x'$), none of which vanishes. Show that if $W_{xx'} = W_{x'x}$, then the unique stationary distribution is $p_1 = p_2 = p_3 = p_4 = 1/4$. Show with an example that this is not necessarily the case if some of the rates vanishes.

The stationary distribution p_m is such that:

$$\sum_{m'} W_{m,m'} p_{m'} = 0.$$

We will first show that $p_1 = p_2 = p_3 = p_4 = 1/4$ is a solution to these equations. This is easy to show, since if we substitute those values of p on the left hand side, we get:

$$\begin{aligned} \sum_{m'} W_{m,m'} p_{m'} &= \frac{1}{4} \sum_{m'} W_{m,m'} \\ &= \frac{1}{4} \sum_{m'} W_{m'm}, \quad \text{because } W_{mm'} = W_{m'm}. \\ &= 0, \quad \text{because the columns of the matrix add to 0.} \end{aligned}$$

Therefore, we have proven that $p_m = 1/4$ is a stationary state. Now, we need to prove that it is the only stationary state.

To do it, imagine there is some other stationary state (p_1, p_2, p_3, p_4) in which not all the probabilities are $1/4$. Then, one of the probabilities will be bigger than all the others (let's say, without loss of generality that said probability is p_1). Then, we can write the equation for $m = 1$, which is:

$$\dot{p}_1 = W_{11}p_1 + W_{12}p_2 + W_{13}p_3 + W_{14}p_4.$$

Then, because of how we define the diagonal elements, we have that:

$$\dot{p}_1 = (-W_{12} - W_{13} - W_{14})p_1 + W_{12}p_2 + W_{13}p_3 + W_{14}p_4.$$

And because of the symmetry of the matrix $W_{mm'} = W_{m'm}$, we conclude:

$$\dot{p}_1 = W_{12}(p_2 - p_1) + W_{13}(p_3 - p_1) + W_{14}(p_4 - p_1).$$

However, given that p_1 is bigger than all the other probabilities, we have that $p_2 - p_1, p_3 - p_1$ and $p_4 - p_1$ are all negative. Furthermore, W_{12}, W_{13}, W_{14} are positive values (and non zero). This implies that the expression we got for \dot{p}_1 is negative, and it is not equal to 0, meaning that the distribution is not stationary. Therefore, we have concluded that if not all the p_i are equal to $1/4$, then the state is not stationary, so that the only stationary distribution is $(1/4, 1/4, 1/4, 1/4)$.

For the second part of the question, we can take for example the following matrix in which some of the terms are zero:

$$W = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & W_{22} & W_{23} & W_{24} \\ 0 & W_{32} & W_{33} & W_{34} \\ 0 & W_{42} & W_{43} & W_{44} \end{pmatrix}$$

Then, in this case, not only is $(1/4, 1/4, 1/4, 1/4)$ a stationary state for the same reason as before, but now $(0, 1/3, 1/3, 1/3)$ is stationary too, since it would lead to the equations:

$$\begin{aligned} \dot{p}_1 &= 0 \\ \dot{p}_2 &= \frac{1}{3} (W_{22} + W_{23} + W_{24}) \\ \dot{p}_3 &= \frac{1}{3} (W_{32} + W_{33} + W_{34}) \\ \dot{p}_4 &= \frac{1}{3} (W_{42} + W_{43} + W_{44}) \end{aligned}$$

And because of the symmetry $W_{mm'} = W_{m'm}$, we have that:

$$\begin{aligned} \dot{p}_1 &= 0 \\ \dot{p}_2 &= \frac{1}{3} (W_{22} + W_{32} + W_{42}) \\ \dot{p}_3 &= \frac{1}{3} (W_{23} + W_{33} + W_{43}) \\ \dot{p}_4 &= \frac{1}{3} (W_{24} + W_{34} + W_{44}) \end{aligned}$$

Finally, since all the columns of the matrix add up to 0, we conclude that $\dot{p}_1 = \dot{p}_2 = \dot{p}_3 = \dot{p}_4 = 0$, so that $(0, 1/3, 1/3, 1/3)$ is indeed a stationary state.

Exercise 2

Consider a system with states $x \in \{1, 2, 3\}$ and energies $\epsilon_1, \epsilon_2, \epsilon_3$ in an environment with temperature T . Calculate the energy, entropy and free energy associated with the equilibrium state of the system. Show that the equilibrium distribution minimizes the free energy.

Since the state is in equilibrium, it follows the Boltzmann distribution, so that the probability of being in state x is proportional to $e^{-\beta\epsilon_x}$, with $\beta = 1/k_B T$. To calculate what we are being asked for, we can begin by writing the partition function $Z = \sum_i e^{-\beta\epsilon_i}$:

$$Z = e^{-\beta\epsilon_1} + e^{-\beta\epsilon_2} + e^{-\beta\epsilon_3}.$$

The energy is then equal to $-\frac{\partial}{\partial\beta} \ln Z$, so:

$$\begin{aligned} E &= -\frac{1}{Z} \frac{\partial}{\partial\beta} Z \\ &= \boxed{\frac{1}{Z} (\epsilon_1 e^{-\beta\epsilon_1} + \epsilon_2 e^{-\beta\epsilon_2} + \epsilon_3 e^{-\beta\epsilon_3})}. \end{aligned}$$

Which of course makes a lot of sense, since this result is just the sum of the energies ϵ_i , each multiplied by its probability $p_i = e^{-\beta\epsilon_i}/Z$.

For the entropy, we can use that $S = -k_B \sum_i p_i \ln p_i$, with the probabilities given by the Boltzmann distribution $p_i = e^{-\beta\epsilon_i}/Z$:

$$\begin{aligned} S &= -k_B \sum_i \frac{e^{-\beta\epsilon_i}}{Z} \ln \left(\frac{e^{-\beta\epsilon_i}}{Z} \right) \\ &= -k_B \sum_i \frac{e^{-\beta\epsilon_i}}{Z} [\ln(e^{-\beta\epsilon_i}) - \ln Z] \\ &= -k_B \sum_i \frac{e^{-\beta\epsilon_i}}{Z} (-\beta\epsilon_i - \ln Z) \\ &= k_B \beta \sum_i \frac{e^{-\beta\epsilon_i}}{Z} \epsilon_i + k_B \frac{\ln Z}{Z} \sum_i e^{-\beta\epsilon_i} \end{aligned}$$

We can see that the first sum is just the energy $E = \sum_i \epsilon_i p_i$, while the second one is the definition of Z . Therefore:

$$\begin{aligned} S &= k_B \beta E + k_B \frac{\ln Z}{Z} Z \\ &= \boxed{\frac{1}{T} E + k_B \ln Z}. \end{aligned}$$

If we want the full expression, we can substitute what we got for $\langle E \rangle$ and the definition of Z :

$$S = \frac{1}{T} \frac{\epsilon_1 e^{-\beta\epsilon_1} + \epsilon_2 e^{-\beta\epsilon_2} + \epsilon_3 e^{-\beta\epsilon_3}}{e^{-\beta\epsilon_1} + e^{-\beta\epsilon_2} + e^{-\beta\epsilon_3}} + k_B \ln (e^{-\beta\epsilon_1} + e^{-\beta\epsilon_2} + e^{-\beta\epsilon_3}).$$

Finally, for the free energy, we can use the definition $F = E - TS$, and use that we found $S = E/T + k_B \ln Z$. Therefore, we find that:

$$\begin{aligned} F &= E - T \left(\frac{1}{T} E + k_B \ln Z \right) \\ &= E - E - k_B T \ln Z \\ &= \boxed{-k_B T \ln Z} \\ &= -k_B T \ln (e^{-\beta\epsilon_1} + e^{-\beta\epsilon_2} + e^{-\beta\epsilon_3}) \end{aligned}$$

Now, for the second part of the exercise, we need to prove that the equilibrium distribution (the boltzmann distribution) minimizes the free energy.

We can begin by writing the free energy $F = E - TS$ using that $E = \sum_i \epsilon_i p_i$ and $S = -k_B \sum_i p_i \ln p_i$ as follows:

$$F = \sum_i \epsilon_i p_i + k_B T \sum_i p_i \ln p_i$$

Then, we wish to find the minimum of this function with respect to the probabilities p_i . However, we also have the additional condition that $\sum p_i = 1$. Therefore, we can find the minimum given the condition by using a Lagrange multiplier λ , writing the Lagrangian function $F - \lambda(\sum_i p_i - 1)$ and setting its partial derivative with respect to every p_j equal to 0:

$$\begin{aligned} \frac{\partial}{\partial p_j} \left(\sum_i \epsilon_i p_i + k_B T \sum_i p_i \ln p_i - \lambda (\sum_i p_i - 1) \right) &= 0 \\ \Rightarrow \frac{\partial}{\partial p_j} (\epsilon_j p_j) + \frac{\partial}{\partial p_j} (p_j \ln p_j) - \lambda \frac{\partial}{\partial p_j} (p_j - 1) &= 0 \\ \Rightarrow \epsilon_j + k_B T (\ln p_j + 1) - \lambda &= 0 \end{aligned}$$

Then, we can isolate p_j from this equation:

$$\begin{aligned} k_B T \ln p_j &= \lambda - \epsilon_j - k_B T \\ \Rightarrow \boxed{p_j} &= e^{-\epsilon_j/k_B T} e^{(\lambda - k_B T)/k_B T} \end{aligned}$$

Now we can use the restriction that $\sum p_j = 1$ to find λ :

$$\begin{aligned}\sum e^{-\epsilon_j/k_B T} e^{(\lambda - k_B T)/k_B T} &= 1 \\ \Rightarrow e^{(\lambda - k_B T)/k_B T} \sum e^{-\epsilon_j/k_B T} &= 1 \\ \Rightarrow e^{(\lambda - k_B T)/k_B T} &= \frac{1}{\sum_j e^{-\epsilon_j/k_B T}} = \frac{1}{Z}\end{aligned}$$

With that equation we could get λ , but it is simpler to just substitute this exponential into the equation we got for p_j :

$$\boxed{p_j = e^{-\epsilon_j/k_B T} \frac{1}{Z}}$$

Therefore, we have recovered the Boltzmann distribution.

Exercise 3

Consider a system with two states $x = 0$ and $x = 1$ with energies initially set to $\epsilon_0 = 0$ and $\epsilon_1 = k_B T$. The system is initially in equilibrium with an environment at temperature T . At time $t = 0$, the energy of ϵ_1 is then quenched to $\epsilon_2 = 2k_B T$. Compute the average total amount of heat, work and entropy production.

Before the quenching, the partition function is:

$$\begin{aligned} Z &= e^{-\beta\epsilon_0} + e^{-\beta\epsilon_1} \\ &= e^0 + e^{-\beta \cdot k_B T} \\ &= 1 + e^{-1}. \end{aligned}$$

Since we are in equilibrium, the equilibrium probabilities follow the Boltzmann distribution: $p_0 = e^{-\beta\epsilon_0}/Z = e^0/Z = 1/Z$ and $p_1 = e^{-\beta\epsilon_1}/Z = e^{-\beta(k_B T)}/Z = e^{-1}/Z$. That is, the probabilities are:

$$p_0 = \frac{1}{1 + e^{-1}} = \frac{e}{1 + e}, \quad p_1 = \frac{e^{-1}}{1 + e^{-1}} = \frac{1}{1 + e}.$$

On the other hand, after the quenching (and after the system relaxes to equilibrium), the partition function becomes $Z' = 1 + e^{-2}$ and the probabilities are now:

$$p'_0 = \frac{e^0}{Z'} = \frac{1}{1 + e^{-2}} = \frac{e^2}{1 + e^2}, \quad p'_1 = \frac{e^{-2}}{1 + e^{-2}} = \frac{1}{1 + e^2}$$

Now we can start by calculating the average amount of work. To do it, we may use the equation $\dot{W} = \sum \dot{\epsilon}_m p_m$. When the quenching happens, we know that the system is in the state with probabilities $p_0 = e/(1 + e)$ and $p_1 = 1/(1 + e)$, so those are the probabilities we should consider for this equation. Therefore, we have:

$$\begin{aligned} \dot{W} &= \sum \dot{\epsilon}_m p_m \\ &= d\epsilon_0 p_0 + d\epsilon_1 p_1 \end{aligned}$$

and since the energy of state zero does not change and that of state 1 changes by $k_B T$, we have that:

$$\dot{W} = k_B T p_1 = k_B T \frac{1}{1 + e} = 0.269 k_B T$$

To find the heat, we can first find the change in energy and then use the first law of thermodynamics. The change in energy is just the final energy $\epsilon_0 p'_0 + \epsilon_2 p'_1$ minus the initial one

$\epsilon_0 p_0 + \epsilon_1 p_1$, therefore:

$$\begin{aligned}
\dot{E} &= (\epsilon_0 p'_0 + \epsilon_2 p'_1) - (\epsilon_0 p_0 + \epsilon_1 p_1) \\
&= 2k_B T p'_1 - k_B T p_1 \\
&= 2k_B T \frac{1}{1+e^2} - k_B T \frac{1}{1+e}. \\
&= -0.031 k_B T
\end{aligned}$$

From there, we can use that $\dot{E} = \dot{W} + \dot{Q}$ and combine it with our result for the work done:

$$\begin{aligned}
\dot{Q} &= \dot{E} - \dot{W} \\
&= -0.3 k_B T.
\end{aligned}$$

Finally, for the entropy production, we may use the expression for entropy $S_{int} = -k_B \sum_i p_i \ln p_i$ before and after quenching:

$$\begin{aligned}
\dot{S}_{int} &= -k_B \sum_i p'_i \ln p'_i - (-k_B \sum_i p_i \ln p_i) \\
&= k_B \left(-\frac{1}{1+e^2} \ln \left(\frac{1}{1+e^2} \right) - \frac{e^2}{1+e^2} \ln \left(\frac{e^2}{1+e^2} \right) + \frac{1}{1+e} \ln \left(\frac{1}{1+e} \right) + \frac{e}{1+e} \ln \left(\frac{e}{1+e} \right) \right) \\
&= -0.217 k_B
\end{aligned}$$

That is the internal entropy production. If we want the complete entropy production, we need to add the entropy exchanged with the reservoir, which is $Q/T = -0.3 k_B T$. Therefore, we have a total of $\dot{S}_{int} + \dot{S}_{res} = -0.517 k_B$.

Exercise 4

Consider a closed system where there is an exchange of energy between the system and a surrounding reservoir. Show that the second law of thermodynamics for the total system ($\Delta_i S = \Delta S_{\text{system}} + \Delta S_{\text{reservoir}} \geq 0$) implies that $\Delta G_{\text{system}} \leq 0$. Use the first law of thermodynamics $Q = T\Delta S_{\text{reservoir}}$, and $W = p\Delta V$, where p is the pressure and V is volume. The following definition may also be useful $G = E + pV - TS$.

From the first law applied to the reservoir, we have that:

$$\Delta E_r = Q_r + W_r.$$

Then, we can use that $Q_r = T\Delta S_r$ and that $W = -p\Delta V_r$ to get:

$$\Delta E_r = T\Delta S_r - p\Delta V_r.$$

From that equation we can get ΔS_r :

$$\Delta S_r = \frac{1}{T}\Delta E_r + \frac{p}{T}\Delta V_r.$$

Therefore, we now have an expression for ΔS_r , which we will use in the second law. However, we would like to have this expression in terms of the variables of the system, and not on those of the reservoir. To do it, we may use the fact that the system is only exchanging energy with the reservoir, so that by conservation of energy, $\Delta E_r = -\Delta E_s$. Furthermore, the change in volume of the reservoir corresponds to a change in volume of the system, so that $\Delta V_r = -\Delta V_s$. Substituting these two expressions into the equation we had for ΔS_r , we obtain that:

$$\Delta S_r = -\frac{1}{T}\Delta E_s - \frac{p}{T}\Delta V_s$$

Now we can make use of the second law, which states $\Delta_i S = \Delta S_s + \Delta S_r \geq 0$, and substitute our expression for ΔS_r .

$$\Delta_i S = \Delta S_s - \frac{1}{T}\Delta E_s - \frac{p}{T}\Delta V_s \geq 0.$$

Now we multiply both sides by $-T$:

$$-T\Delta S_s + \Delta E_s + p\Delta V_s \leq 0.$$

And we are done, because this last expression is ΔG_s . This can be seen from the definition of $G_s = E_s + pV_s - TS_s$, and since the pressure and temperature are held constant by the reservoir, we have that $\Delta G_s = \Delta E_s + p\Delta V_s - T\Delta S_s$. So we have proven that $\Delta G_s \leq 0$.

1 Exercises Week 2

Suppose an experiment was carried out in which an RNA hairpin was unfolded and refolded using the same force loading rate, i.e., the system was switched between two equilibrium states. The following sets of unfolding and refolding total works (given in numerical values) were measured:

UNFOLDING	Work	REFOLDING	Work
Counts	Value [$k_B T$]	Counts	Value [$k_B T$]
3	8.5	6	7.0
6	9.0	11	7.5
11	9.5	14	8.0
16	10.0	10	8.5
9	10.5	5	9.0
2	11.0	1	9.5

1. Use Crook's theorem to find free energy ΔF , of the unfolding process. Explain how you do, preferably illustrating your method with a plot of the unfolding and refolding probability distributions $P_U(W)$ and $P_R(W)$.

Given the data, I plotted the probability distributions for the folding and unfolding processes in the following figure (the probability is just the number of times a given work was measure divided by the total amount of data points).

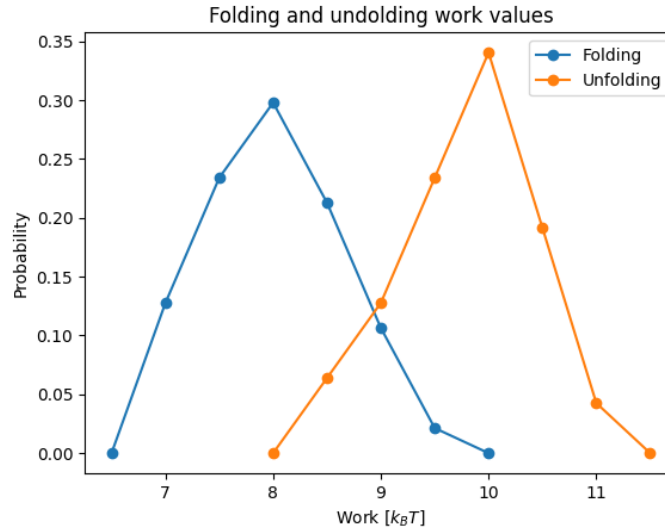


Figure 1: Probability distributions for the work done on the folding and unfolding processes.

To get a value of the free energy difference, we can start from Crook's theorem

$$\frac{P(w)}{\tilde{P}(-w)} = \exp(\beta(w - \Delta F^{eq}))$$

and take the logarithm of both sides:

$$\ln \left(\frac{P(w)}{\tilde{P}(-w)} \right) = \beta(w - \Delta F^{eq}).$$

Therefore, if we plot $\ln \left(\frac{P(w)}{\tilde{P}(-w)} \right)$ as a function of w , we should get a line with slope β and intercept $\beta\Delta F^{eq}$. Because there are only three data points that had both an unfolding and refolding work value, we can only plot these three points. The plot of the three points together with the line that best fits them is shown in the following figure.

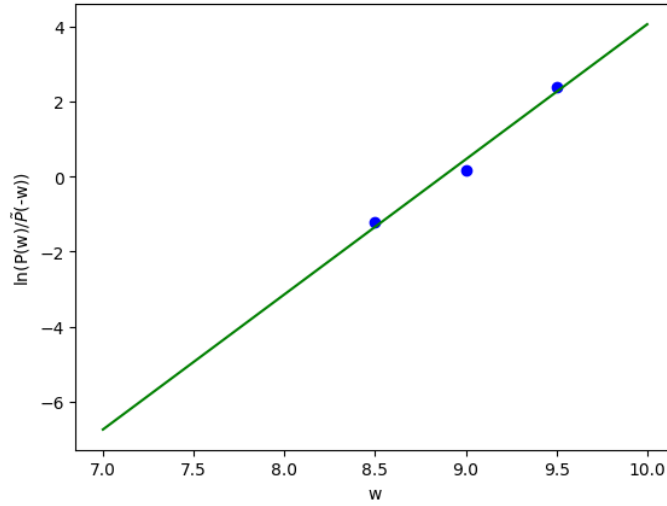


Figure 2: Plot of $\ln \left(\frac{P(w)}{\tilde{P}(-w)} \right)$ vs w and the best fit line for the three points in which we have data for both folding and unfolding data.

The best fit line was found using linear regression with the Python package “sklearn” and the result of the intercept is -8.873 . Therefore, the free energy difference found is

$$\Delta F^{eq} = -8.873k_B T.$$

2. Calculate the Gibbs free energy, ΔF , of the unfolding process using Jarzynski’s equality on the dataset in the table above.

According to Jarzynski’s equality, we have that:

$$\langle \exp(-\beta w) \rangle = \exp(-\beta \Delta F^{eq}).$$

Taking the logarithm of both sides, we obtain:

$$\Delta F^{eq} = -k_B T \ln \langle \exp(-\beta w) \rangle.$$

We can directly calculate this expected value using the given data. We can do it for the folding direction and for the unfolding direction and the results are:

$$\begin{aligned}\Delta F_F^{eq} &= 7.81k_B T \\ \Delta F_U^{eq} &= 9.60k_B T.\end{aligned}$$

To approximate the difference in free energy, we can average both directions, so that the result is:

$$\Delta F^{eq} = \frac{9.6k_B T + 7.81k_B T}{2} = \boxed{8.705k_B T}.$$

3. Does Jarzynski's equality return the same value for ΔF as Crook's theorem? Why or why not is this so?

In theory, both methods should give the same result, however, we can see that the actually give different values.

One reason for this could be that to use Crook's theorem, we could only consider the few data points for which we had information about both the folding and unfolding directions. However, for Jarzynski's equality, we can use all the data points for the folding direction and then use all data points for the unfolding direction.

More generally, we have to keep in mind that the theorems speak about statistical results, averages and probabilities. However, the measurements have some variance because of small fluctuations from one repetition of the experiment to the next. Therefore, the values for the probabilities have some variance and the way we use said probabilities in both methods is different, so that it should be expected that these variances act differently on both methods and give overall different results.