

Name: **Huan Q. Bui**
 Course: **8.333 - Statistical Mechanics I**
 Problem set: **#2**

1. Random deposition.

- (a) Consider a site. Assume that the gold atoms arrive at this site over time via a Poisson process. The deposition rate is d layers per second, so this Poisson process has rate d . Over time t , the average number of deposition at a site is dt . With this, we have

$$\Pr(m \text{ atoms in time } t) = \frac{(dt)^m e^{-dt}}{m!}$$

Glass is not covered if there is no deposition, i.e., $m = 0$. The fraction of the glass not covered by the atoms is the probability of zero deposition:

$$\Pr(0 \text{ atoms in time } t) = e^{-dt}$$

We see that the fraction of the glass not covered decreases exponentially in time.

- (b) From Part (a), we know that the average thickness is $\langle x \rangle = dt$. To find the variance we need to compute the second moment:

$$\langle x^2 \rangle = \sum_{i=0}^{\infty} x^2 \frac{(dt)^x e^{-dt}}{x!} = dt + d^2 t^2.$$

Therefore, the variance in thickness is

$$\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2 = dt$$

2. Semi-flexible polymer in two dimensions.

- (a) It's nicer to work with the ϕ -dependent \mathcal{H} , so let us write $\mathbf{t}_m \cdot \mathbf{t}_n$ in terms of angles:

$$\mathbf{t}_m \cdot \mathbf{t}_n = a^2 \cos(\theta_m + \theta_{m+1} + \cdots + \theta_{n-1}).$$

The summed form for the angles is not very convenient to work with, as there is no clear way to find $\langle \mathbf{t}_m \cdot \mathbf{t}_n \rangle$ written in this form. Instead, let us find $a^2 \langle \exp(i(\theta_m + \theta_{m+1} + \cdots + \theta_{n-1})) \rangle$ and then take the real part. This, written in this form, is still cumbersome. However, we may assume that the angles ϕ_i are independent, and therefore

$$\langle \mathbf{t}_m \cdot \mathbf{t}_n \rangle = \text{Re} \left[a^2 \langle \exp(i(\theta_m + \theta_{m+1} + \cdots + \theta_{n-1})) \rangle \right] = \text{Re} \left[a^2 \prod_{j=m}^{n-1} \langle e^{i\phi_j} \rangle \right] = a^2 \prod_{j=m}^{n-1} \langle \cos \phi_j \rangle.$$

Moreover, since the angles ϕ_i 's are independent, the probability for each configuration is simply the product of the individual probabilities:

$$\Pr(\phi_1, \dots, \phi_{N-1}) = \exp \left[\frac{a^2 \kappa}{k_B T} \sum_{i=1}^{N-1} \cos \phi_i \right] = \prod_{i=1}^{N-1} \exp \left[\frac{a^2 \kappa}{k_B T} \cos \phi_i \right].$$

And so we may write

$$\Pr(\phi_i) = \exp \left[\frac{a^2 \kappa}{k_B T} \cos \phi_i \right]$$

With this we have

$$\langle \mathbf{t}_m \cdot \mathbf{t}_n \rangle = a^2 \prod_{j=m}^{n-1} \frac{\int d\phi \cos \phi \exp \left[\frac{a^2 \kappa}{k_B T} \cos \phi_j \right]}{\int d\phi \exp \left[\frac{a^2 \kappa}{k_B T} \cos \phi_j \right]} = a^2 \left\{ \frac{\int d\phi \cos \phi \exp \left[\frac{a^2 \kappa}{k_B T} \cos \phi \right]}{\int d\phi \exp \left[\frac{a^2 \kappa}{k_B T} \cos \phi \right]} \right\}^{|n-m|}.$$

So $\langle \mathbf{t}_m \cdot \mathbf{t}_n \rangle$ has the form $a^2 [f(T)]^{|n-m|}$ where $f(T)$ is the fraction in the curly brackets. We may write $\langle \mathbf{t}_m \cdot \mathbf{t}_n \rangle$ as an exponential:

$$\langle \mathbf{t}_m \cdot \mathbf{t}_n \rangle = a^2 \exp \left[|n-m| \ln f(T) \right] = a^2 \exp \left[\frac{|n-m|}{1/\ln f(T)} \right] \equiv a^2 \exp \left[\frac{-|n-m|}{\xi} \right],$$

as desired. The persistence length is thus

$$l_p = a\xi = \frac{a}{-\ln f(T)} = \frac{a}{\ln \left[\frac{\int d\phi \exp \left[\frac{a^2 \kappa}{k_B T} \cos \phi \right]}{\int d\phi \cos \phi \exp \left[\frac{a^2 \kappa}{k_B T} \cos \phi \right]} \right]}$$

(b) By definition, we have

$$\mathbf{R} = \sum_{i=1}^N \mathbf{t}_i \implies \langle R^2 \rangle = \langle \mathbf{R} \cdot \mathbf{R} \rangle = \sum_{m,n=1}^N \langle \mathbf{t}_m \cdot \mathbf{t}_n \rangle = \sum_{m,n=1}^N a^2 \exp \left[\frac{-|n-m|}{\xi} \right].$$

Now we consider what happens when $N \rightarrow \infty$. We see that \mathbf{R} has the form

$$\langle R^2 \rangle = a^2 \left[N + N_1 e^{-1/\xi} + N_2 e^{-2/\xi} + N_3 e^{-3/\xi} + \dots \right]$$

where N_1, N_2, N_3, \dots are natural numbers. In the limit $N \rightarrow \infty$, we have $N_j \approx 2N$ for small j 's (swapping n, m gives an extra factor of 2), and N_j 's for large j 's don't really matter because of the exponential decay $e^{-j/\xi}$. So, we may very well write this as

$$\langle R^2 \rangle \approx a^2 \left[N + 2N \left(e^{-1/\xi} + e^{-2/\xi} + e^{-3/\xi} + \dots \right) \right]$$

To see this, make a table with n 's on the columns and m 's on the rows respectively. The content of the table will be $|n-m|$. Obviously, the diagonal contains all N zeros since that's where $n=m$. The adjacent off-diagonals contain terms with $|n-m|=1$, followed by $|n-m|=2$, and so on. Since there are two such off-diagonals for each value of $|n-m|$, the multiplicity for each of the terms is $2N$. When we are sufficiently far off from the diagonal, the exponential decay tells us that we don't have to worry about them.

In any case, we now recall the telescoping sum formula

$$\frac{x}{1-x} = x + x^2 + x^3 + \dots$$

which gives us a more compact formula for the asymptotic behavior of $\langle R^2 \rangle$:

$$\langle R^2 \rangle = a^2 N \left(1 + 2 \frac{e^{-1/\xi}}{1 - e^{-1/\xi}} \right), \quad N \rightarrow \infty$$

(c) \mathbf{R} is a sum of iid's \mathbf{t}_i . In view of the central limit theorem, $p(\mathbf{R})$ is a Gaussian. To determine the form of $p(\mathbf{R})$, we must find the first and second moments. Since each \mathbf{t}_i is random, we can conclude that $\langle \mathbf{R} \rangle = 0$. The second moment is given by Part (b), and so the variance of this distribution is

$\sigma^2 = \langle R^2 \rangle - 0 = \langle R^2 \rangle$ which is what we found in Part (b). To find the normalization constant, we look at the covariance matrix C . Its determinant $|\det(C)|$ will be the product of $\langle R_x^2 \rangle$ and $\langle R_y^2 \rangle$, each of which is $\langle R^2 \rangle/2$ (by symmetry, and the fact that variances of independent variables add). With these,

$$p(\mathbf{R}) = \frac{1}{\sqrt{(2\pi)^2 |\det(C)|}} \exp\left(-\frac{\mathbf{R}^\top C^{-1} \mathbf{R}}{2}\right) = \boxed{\frac{1}{\pi \langle R^2 \rangle} \exp\left(-\frac{\mathbf{R} \cdot \mathbf{R}}{\langle R^2 \rangle}\right)}$$

where we have used the fact that $C = \langle R^2 \rangle \mathbb{I}/2$.

(d) We shall “formally” consider the modified probability weight:

$$\exp(-\mathcal{H}/k_B T) \rightarrow \exp(\mathbf{F} \cdot \mathbf{R}/k_B T) \exp(-\mathcal{H}/k_B T).$$

Taking the average of \mathbf{R} under these new weights yields

$$\langle \mathbf{R} \rangle = \frac{\int \mathbf{R} \exp(\mathbf{F} \cdot \mathbf{R}/k_B T) \exp(-\mathcal{H}/k_B T)}{\int \exp(\mathbf{F} \cdot \mathbf{R}/k_B T) \exp(-\mathcal{H}/k_B T)}.$$

We may treat $\mathbf{R} \exp(\mathbf{F} \cdot \mathbf{R}/k_B T)$ and $\exp(\mathbf{F} \cdot \mathbf{R}/k_B T)$ as input functions whose averages we wish to find and write

$$\langle R \rangle = \frac{\langle \mathbf{R} \exp(\mathbf{F} \cdot \mathbf{R}/k_B T) \rangle'}{\langle \exp(\mathbf{F} \cdot \mathbf{R}/k_B T) \rangle'}$$

where the new average $\langle \cdot \rangle'$ are essentially averages for when $\mathbf{F} = 0$. The denominator becomes unity, while the numerator can be expanded as

$$\langle \mathbf{R}_i \exp(\mathbf{F} \cdot \mathbf{R}/k_B T) \rangle' \approx \langle \mathbf{R}_i \rangle' + \frac{\langle \mathbf{R}_i \mathbf{F} \cdot \mathbf{R} \rangle'}{k_B T} + \frac{\langle \mathbf{R}_i (\mathbf{F} \cdot \mathbf{R})^2 \rangle'}{2k_B T} + \frac{\langle \mathbf{R}_i (\mathbf{F} \cdot \mathbf{R})^3 \rangle'}{6k_B T} + \dots$$

where \mathbf{R}_i 's are the components of \mathbf{R} (i.e., i is x and y), and that $\langle R \rangle$ at $\mathbf{F} = 0$ is simply $\langle R^2 \rangle$ which we already know. Moreover, terms with odd-powered \mathbf{R} 's will vanish by symmetry. We are thus interested in the term

$$\langle \mathbf{R}_i \mathbf{F} \cdot \mathbf{R} \rangle' = \langle \mathbf{R}_i \mathbf{F}_j \mathbf{R}_j \rangle' = \mathbf{F}_j \langle \mathbf{R}_i \mathbf{R}_j \rangle' = \mathbf{F}_j \delta_{ij} \langle R^2 \rangle / 2 = \mathbf{F}_i \langle R^2 \rangle / 2.$$

With this we have

$$\boxed{\langle \mathbf{R} \rangle = \frac{\langle R^2 \rangle}{2k_B T} \mathbf{F} + \mathcal{O}(F^3) = K^{-1} \mathbf{F} + \mathcal{O}(F^3), \quad K = \frac{2k_B T}{\langle R^2 \rangle}}$$

3. Foraging. We have

$$p(r|t) = \frac{r}{2Dt} \exp\left(-\frac{r^2}{4Dt}\right) \quad \text{and} \quad p(t) \propto \exp\left(-\frac{t}{\tau}\right).$$

Normalizing $p(t)$ give the leading factor equal $1/\tau$. So, we can write

$$p(t) = \frac{1}{\tau} \exp\left(-\frac{t}{\tau}\right).$$

The (unconditional) probability of finding the searcher at a distance r from the nest is

$$p(r) = \int_0^\infty p(r|t) p(t) dt = \int_0^\infty \frac{r}{2D\tau t} \exp\left(-\frac{r^2}{4Dt} - \frac{t}{\tau}\right) dt.$$

We may compute this using saddle-point approximation. Let $f(t) = r^2/4Dt + t/\tau$. Then we see that f attains a maximum at $t_0 = r\sqrt{\tau}/2\sqrt{D}$. Let $g(t) = r/2D\tau t$. The saddle point approximation reads

$$p(r) \approx e^{-f(t_0)} g(t_0) \sqrt{\frac{2\pi}{f''(t_0)}} = \exp\left(\frac{-r}{\sqrt{D\tau}}\right) \frac{1}{\sqrt{D\tau}^{3/2}} \sqrt{\frac{2\pi r t_0^{3/2}}{4\sqrt{D}}} \Rightarrow \boxed{p(r) \propto \sqrt{r} \exp\left(\frac{-r}{\sqrt{D\tau}}\right)}$$

We could also say that asymptotically the exponential decay in r dominates over the \sqrt{r} growth, and so in the large r limit, $p(r) \sim \exp\left(-r/\sqrt{D\tau}\right)$.

4. Jensen's inequality and Kullback–Liebler divergence.

- (a) Claim: Jensen's inequality: For a convex function $\langle f(x) \rangle \geq f(\langle x \rangle)$. To prove this, we let a probability density function $p(x)$ be given. Then

$$\langle f(x) \rangle = \int p(x) f(x) dx \geq \int p(x) [f(\langle x \rangle) + f'(\langle x \rangle)(x - \langle x \rangle)] dx = f(\langle x \rangle) + f'(\langle x \rangle)\langle x - \langle x \rangle \rangle = f(\langle x \rangle).$$

And we're done.

- (b) We see that $D(p|q)$ is an expectation taken with respect to $p(x)$. The proof uses Jensen's inequality with the fact that the function $-\ln(x)$ is convex:

$$\begin{aligned} D(p|q) &= \left\langle \ln \frac{p}{q} \right\rangle_p \\ &= \left\langle -\ln \frac{q}{p} \right\rangle_p \\ &\geq -\ln \left\langle \frac{q}{p} \right\rangle_p \\ &= -\ln \left[\int p(x) \frac{q(x)}{p(x)} dx \right] = -\ln 1 = 0. \end{aligned}$$

5. The book of records.

- (a) Suppose we have picked n entries $\{x_1, \dots, x_n\}$. The probability that x_n is the largest is actually the same as the probability that any other entry x_i is the largest. Therefore, $\boxed{P_n = 1/n}$
- (b) Since S_N is the number of records after N attempts, we may write S_N as the sum of the indicators R_i . As a result,

$$\langle S_N \rangle = \sum_{m=1}^N \langle R_m \rangle = \sum_{m=1}^N (1 \times P_m + 0 \times (1 - P_m)) = \sum_{m=1}^N P_m = \sum_{m=1}^N \frac{1}{m}.$$

We may estimate the growth of $\langle S_N \rangle$ by bounding it by two integrals. Choose the integral

$$I = \int_0^N \frac{1}{n} dn = \ln N.$$

to be an estimate for $\langle S_N \rangle$. We see that I is an under-estimation. However, we can find how much $\langle S_N \rangle$ differs from $\ln N$ by taking a limit:

$$\lim_{N \rightarrow \infty} \langle S_N \rangle - \ln N = \text{EulerGamma} \approx 0.5775216 \dots$$

where EulerGamma is the output that I got from evaluating this limit in Mathematica. Therefore, we conclude that $\langle S_N \rangle$ grows like $\ln N$ in the $N \gg 1$ limit. If the number of trials double every year, then $N = N(t) = 2^t$. In this case,

$$\langle S_N \rangle \sim \ln N = \ln 2^t = t \ln 2$$

asymptotically.

(c) By definition, $\langle R_n R_m \rangle = \langle R_n \rangle_c \langle R_m \rangle_c + \langle R_n R_m \rangle_c = \langle R_n \rangle \langle R_m \rangle + \langle R_n R_m \rangle_c$, so

$$\langle R_n R_m \rangle_c = \langle R_n R_m \rangle - \langle R_n \rangle \langle R_m \rangle = P_n P_m - P_n P_m = 0$$

if $m \neq n$. When $m = n$ the covariance term becomes a variance term which is P_n .

(d) **(Optional)** We want to compute all moments and the first three cumulants for S_N . To compute all moments, we may use the method of characteristic function. However, since there is no nice way to explicitly write out the expectation function, we will have to proceed as follows: Starting with $S_1 = R_1 = 1$, we have $\langle e^{-ikS_1} \rangle = e^{-ik}$. Next, using the definition

$$\langle e^{-ikS_N} \rangle = \sum_{n=1}^N \text{Pr}(n) e^{-ikn}$$

we find

$$\begin{aligned} \langle e^{-ikS_2} \rangle &= \frac{1}{2} e^{-ik} + \frac{1}{2} e^{-2ik} = \frac{1}{2!} e^{-ik} (e^{-ik} + 1) \\ \langle e^{-ikS_3} \rangle &= \frac{1}{2} \frac{2}{3} e^{-ik} + \left(\frac{1}{2} \frac{2}{3} + \frac{1}{2} \frac{1}{3} \right) e^{-2ik} + \frac{1}{2} \frac{1}{3} e^{-3ik} = \frac{1}{3!} e^{-ik} (e^{-ik} + 1)(e^{-ik} + 2) \\ \langle e^{-ikS_4} \rangle &= \frac{1}{2} \frac{2}{3} \frac{3}{4} e^{-ik} + \left(\frac{1}{2} \frac{2}{3} \frac{1}{4} + \frac{1}{2} \frac{1}{3} \frac{3}{4} + \frac{1}{2} \frac{2}{3} \frac{1}{4} \right) e^{-2ik} + \left(\frac{1}{2} \frac{1}{3} \frac{1}{4} + \frac{1}{2} \frac{2}{3} \frac{3}{4} + \frac{1}{2} \frac{1}{3} \frac{3}{4} \right) e^{-3ik} + \frac{1}{2} \frac{1}{3} \frac{1}{4} e^{-4ik} \\ &= \frac{1}{4!} e^{-ik} (e^{-ik} + 1)(e^{-ik} + 2)(e^{-ik} + 3) \\ &\vdots \end{aligned}$$

where we have used generated the probabilities by writing out all the record-configurations within the sequence $\{x_n\}$ and finding the probability for each outcome via $P_n = 1/n$. In any case, we see that a pattern emerges(!) and conclude without proving¹ that

$$\langle e^{-ikS_N} \rangle = \frac{1}{N!} \prod_{n=1}^N (e^{-ik} + n - 1) = \frac{1}{N!} [e^{-ik}]^{\overline{N}},$$

where the exponent \overline{N} is used in combinatorics to denote the *rising factorial*. [A trip to Wikipedia²](#) tells us that this expression can be expressed in terms of the (unsigned) Stirling numbers of the first kind:

$$\langle e^{-ikS_N} \rangle = \frac{1}{N!} \prod_{n=1}^N (e^{-ik} + n - 1) = \frac{1}{N!} [e^{-ik}]^{\overline{N}} = \sum_{n=0}^N \binom{N}{n}_S e^{-ikn}.$$

Now it is trivial to find the moments of S_N : we simply expand the exponential into powers of k :

$$\langle e^{-ikS_N} \rangle = \sum_{n=0}^N \binom{N}{n}_S \sum_{m=0}^{\infty} \frac{1}{m!} (-ikn)^m = \sum_{m=0}^{\infty} \frac{(-ik)^m}{m!} \left[\frac{1}{N!} \sum_{n=0}^N \binom{N}{n}_S n^m \right]$$

¹The proof is probably closely related to how the probabilities are generated.

²Never knew my undergrad MA355 would be useful for StatMech... I recognized this as something related to *falling factorials* which are related to Stirling numbers. See Kenneth P. Bogart's *Combinatorics Through Guided Discovery* for details about Stirling numbers as well as a change of basis for going from falling to rising factorials.

So, the n th moment of S_N is given by the term in the square brackets:

$$\langle S_N^m \rangle = \frac{1}{N!} \sum_{n=0}^N \binom{N}{n}_S n^m$$

Next, since R_i 's are independent (in the statement of Part (c)), all correlation terms $\langle R_a R_b R_c \rangle$ vanish. As a result, we have

$$\langle S_N^m \rangle = \sum_{n=1}^N \langle R_n^m \rangle \iff \langle S_N^m \rangle_c = \sum_n \langle R_n^m \rangle_c.$$

We simply compute, using the fact that $\langle R_n \rangle = 1/n$:

$$\begin{aligned} \langle S_N^1 \rangle_c &= \sum_{n=1}^N \langle R_n^1 \rangle_c = \sum_{n=1}^N \frac{1}{n} \\ \langle S_N^2 \rangle_c &= \sum_{n=1}^N \langle R_n^2 \rangle_c = \sum_{n=1}^N (\langle R_n^2 \rangle - \langle R_n \rangle^2) = \sum_{n=1}^N \left[\left(1^2 \times \frac{1}{n} + 0^2 \times \frac{n-1}{n} \right) - \frac{1}{n^2} \right] = \sum_{n=1}^N \left(\frac{1}{n} - \frac{1}{n^2} \right) \\ \langle S_N^3 \rangle_c &= \sum_{n=1}^N \langle R_n^3 \rangle_c = \sum_{n=1}^N (\langle R_n^3 \rangle - 3\langle R_n^2 \rangle \langle R_n \rangle + 2\langle R_n \rangle^3) = \sum_{n=1}^N \left(\frac{1}{n} - \frac{3}{n^2} + \frac{2}{n^3} \right) \end{aligned}$$

where in the last equality we have used

$$\langle R_n^3 \rangle = 1^3 \frac{1}{n} + 0^3 \frac{n-1}{n} = \frac{1}{n}.$$

To answer the last point, since S_N is not quite a sum of iid variables (the R_n 's have different distributions for different n 's.) and that its variance is not finite when $N \rightarrow \infty$, the central limit theorem does not/might not apply to S_N . **I'm not too sure about this.**

- (e) **(Optional)** Suppose the second record occurs at location $m \in \{1, 2, \dots, 8\}$. Then, the probability for this event (in the total sample space) is

$$\begin{aligned} P(n_2 = m) &= P_1(1 - P_2)(1 - P_3) \dots (1 - P_{m-1})P_m(1 + P_{m+1}) \dots P_9 \\ &= \frac{1}{2} \frac{2}{3} \frac{3}{4} \dots \frac{m-2}{m-1} \frac{1}{m} \frac{m}{m+1} \dots \frac{1}{9} \\ &= \frac{1}{8 \times 9 \times (m-1)} \\ &= \frac{1}{72(m-1)} \end{aligned}$$

where in the third equality we have used the fact that multiplying the second line by $8(m-1)$ gives $1/9$. To use this probability in the n_2 sample space, we have to first normalize it:

$$\mathcal{N} = \sum_{m=2}^8 \frac{1}{72(m-1)} = \frac{121}{3360}.$$

Therefore,

$$\langle n_2 \rangle = \frac{1}{\mathcal{N}} \sum_{m=1}^N \frac{m}{72(m-1)} = \frac{1343}{363} \approx \boxed{3.7}$$

Let the third record to fixed at some value. In order for some number x_m to be the fourth record, it has to be bigger than the third record, and all numbers from x_{m-1} to x_{10} inclusive must be less than the third record. Let p denote the probability that we pick a number bigger than the third record, then the probability that a number x_m is the fourth record is

$$(1 - p)^{m-10} p.$$

So, we have

$$\langle n_4 \rangle_{\text{3rd record fixed}} = \sum_{m=10}^{\infty} (1 - p)^{m-10} p = 9 + \frac{1}{p}$$

using Mathematica. We now want to take the average of this quantity over all possible values of the third record. To do this, let us assume that the numbers are picked from some distribution ψ , so that we can write p in terms of the cumulative density function associated with ψ , evaluated at the value of the third record. Then

$$\langle n_4 \rangle = \langle \langle n_4 \rangle_{\text{3rd record fixed}} \rangle_{\text{all third records}} = \int_{-\infty}^{\infty} dx \psi(x) \left(9 + \frac{1}{p(x)} \right).$$

Since $p(x) = 1 - \int_{-\infty}^x \psi(x) dx$, we have $-\psi(x) dx = dp(x)$ by the fundamental theorem of calculus. With this, we have

$$\langle n_4 \rangle = \int_0^1 dp \left(9 + \frac{1}{p} \right)$$

which diverges. So $\langle n_4 \rangle$ is infinite, as expected.

6. Jarzynski equality.

- (a) The new probability density function $p_f(W(\mu))$ is obtained from $p(\mu)$ via a change of variable transformation for which the rule is

$$p_f(W(\mu)) \left| \frac{dW}{d\mu} \right| = p(\mu)$$

where $|dW/d\mu|$ is the Jacobian. Similarly, we have

$$p_b(-W(\mu')) \left| \frac{-dW}{d\mu'} \right| = p'(\mu').$$

Since there is no real thermodynamic reason for $d\mu \neq d\mu'$, we must have

$$\frac{p_f(W)}{p_b(-W)} = \frac{p(\mu)}{p'(\mu')} = \frac{\mathcal{Z}'}{\mathcal{Z}} \frac{\exp(-\beta \mathcal{H}(\mu))}{\exp(-\beta \mathcal{H}(\mu)) \exp(-\beta W(\mu))} = \exp[\beta(W + F - F')],$$

as desired, where we have used $\ln \mathcal{Z} = -\beta F$.

- (b) Let $\Delta F = F' - F$. From Part (a) we have

$$p_f(W)e^{-\beta W} = p_b(-W)e^{-\beta \Delta F} \implies \langle e^{-\beta W} \rangle = \int p_f(W)e^{-\beta W} dW = e^{-\beta \Delta F} \int p_b(-W) dW = e^{-\beta \Delta F}.$$

As a result,

$$\Delta F = -\frac{1}{\beta} \ln \langle e^{-\beta W} \rangle = -k_B T \ln \langle e^{-\beta W} \rangle = -k_B T \ln \left[\int p_f(W)e^{-\beta W} dW \right],$$

as desired.

(c) We use Jensen's inequality and the fact that the function $f(W) = e^{-\beta W}$ is convex:

$$\Delta F = -k_B T \ln \langle e^{-\beta W} \rangle \leq -k_B T \ln e^{-\beta \langle W \rangle} = -k_B T \ln e^{-\langle W \rangle / k_B T} = \langle W \rangle.$$

(d) To do this problem we have to define a probability density function $\rho(\omega)$ associated with violating the second law. Given a particular W , the probability density for second law violation is $p_f(W - \omega)p_b(-W)$. Thus, $\rho(\omega)$ is in some sense a “quasi-convolution” of p_f and p_b , obtained by integrating out W :

$$\rho(\omega) = \int p_f(W - \omega)p_b(-W) dW.$$

We wish to find $\Pr(\omega > 0)$, which is given by

$$\Pr(\omega > 0) = \int_0^\infty \rho(\omega) d\omega.$$

In order for a factor of $e^{-\beta\omega}$ to pop up, we must make use of the fact that

$$\frac{p_f(W)}{p_b(-W)} = e^{\beta(W+F-F')}.$$

In particular,

$$\left(\frac{p_f(W)}{p_b(-W)} \right)^{-1} \frac{p_f(W - \omega)}{p_b(-W + \omega)} = e^{-\beta(W+F-F')} e^{\beta(W - \omega + F - F')} = e^{-\beta\omega}.$$

As a result,

$$p_b(-W)p_f(W - \omega) = e^{-\beta\omega} p_f(W)p_b(-W + \omega) \implies \rho(\omega) = e^{-\beta\omega} \int p_f(W)p_b(-W + \omega) dW.$$

By shifting the integration $dW \rightarrow d(W + \omega)$, we find that

$$\rho(\omega) = e^{-\beta\omega} \int p_f(W + \omega)p_b(-W) d(W + \omega) = e^{-\beta\omega} \int p_f(W + \omega)p_b(-W) dW = e^{-\beta W} \rho(-\omega).$$

Therefore,

$$\Pr(\omega > 0) = e^{-\beta W} \int_0^\infty \rho(-\omega) d\omega \leq e^{-\beta W} \underbrace{\int_{-\infty}^\infty p(-\omega) d\omega}_{=1} \leq e^{-\beta W} = e^{-\beta W},$$

as desired, where we have used the fact that the cumulative probability is bounded above by 1.

7. (Optional) Dice.

(a) The unbiased probabilities are such that the entropy is maximized. Since 6 appears twice as many times as 1, and that entropy must be maximized, we have

$$p_6 = 2p_1 = 2a, \quad p_2 = p_3 = p_4 = p_5 = (1 - 3a)/4.$$

Now we compute:

$$S = - \sum_{i=1}^6 p_i \ln p_i = -a \ln a - 2a \ln 2a - 4 \frac{1-3a}{4} \ln \frac{1-3a}{4} = -a \ln a - 2a \ln 2a - (1-3a) \ln \frac{1-3a}{4}$$

To extremize S , we find dS/da :

$$\frac{dS}{da} = -8 \ln 2 + 3 \ln(1 - 3a) - 3 \log a = 3 \ln \left[\frac{1 - 3a}{2^{8/3} a} \right] = 0 \iff 1 - 3a = 2^{8/3} a \implies a = p_1 = \frac{1}{2^{8/3} + 3}$$

this is an local maximum because dS/da is monotonically decreasing and crosses zero at $a = 1/(2^{8/3} + 3)$. We can now calculate the rest of the probabilities:

$$p_6 = 2p_1 = \frac{2}{2^{8/3} + 3}, \quad p_2 = p_3 = p_4 = p_5 = \frac{1}{4} \left[1 - \frac{3}{2^{8/3} + 3} \right] = \frac{2^{2/3}}{2^{8/3} + 3}$$

(b) The information content is the difference in entropy of a fair dice and this one.

$$I = S_{\text{fair}} - S_{\text{loaded}} = -6 \times \frac{1}{6} \log_2 \frac{1}{6} - \left[-a_0 \log_2 a_0 - 2a_0 \log_2 2a_0 - (1 - 3a_0) \log_2 \frac{1 - 3a_0}{4} \right]$$

where we have defined I so that it is positive (the entropy associated with a fair dice is maximal) and $a_0 = 1/(2^{8/3} + 3)$. With the help of Mathematica, we find

$$I \approx 0.0267 \dots$$

Mathematica code:

```
In[13]:= a0 = 1/(2^(8/3) + 3);
In[17]:= N[-Log2[1/6] - (-a0*Log2[a0] -
2*a0*Log2[2*a0] - (1 - 3*a0)*Log2[(1 - 3*a0)/4])]
Out[17]= 0.0267239
```

8. (Optional) Approach to Equilibrium

(a) By time-translation invariance we have

$$C_{ij}(t) = C_{ij}(t + \tau) = \langle x_i(t + \tau) x_j(\tau) \rangle.$$

Picking $\tau = -t$, we find

$$C_{ij}(t) = \langle x_i(0) x_j(-t) \rangle.$$

By time-reversal invariance we have

$$C_{ij}(t) = C_{ij}(-t) = \langle x_i(0) x_j(t) \rangle = \langle x_j(t) x_i(0) \rangle = C_{ji}(t),$$

as desired.

(b) As appeared in the form

$$\sqrt{\frac{\det(K)}{(2\pi)^n}} \exp \left[-\frac{1}{2} K_{ij} x_i x_j \right],$$

the matrix $[K]$ is the inverse of the covariance matrix associated with this Gaussian, i.e.,

$$[K^{-1}]_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle$$

On the other hand, $C_{ij}(0) = \langle x_i(0) x_j(0) \rangle = \langle x_i x_j \rangle$ forms the autocorrelation matrix. The two matrices are related by the well-known identity in statistics:

$$[K]^{-1} = [C(0)] - \langle \mathbf{x} \rangle \langle \mathbf{x} \rangle^T \implies C[0] = [K]^{-1} + \langle \mathbf{x} \rangle \langle \mathbf{x} \rangle^T$$

where $\langle \mathbf{x} \rangle^T = (\langle x_1 \rangle, \langle x_2 \rangle \dots)^T$.

(c) Given $J_\alpha = -\partial \ln p(\mathbf{x}) / \partial x_\alpha$, we may compute:

$$J_\alpha = -\frac{\partial}{\partial x_\alpha} \ln \left\{ \sqrt{\frac{\det(K)}{(2\pi)^n}} \exp \left[-\frac{1}{2} K_{ij} x_i x_j \right] \right\} = -\frac{\partial}{\partial x_\alpha} \left[-\frac{1}{2} K_{ij} x_i x_j \right] = K_{\alpha j} x_j.$$

So,

$$\langle J_\alpha x_\beta \rangle = \langle K_{\alpha j} x_j x_\beta \rangle.$$

In order to compute $\langle x_j x_\beta \rangle$, we recall the definition for $\langle x_j \rangle$:

$$\langle x_j \rangle = \sqrt{\frac{\det(K)}{(2\pi)^n}} \int \cdots \int x_j \exp \left[-\frac{1}{2} K_{\alpha\beta} x_\alpha x_\beta \right] \prod_m dx_m.$$

By the form of the given Gaussian, we see that $\langle x_j \rangle = 0$. However, let us pretend that $\langle x_j \rangle = \eta_j \neq 0$, so that we can parameterize the Gaussian with these η_j 's as follows:

$$\langle x_j \rangle = \eta_j = \sqrt{\frac{\det(K)}{(2\pi)^n}} \int \cdots \int x_j \exp \left[-\frac{1}{2} K_{\alpha\beta} (x_\alpha - \eta_\alpha)(x_\beta - \eta_\beta) \right] \prod_m dx_m.$$

Now we use Feynman's trick of differentiating under the integral sign with respect to η_k and re-setting $\eta_j = 0$ for all j to find that

$$\frac{\partial}{\partial \eta_k} \eta_j = \sqrt{\frac{\det(K)}{(2\pi)^n}} \int \cdots \int x_j K_{k\alpha} x_\alpha \exp \left[-\frac{1}{2} K_{\alpha\beta} (x_\alpha - \eta_\alpha)(x_\beta - \eta_\beta) \right] \prod_m dx_m = \langle x_j K_{k\alpha} x_\alpha \rangle.$$

Thus, we see that

$$\langle J_\alpha x_\beta \rangle = \langle K_{\alpha j} x_j x_\beta \rangle = \frac{\partial}{\partial \eta_\alpha} \eta_\beta = \delta_{\alpha\beta},$$

as desired.

(d) Starting with $C_{ij}(t) = C_{ji}(t)$, we differentiate both sides with respect to t to find:

$$\begin{aligned} \frac{d}{dt} \langle x_i(t) x_j(0) \rangle &= \frac{d}{dt} \langle x_j(t) x_i(0) \rangle \implies \langle \dot{x}_i(t) x_j(0) \rangle = \langle \dot{x}_j(t) x_i(0) \rangle \\ &\implies \langle \mu_{i\alpha} K_{\alpha\beta} x_\beta(t) x_j(0) \rangle = \langle \mu_{j\alpha} K_{\alpha\beta} x_\beta(t) x_i(0) \rangle. \end{aligned}$$

Setting $t = 0$, we get

$$\begin{aligned} \langle \mu_{i\alpha} K_{\alpha\beta} x_\beta(0) x_j(0) \rangle &= \langle \mu_{j\alpha} K_{\alpha\beta} x_\beta(0) x_i(0) \rangle \implies \langle \mu_{i\alpha} K_{\alpha\beta} x_\beta x_j \rangle = \langle \mu_{j\alpha} K_{\alpha\beta} x_\beta x_i \rangle \\ \text{Definition} &\implies \mu_{i\alpha} \langle J_\alpha x_j \rangle = \mu_{j\alpha} \langle J_\alpha x_i \rangle \\ \text{Part (c)} &\implies \mu_{i\alpha} \delta_{\alpha j} = \mu_{j\alpha} \delta_{\alpha i} \\ &\implies \mu_{ij} = \mu_{ji}. \end{aligned}$$

So, the matrix μ must be symmetric, as desired.

9. (Optional) Simpson's Paradox

(a) Shouldn't the question be: "Show that the country A as a whole can still end up with lower cfrs if its population is **younger**.", as this would seem paradoxical given that $p_o^A > p_o^B > p_y^A > p_y^B$?

To answer this modified question, I could just give an example. Let $p_o^A = 0.9$, $p_o^B = 0.7$, $p_y^A = 0.5$, $p_y^B = 0.3$. We would expect that since A has higher fatality rates than B across all age groups, it would have

a higher overall fatality rate. However, this assumption ignores the effect of population. Suppose that A has 10 olds and 90 youngs, while B has 90 olds and 10 youngs, then we find that

$$p^A = \frac{10 \times 0.9 + 90 \times 0.5}{100} = 54\%$$

whereas

$$p^B = \frac{90 \times 0.7 + 10 \times 0.3}{100} = 66\% > p^A.$$

Note that it is impossible for A to have lower overall fatality rate than B if its population is older. Let η_A denote the old/total ratio in A and η_B in B . Then the overall rates for the countries are defined by two lines

$$\begin{aligned} r_A(\eta_A) &= \eta_A p_o^A + (1 - \eta_A) p_y^A = p_y^A + \eta_A (p_o^A - p_y^A) \\ r_B(\eta_B) &= \eta_B p_o^B + (1 - \eta_B) p_y^B = p_y^B + \eta_B (p_o^B - p_y^B) \end{aligned}$$

defined for $\eta_A, \eta_B \in [0, 1]$. It is obvious that the line corresponding to the overall rate of A lies strictly above the line associated with B . Therefore, for any η , we have

$$r_A(\eta) - r_B(\eta) = (p_y^A - p_y^B) + \eta[(p_o^A - p_o^B) - (p_y^A - p_y^B)] = (1 - \eta)(p_y^A - p_y^B) + \eta(p_o^A - p_o^B) > 0$$

which means if the populations of A and B are similar (age-wise) then it is impossible for A to have a smaller fatality rate. Now, suppose we make A even older, i.e., $\eta_A > \eta_B$ then the fatality rate for B is $r_B(\eta_B) < r_B(\eta_A) < r_A(\eta_A)$ where we have used the fact that the slope of $r_B(\eta)$ is positive. As a result, it is actually **not possible** for A to have a lower fatality rate if its population is **older**.

From this we can also conclude that $r_A < r_B$ can occur only if $\eta_A < \eta_B$, or A is younger than B .

- (b) For simplicity we may consider two pairs of engines, two Carnot (C_1, C_2) and two non-Carnot (N_1, N_2). In view of Part (a), let us define the efficiencies for C_1, C_2, N_1, N_2 as $\eta_{C_1}, \eta_{C_2}, \eta_{N_1}, \eta_{N_2}$ such that

$$\eta_{C_1} > \eta_{N_1} > \eta_{C_2} > \eta_{N_2}.$$

Here, $\eta_{C_i} > \eta_{N_i}$, the Carnot and non-Carnot engines in each pair share the same input and output. The Carnot engines receive some quantity Q of heat, and similarly do the non-Carnot engines. However, since we don't know how Q is shared within the engines, we may call λ_C and λ_N the distribution ratios within the Carnot and non-Carnot engines respectively. Then, the work done by the Carnot and non-Carnot engines are

$$\begin{aligned} W_C &= Q[\lambda_C \eta_{C_1} + (1 - \lambda_C) \eta_{C_2}] \\ W_N &= Q[\lambda_N \eta_{N_1} + (1 - \lambda_N) \eta_{N_2}] \end{aligned}$$

By virtue of the analysis in Part (a), we may consider the case where

$$\eta_{C_1} = 0.9 \quad \eta_{N_1} = 0.7 \quad \eta_{C_2} = 0.5 \quad \eta_{N_2} = 0.3$$

and $\lambda_C = 0.1, \lambda_N = 0.9$. We find

$$W_C = 0.54Q < 0.66Q = W_N.$$

So, the non-Carnot engines actually produce more work! While this seems to violate the second law of thermodynamics, it doesn't. The apparent paradox fails to account for the confounding variable which is the distribution of the input heat Q across the engines. While Q is given to both sets of engines, it is entirely possible that it is simultaneously absorbed by the most efficient non-Carnot engine and least efficient Carnot engine. And since we have made our engines so that $\eta_{C_2} < \eta_{N_1}$, this scenario is thermodynamically allowed.