Overview: The goal of this first problem set is to see how a single macrostate can become exceedingly probable even when all the microstates are equally probable. The dominance of that most probable macrostate emerges in the limit of many independent entities. In problem 1 these entities are independent coin flips, in problem 2 they are independent spins, and in problem 3 they are independent steps of a random walk.

1. **Coin Flips.** Imagine flipping an unbiased coin N times. Let $N_{\rm H}$ be the number of heads results, and $f = N_{\rm H}/N$ be the fraction of such results.

   (i) What is the probability of observing a particular sequence of heads (H) and tails (T) results, e.g., H T T T T H H T T H H T H... ?

   > Each coin flip has two possibilities, so there are 2^N possible sequences of N coin flips. Each sequence is equally likely, so a particular sequence has probability $1/2^N$.

   (ii) How many possible flip sequences yield exactly $N_{\rm H}$ heads results? Your answer should involve the factorial function, $M! \equiv M \times (M-1) \times (M-2) \times \ldots \times 3 \times 2 \times 1$.

   > Think of the N different coin flips as N slots, each of which must be assigned an H or a T. We must place $N_{\rm H}$ heads into those slots. The first H has N possible places to go, the next $N-1$, and so on. Hence we might suspect there will be $N!$ possible ways to put the H's into the N slots. This, however, would be overcounting. Why? Because we do not really keep track of the order. In other words, if the first H is assigned to slot 1 and the second H is assigned to slot 2, it yields the same result as the first H going into 2 and the second H going into 1. We're aiming to count the number of final configurations, not the number of ways to get to those final configurations (in which case the order would matter). To prevent the overcounting, we must divide by the number of ways to shuffle the $N_{\rm H}$ H's around their slots, $N_{\rm H}!$, and by the number of ways to shuffle the $N - N_{\rm H}$ T's around in their slots, $(N-N_{\rm H})!$. Hence the final result is:
   >
   > $$\frac{N!}{N_{\rm H}!(N-N_{\rm H})!}.$$

   (iii) Write an exact equation for the probability $P(N_{\rm H})$ of observing $N_{\rm H}$ heads results when the coin is flipped N times.

   > From (i), we have the probability of each configuration. From (ii), we have the number of configurations yielding $N_{\rm H}$ H's. Combined, we get
   >
   > $$P(N_{\rm H}) = \frac{1}{2^N}\frac{N!}{N_{\rm H}!(N-N_{\rm H})!}.$$

   (iv) Stirling's approximation,

   $$\ln M! \approx M \ln M - M \quad \text{for large } M,$$

   allows you to simplify your result in part (iii) assuming N is very large. First, we consider a hand-wavy way to "derive" Stirling's approximation. We know that the integral of a function $g(x)$ can be

approximated by a Riemann sum:

$$\int_a^b dx\, g(x) \approx \sum_{i=0}^{(b-a)/\Delta x} g(a + i\Delta x)\Delta x$$

when Δx is sufficiently small. If $b - a \gg 1$, $\Delta x = 1$ can be small enough for a good approximation of the integral. Follow this line of argument to show Stirling's approximation. (Hint: you will want to consider $g(x) = \ln x$ and an appropriate choice of a and b.)

$$\ln M! = \ln 1 + \ln 2 + \ldots \ln M \approx \int_0^M dx \ln x = (x \ln x - x)|_0^M = M \ln M - M.$$

The Riemann sum step is effectively doing a right rectangle method.

(v) Armed with Stirling's approximation, show that $P(N_\mathrm{H})$ can be written in the large deviation form

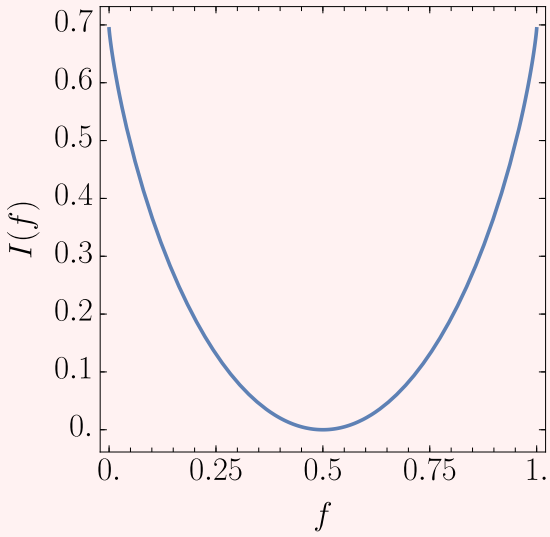
$$P(N_\mathrm{H} = fN) \propto e^{-NI(f)}$$

when N is sufficiently large to justify Stirling's approximation. Identify and plot $I(f)$ as a function of f. [*Please plot this and future plots using a computer. If you feel uncomfortable doing so, see Problem Set 0 and/or* BiasedCoinFlip.ipynb *for additional help.*] Notice that I does not depend on N. In other words the extensive (large) part of the problem has dropped out and only impacts the probability through the factor that multiplies I. This is a major simplification! You might have thought that the term in the exponent should have higher powers of N, but it does not.

From (iv) we have $N! \approx N^N e^{-N}$, so

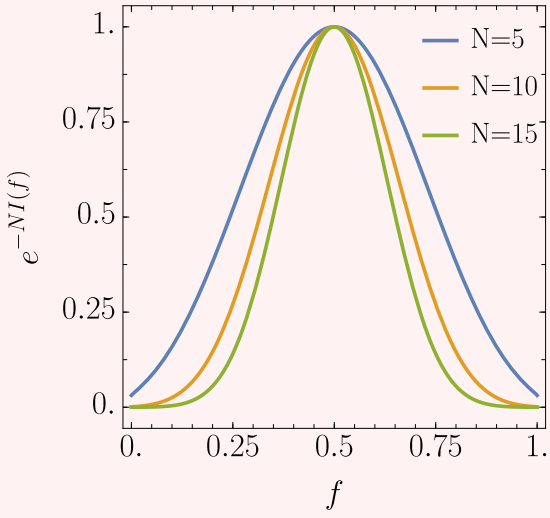
$$\begin{aligned}
P(N_\mathrm{H}) &\approx \frac{1}{2^N} \frac{N^N e^{-N}}{N_\mathrm{H}^{N_\mathrm{H}} e^{-N_\mathrm{H}} (N - N_\mathrm{H})^{N-N_\mathrm{H}} e^{-N+N_\mathrm{H}}} \\
&= \frac{1}{2^N} \frac{N^N}{N_\mathrm{H}^{N_\mathrm{H}} (N - N_\mathrm{H})^{N-N_\mathrm{H}}} \\
&= \frac{1}{2^N} \frac{N^N}{(Nf)^{Nf} (N - Nf)^{N-Nf}} \\
&= \frac{1}{2^N} \frac{N^N}{N^{Nf} N^{N-Nf} f^{Nf} (1-f)^{N-Nf}} \\
&= \frac{1}{2^N} \frac{1}{f^{Nf} (1-f)^{N(1-f)}} \\
&= e^{-N[\ln 2 + f \ln f + (1-f)\ln(1-f)]}.
\end{aligned}$$

The final term in the square brackets is $I(f)$, which we plot here:

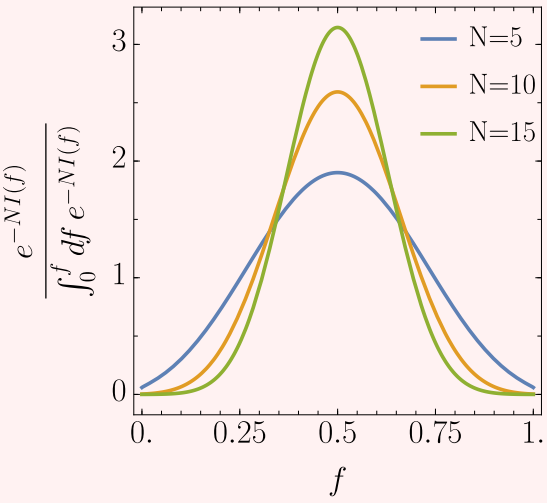


(vi) For $N = 5, 10$, and 15, plot $e^{-NI(f)}$ on the same plot. Observe the very rapid concentration at $f = 0.5$. You will probability find that it is helpful to normalize the curve for each value of N so you can be looking at an approximation for the probability distribution (rather than something which is merely proportional to a distribution). You should see that measurements of f become more and more deterministic as N increases. We explore this point further in the next problem.

The plot I asked for looks like this:



though it is a little more instructive to look at the normalized version:



In retrospect, the point would have been made a little more clearly if I had asked you to plot things for even larger values of N.

2. **A Macroscopic Number of Spins.** Now imagine the physical scenario of making a single measurement (as opposed to repeated coin flips) of $N \gg 1$ noninteracting spin-1/2 particles. In that measurement, the observed z-component of each spin is up or down with equal probability.

   (i) What is the probability $P(f)$ of observing a fraction $f = N_{\text{up}}/N$ of up spins in a given observation? Write your answer in terms of the fraction f and the number of spins N.

The independent coin flips in problem 1 were repeated, one after while these spins are measured all at once. However, if the spins are truly noninteracting, it doesn't matter that they're measured simultaneously. Their states are still indpendent of the other spins, just like the coin flips. Hence the large-N limiting probability distribution is well-approximated by

$$P(N_{\text{up}}) \approx e^{-N[\ln 2 + f \ln f + (1-f)\ln(1-f)]}.$$

(ii) Although $f = 1/2$ is the most likely observation, a typical measurement will not yield *exactly* half the spins pointing up. For Avogadro's number of spins, $N \approx 10^{24}$, estimate the relative probability of a small deviation $\delta = 10^{-7}$ from the ideal fraction, i.e., calculate $P(f = 0.5 + \delta)/P(f = 0.5)$. Your numerical answer need not be highly accurate; just determine the order of magnitude. (For this purpose, Taylor expansion of $\ln P$ about $\delta = 0$ is both permitted and a good idea).

Notice that the relative probability of measuring $f = 0.5 + \delta$ to the probability of measuring $f = 0.5$ is given by

$$\frac{P(f = 0.5 + \delta)}{P(f = 0.5)} \approx e^{-N[I(0.5+\delta) - I(0.5)]}.$$

It is the exponent that we handle with a Taylor series expansion.

$$I(0.5 + \delta) - I(0.5) \approx I'(0.5)\delta + \frac{1}{2} I''(0.5)\delta^2 + \mathcal{O}(\delta^3),$$

where

$$I''(f) = \frac{\partial}{\partial f}[\ln f + 1 - \ln(1-f) - 1] = \frac{\partial}{\partial f}[\ln f - \ln(1-f)] = \frac{1}{f} + \frac{1}{1-f}.$$

Thus $I''(0.5) = 4$ and

$$\frac{P(f = 0.5 + \delta)}{P(f = 0.5)} \approx e^{-10^{24} \cdot 10^{-14} \cdot 2} \approx e^{-2 \cdot 10^{10}}.$$

This is a very tiny number, meaning it is exceedingly unlikely to measure a value of f which deviates appreciably from the expected value of $f = 0.5$.

(iii) For finite N, only discrete values of f are possible, but in the limit of large N, $P(f)$ approaches a Gaussian distribution of the form

$$\rho(f) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(f-\mu)^2}{2\sigma^2}\right)$$

with mean μ and variance σ^2. In that limit, f is continuous rather than being limited to the discrete values $0, 1/N, 2/N, \ldots, 1$. Using your Taylor expansion from (ii), determine μ and σ^2 to obtain the probability distribution for $P(f)$ in the large N limit. With an appropriate change of coordinates, also determine the large N (Gaussian) limit for $P(N_{\text{up}})$. To make sure you have changed coordinates correctly, confirm for yourself that your expression for $\rho(N_{\text{up}})$ is normalized.

Note: passing from discrete to continuous probability distributions can be a little subtle. Technically $\rho(f)$ is not the probability of observing f; rather the probability of observing $a \leq f \leq b$ is given by

$$P(a \leq f \leq b) = \int_a^b df\, \rho(f).$$

We have found from 2(i) and 2(ii) that

$$\rho(f) \propto e^{-NI(f)}, \quad I(f) \approx 2(f-0.5)^2.$$

Combining these results yields

$$\rho(f) \propto e^{-2N(f-0.5)^2}.$$

Recalling that a normalized Gaussian distribution for the random variable X has density

$$\rho_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(x-\mu)^2/(2\sigma^2)},$$

where μ is the mean of X and σ^2 the variance of X, we can construct a similar density for f. By matching terms in the exponential argument, we identify $\sigma^2 = 1/(4N)$ and hence the normalization factor $\sqrt{2N/\pi}$, giving rise to the normalized density

$$\rho(f) = \sqrt{\frac{2N}{\pi}}\, e^{-2N(f-0.5)^2}.$$

In order to find the distribution for N_{up}, we simply perform the change of variables $N_{\text{up}} = Nf$. Notice that we *cannot* simply substitute $f = N_{\text{up}}/N$ in $\rho(f)$ to obtain $\rho(N_{\text{up}})$. This is

because the density is not invariant under a change of variables, unlike the probability, which is. (Probability is conserved.) Thus, when changing variables, the appropriate equation for doing so must involve an equality of probabilities; that is,

$$\rho(f)\,|\mathrm{d}f| = \rho(N_{\text{up}})\,|\mathrm{d}N_{\text{up}}| \quad \Longleftrightarrow \quad \rho(N_{\text{up}}) = \rho(f)\left|\frac{\mathrm{d}f}{\mathrm{d}N_{\text{up}}}\right|,$$

so we must, in addition to substituting $f = N_{\text{up}}/N$, multiply the distribution by an additional factor, essentially the Jacobian for the transformation.
We thus obtain

$$\rho(N_{\text{up}}) = \sqrt{\frac{2}{N\pi}}\,e^{-2(N_{\text{up}}-0.5)^2/N}.$$

If we had forgotten about the Jacobian and had simply substituted $N_{\text{up}} = Nf$, the integral over the distribution would have been N rather than 1—our transformation essentially stretches the density by a factor of N in the x-direction, because a fraction $f = 0.5$ corresponds to a number $N_{\text{up}} = 0.5N$, and naturally the integral over the distribution also stretches by a factor of N. The Jacobian fixes this problem.

(iv) The fraction of up spins f is intensive whereas the total number of up spins N_{up} is extensive. Imagine recording both f and N_{up} from a measurement of a macroscopic system. Does the variance of your measurements increase or decrease as the system is made bigger? Base your answer on your distributions from (iii). You may find that the variance behaves differently for intensive and extensive measurements.

We imagine measuring all of the N spins in one measurement. That returns a single value of f and N_{up}. Now imagine repeating that measurement many times. Those repeated measurements return values of f and N_{up} which are effectively drawn from the distributions $P(f)$ and $P(N_{\text{up}})$ from part (iii). The Gaussian distribution describing $P(f)$ has mean $1/2$ and variance $1/(4N)$, while the Gaussian distribution describing $P(N_{\text{up}})$ has mean $N/2$ and variance $N/4$ (see problem 3). Thus we see that the variance of f measurements *decreases* as N increases, meaning we become more and more certain of the result. Indeed, we saw in (ii) that the chances of deviating very far from $f = 1/2$ become exceedingly unlikely. In contrast, the measurement of N_{up} has a variance which *grows* proportionally to N. These are very general observations. Intensive variables (like f) have fluctuations which decay rapidly, so rapidly that in large systems it makes sense to characterize them by their average value and ignore the fluctuations altogether. The study of those expectation values is effectively the study of thermodynamics. More on this as the course proceeds.

(v) You may have noticed in (iii) that by moving from discrete to continuous f in the large N limit, we have inadvertently allowed f to range from $-\infty$ to ∞. Argue that this is not a problem.

Many possible solutions. Formally and quantitatively, one may integrate the distribution in the interval $(-\infty, 0] \cup [1, \infty)$ to show that its contribution to the density is negligible and decreases exponentially in N. More qualitatively, one might simply calculate the value of the density at $f = 0$ or $f = 1$ and hand-wavingly argue that the density would be so small as to be insignificant beyond the boundary. The purpose of this problem is simply to explore more

deeply the ramifications of the various, seemingly innocuous assumptions that we make.

3. **A Random Walk.** In class we discussed deterministic models for dynamics that had their origins in physics. Suppose, however, that you have a fluorescent protein in solution and every Δt units of time you make a measurement of the protein's location. For simplicity, we will focus on a single dimension, tracking only the x coordinate of the protein. You might reason that the effect of all the solvent molecules is to randomly bump against the protein causing it to move a little bit to the left or a little bit to the right every Δt. That reasoning leads to a probabilistic model for the dynamics which is known as a 1d random walk. With probability $1/2$ the protein moves to the right by a distance l and with probability $1/2$ it moves left by the same distance. (To make things easy on you, I have not allowed the particle to stay at its original position in a step of duration Δt. If this disturbs you, feel free to solve that model as well, and you'll see the same sort of behavior!)

(i) Let the position of the protein at the initial time be 0. Use your results from Problem 1 to determine the probability distribution $P_N(X)$ that the protein is at position X after N steps.

We can map coin tosses to this problem as follows: Flip a coin. If it shows heads, move right, else, move left. The final position is then the difference between the number of heads and tails, divided by the length of each step:

$$\frac{X}{l} = N_H - N_T \quad \Longleftrightarrow \quad N_H = \frac{N}{2} + \frac{X}{2l},$$

where in deriving the expression for N_H we have also used the identity $N_H + N_T = N$. Then, it suffices to plug this result into the exact result of 1.iii,

$$P(X) = P\left(N_H = \frac{N}{2} + \frac{X}{2l}\right) = 2^{-N}\binom{N}{N/2 + X/(2l)}.$$

Notice that, even though fractions appear in this equation for N_H, N_H only takes on integer values because N and X are always of the same parity, which is itself because the walker's position must alternate between even and odd values every step.

(ii) According to Problem 2, we should expect $P_N(X)$ to tend toward a Gaussian distribution in the limit of a large number of steps. Determine the Gaussian $\rho(X)$ in terms of N and l.

From 2(iii),

$$\rho\left(N_{\text{up}} = \frac{N}{2} + \frac{X}{2l}\right) \mathrm{d}N_{\text{up}} = \sqrt{\frac{2}{\pi N}} e^{-X^2/(2Nl^2)} \frac{\mathrm{d}X}{2l} = \sqrt{\frac{1}{2\pi N l^2}} e^{-X^2/(2Nl^2)} \mathrm{d}X.$$

Because $\rho(N_{\text{up}})\,\mathrm{d}N_{\text{up}} = \rho(X)\,\mathrm{d}X$, we must have

$$\rho(X) = \sqrt{\frac{1}{2\pi N l^2}} e^{-X^2/(2Nl^2)}.$$

(iii) After N steps the average position is given by

$$\langle X \rangle = \sum_X X P_N(X),$$

where the sum includes all allowed values of X. In the large N limit, this average becomes the integral:

$$\langle X \rangle = \int dX \, X \rho(X).$$

What is the average position as a function of N and l?

The mean is 0, by symmetry—the protein moves left and right with equal probability. In terms of (fair) coin tosses, heads and tails are equally likely. Mathematically, this is represented by the fact that the integrand for the mean is odd, whereas the range of integration is even, so the integral must vanish. This can be confirmed by evaluating the Gaussian integral directly, which is worth doing at least once.

(iv) After N steps the variance in the position is given by

$$\left\langle \delta X^2 \right\rangle \equiv \left\langle \left(X - \langle X \rangle \right)^2 \right\rangle = \sum_X \left(X - \langle X \rangle \right)^2 P_N(X).$$

In the large N limit, this variance becomes the integral

$$\left\langle \delta X^2 \right\rangle = \int dX \, \left(X - \langle X \rangle \right)^2 \rho(X).$$

What is the variance as a function of N and l?

Notice that this is not zero, because even though the protein isn't biased on average, it does move around. Comparing the exponential argument of the probability distribution derived in 3(ii) to that of a Gaussian yields

$$\sigma_X^2 = \langle \delta X^2 \rangle = N l^2.$$

One can also perform the Gaussian integral directly (easiest by differentiating under the integral sign, though I expect integration by parts might also work). Alternatively, one can notice that each step is independent, X is the sum of all these steps, and the variance of the sum of independent random variables is the sum of the variance of each random variable. The variance of one step is straightforward to calculate by definition (you get l^2), and there are N steps, so the overall variance is Nl^2, as has been derived in a couple of different ways here.

(v) A diffusion constant D is a measure of how quickly the probability distribution for a particle's position spreads out. Specifically (for a one-dimensional problem), $\left\langle \delta X^2 \right\rangle = 2D\tau$, where τ is the total elapsed time. What is D in terms of Δt and l?

From 3(iv), $\langle \delta X^2 \rangle = N l^2$. Realizing that $\tau = N \Delta t$, we have

$$\langle \delta X^2 \rangle = \frac{\tau l^2}{\Delta t} = 2 \left(\frac{l^2}{2 \Delta t} \right) \tau =: 2D\tau \quad \Longleftrightarrow \quad D := \frac{l^2}{2\Delta t}.$$