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... ?

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

   (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.

   (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.)

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

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

   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.

   (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.

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 $f = N_{\rm up}/N$ of up spins in a given observation? Write your answer in terms of the fraction f and the number of spins N.

   (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).

   (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 $P(N_{\rm up})$. To make sure you have changed coordinates correctly, confirm for yourself that your expression for $\rho(N_{\rm up})$ 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 $a \le f \le b$ is given by

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

   (iv) The fraction of up spins f is intensive whereas the total number of up spins $N_{\rm up}$ is extensive. Imagine recording both f and $N_{\rm up}$ 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.

   (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.

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.

(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.

(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?

(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?

(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?