1. **A Warm Up to Establish Notation.** In this course we will be using mathematical tools from probability and statistics to uncover physical principles. To make sure we are all on the same page (and to establish a little notation), let's start by reviewing random variables and expectation values.

   Let X be a discrete random variable that takes the value 1, 2, 3, 4, 5, or 6 based on the roll of a fair die. By fair I mean that each possibility is equally likely, so the probability that X equals any particular value is $1/6$. The expectation value or mean of X computes an average over the possible values, weighted by their probabilities:

   $$\langle X \rangle = \sum_x x P(X = x) = \sum_x \frac{x}{6} = \frac{7}{2}.$$

   Note that I have used a shorthand summation notation where my subscript x indicates that x is being summed over all of its possible values without explicitly listing what those values are.

   Even though X can never take the value 3.5, in a certain sense we "expect" 3.5 to be the outcome. We can also measure deviations from that expectation. Throughout the course we will use a lowercase δ to denote these deviations, $\delta x \equiv x - \langle x \rangle$.

   > You may wonder about the difference between $\langle X \rangle$ above and the $\langle x \rangle$ I used here? There is no real difference, just different notational choices. In a math class when talking about random variables you would quite possibly see the former, but in this class we will typically be less formal. Rather than talking of "the expectation value of the random variable X", I'm more likely to talk about "the average value of x", and I think we all know what that means.

   (i) What is the expected deviation from the mean, $\langle \delta x \rangle$?

   (ii) What is the expected squared value of this deviation, also known as the variance, $\langle \delta x^2 \rangle$?

2. **Fun with Gaussians.** Random variables can also be continuous, for example, x might be allowed to take all possible real values with some probability density $\rho(x)$. Technically you cannot define the probability of a single value of x (it has "measure zero"), but you can instead define the probability that x will fall in a small interval by integrating $\rho(x)$:

   $$P(x \in [a, b]) = \int_a^b dx\, \rho(x).$$

   One of the most important continuous distributions is the Gaussian distribution

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

   You likely already have some familiarity with this distribution, so this problem is either a crash course or a quick refresher. I do not consider it important that you "discover" the standard tricks on your own, but I do think you should be aware of them. Please talk to me or to other students if this is foreign!

   (i) Normalization: Show that the probability density is normalized, meaning $\int dx\, \rho(x) = 1$. [*Standard trick: Convert to a two dimensional integral over a joint distribution of identical, independent Gaussians for the x and y coordinates. You may find you need to remind yourself what a Jacobian is.*]

(ii) Mean: Show that $\langle x \rangle = \mu$. [*Standard trick: Substitute $u = x - \mu$ and notice that an integral cancels by symmetry.*]

(iii) Variance: Show that $\langle \delta x^2 \rangle = \sigma^2$. [*Standard trick: From (a) and symmetry find the integral of $e^{-\alpha x^2}$ from 0 to ∞. Differentiate with respect to α.*]

3. **Coin Flips.** Now we will solve perhaps the most common problem in an introductory statistical mechanics course. On its surface this is a problem about the probability of outcomes from flipping fair coins, but the broader goal 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 which case we will see that it is useful to approximate discrete random variables by continuous ones.

   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. 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 $N = 5, 25$, and 100, plot $e^{-NI(f)}$ on the same plot. You should see that the plots are getting more sharply peaked at $f = 0.5$ as N increases, but the distributions are not normalized. To approximate the *normalized* probability distribution, plot $e^{-NI(f)} / \int_0^1 df\ e^{-NI(f)}$ for these same three values of N. You should see that measurements of f become more and more deterministic

(concentrated around $f = 0.5$) as N increases. We explore this concentration of probability further in Problem 4.

(vii) (OPTIONAL) Normalizing $e^{-NI(f)}$ may have seemed ad hoc and unsatisfying to some of you. You might have expected that the normalization should have already been present since our expression for $P(N_{\mathrm{H}})$ in (iii) was already normalized. The problem was that Stirling's approximation as written in (iv) is an approximation, but this approximation actually could have been improved to more gracefully handle the normalization. The problem leading-order correction is $\ln\sqrt{2\pi n}$, so that Stirling's approximation is stated more precisely as

$$\ln n! = n\ln n - n + \ln\sqrt{2\pi n} + O\left(\frac{1}{n}\right),$$

where the big-O notation implies that higher-order corrections are asymptotically at least as small as n^{-1}. Explain why neglecting this logarithmic contribution affects the normalization, and why neglecting higher-order contributions will not.

(viii) (OPTIONAL) A nice way to derive that leading-order correction of part (vii) uses Laplace's method, which approximates integrals of exponential form associated with some large parameter λ:

$$\int \mathrm{d}x\, e^{\lambda f(x)} \approx \int_{-\infty}^{\infty} \mathrm{d}x\, e^{\lambda(f(x^*)+f''(x^*)(x-x^*)^2/2)},$$

where x^* is the local maximum of f within the original domain of integration. In words, the method simply consists of approximating $f(x)$ as its second-order Taylor expansion where the integrand is peaked. The resulting integral is Gaussian and hence easily evaluated. Apply Laplace's method to the integral representation of the factorial,

$$n! = \int_0^{\infty} \mathrm{d}x\, x^n e^{-x},$$

to derive the stated logarithmic correction. You will likely want to introduce the substitution $x = n\xi$.

4. **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) Using your work on the previous problem, approximate the probability $P(f)$ of observing a fraction $f = N_{\mathrm{up}}/N$ 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 $P(f = 0.5 + \delta)/P(f = 0.5)$. 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 $\rho(f)$ in the large N limit. With an appropriate change of coordinates, also determine the large N (Gaussian) limit for $\rho(N_{\rm up})$. To make sure you have changed coordinates correctly, confirm for yourself that your expression for $\rho(N_{\rm up})$ is normalized.

(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.