1. **An ideal gas on a lattice.** In class we introduced a microscopic definition of entropy due to Boltzmann:

   $$S_{\mathrm{B}} = k_{\mathrm{B}} \ln \Omega.$$

   A more general expression for entropy due to Gibbs is:

   $$S_{\mathrm{G}} = -k_{\mathrm{B}} \sum_{\nu} P(\nu) \ln P(\nu),$$

   where $\nu$ is a microstate.

   (i) For an isolated system (with fixed energy $E$, number of molecules $N$, and volume $V$) show that $S_{\mathrm{B}}$ and $S_{\mathrm{G}}$ are identical. Recall that in such a system, all allowed microstates are equally likely.

   (ii) For the remaining parts of this problem, consider a collection of $N$ indistinguishable particles arranged on a lattice of $M$ cells. Each cell can be occupied by at most one particle, and particles in different cells do not interact. Calculate the total number $\Omega(M, N)$ of possible configurations for this system.

   (iii) Assuming $M, N$, and $M - N$ are all very large, use Stirling's approximation to write the Boltzmann entropy per cell, $S_{\mathrm{B}}/M$, as a function of $f \equiv N/M$ alone.

   (iv) The occupation state of one cell in this lattice system is not affected by that of any other cell. As a result, the total entropy can be written as $S = M s_{\mathrm{cell}}$, where $s_{\mathrm{cell}}$ is the entropy of a single cell. Using Gibbs' definition of entropy, calculate $s_{\mathrm{cell}}$ in terms of $p_1$ and $p_0$, the probabilities of finding a particular cell occupied or unoccupied, respectively.

   (v) By expressing $p_0$ and $p_1$ in terms of $f$, the fraction of occupied cells, demonstrate that the Boltzmann and Gibbs lattice gas entropies are the same.

   (vi) Show that for the low density lattice gas ($f \ll 1$),

   $$S \approx -k_{\mathrm{B}} V [\rho \ln(\rho v) - \rho],$$

   where $v$ is the volume of a single lattice cell, and $\rho = N/V = N/(Mv)$ is the density.

   (vii) From macroscopic thermodynamics we saw that

   $$dS = \frac{1}{T} dE + \frac{p}{T} dV - \frac{\mu}{T} dN,$$

   so

   $$\frac{p}{T} = \left( \frac{\partial S}{\partial V} \right)_{E,N}.$$

   Use the expression you just obtained for the entropy of a low density lattice gas to explicitly compute the partial derivative. Express your result as the familiar ideal gas law: $pV = N k_{\mathrm{B}} T$. (You may be more familiar with $pV = nRT$. Make sure you think through the distinction between the two forms if it's not immediately obvious to you.)

   (viii) An ideal gas is one with particles which do not interact. Is the lattice gas model whose entropy you found in parts (iii) through (v) an ideal gas? In other words, are there any interactions between the particles? Explain how your answer does or does not agree with your derivation of an ideal gas law in (vii).

2. **A polymer off a lattice.** There are many ways in which you might argue that the polymer model from last week's problem set is lacking. For one thing, you might complain that real polymers don't live on a grid; we should have let each step move in a random direction, not just along a Cartesian axis. You might also complain that polymers can stretch. Perhaps we should have allowed the size of each link between monomers to vary. In this problem we will explore a model of a Gaussian polymer (in three dimensions) which addresses both of these complaints. It's a model we'll refer back to in lectures as we discuss heat, work, and free energy.

   We denote the position of the $n+1$ monomers (still $n$ segments as in last week's lattice polymer) by $\boldsymbol{R}_0, \boldsymbol{R}_1, \ldots \boldsymbol{R}_n$ and define our coordinate system by putting the origin at the location of the initial monomer, so $\boldsymbol{R}_0 = 0$. The next monomer's position is randomly selected from the Gaussian distribution

$$P(\boldsymbol{R}_1) = \left(\frac{3}{2\pi\ell^2}\right)^{\frac{3}{2}} \exp\left(-\frac{3\left(\boldsymbol{R}_1\right)^2}{2\ell^2}\right). \tag{1}$$

   Observe that the bond between monomers 0 and 1 can be of variable length and can point in any direction in three-dimensional space, in contrast to Problem 1. Monomer 2's position is similarly selected from a Gaussian, but now a Gaussian centered at the location of the second monomer:

$$P(\boldsymbol{R}_2|\boldsymbol{R}_1) = \left(\frac{3}{2\pi\ell^2}\right)^{\frac{3}{2}} \exp\left(-\frac{3\left(\boldsymbol{R}_2-\boldsymbol{R}_1\right)^2}{2\ell^2}\right).$$

   The left-hand-side is read as "the probability of monomer 2 position $\boldsymbol{R}_2$ given that monomer 1 is at $\boldsymbol{R}_1$." The pattern continues, so the position of monomer $i+1$ is randomly selected from

$$P(\boldsymbol{R}_{i+1}|\boldsymbol{R}_i) = \left(\frac{3}{2\pi\ell^2}\right)^{\frac{3}{2}} \exp\left(-\frac{3\left(\boldsymbol{R}_{i+1}-\boldsymbol{R}_i\right)^2}{2\ell^2}\right).$$

   The goal of this problem is to see that this model has the exact same root mean squared end-to-end distance as the simpler polymer in Problem 1. (The secondary goal is to become familiar with this model since it will come up in lecture.)

   (i) First, let's see the result by sampling polymers with a computer. Adapt your polymer sampling code from last week to generate random Gaussian polymers with $n$ segments. If you're using Python, you'll probably find the function np.random.normal to be useful. By generating a collection of many random polymers with the same length, plot for yourself a histogram for the $x$-component of $\boldsymbol{R}_n$. What is the mean of this histogram? In terms of $l$, what is the variance? Repeat these calculations for various $n$ from 1 to 100 and plot the variance of the $x$-component of $\boldsymbol{R}_n$ as a function of $n$. As in experimental plots, it is best to include error bars. Your error bars should indicate how much uncertainty there is due to the fact that you only sampled some of the possible polymer configurations.

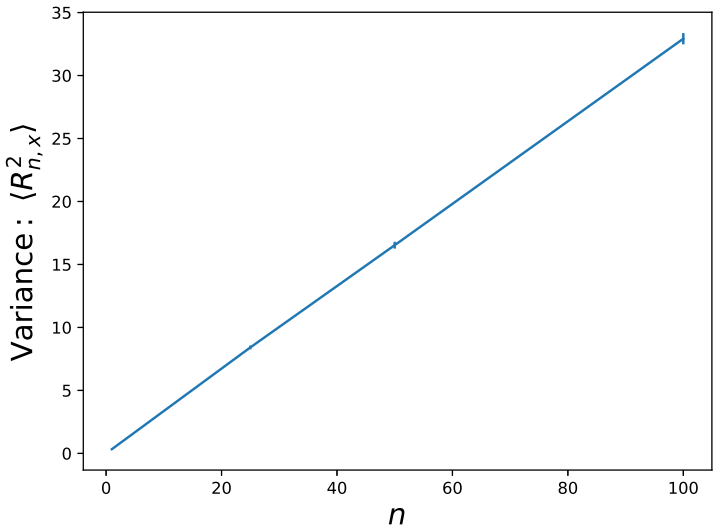
   It is a little more painful to include error bars. The simplest way to get error bars is to do $N$ independent experiments, compute the quantity of interest from each experiment, then find the standard deviation of the values from those $N$ experiments. If the experiments truly are independent of each other, then the standard deviation will drop like $1/\sqrt{N}$. Here, I generate 10,000 polymer samples then split those into 10 independent subsamples (independent experiments). The standard deviation I would expect from 10,000 samples is $1/\sqrt{10}$ times as much as the standard deviation I would expect from 1,000 samples. I am able to estimate this 1,000 sample standard deviation from my 10 subsamples, then rescale it to plot 10,000 sample error bars. Yes, it's tedious to do this. Yes, the error bars are very small in this case. But this type of manipulation will probably be something you'll want to use throughout graduate school when responsibly analyzing data.

```
ErrorBar_List = np.zeros(len(Num_Steps_List));

Num_Subsamples = 10;

for k in range(len(Num_Steps_List)):
  subsamples = np.split(np.transpose(x_distances)[k], Num_Subsamples);
  Subsample_Variances = np.zeros(Num_Subsamples)
  for m in range(Num_Subsamples):
    Subsample_Variances[m] = np.var(subsamples[m])
  ErrorBar_List[k] = np.sqrt(np.var(Subsample_Variances) \
                             / float(Num_Subsamples))

plt.errorbar(Num_Steps_List, Variance_List,yerr=ErrorBar_List);
plt.xlabel(r'$n$', fontsize=18);
plt.ylabel(r'$\mathrm{Variance: \ }\langle R_{n,x} ^2 \rangle$', fontsize=18);
plt.tight_layout()
plt.savefig('PS3P2p1MSDErrorBars.eps',format='eps',transparent=True);
from google.colab import files
files.download('PS3P2p1MSDErrorBars.eps')
```




(ii) In the last problem set, you found that $\langle R_n^2 \rangle = nl^2$. Compare this to your result from (i). Explain any notable similarities and differences between the two results.

(iii) To estimate an average value, we have used the average of a set of sampled polymers, but the true average involves an integration over all possible polymers. Typically such an integral cannot be performed analytically, but in the special case of Gaussian distributions, the integrals are solvable (as you explored on the previous problem set). Using those Gaussian integrals (you don't need to rederive these!), compute the root mean squared distance of the connection between monomer 0 and 1 (it will be the same between any monomers $i$ and $i+1$). You will probably want to write integrals in Cartesian coordinates, i.e.,

$$\begin{aligned}\langle R_1^2 \rangle &= \left(\frac{3}{2\pi\ell^2}\right)^{\frac{3}{2}} \int d\boldsymbol{R}_1 \ (\boldsymbol{R}_1)^2 \, e^{-\frac{3(\boldsymbol{R}_1)^2}{2\ell^2}} \\ &= \left(\frac{3}{2\pi\ell^2}\right)^{\frac{3}{2}} \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dy_1 \int_{-\infty}^{\infty} dz_1 \ (x_1^2 + y_1^2 + z_1^2) \exp\left(-\frac{3(x_1^2+y_1^2+z_1^2)}{2\ell^2}\right),\end{aligned}$$

with $\boldsymbol{R}_1 = (x_1, y_1, z_1)$. If you want, you can do integrals like this in Mathematica as well.

(iv) We now consider $n = 2$. Again, the first monomer is set to be at the origin, the second monomer is at $\boldsymbol{R}_1$ and the final monomer is at $\boldsymbol{R}_2$. In terms of $\boldsymbol{R}_1$, $\boldsymbol{R}_2$, and $l$, what is the (normalized) probability distribution for a microstate of the polymer $P(\boldsymbol{R}_1, \boldsymbol{R}_2)$? We said all microstates are equally likely, but you probably found that these are not. Why not?

(v) Suppose you only intend to measure the location of $\boldsymbol{R}_0$ (the origin) and $\boldsymbol{R}_2$ but will never explicitly observe $\boldsymbol{R}_1$. Consequently, you would like to know the *marginalized* distribution that averages over the possible values of $\boldsymbol{R}_1$:

$$P(\boldsymbol{R}_2) = \int d\boldsymbol{R}_1 \, P(\boldsymbol{R}_1, \boldsymbol{R}_2).$$

Perform that integration (Mathematica or citing last week's results are fine) to determine the marginal distribution $P(\boldsymbol{R}_2)$. Compare this distribution to the distribution in Eq. (1). What is different? What is the same?

(vi) A similar strategy could work for larger $n$. You could write down a Gaussian distribution for $P(\boldsymbol{R}_1, \boldsymbol{R}_2, \ldots \boldsymbol{R}_n)$, the probability of the microstate then get the marginal distribution for the final monomer position as:

$$P(\boldsymbol{R}_n) = \int d\boldsymbol{R}_1 \int d\boldsymbol{R}_2 \ldots \int d\boldsymbol{R}_{n-1} P(\boldsymbol{R}_1, \boldsymbol{R}_2, \ldots \boldsymbol{R}_n).$$

Doing so many Gaussian integrals could get painful, but the result must still be a Gaussian. Based on what you learned in part (i) and (ii) about the mean and variance of $\boldsymbol{R}_n$, determine this (normalized) Gaussian distribution $P(\boldsymbol{R}_n)$.

(vii) [Optional] Prove your answer to (vi) without asserting a priori that the distribution will be Gaussian.