1. **A simple model of a polymer.** You're no doubt aware that the shape a protein folds up into is influenced by energetic interactions between amino acids. You may not have given as much thought to the influence of entropy on a protein's shape. To focus on that concept, we consider the simplest model for conformational fluctuations of a long chain molecule, formed from n polymer segments connected end to end. Each segment $i = 1, 2, \ldots, n$ (perhaps comprising many chemical units) has a fixed length ℓ and an orientation $\hat{b}_i$ that is parallel to one of d Cartesian axes ($\hat{x}, \hat{y}$, or $\hat{z}$ in 3 dimensions). In other words, the molecular configuration traces a random walk on a $d-$dimensional cubic lattice:
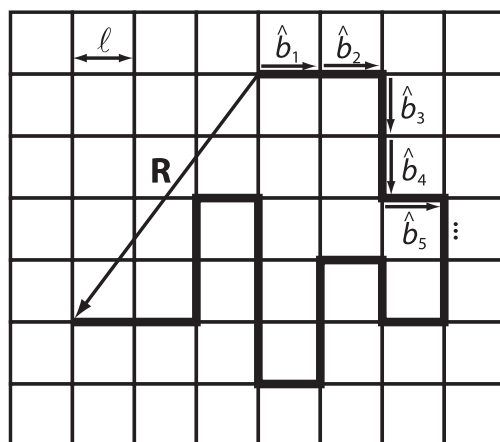
Imagine that the orientations of different segments are statistically independent, and that there is no preferred orientation, $\langle \hat{b}_i \rangle = 0$ and $\langle \hat{b}_i \cdot \hat{b}_j \rangle = \delta_{ij}$, where $\delta_{ij} = 1$ if $i = j$ and vanishes otherwise.

(i) Show that the entropy of such an ideal chain molecule has the form $S = k_{\mathrm{B}} n \ln a$. Determine the parameter a as a function of dimensionality d.

(ii) What is the expected value of $\boldsymbol{R}$, that is to say what is $\langle \boldsymbol{R} \rangle$, where $\boldsymbol{R} = \ell \sum_{i=1}^{n} \hat{b}_i$ is the end-to-end vector?

(iii) Is this the most likely value of $\boldsymbol{R}$? In other words, there is some probability distribution $P(\boldsymbol{R})$ describing the likelihood of every possible vector $\boldsymbol{R}$. Does $\langle \boldsymbol{R} \rangle$ coincide with the peak of that probability distribution?

(iv) Imagine n is very large and I use a computer program to take random steps on the lattice, thereby generating a single sample of the polymer. (I have provided a Mathematica notebook, ps3Mathematica.nb, that does such a simulation!) Do you think the sample will have sections that looked bunched up as though the polymer is partially "folded" or do you expect the polymer to be completely spread out and "unfolded"? Notice that there is nothing like a Coulomb attraction pulling any of the monomers closer together. The expected behavior of this model is purely entropic (due to counting the number of possibilities).

(v) Do you think my randomly generated polymer sample will start and stop at the exact same grid site? Rationalize your answer in light of your responses to (ii) and (iii).

(vi) Given your response to part (v), suggest a reason that the root mean squared end-to-end distance $\sqrt{\langle R^2 \rangle}$ would be a better measure of the size of the polymer than the average displacement vector $\langle \boldsymbol{R} \rangle$.

(vii) Calculate the mean squared end-to-end distance of the chain molecule, $\langle R^2 \rangle$. Your result should indicate that the typical distance between ends of the molecule grows with chain length as $\sqrt{\langle R^2 \rangle} \propto n^\nu$. Identify the exponent ν. How does your result depend on dimensionality d? [Hint: You should not need to use generating functions or other fancy tricks.]

(viii) For a (three-dimensional) polymer in "good" solvent, experiments yield $\nu \approx 3/5$. (A "good" solvent is one that prevents the molecule from collapsing onto itself, i.e., effective interactions among different segments are repulsive.) Compare this measured value with the one you calculated, and comment on the discrepancy.

2. **A slightly less simple model of a polymer.** There are many ways in which you might argue that the polymer model from Problem 1 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 future lectures as we discuss heat, work, and free energy.

We denote the position of the $n + 1$ monomers (still n segments as in Problem 1) by $\boldsymbol{R}_0, \boldsymbol{R}_1, \dots \boldsymbol{R}_n$ 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

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

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:

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

The pattern continues, so the position of monomer $i + 1$ is randomly selected from

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

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 I'll refer to it in class at a later date.)

(i) Using the Gaussian integrals from Problem Set 1 (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.,

$$\langle R_1^2 \rangle = \left(\frac{3}{2\pi\ell^2} \right)^{\frac{3}{2}} \int d\boldsymbol{R}_1 \; (\boldsymbol{R}_1)^2 \, e^{-\frac{3(\boldsymbol{R}_1)^2}{2\ell^2}}$$

$$= \left(\frac{3}{2\pi\ell^2} \right)^{\frac{3}{2}} \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dy_1 \int_{-\infty}^{\infty} dz_1 \; (x_1^2 + y_1^2 + z_1^2) \exp \left(-\frac{3(x_1^2 + y_1^2 + z_1^2)}{2\ell^2} \right),$$

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

(ii) We can view the polymer as being constructed by a series of steps. You start at monomer 0, then take a random step to get to monomer 1, etc. We saw in our study of random walkers that when the individual steps were *independent* and drawn from the identical distribution (in this case the Gaussian), we could write the cumulant generating function for a variable $X_T = \sum_{i=1}^{T} x_i$ in the form

$$\ln Z_T(\beta) = \ln \left\langle e^{-\beta X_T} \right\rangle = T \ln \left\langle e^{-\beta X_1} \right\rangle$$

where $\left\langle e^{-\beta X_1} \right\rangle$ was an expectation value taken after a single step (which was comparatively easy to compute). The consequence was that the cumulants after T steps could be expressed in terms of the cumulants after a single step. Use this idea to determine the root mean squared end-to-end distance of the Gaussian polymer, $\sqrt{\langle R_n^2 \rangle}$.

(iii) You should similarly be able to compute the first cumulant $\langle R_n \rangle$ for the n step Gaussian polymer. It turns out that all the other cumulants are exactly zero so that the distribution for the position of the nth monomer (remember monomer 0 is defined to be at the origin, so this is the same as the end-to-end vector) is a Gaussian. Write down that Gaussian probability density function $P(\boldsymbol{R}_n)$.

(iv) [Completely Optional!] One way to convince yourself that the distribution in (iii) must be a Gaussian is to compute the cumulant generating function, taking into account that each polymer step occurs in three dimensions. In place of the scalar β, let us use a vector $\boldsymbol{k} = (k_x, k_y, k_z)$ which has one component for each dimension:

$$\ln Z_n(\boldsymbol{k}) = \ln \left\langle e^{-\boldsymbol{k} \cdot \boldsymbol{R}_n} \right\rangle = n \ln \left\langle e^{-\boldsymbol{k} \cdot \boldsymbol{R}_1} \right\rangle .$$

Now $\ln Z_n(\boldsymbol{k})$ is a function of three variables (k_x, k_y, and k_z), and we could take derivatives with respect to any of these. Argue that the generating function gives access to the statistical properties of the x, y, and z components of $\boldsymbol{R}_n$:

$$\left. \frac{\partial^2 \ln Z_n(\boldsymbol{k})}{\partial k_x^2} \right|_{\boldsymbol{k}=0} = \left\langle x_n^2 \right\rangle , \quad \left. \frac{\partial^2 \ln Z_n(\boldsymbol{k})}{\partial k_y^2} \right|_{\boldsymbol{k}=0} = \left\langle y_n^2 \right\rangle , \quad \left. \frac{\partial^2 \ln Z_n(\boldsymbol{k})}{\partial k_z^2} \right|_{\boldsymbol{k}=0} = \left\langle z_n^2 \right\rangle .$$

In Problem Set 1, Problem 3, Part iv, you computed the moment generating function for a single variable Gaussian. The d-dimensional analog is

$$\left\langle e^{-\boldsymbol{k} \cdot \boldsymbol{R}_1} \right\rangle = \exp \left(\boldsymbol{k} \cdot \boldsymbol{\mu} + \frac{\sigma^2 k^2}{2d} \right) ,$$

where $\boldsymbol{\mu} = \langle \boldsymbol{R}_1 \rangle$ and $\sigma^2 = \left\langle R_1^2 \right\rangle$. Confirm this result then use it to find $\ln Z_n(\boldsymbol{k})$.

Confirm that this gives the same root mean squared end-to-end distance of the Gaussian polymer, $\sqrt{\langle R_n^2 \rangle}$ you calculated in (ii). Use your form for the cumulant generating function and the fact that $\left\langle R_n^2 \right\rangle = \left\langle x_n^2 \right\rangle + \left\langle y_n^2 \right\rangle + \left\langle z_n^2 \right\rangle$.

Finally, explain why the form of the cumulant generating function $\ln Z_n(\boldsymbol{k})$ ensures that all cumulants beyond the second cumulant will be zero [Hint: Problem Set 1, Problem 3, Part v].