Chemistry C3102-2006: Polymers Section

Dr. Edie Sevick, Research School of Chemistry, ANU

These notes are meant to supplement the lectures and to provide example problems. Homework problems will be provided separately As this an evolving draft, there may be some errors. I would appreciate it if you would provide me with feedback: point out errors, even simple grammatical errors; indicate which passages are unclear; and importantly, indicate where you feel there are missing "steps" in derivations.

1.0 The ideal chain: a statistical model

The simplest possible model of a chain is a series of "bonds" which are linearly and freely joined into a chain and may intersect itself any number of times. By freely jointed, we mean that the angle between any two consecutive bonds is completely free and unhindered, unlike consecutive carbon single bonds in polyethylene where the valence angles are restricted to around 109° and the torsional angles are localised around three different values. If each bond is denoted by the vector $\mathbf{a}$, then the end-to-end vector, $\mathbf{R}$ of such a freely jointed chain (FJC) model is

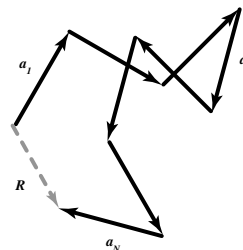$$\mathbf{R} = \sum_{i=1}^{N} \mathbf{a}_i. \qquad (1)$$

Figure 1: Schematic of a Freely Jointed Chain Model (FJC) in 2-dimensions. Unit "bond" vectors, designated by $\mathbf{a}_i$ are placed head to tail consecutively and without orientational bias. The end-to-end vector of the chain, $\mathbf{R}$, is given by the dotted grey line.

For a freely fluctuating chain the average or mean end-to-end vector, $\langle \mathbf{R} \rangle$, will be $\mathbf{0}$, as the chain can traverse in the positive and negative directions with equal probability. For example, if we label the ends of a chain A and B, then for any chain conformation with an end-to-end vector $\mathbf{R}_{AB}$, there will exist an identical conformation whose end-to-end vector is $\mathbf{R}_{BA} = -\mathbf{R}_{AB}$, and the probability of occurence of these two chains is equivalent. Thus, $\langle \mathbf{R} \rangle$ does not serve as a good measure of the chain size as it will be zero irrespective of the number of monomers, N, in the chain. Another measure is the mean of the square of the end-to-end distance, or simply the "mean square end-to-end distance", denoted by $\langle R^2 \rangle \equiv \langle \mathbf{R} \cdot \mathbf{R} \rangle$. This is not a vector quantity, but rather a scalar - it is the magnitude of the end-to-end vector. For consistency in dimensions, we will refer to the size of the polymer as R, where $R = \langle \mathbf{R} \cdot \mathbf{R} \rangle^{1/2}$. Because R is the mean *magnitude* of the vector, it is not zero but rather a finite number.

We can derive a simple formula from eq. 1 for the mean square end-to-end distance, R^2, as

$$R^2 = \langle \mathbf{R}^2 \rangle = \sum_{i}^{N} \sum_{j}^{N} \langle \mathbf{a}_i \cdot \mathbf{a}_j \rangle. \qquad (2)$$

We can rewrite the sums as

$$R^2 = \sum_{i=1}^{N} \langle \mathbf{a}_i \cdot \mathbf{a}_i \rangle + 2 \sum_{i}^{N} \sum_{j>i}^{N} \langle \mathbf{a}_i \cdot \mathbf{a}_j \rangle. \qquad (3)$$

You can understand this equation as the sum of all elements in a two dimensional matrix of elements $\mathbf{a}_i \cdot \mathbf{a}_j$: the first term is the sum of all elements on the diagonal and the second sum is the sum of the rest of the matrix terms, noting that the sum of terms above the diagonal is equal to the sum of terms below the diagonal. Remember that the dot product of two vectors $\mathbf{x}$ and $\mathbf{y}$, or the projection of $\mathbf{x}$ onto $\mathbf{y}$, is simply the product of the magnitudes of the vectors times the cosine of the angle between the vectors: $\mathbf{x} \cdot \mathbf{y} \equiv \mid \mathbf{x} \mid\mid \mathbf{y} \mid \cos\theta$. Thus, eq. 3 is

$$R^2 = \sum_{i=1}^{N} a^2 + 2 \sum_{i=1}^{N} \sum_{j>i}^{N} \langle \mathbf{a}_i \cdot \mathbf{a}_j \rangle \tag{4}$$

or

$$R^2 = Na^2 + 2 \sum_{i=1}^{N} \sum_{j>i}^{N} \langle \mathbf{a}_i \cdot \mathbf{a}_j \rangle. \tag{5}$$

Let's focus upon eq. 5. For a FJC chain, we said that consecutive bonds are completely uncorrelated; that is, the bonds can adopt *any* angle θ with equivalent probability. This means that $\langle \mathbf{a}_i \cdot \mathbf{a}_j \rangle \equiv a^2 \langle \cos\theta_{ij} \rangle$. And as the angle between $\mathbf{a}_i$ and $\mathbf{a}_j$ can take on any angle in the range $-\pi < \theta_{ij} \leq \pi$ with equal probability, then $\langle \cos\theta_{ij} \rangle = \int_{-\pi}^{\pi} \cos\theta_{ij} d\theta_{ij} = 0$. Thus, for a freely jointed chain

$$R^2 = Na^2. \tag{6}$$

This is an important result as it tells you that the size of a FJC scales as the square root of the number of monomers, $R \sim \sqrt{N}$. The symbol $\sim$ should be read as "scales as" and eliminates the need to include constants, as for example, the size of the bond, a.

For a jointed chain whose consecutive bonds *are* correlated, eq. 5 provides us with a simple expression for the chain size. Imagine that the chain is not freely jointed (we often refer to this as a chain which is not fully flexible) and, for simplicity, confined to two dimensions. Let the angle between any two adjacent bonds i and $i+1$ be $\theta_{i,i+1} = 109° + \delta$, the carbon-carbon valence angle plus a small angle variation which depends upon the temperature. As you know, δ is very small in comparison to 109° and consequently $\langle \mathbf{a}_i \cdot \mathbf{a}_{i+1} \rangle = a^2 \times \cos(109°) \neq 0$. The average angle between the bonds i and $i+2$ will also be fairly fixed by the valence angles; however the fluctuations in the $\theta_{i,i+2}$ will roughly twice that of $\theta_{i,i+1}$, but still a small angle. However, the fluctuations in the angle between bond i and $i+m$, $\theta_{i,i+m}$, become greater as m becomes larger, until at a sufficiently large m, these fluctuations far exceed those in the valence angle. In other words, the orientational correlation between bonds is "lost" or "forgotten" for bonds which are widely separated along the backbone of the chain. We say that in the limit of large m, $\langle \mathbf{a}_i \cdot \mathbf{a}_{i\pm m} \rangle = 0$. The "plus or minus" indicates that our argument is independent of the direction along the chain, and depends only upon the number of bonds, $m-1$, separating the bonds. Thus, the distance along the backbone over which the bonds are correlated is simply the distance that is spanned by $2m$ consecutive bonds, where m is the smallest number for which $\langle \mathbf{a}_i \cdot \mathbf{a}_{i\pm m} \rangle = 0$. This distance over which a 2- or 3-dimensional chain is inflexible or "stiff" is characterised by the persistence length, ℓ_p, and is defined by

$$\ell_p \equiv a + \frac{1}{a} \sum_{j>i}^{N} \langle \mathbf{a}_i \cdot \mathbf{a}_j \rangle. \tag{7}$$

Here you see that for the FJC, the persistence length is simply the length of a single bond, a. As the orientational correlation persists over larger and larger number of consecutive bonds, the persistence length increases. In the mathematical limit where there is absolutely *no* fluctuations in the valence angle, $\delta = 0$, then a 2-dimensional chain will be completely rigid (*i.e.,* a rigid rod) over its entire contour length, Na, and the persistence length and contour length are identical. The persistence length of the chain is determined by the chemistry of the individual chemical monomers as well as how these monomers interact with surrounding small solvent molecules. The persistence length of polystyrene is roughly a nanometre and corresponds to the length of 4 to 5 backbone bonds. In contrast, *ds*-DNA has a persistence length on the order of 50 nanometres, corresponding to about 150 base pairs.

When the definition of the persistence length, eq. 7, is inserted into eq. 5, an important conclusion results. Rearranging eq. 7 leads to $\sum_{j>i}^{N}\langle \mathbf{a}_i \cdot \mathbf{a}_j \rangle = (\ell_p - a)a$. Replacing the second sum of the *LHS* of eq. 5 yields

$$R^2 = Na^2 + 2\sum_{i=1}^{N}(\ell_p - a)a, \tag{8}$$

and as the argument of the sum (*i.e.* $(\ell - a)a$) is a constant, this reduces to

$$R^2 = Na^2(2\frac{\ell_p}{a} - 1) \tag{9}$$

Eq. 9 conveys much information. It shows that for the general model of a jointed chain with orientational correlation, $R^2 \sim Na^2$, as the term in the brackets is a constant depending upon the persistence length. This is important as we can now say that the FJC model, as well as its more general model that takes account orientational correlation, $R^2 \sim Na^2$. However for the FJC, $R^2 = Na^2$ where N is the number of bonds and a is the magnitude of each bond.

More importantly, we can show that any ideal chain has a size which can be described using a formula very similar to $R^2 = Na^2$, whether the chain is freely jointed or is a jointed chain with orientational correlation. We can demonstrate this by defining N_{stat} and a_{stat} as the number and size of statistical units within the chain and representing the chain of N bonds of size a by these statistical quantities. The size of a statistical unit is simply twice the persistence length, $a_{stat} = 2\ell_p$ and as the chain is uncorrelated over the persistence length, we can represent a series of correlated bonds by a single "statistical" link of size a_{stat}. These statistical segments are linked in a freely jointed manner and consequently, we can say that the size of the chain is $R^2 = N_{stat}a_{stat}^2$. The size of a statistical segement, a_{stat}, is often referred to as the "Kuhn step length" and N_{stat} as the number of Kuhn step lengths in the chain. Often times, we drop the subscript and simply write $R^2 = Na^2$ and assume that we are talking about Kuhn or statistical units. In this way, we do not need to distinguish between a FJC chain or one with orientational correlation, only that the chain is ideal.

• **Give examples of specific polymers which are fully flexible, semi-flexible, and "rigid-rod"-like.**

Polyethylene is a fully flexible polymer. Chain flexibility results from rotational isomerisation. That is, while the valence angles are 109° plus or minus a small fluctuation, the torsional angles can rotate and take on three different values, yielding (for a simple butane molecule) trans, gauche$^+$, and gauche$-$ conformations.

$$-CH_4 - CH_4 - CH_4-$$

Polyethylene

DNA is an example of a semi-flexible polymer. Alternatively, you can substitute double or triple bonds into a fully flexible carbon backbone chains to reduce the backbone flexibility or increase the persistence length. An obvious construction would be

$$-CH = CH - CH = CH - CH =$$

By placing placing double bonds in the backbone, you are restricting flexibility by removing rotational isomerisation.

Polyacetylene is a "rigid-rod" like polymer.

$$-C \equiv C - C \equiv C-$$
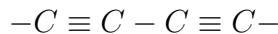
Polyacetylene

The degree of flexibility is characterised by the persistence lengths. The persistence lengths, from small (flexible) to large (inflexible) are ordered from polyethylene, DNA, to polyacetylene. Synthetic chemists make stiff or inflexible polymers by incorporating double and triple carbon-carbon bonds in their backbone, and bulk side groups groups that sterically hinder chain flexibility. .

"Concatenate" means to "link together". A concatemer of DNA is made of repeated sequences linked end to end. Consider λ-DNA: this is commercially available "standard" biomolecule that is isolated from the *E. Coli* bacteriophage lambda. It is a linear, double-stranded molecule that is 48,502 bp (base pairs) in length. The molecule also contains 10-16 base single-stranded regions at its ends that are self-complementary and these serve to circularise the molecule. In its circular form, λ-DNA can replicate continuously, forming concatemers of the λ-DNA. How does the size of the linear concatemer scale with the number of λ-DNA sequences?

Assuming that the ds-DNA molecule obeys ideal, or random-walk statistics, the mean square end-to-end distance varies with the number of monomers, or the number of λ-DNA sequences

in the concatemer. Let C be the number of λ-DNA sequences in the concatemer, such that a concatemeter with $C = 1$ has 48 kbp, $C = 2$ has 72 kbp, $C = 3$ has 144 kbp, $C = c$ has $48 \times c$ kbp. Then $\langle R^2 \rangle \sim c$

- **Consider that the fluctuations in valence angles are zero, $\delta = 0$. What is the size of the statistical segment of a polyacetylene having N chemical monomers each of size a? Compare your answer with the size of the statistical segment of a polyethylene chain with N chemical monomers**

The angle between consecutive carbon-carbon bonds in polyacetylene is 180° plus some fluctuation, say δ. If you assume $\delta = 0$, *i.e.* there are no fluctuations in the chain, then the chain is a rigid rod whose end-to-end distance is the contour length. In this case, the statistical segment is identical to the contour length of the chain, Na. Realistically, $\delta \neq 0$; that is, there will always be some fluctuation in the bond angles. You can determine the average magnitude of this fluctuation by considering the energy penalty of the fluctuation, relative to $k_B T$. As δ can be small compared to 180°, this fluctuation contributes negligibly to the overall size of short polyacetylene chains; however, the longer the chain, the more these fluctuations influence the size and shape of the chain. As the average magnitude of δ increases, the size of the statistical segment becomes smaller and the chain is more flexible.

In contrast, polyethylene remains flexible, even if are no fluctuations in the valence angles. This is because of fluctuations in the torsional angles, or rotational isomersiation.

- - **How dilute must a polymer solution be in order that each polymer chain be non-intersecting?**

In other words, we want to know the highest concentration at which the chains do not intersect other chains. First, we need to decide how to specify concentration. There are many ways of specifying concentration, but a particularly easy one is the volume fraction of statistical units within the spherical volume which contains the chain: $\phi = N\nu/V$, where N is the number of statistical units in the chain, ν is the volume of a single statistical unit and V is the spherical volume that circumscribes or encloses the single chain. Second, the sought concentration is found by adding as many of these non-intersecting chains as possible to the solution; or in other words, to let every point within the solution volume to be a point within a volume V containing a chain. Thus, the sought solution volume fraction, ϕ^*, is simply the volume fraction inside the volume circumscribing a single chain, ϕ. So to complete the problem, all we need to do is express
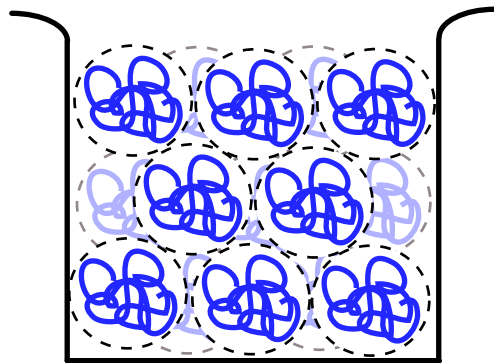
Figure 2: Schematic of a container of ideal polymer chains at a semi-dilute concentration, *i.e.*, at the highest concentration at which the polymers are isolated single chains and do not interpenetrate each other. The size of an isolated, single chain is given by $R \sim a\sqrt{N}$, and the dotted lines identify spheres of this radius that circumscribe the ideal chains. Higher concentrations require that the chains interpenetrate, i.e. the circumscribing spheres would intersect.

ϕ in terms of characteristics of an individual chain: N and a. The volume of a statistical segment is that of a cylinder of length a and diameter d, or $\nu \sim ad^2$. The spherical volume containing the chain $V \sim R^3$ where $R^2 = Na^2$. Thus

$$\phi^* = \phi = \frac{Nad^2}{a^3 N^{3/2}} \sim \left(\frac{d}{a}\right)^2 \frac{1}{\sqrt{N}}.$$

Now the ratio a/d varies between 1 and 50. $a/d = 1$ corresponds to a very flexible chain as the statistical segment is as short as it can be. $a/d \sim 50$ corresponds to a relatively stiff polymer, such as double stranded DNA. This result is important in that it says that ϕ^* decreases as the size of the chain increases and as the chain becomes more stiff or less flexible. or in other words, the larger the chain, N, the fewer the number of molecules that will just fill the container without overlapping.
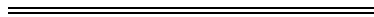
1.1 How the end-to-end distance represents the size of the chain

So far, we have very loosely related the magnitude of the end-to-end vector of the chain with the size of the chain. An often-used measure of the size of an odd-shaped or fluctuating object is the radius of gyration, s. In this section we show that $\langle s^2 \rangle = \langle R^2 \rangle / 6$ or simply that the mean square radius of gyration of an ideal chain is equivalent, to within a small constant, the mean square end-to-end distance of the chain. Thus $\langle \mathbf{R} \cdot \mathbf{R} \rangle$ is an appropriate measure of the chain's size. *The derivation below is optional reading.*

The definition of the radius of gyration, s, is the mass-weighted root mean square (rms) average of the magnitudes of the vectors leading from the center of mass to the points of mass making up the object. For our fluctuating ideal chain, we can simplify this definition by assuming that the chain is comprised of N "beads" linked by $N-1$ statistical segments, where these beads contain the equally distributed mass of the chain. The center of mass of the chain, $\mathbf{R}_{com}$, is defined as

$$\mathbf{R}_{com} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{r}_i \tag{10}$$

where $\mathbf{r}_i$ is the vector position of the ith bead. The square of the radius of gyration is defined as

$$s^2 = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{r}_i - \mathbf{R}_{com})^2. \tag{11}$$

From these two definitions, it is easy to show that $\langle s^2 \rangle \sim \langle R^2 \rangle$. First, insert eq. 10 into 11,

$$s^2 = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{r}_i - \frac{1}{N} \sum_{j=1}^{N} \mathbf{r}_j)^2, \tag{12}$$

and rewrite $\mathbf{r}_i$ as $\frac{1}{N} \sum_{j=1}^{N} \mathbf{r}_i$, to yield

$$s^2 = \frac{1}{N} \sum_{i=1}^{N} (\frac{1}{N} \sum_{j=1}^{N} \mathbf{r}_i - \frac{1}{N} \sum_{j=1}^{N} \mathbf{r}_j)^2. \tag{13}$$

This reduces to

$$s^2 = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} (\mathbf{r}_i - \mathbf{r}_j)^2, \tag{14}$$

or

$$s^2 = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} (\mathbf{r}_{ij})^2, \tag{15}$$

where $\mathbf{r}_{ij} \equiv \mathbf{r}_i - \mathbf{r}_j$ is the vector between the ith and j^{th} beads. Eq. 15 is also an alternative definition of the s^2: it is the mass-weighted sum of the magnitudes of the vectors between all mass points making up the object. As our object is not rigid, but a fluctuating chain, we can re-express eq. 15 in terms of the *mean* square radius of gyration and the *mean* square distances within the chain

$$\langle s^2 \rangle = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \langle \mathbf{r}_{ij}^2 \rangle. \tag{16}$$

Now the final part of the proof is very simple: all we need to do is insert two series identities into eq. 16 . First $\langle \mathbf{r}_{ij}^2 \rangle = \mid i - j \mid a^2$. This is simply restating $\langle R^2 \rangle = Na^2$ for a portion of the chain spanning beads i to j, rather than the whole chain. The absolute value signs ensure that the number of segments and the square distance is positive. With this substitution, we find

$$\langle s^2 \rangle = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \mid i - j \mid a^2, \tag{17}$$

which we can rewrite as

$$\langle s^2 \rangle = \frac{2a^2}{N^2} \sum_{i=1}^{N} \sum_{k=1}^{N-i} k \tag{18}$$

where we have simply summed over $j > i$ and substituted $k = j - i$. The two series of interest are

$$\sum_{k=1}^{m} k = \frac{m(m+1)}{2} \tag{19}$$

and

$$\sum_{k=1}^{m} k^2 = \frac{m(m+1)(2m+1)}{2}. \tag{20}$$

Inserting the first series identitify for the second sum on the RHS of eq. 18 yields

$$\langle s^2 \rangle = \frac{2a^2}{N^2} \sum_{i=1}^{N} \frac{(N-i)(N-i+1)}{2} \tag{21}$$

or, multiplying this out,

$$\langle s^2 \rangle = \frac{a^2}{N^2} \sum_{i=1}^{N} (N^2 - 2Ni + i^2 + N - i). \tag{22}$$

Applying the both series, eqs. 19 and 20 to each term on the RHS of eq. 22 gives

$$\langle s^2 \rangle = \frac{a^2}{N^2}[N^3 - \frac{2N^2(N+1)}{2} + \frac{(N(N+1)(2N+1)}{2} + N^2 - \frac{N(N+1)}{2}] = \frac{a^2}{N^2}[N^3 + N^2]. \tag{23}$$

The result is $\langle s^2 \rangle = (N+1)a^2$, for a chain of N bonds linking $N+1$ beads. So to conclude, the mean square end-to-end distance, $\langle \mathbf{R}^2 \rangle \sim Na^2$ is proportional to the mean square radius of gyration, a standard measure of the size of an object.

<u>*EXAMPLE PROBLEM*</u>

•• **What is the radius of gyration of a ds-DNA molecule if its contour length is 20 μm**

The contour length of a chain having N_{stat} number of statistical units, each of size a_{stat} is $N_{stat} \times a_{stat}$. The statistical monomer size is twice the persistence length, $a_{stat} = 2\ell_p$. The mean square end-to-end distance of the chain is $< R^2 >= N_{stat} a_{stat}^2$ and the radius of gyration

is $< s^2 >=< R^2 > /6$. The persistence length of ds-DNA is 50nm. Thus, the natural size is given by the radius of gyration, or

$$
\begin{aligned}
< s^2 > \ &= \ \frac{1}{6} < R^2 > \\
&= \ \frac{1}{6} N_{stat} a_{stat}^2 \\
&= \ \frac{1}{6} (N_{stat} a_{stat}) \times a_{stat} \\
&= \ \frac{1}{6} (20\mu m) \times 2\ell_p \\
&= \ \frac{20\mu m}{3} \times 50nm,
\end{aligned}
$$

or $< s^2 >= 1/3\mu m^2$ or $\sqrt{< s^2 >} = 1/\sqrt{3}\mu m$

1.2 Gaussian statistics to describe the probability of finding a chain of a given end-to-end distance

In Section 1.0, we derived a scaling relationship for the mean size of a single chain. However, the chain is constantly fluctuating, so we need to know how likely we can expect to observe the chain possessing this particular size. In other words, we need to know the full distribution of chain sizes, and not just the mean size. As you will see, the distribution of chain sizes allows us to make thermodynamic-based predictions of how an ideal chain behaves when stretched or compressed. In this section, we will derive an expression for the probability distribution of the square end-to-end distance, R^2, of an ideal chain.

First, for simplicity, we will consider a chain in one-dimensional space, only. (After we finish the derivation, we will extend our result to the usual three-dimensional space, but by limiting ourselves to one dimension, we greatly reduce the number of equations that we must write.) One end of the chain is tethered to $x = 0$, and the equi-sized links point either the right or to the left, with equivalent probability. The problem is similar to a drunken walker in one dimension: starting at the origin, the walker takes a random, unit-sized step to the right or to the left. We want to know the probability that, after N equi-sized steps, the random walker will arrive at a distance m steps from his starting position. Let n^+ be the number of links or "steps" that move to the right and n^- be the number that move to the left, such that $N = n^+ + n^-$. There will be 2^N possible trajectories, each of them equally likely but with many of them "arriving" at the same place m. The arrival destination $m = n^+ - n^-$, depends only upon the number of right-steps and left-steps and NOT the ordering of these steps in the trajectory. Out of the 2^N trajectories, the number of trajectories that are comprised of precisely n^+ right-steps and n^- left-steps is $N!/(n^+!n^-!)$. Since knowing the number of right-steps and left-steps implies that you know the destination, then the number of trajectories that arrive at destination m is

$$\frac{N!}{(\frac{N+m}{2})!(\frac{N-m}{2})!}. \tag{24}$$

Therefore the probability of the random walker taking N steps and arriving at m is simply the number of trajectories that arrive at m divided by the total number of trajectories, or

$$p(m, N) \equiv \left(\frac{N!}{(\frac{N+m}{2})!(\frac{N-m}{2})!} \right) / 2^N. \tag{25}$$

(This discretised distribution is normalised, *i.e.*, $\sum_{(n^+,n^-)} p(m, N) = 1.0$ You might also recognise this from the Binomial Theorem

$$(\frac{1}{2})^N \sum_{(n^+,n^-)} \frac{N!}{n^+!n^-!}(x^{n^+} y^{n^-}) = (\frac{1}{2})^N (x + y)^N. \tag{26}$$

If $x = 1$ and $y = 1$, the *LHS* is simply the sum over all $p(m, N)$, and the *RHS* is unity.) To simplify eq. 25, apply Stirling's formula,

$$n! \approx \sqrt{2\pi} n^{n+1/2} exp(-n), \tag{27}$$

to get

$$p(m, N) = (\frac{2}{\pi N})^{\frac{1}{2}} \exp{(-\frac{m^2}{2N})}. \tag{28}$$

Now, for a discrete random walker, the destination does not take on all integer values between 0 and N: if N is odd, then the walker will arrive at destinations that are only odd (*i.e.*, m is odd), and if N is even, then m is restricted to take on only even values. However, this is not captured in the discrete distribution $p(m, N)$, eq 28. To get this into a continuous form, let ℓ be the size of a step in 1-dimension, so that the destination is given not by the integer number, m, but rather the continuous, real distance x where $x = m\ell$. Now, the trajectories counted for a specific value of m in the discrete distribution $p(m, N)$ are now accounted in the continuous distribution $p(x)$ from $x = (m-1)\ell$ to $x = (m+1)\ell$, or $p(x)\Delta x = p(m, N)$ where $\Delta x \equiv 2\ell$. Thus, the continuous distribution for one dimension is

$$p(x) = (2\pi N\ell^2)^{-\frac{1}{2}} \exp\left(-\frac{x^2}{2N\ell^2}\right). \tag{29}$$

Having developed this probability distribution in one-dimension, we can now extend this to three dimensions. Let a single step be comprised of a random step of size ℓ in each of the x, y, and z-directions, and the destination be $R = \sqrt{x^2 + y^2 + z^2}$.

$$
\begin{aligned}
p(R) &= p(x) \times p(y) \times p(z) \\
&= (2\pi N\ell^2)^{-\frac{1}{2}} \exp\left(-\frac{x^2}{2N\ell^2}\right) \times (2\pi N\ell^2)^{-\frac{1}{2}} \exp\left(-\frac{y^2}{2N\ell^2}\right) \times (2\pi N\ell^2)^{-\frac{1}{2}} \exp\left(-\frac{z^2}{2N\ell^2}\right).
\end{aligned}
$$

The size of a singe step in three dimensions is therefore $a = \sqrt{a^2} = \sqrt{(\pm\ell)^2 + (\pm\ell)^2 + (\pm\ell)^2} = \sqrt{3}\ell$, so that the distribution function is

$$p(R) = \left(\frac{3}{2\pi Na^2}\right)^{-\frac{3}{2}} \exp\left(-\frac{3R^2}{2Na^2}\right) \tag{30}$$

This is, of course, a Gaussian distribution, named after the famous mathematician K.F. Gauss (1777-1855) who studied the distribution of a sum of random numbers. The Gaussian distribution is exact only in the limit of $N \to \infty$; however, it is very accurate for smaller values of N. In its general form, the normal or Gaussian distribution is

$$p(x) = \left(\frac{d}{2\pi\sigma^2}\right)^{d/2} \exp\left(-\frac{d(x - \langle x \rangle)^2}{2\sigma^2}\right) \tag{31}$$

where d is the dimension, $\langle x \rangle$ is the mean, and σ^2 is the variance (or σ is the standard deviation that characterises the width of the distribution).

EXAMPLE PROBLEMS

● **When reporting an experimental result, researchers normally repeat the experiments n times, sum the measurements and take an average. Assuming that the error in experimental measurement follows a normal distribution, explain why (a) the error in the sum of measurements is proportional to $n^{1/2}$ and not n, and (b) the error in the average scales as $n^{-1/2}$.**AIM: to demonstrate use of Gaussian distribution in collection of experimental data

(a) Each measurement contains some error. So when you sum up the measurements, you are also summing up the errors contained within. However, the errors are positive as well as negative, similar to the steps of a 1-dimensional random walker. The sum of the magnitude of

errors is analogous to the random walker's steps, or the destination. So, if the mean destination or $\sqrt{< R^2 >}$ scales with number of steps according to $N^{1/2}$, then analogously, the sum of the errors scales with the number of measurements according to $n^{1/2}$. If the error was only negative, or only positive, then we would expect the sum of the errors to scale with n^1.

(b) To find the average measure, you take the sum of measurements and divide by the number of measurements. So therefore, the error in the average is the error in the sum of the measurements (which scales as $n^{1/2}$ divided by the number of measures, n, More concisely, the error in the average scales as $n^{1/2}/n \equiv n^{-1/2}$

• **A child builds a lattice "sculpture" by randomly gluing together ends of 120 identically sized sticks. For his first structure, the sticks are glued onto paper (*i.e.* it is two dimensional), but his later sculpture is a 3-dimensional random network. Two thousand children Canberra-wide complete similar sculptures. Assuming that the sticks are glued together with arbitrary angle, plot the probabillity distributions for the end-to-end distance for the 2- and 3-dimensional structures that result.**

AIM: to apply expressions for Gaussian distributions to an experiment

• **Consider three linear, freely-jointed chains having the same number of steps, N, with monomers of the same size a, but in different dimensions. One chain is 1-dimensional, another is 2-dimensional, and the last is in 3-dimensions. Order the chains in terms of increasing**

1. **mean end-to-end distance, and**

2. **probability of finding largest end-to-end distance.**

Does your answer agree or contravene your intuition?

• **Which random walk sequence has a larger mean size: $N = 100$ steps of $\ell = 1$ in 1-dimension, or $N = 50$ steps of size $\ell = \sqrt{(2/3)}$ in each of 3 directions.**

The step size in 1-D is $a = \ell = 1$, so the mean square end-to-end distance of the 1-D chain is $Na^2 = 100$. The step size in 3-D is

$$a = \sqrt{(3\ell^2)} = \sqrt{(3 \times (2/3))} = \sqrt{2},$$

so the mean square end-to-end distance of the 3-D chain is $Na^2 = 100$. Both chains share the same mean size.

• **Given a Gaussian chain (another name for an ideal chain) of N monomers, each of size a with one end of the chain tethered at an origin in 3-dimensional space, what is the probability that you will find the other, free end of the chain located a distance greater than $\sqrt{N}a$ from the tethered chain end ?**

The normalised probability distribution is

$$P(\mathbf{R}) = (\frac{3}{2\pi Na^2}) \exp{(-\frac{3R^2}{2Na^2})}.$$

That is, the probability of finding the ends of the chain separated a distance between $\mathbf{R}$ and $\mathbf{R}+d\mathbf{R}$ is $P(\mathbf{R})d\mathbf{R}$ or $P(\mathbf{R})d^3R$. So, to find the probability of observing the end-to-end distance *less than* $\sqrt{N}a$, we need to evaluate

$$I = \int_0^{\sqrt{N}a} P(\mathbf{R})d^3R$$

Here, d^3R represents a differential volume element. As we seek the probability of end-to-end *distances*, we want to recast this in terms of scalar distance R. This is simply done by noting that $d^3R \equiv d(\text{Volume}) = d(\frac{4}{3}\pi R^3) = 4\pi R^2 \times dR$. Thus, we need to evaluate

$$\int_0^{\sqrt{N}a} 4\pi R^2 P(\mathbf{R})dR,$$

or inserting the probability distribution

$$\int_0^{\sqrt{N}a} 4\pi R^2 (\frac{3}{2\pi Na^2}) \exp{(-\frac{3R^2}{2Na^2})}dR.$$

Now before turning to Maple, Mathematica, or tables of integrals, it is best to simplify this expression. I do that by collecting constants outside the integral and making a change in the integration variable. Rather than R, I will define a new integration variable $x \equiv R/(\sqrt{N}a)$. Substitution of this new integration variable yields (using $dx = dR/(\sqrt{N}a)$, and changing the integration limits from $(R = 0, R = \sqrt{N}a)$ to $(x = 0, x = 1)$)

$$
\begin{aligned}
I &= 4\pi(\frac{3}{2\pi Na^2})^{3/2} \int_0^1 x^2(Na^2)^{3/2} \exp{(-\frac{3x^2}{2})}dx \\
&= 4\pi(\frac{3}{2\pi})^{3/2} \int_0^1 x^2 \exp{(-\frac{3x^2}{2})}dx.
\end{aligned}
$$

To check to see if this last expression is correct, we should try to integrate x between 0 and ∞: as the probability distribution is normalised, this should result in unity. So, I type into Maple the above equation with the appropriate limits:

4*pi * (3/(2*pi))^(3/2)* int(x^2 * exp(-3*x^2/2), x=0..infinity);

and get the answer

$$\pi(\frac{1}{\pi})^{3/2}\sqrt{\pi},$$

13

which is 1. When you do this, you might get a another product of cancelling factors which, although the factors are different that shown here, they still reduce to unity. Now I can change the upper integration limit back to unity to find the probability that the end-to-end distance will be *less* than $\sqrt{N}a$. I type into Maple

<center>4*pi * (3/(2*pi))^(3/2)* int(x^2 * exp(-3*x^2/2), x=0..1);</center>

and get

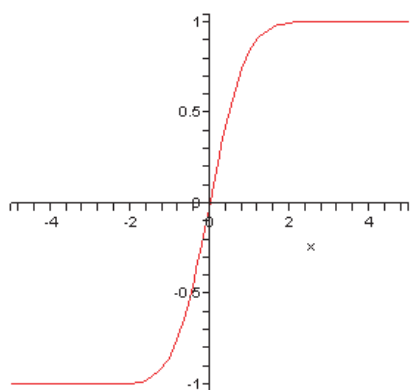$$3\pi\sqrt{3}\sqrt{2}(\frac{1}{\pi})^{3/2}(-\frac{1}{3}\exp{(-3/2)} + \frac{1}{18}\sqrt{\pi}\sqrt{6}erf(\frac{1}{2}\sqrt{6})).$$

Figure 3: Error Function, $erf(x)$ versus x, as plotted by Maple. Note that for large values of x or $-x$, the error function returns $x/\mid x \mid$.

erf is the error function; it is 1.0 or -1.0, for large positive or large negative arguments. To see the functional form of the error function. use the plot facility on Maple or any other similar program. In Maple you simply type: plot (erf(x), x = -5 . . 5); to get output shown in figure (3)

To get a numerical evaluation of this expression containing the error function, you can most simply rerun the integration within an "evaluate function" call: evalf(4*pi * (3/(2*pi))^(3/2)* int(x^2 * exp(-3*x^2/2), x=0..1)); This gives you $1.078316300/\sqrt{\pi}$; to get a number I type into Maple 1.07831630/ \sqrt(3.14159);, or alternatively I capitalise pi to "Pi", and get out $I = 0.6083750812$.

Thus, the probability of finding the ends of the chain separated a distance *less* than $\sqrt{N}a$ is 0.61. Therefore, the probability of finding the chain ends located a distance *greater* than $\sqrt{N}a$ is $1.00 - 0.61$ or

$$1.0 - I = 0.39 \tag{32}$$

• What is the probability of observing a chain in three dimensions whose end to end distance, R is within one standard deviation of the mean end-to-end distance?

••• What is the distribution of monomer-monomer distances of an ideal chain of N monomers whose first and last monomers are joined to form a ring?

First, consider a Gaussian ring chain where the monomers are labelled 1 to N. Let $\mathbf{R}_{mn}$ be the end-to-end distance between any two arbitrary monomers n and m within the Gaussian ring. Clearly $\mathbf{R}_{1,N} \equiv \mathbf{R} = 0$. Let $P_N(\mathbf{R}_{mn}; \mathbf{R} = 0)$ be the distribution of end-to end distances $\mathbf{R}_{mn}$

<center>14</center>

given that the freely jointed chain of N monomers is a ring. We can write an expression from this distribution in terms of the distribution of linear chains:

$$P_N(\mathbf{R}_{m,n}; \mathbf{R} = 0) = P_{n-m}(\mathbf{R}_{mn})P_{N-n+m}(\mathbf{R}_{mn})$$

where $P_{n-m}(\mathbf{R}_{mn})$ is the distribution of distances for a linear chain of $n - m$ monomers and $P_{N-n+m}(\mathbf{R}_{mn})$ is the distribution associated with a linear chain of $N - n + m$ monomers. This equation results from the fact that the sub-components of a Gaussian chain remain Gaussian, as long as they contain a large number of monomers.
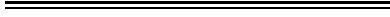
You can then show, by replacing the RHS of the above equation with Gaussian distributions (remembering that a linear chain of $n - m$ monomers has $\langle R^2 \rangle = (n - m)a^2$ and that a chain of $N - n + m$ monomers has $\langle R^2 \rangle = (N - n + m)a^2$) that

$$P_N(\mathbf{R}_{mn}; \mathbf{R} = 0) = \left(\frac{3}{2\pi\mu^2}\right)^{d/2} \exp\left(-\frac{3R_{mn}^2}{2\mu^2}\right),$$

where

$$\mu^2 = \mid m - n \mid a^2\left(\frac{N - m + n}{N}\right).$$

That is, for the Gaussian ring, $\langle R_{mn}^2 \rangle = \mu^2$.

Summary

So far, what we've done is construct a model of an Ideal chain, focussing upon its size and its distribution of sizes

- An ideal chain is self-intersecting, lacks interactions with itself and other chains and is modelled by a series of freely jointed links (referred to as a Freely Jointed Chain or FJC

- The mean size of a FJC follows the scaling relation $< R^2 >= Na^2$, where N and a are the number and size of monomers or links. If there were orientaitional correlation in the jointed chain, the scaling relation reduces to $< R^2 > \sim Na^2$ or $< R^2 >= (2\frac{\ell_p}{a} - 1)Na^2$. Alternatively, we can assume that N and a describe statistical monomers and describe all ideal chains using $< R^2 > \sim Na^2$

- The mean square radius of gyration of an ideal chain is $< s^2 >=< R^2 > /6$.

- An ideal chain's end-to-end distance follows a normal or Gaussian distribution. The spread in the distribution is on the order of the size, reflecting that the ideal chain fluctuates greatly.