A one-dimensional random walk model for polymer chains.

Consider a chain consisting of N segments each of length 1. Assume that the x-axis is in the direction of the chain, and let us place the beginning of the chain in the origin. We are interested in finding the probability that the end of the chain will be at a distance L from the origin, i.e. its coordinate along the x-axis will be L. L is also called the "end-to-end" distance. This probability equals the number of possible ways to arrange the chain so that its end is at the distance L (let's call it $\Omega(N,L)$) *divided* by the total number of all possible arrangements for a chain of length N (let's call it $\Omega(N)$).

$$P(N,L) = \Omega(N,L)/\Omega(N) \tag{1}$$

In a 1D chain, each segment can be considered as a vector of length 1 pointing in either positive (+) or negative (-) direction along the x-axis. If N_+ is the number of segments oriented in the "+" direction and N_- is the number of segments in the "–" direction for a given conformation of the chain, then the end-to-end distance for this conformation is

$$L = N_+ - N_- \tag{2}$$

Note that the total number of segments equals N:

$$N_+ + N_- = N. \tag{3}$$

From Equation (3) we obtain $N_- = N-N_+$, and substituting it for N- in Eq.(2) we get:

$$L = 2N_+ - N = 2(N_+ - \tfrac{1}{2} N) \tag{4}$$

This means that for a given N, L is directly determined by the number of "+" segments in the chain. Therefore, $\Omega(N,L)$ is the same as $\Omega(N, N_+)$, the number of possible ways to arrange N_+ plus-segments in a N-segment chain:

$$\Omega(N,L) = \Omega(N,N_+). \tag{5}$$

We assume that for each segment there is an *equal probability* to be oriented in the "+" or the "-" direction, *independent* of any other segment in the chain (a <u>random-walk</u> model). Naturally, we can expect that for the most probable conformation, $N_+ = N_- = \tfrac{1}{2} N$ (assuming N is a large number).

For convenience, we can label each segment by a "+" or a "–" depending on its orientation, so that each conformation of the chain is characterized by a string of alternating pluses and minuses, the total length of this string being N). For example, a chain of 5 segments with $N_+ = 2$ (hence $N_- = 3$) can be obtained if the following 10 ways:

```
( + + − − − )     ( − + + − − )     ( − − + + − )     ( − − − + + )
( + − + − − )     ( − + − + − )     ( − − + − + )
( + − − + − )     ( − + − − + )
( + − − − + )
```

(The relative arrangement of these 10 possible conformations does not matter: I grouped them merely for convenience of keeping track of where I put the pluses).

Then the problem of counting all possible conformations for a chain consisting of N_+ segments in the + direction and N_- segments in the – direction is the same problem as asking how many ways are there to distribute N_+ pluses over N positions in a string of N characters. (We don't have to worry about distributing the minuses – they will fill empty spaces automatically!). This is *exactly* the same problem as distributing N_+ balls over N empty boxes such that each box can be occupied by only one ball. We can start with a set of N empty boxes (analog = a string of all minuses). There are N possibilities to place the first ball. After it is placed (analog: a minus is replaced by a plus), there are now N-1 boxes available for placing the second ball. Similarly, after the first two balls were placed, there will remain N-2 possibilities to place the third ball and so on.
Then the total number of possibilities to place N_+ balls, taking them one by one, is a product of the number of possibilities we have for placing each ball, i.e. (let's call it $P_{N,N+}$)

$$P_{N,N+} = N*(N-1)*(N-2)*\ldots*(N-N_++1) = N!/(N-N_+)!$$

In the above mentioned case of N=5 and $N_+=2$ this gives: $5*4 = 5!/(5-2)!=5!/3!=20$. We are not done yet. Because all balls (or pluses) are *identical*, we have overestimated the number of possible arrangements. For example, the fist conformation, $(++---)$, has occurred in our method of counting twice:
as

([first ball],[second ball],[empty],[empty],[empty]) {or ([first +],[second +],[-],[-],[-])}

and as

([second ball],[first ball],[empty],[empty],[empty]) {or ([second +],[first +],[-],[-],[-])}.

And therefore it has been counted twice. These two conformations are **indistinguishable**! To correct the number we obtained for the fact that the balls (pluses) are all indistinguishable, we have to divide $P_{N,N+}$ by the number of possible permutations for N_+ objects, which is $N_+!$. So, the correct number of possible (distinguishable) arrangements of N_+ balls (pluses) among N places is

$$\Omega(N,N_+) = \frac{P_{N,N_+}}{N_+!} = \frac{N!}{N_+!(N-N_+)!} = \frac{N!}{N_+!\,N_-!} \qquad (6)$$

In the example above, we get $\Omega(N,N_+)= 5!/(2! *3!)=10$. Finally, we got the correct count!

Note that pluses are not better than minuses. You can do a similar counting job assuming that you distribute N_- minuses over a string of N spaces (initially filled with pluses). You will get the same result. Note that the end-result in Eq(6) is symmetric over N_+ and N_-, i.e. it will not change if you replace N_+ with N_- and N_- with N_+ at the same time.

Now let's do some math. Let's first express N_+ and N_- in terms of L (in the EC textbook they use x which is $L/2$). From Eq.(4) we get $N_+ = (N+L)/2$; and then from Eq.(3): $N_- = (N-L)/2$, so that

$$\Omega(N,L) = \Omega(N,N_+) = \frac{N!}{\left(\dfrac{N+L}{2}\right)! \left(\dfrac{N-L}{2}\right)!} \tag{7}$$

This expression is very compact but not very informative, isn't it? What we would like to do is to get rid of the factorials here. Let us introduce a new variable (let's call it Y):

$$Y = \ln\{\Omega(N,L)\} \tag{8}$$

Then we can always obtain $\Omega(N,L)$ from Y by an inverse transformation:

$$\Omega(N,L) = e^Y \quad . \tag{9}$$

Substituting Eq.(7) into Eq.(8) and using Stirling's approximation (lnN!=NlnN-N)

we get:

$$Y = N \ln N - N - \frac{N+L}{2}\ln\frac{N+L}{2} + \frac{N+L}{2} - \frac{N-L}{2}\ln\frac{N-L}{2} + \frac{N-L}{2}$$
$$\text{(the terms not containing the logarithm cancel!)}$$
$$= N \ln N - \frac{N+L}{2}\ln\frac{N+L}{2} - \frac{N-L}{2}\ln\frac{N-L}{2}$$
$$\text{(Note that } \ln\{(N+L)/2\} = \ln(N+L)-\ln2, \text{ etc.)}$$
$$= N \ln N - \frac{N+L}{2}\ln(N+L) - \frac{N-L}{2}\ln(N+L) + N \ln 2$$

(we actually don't care much about the terms that don't contain L but let's keep them for now)

$$= N \ln N + N \ln 2 - \frac{N+L}{2}\ln(N+L) - \frac{N-L}{2}\ln(N+L)$$

Here we take advantage of the fact that N is a big number and, as long as $L <<N$, L/N is a very small number. Let's call it ε, for simplicity:

$$L/N = \varepsilon. \tag{10}$$

We will use the following property of the logarithm (recall Taylor series expansion):

$$\ln(1+\varepsilon) = \varepsilon - \varepsilon^2/2 + O(\varepsilon^3).$$

3

Then $N+L = N(1+\varepsilon)$, and $\ln(N+L) = \ln N + \ln(1+\varepsilon) = \ln N + \varepsilon - \varepsilon^2/2$ (we will keep only terms up to the ε^2 and ignore higher powers of ε, as these are much smaller numbers.. Similarly, $\ln(N-L) = \ln N + \ln(1-\varepsilon) = \ln N - \varepsilon - \varepsilon^2/2$.

We then obtain:

$$Y = N \ln N + N \ln 2 - \frac{N}{2}\left[(1+\varepsilon)\left(\ln N + \varepsilon - \frac{\varepsilon^2}{2}\right) + (1-\varepsilon)\left(\ln N - \varepsilon - \frac{\varepsilon^2}{2}\right)\right]$$

The expression in the square brackets can be simplified as follows:

[…]= $\ln N + \varepsilon \ln N + \varepsilon + \varepsilon^2 - \varepsilon^2/2 - \varepsilon^3/2 + \ln N - \varepsilon \ln N - \varepsilon + \varepsilon^2 - \varepsilon^2/2 + \varepsilon^3/2 = 2\ln N + \varepsilon^2$ (nice and compact, isn't it?)

and then we get

$Y = N*\ln N + N*\ln2 - N*\ln N - N\varepsilon^2/2 = N*\ln2 - L^2/(2N)$. 　　　　　　　(11)

The last expression was obtained using the definition of ε (Eq.(10)). Recalling Eq.(9) we finally get:

$$\Omega(N,L) = 2^N e^{-\frac{L^2}{2N}}　.$$　　　　　　　(12)

This is the main result that we wanted to obtain. It states that for a random-walk 1D chain composed of N segments of length 1 **the number of possible conformations with the end-to-end distance L scales with L as $\exp(-L^2/(2N))$**.

Then the probability that a randomly picked conformation of such a chain has the end-to-end distance L is (recall Eq.(1))

$$P(N,L) = C_N e^{-\frac{L^2}{2N}},$$　　　　　　**(the final result!)**　　　　(13)

where C_N is a normalization factor that does not depend on L and can be determined from the condition that the sum of the probabilities over all possible states of a chain (i.e. over all possible L values) should equal 1: $\sum_L P(N,L) = \int P(N,L)dL = 1$.

For the 1D problem considered here, this condition gives (recall, L is a projection of the end-to-end vector): $C_N \sum_{L=-N}^{N} e^{-\frac{L^2}{2N}} \approx C_N \int_{-\infty}^{\infty} e^{-\frac{L^2}{2N}} dL = C_N \sqrt{2\pi N} = 1$; so $C_N = (2\pi N)^{-\frac{1}{2}}$

------------------For those who are curious or care about the numbers------------------------
Note that the total number of all possible conformations (independent of what the end-to-end distance is) of a 1D chain of N segments is $\Omega(N) = 2^N$. You can easily obtain this number by realizing that each segment can be in two states: either in the + or in the – direction.

If we put this result, together with Eq(12), into Eq(1), we get Eq(13) with $C_N=1$. This C_N is different from the correct normalization: this is due to an oversimplified form for the Stirling's approximation that we used to derive $\Omega(N,L)$: we have ignored terms containing $N^{-\frac{1}{2}}$. The most accurate form of the Stirling's approximation is $N! = N^{N+\frac{1}{2}} e^{-N} (2\pi)^{\frac{1}{2}}$ (this is before taking the logarithm!): it should recover the factor $(2\pi N)^{\frac{1}{2}}$ in $\Omega(N,L)$ that we lost when using a less accurate formula.