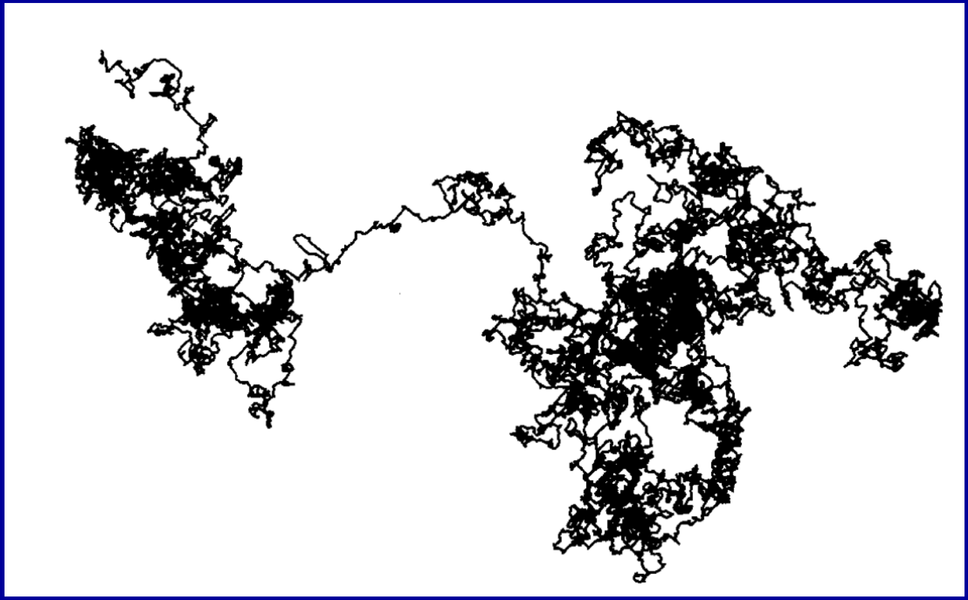
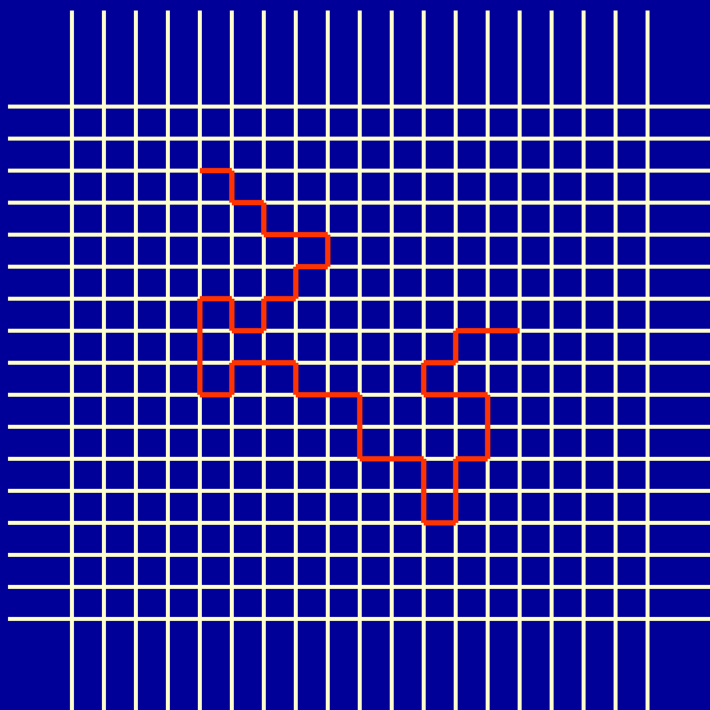
Question??

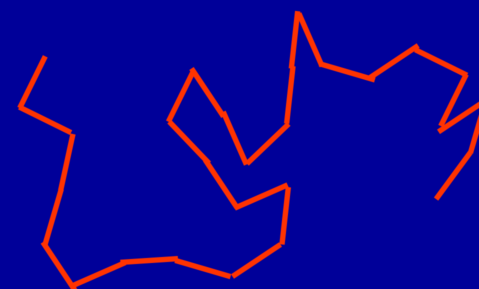
Which systems can be linked with random walk?



A random walk can be achieved by Monte Carlo simulation on lattice or off lattice



on lattice



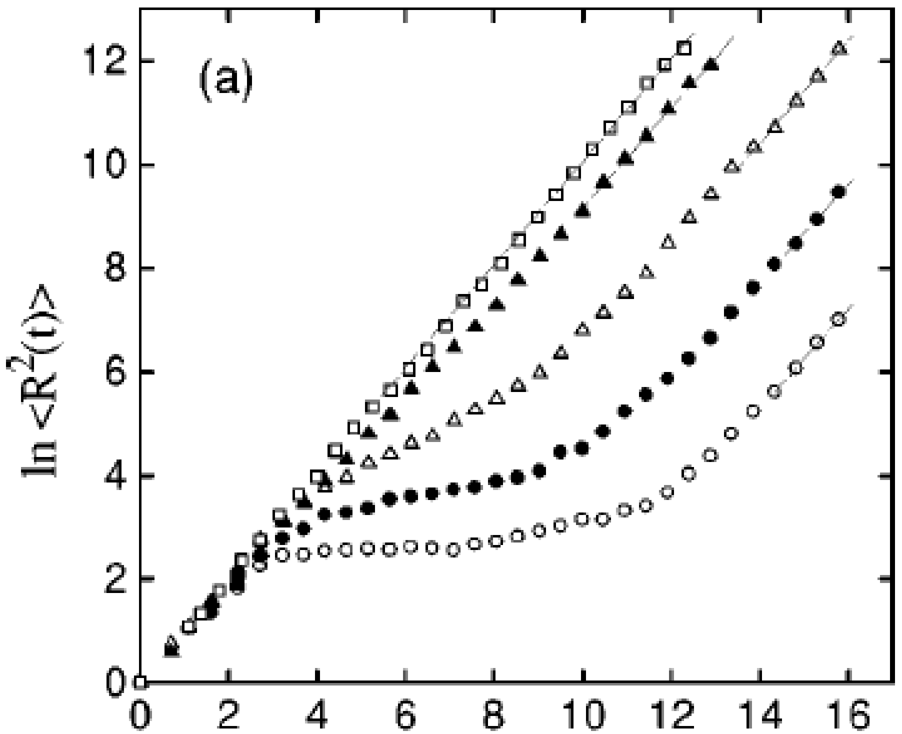
off lattice

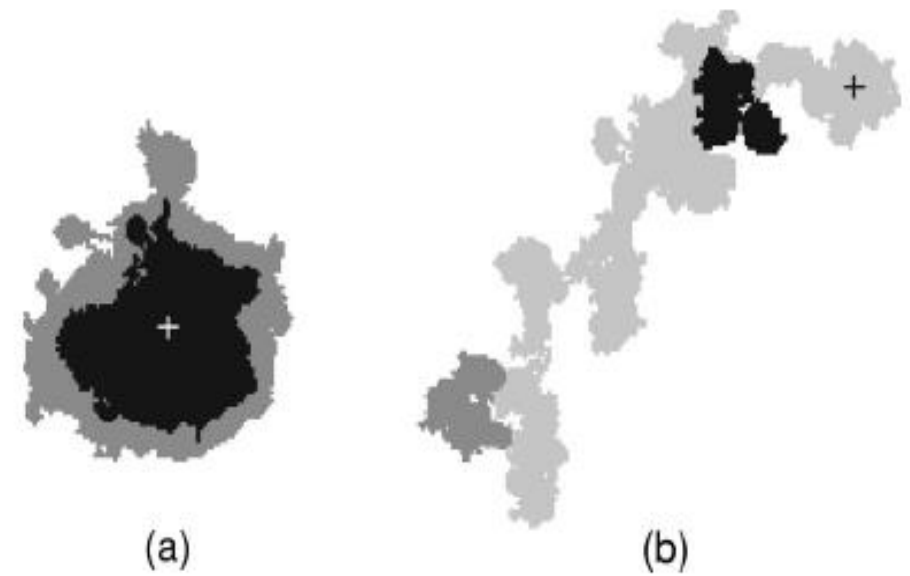
PHYSICAL REVIEW E, VOLUME 65, 041101

Random walk with memory enhancement and decay

Zhi-Jie Tan, Xian-Wu Zou,* Sheng-You Huang, Wei Zhang, and Zhun-Zhi Jin
Department of Physics, Wuhan University, Wuhan 430072, China

(Received 28 September 2001; published 25 March 2002)





Chapter 1:
Introduction to statistical methods

Zhi-Jie Tan
Wuhan University

2019 spring semester

Historical overview: thermodynamics

Count Rumford (1798), Davy (1799) Explicitly by R.J. Mayer (1842) Acceptance after Joule (1843-1849)	Heat is a form of energy
S. Carnot, 1824	First analysis of heat engine
Clausius and Lord Kelvin around 1850	Formulate thermodynamics theory
J.W. Gibbs, 1876-1878	Theory greatly developed

Historical overview-statistical mechanics

Maxwell, 1859	Discover distribution law of molecular velocity
Boltzmann, 1872	The fundamental integro-Differential equation (boltzmann)
Chapman and Enskog (1916-1917)	The kinetic theory of gases
Boltzmann, 1872	More general discipline of Statistical mechanics
J.W. Gibbs, 1902	Greatly enhanced

1, random walk and binomial distribution

1.1 Elementary statistical concepts and examples

Statistical probability and ensemble

It is important to keep in mind that whenever it is desired to describe a situation from a statistical point of view (i.e., in terms of probabilities), it is always necessary to consider an assembly (or "ensemble") consisting of a very large number $\mathfrak{N}$ of similarly prepared systems. The probability of occurrence of a particular event is then defined with respect to this particular ensemble and is given by the fraction of systems in the ensemble

consisting of a very large number $\mathfrak{N}$ of similarly prepared systems.

Statistical probability depends on the nature of ensemble

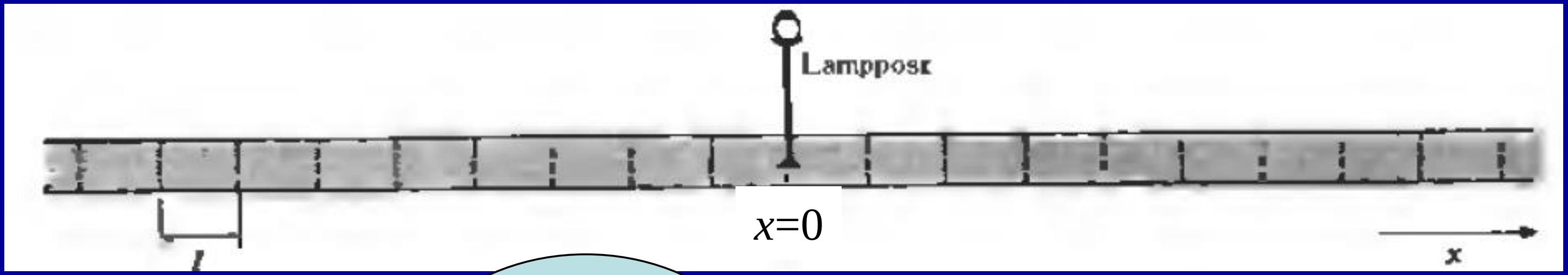
1, random walk and binomial distribution

Which systems are related to random walk

Brownian motion; polymer; 1D Ising Ferromagnetism model;

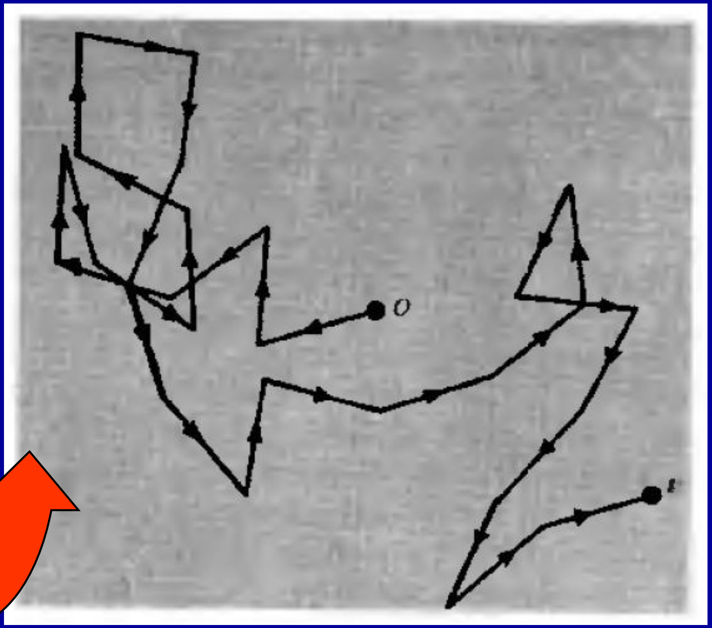
Noise; chaos; drunk; stock market; growth of polymer; DNA denaturation; throw dice; conformation change of polymer; gene mutation; movement of paramecium

What is the probability of a drunk at x=ml from his origin After N step in a street?

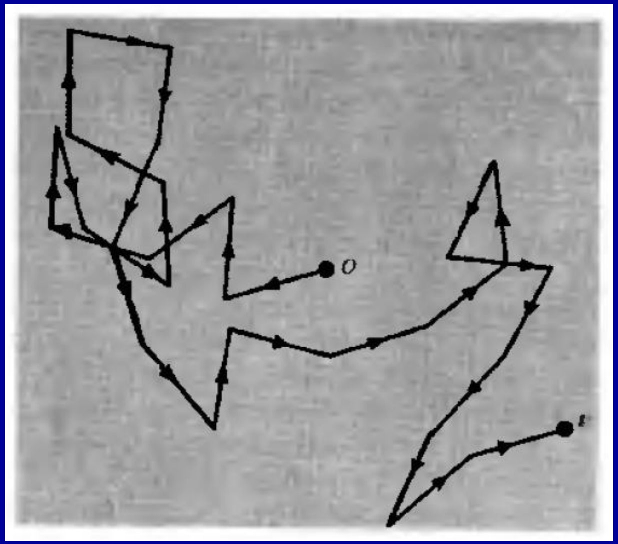


No alcohol

Random walk problem



Random walk is a paradigm problem



As typical examples:

- **Polymer** in solutions. The trajectories of random walks correspond to the ensemble of a polymer.

- **Magnetism**. A atom can have a spin ½ up or down, which corresponds to the probability to the right or to the left.

- **Diffusion** of a particle in gas or in solution. The walk step corresponds to l (mean distance before collision with other molecules)

1.2 The simple random walk problem in 1 dimension

Instead of alcohol or drunk, we use the random walk.

in one dimension. After a total of N such steps, each of length l, the particle is located at

$$x = ml$$

where m is an integer lying between

$$-N \le m \le N$$

We want to calculate the probability $P_N(m)$ of finding the particle at the position $x = ml$ after N such steps.

Probability $P_N(m)$ of finding the particle at $x=ml$?

Probability $P_N(m)$ of finding the particle at $x=ml$?

Let n_1 denote the number of steps to the right and n_2 the corresponding number of steps to the left. Of course, the total number of steps N is simply

$$N = n_1 + n_2 \tag{1·2·1}$$

The net displacement (measured to the right in units of a step length) is given by

$$m = n_1 - n_2 \tag{1·2·2}$$

If it is known that in some sequence of N steps the particle has taken n_1 steps to the right, then its net displacement from the origin is determined. Indeed, the preceding relations immediately yield

$$m = n_1 - n_2 = n_1 - (N - n_1) = 2n_1 - N \tag{1·2·3}$$

This shows that if N is odd, the possible values of m must also be odd. Conversely, if N is even, m must also be even.

Finding the probability $P_N(m)$?

Our fundamental assumption was that successive steps are statistically independent of each other. Thus one can assert simply that, irrespective of past history, each step is characterized by the respective probabilities

$$p = \text{probability that the step is to the right}$$

and

$$q = 1 - p = \text{probability that the step is to the left}$$

Now, the probability of any *one* given sequence of n_1 steps to the right and n_2 steps to the left is given simply by multiplying the respective probabilities, i.e., by

Probability of the move

$$\underbrace{p\,p \cdots p}_{n_1 \text{ factors}}\;\underbrace{q\,q \cdots q}_{n_2 \text{ factors}} = p^{n_1} q^{n_2} \tag{1·2·4}$$

But there are many different possible ways of taking N steps so that n_1 of them are to the right and n_2 are to the left (see illustration in Fig. 1·2·1). Indeed, the number of distinct possibilities (as shown below) is given by

Number for the move:

$$\frac{N!}{n_1!\,n_2!} \tag{1·2·5}$$

Finding the probability $P_N(m)$?

Hence the probability $W_N(n_1)$ of taking (in a total of N steps) n_1 steps to the right and $n_2 = N - n_1$ steps to the left, in any order, is obtained by multiplying the probability (1·2·4) of this sequence by the number (1·2·5) of possible sequences of such steps. This gives

$$W_N(n_1) = \frac{N!}{n_1!n_2!} p^{n_1} q^{n_2} \tag{1·2·6}$$

The probability function (1·2·6) is called the binomial distribution. The reason is that (1·2·5) represents a typical term encountered in expanding $(p + q)^N$ by the binomial theorem. Indeed, we recall that the binomial expansion is given by the formula

$$(p + q)^N = \sum_{n=0}^{N} \frac{N!}{n!(N - n)!} p^n q^{N-n} \tag{1·2·7}$$

Finding the probability $P_N(m)$?

By (1·2·1) and (1·2·2) one finds explicitly*

$$n_1 = \tfrac{1}{2}(N + m), \qquad n_2 = \tfrac{1}{2}(N - m) \tag{1·2·9}$$

Substitution of these relations in (1·2·6) thus yields

$$P_N(m) = \frac{N!}{[(N+m)/2]![(N-m)/2]!} p^{(N+m)/2}(1-p)^{(N-m)/2} \tag{1·2·10}$$

In the special case where $p = q = \frac{1}{2}$ this assumes the symmetrical form

$$P_N(m) = \frac{N!}{[(N+m)/2]![(N-m)/2]!} \left(\frac{1}{2}\right)^N$$

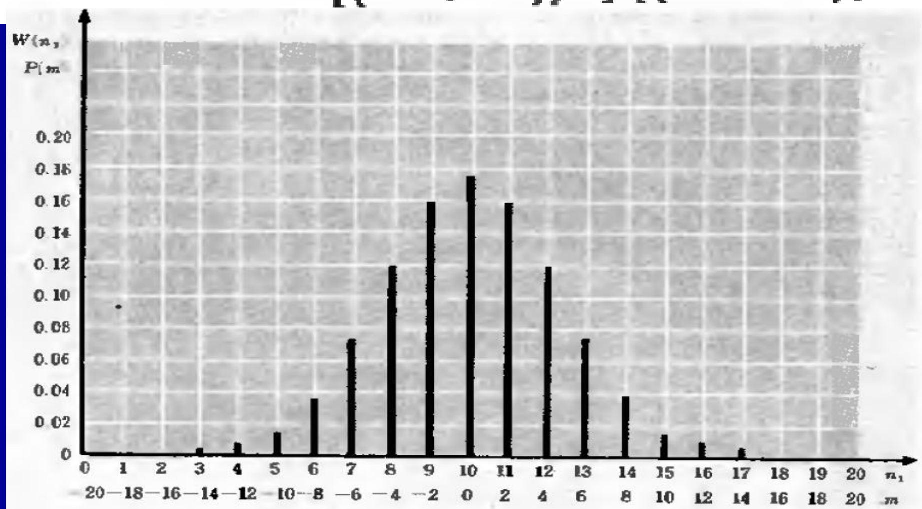


Fig. 1·2·3 Binomial probability distribution for $p = q = \frac{1}{2}$ when $N = 20$ steps. The graph shows the probability $W_N(n_1)$ of n_1 right steps, or equivalently the probability $P_N(m)$ of a net displacement of m units to the right.

1.3 General discussion of mean values

How to calculate the mean values

Let u be a variable which can assume any of the M discrete values

$$u_1, u_2, \ldots, u_M$$

with respective probabilities

$$P(u_1), P(u_2), \ldots, P(u_M)$$

The mean (or average) value of u is denoted by $\bar{u}$ and is defined by

$$\bar{u} = \frac{P(u_1)u_1 + P(u_2)u_2 + \cdots + P(u_M)u_M}{P(u_1) + P(u_2) + \cdots + P(u_M)}$$

or, in shorter notation, by

General formulas:

$$\boxed{\bar{u} \equiv \frac{\sum_{i=1}^{M} P(u_i)u_i}{\sum_{i=1}^{M} P(u_i)}} \qquad (1 \cdot 3 \cdot 1)$$

How to calculate the mean values in general

More generally, if $f(u)$ is any function of u, then the mean value of $f(u)$ is defined by

Pre-normalized:

$$\overline{f(u)} \equiv \frac{\sum_{i=1}^{M} P(u_i) f(u_i)}{\sum_{i=1}^{M} P(u_i)} \qquad (1 \cdot 3 \cdot 2)$$

This expression can be simplified. Since $P(u_i)$ is defined as a probability, the quantity

$$P(u_1) + P(u_2) + \cdots + P(u_M) \equiv \sum_{i=1}^{M} P(u_i)$$

represents the probability that u assumes any one of its possible values and this must be unity. Hence one has quite generally

▶**Normalization:**

$$\sum_{i=1}^{M} P(u_i) = 1 \qquad (1 \cdot 3 \cdot 3)$$

This is the so-called "normalization condition" satisfied by every probability. As a result, the general definition (1·3·2) becomes

▶**After Normalization:**

$$\overline{f(u)} \equiv \sum_{i=1}^{M} P(u_i) f(u_i) \qquad (1 \cdot 3 \cdot 4)$$

How to calculate the mean values of interests

Note the following simple results. If $f(u)$ and $g(u)$ are any two functions of u, then

$$\overline{f(u) + g(u)} = \sum_{i=1}^{M} P(u_i)[f(u_i) + g(u_i)] = \sum_{i=1}^{M} P(u_i)f(u_i) + \sum_{i=1}^{M} P(u_i)g(u_i)$$

or

$$\boxed{\overline{f(u) + g(u)} = \overline{f(u)} + \overline{g(u)}} \tag{1·3·5}$$

Furthermore, if c is any constant, it is clear that

$$\boxed{\overline{cf(u)} = c\overline{f(u)}} \tag{1·3·6}$$

Some simple mean values are particularly useful for describing characteristic features of the probability distribution P. One of these is the mean value $\bar{u}$ (e.g., the mean grade of a class of students). This is a measure of the central value of u about which the various values u_i are distributed. If one measures the values of u from their mean value $\bar{u}$, i.e., if one puts

$$\Delta u \equiv u - \bar{u} \tag{1·3·7}$$

then

$$\overline{\Delta u} = \overline{(u - \bar{u})} = \bar{u} - \bar{u} = 0 \tag{1·3·8}$$

This says merely that the mean value of the deviation from the mean vanishes.

How to calculate the mean values of interests

$\overline{(\Delta u)^n}$: nth moment of u about its mean

Another useful mean value is

$$\overline{(\Delta u)^2} = \sum_{i=1}^{M} P(u_i)(u_i - \bar{u})^2 \geq 0 \qquad (1\cdot3\cdot9)$$

which is called the "second moment of u about its mean," or more simply the "dispersion of u." This can never be negative, since $(\Delta u)^2 \geq 0$ so that each term in the sum contributes a nonnegative number. Only if $u_i = \bar{u}$ for *all* values u_i will the dispersion vanish. The larger the spread of values of u_i about $\bar{u}$, the larger the dispersion. The dispersion thus measures the amount of scatter of values of the variable about its mean value (e.g., scatter in grades about the mean grade of the students). Note the following general relation, which is often useful in computing the dispersion:

$$\overline{(u - \bar{u})^2} = \overline{(u^2 - 2u\,\bar{u} + \bar{u}^2)} = \overline{u^2} - 2\bar{u}\bar{u} + \bar{u}^2$$

or

$$\boxed{\overline{(u - \bar{u})^2} = \overline{u^2} - \bar{u}^2} \qquad (1\cdot3\cdot10)$$

Since the left side must be positive it also follows that

$$\overline{u^2} \geq \bar{u}^2 \qquad (1\cdot3\cdot11)$$

1.4 Calculation of mean values for the random walk

In (1·2·6) we found that the probability, in a total of N steps, of making n_1 steps to the right (and $N - n_1 \equiv n_2$ steps to the left) is

Recall:

$$W(n_1) = \frac{N!}{n_1!(N-n_1)!} p^{n_1} q^{N-n_1} \qquad (1\cdot4\cdot1)$$

Let us first verify the normalization, i.e., the condition

Normalized?

$$\sum_{n_1=0}^{N} W(n_1) = 1 \qquad (1\cdot4\cdot2)$$

which says that the probability of making any number of right steps between 0 and N must be unity. Substituting (1·4·1) into (1·4·2), we obtain

$$\sum_{n_1=0}^{N} \frac{N!}{n_1!(N-n_1)!} p^{n_1} q^{N-n_1} = (p+q)^N \qquad \text{by the binomial theorem}$$

Yes!

$$= 1^N = 1 \qquad \text{since } q \equiv 1 - p$$

which verifies the result.

Finding mean number n_1 of steps to the right ?

What is the mean number $\bar{n}_1$ of steps to the right? By definition

Generally:

$$\bar{n}_1 \equiv \sum_{n_1=0}^{N} W(n_1)n_1 = \sum_{n_1=0}^{N} \frac{N!}{n_1!(N-n_1)!} p^{n_1} q^{N-n_1} n_1 \tag{1·4·3}$$

Then one observes that the extra factor n_1 can be produced by differentiation so that

Useful relation:

$$\boxed{n_1 p^{n_1} = p \frac{\partial}{\partial p}(p^{n_1})}$$

$$n_1{}^a p^{n_1} = (p\frac{\partial}{\partial p})^a (p^{n_1})$$

More general

Hence the sum of interest can be written in the form

$$\sum_{n_1=0}^{N} \frac{N!}{n_1!(N-n_1)!} p^{n_1} q^{N-n_1} n_1 = \sum_{n_1=0}^{N} \frac{N!}{n_1!(N-n_1)!} \left[p \frac{\partial}{\partial p}(p^{n_1})\right] q^{N-n_1}$$

$$= p \frac{\partial}{\partial p} \left[\boxed{\sum_{n_1=0}^{N} \frac{N!}{n_1!(N-n_1)!} p^{n_1} q^{N-n_1}} \right]$$ by interchanging order of summation and differentiation

$$= p \frac{\partial}{\partial p}(p+q)^N$$ by the binomial theorem

$$= pN(p+q)^{N-1}$$

Since this result is true for arbitrary values of p and q, it must also be valid in our particular case of interest where p is some specified constant and $q = 1 - p$. Then $p + q = 1$ so that (1·4·3) becomes simply

$$\boxed{\bar{n}_1 = Np} \tag{1·4·4}$$

Finding mean number *m* of steps to the right ?

Clearly, the mean number of left steps is similarly equal to

Similarly:

$$\bar{n}_2 = Nq \qquad (1\cdot4\cdot5)$$

Of course

$$\bar{n}_1 + \bar{n}_2 = N(p+q) = N$$

add up properly to the total number of steps.

The displacement (measured to the right in units of the step length l) is $m = n_1 - n_2$. Hence we get for the mean displacement

Displacement: $$\bar{m} = \overline{n_1 - n_2} = \bar{n}_1 - \bar{n}_2 = N(p-q) \qquad (1\cdot4\cdot6)$$

If $p = q$, then $\bar{m} = 0$. This must be so since there is then complete symmetry between right and left directions.

Calculation of the dispersion?

one has

$$\overline{(\Delta n_1)^2} \equiv \overline{(n_1 - \bar{n}_1)^2} = \overline{n_1^2} - \bar{n}_1^2 \qquad (1\cdot4\cdot7)$$

We already know $\bar{n}_1$. Thus we need to compute $\overline{n_1^2}$.

$$\begin{aligned} \overline{n_1^2} &\equiv \sum_{n_1=0}^{N} W(n_1) n_1^2 \\ &= \sum_{n_1=0}^{N} \frac{N!}{n_1!(N-n_1)!} \underline{p^{n_1} q^{N-n_1} n_1^2} \end{aligned} \qquad (1\cdot4\cdot8)$$

Considering p and q as arbitrary parameters and using the same trick of differentiation as before, one can write

$$\boxed{n_1^2 p^{n_1} = n_1 \left(p \frac{\partial}{\partial p}\right)(p^{n_1}) = \left(p \frac{\partial}{\partial p}\right)^2 (p^{n_1})}$$

$$n_1^{\,a} p^{n_1} = \left(p \frac{\partial}{\partial p}\right)^a (p^{n_1})$$

Calculation of the dispersion?

Hence the sum in (1·4·8) can be written in the form

$$\sum_{n_1=0}^{N} \frac{N!}{n_1!(N-n_1)!}\left(p\frac{\partial}{\partial p}\right)^2 p^{n_1}q^{N-n_1}$$

$$= \left(p\frac{\partial}{\partial p}\right)^2 \sum_{n_1=0}^{N} \frac{N!}{n_1!(N-n_1)!} p^{n_1}q^{N-n_1}$$ by interchanging order of summation and differentiation

$$= \left(p\frac{\partial}{\partial p}\right)^2 (p+q)^N$$ by the binomial theorem

$$= \left(p\frac{\partial}{\partial p}\right)[pN(p+q)^{N-1}]$$

$$= p[N\underline{(p+q)}^{N-1} + pN(N-1)\underline{(p+q)}^{N-2}]$$

The case of interest in (1·4·8) is that where $p + q = 1$. Thus (1·4·8) becomes simply

$$\begin{aligned}\overline{n_1^2} &= p[N + pN(N-1)] \\ &= Np[1 + pN - p] \\ &= (Np)^2 + Npq \qquad \text{since } 1 - p = q \\ &= \bar{n}_1^2 + Npq \qquad \text{by } (1\cdot4\cdot4)\end{aligned}$$

Hence (1·4·7) gives for the dispersion of n_1 the result

$$\boxed{\overline{n_1^2} - \overline{n_1}^2 = \overline{(\Delta n_1)^2} = Npq} \qquad (1\cdot4\cdot9)$$

Calculation of the mean net displacement m?

Relative dispersion: $$\frac{\Delta^* n_1}{\bar{n}_1} = \frac{\sqrt{Npq}}{Np} = \sqrt{\frac{q}{p}}\frac{1}{\sqrt{N}}$$

In particular,

for $p = q = \frac{1}{2}$,
$$\frac{\Delta^* n_1}{\bar{n}_1} = \frac{1}{\sqrt{N}}$$

Note that as N increases, the mean value $\bar{n}_1$ increases like N, but the width $\Delta^* n_1$ increases only like $N^{\frac{1}{2}}$. Hence the *relative* width $\Delta^* n_1/\bar{n}_1$ decreases with increasing N like $N^{-\frac{1}{2}}$.

One can also compute the dispersion of m, i.e., the dispersion of the net displacement to the right. By (1·2·3)

$$\underline{m = n_1 - n_2 = 2n_1 - N} \qquad (1\cdot4\cdot10)$$

Hence one obtains

$$\underline{\Delta m = m - \bar{m} = (2n_1 - N) - (2\bar{n}_1 - N) = 2(n_1 - \bar{n}_1) = 2\Delta n_1} \qquad (1\cdot4\cdot11)$$

and
$$(\Delta m)^2 = 4(\Delta n_1)^2$$

Taking averages, one gets by (1·4·9)

$$\boxed{\overline{(\Delta m)^2} = 4\overline{(\Delta n_1)^2} = 4Npq} \qquad (1\cdot4\cdot12)$$

In particular,

for $p = q = \frac{1}{2}$,
$$\overline{(\Delta m)^2} = N$$

1.5 Probability distribution for large *N* (?)

If N is large and we consider regions near the maximum of W where n_1 is also large, the fractional change in W when n_1 changes by unity is relatively quite small, i.e.,

$$|W(n_1 + 1) - W(n_1)| \ll W(n_1) \tag{1·5·1}$$

Condition for continuous approximation

Thus W can, to good approximation, be considered as a continuous function of the continuous variable n_1, although only integral values of n_1 are of physical relevance. The location $n_1 = \tilde{n}$ of the maximum of W is then approximately determined by the condition

Maximizing W

$$\frac{dW}{dn_1} = 0 \quad \text{or equivalently} \quad \frac{d \ln W}{dn_1} = 0 \tag{1·5·2}$$

where the derivatives are evaluated for $n_1 = \tilde{n}_1$. To investigate the behavior of $W(n_1)$ near its maximum, we shall put

$$n_1 \equiv \tilde{n}_1 + \eta \tag{1·5·3}$$

and expand $\ln W(n_1)$ in a Taylor's series about $\tilde{n}_1$. The reason for expanding $\ln W$, rather than W itself, is that $\ln W$ is a much more slowly varying function of n_1 than W. Thus the power series expansion for $\ln W$ should converge much more rapidly than the one for W.

1.5 Probability distribution $W(n_1)$ for large N (?)

Expanding $\ln W$ in Taylor's series, one obtains

$$\ln W(n_1) = \ln W(\tilde{n}_1) + B_1\eta + \tfrac{1}{2}B_2\eta^2 + \tfrac{1}{6}B_3\eta^3 + \cdots \qquad (1\cdot5\cdot4)$$

where **Coefficients:**

$$B_k \equiv \frac{d^k \ln W}{dn_1^k} \qquad (1\cdot5\cdot5)$$

is the kth derivative of $\ln W$ evaluated at $n_1 = \tilde{n}_1$. Since one is expanding about a maximum, $B_1 = 0$ by $(1\cdot5\cdot2)$. Also, since W is a maximum, it follows that the term $\frac{1}{2}B_2\eta^2$ must be negative, i.e., B_2 must be negative. To make this explicit, let us write $B_2 = -|B_2|$. Hence $(1\cdot5\cdot4)$ yields, putting $\tilde{W} = W(\tilde{n}_1)$,

$$W(n_1) = \tilde{W}\, e^{\frac{1}{2}B_2\eta^2 + \frac{1}{6}B_3\eta^3 \cdots} = \tilde{W}\, e^{-\frac{1}{2}|B_2|\eta^2} e^{\frac{1}{6}B_3\eta^3 \cdots} \qquad (1\cdot5\cdot6)$$

In the region where η is sufficiently small, higher-order terms in the expansion can be neglected so that one obtains in first approximation an expression of the simple form

$$W(n_1) = \tilde{W}\, e^{-\frac{1}{2}|B_2|\eta^2} \qquad (1\cdot5\cdot7)$$

W ?

B_2 ?

Finding B_2 for $W(n_1)$??

Let us now investigate the expansion (1·5·4) in greater detail. By (1·4·1) one has

$$\boxed{\ln W(n_1) = \ln N! - \ln n_1! - \ln(N - n_1)! + n_1 \ln p + (N - n_1) \ln q} \quad (1\cdot5\cdot8)$$

But, if n is any large integer so that $n \gg 1$, then $\ln n!$ can be considered an almost continuous function of n, since $\ln n!$ changes only by a small fraction of itself if n is changed by a small integer. Hence

$$\frac{d \ln n!}{dn} \approx \frac{\ln (n+1)! - \ln n!}{1} = \ln \frac{(n+1)!}{n!} = \ln (n+1)$$

Thus **Approximation for n>>1**

for $n \gg 1$,

$$\frac{d \ln n!}{dn} \approx \ln n \quad (1\cdot5\cdot9)$$

Hence (1·5·8) yields

$$\boxed{\frac{d \ln W}{dn_1} = -\ln n_1 + \ln (N - n_1) + \ln p - \ln q} \quad (1\cdot5\cdot10)$$

Finding B_2 for $W(n_1)$??

By equating this first derivative to zero one finds the value $n_1 = \tilde{n}_1$, where W is maximum. Thus one obtains the condition

$$\frac{dW(n_1)}{dn_1} = 0 \Rightarrow \ln\left[\frac{(N-\tilde{n}_1)}{\tilde{n}_1}\frac{p}{q}\right] = 0$$

or

$$(N-\tilde{n}_1)p = \tilde{n}_1 q$$

so that

$$\tilde{n}_1 = Np \tag{1·5·11}$$

since $p + q = 1$.

$$\boxed{B_k \equiv \frac{d^k \ln W}{dn_1{}^k}}$$

Further differentiation of (1·5·10) yields

$$\frac{d^2 \ln W}{dn_1{}^2} = -\frac{1}{n_1} - \frac{1}{N-n_1} = B_2 \tag{1·5·12}$$

Evaluating this for the value $n_1 = \tilde{n}_1$ given in (1·5·11), one gets

$$B_2 = -\frac{1}{Np} - \frac{1}{N-Np} = -\frac{1}{N}\left(\frac{1}{p}+\frac{1}{q}\right)$$

or

$$\boxed{B_2 = -\frac{1}{Npq}} \tag{1·5·13}$$

since $p + q = 1$. Thus B_2 is indeed negative, as required for W to exhibit a maximum.

Finding W for $W(n_1)$??

Normalization:

The value of the constant $\tilde{W}$ in (1·5·7) can be determined from the normalization condition (1·4·2). Since W and n_1 can be treated as quasicontinuous variables, the sum over all integral values of n_1 can be approximately replaced by an integral. Thus the normalization condition can be written

$$\sum_{n_1=0}^{N} W(n_1) \approx \int W(n_1)\, dn_1 = \int_{-\infty}^{\infty} W(\tilde{n}_1 + \eta)\, d\eta = 1 \qquad (1\cdot 5\cdot 17)$$

Here the integral over η can to excellent approximation be extended from $-\infty$ to $+\infty$, since the integrand makes a negligible contribution to the integral wherever $|\eta|$ is large enough so that W is far from its pronounced maximum value. Substituting (1·5·7) into (1·5·17) and using (A·4·2), one obtains

$$\tilde{W} \int_{-\infty}^{\infty} e^{-\frac{1}{2}|B_2|\eta^2}\, d\eta = \tilde{W}\sqrt{\frac{2\pi}{|B_2|}} = 1$$

$$\boxed{\tilde{W} = \sqrt{\frac{|B_2|}{2\pi}}}$$

Obtained formula for W(n_1) ??

Thus (1·5·7) becomes

$$W(n_1) = \sqrt{\frac{|B_2|}{2\pi}}\, e^{-\frac{1}{2}|B_2|(n_1-\tilde{n}_1)^2} \tag{1·5·18}$$

The reasoning leading to the functional form (1·5·7) or (1·5·18), the so-called "Gaussian distribution," has been very general in nature. Hence it is not surprising that Gaussian distributions occur very frequently in statistics whenever one is dealing with large numbers. In our case of the binomial distribution, the expression (1·5·18) becomes, by virtue of (1·5·11) and (1·5·13),

$$W(n_1) = (2\pi Npq)^{-\frac{1}{2}} \exp\left[-\frac{(n_1 - Np)^2}{2Npq}\right] \tag{1·5·19}$$

Note that (1·5·19) is, for large values of N and n_1, much simpler than (1·4·1), since it does not require the evaluation of large factorials. Note also that, by (1·4·4) and (1·4·9), the expression (1·5·19) can be written in terms of the mean values $\bar{n}_1$ and $\overline{(\Delta n_1)^2}$ as

$$W(n_1) = [2\pi\overline{(\Delta n_1)^2}]^{-\frac{1}{2}} \exp\left[-\frac{(n_1 - \bar{n}_1)^2}{2\overline{(\Delta n_1)^2}}\right]$$

$$\overline{\Delta n^2} = Npq$$

1.6 Gaussian probability distributions P(m)

The Gaussian approximation (1·5·19) also yields immediately the probability $P(m)$ that in a large number of N steps the net displacement is m. The corresponding number of right steps is, by (1·2·9), $n_1 = \frac{1}{2}(N + m)$. Hence (1·5·19) gives

$$P(m) = W\left(\frac{N+m}{2}\right) = [2\pi Npq]^{-\frac{1}{2}} \exp\left\{-\frac{[m - N(p-q)]^2}{8Npq}\right\} \tag{1·6·1}$$

since $n_1 - Np = \frac{1}{2}[N + m - 2Np] = \frac{1}{2}[m - N(p-q)]$. By (1·2·3) one has $m = 2n_1 - N$, so that m assumes here integral values separated by an amount $\Delta m = 2$.

We can also express this result in terms of the actual displacement variable x,

$$x = ml \tag{1·6·2}$$

where l is the length of each step. If l is small compared to the smallest length of interest in the physical problem under consideration,* the fact that x can only assume values in discrete increments of $2l$, rather than all values continuously, becomes unimportant. Furthermore, when N is large, the probability $P(m)$ of occurrence of a displacement m does not change significantly from

Finding $\mathcal{P}(x)$ in general form

Under these circumstances it is possible to regard x as a continuous variable on a macroscopic scale and to ask for the probability that the particle is found after N steps in the range between x and $x + dx$.* Since m assumes only integral values separated by $\Delta m = 2$, the range dx contains $dx/2l$ possible values of m, all of which occur with nearly the same probability $P(m)$. Hence the probability of finding the particle anywhere in the range between x and $x + dx$ is simply obtained by summing $P(m)$ over all values of m lying in dx, i.e., by multiplying $P(m)$ by $dx/2l$. This probability is thus proportional to dx (as one would expect) and can be written as

$$\boxed{\mathcal{P}(x)\,dx = P(m)\,\frac{dx}{2l}} \qquad (1\cdot6\cdot3)$$

where the quantity $\mathcal{P}(x)$, which is independent of the magnitude of dx, is called a "probability *density*." Note that it must be multiplied by a differential element of length dx to yield a probability.

By using (1·6·1) one then obtains

▶ $$\boxed{\mathcal{P}(x)\,dx = \frac{1}{\sqrt{2\pi}\,\sigma}\,e^{-(x-\mu)^2/2\sigma^2}\,dx} \qquad (1\cdot6\cdot4)$$

where we have used the abbreviations

$$\mu \equiv (p - q)Nl \qquad (1\cdot6\cdot5)$$

and

$$\sigma \equiv 2\sqrt{Npq}\,l \qquad (1\cdot6\cdot6)$$

Verifying $\mathcal{P}(x)$ is normalized

Using (1·6·4), one can quite generally compute the mean values $\bar{x}$ and $\overline{(x-\bar{x})^2}$. In calculating these mean values, sums over all possible intervals dx become, of course, integrations. (The limits of x can be taken as $-\infty < x < \infty$, since $\mathcal{P}(x)$ itself becomes negligibly small whenever $|x|$ is so large as to lead to a displacement inaccessible in N steps.)

First we verify that $\mathcal{P}(x)$ is properly normalized, i.e., that the probability of the particle being somewhere is unity. Thus

$$\begin{aligned}\int_{-\infty}^{\infty} \mathcal{P}(x)\,dx &= \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2}\,dx \\ &= \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} e^{-y^2/2\sigma^2}\,dy \\ &= \frac{1}{\sqrt{2\pi}\,\sigma}\sqrt{\pi 2\sigma^2} \\ &= 1 \end{aligned} \qquad (1\cdot6\cdot7)$$

Appendix

Calculating mean values of x et al

Next we calculate the mean value

$$\bar{x} \equiv \int_{-\infty}^{\infty} x\mathcal{P}(x)\,dx$$
$$= \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} x\, e^{-(x-\mu)^2/2\sigma^2}\,dx$$
$$= \frac{1}{\sqrt{2\pi}\,\sigma}\left[\underline{\int_{-\infty}^{\infty} y\, e^{-y^2/2\sigma^2}\,dy} + \mu \underline{\int_{-\infty}^{\infty} e^{-y^2/2\sigma^2}\,dy}\right]$$

Appendix

Since the integrand in the first integral is an odd function of y, the first integral vanishes by symmetry. The second integral is the same as that in (1·6·7), so that one gets

$$\boxed{\bar{x} = \mu} \qquad \boxed{\mu \equiv (p - q)Nl} \qquad (1\cdot6\cdot8)$$

This is a simple consequence of the fact that $\mathcal{P}(x)$ is only a function of $|x-\mu|$ and is thus symmetric about the position $x = \mu$ of its maximum. Hence this point corresponds also to the mean value of x.

Calculating mean values of dispersion x^2

The dispersion becomes

$$\begin{aligned}
\overline{(x-\mu)^2} &= \int_{-\infty}^{\infty} (x-\mu)^2 \mathcal{P}(x)\, dx \\
&= \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} y^2 e^{-y^2/2\sigma^2}\, dy \\
&= \frac{1}{\sqrt{2\pi}\,\sigma} \left[\frac{\sqrt{\pi}}{2} (2\sigma^2)^{\frac{3}{2}} \right] \\
&= \sigma^2
\end{aligned}$$

Appendix

where we have used the integral formulas (A·4·6). Thus

$$\boxed{\overline{(\Delta x)^2} = \overline{(x-\mu)^2} = \sigma^2} \qquad (1\cdot6\cdot9)$$

Hence σ is simply the root-mean-square deviation of x from the mean of the Gaussian distribution.

By (1·6·5) and (1·6·6), one then obtains for the random walk problem the relations

$$\boxed{\begin{aligned} \bar{x} &= (p-q)Nl \qquad (1\cdot6\cdot10) \\ \overline{(\Delta x)^2} &= 4Npql^2 \qquad (1\cdot6\cdot11) \end{aligned}}$$

For random polymer

In 3 dimension:

$$\boxed{\begin{aligned} P(x,y,z,N)\,dx\,dy\,dz &= P(x,N)P(y,N)P(z,N)\,dx\,dy\,dz \\ &= \left[\frac{\beta}{\pi}\right]^{3/2} e^{-\beta(x^2+y^2+z^2)}\,dx\,dy\,dz. \end{aligned}}$$

In terms of the vector **r**, you have

$$\boxed{P(\mathbf{r},N) = P(x,y,z,N) = \left[\frac{\beta}{\pi}\right]^{3/2} e^{-\beta r^2}.}$$

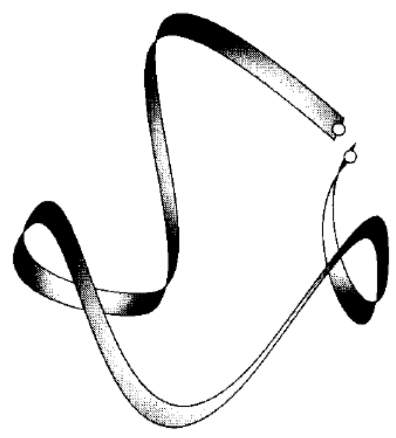


Figure 32.8 For polymer cyclization, the two chain ends must be close together.

Probability of finding a N-mer polymer with end-to-end distance r in 3D

$$\boxed{4\pi r^2 P(\mathbf{r},N)}$$

Polymer cyclization (Jacobson–Stockmayer theory).

$$\begin{aligned} P_{\text{cyclization}} &= \int_0^b P(r,N)\,dr \\ &= \left[\frac{3}{2\pi N b^2}\right]^{3/2} \int_0^b e^{-3r^2/2Nb^2} 4\pi r^2\,dr. \qquad (32.21) \end{aligned}$$

Further reading: 《**Molecular driving force**》 **Dill et al，Taylor , 2010**

To be continued…..

Question?? <<< Answer

A random linear polymer can be represented by a trajectory of a random walker?

1: what properties can be used to describe the structure of polymer in a free space?

2: is the pure random walk adequate for a polymer? If now, do you have a suggestion?

3: Can you simulate a simple polymer? How? (at home?)

Homework: Textbook page 40-43

1.1 1.6 1.8 1.9 1.14