CAMBRIDGE

Cambridge University Press
978-1-107-05378-6 - Introduction to Statistical Physics and to Computer Simulations
João Paulo Casquilho and Paulo Ivo Cortez Teixeira
Excerpt
More information

1 Random walks

1.1 Introduction

We shall start our study of statistical physics with *random walks*. As their name suggests, random walks are sequences of steps taken in random directions, i.e., where the direction of a step is chosen according to a given probability distribution. In the most general case, the length of a step is also random. The probability that each step in a sequence of N steps has a given length and is taken in a given direction is independent of the lengths and directions of all other steps in the sequence; in other words, the N steps are statistically independent. We shall study random walks to find out how probable it is that a random walker ends up a certain distance from its starting point after N steps, and how far on average the random walker strays from its starting point.

Random walks are important for a number of reasons. Firstly, they illustrate some basic results of probability theory. Secondly, their statistics are formally identical to those of many key problems in physics and other disciplines, such as diffusion in fluids and solids and the conformations of polymer molecules, some of which will be addressed later in this book. Finally, random walks are particularly suitable vehicles for introducing Monte Carlo simulation methods into statistical physics as they are simple but require many techniques employed in the study of more complex systems.

1.2 Probability: basic definitions

When seeking to describe a system statistically (i.e., in probabilistic terms), it is often useful to consider a *statistical ensemble*. This consists of a very large number N of virtual copies of the system under study. The probability of some particular event occurring is then defined with respect to the ensemble: it is the fraction of systems in the ensemble in which the event actually occurs. Consider, for example, a die throw. One possible statistical description is to assume that the die is thrown N times in succession, under identical conditions. Alternatively, we can imagine a very large number N of identical dice (the representative ensemble of the die) and that each of these is thrown once, all under identical conditions.

We define the probability P_r of a random event r as the limiting value of the relative frequency of r, when the number of trials $N \to \infty$:

$$P_r = \lim_{N\to\infty} f(r), \qquad f(r) = \frac{N_r}{N}, \tag{1.1}$$

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-107-05378-6 - Introduction to Statistical Physics and to Computer Simulations
João Paulo Casquilho and Paulo Ivo Cortez Teixeira
Excerpt
More information

where N_r is the number of times that r occurs in N trials. This is the so-called 'frequencist' definition of probability, which is the one most commonly used in the physical sciences. Coming back to the preceding example, the probability of a certain outcome is given by the fraction of trials with that particular outcome. This *ensemble probability* is always defined with respect to a specific ensemble. For example, the probability that the outcome of a die throw is 6, defined with respect to the ensemble of perfect six-sided dice, is different from the probability that the outcome is 6, defined with respect to the ensemble of perfect ten-sided dice.

Consider now an experiment with L mutually exclusive possible outcomes (e.g., a die throw), and label each of these outcomes with the index r: thus $r = 1, 2, \ldots, L$. After $N \gg 1$ trials, outcome 1 was found N_1 times, outcome 2 N_2 times, ..., and outcome L N_L times. Because the outcomes are mutually exclusive, we must have $N_1 + N_2 + \cdots + N_L = N$. Dividing both sides of this equation by N and using Eq. (1.1) we obtain

$$\sum_{r=1}^{L} P_r = 1, \tag{1.2}$$

i.e., the probability distribution is *normalised to unity*. If the outcomes are *equally probable*, it follows straightforwardly from Eq. (1.2) that

$$P_r = \frac{1}{L}, \quad \text{for all } r. \tag{1.3}$$

The probability of that the outcome is r **or** s is

$$P(r \text{ or } s) = \frac{N_r + N_s}{N} = P_r + P_s, \tag{1.4}$$

which is straightforwardly generalised to more than two outcomes. For example, when throwing a die the probability that the outcome is 4 or 5 is $1/6 + 1/6 = 1/3$

Now suppose that two separate experiments are performed: one with L mutually exclusive possible outcomes $r = 1, 2, \ldots, L$, and the other with M mutually exclusive possible outcomes $s = 1, 2, \ldots, M$. An example might be throwing a die ($L = 6$) and flipping a coin ($M = 2$) simultaneously. The probability that outcomes r and s occur simultaneously is called the *joint probability* of r and s. For each of the N_r possible trials of the first experiment with outcome r, there are N_s possible trials of the second experiment with outcome s, so the total number of trials with outcomes r and s is $N_{rs} = N_r N_s$. The joint probability is then

$$P_{rs} = \frac{N_{rs}}{N} = \frac{N_r N_s}{N}. \tag{1.5}$$

Two events are said to be *statistically independent* when the probability that one event has a particular outcome does not depend on the probability of the other event having a particular outcome. In this case, for all N_r trials with outcome r of the first event, there will be a fraction P_s for which the outcome of the second event is s, so that $N_{rs} = N_r P_s$, leading to

$$P_{rs} = \frac{N_{rs}}{N} = \frac{N_r}{N} P_s = P_r P_s. \tag{1.6}$$

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-107-05378-6 - Introduction to Statistical Physics and to Computer Simulations
João Paulo Casquilho and Paulo Ivo Cortez Teixeira
Excerpt
More information

This is easily generalised to more than two independent events as: *The joint probability of statistically independent events is the product of the probabilities of the individual events*.

1.3 Random variables and distribution functions

A *random variable* is a variable X which takes values x in a set B according to a given *probability law* or *distribution function* F_X such that

$$F_X(x) = P[X \leq x], \quad \text{for all } x \in B. \tag{1.7}$$

In other words, the distribution function of X evaluated at x equals the probability that $X \leq x$. Two random variables are said to be *equal in distribution* if they have the same distribution function.

Random variables may be discrete or continuous.

For a *discrete random variable*, there exists a function p_X, called *probability mass function*, such that

$$P[X \leq x] = \sum_{t \leq x} p_X(t), \tag{1.8}$$

$$P[X = x] = p_X(x), \tag{1.9}$$

i.e., $p_X(x)$ equals the probability that $X = x$ and the sum is over all possible values t of X not greater than x. In this case, normalisation gives

$$\sum_{x \in B} p_X(x) = 1, \tag{1.10}$$

where the sum is over all possible values x of X.

For a *continuous random variable* there exists a function $f_X(x)$ called *probability density function* such that

$$P[X \leq x] = \int_{-\infty}^{x} f_X(t)dt, \tag{1.11}$$

$$P[X \in (x, x + dx)] = f_X(x)dx, \tag{1.12}$$

where in this case we have assumed that X may take any value between $-\infty$ and $+\infty$. Normalisation now gives

$$\int_{-\infty}^{+\infty} f_X(x)dx = 1. \tag{1.13}$$

We define the *most probable value* of a random variable X as that for which the probability mass function of X, or the probability density function of X, is maximised (for discrete and continuous X, respectively).

The *mean*, *mean value*, *expected value* or *mathematical expectation* of a random variable X is defined as

$$E[X] = \sum_{k=1}^{\infty} x_k p_k, \qquad p_k = P[X = x_k], \tag{1.14}$$

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-107-05378-6 - Introduction to Statistical Physics and to Computer Simulations
João Paulo Casquilho and Paulo Ivo Cortez Teixeira
Excerpt
More information

if X is discrete, and

$$E[X] = \int_{-\infty}^{+\infty} x f_X(x) dx, \tag{1.15}$$

if X is continuous. The values of X cluster around its mean, which is therefore a measure of the localisation of X, or of its distribution. Note that the mean is the weighted average of all values of X.

Henceforth we shall often use x to denote both the random variable X and its values. The mean of X will then be denoted $\overline{x}$. For consistency we shall employ this notation for discrete as well as continuous variables, although in probability theory $\overline{x}$ usually denotes the arithmetic mean of a sample of X. Thus

$$E[X] = \overline{x}. \tag{1.16}$$

The nth *central moment* of a random variable X is defined as $E[(X - E[X])^n]$. If the distribution function of X is known, then the mean and all moments of X can be calculated. The converse is also true in the case of smooth distributions: this result is often used when the mean and a few moments of a distribution are known, experimentally or from computer simulations ('numerical experiments'), to extract an approximation to the probability distribution function.

The *deviation from the mean* is defined as $X - E[X]$, which in the notation of Eq. (1.16) may also be written

$$X - E[X] = x - \overline{x} \equiv \Delta x. \tag{1.17}$$

If a random variable X deviates very little from its mean, then the true value of the quantity X may be replaced by the mean of X, with a very small error. This is a key issue in statistical physics: in order to quantify how much the true value of X deviates from its mean, we shall calculate the first and second centred moments of a continuous random variable. The same results can easily be obtained for a discrete random variable. Because $X - E[X]$ is also a random variable, we find, from Eqs (1.13)–(1.17):

$n = 1$ (*mean deviation from the mean*):

$$E[X - E[X]] \equiv \overline{\Delta x} = \int_B (x - \overline{x}) f_X(x) dx = 0. \tag{1.18}$$

$n = 2$ (*variance, mean square deviation or scatter*):

$$\begin{aligned} \mathrm{Var}(X) \equiv \sigma^2(X) = E\left[(X - [X])^2\right] &= \overline{(\Delta x)^2} = \int_B (x - \overline{x})^2 f_X(x) dx \\ &= \int x^2 f_X(x) dx - 2\overline{x} \int x f_X(x) dx + \overline{x}^2 \int f_X(x) dx \\ &= \overline{x^2} - \overline{x}^2. \end{aligned} \tag{1.19}$$

Clearly the variance cannot be negative, as $(\Delta x)^2 \geq 0$. Equation (1.19) then implies that $\overline{x^2} \geq \overline{x}^2$. The variance vanishes only if $x = \overline{x}$ for all x. The larger the variance, the greater the scatter of x about its mean. If $\sigma^2(X)$ is small, then X will always be close to $\overline{x}$, and we

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-107-05378-6 - Introduction to Statistical Physics and to Computer Simulations
João Paulo Casquilho and Paulo Ivo Cortez Teixeira
Excerpt
More information

can replace the true value of X by its mean $\overline{x}$. The relative error of this procedure is the *relative fluctuation*:

$$\delta(X) = \sigma(X)/E[X], \tag{1.20}$$

where $\sigma(X) = \sqrt{\sigma^2(X)}$ is the *standard deviation* of X.

If a discrete random variable X can take only a finite number N of values all with the same probability p_k, normalisation implies that $p_k = 1/N$ as in Eq. (1.3). By Eq. (1.14) the statistical mean then coincides with the arithmetic mean:

$$\overline{x} = \frac{1}{N}\sum_{k=1}^{N} x_k. \tag{1.21}$$

In this case the discrete version of Eq. (1.18) is just that *the deviations from the mean add up to zero*.

In the limit $N \to \infty$, the statistical mean and the arithmetic mean of a discrete random variable are related through a theorem known as the *law of large numbers*, which can be stated as follows (Apostol, 1969): Let $X_1, X_2, \ldots, X_N$ be a sequence of N independent random variables, all equal in distribution to some random variable X of finite mean and variance. If we define a new random variable $\overline{X}$, the *arithmetic mean* of $X_1, X_2, \ldots, X_N$, as

$$\overline{X} = \frac{1}{N}\sum_{k=1}^{N} X_k, \tag{1.22}$$

then

$$\lim_{N\to\infty} \overline{X} = E[X]. \tag{1.23}$$

We close this section with a brief reference to *functions of random variables*. For simplicity we shall drop the X subscript. The mean of a function $g(X)$ of random variable X is then defined as

$$\overline{g(X)} = \sum_x g(x)p(x) \quad \text{or} \quad \overline{g(X)} = \int g(x)f(x)dx, \tag{1.24}$$

where $p(x)$ and $f(x)$ are given by Eqs. (1.9) and (1.12), respectively.

1.4 The simple random walk

1.4.1 The binomial distribution

We shall start by deriving the probability distribution for a one-dimensional random walk. Here steps are all the same length ℓ and can be taken either to the right or to the left. Out of N steps, n_1 are rightward steps and $n_2 = N - n_1$ are leftward steps. Assuming that consecutive steps are statistically independent, then at each moment in time the probability of taking a step right is p and the probability of taking a step left is $q = 1 - p$. If we view a random walk as N trials of a *Bernoulli experiment* where a step right is a success and a step

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-107-05378-6 - Introduction to Statistical Physics and to Computer Simulations
João Paulo Casquilho and Paulo Ivo Cortez Teixeira
Excerpt
More information

left is a failure, then the probability of n_1 successes out of N trials is given by the *binomial distribution*:

$$\begin{aligned} &P(X=n_1)\equiv P(n_1)=\binom{N}{n_1}p^{n_1}q^{n_2}, && n_1=0,1,\ldots,N,\\ &P(n_1)=0, && \text{any other value of } n_1, \end{aligned} \tag{1.25}$$

where X is the random variable equal to the number of successes in N independent Bernoulli experiments, $n_2=N-n_1$ is the number of failures, and

$$\binom{N}{n_1}=\frac{N!}{n_1!(N-n_1)!} \tag{1.26}$$

is the binomial coefficient.

Proof of Eq. (1.25) For N Bernoulli experiments, the probability of n_1 consecutive successes is p^{n_1} (because all outcomes are independent). Likewise, the probability of n_2 consecutive failures is q^{n_2}. Hence the probability that n_1 consecutive successes are followed by n_2 consecutive failures is $p^{n_1}q^{n_2}$. The probability of n_1 successes and n_2 failures in any other order is also $p^{n_1}q^{n_2}$, since each of the N outcomes is either a success or a failure. Then the probability of n_1 successes and n_2 failures regardless of order equals the probability of n_1 successes and n_2 failures in some order, times the number of possible ways in which n_1 out of N experiments are successes, which is given by the binomial coefficient.

The mean of n_1 can be found using Eqns. (1.14) and (1.25):

$$\overline{n_1}=\sum_{n_1=0}^{N}n_1P(n_1)=\sum_{n_1=0}^{N}n_1\binom{N}{n_1}p^{n_1}q^{n_2}. \tag{1.27}$$

Noting that

$$n_1p^{n_1}q^{n_2}=p\frac{\partial}{\partial p}\left(p^{n_1}q^{n_2}\right), \tag{1.28}$$

that the binomial theorem implies

$$\sum_{n_1=0}^{N}P(n_1)=(p+q)^N, \tag{1.29}$$

and that $P(n_1)$ is normalised to unity, Eq. (1.10), we obtain

$$\begin{aligned} \overline{n_1}&=\sum_{n_1=0}^{N}p\frac{\partial}{\partial p}P(n_1)=p\frac{\partial}{\partial p}\left(\sum_{n_1=0}^{N}P(n_1)\right)\\ &=p\frac{\partial}{\partial p}\left[(p+q)^N\right]=Np(p+q)^{N-1}=Np. \end{aligned} \tag{1.30}$$

Along the same lines, it may be easily shown that

$$\overline{n_1^2}=Np+N(N-1)p^2, \tag{1.31}$$

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-107-05378-6 - Introduction to Statistical Physics and to Computer Simulations
João Paulo Casquilho and Paulo Ivo Cortez Teixeira
Excerpt
More information

whence from Eq. (1.19) with Eqs. (1.30) and (1.31) we find for the variance:

$$\sigma^2(n_1) = \overline{n_1^2} - \overline{n_1}^2 = Npq. \tag{1.32}$$

Up till now we have concentrated on a description of the one-dimensional random walk in terms of the random variable n_1, the number of rightward steps out of N total steps. If, however, we wish to study the *displacement*, it is convenient to introduce a new variable, the *effective number of rightward steps*, as (Reif, 1985)

$$n = n_1 - n_2, \tag{1.33}$$

where $n_1 + n_2 = N$ (recall that n_2 is the total number of leftward steps). We can then define the *effective rightward displacement* as

$$L = nl, \tag{1.34}$$

where l is the step length.

The mean of n is easily calculated from Eq. (1.30):

$$\overline{n} = \overline{n_1 - n_2} = \overline{n_1} - \overline{n_2} = N(p - q). \tag{1.35}$$

Because $n = n_1 - n_2 = 2n_1 - N$, n takes only integer values with spacing $\delta n = 2$, whence

$$\Delta n \equiv n - \overline{n} = (2n_1 - N) - (2\overline{n_1} - N) = 2\Delta n_1, \tag{1.36}$$

where $\Delta n_1 = n_1 - \overline{n_1}$. From Eq. (1.32) we then get the variance of n:

$$\sigma^2(n) = \overline{\Delta n^2} = 4Npq. \tag{1.37}$$

The mean and variance of the displacement then follow easily from Eqs. (1.34)–(1.37):

$$\overline{L} = Nl(p - q), \tag{1.38}$$

$$\sigma^2(L) = 4Nl^2pq. \tag{1.39}$$

The variance of n is thus quadratic in l, whence the standard deviation $\sigma(L)$ is a measure of the linear scatter of L values.

As mentioned before, a suitable measure of the width of a distribution is its relative fluctuation, given by Eq. (1.20). For n_1 this is, from Eqs. (1.30) and (1.32):

$$\delta(n_1) \equiv \frac{\sigma(n_1)}{\overline{n_1}} = \sqrt{\frac{q}{p}}\frac{1}{\sqrt{N}}, \tag{1.40}$$

i.e., the relative fluctuation decays with $N^{-1/2}$ when N increases (see Figure 1.1).

In the special case where $p = q = 1/2$, i.e., rightward and leftward steps are equally probable, we have the following results:

$$\overline{n} = 0; \qquad \sigma^2(n) = N; \qquad \sigma(n) = \sqrt{N}; \tag{1.41}$$

$$\overline{L} = 0; \qquad \sigma^2(L) = Nl^2; \qquad \sigma(n) = \sqrt{N}l; \tag{1.42}$$

$$\delta(n_1) = \frac{1}{\sqrt{N}}. \tag{1.43}$$

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-107-05378-6 - Introduction to Statistical Physics and to Computer Simulations
João Paulo Casquilho and Paulo Ivo Cortez Teixeira
Excerpt
More information

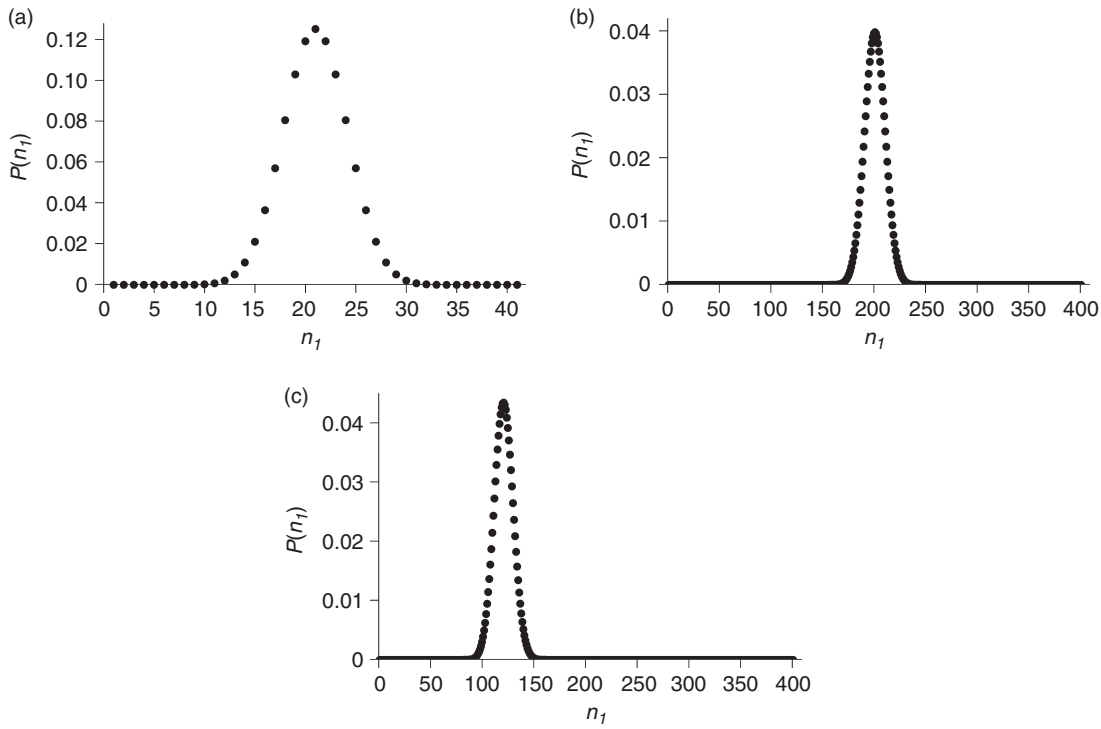

Figure 1.1 Examples of binomial distributions, Eq. (1.25.) (a) $N = 40$; $p = 0.5$; $\overline{n}_1 = 20$; $\sigma^2(n_1) = 10$; $\delta(n_1) = 0.16$. (b) $N = 400$; $p = 0.5$; $\overline{n}_1 = 200$; $\sigma^2(n_1) = 100$; $\delta(n_1) = 0.05$. (c) $N = 400$; $p = 0.3$; $\overline{n_1} = 120$; $\sigma^2(n_1) = 84$; $\delta(n_1) = 0.076$.

We thus conclude that, for the symmetric one-dimensional random walk, N is a measure of the variance, $N^{1/2}$ is a measure of the standard deviation, and $N^{-1/2}$ is a measure of the relative fluctuation, with l a scale factor. Moreover, because rightward and leftward steps are equally probable, the mean effective displacement vanishes, as might be intuitively expected.

1.4.2 The Gaussian distribution

For large N, the binomial distribution $P(n_1)$ given by Eq. (1.25) acquires a very pronounced maximum at $n_1 = \overline{n_1}$, dropping rapidly as n_1 moves away from $\overline{n_1}$ (see Figure 1.1). This enables us to find an approximate expression for $P(n_1)$ when $N \to \infty$:

$$P(n_1) = \frac{1}{\sqrt{2\pi\sigma^2(n_1)}} \exp\left[-\frac{1}{2}\frac{(n_1 - \overline{n_1})^2}{\sigma^2(n_1)}\right], \tag{1.44}$$

where $\overline{n_1}$ and $\sigma^2(n_1)$ are given by Eqs. (1.30) and (1.32), respectively. Equation (1.44) is known as the *Gaussian distribution* or *standard normal distribution*. Gaussian distributions

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-107-05378-6 - Introduction to Statistical Physics and to Computer Simulations
João Paulo Casquilho and Paulo Ivo Cortez Teixeira
Excerpt
More information

occur very often (but by no means in all cases!) in statistics when one has to deal with large numbers of trials.

Proof of Eq. (1.44) Start by making the change of variables $n_1 = \overline{n_1} + k$, with $\overline{n_1}$ given by Eq. (1.30), in Eq. (1.25):

$$P(k) = \frac{N!}{(Np+k)!(Np-k)!} p^{Np+k} q^{Np-k}.$$

Now use Stirling's formula, Eq. (1.126), for $N!$ when $N \gg 1$:

$$N! \simeq \sqrt{2\pi N} N^N e^{-N},$$

with the result

$$\begin{aligned} P(k) &= \sqrt{2\pi N} N^N e^{-N} \times \frac{p^{Np+k}}{\sqrt{2\pi(Np+k)}(Np+k)^{Np+k} e^{-(Np+k)}} \\ &\quad \times \frac{q^{Nq-k}}{\sqrt{2\pi(Nq-k)}(Nq-k)^{Nq-k} e^{-(Nq-k)}} \\ &= \frac{1}{\sqrt{2\pi N(p+k/N)(q-k/N)}(1+k/Np)^{Np+k}(1-k/Nq)^{Nq-k}}, \end{aligned}$$

whence

$$\begin{aligned} \ln P(k) = &-\frac{1}{2}\ln\left[2\pi N(p+k/N)\,(q-k/N)\right] \\ &-(Np+k)\ln(1+k/Np) - (Nq-k)\ln(1-k/Nq). \end{aligned}$$

For $k/Np \ll 1$ and $k/Nq \ll 1$ we can truncate the series expansions of the logarithms at low order:

$$\begin{aligned} \ln\left(1+\frac{k}{Np}\right) &\simeq \frac{k}{Np} - \frac{k^2}{2N^2p^2} + \frac{k^3}{3N^3p^3}, \\ \ln\left(1-\frac{k}{Nq}\right) &\simeq -\frac{k}{Nq} - \frac{k^2}{2N^2q^2} - \frac{k^3}{3N^3q^3}. \end{aligned}$$

Substitution into the expression for $\ln P(k)$ then yields:

$$\begin{aligned} \ln P(k) = &-\frac{1}{2}\ln\left\{2\pi N\left[pq - \frac{k}{N}(p-q) - \frac{k^2}{N^2}\right]\right\} \\ &-\frac{k^2}{2Np} - \frac{k^2}{2Nq} + \frac{k^3}{6N^2p^2} - \frac{k^3}{6N^2q^2}, \end{aligned}$$

or, equivalently,

$$P(k) = \frac{\exp\left(-\frac{k^2}{2Npq} - \frac{k^3}{6N^2}\frac{p-q}{p^2q^2}\right)}{\sqrt{2\pi N\left[pq - \frac{k}{N}(p-q)\frac{k^2}{N^2}\right]}}.$$

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-107-05378-6 - Introduction to Statistical Physics and to Computer Simulations
João Paulo Casquilho and Paulo Ivo Cortez Teixeira
Excerpt
More information

Further neglecting terms of order k/N or higher, we find:

$$P(k) = \frac{\exp\left(-\frac{k^2}{2Npq}\right)}{\sqrt{2\pi Npq}}$$

Changing variables back to n_1 then gives Eq. (1.44).

The probability that, after a very large number of steps N, the effective number of rightward steps, as defined by Eq. (1.33), is n, follows from Eq. (1.44) combined with Eqs. (1.30) and (1.32), noting that $n_1 = (N+n)/2$ (Reif, 1985):

$$P(n) = \frac{1}{\sqrt{2\pi Npq}} \exp\left\{-\frac{[n-N(p-q)]^2}{8Npq}\right\}, \tag{1.45}$$

where we have used the fact that $n_1 - Np = (N+n-2Np)/2 = [n-N(p-q)]/2$. This result may be seen as a special case of the *central limit theorem*, to which we shall come back later: if a random variable (the effective number of rightward steps) is the sum of a very large number of other, statistically independent, random variables (all individual steps), then its distribution is Gaussian.

The above distribution can be re-expressed in terms of the effective rightward displacement, defined by Eq. (1.34). If the step length l is much smaller than the relevant lengthscale of the system under study (e.g., its linear dimension), the discrete variable L with increment $\delta L = 2l$ may be replaced by a continuous, 'macroscopic' variable x with increment $dx \gg l$. Then the probability that after $N \gg 1$ steps the effective rightward displacement lies between x and $x + dx$ is, by definition (1.11), the sum of $P(n)$ over all n in dx, of which there are $dx/2l$. Because $P(n)$ is approximately constant in such a narrow interval, to a good approximation the probability we seek is just $P(n)$ times $dx/2l$, whence we can write

$$f(x)dx = P(n)\frac{dx}{2l}, \tag{1.46}$$

where $f(x)$ is the probability density, according to definition (1.12). Equations (1.45) and (1.46) finally yield

$$f(x)dx = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right] dx, \tag{1.47}$$

where

$$\mu = Nl(p-q), \tag{1.48}$$

$$\sigma^2 = 4Nl^2pq. \tag{1.49}$$

Equation (1.47) is the most common form of the Gaussian distribution for a continuous variable. It gives *the probability that after N steps of size l each, the random walker will find itself at a distance between x and x + dx from its starting point*. This distribution is symmetric about μ, which as we shall see shortly is its mean value. Figure 1.2 shows some exemplary Gaussian distributions.

© in this web service Cambridge University Press www.cambridge.org