The little book of kinetic theory

all you need to know for undergraduate physics

draft version: NOT FOR DISTRIBUTION

Andrew Steane

CLARENDON PRESS • OXFORD
2026

Contents

1 Introduction **1**

2 Models of a gas **3**
2.1 The quantum limit 5

3 Probability **7**
3.1 Discrete probabilities 8
3.1.1 Symmetry and equal likelihood 8
3.1.2 Mean, variance and standard deviation 10
3.1.3 Binomial distribution 13
3.2 Continuous probability distribution 14
3.2.1 Mean, variance, standard deviation 16
3.2.2 Working with distribution functions 17
3.3 The normal distribution 18
3.4 Cauchy distribution 20
3.5 Exponential and Poisson distributions 20
3.6 Independence and correlation 22
3.7 Relating one distribution to another 24
3.8 Afterword: On coincidence involving large numbers 26

4 The pressure in a gas **31**
4.1 Pressure in very simple models 31
4.1.1 Relation between pressure and energy density 34
4.2 Pressure in a more realistic model 35
4.3 Summary 40

5 Speed and velocity **43**
5.1 Notation for a distribution of velocity 43
5.2 Isotropic velocity distribution 44
5.3 Conclusion 47

6 Effusion **49**
6.1 Pressure calculation using polar coordinates 49
6.2 Effusion calculation 51
6.2.1 Effusion between regions at different temperature 51
6.3 The distribution over speed and angle 52
6.4 Non-effusive flow 54

7 The Maxwell-Boltzmann distribution **57**
7.1 State space and Boltzmann factor 59

7.2 The distribution of each component 60
7.2.1 Doppler broadening 62
7.3 Exercises 63

8 The heat capacity of a gas 65
8.1 Main features for any gas 66
8.2 Equipartition theorem 66
8.3 Exercises 72

9 Mean free path 73
9.1 The collision cross-section 73
9.2 Average free path length as a function of speed 74
9.3 The distribution of path lengths 79
9.3.1 Some properties of the distribution of path lengths 83
9.4 Variable hard spheres and the Lennard-Jones potential 84
9.5 Calculation of $\lambda_v(v)$ 86
9.6 Exercises 89

10 Is the physical world deterministic? 93
10.1 Determinism and Newtonian physics 93
10.2 Other considerations 96
10.3 Exercises 98

11 Transport properties 99
11.1 Precision of calculation method 99
11.2 Relation of continuity equation to diffusion equation 100
11.3 Diffusion 103
11.3.1 Flux down a gradient 104
11.4 Viscosity 107
11.4.1 Simple argument 109
11.5 Thermal conductivity 112
11.5.1 Why the simple energy-flow argument cannot be trusted 113
11.5.2 Euken–Hirschfelder formula 115
11.5.3 Prandtl number 117
11.6 Exercises 118

12 Random walks 121
12.1 The basic random walk 121
12.1.1 Distribution of S_n 123
12.2 Distribution for a general walk 125
12.2.1 Generalization to any number of dimensions 128
12.3 Knudsen flow 130
12.4 Brownian motion 132
12.4.1 Thermal equilibrium with an applied force 134
12.4.2 Average motion under a stochastic and viscous force 136
12.5 Exercises 138

13 Transport calculation revisited 141

13.1 Diffusion 141
13.2 Viscosity 143
13.3 Thermal conductivity 145
13.4 Exercises 146

14 The Boltzmann transport equation 147
14.1 Fluid equations 149
14.2 Solving the Boltzmann equation 151
14.2.1 Relaxation time approximation 152
14.3 Chapman-Enskog theory 157

15 The relativistic gas 159
15.1 Basic relativistic formulae 161
15.1.1 Pressure and effusion 161
15.1.2 Distribution of momentum and energy 162
15.2 Energy and equation of state 163
15.3 Transport properties 165
15.4 Exercises 165

A Rectangular and polar coordinates 167
A.1 Coordinates in a plane 167
A.2 Coordinates in three dimensions 167

B Kinetic integrals and how to find them 169
B.1 Exponential functions with powers 169
B.2 Gaussian functions with powers 170

C Random walk in multiple dimensions 171

D Principle heat capacities of ideal gas 173

E Normal modes of vibration of molecules 175

F Conservation laws and the Boltzmann equation 177
F.1 Conservation of particles 177
F.2 Conservation of momentum 177
F.3 Conservation of energy 180

G The Boltzmann collision operator 183

Introduction

1

Kinetic theory is at once a humble and a sophisticated part of modern physics. Humble because it starts with simple ideas: little point-like particles whizzing about and bumping off things. Sophisticated because it ends in complex dynamical systems such as shock-waves near aircraft wings, exploding stars, and a myriad other examples which are described by multi-dimensional partial differential equations which no-one has ever fully solved.

The present book is written with the first-year undergraduate student in mind. Ideas such as the mathematics of probability are introduced without assuming any prior knowledge. Also, the text aims to show how science is often done: how we creep up on complex things by introducing simple models to begin with and then developing them further. The early chapters treat pressure and effusion (leakage) of gases without requiring one to know any particular velocity distribution (such as the Maxwell-Boltzmann distribution) of the particles in a gas, or any model of collisions. This makes it clear that some of the results are quite general and not restricted to conditions of thermal equilibrium alone.

The text then describes the Maxwell-Boltzmann distribution and gives brief comments on the statistical methods used to derive it. The heat capacity of a gas is then discussed by using the equipartition theorem. Next we consider collisions and the mean free path, and use these concepts in a simple treatment of diffusion, viscosity and thermal conduction.

We then describe in detail the mathematical and physical idea of a random walk, and use it both to treat Brownian motion, and also to develop a more precise treatment of diffusion, viscosity and thermal conduction without requiring any further mathematical sophistication.

Finally, the text turns to what may be called advanced topics: the Boltzmann transport equation, and the relativistic gas. Many university courses would only wish to touch lightly on these. In any case they are beyond first-year level. The treatment here is suitable for undergraduate courses at second- or third-year level.

Why do we need this new book on a subject which has been part of the standard physics curriculum for many years? The reason is that the existing introductory treatments are just a bit too simplistic, in my opinion, whereas the graduate- or research-level monographs do not provide the best way in. By 'too simplistic' I mean, for example, the following:

(1) Letting students think that molecules bounce off walls like light reflecting off a mirror without warning them that this is not true

and quite misleading.

(2) Omitting the case of anisotropic velocities when it is easily included.

(3) Incorrectly asserting that the velocity distribution in an effusing beam differs from that in a gas.

(4) Ignoring the speed-dependence of the mean free path.

(5) A widely-used rough treatment of thermal conduction which has built into it an inconsistency which renders it unreliable.

(6) Assertions about temperature-dependence of transport coefficients which ignore the temperature-dependence of the collision cross section and consequently are wrong.

(7) Either no or very little information about random walks.

So this book, in addition to being, I hope, a clear and helpful resource for students, will also introduce into the pedagogy methods to treat the above more fully or more correctly, without the explicit use of the Boltzmann transport equation. The text also shows how the transport equation is used, with a fully worked example adopting the relaxation time approximation. Some of the material on the relativistic gas is also original.

Models of a gas

2

In science generally, a *model* is a set of ideas which is intended to capture the essence of some physical behaviour, and allow it to be understood, at least in some respects. For example, we can model the solar system by supposing that it consists of precisely spherical planets orbiting a precisely spherical Sun which does not itself move, and there are no moons, comets or asteroids. Such a model gives a good basic insight into the solar system, but we know from the outset that it will not be perfectly precise because in fact planets and the Sun are not precisely spherical, the Sun does move, and there are moons, comets and asteroids. In a similar way, in the kinetic theory of gases we propose a model of a gas, quite a simple one in the first instance, and we seek to find out what we can about the gas based on our model. We then develop the model by introducing more sophistication, and we keep testing our ideas by checking their internal coherence and by comparing with experimental results.

In the kinetic theory of gases, we aim to understand some of the properties of ordinary gases, by considering that a gas consists of a large number of very small things (typically either atoms or molecules) flying about randomly.

The very small entities which make up most gases are molecules, but there are gases (e.g. the noble gases) composed of single atoms. They are called *monatomic* gases. A single atom could be regarded as a simple kind of molecule (one with only one atom) so we could just use the word 'molecule' for the parts which make up any gas, but this is not the standard terminology. Sometimes the word 'particle' is used, and sometimes the term 'quasi-particle', if it is desirable to make it clear that we do not mean a fundamental particle with no diameter or sub-structure. In this book we will mostly use the terminology of 'particle' and 'molecule'. It is not necessary to assume that these 'particles' are strictly point-like, only that their size is small compared to the average distance between them.

At first we imagine the particles of the gas can be regarded as so small that we need not concern ourselves with their size, and each is treated as a point particle with a given mass and velocity (Fig. 2.1a). This picture is enough to teach us some basic facts about the pressure in a gas (chapter 4) and about the way particles emerge from a small hole in a container of gas (chapter 6). We can also find the velocity distribution (chapters 5 and 7).

A weakness of the point-particle model is that it cannot account for

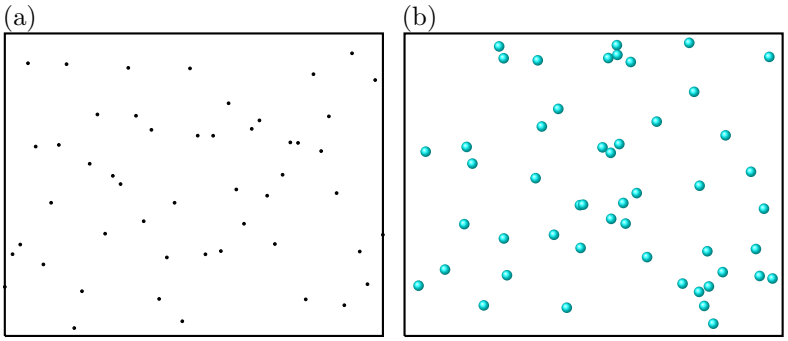


Fig. 2.1 Two models of a gas. (a) point particle model, (b) hard spheres model.

the way the gas particles collide with one another as they move about. But these collisions are an important feature of a gas, as we shall see. So we develop our model by introducing a little more sophistication. Now we suppose that the particles have some finite size, so that as the move about they can hit one another, but we suppose that otherwise they do not interact. This model is called the 'hard spheres' model, because it pictures the gas particles as little spheres that do not interact at all when they are not touching, but which interact like hard billiard balls when they collide (Fig. 2.1b). Like the point particle model, this is an approximation, because in fact gas particles exert small but non-zero forces on each other no matter how far apart they are. Also, although single atoms are spherical to good approximation, most molecules are not. But the hard spheres model is adequate for approximate calculations, and it gives a very good general insight into the behaviour. We will use it in chapter 11 to find out about diffusion, viscosity and conduction of heat.

One reason why the hard spheres model works for a gas which is not too dense is that there is a *separation of distance scales*. The distance scales relevant to a gas include the diameter L of some vessel containing the gas, the average separation s between the molecules, the radius r_c of a single molecule, and the average distance λ travelled by any given molecule before it hits another. The typical situation is

$$L \gg \lambda \gg s \gg r_c. \tag{2.1}$$

For example, in ordinary conditions one might have the values (in metres),

$$10^{-2} \gg 10^{-7} \gg 3 \times 10^{-9} \gg 10^{-10}.$$

When the density of the gas is very high or very low, one or more of these inequalities may not hold, and in that case we have to reconsider whether the model can be trusted.

In all these models, the question arises as to whether we should allow for internal states of the particles of the gas. For example, a simple particle has no internal states, but a molecule has: it can rotate and

vibrate, and there are many different states of rotation and vibration. So it looks at first as if allowing for this will be difficult. It turns out, however, that in conditions of thermal equilibrium the internal states can be allowed for quite easily, by averaging.

2.1 The quantum limit

Finally, the question arises as to whether we can treat the gas particles using classical (Newtonian) physics, or whether quantum theory is needed. Once again this a question of degree of approximation. In ordinary circumstances, which is to say conditions of thermal equilibrium or near to thermal equilibrium (not some very special state brought about by an experimenter, for example) then the classical treatment yields accurate results as long as the particles are not so densely packed that their de Broglie waves cannot fail to overlap. A good rule of thumb for this can be obtained as follows. Consider the thermal energy $k_{\mathrm{B}}T$ where k_{B} is Boltzmann's constant. A particle of mass m with this amount of energy will have momentum

$$p_{\mathrm{th}} = \sqrt{2mk_{\mathrm{B}}T} \tag{2.2}$$

and consequently a de Broglie wavelength

$$\lambda_{\mathrm{dB}} = \frac{h}{p_{\mathrm{th}}} = \frac{2\pi\hbar}{\sqrt{2mk_{\mathrm{B}}T}}. \tag{2.3}$$

Let n be the number density of gas particles. This is the number of gas particles per unit volume. A gas at temperature T and number density n is far from the quantum limit as long as

$$n\lambda_{\mathrm{dB}}^3 \ll 1. \tag{2.4}$$

In other words, if this inequality holds then Newtonian physics is sufficient to treat the motion of the particles, so as to find out about bulk properties such as pressure, viscosity, thermal conduction and the like. All the calculations presented in this book are comfortably in this Newtonian regime, with the exception of exercise 3 below.

©A.M. Steane 2026, not for distribution.

Exercises

(1) What are the main distance scales relevant to the treatment of a gas by kinetic theory?

(2) What is the 'hard spheres model' for a gas?

(3) Let n be the number of particles per unit volume. We will show in chapter 7 that for a gas in thermal equilibrium, this is related to pressure P and temperature T by $P = nk_{\mathrm{B}}T$. Find n and λ_{dB} given by eq. (2.3), and hence evaluate $n\lambda_{\mathrm{dB}}^3$ for (i) nitrogen at $10^5\,\mathrm{Pa}$ and $300\,\mathrm{K}$; (ii) helium at $10^5\,\mathrm{Pa}$ and $4\,\mathrm{K}$, and comment. [*Ans.* 9.29×10^{-7}; 0.837]

Probability

3

In this chapter we present basic mathematical aspects of probability.

Probability concerns the quantitative treatment of situations where more than one outcome may occur. For example, when we flip an ordinary coin, two outcomes may occur—heads or tails (here we assume that if a coin were to land balanced on its edge then it would simply be flipped again). In the throw of an ordinary dice, six outcomes may occur. In the length of a human nose, the outcomes that may occur fall in a continuous range.

In examples like these it can happen that if we had further information then the concept of probability might not arise. For example, if we knew the precise amount, direction and location of the forces on a coin when it was flipped, and we also knew the local gravity and atmospheric conditions, then it can happen that we could predict with close to certainty on which side the coin will settle after it landed on some fixed surface such as a table. In this case the probabilities (for a person having such further information) are not half and half for heads and tails, but close to one for the predicted outcome, and close to zero for the other (not quite exactly one and zero because of the chance that we made a mistake or failed to account for some further influence). This shows that sometimes probability can reflect an absence of information or knowledge. But it can also happen that it would be impossible to gather enough information to predict some outcome. Examples of this kind occur in chaos theory in classical mechanics, and in quantum theory; this point is discussed in chapter 10.

The precise meaning of the term 'probability' has been pondered and argued-over at length by philosophers. The two main views are termed *frequentist* and *Bayesian*. Sometimes the term 'likelihood' is reserved for the second view.

On a frequentist view, the probability of some outcome is equal to the proportion of occasions on which that outcome takes place, if the same process is repeated many times, in the limit of a large number of trials. So if we roll a dice 6000 times, and of these, 1009 rolls produced a 2, then we may say that on each occasion, the probability of throwing a 2 is close to $1009/6000 \simeq 1/6$. Or if we consider an idealized dice, which is a cube which by definition has equal likelihood of showing any one of its 6 faces, then the probability of obtaining a 2 upon rolling such a dice is exactly $1/6$.

On a Bayesian view, probability is all about reasonable expectation. It is a way of quantifying what it would be reasonable to bet, in view of

the evidence one has. For example, if a dice has cubic symmetry and is not weighted, then it would be reasonable to expect that it would show any of its faces with equal probability, so the probability of obtaining a two is 1/6, and we do not need to consider any trials at all. This way of thinking about probability is especially useful when it is not in fact possible to produce many trials. For example, it is useful when assessing the safety of nuclear power plants or the probability of one-off events in human lives or in the evolution of the universe at large.

Although the frequentist versus Bayesian debate is interesting philosophical territory, we will not need to take a view on it for the purposes of this book. For our purposes we will cheerfully switch between the views, sometimes adopting a more frequentist style, sometimes more Bayesian. They are not mutually contradictory but rather illuminate different aspects of what probability can involve.

3.1 Discrete probabilities

3.1.1 Symmetry and equal likelihood

The easiest example of probability is one we already considered above when we proposed an idealized dice. This is the case of equal likelihood owing to a symmetry. If each of N outcomes can occur, and the situation is balanced, or symmetric, such that nothing prefers one outcome over another, then the probability, P_j, of any given outcome, j, is $1/N$. This illustrates an important feature of probability: once one has accounted for all possible outcomes, *the probabilities of independent outcomes must sum to 1*:

$$\sum_{j=1}^{N} P_j = 1.$$

When all the outcomes are equally likely we have $P_j = p$ for some p, and this sum takes the form

$$\sum_{j=1}^{N} p = 1.$$

Hence $Np = 1$ so $p = 1/N$.

An important fact about probability is illustrated by a human situation that often catches out people who are drawn to gambling for money. If it has already been established that some given coin is fair, and we observe, flipping it, a string of 5 *heads*, then there is a natural human tendency to feel that the coin 'ought' to produce a *tails* next, or at least that the chance of getting tails is now greater than one half. This is a feeling that many a gambler at a roulette wheel has had after observing several instances of *black* in a row. This feeling must be resisted. If the coin is tossed randomly on each occasion then the chances of what it will do next are utterly unrelated to what it has done up till now. The probability of *tails* remains strictly one half and we ought not to

imply that the coin has any 'duty' to make up for past 'prejudice' by now leaning in favour of tails until the balance of outcomes is restored.[1]

When the behaviour at the next trial is not affected by those that went before we say the trials are *independent*.

Next let's consider the example of a pack of playing cards. The pack consists of 52 cards, in four types called 'suits' going by the names *hearts, spades, diamonds and clubs*, with thirteen cards in each suit. We make the reasonable idealization that after the pack has been shuffled, and a single card is drawn, the probability of drawing any given card will be $p = 1/52$, no matter which card it happens to be. Hence the answer to the question, "what is the probability of drawing the ace of spades"? is $1/52$. And the probability of drawing the two of clubs, or any other card, is the same.

Now suppose we ask "what is the probability of drawing a two?" We can answer this by appeal to the frequentist view of probability. We imagine an endless sequence of draws, where each time a card is drawn it is then returned and the pack is reshuffled. In this endless sequence we find the number of twos by *adding*. Let $N_{2\heartsuit}$ be the number of occurrences of two of hearts, and let $N_{2\clubsuit}$, $N_{2\diamondsuit}$, $N_{2\spadesuit}$ be the numbers of two of clubs, diamonds, spades, respectively. Then the total number of twos in the sequence of draws is

$$N_2 = N_{2\heartsuit} + N_{2\clubsuit} + N_{2\diamondsuit} + N_{2\spadesuit}.$$

If there are N draws altogether, then the proportion giving a two, and therefore the probability of getting a two on any given occasion, is

$$P(2) = \frac{N_2}{N}.$$

Hence

$$\begin{aligned} P(2) &= \frac{N_{2\heartsuit} + N_{2\clubsuit} + N_{2\diamondsuit} + N_{2\spadesuit}}{N} \\ &= \frac{N_{2\heartsuit}}{N} + \frac{N_{2\clubsuit}}{N} + \frac{N_{2\diamondsuit}}{N} + \frac{N_{2\spadesuit}}{N} \\ &= P(2\heartsuit) + P(2\clubsuit) + P(2\diamondsuit) + P(2\spadesuit) \qquad (3.1) \\ &= \frac{4}{52} = \frac{1}{13}. \end{aligned}$$

I expect this result is already familiar to most readers, but it has served as a useful exercise in seeing how one of the rules concerning probabilities comes about. This is the rule that *probabilities of independent outcomes add*. That is to say, the total probability of something that can happen in several ways independently is obtained by adding the probabilities of the various ways.

This naturally leads to more general probability distributions where not all the events are equally likely. We could for example group the

[1] If a real coin gave a string of very many heads then a correct physical interpretation would include for consideration the possibility that the coin is not in fact a fair coin. In that case it may be assessed more likely that it will continue to show further heads.

playing cards into those numbered 1 to 2, those 3 to 10, and the others. The probabilities of drawing a card in each of these groups are, respectively,

$$\frac{2}{13}, \qquad \frac{8}{13}, \qquad \frac{3}{13}.$$

3.1.2 Mean, variance and standard deviation

The *average* of a set of N numbers x_j is defined to be the sum of the numbers divided by N:

$$\text{average} = \frac{\sum_{j=1}^{N} x_j}{N}$$

For example, the average of 3, 5 and 10 is $18/3 = 6$. This concept also applies when some of the x_j are equal. For example the average of $\{3, 3, 4, 5, 7, 7, 7, 12\}$ is 6. Now let's suppose each of those numbers refers to the measured value of some physical quantity such as the length, in centimetres, of a dinosaur tooth. A useful way to write the sum is to gather the values into groups:

$$\begin{aligned} 3+3+4+5+7+7+7+12 &= (3+3)+4+5+(7+7+7)+12 \\ &= (2\times 3)+(1\times 4)+(1\times 5)+(3\times 7)+(1\times 12) \end{aligned}$$

Generalizing, we have that

$$\text{average} = \frac{1}{N}\sum_{k=1}^{g} g_k x_k$$

where g_k is the size of the k'th group and g is the number of groups. (In the example just considered, we have $g = 5$, $g_k = \{2, 1, 1, 3, 1\}$, $x_k = \{3, 4, 5, 7, 12\}$). Now we observe that if the above refers to a sequence of measurements of teeth, and the frequencies with which the various lengths are found reflect their probabilities, then the probabilities are $P_k = g_k/N$, so have

$$\text{average length} = \sum_{k=1}^{g} P_k x_k. \tag{3.2}$$

In practice, if we only measure 8 specimen teeth, then we cannot obtain precise knowledge of the probabilities, so we cannot claim to know that P_k is exactly equal to g_k/N when $N = 8$. However, we can imagine a larger set of measurements, giving a large N, and in that larger set we will obtain $g_k = NP_k$ occurrences of the value x_k, because this is the definition of probability. So on this basis we can expect that the average outcome will be given by eq. (3.2). We therefore adopt this expression as the *definition* of the average, when a probability distribution is involved. Such an average is usually referred to by one of two names: either **mean value** or **expected value**. The result is commonly written using an angle bracket, thus:

$$\langle x \rangle \equiv \sum_{k} P_k x_k. \tag{3.3}$$

Another notation is the overbar: $\bar{x} \equiv \langle x \rangle$. Both these notations are widely employed in the physical sciences and in this book we shall use them both. They are exact synonyms: just different ways of writing the same thing. In mathematical discussions of probability the notation $E[x]$ is also used.

We can also take an interest in quantities related to x, such as x^2 and $1/x$. The above argument will carry through precisely as before, and we shall find that the mean value of x^2, for example, is given by

$$\langle x^2 \rangle = \sum_k P_k x_k^2. \tag{3.4}$$

By the same argument, we must find

$$\langle x^0 \rangle = \sum_k P_k x_k^0. \tag{3.5}$$

But $x^0 = 1$ and $x_k^0 = 1$. Hence

$$\langle 1 \rangle = \sum_k P_k = 1. \tag{3.6}$$

This makes sense: the average value of 1 is 1.

Now suppose we ask for the mean value of mx where m is some constant. We shall find

$$\langle mx \rangle = \sum_k P_k m x_k = m \sum_k P_k x_k = m \langle x \rangle. \tag{3.7}$$

We can also ask for the mean value of a constant such as π or the speed of light. We must get simply that constant. The formula works as follows:

$$\langle c \rangle = \langle c x^0 \rangle = c \langle x^0 \rangle = c. \tag{3.8}$$

Observe also that if c is some constant having the same physical dimensions as x, then

$$\begin{aligned} \langle x - c \rangle &= \sum_k P_k (x_k - c) = \sum_k (P_k x_k - P_k c) \\ &= \left(\sum_k P_k x_k \right) - \left(\sum_k P_k c \right) = \langle x \rangle - c. \end{aligned} \tag{3.9}$$

An example of such a constant is the mean value itself, $\langle x \rangle$. With this in mind you can immediately find that $\langle x - \langle x \rangle \rangle = 0$. This says that "the average of value of the difference of x from its mean is zero".

Observe next that the mean of a sum is the sum of the means:

$$\begin{aligned} \langle x + y \rangle &= \sum_k P_k (x_k + y_k) = \left(\sum_k P_k x_k \right) + \left(\sum_k P_k y_k \right) \\ &= \langle x \rangle + \langle y \rangle. \end{aligned} \tag{3.10}$$

The algebra here is straightforward, but we need to check the meaning. The first step makes sense when the two properties x and y are associated with the physical system in such a way that when the system is in the k'th state then $x = x_k$ and $y = y_k$. For example, consider a dice with two numbers written on each face. Or, for a more practical example, let k count the physical states available to molecules in a gas, and let x and y be **independent** physical properties such as translational kinetic energy and rotational kinetic energy. Then the mean total energy will be the sum of the mean translational kinetic energy and the mean rotational kinetic energy.

We often take an interest in the amount by which some set of values is spread around the mean. We would like to indicate whether the values mostly lie close to the mean, like the frequencies of sounds emitted by a tuning fork, or whether they are spread widely around the mean, like the frequencies of sounds emitted by a piano. The deviation of any one value x_k from the mean is $(x_k - \langle x \rangle)$. The mean squared deviation is called the **variance**:

$$\mathrm{Var}\,(x) \equiv \langle (x - \langle x \rangle)^2 \rangle = \sum_k P_k (x_k - \langle x \rangle)^2.$$

An important result is obtained by expanding the bracket in this expression:

$$\begin{aligned} \mathrm{Var}\,(x) &= \sum_k P_k (x_k - 2x_k \langle x \rangle + \langle x \rangle^2) \\ &= \langle x^2 \rangle - 2\langle x \rangle \langle x \rangle + \langle x \rangle^2 \\ &= \langle x^2 \rangle - \langle x \rangle^2. \end{aligned} \tag{3.11}$$

Perhaps a nicer way to write the algebra here is not to write down the sum explicitly, but rather appeal to our previous results, thus:

$$\begin{aligned} \mathrm{Var}\,(x) &= \langle (x - \bar{x})^2 \rangle \\ &= \langle x^2 - 2x\bar{x} + \bar{x}^2 \rangle \\ &= \langle x^2 \rangle - \langle 2x\bar{x} \rangle + \langle \bar{x}^2 \rangle \\ &= \langle x^2 \rangle - 2\langle x \rangle \bar{x} + \bar{x}^2 \\ &= \langle x^2 \rangle - \langle x \rangle^2, \end{aligned} \tag{3.12}$$

where, as was explained above, $\bar{x}$ and $\langle x \rangle$ are two ways to write the same quantity. I employed both notions here because I think this makes the algebra easier to read. It is a useful educational exercise to make sure you are clear about the validity of each of these steps, and exactly what is going on. Keep in mind that x is a random variable, but $\bar{x}$ is a constant, i.e. not changing from one trial to the next for a given probability distribution.

The final expression in eq. (3.12) is often convenient in practice when one wants to calculate the variance.

The square root of the variance is called the **standard deviation**, written Δx or σ_x:

$$\Delta x = \sqrt{\mathrm{Var}\,(x)} = \sqrt{\langle (x - \langle x \rangle)^2 \rangle}.$$

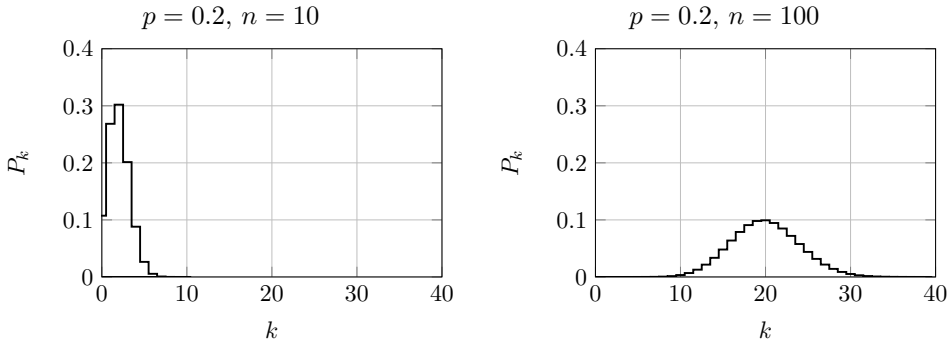


Fig. 3.1 The binomial distribution for $p = 0.2$ with $n = 10$ trials (left) and $n = 100$ trials (right).

The standard deviation is thus the *root-mean-squared deviation* or r.m.s. deviation. Note that the standard deviation has the same physical dimensions as x (whereas the variance has the physical dimensions of x^2.) If the standard deviation is small compared to the mean then the distribution of values of x is comparatively narrow. If the standard deviation is similar to the mean, or larger, then the distribution is broad.

> **Example 3.1** A certain task takes from 1 to 7 days to complete, all equally likely. What is the mean and standard deviation of the completion-time distribution?
> *Solution.*
> $(1 + 2 + 3 + 4 + 5 + 6 + 7)/7 = 4$ so the mean is 4 days.
> $(1^2 + 2^2 + \ldots + 7^2)/7 = 20$ so the standard deviation is $\sqrt{20 - 16} = 2$ days.

3.1.3 Binomial distribution

The *binomial distribution* is a probability distribution for a discrete random variable which often arises in simple physical and engineering situations. Suppose a random variable has two possible values, occurring on each trial with probabilities p and $(1 - p)$ respectively. Then in n trials the probability to obtain k occurrences of the first outcome (and therefore $n - k$ of the other outcome) is

$$P_k(p, n) = \binom{n}{k} p^k (1 - p)^{n-k} \tag{3.13}$$

where the bracket notation signifies the binomial coefficient:

$$\binom{n}{k} \equiv \frac{n!}{k!(n - k)!}. \tag{3.14}$$

This distribution can be applied, for example, to tosses of a biased coin, and to sending binary digits down a noisy channel (if the noise is the

same on each transmission), and to sampling from a bag containing balls of two colours, with replacement (that is, the ball is replaced after each trial so that the probability p remains fixed).

We typically apply this distribution to a case where n and p are fixed, and we take an interest in the number of occasions on which the first outcome takes place. Let us call this number x. It could be, for example, the number of 'heads' in a sequence of n coin tosses. We obtain for the mean and variance of x:

$$\langle x \rangle = \sum_k x_k P_k = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = np \tag{3.15}$$

$$\mathrm{Var}\,(x) = np(1-p) \tag{3.16}$$

(exercise 1).

The distribution is shown in Fig. 3.1 for a fixed value of p and two example values of n. The binomial distribution tends to the normal distribution (described in section 3.3) in the limit $n \to \infty$; the plotted examples illustrate this.

3.2 Continuous probability distribution

It often happens that we are interested in a random variable x which can take a continuous range of values. We can handle this by introducing the concept of *probability density*. We introduce a function $f(x)$ whose value is defined such that

$$\mathrm{Prob}(A \le x \le B) = \int_A^B f(x)\,\mathrm{d}x. \tag{3.17}$$

The function f defined by this expression is called the **probability density**, or the *probability density function*, often shortened to 'pdf'. Its integral over a given range gives the probability that a trial observation of x will find the value to be in the given range.[2]

We can immediately make two important observations about the probability density function, by considering a wide and then a very narrow range of values. First suppose we take as A and B the smallest and largest values that x can possibly attain. Suppose for the sake of argument that these values are $-\infty$ and $+\infty$. Then we shall find that

$$\mathrm{Prob}(-\infty \le x \le \infty) = \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1.$$

The result must be 1 since x cannot fail to be somewhere in the entire allowed range. Thus we find that *the probability density function is normalized*.

[2] In (3.17) we included both limits in the range specified for x on the left. In ordinary circumstances it will not matter whether one specifies the range as $A < x < B$ or $A \le x < B$ or $A < x \le B$ or $A \le x \le B$. However if one or both of the end-points gives a special case then one might have to reconsider this.

Next suppose we take a very narrow range of values, from A to $A+\delta x$ where δx is small:

$$\text{Prob}(A \le x < A + \delta x) = \int_A^{A+\delta x} f(x)\,\mathrm{d}x = f(A)\delta x + O(\delta x^2).$$

The idea is that when the range of integration is small, f is approximately constant over that range (if you like you can use a Taylor expansion of f about A in order to check). In the limit of small quantities we can ignore the $O(\delta x^2)$ term and then we find

$$\boxed{\text{Prob(the value is in the range } x \text{ to } x + \mathrm{d}x) = f(x)\mathrm{d}x.} \quad (3.18)$$

This statement is used very often in the physical sciences; it is the central statement from which most of our further reasoning will follow. We used the phrase 'the value' on the left here because we wanted to use x on the right. For clear thinking one should distinguish between the random variable and the values it may take. In the treatment of discrete distributions above we used x for the variable and x_k for the values. In the case of a continuous distribution one may use X for the variable, and x for the values. Thus one may discuss "the probability that X is in the range x to $x + \mathrm{d}x$". It is good to make this distinction between X and x because probability is a subtle subject and requires precise thinking. However it often happens that the situation is clear enough that the distinction does not need to be made explicit in the mathematical expressions. For example, $\text{Prob}(a < X < b)$ can equally well be expressed as $\text{Prob}(a < x < b)$.

Equation eq. (3.18) is an extremely important statement about probability density functions, and the reader should ponder it at length. Notice for example that $f(x)$ **is not itself a probability**. Rather, f is a probability-per-unit-range-of-x. The combination $(f\mathrm{d}x)$ is a probability. For example, if x is a length then $f(x)$ has the physical dimensions of inverse length and $f(x)\mathrm{d}x$ is dimensionless. If x is a mass then f has the dimensions of inverse mass. If x is a momentum then f has dimensions of inverse momentum. And so on.

In the case of a continuous variable, we have to be careful how we phrase the questions we may be interested in. For example, if x is a length, and we ask 'what is the probability that x is equal to 2 cm?' then the answer comes back 'zero', no matter what the distribution may be. This is because

$$\text{Prob}(2 \le x \le 2) = \int_2^2 f(x)\,\mathrm{d}x = 0.$$

The problem is that we have asked for a precise value, but the probability that a continuous variable has a value in a narrow range tends to zero as the range tends to zero. That is,

$$\lim_{\delta x \to 0} f(x)\delta x = 0.$$

To get some information about the probability distribution, therefore, it is no use asking for the probability of some perfectly precise outcome. We have to ask always for the probability that our variable lies in some non-zero range.

3.2.1 Mean, variance, standard deviation

We define the mean, variance and standard deviation for a continuous variable by constructing expressions which reproduce those applying to discrete variables in the appropriate limit.

Consider the sum

$$\sum_k x_k P_k.$$

This would be the mean of x if x were discrete. We can now envisage that the k'th result here contains those cases where the variable lies in the range x_k to $x_k + \delta x$. Then in the limit where δx is small, P_k in this expression must be $f(x_k)\delta x$, and we obtain

$$\sum_k x_k f(x_k)\delta x.$$

In the continuous limit this sum becomes an integral. If $f(x)$ is the probability density function, then we define the mean value of our continuous variable as

$$\langle x \rangle \equiv \int x f(x) \mathrm{d}x$$

where the integral is over the entire range of possible values of x. Note the natural reading of this expression: each value of x in the integral is being 'weighted' by the probability $f(x)\mathrm{d}x$.

It is straightforward to write expressions for the mean of other quantities related to x:

$$\langle x^2 \rangle \equiv \int x^2 f(x) \mathrm{d}x,$$

$$\langle g(x) \rangle \equiv \int g(x) f(x) \mathrm{d}x$$

where g is any function of x.

The variance can now be defined as before, and the argument leading to eq. (3.12) still applies.

> **Example 3.2** Points are chosen randomly within a flat circle of radius R centred at the origin, with a uniform distribution. That is, the probability that one of the points lies inside any region of area A lying wholly inside the circle is proportional to A. Find the mean distance from the origin of points drawn from such a set, and the mean squared distance.
>
> *Solution.*
>
> Let r be the distance from the origin. We would like to find the distribution function $f(r)$ for the randomly chosen

points. The distribution is such that the probability a point lies in a ring of inner radius r and width dr is proportional to the area of the ring, so we must have

$$f(r)\mathrm{d}r = \alpha 2\pi r \mathrm{d}r$$

where α is a constant of proportionality. A convenient way to find this constant is to ensure that f is normalized:

$$\int_0^R f(r)dr = 1$$

which gives $\alpha\pi R^2 = 1$ hence $\alpha = 1/\pi R^2$. Hence we have

$$f(r) = \frac{2\pi r}{\pi R^2} = \frac{2r}{R^2}.$$

Now we can find the quantities asked about:

$$\langle r\rangle = \int_0^R r f(r)\mathrm{d}r = \frac{2}{3}R$$

$$\langle r^2\rangle = \int_0^R r^2 f(r)\mathrm{d}r = \frac{1}{2}R^2.$$

3.2.2 Working with distribution functions

If often happens that we have the functional form of some probability density function, but it has not been normalized. Here is an example:

$$f(x) \propto x^2 e^{-x^2/2}. \tag{3.19}$$

Suppose we would like to find the mean value of x. There are two ways to proceed. We could introduce a normalization constant N and find its value as follows:

$$f(x) = Nx^2 e^{-x^2/2} \implies N = \frac{1}{\int x^2 e^{-x^2/2}\mathrm{d}x}. \tag{3.20}$$

Then we write $\langle x\rangle = \int xNx^2 e^{-x^2/2}\mathrm{d}x$ and since N has been found, the integral can be done.

The second way to proceed is to notice that the above method is equivalent to writing

$$\langle x\rangle = \frac{\int xx^2 e^{-x^2/2}\mathrm{d}x}{\int x^2 e^{-x^2/2}\mathrm{d}x}. \tag{3.21}$$

In this second approach the normalization is taken care of by dividing by the integral of the distribution function, whatever that may be. This second approach sometimes offers a short-cut, since it can happen that the ratio of the integrals can be found without the need to find either of the integrals on their own (for an example see eqns (7.9),(7.10)). One can also find other averages this way:

$$\langle x^2\rangle = \frac{\int x^2 f(x)\mathrm{d}x}{\int f(x)\mathrm{d}x}, \qquad \langle g(x)\rangle = \frac{\int g(x)f(x)\mathrm{d}x}{\int f(x)\mathrm{d}x}, \tag{3.22}$$

where $g(x)$ is any function of x.

Moments of a probability distribution

The term 'moment' has several technical meanings in mathematics and physics. In the context of probability distributions it is employed as follows.

If c is a constant and $f(x)$ is a real-valued function then **the n'th moment of the function $f(x)$ about c** is

$$\mu_n \equiv \int_{-\infty}^{\infty} (x-c)^n f(x) \mathrm{d}x. \tag{3.23}$$

Applied to probability, we observe that the zeroth moment of a probability density function is 1, the first moment about zero is the mean value, and second moment about the mean is the variance. Further moments give further information about the overall shape of the function. The third moment about the mean is called *skewness*. It gives information about the degree to which the distribution is not symmetric about the mean. The fourth moment about the mean is called *kurtosis*. It gives information about the degree to which the tails of one distribution are larger or wider than those of another.

3.3 The normal distribution

An important distribution function widely encountered in science is the

Normal distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)} \tag{3.24}$$

This is also called a **Gaussian** function. It is plotted in Fig. 3.2. The two constants μ and σ characterise the location of the peak, and its width. If this is the probability density function for some random variable x, then μ is the mean and σ is the standard deviation of x. This can be deduced from

$$\langle x \rangle = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-(x-\mu)^2/(2\sigma^2)} \mathrm{d}x = \mu, \tag{3.25}$$

$$\langle x^2 \rangle = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-(x-\mu)^2/(2\sigma^2)} \mathrm{d}x = \sigma^2 + \mu^2 \tag{3.26}$$

$$\Delta x = \sqrt{\langle x^2 \rangle - \langle x \rangle^2} = \sigma. \tag{3.27}$$

(To obtain the first result here, one does not need to do any integrals; it suffices to notice that the Gaussian function is symmetrically distributed about its peak, and that peak occurs at $x = \mu$. The second result is obtained in the appendix.)

Most of the normal distribution lies within a few standard deviations of the mean. To state this in a quantitative fashion, one may give the probability that x lies within some multiple of the standard deviation from the mean. Let n be an integer. The probability that x lies in the

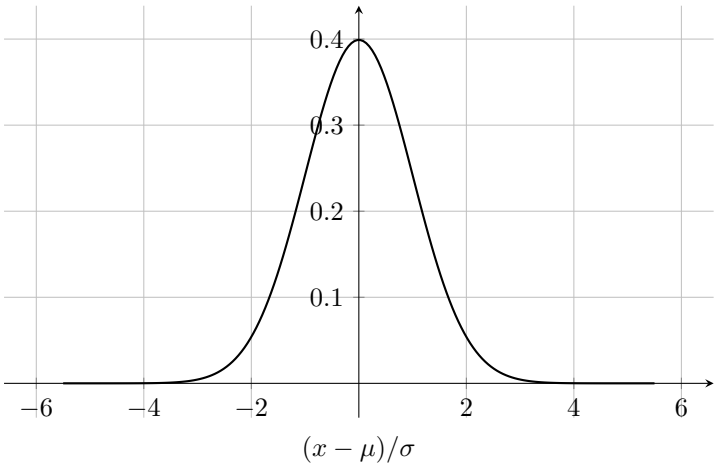


Fig. 3.2 The normal distribution.

range $-n\sigma < (x-\mu) < n\sigma$ is given by the following table for small values of n:

n	Prob$(-n\sigma < (x-\mu) < n\sigma)$
0	0
1	0.683
2	0.9545
3	0.9973
4	0.999937
5	$1 - 5.7 \times 10^{-7}$

Notice that about a third of the distribution lies outside the range of one standard deviation around the mean, and almost all of it lies within three standard deviations of the mean.

It often happens that the normal distribution is appropriate to model the effects of experimental imprecision.[3] In this context, when the experimental uncertainty has been estimated correctly, one expects to find about 68% of the data within one 'error bar' of the mean, so about a third of the data should be further out. If one thinks that the uncertainty of each data point is Δx, but almost all the data is within $\pm\Delta x$ of the mean, then this Δx is likely to be larger than the actual standard deviation. Equally, if one finds a data point at 5 notional standard deviations from the mean, then unless this is a one-off result in a million data points, the correct interpretation is either that the standard deviation has been underestimated, or that the experimental uncertainties are not in fact giving normally distributed outcomes.

[3]The reason for this is explored in chapter 12.

3.4 Cauchy distribution

The distribution

$$f(x) = \frac{1}{\pi}\frac{1}{x^2+1} \tag{3.28}$$

is called by both the names *Cauchy distribution* and *Lorentzian function*. It often occurs in frequency distributions of emitted electromagnetic radiation, because it is the Fourier transform of an exponentially decaying function. It is notable in probability theory as an example of a distribution which can be normalized but which has infinite variance. More generally, distribution functions which do not fall off exponentially with x are termed *heavy-tailed*.

3.5 Exponential and Poisson distributions

Consider a collection of radioactive atomic nuclei in ordinary circumstances where the decay of one nucleus does not affect the others. Let $t = 0$ be some initial time, at which some given nucleus has not yet decayed. The probability that this given nucleus will first survive for a time interval t, and then decay in the short interval between t and $t+\mathrm{d}t$, is

$$f(t;\gamma)\,\mathrm{d}t = \gamma e^{-\gamma t}\,\mathrm{d}t \tag{3.29}$$

where γ is a constant called the decay rate.

In the limit of a large number of nuclei in the sample, the probability that k of them decay during the time interval between 0 and t is given by[4]

Poisson distribution

$$P_k(\mu) = \frac{1}{k!}\mu^k e^{-\mu} \tag{3.30}$$

where $\mu = \gamma t$.

The distribution $f(t;\gamma)$ is the **exponential distribution**. The distribution $P_k(\mu)$ is called the **Poisson distribution**. For the Poisson distribution the random variable takes the values $x_k = k$ so you can regard k itself as the random variable, and μ is a parameter.

You can show (exercise) that for the exponential distribution,

$$\langle t\rangle = \gamma^{-1} \qquad \langle t^2\rangle = 2\gamma^{-2} \qquad \mathrm{Var}\,(t) = \gamma^{-2} \tag{3.31}$$

and for the Poisson distribution

$$\langle k\rangle = \mu, \qquad \mathrm{Var}\,(k) = \mu. \tag{3.32}$$

[4]This is an approximation since in practice there will be a finite number of nuclei but the Poisson distribution allows for all values of k between zero and infinity. Also when the decays are not independent (such as in a fission reactor) then the result does not apply.

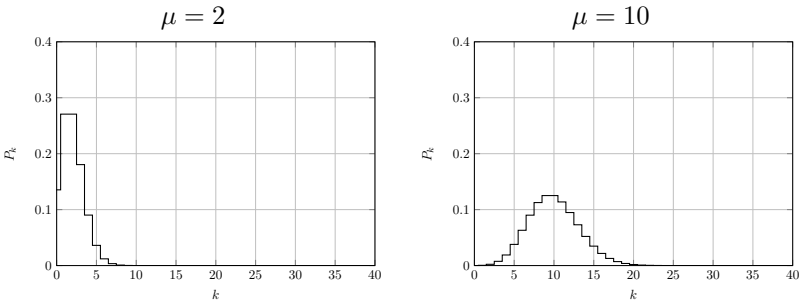


Fig. 3.3 The Poisson distribution for two values of μ.

It follows that for the Poisson distribution the standard deviation is the square root of the mean. (The property that the variance is equal to the mean in this case is not forbidden by dimensional considerations, because for the Poisson distribution k and μ are dimensionless numbers.)

In general the Poisson distribution gives the probability of a given number of events occurring in a fixed interval of time or region of space if these events occur with a constant mean rate and independently of the time or distance since the last event. It is the archetypical distribution for independent point processes and has widespread application. For example: number of photons hitting a detector in a given exposure time, number of people injured by horses in the army in a year, number of quasars in a given solid angle across the sky, ...

We already showed that the exponential distribution $f(t;\gamma)$ is related to the Poisson distribution, by introducing them in a physical context where both are relevant. Note that

$$\begin{aligned} f(t;\gamma)\mathrm{d}t &= P_0(\gamma t)\,(1 - P_0(\gamma \mathrm{d}t)) \quad + O(\mathrm{d}t^2) \\ &= P_0(\gamma t) P_1(\gamma \mathrm{d}t) \quad + O(\mathrm{d}t^2) \end{aligned} \tag{3.33}$$

We should expect this from the definitions. For: the right hand side gives the probability to survive for a time interval t multiplied by the probability to then decay in the time interval $\mathrm{d}t$, both calculated in the case of a mean decay rate γ, hence a single radioactive nucleus, and this same probability is what the left hand side also gives. This is one way to derive the Poisson distribution. It can also be derived as a limiting case of the binomial distribution (see exercises).

The probability of 1 or more Poisson-distributed events occurring is

$$P(k \geq 1) = \sum_{k=1}^{\infty} P_k(\mu) = \sum_{k=1}^{\infty} \frac{\mu^k}{k!} e^{-\mu} = (e^{\mu} - 1)\, e^{-\mu} = 1 - e^{-\mu} \tag{3.34}$$

... but it is easier to calculate this another way. The probability is $1 - P_0(\mu)$, which is immediately obtained from

$$P_0(\mu) = e^{-\mu}. \tag{3.35}$$

3.6 Independence and correlation

So far in this chapter we have discussed a single random variable, x, and functions of it, such as x^2, and we have considered the case where different observations of x (or 'trials') are independent. It often happens that we have more than one random variable under consideration. For example, there may be random velocities, and also random locations. Or there may be random energies, and also random directions. In each case, the physical quantities in play may be either independent or not. Velocity can be independent of position, but in the case of a gas with a temperature gradient, the particles will be moving faster on average in the regions at higher temperature. Energy also typically depends on speed through the kinetic energy, but internal energy of a molecule is often independent of the speed at which the centre of mass of the molecule may be moving.

First let's consider two quantities x and y which are independently distributed, with probability density functions f_x and f_y respectively. In this case the probability that the first quantity is the range x to $x + \mathrm{d}x$ *and* the second quantity is in the range y to $y + dy$ is

$$f_x(x)\mathrm{d}x f_y(y)\mathrm{d}y.$$

The mean value of some function $z(x, y)$ is then given by

$$\langle z\rangle = \iint z(x,y) f_x(x) f_y(y)\,\mathrm{d}x\mathrm{d}y.$$

For example, if we want the average value simply of x, irrespective of y, then this expression (using $z(x, y) = x$) gives

$$\begin{aligned}\langle x\rangle &= \iint x f_x(x) f_y(y)\mathrm{d}x\mathrm{d}y = \int x f_x(x)\mathrm{d}x \int f_y(y)\mathrm{d}y \quad (3.36)\\ &= \int x f_x(x)\mathrm{d}x \quad (3.37)\end{aligned}$$

where we used that $\int f_y(y)dy = 1$.

The mean value of the product xy is obtained by substituting $z(x, y) = xy$ and we find

$$\begin{aligned}\langle xy\rangle &= \iint xy f_x(x) f_y(y)\mathrm{d}x\mathrm{d}y = \int x f_x(x)\mathrm{d}x \int y f_y(y)\mathrm{d}y\\ &= \langle x\rangle\langle y\rangle. \quad (3.38)\end{aligned}$$

Thus for quantities whose probability distributions are independent of one another the mean value of the product is equal to the product of the mean values.

More generally, when the quantities are not independent, this will not necessarily happen.

Consider for example, the case where each of x and y can take one of two values, say 2 or 4 for x and 3 or 6 for y, but the combination $(4, 6)$

never happens. The table of possible values is

x	y
2	3
2	6
4	3

In this situation the values of x and y are not independent of one another. If y is equal to 6 then there is only one possibility for x: it must be 2, whereas if y is equal to 3 then x is not always equal to 2 but has another distribution (namely, values 2 and 4 with their probabilities). A situation like this cannot be described by assigning any distribution to x irrespective of y. We have to consider a distribution over both variables together. To illustrate what can happen, let's consider the case where the three outcomes listed in the table are equally likely. Then each row has probability $1/3$. In a large number N of trials we shall find that $x = 2$ in $(2/3)N$ cases, and $x = 4$ in $(1/3)N$ cases, so the mean value of x is $2 \times (2/3) + 4 \times (1/3) = 8/3$. By a similar calculation, the mean value of y is 4. But the mean value of xy is $6 \times (1/3) + 12 \times (2/3) = 10$, which is not equal to $4 \times 8/3$.

The above example considered a discrete case. For a continuous case we describe the overall distribution using a probability density function $f(x, y)$. Now the mean value of some $z(x, y)$ is

$$\langle z(x,y)\rangle = \iint z(x,y) f(x,y) \mathrm{d}x\mathrm{d}y \tag{3.39}$$

and since the distribution function $f(x, y)$ does not factorize, we cannot usually factorize this double integral into two single integrals. In this case it will usually happen that

$$\langle xy\rangle \neq \langle x\rangle\langle y\rangle. \tag{3.40}$$

If one has a set of data concerning two physical quantities both having a range of values, and the mean of the product is not equal to the product of the means, then one can deduce that the two quantities are not independent. If one were to plot a graph of y verses x for such quantities, the graph would show some non-zero correlation between them. If, conversely, one has $\langle xy\rangle = \langle x\rangle\langle y\rangle$ then this is a suggestion but not a proof that the variables are independent.

Correlation can be quantified by using the Pearson **correlation coefficient** r. If x and y are two random variables then the correlation coefficient is defined to be

$$r \equiv \frac{\langle (x-\bar{x})(y-\bar{y})\rangle}{\sigma_x \sigma_y}. \tag{3.41}$$

The quantity in the numerator here is called the **covariance**:

$$\mathrm{Cov}(x,y) \equiv \langle (x-\bar{x})(y-\bar{y})\rangle = \langle xy\rangle - \langle x\rangle\langle y\rangle. \tag{3.42}$$

In the case where both variables have zero mean, this becomes

$$\mathrm{Cov}(x, y) = \iint xyf(x, y)\,\mathrm{d}x\mathrm{d}y \tag{3.43}$$

and then

$$r = \frac{\langle xy\rangle}{\sqrt{\langle x^2\rangle\langle y^2\rangle}}. \tag{3.44}$$

r is always in the range $-1 \leq r \leq 1$. A value 1 shows the variables are perfectly correlated, -1 show they are anti-correlated, and 0 shows they are uncorrelated (and may or may not be independent).

3.7 Relating one distribution to another

It often happens that we have a probability density as a function of one quantity such as speed v, and we would like to know the distribution as a function of a related quantity such as kinetic energy $\epsilon = \frac{1}{2}mv^2$. Since these two physical quantities are directly related to one another, so are the distribution functions.

Let us define probability density functions f and g such that

$$\begin{aligned} \text{Prob(speed is in the range } v \text{ to } v + \mathrm{d}v) &= f(v)\mathrm{d}v \\ \text{Prob(energy is in the range } \epsilon \text{ to } \epsilon + \mathrm{d}\epsilon) &= g(\epsilon)\mathrm{d}\epsilon \end{aligned}$$

It follows that, for any v_1, v_2 and ϵ_1, ϵ_2,

$$\begin{aligned} \text{Prob(speed is in the range } v_1 \text{ to } v_2) &= \int_{v_1}^{v_2} f(v)\mathrm{d}v \\ \text{Prob(energy is in the range } \epsilon_1 \text{ to } \epsilon_2) &= \int_{\epsilon_1}^{\epsilon_2} g(\epsilon)\mathrm{d}\epsilon \end{aligned}$$

Now consider the situation where $\epsilon_1 = \frac{1}{2}mv_1^2$ and $\epsilon_2 = \frac{1}{2}mv_2^2$. In this situation the two cases just described (speed in the given range; energy in the given range) are one and the same, so they must have the same probability:

$$\int_{v_1}^{v_2} f(v)\mathrm{d}v = \int_{\epsilon_1}^{\epsilon_2} g(\epsilon)\mathrm{d}\epsilon = \int_{v_1}^{v_2} g(\epsilon)\frac{\mathrm{d}\epsilon}{\mathrm{d}v}\mathrm{d}v. \tag{3.45}$$

Since these integrals agree for any v_1, v_2, the integrands must agree:

$$f(v) = g(\epsilon)\frac{\mathrm{d}\epsilon}{\mathrm{d}v}. \tag{3.46}$$

A memorable way to write this is:

Relation between probability densities

$$f(v)\mathrm{d}v = g(\epsilon)\mathrm{d}\epsilon \tag{3.47}$$

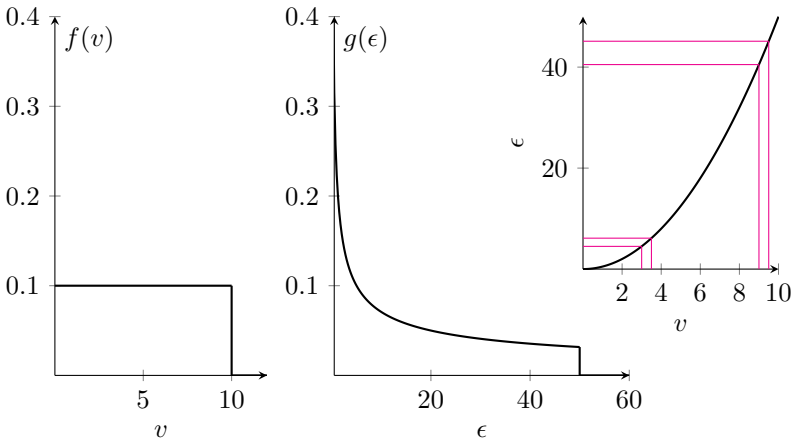


Fig. 3.4 Example of related probability distributions. If the speed distribution is uniform (left graph) then the energy distribution is not (right main graph) owing to the quadratic relationship between energy and speed (inset graph). The lines added to the inset graph show how two example ranges of speed map to the corresponding ranges of energy. Overall, since half the speeds are less than 5 then half the energies are less than 12.5.

In the example of kinetic energy and speed, we have $\mathrm{d}\epsilon/\mathrm{d}v = mv$. Suppose for example that the speed distribution is uniform over some range 0 to $v_{\max}$. Then

$$f(v) = \frac{1}{v_{\max}}, \qquad g(\epsilon) = f(v)\frac{\mathrm{d}v}{\mathrm{d}\epsilon} = \frac{1}{\sqrt{2m}\,v_{\max}}\epsilon^{-1/2} \tag{3.48}$$

These functions are plotted in Fig. 3.4 for the values $v_{\max} = 10$, $m = 1$. The energy distribution in this example is strongly peaked at low energy. This is because the relation $\epsilon \propto v^2$ means that a given range δv corresponds to a smaller range of energies when v is small, so if the probability is uniformly spread in v then it will give more low than high energy values. *The reader should ponder this point by contemplating figure 3.4 at length.* See exercise 9 for another example.

So far we considered a case where the second quantity increased when the first quantity increased ($\mathrm{d}\epsilon/\mathrm{d}v > 0$). More generally one should replace (3.46) with

$$f(x) = g(y)\left|\frac{\mathrm{d}y}{\mathrm{d}x}\right|. \tag{3.49}$$

This relationship applies whenever the quantities x and y are related to one another in a one-to-one fashion, so that $y = y(x)$. The absolute value is required owing to the way probability density functions are defined, which affects the placement of the limits on the integrals in (3.45).

Similar ideas occur in the case of two or more independent random variables. For example, to describe a velocity distribution in a plane we may use the pair of random variables (v_x, v_y) (the components of the velocity vector) or the pair of random variables $(v,\ \theta)$ (the speed and

direction). We may then wish to relate a probability density $f(v_x, v_y)$ to a probability density $g(v, \theta)$. By following the above argument, but now in two dimensions, one finds that in order for the probability integrals to agree one must have

$$f(v_x, v_y) = g(v, \theta) \left| \frac{\partial(v, \theta)}{\partial(x, y)} \right| \tag{3.50}$$

where the quantity multiplying g on the right is the absolute value of the *Jacobian* of the transformation between (v_x, v_y) and (v, θ). The Jacobian is given by

$$\frac{\partial(v, \theta)}{\partial(x, y)} \equiv \det \begin{pmatrix} \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \\ \frac{\partial \theta}{\partial x} & \frac{\partial \theta}{\partial y} \end{pmatrix} \tag{3.51}$$

where 'det' signifies the determinant of the matrix. The general concept of the Jacobian will not be needed in this book. In chapters to follow, we will often move between rectangular and polar coordinates by invoking volume elements; these will be examples of Jacobians but the reader does not need to know this.

3.8 Afterword: On coincidence involving large numbers

This section is not needed for the main development of ideas in the book. It is included in order to point out a widely misunderstood aspect of probability involving coincidence of many events.

Physical processes such as radioactive decay, collisions of atoms in a gas, and many types of thermal motion, are often modelled as independent, in the sense of uncorrelated, from one event to another. This is an accurate model in many situations because, for example, the decay of one nucleus only has a very small effect on the decay of others, and the collision of one pair of atoms in a gas only has a very small effect on the rest of the gas, and so on.

There are two cases where such a model is not accurate. First, the case where the events are not even approximately independent, and secondly the case where the events are independent to good approximation but not completely independent, and we take an interest in coincidences involving a large number of events.

The first case occurs, for example, in the radioactive chain reaction that goes on in inside fissile material above the critical mass, and in the sequence of collisions in a gas which is undergoing a phase transition to liquid. In these examples the model of independent random events is completely inaccurate.

The second case—large coincidences of nearly independent events—is interesting in part because it is widely misunderstood by science students and teachers. In this situation one should not adopt the independent-event model and suppose it to be approximately accurate: it is simply

wrong. For example, suppose there is a type of radioactive nucleus with a half-life of one minute, and we observe a set of n such nuclei for one minute. We would expect around $n/2$ nuclei to decay. The probability for k of them to decay is given, to good approximation, by the Poisson distribution for values of k in the region of $n/2$. But suppose we enquire into the probability for *all* of the nuclei to decay during the one minute observation period. It is tempting to assert that this probability is $(1/2)^n$, since that is the answer if the decay events were truly independent of one another. For large n this answer is wrong, however, and by a very large factor! The reason is that no two processes in nature are fully independent if there is some means of communication between them. In the case of radioactivity, for example, the emission of an electron or a photon or another particle from a nucleus will have a small but non-zero effect on the evolution of other nearby nuclei, and this effect, small as it is, will be much larger than $(1/2)^n$ for a sufficiently large value of n. The lesson is that *correlated effects dominate uncorrelated ones in the tails of the probability distribution.*

We will discuss two simple examples of correlated influences, in order to illustrate the possibilities.

Suppose we have a set of entities such as atoms and, at the outset, a given type of event such as a collision or a decay can occur with probability p for each atom in some given interval of time. Suppose further that if the event happens then the probability for a subsequent event nearby of a similar type is $(1+\epsilon)p$ per atom where ϵ is small. If the events were all independent with probability p then the probability to observe n events in n trials would be p^n. But in the presence of the small correlation the probability is approximately

$$p \times (1+\epsilon)p \times (1+\epsilon)^2 p \times \cdots \simeq p^n (1+\epsilon)^{n^2/2} \simeq p^n \exp(n^2\epsilon/2)$$

where the approximations are valid in the limit $n \gg 1$ and $n\epsilon < \ln(1/p)$. This estimate is not intended to be precise, since there are many details which will depend on the particular physical effect in question. The purpose is merely to point out that the estimate p^n for the probability is incorrect by a factor of order $\exp(n^2\epsilon/2)$ and this could be a large factor even when ϵ itself is tiny. For example if $n = 10^6$ then values of ϵ of order 10^{-12} are significant.

The above method of calculation does not treat the case where probabilities become close to 1. As an illustration of that case, consider a situation where events are almost independent, but the influence of one event on another is such that 100 events occurring in a burst would be sufficient to cause further events to take place with near-certainty. In this case the probability of n events is approximately p^{100} for any $n > 100$. For example, if $p = 0.99$ and $n = 10^6$ then p^n is of order 10^{-4365}, and $p^{100} \simeq 0.4$. The first number is so low that such an outcome would be deemed impossible if the events were independent, but the second number shows that a small (about one part in 100) amount of correlation is sufficient to make what seemed impossible quite likely to occur.

It is not important to learn the details of these example calculations. The main lesson is that even a very small effect, one which could be ignored to good approximation for small numbers of events, can entirely change the outcome when it concerns a correlated influence and a large number of events.

©A.M. Steane 2026, not for distribution.

Exercises

(1) Find the mean and variance of the binomial distribution, confirming (3.15) and (3.16).

(2) Confirm that the Poisson distribution (3.30) is normalized, and find the mean and standard deviation.

(3) A ruler is used to measure a length. The measured value is slightly over 1 metre, with an uncertainty owing to the difficulty of positioning the ruler and reading the measured value. It is decided that the true value is equally likely to lie anywhere in the range 1.001 m to 1.002 m. Find the standard deviation of this length measurement.

(4) The random variable ϕ is uniformly distributed between 0 and 2π. Find: $\langle\phi\rangle, \langle\phi^2\rangle, \langle\cos(\phi)\rangle, \langle\cos^2\phi\rangle$ and the variance and standard deviation of ϕ.

(5) The masses m of stars in a distant galaxy are distributed randomly over a continuous range, such that the distribution is described by a probability density function $f(m)$. What S.I. units would be suitable for reporting values of $f(m)$?

(6) **Obtaining Poisson distribution from binomial distribution**. Consider a process with two outcomes which we shall call 'click' or 'no-click', and suppose the probability of 'click' is p. The mean number of clicks in n trials is $\mu = np$ and the distribution is the binomial distribution, eqn (3.13). We can apply this to the case of a process such as radioactive decay where p signifies the probability of decay in some fixed small time interval δt, given that the decay has not happened yet. Or in another example it could be the probability of an event occurring in some region of space of given size. In an interval t there are $n = t/\delta t$ 'trials'. Writing $p = \mu/n$ we have that the binomial distribution function is

$$P_k(\mu/n, n) = \frac{n!}{k!(n-k)!}\left(\frac{\mu}{n}\right)^k\left(1-\frac{\mu}{n}\right)^{n-k}.$$

(i) Prove the following:

$$\begin{aligned}\lim_{n\to\infty}\frac{n!}{(n-k)!n^k} &= 1,\\ \lim_{n\to\infty}(1-\mu/n)^n &= e^{-\mu},\\ \lim_{n\to\infty}(1-\mu/n)^{-k} &= 1.\end{aligned}$$

(ii) Hence show that

$$\lim_{n\to\infty} P_k(\mu/n, n) = \frac{\mu^k}{k!}e^{-\mu}. \tag{3.52}$$

This shows that if we have an infinite number of independent trials in some given interval of time or region of space, with a fixed mean number of clicks, then the probability of obtaining k clicks is as given by the Poisson distribution function.

(7) The speeds of particles in a gas are described by the speed distribution function

$$f(v) = Nv^2e^{-v^2} \tag{3.53}$$

Find the kinetic energy distribution function, $g(\epsilon)$.

(8) The spectrum of light from the Sun can be considered either as a function of frequency or as a function of wavelength. Considered as function of wavelength, it peaks in the visible region, at around 500 nm. Considered as a function of frequency, it peaks in the infrared region, at around 880 nm. How can it be that the location of the maximum thus moves depending on which measure is employed?

(9) A particle is undergoing simple harmonic motion $x = a\sin\omega t$. Observations are made at random times uniformly distributed over a period of the motion. Find the probability density function describing the probability that the particle will be observed between x and $x + \mathrm{d}x$ (taking care with the normalization). Plot the function on a graph and discuss any interesting features. What is the probability that the particle will be observed to lie between $x = 0$ and $x = 0.01\,a$? What is the probability that the particle will be observed to lie between $x = 0.99x_0$ and $x = a$? [*Ans.* 0.00318, 0.0451]

The pressure in a gas

4

We now apply the statements about modelling from chapter 2 and the treatment of probability in chapter 3 to learn about pressure in a gas. In this chapter we will take as our model that a gas is composed of many small entities, which we shall call either molecules or particles. Each molecule has a fixed mass but we will not be concerned with the spatial size or shape of the molecules. Our aim is to find out how to relate the pressure in a gas to the average motion of the molecules.

4.1 Pressure in very simple models

A single particle moving along a line

As a warm-up exercise, consider a single particle of mass m moving in one dimension between two walls separated by a distance L, and undergoing elastic collisions with the walls (Fig. 4.1). If the particle's velocity is v then each collision with the wall changes the particle's momentum from mv to $-mv$. If the collision with the wall takes a time τ then the force of this impact is $2mv/\tau$. Strictly speaking, this is the time-averaged value of the force after averaging over the time τ. We would like to calculate the average force over some longer time. Between impacts the force on the wall is zero. The next impact with a given wall happens after the particle has travelled twice between the walls, covering a distance $2L$. Therefore for each time $2L/v$ there is a momentum $2mv$ delivered to any given wall. After N such collisions a total time $\Delta t = 2NL/v$ has passed and a total momentum $\Delta p = 2Nmv$ has been delivered, so the time-averaged force on the wall is

$$\langle F \rangle = \langle \frac{\Delta p}{\Delta t} \rangle = \frac{2Nmv}{2L/v} = \frac{N}{L} mv^2. \tag{4.1}$$

A group of particles moving along a line

Now let's consider a situation like the one just discussed, only now there is a large number of particles, not all of the same speed. The speeds

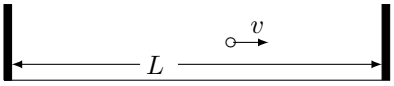


Fig. 4.1 A single particle moving in one dimension between walls separated by distance L.

are distributed randomly with some given distribution $f(v)$ (note, we shall be using $f(v)$ for the probability density, and F for the force). To calculate the average force, we shall average over both time and speed. In order to do this, consider first those particles whose speed lies in the range $(v,\ v+\mathrm{d}v)$. The momentum change, at a collision with the wall, of each such particle is $2mv$, and the number of such particles is $Nf(v)\mathrm{d}v$, if N is the total number of particles. Hence the momentum exchange during collisions with one wall of particles in this group, if each particle has one such collision, is

$$\mathrm{d}p = 2mvNf(v)\mathrm{d}v.$$

This is an infinitesimal because the number of particles in the group is an infinitesimal.

If we ignore collisions among the particles (that is, we adopt a simple model in which there are no such collisions) then each particle takes a time $t = 2L/v$ to travel twice across the box. Therefore the rate of exchange of momentum with the wall, owing to impacts on the wall of particles in this group, is

$$\frac{\mathrm{d}p}{t} = \frac{mv^2Nf(v)\mathrm{d}v}{L}.$$

Hence the total rate of exchange of momentum with the wall, owing to the impacts of particles of all speeds, is

$$\int \frac{1}{t}\mathrm{d}p = \frac{N}{L}m\int v^2 f(v)\mathrm{d}v = \frac{N}{L}m\langle v^2\rangle \tag{4.2}$$

This is the time-averaged force on the wall. Note the similarity with eq. (4.1). Note especially that on the right we have $\langle v^2\rangle$ not $\langle v\rangle^2$.

Three-dimensional box, mirror-like reflections

We now extend our model to three dimensions. We suppose a rectangular box of diameters L_x, L_y, L_z and we suppose particles move in all directions inside the box. For the moment we shall adopt a simple model of the collisions of particles with the walls, in which each collision is a mirror-like reflection. That is, during such a collision the component of the particle's momentum parallel to the wall is unchanged, and the component perpendicular to the wall changes sign (Fig. 4.2). (We will see in the next section that this mirror-like reflection idea is almost never true for real gases in ordinary containers, but it is a useful model to consider nonetheless.) In this model, for a wall whose normal is along the x direction, the change of momentum of the particle, during one collision with the wall, is

$$2mv_x.$$

In this situation, and maintaining our simple model in which the particles do not collide with one another, the motions in the three dimensions x, y, z are entirely oblivious to one another. We say they are *uncoupled*.

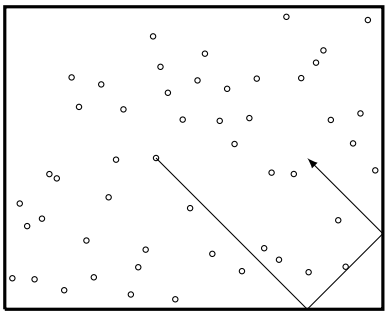


Fig. 4.2 Particles in a box, with mirror-like elastic collisions with the walls (unrealistic!).

The motion along x is not affected when the particles hit the walls parallel to x. So we can find the total force on a wall whose surface is perpendicular to the x direction by adopting the calculation of the previous section. We shall find

$$\langle F_x \rangle = \frac{N}{L_x} m \langle v_x^2 \rangle. \tag{4.3}$$

The area of the wall under consideration is $L_y L_z$. The pressure on the wall is the force per unit area. Hence we obtain, for the time-average of the pressure on a wall normal to x:

$$P_x = \frac{N}{L_x L_y L_z} m \langle v_x^2 \rangle = nm \langle v_x^2 \rangle \tag{4.4}$$

where

$$n \equiv \frac{N}{V} \tag{4.5}$$

is the number of particles per unit volume, called the **number density**.

Similar calculations for the pressures on the other walls give

$$P_y = nm \langle v_y^2 \rangle, \qquad P_z = nm \langle v_z^2 \rangle. \tag{4.6}$$

Notice that we have not had to assume any particular distribution of velocities in order to arrive at these predictions, except that we assumed the distribution is constant over time. We do not need to assume conditions of thermal equilibrium, for example. But if the conditions are isotropic (whether in equilibrium or otherwise) then the distributions of the three components of velocity are the same, and in that case

$$\langle v_x^2 \rangle = \langle v_y^2 \rangle = \langle v_z^2 \rangle. \tag{4.7}$$

Furthermore, the speed v of a particle is related to the components by

$$v = (v_x^2 + v_y^2 + v_z^2)^{1/2} \tag{4.8}$$

and therefore

$$\langle v^2 \rangle = \langle v_x^2 + v_y^2 + v_z^2 \rangle = \langle v_x^2 \rangle + \langle v_y^2 \rangle + \langle v_z^2 \rangle. \tag{4.9}$$

By combining this with eq. eq. (4.7) we have

$$\langle v_x^2 \rangle = \langle v_y^2 \rangle = \langle v_z^2 \rangle = \frac{1}{3}\langle v^2 \rangle \tag{4.10}$$

and therefore our equations for the pressure in each direction all give

Pressure in isotropic conditions

$$P = \frac{1}{3} nm\langle v^2 \rangle. \tag{4.11}$$

We have put this equation in a box because it is an important result, but for the moment the reader should regard it with a pinch of salt because it has only been obtained with an extremely simplistic model of the gas. Section 4.2 will show that it still applies for more realistic models.

4.1.1 Relation between pressure and energy density

Noting that the translational kinetic energy of a particle of mass m and speed v is $K = (1/2)mv^2$, we can write the right hand side of eq. eq. (4.11) as $(2/3)n\langle K \rangle$. Since n is the number density of particles, and $\langle K \rangle$ the mean kinetic energy of any one particle, it follows that $n\langle K \rangle$ is the translational kinetic energy per unit volume, or **energy density**. Let us introduce the symbol u to stand for total energy density:

$$u = \frac{U}{V} \tag{4.12}$$

where V is the volume and U is the total internal energy of the gas, which includes both the translational kinetic energy of the molecules, and any further energy each molecule may have owing to its rotation and internal vibrations. Then we have

$$\frac{1}{2} m\langle v^2 \rangle n = u - u_{\text{other}} \tag{4.13}$$

where u_{other} is the energy-density of other forms of energy (i.e. other than translational kinetic energy), such as that owing to rotation and vibration of the molecules. Hence our equation for the pressure can be written

$$P = \frac{2}{3}\left(u - u_{\text{other}}\right). \tag{4.14}$$

For a monatomic gas (a gas composed of atoms not molecules, e.g. a noble gas), $u_{\text{other}} = 0$ and then the equation reads $P = (2/3)u$. More generally, in conditions of thermal equilibrium and not too low temperature, it often happens that $u_{\text{other}} \propto u$, where the constant of proportionality is independent of temperature over a wide range of temperature. Therefore the relationship

$$\text{(Pressure)} \propto \text{(energy density)} \tag{4.15}$$

applies to a wide range of cases in thermal equilibrium.

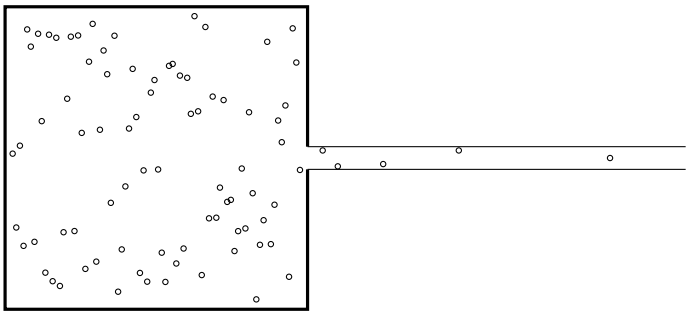


Fig. 4.3 Particles in a gas escaping down a narrow tube.

4.2 Pressure in a more realistic model

Let us revisit now our picture of gas particle collisions with the walls of a chamber. The experimental evidence is that the mirror-like reflection model is wrong. This can be seen by considering a flow of gas down a long narrow tube at low pressure (Fig. 4.3). If the gas particles simply bounced off the walls of such a tube, then each particle would move rapidly down the tube, maintaining its velocity component $v_{\|}$, so the time taken to complete a journey down a tube of length L would be $t = L/v_{\|}$. Typical numbers could be $L = 1$ cm, $v_{\|} = 100\,\mathrm{m/s}$, giving $t = 100$ microseconds. However in practice the observed time can be very much longer, depending on the diameter of the tube. It can be minutes or hours or days!

What is going on?

It turns out that gas particles typically do not reflect from ordinary surfaces. Rather, when they arrive at the surface they stick. In steady-state conditions they can also be released from the surface. These processes are called *adsorption* and *desorption*. The average time a given adsorbed molecule stays on the surface is called the *dwell time*. This dwell time can itself be quite short, so the dwell time is not the main reason why gases only move slowly down narrow tubes. The main reason is that each molecule tends to re-thermalize while it is adsorbed onto the surface, 'forgetting' whatever velocity it had before, and then it is re-emitted isotropically (i.e. with equal likelihood in all directions) into the hemisphere outside the surface. After just a single collision with the wall of a tube, therefore, each gas particle has not a large velocity down the tube, but, on average, zero velocity down the tube! The way a gas moves down a narrow tube at low pressure is by a process called *diffusion* which will be considered in chapter 12. For the present discussion we do not need to consider that process. All we need to do is think more cleverly about the way pressure comes about.

The concept we need, and it is a central concept in kinetic theory, is that of **flux**. The term 'flux' refers to a flow of some kind, and it means the amount of some quantity crossing unit area per unit time. The area envisaged is that of a surface perpendicular to the flow direction. For

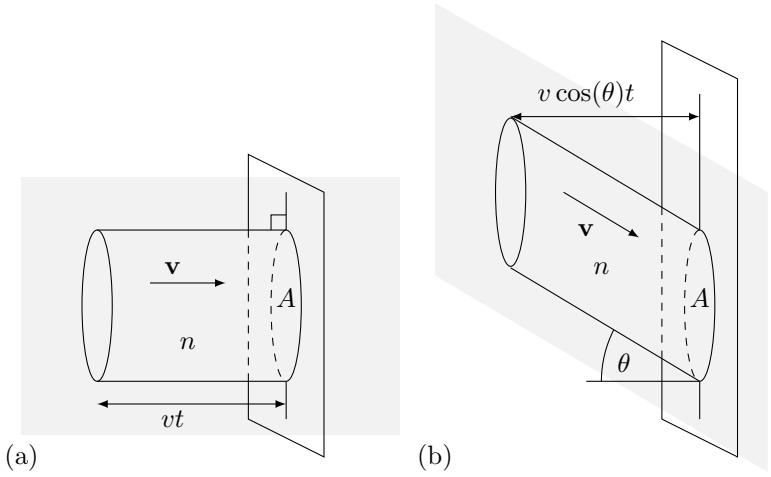


Fig. 4.4 A uni-directional flow crossing a plane. (a) A plane orthogonal to the flow direction; (b) a plane at a general angle to the flow direction.

example, a flux of energy is an amount of energy per unit area per unit time (a power per unit area). A flux of electric charge is an amount of charge per unit area per unit time, also called current density. A particle flux is a number of particles per unit area per unit time.

Consider a beam of particles all moving along together with velocity $\mathbf{v}$. Let n be the number density of particles in the beam. Consider an area A of a plane placed perpendicular to the beam (Fig. 4.4(a)). In a time t, all the particles in a region of length vt and cross-section A will pass through that part of the plane. The number of particles in this region is

$$(\text{volume}) \times (\text{number density}) = Avt\, n.$$

Therefore the number passing through the plane, per unit area and per unit time, is

$$nv.$$

This is the size of the particle flux. The direction of the flux is that of the velocity $\mathbf{v}$, so we have

Flux of particles

$$\mathbf{j} = n\mathbf{v}. \tag{4.16}$$

Notice that the flux is a vector quantity. The direction is the direction of the flow, and the magnitude is the number of particles passing through a plane perpendicular to the flow, per unit area per unit time.

Next let's consider a plane which whose normal makes some angle θ with the flow velocity (Fig. 4.4(b)). How many particles cross an area A of this plane in time t? The answer is: all those that lie in the region shown in the figure. The volume of this region is $Avt\cos\theta$, therefore the

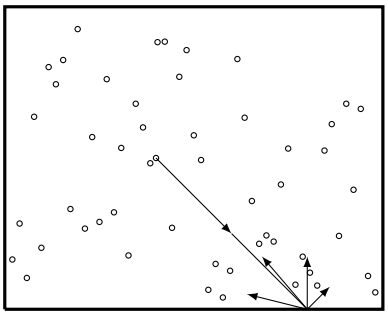


Fig. 4.5 At last! A reasonably realistic model. The arrows indicate the idea of a probability distribution of directions and velocities after a collision of any given molecule with the wall. We merely need to assume this distribution is the same as the one elsewhere in the gas.

number crossing, per unit area per unit time, is

$$nv\cos\theta = nv_\perp = j_\perp \tag{4.17}$$

where, on the right hand side, $v_\perp$ is the component of velocity perpendicular to the plane under consideration, and similarly for $j_\perp$. Hence, to obtain the net flux through a plane of any orientation, *use the component of the flux vector in the direction normal to the plane.*

Calculating pressure from momentum flux

We shall now apply the concept of flux to the calculation of pressure in a gas. The idea is to consider the amount of momentum brought up to any given wall by the gas molecules, per unit area per unit time. In other words, we consider the flux of momentum. The assumptions about the collisions with the wall will now be realistic (Fig. 4.5).

For the sake of argument let's consider a wall perpendicular to the x direction. We consider those gas particles whose velocities are in the range (v_x, v_y, v_z) to $(v_x + \mathrm{d}v_x,\ v_y + \mathrm{d}v_y,\ v_z + \mathrm{d}v_z)$. The number density of such particles is

$$nF(v_x, v_y, v_z)\mathrm{d}v_x\mathrm{d}v_y\mathrm{d}v_x$$

where F is the probability distribution function over all three velocity components, and n is the total number density of particles in the gas. Let us now introduce the useful shorthand

$$\mathrm{d}^3\mathbf{v} \equiv \mathrm{d}v_x\mathrm{d}v_y\mathrm{d}v_z. \tag{4.18}$$

Then the previous expression can be written

$$nF\,\mathrm{d}^3\mathbf{v} \tag{4.19}$$

where it is taken as understood that F is a function of the three velocity components. Using eq. (4.17), the flux of particles in the velocity range under consideration, through a plane perpendicular to x, is

$$nF\,\mathrm{d}^3\mathbf{v}\,v_x. \tag{4.20}$$

This is the number of particles of the given velocity which hit the wall, per unit area per unit time. These particles each bring with them their momentum $m\mathbf{v}$. If we imagine for a moment that they all stick to the wall, then the wall will acquire this momentum. But we do not need to assume the particles stick to the wall, nor that they are reflected in any simple way. All we need to assume is that in steady state conditions the net effect of the interactions with the wall is that there is also this same flux of particles in the direction away from the wall. Therefore if we just calculate the rate per unit area at which momentum is brought up to the wall by the particles as they arrive, then we can claim that the overall pressure will be twice this, because the rate at which momentum is given back by the wall to the particles is the same. In this approach we need make no assumption about the process on the wall except that it has settled to a steady state where the flux away from the wall is equal to the flux towards the wall.

The momentum *brought up* to the wall, per unit area per unit time, by the particles under consideration is

$$dP_{x,\,\text{incident}} = (nF\mathrm{d}^3\mathbf{v})\, v_x \cdot mv_x.$$

(Since momentum change per unit time is force, and force per unit area is pressure, this quantity is a pressure.) Therefore the total momentum perpendicular to the wall brought by the incident particles, per unit area per unit time, is

$$P_{x\,\text{incident}} = \int_{v_x=0}^{\infty}\int_{v_y=-\infty}^{\infty}\int_{v_z=-\infty}^{\infty} mnF(v_x,v_y,v_z)v_x^2\,\mathrm{d}v_x\mathrm{d}v_y\mathrm{d}v_x. \quad (4.21)$$

Finally, since this same amount of momentum is also given back to the particles as they leave the wall, the total pressure on the wall is

$$\begin{aligned} P_x &= 2P_{x,\text{incident}} \\ &= \int_{v_x=0}^{\infty}\int_{v_y=-\infty}^{\infty}\int_{v_z=-\infty}^{\infty} 2mnF(v_x,v_y,v_z)v_x^2\,\mathrm{d}v_x\mathrm{d}v_y\mathrm{d}v_x. \end{aligned} \quad (4.22)$$

Note the limits on the three integrals here. To avoid double-counting we must include in the integral over v_x only the velocities towards the wall. The argument concerns the particles which approach the wall, and calculates what will happen to their momentum. The other two integrals cover the entire range.

To simplify the integral we now make the following assumption about the probability distribution:

$$F(v_x,v_y,v_z) = F(-v_x,v_y,v_z).$$

This asserts that a given x-component of velocity is equally likely as the one of opposite sign. In other words, particles in the entire gas are equally likely to be moving one way as the opposite way. Notice that the function F gives the distribution of velocities in the entire gas, so it will have this symmetry if the gas is not moving anywhere overall, as

long as there are no internal flows such as vortices and vibrations. We don't necessarily require the distribution to be isotropic here, just that it is an even function of v_x. With this assumption we can convert the integral in eq. (4.22) to one over the entire range of velocities[1]:

$$\begin{aligned} P_x &= mn \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(v_x, v_y, v_z) v_x^2 \, \mathrm{d}v_x \mathrm{d}v_y \mathrm{d}v_x \\ &= mn\langle v_x^2 \rangle \end{aligned} \tag{4.23}$$

where the last step used the definition of the mean.

We have now reproduced eq. (4.4), but by a much more general and realistic argument. The same argument applied to a wall with unit normal vector $\hat{\mathbf{n}}$ gives

$$P_{\hat{\mathbf{n}}} = mn\langle (\mathbf{v}\cdot\hat{\mathbf{n}})^2 \rangle. \tag{4.24}$$

In the case of an isotropic distribution of velocities, so that we can use eq. (4.7), we shall obtain (4.11) and (4.15).

Example 4.1 Find the r.m.s. speed of molecules in a gas of density $1\,\mathrm{kg/m^2}$ at pressure $10^5\,\mathrm{Pa}$.
Solution.

$$\sqrt{\langle v^2 \rangle} = \sqrt{\frac{3P}{\rho}} = 547.7\,\mathrm{m/s}.$$

Example 4.2 A gas of magnetic dipoles is situated in a uniform magnetic field $\mathbf{B}$ and is in thermal equilibrium. (a) Express the pressure on a wall perpendicular to $\mathbf{B}$ in terms of a suitable average of particle properties. (b) Do the same for a wall whose normal $\hat{\mathbf{n}}$ is at an angle θ to $\mathbf{B}$.
Solution.
(a) Let the z axis be aligned with $\mathbf{B}$. For the pressure on a wall whose normal is along z, we can use the method of calculation presented in this section. The distribution of velocity components v_x and v_y is irrelevant. One finds:

$$P_z = mn\langle v_z^2 \rangle$$

(b) Although the velocity distribution here will not be isotropic, the influence of the magnetic field will not give a net translational motion to the gas in thermal equilibrium. Therefore we expect the velocity distribution to be an even function of the three velocity components and therefore the reasoning above applies again, and one must find

$$P = mn\langle (\mathbf{v}\cdot\hat{\mathbf{n}})^2 \rangle$$

[1]This is an example of $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 2\int_0^{\infty} f(x)\mathrm{d}x$ for an even function $f(x)$.

Without loss of generality, we keep the z axis along $\mathbf{B}$ and align the y axis perpendicular to $\hat{\mathbf{n}}$. Then

$$\mathbf{v} \cdot \hat{\mathbf{n}} = v_z \cos\theta + v_x \sin\theta.$$

Hence

$$\begin{aligned} P &= mn\langle (v_z \cos\theta + v_x \sin\theta)^2 \rangle = mn\left(\langle v_z^2 \rangle \cos^2\theta + \langle v_x^2 \rangle \sin^2\theta\right) \\ &= mn\langle v_z^2 \rangle \cos^2\theta + \frac{1}{2} mn\langle v_x^2 + v_y^2 \rangle \sin^2\theta \end{aligned}$$

where we used $\langle v_z v_x \rangle = 0$ and in the final expression we used $\langle v_x^2 \rangle = \langle v_y^2 \rangle$ so as to write the result in a more general form.

4.3 Summary

The central ideas of this chapter are the concept of *flux* and the relation between pressure, density and mean squared velocity for particles in a gas. To find this relation it has not been necessary to know the distribution function F, nor that the gas is necessarily in thermal equilibrium. To obtain eq. (4.23) we assumed that the velocity distribution has equal numbers of particles with v_x of either sign, and we assumed that after averaging over many impacts with the wall, the net effect on the momentum of each group of particles is to reverse the momentum component perpendicular to the wall. These assumptions will hold in many steady state situations. In particular they will hold when the gas is in thermal equilibrium.

We have not yet considered collisions among the gas particles, but even in the presence of such collisions eq. (4.11) will still hold, for a gas in steady state, because the collisions merely swap molecules between velocity classes without changing the distribution function. This is on a hard sphere model where the particles do not attract one another when they are not colliding.

For real gases the particles do attract or repel one another, and for this reason our pressure equation will no longer hold for a high-density gas, close to the conditions where it may turn into a liquid or solid, and the results will not apply to liquids and solids. A solid such as a spring can exert a pressure even when its atoms are not moving at all.

©A.M. Steane 2026, not for distribution.

Exercises

(1) Estimate (i) the flux of water in a river (using mass as a measure of amount of water), (ii) the flux of cars on a busy motorway (here count the cars not their mass), (iii) the flux of ants walking on a forest floor near a busy ant colony, (iv) the flux of molecules in a gas at standard temperature and pressure (that is, 1 atmosphere and 300 K) incident on one side of a plane in the gas.

(2) An ideal gas in thermal equilibrium has the equation of state $P = nk_{\mathrm{B}}T$ where P is the pressure, n is the number of particles per unit volume and T is the temperature. (i) Use this to find the relationship between $\langle v^2 \rangle$ and T for such a gas. (ii) Show that the mean kinetic energy per molecule is $(3/2)k_{\mathrm{B}}T$.

Speed and velocity

5

In this chapter we look at the mathematical ideas which are needed to consider a distribution of speeds and a distribution of velocities. Many of the results apply to any distribution; some only apply to an isotropic distribution (defined in section 5.2).

5.1 Notation for a distribution of velocity

When discussing motions of particles in a gas, sometimes we are interested in their speeds, sometimes their directions, and sometimes both. When considering speed we require just one real number to specify the quantity of interest: v. When considering velocity we have three real numbers. These could, for example, be the components (v_x, v_y, v_z). But we could alternatively use the speed and direction, specified by size and angles in a polar system: (v, θ, ϕ). These different choices lead to the fact that the *same* distribution of velocities can be described by *different* mathematical functions, depending on which variables the functions are applied to. It also leads to a plethora of mathematical notations, and the possibility of confusion.

In this book, the function $f(v)$ always refers to a distribution of *speed*, and it is normalized in such a way that

$$\int_0^\infty f(v)\mathrm{d}v = 1. \tag{5.1}$$

In the previous chapter we introduced a distribution of velocity by using the three components in a rectangular system, giving the function $F(v_x, v_y, v_z)$ which first appeared in equation (4.2). This is normalized in such a way that

$$\int_{-\infty}^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty F(v_x, v_y, v_z)\,\mathrm{d}v_x\mathrm{d}v_y\mathrm{d}v_z = 1. \tag{5.2}$$

It should be obvious that the functions f and F are different functions (after all, one has a single parameter while the other has three), even when they describe the relevant distribution for the same gas. But it can happen that in the function F the variables v_x, v_y and v_z only appear in the combination $v_x^2 + v_y^2 + v_z^2$. So in this case even though the function is really a function of three things, it might appear to be a function of one thing. Here is an example:

$$F(v_x, v_y, v_z) = \pi^{-3/2}\exp(-v_x^2)\exp(-v_y^2)\exp(-v_z^2) \tag{5.3}$$

which gives

$$F = \pi^{-3/2} e^{-v^2} \tag{5.4}$$

where $v = (v_x^2 + v_y^2 + v_z^2)^{1/2}$. The confusion might now arise if someone thought that the distribution shown in (5.4) is the distribution of speed in this example. It is not! To make this clear one could write

$$F(v_x,\, v_y,\, v_z) = \pi^{-3/2} e^{-v^2} \tag{5.5}$$

which is less likely to cause confusion, but sometimes it is convenient to drop the explicit listing of the parameters: so, beware! Another way to keep the variables in view is to write $F\,\mathrm{d}v_x\mathrm{d}v_y\mathrm{d}v_z$ or $F\mathrm{d}^3\mathbf{v}$, c.f. eq. (4.19).

In the example of a velocity distribution given by (5.3), the speed distribution turns out to be (we will prove this in the next section):

$$f(v) = \frac{4}{\sqrt{\pi}} v^2 e^{-v^2}. \tag{5.6}$$

Another function of three variables is the one which describes the distribution in polar coordinates. We shall adopt the notation $f_{3\mathrm{p}}$ for this function, where the subscript 'p' stands for 'polar'. It is normalized in such a way that

$$\int_{v=0}^{\infty}\int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi} f_{3\mathrm{p}}(v,\, \theta,\, \phi)\,\mathrm{d}v\mathrm{d}\theta\mathrm{d}\phi = 1. \tag{5.7}$$

Be careful to note the range on the angles θ and ϕ here. θ ranges from 0 to π; whereas ϕ ranges from 0 to 2π. In this way together they indicate directions covering the entire surface of a sphere exactly once; c.f. Figs 5.1 and 5.2. We will see how this function relates to $f(v)$ and $F(v_x, v_y, v_z)$ in the next section.

5.2 Isotropic velocity distribution

The term *isotropic* means *the same in all directions*. If a density distribution has spherical symmetry, for example, then it is isotropic with respect to the centre of symmetry. If particles in a gas are as likely to be moving in one direction as another, then their velocity distribution is isotropic. We already met one property of isotropic velocity distributions in the previous chapter: $\langle v_x^2\rangle = \langle v_y^2\rangle = \langle v_z^2\rangle = (1/3)\langle v^2\rangle$. Another property is $\langle v_x\rangle = \langle v_y\rangle = \langle v_z\rangle = 0$ which follows from the fact that for any given v_x, the probability densities for v_x and $-v_x$ are equal (and similarly for v_y, v_z).

To describe an isotropic velocity distribution it is convenient to adopt spherical polar coordinates (see appendix A). That is, we treat the *velocity space* by using coordinates (v, θ, ϕ) where v is the speed (i.e. the magnitude of the velocity), and spherical polar angles θ, ϕ specify the direction. The *volume element* in such a space is a region extending

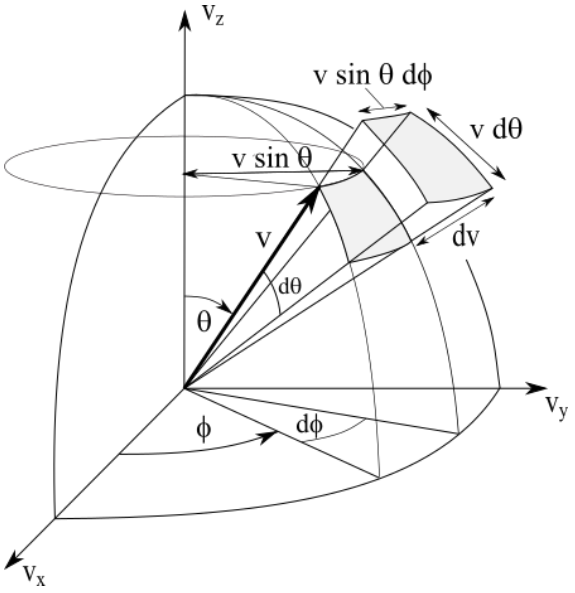


Fig. 5.1 The volume element in spherical polar coordinates, in velocity-space.

from (v, θ, ϕ) to $(v + \mathrm{d}v, \theta + \mathrm{d}\theta, \phi + \mathrm{d}\phi)$ as shown in Fig. 5.1. We are here using the word 'volume' for a region in velocity space, so it is not a volume in the ordinary sense of the word: its physical dimension is speed cubed, not length cubed. However we call it a volume element of the space because it fills up a region of the space, and the entire space is made of many such regions. The 'volume' in velocity-space of such a volume element is obtained by multiplying the lengths of its sides together, since they are orthogonal in the limit of small quantities:

volume element in spherical polar coordinates

$$\mathrm{d}v\,(v\mathrm{d}\theta)\,(v\sin(\theta)\,\mathrm{d}\phi) = v^2\sin(\theta)\,\mathrm{d}v\mathrm{d}\theta\mathrm{d}\phi. \tag{5.8}$$

It is useful to reflect on the role of $\sin(\theta)$ in this formula. Suppose we consider all the velocities at some given v and within a small range of θ, from θ to $\theta + \mathrm{d}\theta$, with any value of ϕ. These velocities lie on a ring in velocity-space as shown in figure Fig. 5.2. Note that this ring has radius $v\sin(\theta)$ not v. This is because it follows a circular loop at a fixed θ, like a line of latitude on a globe. $v\sin\theta$ is the distance of this ring from the axis where its centre lies. The width of the ring-shaped band meanwhile is $v\mathrm{d}\theta$. It follows that the area of the ring-shaped band in velocity space (that is, the area of the band shown shaded in figure 5.2, not the area of the enclosed circle) is

$$(2\pi v\sin\theta)\,(v\mathrm{d}\theta) = 2\pi v^2\sin(\theta)\,\mathrm{d}\theta \tag{5.9}$$

As θ becomes small, the circumference of the ring shrinks, and consequently so does the area of the band. In other words, *a given range* $(\theta, \theta + \mathrm{d}\theta)$ *of the first polar angle corresponds to less and less of velocity-*

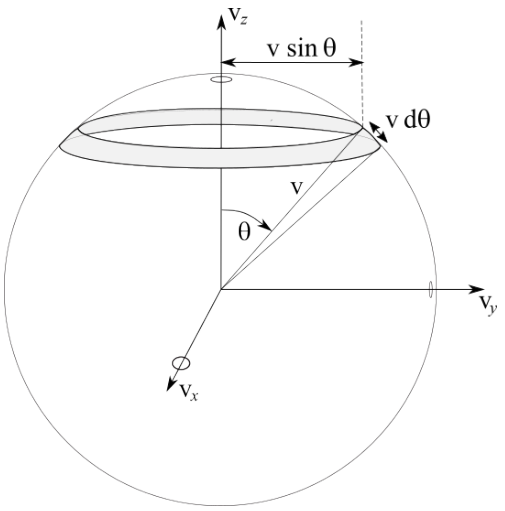


Fig. 5.2 Velocities with polar angle in the range θ to $\theta + \mathrm{d}\theta$ take up an amount of velocity-space proportion to $\sin\theta$. This is a smaller amount, for example, when θ approaches 0 or π. Therefore to describe an isotropic distribution, the probability density function must be proportional to $\sin\theta$.

space as θ *approaches* 0 *or* π. This is what the $\sin(\theta)$ factor in (5.8) is indicating.

We would now like to construct a probability-density function that describes an isotropic velocity distribution. One might think, at first, that this is achieved by a function which does not depend on θ or ϕ. That would be wrong. In view of the way spherical polar coordinates work, and in particular the issue discussed in the previous paragraph, a probability-density function describing an isotropic distribution must be proportional to $\sin\theta$, when θ is the polar angle in spherical polar coordinates, so that the distribution takes the form

$$f_{3\mathrm{p}}(v,\theta,\phi) = \frac{1}{4\pi} f(v)\sin(\theta). \tag{5.10}$$

In this way we will find equal amounts of probability in equal amounts of solid angle on the sphere. The factor 4π here is a normalization factor, as we now show. The normalization over the polar coordinates is as shown in eq. (5.7). It follows that

$$\frac{1}{4\pi}\int_0^\infty f(v)\,\mathrm{d}v \int_0^\pi \sin(\theta)\,\mathrm{d}\theta \int_0^{2\pi} \mathrm{d}\phi = 1.$$

The ϕ integral here evaluates to 2π, and the θ integral evaluates to 2, so we have

$$\int_0^\infty f(v)\,\mathrm{d}v = 1,$$

which is as it should be if $f(v)$ is the speed distribution function.

We now have two functions, $f_{3\mathrm{p}}$ and f, giving two ways to describe the same set of velocities. The function $f_{3\mathrm{p}}$ has the interpretation

$$f_{3\mathrm{p}}(v,\theta,\phi)\,\mathrm{d}v\mathrm{d}\theta\mathrm{d}\phi \;=\; \mathrm{Prob}\left(\begin{array}{l}\text{velocity is in the range}\\ (v,\theta,\phi)\text{ to }(v+\mathrm{d}v,\theta+\mathrm{d}\theta,\phi+\mathrm{d}\phi)\end{array}\right).$$

Using $f(v)$ this same quantity can be written

$$f(v)\mathrm{d}v\frac{\sin(\theta)\,\mathrm{d}\theta\mathrm{d}\phi}{4\pi}. \tag{5.11}$$

The function $f(v)$ has the interpretation

$$f(v)\,\mathrm{d}v = \mathrm{Prob}\,(\text{speed is in the range } v \text{ to } v+\mathrm{d}v)\,.$$

Note that $\sin(\theta)\mathrm{d}\theta\mathrm{d}\phi$ is an element of solid angle, and 4π is the total solid angle of a sphere.

We could also describe the distribution using rectangular coordinates. We must then find

$$f_{3\mathrm{p}}(v,\theta,\phi)\mathrm{d}v\mathrm{d}\theta\mathrm{d}\phi = F(v_x,v_y,v_z)\mathrm{d}v_x\mathrm{d}v_y\mathrm{d}v_z. \tag{5.12}$$

This is an example of the same idea as the one expressed in eq. (3.47), but now with three variables not one.[1] Hence, after using $\mathrm{d}v_x\mathrm{d}v_y\mathrm{d}v_z = v^2\sin\theta\,\mathrm{d}v\mathrm{d}\theta\mathrm{d}\phi$ (c.f. (5.8)), we have, for any distribution,

$$f_{3\mathrm{p}}(v,\theta,\phi) = v^2\sin(\theta)F(v_x,y_y,v_z) \tag{5.13}$$

and for the isotropic case

isotropic

$$f_{3\mathrm{p}}(v,\theta,\phi) = v^2\sin(\theta)F(v_x,y_y,v_z) \;=\; f(v)\frac{\sin\theta}{4\pi} \tag{5.14}$$

so in this case $f(v) = 4\pi v^2 F$.

5.3 Conclusion

To conclude: for an isotropic distribution, the probability for the direction to lie in some given range of θ and ϕ is proportional to the solid angle subtended by that range. Any other factors affecting the probability are contained in $f(v)$. The function $f(v)$ has the dimensions of an inverse speed and one integrates it over speed to get a probability. In terms of this function, the mean speed and the mean squared speed are

$$\langle v\rangle \;=\; \int_0^\infty v f(v)\mathrm{d}v \tag{5.15}$$

$$\langle v^2\rangle \;=\; \int_0^\infty v^2 f(v)\mathrm{d}v \tag{5.16}$$

[1] Using a Jacobian one would write $f_{3\mathrm{p}} = \left|\frac{\partial(v_x,v_y,v_z)}{\partial(v,\theta,\phi)}\right| F$.

More generally, the average of any quantity $Q(\mathbf{v})$ can be obtained from either of

$$\langle Q\rangle = \iiint_{-\infty}^{\infty} Q(\mathbf{v})F(v_x, v_y, v_z)\,\mathrm{d}^3\mathbf{v} \tag{5.17}$$

or

$$\langle Q\rangle = \int_0^{\infty}\int_0^{\pi}\int_0^{2\pi} Q(\mathbf{v})f_{3\mathrm{p}}(v,\theta,\phi)\,\mathrm{d}v\mathrm{d}\theta\mathrm{d}\phi. \tag{5.18}$$

The functions $f_{3\mathrm{p}}$ and F are related to one another by eq. (5.13). These expressions hold for any distribution. In the case of an isotropic distribution we may also use

$$\langle Q\rangle = \int_0^{\infty}\int_0^{\pi}\int_0^{2\pi} Q(\mathbf{v})f(v)\mathrm{d}v\frac{\sin\theta\,\mathrm{d}\theta\mathrm{d}\phi}{4\pi}. \tag{5.19}$$

The quantity Q here need not itself be isotropic.

©A.M. Steane 2026, not for distribution.

Exercises

(1) Starting from (5.4), obtain (5.6).

(2) For the distribution described by (5.6), obtain $\langle v^2\rangle$ by two methods:
(i) Use polar coordinates and $f(v)$.
(ii) Use rectangular coordinates and F given by (5.3), employing $v^2 = v_x^2 + v_y^2 + v_z^2$.
(You may wish to use the results listed in appendix B, paying attention to the ranges of integration).

(3) (i) Obtain $\langle\cos^2(\theta)\rangle$ by integration for an isotropic distribution in three dimensions.
(ii) The answer is 1/3. How could you have predicted this, without doing any integrals, by using $v^2 = v_x^2 + v_y^2 + v_z^2$?

(4) For velocities in three dimensions, consider the distribution function

$$g(v,\theta) = f(v)\frac{1}{2}\sin\theta$$

where v and θ are the first two coordinates in the system of spherical polar coordinates (v,θ,ϕ). Explain how this distribution function relates to $f_{3\mathrm{p}}$. Does it describe an isotropic distribution?

(5) If you are familiar with Jacobians, then confirm eq. (5.13) using a Jacobian method.

Effusion

6

When a hole is opened in the wall of a container of gas, with vacuum or a gas at lower pressure on the other side of the hole, then the gas will begin to move through the hole. If the hole is large then the gas flows in bulk, the pressure behind pushing the gas in front so that it accelerates through the hole. If the hole is small then a different kind of flow takes place, called *molecular effusion*. In effusion the gas is almost unaffected by the presence of the hole: there is negligible disturbance of the gas except that molecules that happen to be near the hole may fly through it. They are not accelerated but simply pass through the hole with whatever velocity they already have. For a circular hole of radius a the condition for this kind of flow is

$$a \ll \lambda \tag{6.1}$$

where λ is the **mean free path**. This is the average distance travelled by any molecule between collisions with other molecules. We shall study the mean free path more fully in chapter 9; all we need for now is to mention the condition 6.1.

In this chapter we shall first calculate effusion while ignoring the possibility of collisions between molecules. Then we shall briefly discuss at the end how collisional effects modify the results. In order to calculate effusion we will be using the results described in the previous chapter for the isotropic velocity distribution. As a warm-up exercise we first consider the calculation of pressure using those ideas.

6.1 Pressure calculation using polar coordinates

Consider a small region A on the wall of chamber, and construct spherical polar coordinates such that the axis of the coordinate system is perpendicular to the wall, see Fig. 6.1. Consider the particles travelling in the directions close to (θ, ϕ) at speeds close to v, with an isotropic velocity distribution. Using (5.14), the number density of such particles anywhere in the gas is

$$nf(v)\mathrm{d}v\frac{\sin(\theta)\,\mathrm{d}\theta\mathrm{d}\phi}{4\pi},$$

where, as usual, n is the number of particles per unit volume and $f(v)$ is the speed distribution function. Among all the particles in this velocity

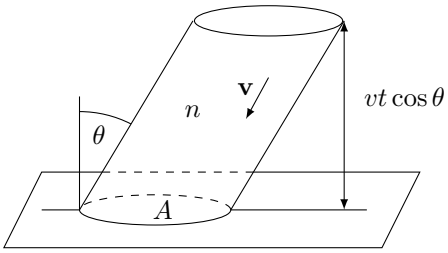


Fig. 6.1 Particles effusing through a hole of small but non-infinitesimal area A. The diagram shows the region occupied by the particles of given velocity (v, θ, ϕ) which will exit the hole in the next time interval t. This is a skew cylinder of volume $Av\cos(\theta)t$. The corresponding region for the particles at velocity $(v+\mathrm{d}v,\ \theta+\mathrm{d}\theta,\ \phi+\mathrm{d}\phi)$ also has this same volume in the limit In the limit $\{\mathrm{d}v, \mathrm{d}\theta, \mathrm{d}\phi\} \to 0$ (since A is not itself infinitesimal.)

class (v, θ, ϕ), now consider just those which are positioned such that they are moving towards the region A. In a time t those lying in the region shown in the diagram arrive at the section of wall. The volume of this region is $Avt\cos\theta$. Multiplying the volume by the number density, we deduce that in the time t the number of particles in the given velocity-class which hit the section of wall is

$$\mathrm{d}N = nf(v)\mathrm{d}v\frac{\sin(\theta)\,\mathrm{d}\theta\mathrm{d}\phi}{4\pi}Avt\cos\theta$$

The momentum component in the direction perpendicular to the wall, for each such molecule, is $mv\cos\theta$, so the rate at which perpendicular momentum is *brought up* to the wall, per unit area, is

$$\frac{\mathrm{d}N mv\cos\theta}{At} = nf(v)\mathrm{d}v\frac{\sin(\theta)\,\mathrm{d}\theta\mathrm{d}\phi}{4\pi}v\cos(\theta)\,mv\cos(\theta).$$

On average the collision with the wall reverses the momentum component perpendicular to the wall (see section 4.2), so the total momentum exchanged with the wall, per unit area per unit time, is twice this. Therefore the pressure is

$$P = 2nm\int_0^\infty v^2 f(v)\,\mathrm{d}v\int_0^{\pi/2}\sin(\theta)\cos^2(\theta)\mathrm{d}\theta\int_0^{2\pi}\frac{\mathrm{d}\phi}{4\pi}$$

Notice the range of the integrals here. We only integrate θ up to $\pi/2$ because molecules only approach the wall from one side. The integral $\int_0^{2\pi}\mathrm{d}\phi$ evaluates to 2π. We have also

$$\int_0^{\pi/2}\sin(\theta)\cos^2(\theta)\mathrm{d}\theta = \left[-\frac{1}{3}\cos^3(\theta)\right]_0^{\pi/2} = \frac{1}{3}$$

hence

$$P = \frac{1}{3}nm\langle v^2\rangle$$

which is eq. (4.11).

6.2 Effusion calculation

The calculation of the rate of effusion of gas through a hole goes much like the pressure calculation presented in the previous section. We model an effusion process by arguing that, in any given time t, the molecules lying in the region shown in Fig. 6.1 emerge through the hole. Therefore the rate at which molecules in the velocity class at (v, θ, ϕ) emerge is

$$\begin{aligned} \text{rate} &= \frac{(\text{number density}) \times (\text{volume})}{(\text{time})} \\ &= nf(v)\mathrm{d}v \frac{\sin(\theta)\mathrm{d}\theta\mathrm{d}\phi}{4\pi} \frac{Avt\cos(\theta)}{t}. \end{aligned} \tag{6.2}$$

The total flux Φ (that is, the total rate per unit area) is the integral of this quantity after dividing by A, giving:

$$\Phi = n\int_0^\infty vf(v)\mathrm{d}v \int_0^{\pi/2} \sin(\theta)\cos(\theta)\mathrm{d}\theta \int_0^{2\pi} \frac{\mathrm{d}\phi}{4\pi}. \tag{6.3}$$

The first integral is $\langle v\rangle$, the second is $\frac{1}{2}$ and the third is $\frac{1}{2}$, so we find

Effusion: flux for isotropic gas

$$\Phi = \frac{1}{4}n\langle v\rangle \tag{6.4}$$

Note the similarity with eq. (4.16). The factor 1/4 and the average account for the fact that we are here dealing with a range of angles and speeds. For this result we have assumed the velocity distribution in the gas is isotropic, but no other assumption about the velocity distribution was needed.

6.2.1 Effusion between regions at different temperature

Consider a vessel divided into two parts by a fixed partition wall which has in it some holes whose radius is small compared the mean free path and large compared to the molecular diameter. The vessel is filled with gas. The region on one side of the partition is maintained at temperature T_1; the region on the other side is maintained at temperature T_2. Let n_1, n_2 be the number densities in the two regions. We have

$$\begin{aligned} \text{effusion rate from 1 to 2} &= \frac{1}{4}An_1\langle v_1\rangle, \\ \text{effusion rate from 2 to 1} &= \frac{1}{4}An_2\langle v_2\rangle. \end{aligned} \tag{6.5}$$

where A is the total area of the holes. If these two rates do not agree, then there will be a net flow of gas, so the density on one side will fall while that on the other increases. The resulting changes in n_1 and n_2 will tend to reduce the imbalance of the flow rates, so eventually a dynamic equilibrium can be established, in which

$$n_1\langle v_1\rangle = n_2\langle v_2\rangle. \tag{6.6}$$

Note, this is not a complete thermal equilibrium of the whole gas, since the two parts are being maintained at different temperatures. But each part can be considered to be in internal equilibrium, and in that case it is found that the mean speed for that part is proportional to the square root of its temperature. Hence we have

$$n_1\sqrt{T_1} = n_2\sqrt{T_2}. \tag{6.7}$$

It follows that the pressures in the two parts will not agree. This is in contrast to what would happen if the holes were large. In that case gas would continue to move until the pressure was uniform.

6.3 The distribution over speed and angle

Let us now rewrite eq. (6.2), after dividing by A so as to get a flux:

$$\mathrm{d}\Phi = vf(v)\mathrm{d}v\frac{\sin(\theta)\cos(\theta)\mathrm{d}\theta\mathrm{d}\phi}{4\pi}. \tag{6.8}$$

This formula gives the flux distribution in a particle beam formed by effusion from a hole, for example. The presence of $vf(v)$ in this formula causes widespread confusion among students and teachers. It is commonly said (even in textbooks) that 'the distribution in a beam is different from the one in gas' but this is not true. The difference between $vf(v)$ and $f(v)$ is not owing to a differences between a beam and a gas. It is owing to the difference between a flux and a density.

To clarify this, let's carefully define two sets of molecules. Let $\mathcal{R}$ stand for the range of velocities (v, θ, ϕ) to $(v+\mathrm{d}v, \theta+\mathrm{d}\theta, \phi+\mathrm{d}\theta)$. One can pick out two different (but overlapping) sets of molecules. In each system (a gas or a beam), define:

- (Set A) = all those molecules which are situated somewhere in the total volume V (of the gas or the beam), and whose velocities are in the range $\mathcal{R}$.
- (Set B) = all those molecules which pass through a fixed plane (in the gas or the beam) in a given time interval t, and whose velocities are in the range $\mathcal{R}$.

All we need to do is note that if the probability density function for set A is $f(v)$, then the probability density function for set B is proportional to $vf(v)$. This is because for given t the region containing molecules contributing to set B has a length proportional to v, so set B picks out more of the faster molecules than the slower ones from set A. Table 6.1 summarizes the results. The table gives the distribution functions for the case where the gas is in thermal equilibrium and the effusion is gradual enough not to disturb this equilibrium. This makes it clear that there are indeed two possible distribution functions, depending on which method is used to specify a group of molecules, but the distributions agree for the two type of physical systems (gas or effusing beam).

	set A (given volume)	set B (given time interval)
gas in a chamber	$v^2e^{-\alpha v^2}$	$v^3e^{-\alpha v^2}$
particle beam produced by effusion	$v^2e^{-\alpha v^2}$	$v^3e^{-\alpha v^2}$

Table 6.1 Probability distribution functions for sets of molecules in either a gas or an effusing beam. The sets A and B are specified in the text. The table gives the unnormalized probability distribution function in the case of a gas in thermal equilibrium and a beam effusing from such a gas. To obtain the probability density one normalizes this by diving by its integral over all v. The constant $\alpha = m/(2k_{\rm B}T)$ where m is the mass of one molecule, $k_{\rm B}$ is Boltzmann's constant and T is the temperature of the gas. The distributions for a gas and for a beam are exactly the same. However, when calculating properties of a gas one usually considers set A, whereas for a beam one often takes an interest in set B.

Example 6.1 For a gas in thermal equilibrium, compare the mean kinetic energy of particles anywhere in the gas with the mean kinetic energy of particles which cross a plane in some short time interval t.
Solution.
The mean kinetic energy of the particles anywhere in the gas is

$$\begin{aligned}\frac{1}{2}m\langle v^2\rangle_A &= \frac{\int \frac{1}{2}mv^2(v^2e^{-\alpha v^2})\mathrm{d}v}{\int (v^2e^{-\alpha v^2})\mathrm{d}v} \\ &= \frac{3}{2}k_{\rm B}T \qquad (6.9)\end{aligned}$$

where the integrals are obtained in appendix B, and we used $\alpha = m/(2k_{\rm B}T)$. The mean kinetic energy of the particles crossing a plane in a fixed short time is

$$\begin{aligned}\frac{1}{2}m\langle v^2\rangle_B &= \frac{\int \frac{1}{2}mv^2(v^3e^{-\alpha v^2})\mathrm{d}v}{\int (v^3e^{-\alpha v^2})\mathrm{d}v} \\ &= 2k_{\rm B}T \qquad (6.10)\end{aligned}$$

Hence the average kinetic energy of the second group is larger than that of the first by a factor 4/3. This is because the selection by crossing a plane in a fixed time preferentially selects the faster molecules.

A similar issue arises for the distribution over angle. For set A, the distribution is isotropic, so the distribution function over angles θ, ϕ is proportional to $\sin\theta$. For set B the distribution function has an extra factor of $\cos\theta$. So if we are interested in *flux*, which is all about *rate*, or *how many in a given time*, then we shall encounter this factor of $\cos\theta$. In that case we deduce that flux per unit solid angle is greatest in the direction $\theta = 0$ and falls away at larger angles, reaching zero at $\theta = \pi/2$.

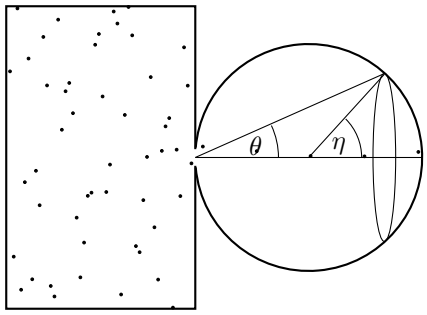


Fig. 6.2 Geometry for paint effusing from a low-pressure gas into an evacuated sphere.

Example 6.2 Particles of paint effuse from a gas in equilibrium into an evacuated sphere, through a small hole in the side of the sphere. They then fly across the sphere and stick wherever they land. Show that the internal surface of the sphere will be painted uniformly.
Solution.
The geometry is shown in Fig. 6.2. It is obvious by symmetry that the painting is independent of ϕ, so we focus on the distribution over θ. During some fixed time t, the number of particles passing through the hole in directions between θ and $\theta + \mathrm{d}\theta$ is proportional to

$$\sin(\theta)\cos(\theta)\mathrm{d}\theta$$

(as given by eq. (6.8)). These particles land on a ring-shaped strip on the interior of the sphere, as shown in the diagram. The radius of this strip is $R\sin\eta$ and its width is $R\mathrm{d}\eta$, so the amount of paint per unit area is proportional to

$$\frac{\sin(\theta)\cos(\theta)\mathrm{d}\theta}{\sin(\eta)\mathrm{d}\eta} \tag{6.11}$$

where the proportionality constant involves π and R and other constants, but not θ or η. But one can easily show from the circular geometry that $\eta = 2\theta$, so $\sin(\eta) = \sin(2\theta) = 2\sin(\theta)cos(\theta)$. Therefore all the functional dependence in (6.11) cancels and the amount of paint per unit area is independent of θ. QED.

6.4 Non-effusive flow

The process called effusion is the exception not the rule for the flow of fluids in general. The more common case is *bulk flow*, in which collisions

are non-negligible. In this case each small part of the fluid exerts forces on its neighbours. The general case can exhibit acceleration and turbulence. If there is a pressure gradient in a fluid then there is a difference in the forces acting on each side of any small part of the fluid, with the result that that part accelerates. There may also be viscous forces which oppose the motion, so that a steady-state velocity may be reached. We shall consider some of these effects in chapter 11.

©A.M. Steane 2026, not for distribution.

Exercises

In the following exercises, you may use that the mean speed of molecules in an ideal gas in thermal equilibrium is $\langle v\rangle = (8k_{\rm B}T/\pi m)^{1/2}$, where m is the mass of one molecule, and the relation between pressure and temperature is $P = nk_{\rm B}T$ where n is number density.

(1) A closed vessel is partially filled with liquid mercury; there is a hole of area $10^{-7}\,{\rm m}^2$ above the liquid level. The vessel is placed in a region of high vacuum at 273 K and after 30 days is found to be lighter by $2.4\times10^{-5}\,{\rm kg}$. Estimate the vapour pressure of mercury at 273 K. (The relative molecular mass of mercury is 200.59.) This exercise illustrates the **Knudsen effusion method** for measuring vapour pressure.

(2) A bottle of gas at high pressure develops a leak through a small hole of area A. Assuming the pressure outside the bottle is negligible, and the temperature of the gas remaining in the bottle stays constant, show that the gas pressure falls exponentially with time, and find an expression for the time constant.

(3) A gas at temperature T leaks through a small hole into a vacuum. A thermally insulated box is opened for a short time and catches some of the molecules. Neglecting the thermal capacity of the box, calculate the temperature of the gas trapped in the box after it reaches equilibrium.

(4) A beam of paint effusing from a hole is incident on a flat surface where it sticks. Describe the distribution of the thickness of the paint after some given elapsed time.

(5) A gas is a mixture of H_2 and HD in the proportion 8000 : 1, where H is hydrogen and D is deuterium. As the gas effuses through a small hole from a vessel at constant temperature into a vacuum the composition of the remaining mixture changes. Why? After some time, the remaining mixture consists of H_2 and HD in the proportion 800 : 1. By what factor will the pressure in the vessel have then fallen? (You may neglect the contribution of HD to the pressure, and take the relative molecular masses to be 2 and 3 for H_2 and HD, respectively). (This application of effusion has some practical importance.)

(6) Consider the two-chamber vessel described in section 6.2.1. Let n_1, n_2 and P_1, P_2 be the number densities and pressures in the two

parts. If the gas initially has some other value of either or both of the ratios n_1/n_2 and P_1/P_2, then as the system adjusts towards equilibrium, (i) Is it possible for there to be a net flow from a region of lower to higher pressure? (ii) Is it possible for there to be a net flow from a region of lower to higher temperature? (iii) Once the dynamic equilibrium is established, at what rate is energy continuously transferred between the two regions by the two flows?

(7) In practice a wall cannot be infinitely thin so what is called a 'hole' is really a short tube. It is necessary to allow for this when using effusion for vapour pressure measurements; the associated factor is called the **Clausing factor**. (i) For a cylindrical tube of radius a and length L, what is the probability that a molecule arriving at the centre of the opening will reach the other end without hitting the tube wall? (Assume the arriving molecules are drawn from a gas with isotropic velocity distribution). (ii) If a molecule strikes the tube wall half way along, what is the probability that that molecule will eventually re-enter the gas?

(8) In the presence of effusion the density of the gas near the hole is slightly diminished. Estimate the size of this effect at a distance $\lambda \gg a$ from a hole of radius a, ignoring collisions. [Hint: the density at any given location is maintained by the flux of molecules arriving there from other locations.]