Mathematical methods for ...
By: Straub, John Edward

9 Fundamentals of probability and statistics

9.1 Probability distributions of discrete variables 205
9.2 Probability distributions of continuous variables 214
9.3 Probability distributions in the physical sciences 224
A9 Connecting the gaussian and binomial probability distributions 230
B9 Uniform distributions of independent random variables 231
C9 Gaussian distributions of independent random variables 233
D9 Three definitions of Pythagorean means 234
E9 Propagation of error through total differentials and Taylor series 235
F9 End-of-chapter problems 238

9.1 Probability distributions of discrete variables

Functions of discrete variables arise in problems in which a number of observations are made of a property of the system consisting of individually separate and distinct states. Matter at the atomic level is discrete. Energy levels in quantum mechanical systems are discrete. Simple models of classical phenomena such as gases, liquids, and solids often assume a discrete set of possible states of the system. This section explores functions of discrete variables commonly used in the physical sciences.

9.1.1 *Factorials*

Consider the set of numbers $\{1,2,3,4\}$. The number of ways to order the four numbers is

$$4 \times 3 \times 2 \times 1$$

since there are four possible choices for the first number in the sequence, leaving three choices for the second number, two choices for the third number, and one choice for the fourth number. This sequential product is known as a *factorial* and defined as

$$n! = n \times (n-1) \times (n-2) \times \ldots \times 1$$

where we pronounce $n!$ as *n factorial*. We say that there are $4! = 4 \times 3 \times 2 \times 1 = 24$ ways to order the four numbers.[1] The dependence of $n!$ on n is displayed in Figure 9.1 on a semi-log plot.

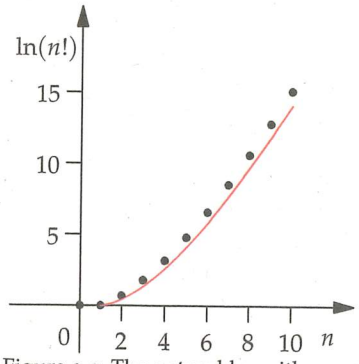


Figure 9.1: The natural logarithm of $n!$ (black dots) for $n \in [0,10]$. Shown for comparison is Sterling's approximation $\ln(n!) \approx n \ln n - n + 1$ (red line).

Let's explore a few examples to appreciate the size of factorials. Suppose we have a peptide consisting of a sequence of 10 unique amino acids. There are $10! = 3.6 \times 10^6$ possible unique peptide sequences. Consider a deck of 52 playing cards. There are $52! \approx 8.1 \times 10^{67}$ ways of ordering the deck of cards. To appreciate how large this number is, consider that there may be 10^{22} stars in the observable Universe and 10^{50} atoms composing the Earth. If we shuffled our deck of cards once every four seconds, we would explore a meager 10^{17} shuffles over the age of the Universe. Now consider a crystal consisting of $N_0 = 6 \times 10^{23}$ indistinguishable atoms. There are $N_0!$ ways of arranging the atoms in the crystal, a number that seems too large to imagine.

[1] Note that as $n! = n \times (n-1)!$ it must be that $1! = 1 \times 0! = 1$ so that $0! = 1$.

9.1.2 *The binomial probability distribution*

Consider an event with two possible outcomes, such as the flip of a coin. Further consider the probability of observing an outcome of interest, referred to as a favorable outcome, such as observing two heads in two tosses of a coin.[2] For a fair coin, for each coin flip the probability of observing a *head* is $\frac{1}{2}$ and the probability of observing a *tail* is $\frac{1}{2}$.

We assume that the probability of observing a head or tail is independent of whether a head or tail was observed the prior flip. The outcome of the coin toss is treated as an *independent random variable*. When an independent random variable is sampled (say by observing the flip of a coin), the result of the next observation (a second flip of the coin) is independent of all prior observations.[3]

It follows that if we have two coin flips, the probability of a specific outcome such as a head followed by a tail (written HT) is the *joint probability* of flipping a head first followed by a tail:

$$\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

In general, for N coin flips each with one of two possible outcomes, the total number of possible outcomes is 2^N and the probability of observing any one specific outcome is

$$\frac{1}{2^N}$$

Now suppose that we don't care about the order of outcomes. We only care how many heads and how many tails occur in N coin flips. Note that for two coin flips there is a total of $2^2 = 4$ possible outcomes:

HH HT TH TT

If we ignore the order, there are three unique outcomes, one with two heads, one with a head and a tail, and one with two tails. There is one way of observing two heads (HH), two ways of observing one head and one tail (HT, TH), and one way of observing two tails (TT). The results are distributed according to a *binomial distribution*, a distribution composed of two terms H and T.

The number of ways to order N outcomes is $N!$. For $N = 2$ coin flips, there will be $N! = 2! = 2$ ways of ordering any outcome. However, if we flip $n_1 = 2$ heads, there is really only one unique sequence. So we must divide 2! by the number of ways of permuting the 2 heads, which is $n_1! = 2! = 2$, times the number of ways of permuting 0 tails, which is $n_2! = 0! = 1$. We say there is one way of flipping two heads:

$$\frac{N!}{n_1!n_2!} = \frac{2!}{2!\,0!} = 1$$

Similarly, if we flip two tails there are $n_1 = 0$ heads, $n_2 = 2$ tails, and $\frac{N!}{n_1!n_2!} = \frac{2!}{0!\,2!} = 1$ way of flipping two tails. The number of ways to flip one head and one tail is the total number of ways $N!$ divided by the number of ways of permuting one head, $n_1! = 1$, times the number of ways of permuting

[2] The *probability* of an event is a measure of the likelihood of the event happening. Mathematically, the probability of a favorable event is the ratio of the number of favorable outcomes to the total number of possible outcomes. As such, the probability is expressed as a number between 0 and 1, with 0 representing impossibility and 1 representing certainty. The larger the probability of an event, the more likely the event is to occur.

[3] A *random process* is a series of actions having no defined method and leading to no discernable pattern of outcomes.

Mathematical methods for molecular science : theor
By: Straub, John Edward

one tail, $n_2! = 1$, or $\frac{N!}{n_1!n_2!} = \frac{2!}{1!\,1!} = 2$ ways to flip a head and a tail.

If we have three coin flips, there are $2^3 = 8$ possible outcomes:

HHH HHT HTH HTT THH THT TTH TTT

The number of ways of flipping three heads is $\frac{N!}{n_1!} = \frac{3!}{3!} = 1$, the number of ways of flipping two heads and one tail is $\frac{N!}{n_1!n_2!} = \frac{3!}{2!1!} = 3$, the number of ways of flipping one head and two tails is $\frac{N!}{n_1!n_2!} = \frac{3!}{1!2!} = 3$, and the number of ways of flipping three tails is $\frac{N!}{n_2!} = \frac{3!}{3!} = 1$. The resulting distribution is shown in Figure 9.2.

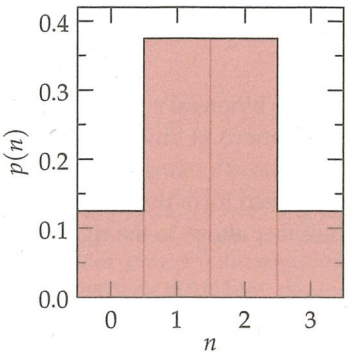


Figure 9.2: The binomial distribution for the relative probability of the four possible outcomes of $N = 3$ flips of a coin. Here we assume the order of events does not matter. As such, TTH, THT, and HTT are identical outcomes as each has $n = 1$ head and 2 tails. The same holds for THH, HTH, and HHT, which are three identical outcomes each having $n = 2$ heads and 1 tail.

In general, for N coin flips each having two possible outcomes, the number of ways of having a specific outcome with n_1 heads and n_2 tails is

$$W(n_1, n_2; N) = \frac{N!}{n_1!n_2!} = \frac{N!}{n_1!(N-n_1)!} \tag{9.1}$$

where we recognize that $N = n_1 + n_2$. The factor $W(n_1, n_2; N)$ is known as the *binomial coefficient*.

We can now consider the probability of observing a given outcome of interest in a series of coin flips. Recall that for N coin flips there are 2^N possible outcomes with each individual outcome occurring with a probability of $\frac{1}{2^N}$. As such, the probability of having N coin flips with n_1 heads and $n_2 = N - n_1$ tails is the probability $\frac{1}{2^N}$ times the number of outcomes having n_1 heads and n_2 tails:

$$p(n_1, n_2; N) = \frac{1}{2^N}\, W(n_1, n_2; N) = \frac{1}{2^N}\, \frac{N!}{n!(N-n)!} \tag{9.2}$$

In the final form of the equation we have set $n = n_1$. This result is known as the *binomial probability distribution*.[4]

[4] The binomial theorem was first mentioned by Greek mathematician *Euclid* (*circa* 300 BCE).

The resulting distribution of probabilities for the case $N = 6$ is shown in Figure 9.3. Note that the mean value of the distribution is $\frac{N}{2} = 3$. Further note that in comparison to the distribution for $N = 3$ (see Figure 9.2) the distribution for $N = 6$ is somewhat more sharply peaked.

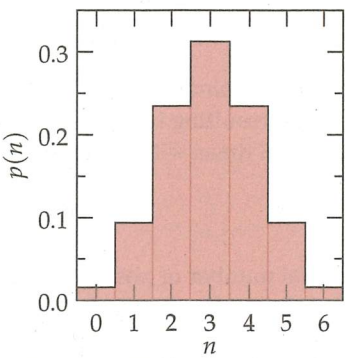


Figure 9.3: The binomial distribution for the relative probability of seven possible outcomes from $N = 6$ flips of a coin, where n is the total number of heads.

Let's evaluate the probabilities defined by Equation 9.2 for a few specific cases. For $N = 2$ events, the probabilities for the three possible outcomes (HH), (HT or TH), and (TT) are

$$\frac{1}{4} \quad \frac{2}{4} \quad \frac{1}{4}$$

Each probability is equal to $\frac{1}{2^2}$ times the binomial coefficient, 1, 2, or 1, expressing the number of ways of observing a given outcome (HH), (HT or TH), or (TT). The binomial coefficients form a series $1, 2, 1$.

For $N = 3$ events, the probabilities for the four possible outcomes (HHH), (HHT or HTH or THH), (TTH or THT or HTT), and (TTT) are

$$\frac{1}{8} \quad \frac{3}{8} \quad \frac{3}{8} \quad \frac{1}{8}$$

where the binomial coefficients form a series $1, 3, 3, 1$. For $N = 4$ events, the

probabilities for the five possible outcomes are

$$\frac{1}{16} \quad \frac{4}{16} \quad \frac{6}{16} \quad \frac{4}{16} \quad \frac{1}{16}$$

where the binomial coefficients form a series $1, 4, 6, 4, 1$. And so on.

The pattern of binomial coefficients can be captured graphically in the form of *Pascal's triangle* shown in Figure 9.4. The pattern starts with 1 at the top. We then form the next line by summing the number above to the left and the number above to the right.

$$\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & \boxed{1} & & \\
 & 1 & & 4 & & \boxed{6} & & 4 & & 1 & \\
1 & & \boxed{5} & & 10 & & 10 & & 5 & & 1 \\
\boxed{1} & & 6 & & 15 & & 20 & & 15 & & 6 \\
1 & & 7 & & 21 & & 35 & & 35 & & 21 \\
1 & & 8 & & 28 & & 56 & & 70 & & 56 \\
1 & & 9 & & 36 & & 84 & & 126 & & 126 \\
1 & & 10 & & 45 & & 120 & & 210 & & 256
\end{array}$$

$$\begin{array}{ccccccccccc}
 & & & & & & & & & & \\
 & & & & & & & & & & \\
 & & & & & & & & & & \\

Mathematical methods for molecular science : theor
By: Straub, John Edward

distinguishable, there are $2^N = 2^{10} = 1024$ possible ways of ordering the 10 spins.

And what if the spins are indistinguishable, rather than distinguishable? If 9 spins are up and 1 spin is down, the 1 down spin could be any of the 10 spins. In that case, the total number of ways of ordering $n_1 = 9$ up spins and $n_2 = 1$ down spin is $\frac{N!}{n_1!n_2!} = \frac{10!}{9!1!} = 10$. It follows that the total probability of observing 9 up spins and 1 down spin is $p(9,1;10) = \frac{10}{1024} = 0.00976$ or just less than 1%.[7]

[7] Consider this alternative derivation. There are 10 ways of ordering the 1 down spin. For a given down spin, the order of the 9 indistinguishable up spins does not matter. As a result, there is a total of 10 unique ways to order the up and down spins.

9.1.4 *The multinomial probability distribution*

Suppose we randomly choose a number from a set $\{1,2,3,4\}$. There are $k = 4$ numbers in the set and the probability of choosing any one number is $\frac{1}{4}$, which is 1 specific outcome divided by 4 possible outcomes. If we return to the full set $\{1,2,3,4\}$ and again choose a number at random, the probability of choosing any one number is again $\frac{1}{4}$. For the two choices, there are 4^2 possible outcomes so the probability of observing 1 specific outcome will be $\frac{1}{4^2} = \frac{1}{16} \approx 6.2\%$.

If we return to the full set $\{1,2,3,4\}$ a third time, the probability of choosing any one number is $\frac{1}{4}$. It follows that the probability of choosing any particular sequence of three numbers such as $2,3,4$ or $1,4,2$ or $3,1,1$ will be $\frac{1}{4^3} \approx 1.5\%$. In general, the probability of observing a particular sequence of N choices each with k possible outcomes is

$$\frac{1}{k^N}$$

Now let's ask what the probability of choosing a sequence of N numbers will be if we ignore the order in which the numbers are chosen. For example, for $N = 3$ choices of $k = 4$ numbers from the set $\{1,2,3,4\}$ the outcomes $2,3,4$ and $1,3,4$ are unique but $2,3,4$ and $3,2,4$ are taken to be the same. The probability of choosing a unique sequence of numbers will be the probability of choosing that sequence times the multiplicity of ways that the sequence can be chosen.

For example, there is one way a sequence with 3 twos can be chosen $(2,2,2)$, but there are three ways a sequence with 1 one and 2 twos can be chosen:

$$1,2,2 \qquad 2,1,2 \qquad 2,2,1$$

For a sequence in which each number is different such as $3,2,4$ there are six unique ways the numbers can be ordered:

$$3,2,4 \qquad 3,4,2 \qquad 2,3,4 \qquad 2,4,3 \qquad 4,3,2 \qquad 4,2,3$$

Let's generalize these observations. For a sequence of N choices from a set $\{1,2,3,4\}$ having n_1 ones, n_2 twos, n_3 threes, and n_4 fours, there will be

$$n_1!n_2!n_3!n_4!$$

ways of ordering the sequence where $n_1 + n_2 + n_3 + n_4 = N$. It follows that the number of unique sequences that can be chosen, having n_1 ones, n_2 twos, n_3 threes, and n_4 fours, is the total number of possible sequences $N!$ divided

by the number of non-unique ways of ordering the sequence $n_1!n_2!n_3!n_4!$, or

$$\frac{N!}{n_1!n_2!n_3!n_4!}$$

In general, for N choices from k distinguishable outcomes with n_j choices of the jth specific outcome, the number of ways a unique sequence can be chosen is

$$W(n_1, n_2, n_3, \ldots, n_k; N) = \frac{N!}{n_1!n_2!n_3!\ldots n_k!} \tag{9.3}$$

where $N = n_1 + n_2 + n_3 + \ldots + n_k$. The function $W(n_1, n_2, n_3, \ldots, n_k; N)$ is known as the *multinomial coefficient*.

The probability of observing N choices leading to an outcome defined by $n_1, n_2, n_3, \ldots, n_k$ will be

$$p(n_1, n_2, n_3, \ldots, n_k; N) = \frac{1}{k^N} W(n_1, n_2, n_3, \ldots, n_k; N)$$

which is $\frac{1}{k^N}$ times the multiplicity of ways the sequence can be chosen.

9.1.5 *Applications of the multinomial probability distribution*

The multinomial coefficient can be used to determine the number of ways one can roll a six-sided die N times with the outcome n_1 ones, n_2 twos, n_3 threes, n_4 fours, n_5 fives, and n_6 sixes. For example, the number of ways to roll the die $N = 5$ times, resulting in $n_1 = 2$ ones and $n_6 = 3$ sixes, is

$$W(2,0,0,0,0,3;5) = \frac{5!}{2!\,0!\,0!\,0!\,0!\,3!} = 10$$

The total number of possible outcomes rolling the die $N = 5$ times is $k^N = 6^5 = 7{,}776$. Therefore, the probability of rolling the die $N = 5$ times and having $n_1 = 2$ ones and $n_6 = 3$ sixes[8] is

$$\begin{aligned} p(2,0,0,0,0,3;5) &= \frac{1}{6^5} W(2,0,0,0,0,3;5) \\ &= \frac{10}{7{,}776} \approx 1.28 \times 10^{-3} \end{aligned}$$

For another example, consider $k = 6$ different levels of varying energy, ε, that can be occupied by $N = 24$ indistinguishable particles. The total number of ways the particles can be arranged over the six energy levels is given by the multinomial coefficient

$$W(n_1, n_2, n_3, \ldots, n_6; N) = \frac{N!}{n_1!n_2!n_3!\ldots n_6!}$$

where n_j is the number of particles occupying the jth energy level. Consider the distribution shown in Figure 9.5. For this case, $n_1 = 10$, $n_2 = 7$, $n_3 = 4$, $n_4 = 2$, $n_5 = 1$, and $n_6 = 0$. The number of ways the N indistinguishable particles can be arranged is

$$W(10,7,4,2,1,0;24) = \frac{24!}{10!\,7!\,4!\,2!\,1!\,0!} \approx 7.0 \times 10^{11} \tag{9.4}$$

[8] There are $6 \times 5 = 30$ different pairs that can be formed from the numbers $\{1,2,3,4,5,6\}$. As such, the probability of rolling a *full house* of 2 of one kind and 3 of another in 5 rolls is $30\,(1.28 \times 10^{-3}) = 3.85\%$.

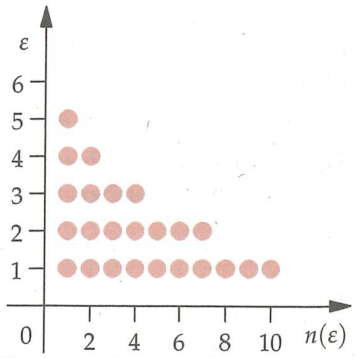


Figure 9.5: An arrangement of $N = 24$ particles (red balls) occupying $k = 6$ energy levels of varying energy ε. $n(\varepsilon)$ represents the number of particles occupying a level of energy ε.

Mathematical methods for molecular science : theor
By: Straub, John Edward

9.1.6 Moments of the binomial and multinomial distributions

Suppose there is a measurement on an observed property of a system. There are n possible outcomes sampled over N trials. The probability of observing a specific outcome is

$$p_k = \frac{N_k}{N} \qquad k = 1, 2, \ldots, n$$

where N_k is the number of times that the kth possible outcome is observed in N trials. We know that the total number of observations must equal the number of trials

$$\sum_{k=1}^{n} N_k = N$$

and the sum

$$\sum_{k=1}^{n} p_k = \sum_{k=1}^{n} \left(\frac{N_k}{N}\right) = \frac{1}{N}\sum_{k=1}^{n} N_k = \frac{1}{N}N = 1$$

Now suppose there is a property x that has a specific value x_k for each of the n possible outcomes. We can write the average or the *mean value* of x as the probability of a specific outcome p_k times the measured value of x for that outcome x_k summed over all possible outcomes. This can be written

$$\overline{x} = \sum_{k=1}^{n} x_k p_k$$

where $\overline{x}$ is the mean value of x.

We can also define the average of the square of the observable x, known as the *mean square value* of x and written[9]

$$\overline{x^2} = \sum_{k=1}^{n} (x_k)^2 p_k$$

[9] The averages $\overline{x}$ and $\overline{x^2}$ are also known as *moments* and provide insight into the shape of the distribution.

Now consider the mean square deviation of x from its average value $\overline{x}$:

$$\overline{(x-\overline{x})^2} = \sum_{k=1}^{n} (x_k - \overline{x})^2 p_k$$

This is known as the *variance*, and is often written

$$\sigma_x^2 \equiv \overline{(x-\overline{x})^2}$$

The variance provides a measure of the size of fluctuations in observations of x around its mean value $\overline{x}$.

If we expand the square, we find

$$\begin{aligned}
\overline{(x-\overline{x})^2} &= \sum_{k=1}^{n} \left((x_k)^2 - 2\overline{x}x_k + \overline{x}^2\right) p_k \\
&= \sum_{k=1}^{n} (x_k)^2 p_k - 2\overline{x}\sum_{k=1}^{n} x_k p_k + \overline{x}^2 \sum_{k=1}^{n} p_k \\
&= \overline{x^2} - 2\overline{x}^2 + \overline{x}^2 = \overline{x^2} - \overline{x}^2 \geq 0
\end{aligned}$$

This expression relates the mean-square difference to the averages $\overline{x^2}$ and $\overline{x}$.

The resulting identity for the variance is

$$\sigma_x^2 = \overline{(x-\overline{x})^2} = \overline{x^2} - \overline{x}^2 \tag{9.5}$$

This is a fundamental result that we will return to in many contexts in the physical sciences, including quantum theory and thermodynamics.

An interesting quantity is the square root of the variance, σ_x, known as the *standard deviation*. The ratio of the standard deviation to the mean value of the distribution

$$\frac{\sigma_x}{\overline{x}} = \frac{\sqrt{\overline{(x-\overline{x})^2}}}{\overline{x}}$$

provides a useful measure of the size of the fluctuations about the mean value relative to the mean value itself.

Consider the binomial probability distribution. There are two outcomes, each assumed to have probability $\frac{1}{2}$, so that

$$p(n;N) = \frac{1}{2^N}\frac{N!}{n!(N-n)!}$$

where

$$\sum_{n=0}^{N} p(n;N) = \frac{1}{2^N}\sum_{n=0}^{N}\frac{N!}{n!(N-n)!} = \frac{1}{2^N}2^N = 1 \tag{9.6}$$

The mean value of n for the binomial distribution is

$$\begin{aligned}\overline{n} &= \sum_{n=0}^{N} n\,p(n;N) = \frac{1}{2^N}\sum_{n=0}^{N} n\frac{N!}{n!(N-n)!} \\ &= \frac{1}{2^N}\sum_{n=1}^{N} n\frac{N!}{n!(N-n)!}\end{aligned} \tag{9.7}$$

where we note that the $n = 0$ term in the sum is zero. Rewriting the coefficient as

$$\begin{aligned}n\frac{N!}{n!(N-n)!} &= \frac{N!}{(n-1)!(N-n)!} \\ &= N\frac{(N-1)!}{(n-1)!((N-1)-(n-1))!}\end{aligned}$$

and substituting $r = n-1$, we find

$$\overline{n} = \sum_{n=0}^{N} n\,p(n;N) = \frac{1}{2^N}\sum_{r=0}^{N-1} N\frac{(N-1)!}{r!((N-1)-r)!} \tag{9.8}$$

In a final step, we note that the sum appearing in Equation 9.6 can alternatively be written in terms of the index r, with the sum extending from $r = 0$ to $r = N-1$:

$$\sum_{r=0}^{N-1}\frac{(N-1)!}{r!((N-1)-r)!} = 2^{N-1}$$

Mathematical methods for molecular science : theor
By: Straub, John Edward

Substituting this relation in Equation 9.8, we arrive at our final result

$$\overline{n} = \frac{N}{2^N} \sum_{r=0}^{N-1} \frac{(N-1)!}{r!((N-1)-r)!} = \frac{N}{2^N} 2^{N-1} = \frac{N}{2} \tag{9.9}$$

This result makes sense. The binomial distribution has a single peak, and the average $\overline{n}$ is the maximum value and midpoint $N/2$.

With additional effort we can also prove that the mean square value of n is

$$\overline{n^2} = \sum_{n=0}^{N} n^2 p(n;N) = \frac{1}{2^N} \sum_{n=0}^{N} n^2 \frac{N!}{n!(N-n)!} = \frac{N^2}{4} + \frac{N}{4}$$

so that the variance is

$$\sigma_n^2 = \overline{(n-\overline{n})^2} = \left(\frac{N^2}{4} + \frac{N}{4}\right) - \left(\frac{N}{2}\right)^2 = \frac{N}{4} \tag{9.10}$$

As a result, the standard deviation is

$$\sigma_n = \frac{1}{2}\sqrt{N}$$

and the ratio of the standard deviation to the mean is[10]

$$\frac{\sigma_n}{\overline{n}} = \frac{\sqrt{N}}{N} = \frac{1}{\sqrt{N}}$$

[10] For the general case of the binomial process in which one outcome has probability p and the other probability $(1-p)$ we find the mean $\overline{n} = Np$ and variance $\sigma_n^2 = Np(1-p)$.

The mean and standard deviation are depicted for a binomial distribution with $N = 10$ in Figure 9.6.

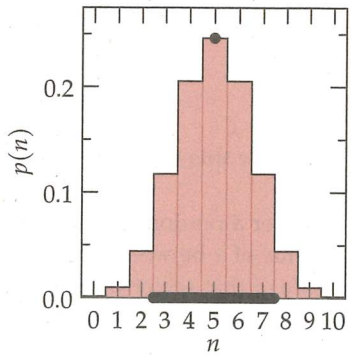


Figure 9.6: The binomial distribution for the relative probability of 11 possible outcomes resulting from $N = 10$ flips of a coin, where n is the total number of heads. The black dot marks the mean $\overline{n}$ and the thick black bar marks the range $\overline{n} - \sigma_n$ to $\overline{n} + \sigma_n$.

This final result states that as the number of observations N increases, the magnitude of fluctuations in the average value of n from the mean value $\overline{n}$ will decrease according to $N^{-1/2}$. This is a fundamental property of the binomial probability distribution that has important implications in the kinetic theory of matter.

Finally, consider the multinomial distribution in which there are k possible outcomes each with probability p_k. When there are N observations, the expected value of the kth outcome is

$$n_k = Np_k$$

The variance, defined as the mean square fluctuation in the kth outcome relative to the average, is

$$\sigma_k^2 = Np_k(1-p_k)$$

Let's consider the binomial distribution as a special case of the multinomial distribution. We take the probability of one outcome to be $p_1 = p$ and that of the second outcome to be $p_2 = 1 - p = q$. In that case, the mean value of observing the first outcome is $Np_1 = Np$ and the associated variance is $\sigma_1^2 = Np_1(1-p_1) = Np(1-p) = Npq$. These results agree with our prior results in Equation 9.9 and Equation 9.10 for the binomial distribution. For equal probabilities $p_1 = p_2 = \frac{1}{2}$, we found a mean of $\frac{N}{2}$ and variance $\frac{N}{4}$.

In the next section, we explore the properties of probability distributions of continuous variables that are commonly used in the physical sciences.

9.2 Probability distributions of continuous variables

FUNCTIONS OF CONTINUOUS VARIABLES arise in problems in which a number of observations are made of a property of the system consisting of a continuity of states. The position of atoms in space is defined in terms of continuous variables. The kinetic energy of atoms modeled as classical particles is taken to be continuous. The variables defining the pressure and temperature of a system are taken to be continuous. This section introduces probability distributions for continuous variables and explores a few functions of continuous variables commonly encountered in the physical sciences.

9.2.1 *Properties of probability distributions of a continuous variable*

For discrete variables such as a number of coin flips, the probability of a favorable event is the ratio of the number of favorable outcomes to the total number of possible outcomes. A sum over all possible outcomes must equal unity, the probability of any event happening. For a continuous variable, such as a particle's position or momentum, we need to generalize our concept of probability. The probability of observing a discrete state, p_k, is replaced by the probability of finding the variable within a range of values, $x \in [x, x+dx]$,[11] such that

$$p_k \to p(x)dx$$

where $p(x)$ is a *probability distribution function*. The sum over the probabilities of all discrete states is replaced by an integral

$$\sum_k p_k \to \int p(x)dx$$

Let's see how this works for the specific case of the *uniform probability distribution*.

Consider a random variable x. There is a constant probability C of observing a value of x between $x = a$ and $x = b$ and zero probability of observing x outside of that region. We define the corresponding uniform probability distribution function as

$$p(x) = \begin{cases} 0 & x < a \\ C & x \in [a, b] \\ 0 & x > b \end{cases}$$

The probability of selecting a value of x in the region $[x, x+dx]$ at random is $p(x)\ dx$, where dx represents a small increment of x.[12] The probability of randomly choosing a value of x somewhere between $x = a$ and $x = b$ must be unity. To reflect this, we say that the probability distribution is a *normalized distribution*, meaning that

$$\text{probability of finding } x \in [a, b] = \int_a^b p(x)\ dx = 1$$

The sum over all the probabilities of all possible outcomes must equal one.

[11] Note that p_k is dimensionless while $p(x)$ has units of $1/x$ making the product $p(x)dx$ dimensionless.

[12] Note that the increment dx has the units of x while the probability distribution function $p(x)$ has units of $1/x$. The product $p(x)dx$ is dimensionless.

Mathematical methods for molecular science : theor
By: Straub, John Edward

This normalization constant can be used to determine the value of C as

$$\int_a^b p(x)\,dx = \int_a^b C\,dx = C\int_a^b dx = C(b-a) = 1$$

From this result, we find $C = 1/(b-a)$. We call C the *normalization constant*. The resulting normalized uniform probability distribution function is

$$p(x) = \begin{cases} 0 & x < a \\ \dfrac{1}{b-a} & x \in [a,b] \\ 0 & x > b \end{cases}$$

This function is depicted in Figure 9.7.

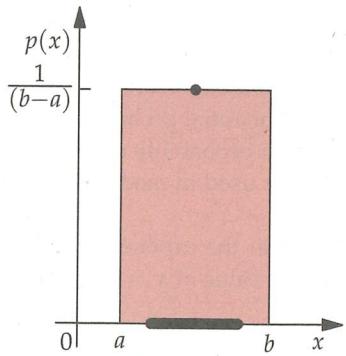


Figure 9.7: The uniform probability distribution function $p(x)$ as a function of x. The distribution function is normalized so that the total shaded area is unity. The black dot marks the mean $\overline{x}$ and the black bar marks the range $\overline{x} - \sigma_x$ to $\overline{x} + \sigma_x$.

The general definition of the mean value or average of x is

$$\overline{x} = \int_a^b xp(x)\,dx$$

For the uniform distribution, we evaluate $\overline{x}$ as

$$\begin{aligned}\overline{x} &= \frac{1}{(b-a)}\int_a^b x\,dx = \frac{1}{(b-a)}\left[\frac{1}{2}x^2\right]_a^b \\ &= \frac{1}{2}\frac{1}{(b-a)}(b^2-a^2) = \frac{1}{2}(b+a)\end{aligned}$$

We can reform this result as

$$\overline{x} = a + \frac{1}{2}(b-a)$$

This is a very intuitive result. The average value of x is the midpoint of the range $[a,b]$. For example, for the specific values $a = 0$ and $b = 1$ the mean, $\overline{x} = \frac{1}{2}$.

The average value of x^2 is defined as

$$\overline{x^2} = \int_a^b x^2 p(x)\,dx$$

which we evaluate for the uniform distribution as

$$\begin{aligned}\overline{x^2} &= \frac{1}{(b-a)}\int_a^b x^2\,dx = \frac{1}{(b-a)}\left[\frac{1}{3}x^3\right]_a^b = \frac{1}{3}\frac{1}{(b-a)}(b^3-a^3) \\ &= \frac{1}{3}(b^2+ab+a^2)\end{aligned}$$

For the specific values $a = 0$ and $b = 1$, we find $\overline{x^2} = \frac{1}{3}$.

With this result we can evaluate the variance of x as

$$\begin{aligned}\sigma_x^2 = \overline{(x-\overline{x})^2} = \overline{x^2} - \overline{x}^2 &= \frac{1}{3}(b^2+ab+a^2) - \left(\frac{1}{2}(b+a)\right)^2 \\ &= \frac{1}{12}(b-a)^2\end{aligned}$$

For the specific values $a = 0$ and $b = 1$ we find $\sigma_x^2 = \overline{x^2} - \overline{x}^2 = \frac{1}{12}$. Finally, the

ratio of the root-mean square value of $(x - \overline{x})$ and the mean $\overline{x}$ is

$$\frac{\sigma_x}{\overline{x}} = \frac{\sqrt{\overline{x^2} - \overline{x}^2}}{\overline{x}} = \frac{1}{\sqrt{3}}\left(\frac{b-a}{b+a}\right)$$

For the specific values $a = 0$ and $b = 1$ we find the ratio $\frac{1}{\sqrt{3}}$.

9.2.2 *Properties of exponential probability distributions*

Many measured observables in the physical sciences are well described by an exponential probability distribution function. Examples include the Boltzmann probability distribution, which is an exponential function of the energy used in modeling chemical equilibria and the rates of chemical reactions.

Consider the exponential distribution function $p(x)$, characterized by a maximum value at $x = 0$ and an exponential decrease with increasing x:

$$p(x) = \mathcal{C}\ \exp(-\gamma x)$$

where $\gamma > 0$. The probability distribution function is normalized, so that

$$\int_0^\infty p(x)\ dx = 1$$

which means that

$$\mathcal{C}\int_0^\infty \exp(-\gamma x)\, dx = \mathcal{C}\left[-\frac{1}{\gamma}e^{-\gamma x}\right]_0^\infty = \frac{1}{\gamma}\mathcal{C} = 1$$

This results in a normalization constant

$$\mathcal{C} = \gamma$$

With this result, we find the normalized exponential probability distribution function to be

$$p(x)\ dx = \gamma\ \exp(-\gamma x)\ dx$$

which is shown in Figure 9.8. We can define the probability of randomly selecting a value of x between $x = a$ and $x = b$:

$$\text{probability of finding } x \in [a, b] = \int_a^b p(x)\ dx$$

This is depicted as the shaded area in Figure 9.8.

The general definition of the mean value or average of x is

$$\overline{x} = \int_0^\infty x p(x)\ dx$$

For the exponential probability distribution we find

$$\overline{x} = \gamma\int_0^\infty x \exp(-\gamma x)\, dx$$

We can evaluate this integral using either integration by parts or our trick of exploiting differentiation with respect to the constant γ. In either case, we

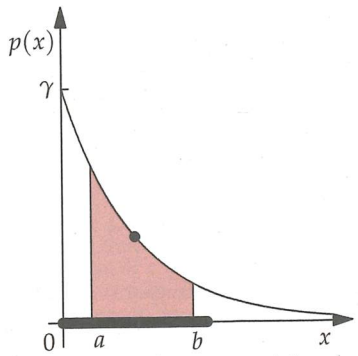


Figure 9.8: The exponential probability distribution function $p(x)$ as a function of x. The distribution function is normalized so that the total area under the curve is unity. The probability of randomly choosing a value of $x \in [a, b]$ is equal to the shaded area.

BRO0914

Mathematical methods for molecular science : theor
By: Straub, John Edward

find the mean:

$$\overline{x} = \gamma \int_0^\infty x \exp(-\gamma x)\, dx = \gamma \frac{1}{\gamma^2} = \frac{1}{\gamma}$$

The average value of x^2 is defined as

$$\overline{x^2} = \int_0^\infty x^2 p(x)\, dx$$

For the exponential probability distribution function we find

$$\overline{x^2} = \gamma \int_0^\infty x^2 \exp(-\gamma x)\, dx$$

We can evaluate this integral using either integration by parts or our trick of exploiting differentiation with respect to the constant γ, discussed in Chapter 7 in the complements. This leads to

$$\overline{x^2} = \gamma \int_0^\infty x^2 \exp(-\gamma x)\, dx = \gamma \frac{2}{\gamma^3} = \frac{2}{\gamma^2}$$

With this result, we can evaluate the variance

$$\sigma_x^2 = \overline{(x-\overline{x})^2} = \overline{x^2} - \overline{x}^2 = \frac{2}{\gamma^2} - \left(\frac{1}{\gamma}\right)^2 = \frac{1}{\gamma^2}$$

The ratio of the root-mean square value of x to the mean of x is

$$\frac{\sigma_x}{\overline{x}} = \frac{\sqrt{\overline{x^2} - \overline{x}^2}}{\overline{x}} = \frac{\sqrt{1/\gamma^2}}{1/\gamma} = 1$$

The exponential probability distribution has the special property that the root-mean square width of the distribution is equal to the mean.

[13] Note that $N(t)\, dt$ has units of number while $N(t)$ has units of number per unit time.

9.2.3 *Applications of the exponential probability distribution*

Consider the case of a sample of N_0 *radioactive atomic nuclei*. Observations of the system demonstrate that the rate of radioactive decay of the nuclei varies with time[13] according to the relation

$$N(t)\, dt = \frac{N_0}{\tau} \exp(-t/\tau)\, dt$$

where $N(t=0) = N_0/\tau$ is the initial rate of decay at $t = 0$ and $N(t)dt$ is the number of nuclei that decay between times t and $t + dt$. Over all time, the number of nuclei that will decay is

$$\int_0^\infty N(t)\, dt = \frac{N_0}{\tau} \int_0^\infty \exp(-t/\tau)\, dt = N_0$$

which is the total number of nuclei in the sample. The form of $N(t)$ is shown in Figure 9.9.

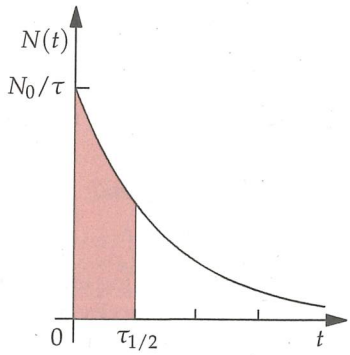


Figure 9.9: The exponential probability distribution function

$$N(t) = \frac{N_0}{\tau} \exp(-t/\tau)$$

as a function of time t. The shaded area represents half of the total area under the curve reached at the half-life time $t = \ln(2)\tau$.

Suppose we would like to know how many nuclei decay between a time t_a and t_b. We can calculate this as

$$\int_{t_a}^{t_b} N(t)\, dt = \frac{N_0}{\tau} \int_{t_a}^{t_b} \exp(-t/\tau)\, dt = N_0 \left[e^{-t_a/\tau} - e^{-t_b/\tau} \right]$$

If we set $t_a = 0$ and $t_b = t$, we find

$$\text{number of nuclei decayed} = N_0 \left[1 - \mathrm{e}^{-t/\tau}\right]$$

It follows that the *survival probability*, which is the probability that a nucleus has not decayed after time t, is

$$\text{survival probability} = \frac{N_0 - N(t)}{N_0} = \mathrm{e}^{-t/\tau}$$

For example, after a time $t = \ln(2)\tau = 0.693\tau$, the survival probability is $\frac{1}{2}$. We call this time the *half-life* of the exponential distribution:

$$\tau_{1/2} = \ln(2)\tau$$

After a time $\tau_{1/2}$, one-half of the nuclei have decayed and one-half have *survived*.

Let's consider another application involving an exponential distribution of energy:[14]

$$p(E) = C \exp(-E/k_B T)$$

[14] Note that $p(E)\,dE$ has units of number while $p(E)$ has units of number per unit energy.

To normalize this probability distribution, we demand that

$$\int_0^\infty p(E)\,dE = C \int_0^\infty \exp(-E/k_B T)\,dE = C k_B T = 1$$

so that $C = 1/k_B T$, and

$$p(E)\,dE = \frac{1}{k_B T} \exp(-E/k_B T)\,dE$$

What is the probabilty of having energy less than or equal to the thermal energy $k_B T$? We can evaluate this as

$$\frac{1}{k_B T} \int_0^{k_B T} \exp(-E/k_B T)\,dE = -\frac{1}{k_B T} k_B T \Big[\exp(-E/k_B T)\Big]_0^{k_B T}$$
$$= 1 - 1/e = 0.632$$

We can further ask for the fraction having energy less than $2k_B T$. That turns out to be

$$\frac{1}{k_B T} \int_0^{2k_B T} \exp(-E/k_B T)\,dE = 1 - \frac{1}{\mathrm{e}^2} = 0.865$$

or for the fraction having energy less than $3k_B T$, which is $1 - \frac{1}{\mathrm{e}^3} = 0.950$. This result is shown graphically in Figure 9.10.

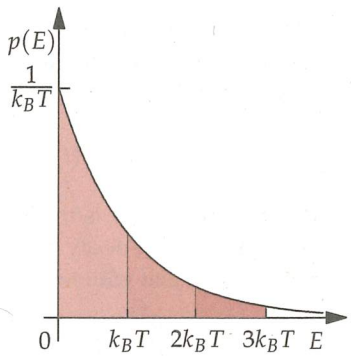


Figure 9.10: The exponential probability distribution function

$$p(E) = \frac{1}{k_B T} \exp(-E/k_B T)$$

as a function of energy E. The shaded areas represent the fraction of atoms having energy E less than $k_B T$, $2k_B T$, or $3k_B T$.

9.2.4 *Properties of gaussian probability distributions*

In the physical sciences, we commonly employ the gaussian function to represent a probability distribution function that is centered at $\overline{x} = x_0$ with mean square width σ^2, such as

$$p(x) = C \exp\left[-\frac{(x - x_0)^2}{2\sigma^2}\right]$$

BRO0914

Mathematical methods for molecular science : theor
By: Straub, John Edward

where the distribution function is normalized as

$$\int_{-\infty}^{\infty} p(x)\, dx = 1$$

Imposing the normalization condition, we find

$$\int_{-\infty}^{\infty} p(x)\, dx = C \int_{-\infty}^{\infty} \exp\left[-\frac{(x-x_0)^2}{2\sigma^2}\right] dx = 1$$

To perform this integral, we want to center the gaussian at the origin. We do this with the algebraic substitution $y = (x - x_0)$, so that $dy = dx$. We find

$$C \int_{-\infty}^{\infty} \exp\left[-\frac{(x-x_0)^2}{2\sigma^2}\right] dx = C \int_{-\infty}^{\infty} \exp\left(-\frac{y^2}{2\sigma^2}\right) dy$$

Using the identity

$$\int_{-\infty}^{\infty} e^{-\alpha y^2} dy = \sqrt{\frac{\pi}{\alpha}}$$

we find

$$C \int_{-\infty}^{\infty} \exp\left(-\frac{y^2}{2\sigma^2}\right) dy = C\sqrt{2\pi\sigma^2} = 1$$

so that the normalization constant $C = 1/\sqrt{2\pi\sigma^2}$. The resulting normalized gaussian probability distribution function is

$$p(x)\, dx = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-x_0)^2}{2\sigma^2}\right] dx$$

This result is shown in Figure 9.11.

Note that the probability of randomly selecting a value of x between $x_0 - \sigma$ and $x_0 + \sigma$ is

$$\text{probability of finding } x \in [x_0 - \sigma, x_0 + \sigma] = \int_{x_0-\sigma}^{x_0+\sigma} p(x)\, dx \approx 0.6827$$

The total probability of observing x within one standard deviation σ from the mean x_0 is depicted as the shaded area in Figure 9.11.

The probability of randomly selecting a value of x within two standard deviations of the mean, $x \in [x_0 - 2\sigma, x_0 + 2\sigma]$, is roughly 0.9545, and within three standard deviations of the mean, $x \in [x_0 - 3\sigma, x_0 + 3\sigma]$, is approximately 0.9973. This is the origin of the *68-95-99.7 rule* that is often evoked in discussions of the probability and statistics of the gaussian distribution.

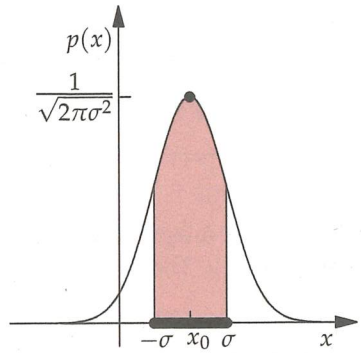


Figure 9.11: The normalized gaussian probability distribution function $p(x)$ as a function of x. The probability of randomly choosing a value of $x \in [-\sigma, \sigma]$ is equal to the shaded area.

The mean of x is defined as

$$\overline{x} = \int_{-\infty}^{\infty} x p(x)\, dx$$

For the gaussian distribution centered at $x = x_0$ with variance σ^2, we find

$$\overline{x} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x \exp\left[-\frac{(x-x_0)^2}{2\sigma^2}\right] dx$$

With the algebraic substitution $y = (x - x_0)$ so that $x = y + x_0$ and $dy = dx$,

we find

$$\begin{aligned}\overline{x} &= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}(y+x_0)\exp\left(-\frac{y^2}{2\sigma^2}\right)dy \\ &= \frac{1}{\sqrt{2\pi\sigma^2}}\left[\int_{-\infty}^{\infty} y\exp\left(-\frac{y^2}{2\sigma^2}\right)dy + x_0\int_{-\infty}^{\infty}\exp\left(-\frac{y^2}{2\sigma^2}\right)dy\right] \\ &= x_0\end{aligned}$$

The first integral is zero

$$\int_{-\infty}^{\infty} y\exp(-ay^2)\,dy = 0$$

as the integrand $y\ \exp(-ay^2)$ is an odd function of y. As such, the integral from $-\infty$ to 0 is equal in magnitude and opposite in sign to the integral from 0 to ∞.

The average value of x^2 is defined as

$$\overline{x^2} = \int_{-\infty}^{\infty} x^2 p(x)\,dx$$

For the gaussian distribution centered at $x = x_0$ with variance σ^2, we find

$$\overline{x^2} = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} x^2\exp\left[-\frac{(x-x_0)^2}{2\sigma^2}\right]dx$$

With the algebraic substitution $y = (x - x_0)$ so that $x^2 = y^2 + 2yx_0 + x_0^2$ and $dy = dx$, we find

$$\overline{x^2} = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}(y^2 + 2yx_0 + x_0^2)\exp\left(-\frac{y^2}{2\sigma^2}\right)dy$$

The first integral is

$$\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} y^2\exp\left(-\frac{y^2}{2\sigma^2}\right)dy = \sigma^2$$

The second integral is zero by symmetry since the integrand is an odd function of y. The third integral is equal to x_0^2. This leads to the final result

$$\overline{x^2} = x_0^2 + \sigma^2$$

Using this result we can determine the average of the square of the deviation in x from the mean value $\overline{x} = x_0$ as

$$\begin{aligned}\overline{(x-x_0)^2} &= \int_{-\infty}^{\infty}(x-x_0)^2 p(x)\,dx \\ &= \overline{x^2} - 2x_0\overline{x} + x_0^2 \\ &= \overline{x^2} - x_0^2 = \sigma^2\end{aligned}$$

This demonstrates that σ^2 is the variance of the gaussian distribution. In the next section, we explore a variety of applications of the gaussian distribution in the physical sciences.

BRO0914

Mathematical methods for molecular science : theor
By: Straub, John Edward

9.2.5 Applications of the gaussian probability distribution

Consider the kinetic energy of an ideal gas of atoms. In one dimension, the kinetic energy of an atom is

$$E = \frac{1}{2}mv^2$$

where m is the mass of the atom and v is the velocity. The distribution of kinetic energy is given by the Boltzmann exponential probability distribution

$$p(E)\,dE = \frac{1}{k_B T}\exp\left(-\frac{E}{k_B T}\right)\,dE$$

Since the energy varies in proportion to the square of the velocity, the distribution of velocities is the gaussian probability distribution function

$$p(v)\,dv = \sqrt{\frac{m}{2\pi k_B T}}\,\exp\left(-\frac{mv^2}{2k_B T}\right)\,dv$$

This distribution is shown graphically in Figure 9.12. The mean velocity is $\overline{v} = 0$, and the variance in the distribution of velocities is $\sigma_v^2 = \frac{k_B T}{m}$.

Suppose we want to know the fraction of atoms with kinetic energy less than or equal to $\frac{k_B T}{2}$, so that $\frac{1}{2}mv^2 \leq \frac{k_B T}{2}$. The velocities of atoms in this fraction are found in the range $v \in [-\sigma_v, \sigma_v]$ where $\sigma_v = \sqrt{\frac{k_B T}{m}}$. That magnitude of that fraction can be quantified using the integral

$$\sqrt{\frac{1}{2\pi\sigma_v^2}}\int_{-\sigma_v}^{\sigma_v}\exp\left(-\frac{v^2}{2\sigma_v^2}\right)\,dv = \mathrm{erf}\left(\frac{1}{\sqrt{2}}\right) = 0.683$$

where $\mathrm{erf}(x)$ is the *error function* introduced in Chapter 7 through Equation 7.18 and Figure 7.6.[15] It follows that the fraction of atoms having kinetic energy greater than $\frac{k_B T}{2}$ is $1 - \mathrm{erf}\left(\frac{1}{\sqrt{2}}\right) = \mathrm{erfc}\left(\frac{1}{\sqrt{2}}\right)$, where $\mathrm{erf}(x)$ is the *error function* and $\mathrm{erfc}(x)$ is the *complementary error function*.

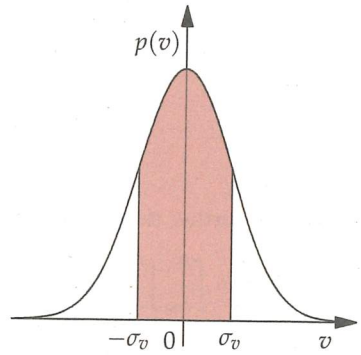


Figure 9.12: The integral over $p(v)$ between $-\sigma_v$ and σ_v is the fraction of atoms with kinetic energy $E \leq \frac{k_B T}{2}$, shown as the shaded area under the curve.

[15] The error function is defined as

$$\mathrm{erf}(x) = \frac{1}{\sqrt{\pi}}\int_{-x}^{x} e^{-t^2}\,dt$$

and is positive for positive values of x. In this case, $t^2 = \frac{v^2}{2\sigma_v^2}$. Since $v = \sigma_v$ corresponds to $t^2 = \frac{1}{2}$, the limit of integration $x = \frac{1}{\sqrt{2}}$.

9.2.6 Probability distributions in many dimensions

The concept of the probability distribution can be extended to many dimensions. Consider two random variables, x and y, where the probability of measuring a value of $x \in [x, x+dx]$ and $y \in [y, y+dy]$ is written

$$p(x,y)\,dxdy$$

where $p(x,y)$ is a two-dimensional probability distribution function. An example is shown graphically in Figure 9.13.

The integral over all x and y is the total probability of observing an outcome (x,y) somewhere on the xy-plane, which must equal unity:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\,dxdy = 1$$

This is our normalization condition. The probability of measuring $x \in [x, x+dx]$ and $y \in [y, y+dy]$ is the volume of the column over the area $dx\,dy$ under the surface defined by $p(x,y)$.

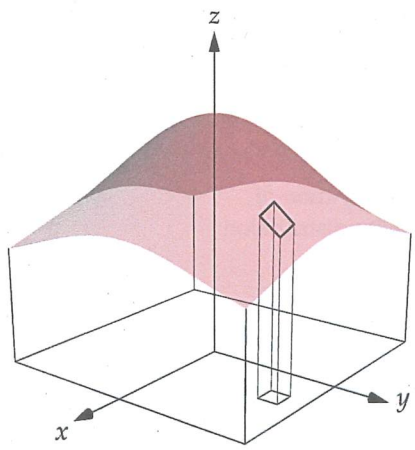


Figure 9.13: A two-dimensional probability distribution $p(x,y)$ shown as a surface over the xy-plane. The probability of observing a value of $x \in [x, x+dx]$ and $y \in [y, y+dy]$ is given by the volume of the vertical column over the area $dx\ dy$ and beneath the surface defined by $p(x,y)$.

We can further define the mean values of the x and y as

$$\overline{x} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\ p(x,y)\ dxdy \qquad \overline{y} = \int_{=\infty}^{\infty}\int_{-\infty}^{\infty} y\ p(x,y)\ dxdy$$

In defining the variances of this two-dimensional probability distribution function, we encounter a new concept. We define the variances σ_x^2 and σ_y^2 as

$$\sigma_x^2 = \overline{\overline{x^2} - \overline{x}^2} \qquad \sigma_y^2 = \overline{\overline{y^2} - \overline{y}^2}$$

where the mean square values of x and y are defined as

$$\overline{x^2} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2\ p(x,y)\ dxdy \qquad \overline{y^2} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y^2\ p(x,y)\ dxdy$$

We can also define a measure of the *cross correlation* between a deviation from the mean in x, $(x - \overline{x})$, and a deviation from the mean in y, $(y - \overline{y})$, as

$$\sigma_{xy} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - \overline{x})(y - \overline{y})\ p(x,y)\ dxdy$$

This is known as the *covariance*.

The value of σ_{xy} provides a measure of the correlation between observations of the two variables. We can define a useful measure of the magnitude of the correlation as $\sigma_{xy} = \rho_{xy}\ \sigma_x\sigma_y$ where ρ_{xy} is the *correlation coefficient*. When measurements of x and y are uncorrelated, the probability distribution $p(x,y)$ is separable and can be written as a product of the probability of observing a value of x times the independent probability of observing a value of y:

$$p(x,y)\ dxdy = p_x(x)\ dx \times p_y(y)\ dy$$

In that case, we find

$$\sigma_{xy} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - \overline{x})(y - \overline{y})\ p(x,y)\ dxdy$$

BRO0914

Mathematical methods for molecular science : theor
By: Straub, John Edward

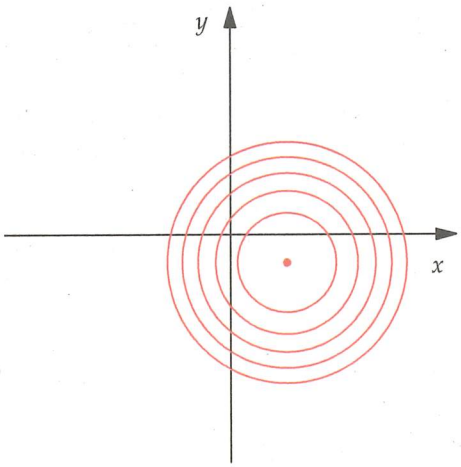


Figure 9.14: A two-dimensional probability distribution $p(x, y)$ over the xy-plane with correlation coefficient $\rho_{xy} = 0$. Contours define constant values of the probability and the dot marks the mean value of the distribution.

which can be evaluated as

$$\sigma_{xy} = \int_{-\infty}^{\infty} (x - \overline{x}) p_x(x)\, dx \times \int_{-\infty}^{\infty} (y - \overline{y})\, p_y(y)\, dy = 0$$

and the correlation coefficient $\rho_{xy} = 0$.

An example of a distribution with $\rho_{xy} = 0$ is shown in Figure 9.14. In this case, we say that the observations of x and y are uncorrelated. In general, the correlation coefficient will vary over the range: $-1 \leq \rho_{xy} \leq 1$ where positive values of ρ_{xy} indicate positive correlation and negative values of ρ_{xy} indicate negative correlation. When $\rho_{xy} = 0$, the variables x and y are independent. Examples of distributions with positive and negative values of ρ_{xy} are provided in Figure 9.15.

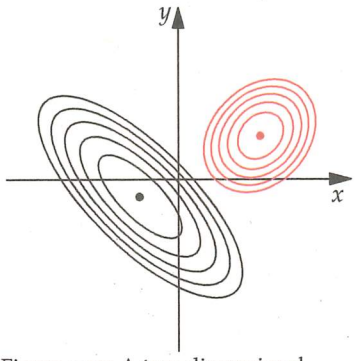


Figure 9.15: A two-dimensional probability distribution $p(x, y)$ over the xy-plane with correlation coefficient $\rho_{xy} = 1$ (red) and $\rho_{xy} = -1$ (black). Contours define constant values of the probability and dots mark the mean values of the distributions.

Consider the case of two fair coins. We expect the outcome of the flip of the first coin to be independent of the outcome of a flip of the second coin. In that case, the value of $\rho_{xy} = 0$ and the observations are uncorrelated. Now consider the case of an ideal gas for which the pressure can be written

$$p = \frac{nRT}{V}$$

Suppose the gas is at constant volume. The average pressure is measured to be $\overline{p}$ and the average temperature $\overline{T}$. If the pressure is higher than its mean value, so that $(p - \overline{p}) > 0$, we expect the temperature to be higher than its mean value, so that $(T - \overline{T}) > 0$. In this case we expect $\sigma_{pT} > 0$ and $\rho_{pT} > 0$. In such a case, we say that the pressure and temperature are positively correlated.

Now suppose the gas is at constant temperature. The average pressure is measured to be $\overline{p}$ and the average volume $\overline{V}$. If the pressure is higher than its mean value, then $(p - \overline{p}) > 0$. Given the inverse dependence of volume on pressure, if the pressure is greater than its mean value, the volume will be lower than its mean value and $(V - \overline{V}) < 0$. In this case we expect $\sigma_{pV} < 0$ and $\rho_{pV} < 0$. In such a case, we say that the pressure and volume are negatively correlated.

9.3 Probability distributions in the physical sciences

THE MOST COMMONLY ENCOUNTERED MULTIPLE INTEGRALS in the physical sciences involve integration over multivariate probability distributions. Examples include the Maxwell-Boltzmann probability distribution in kinetic theory, the Boltzmann probability distribution in thermodynamics, and square of the modulus of the wave function in quantum theory. The functions are typically complicated by the presence of many physical constants. However, careful simplification of the expressions by algebraic substitution followed by the right choice of coordinate system can convert an intimidating multiple integral into a readily evaluated expression. This section will explore some of the more common functional forms encountered in the physical sciences.

9.3.1 *Properties of the Maxwell-Boltzmann probability distribution*

Probability distributions are common in the physical sciences. We are often challenged to describe the properties of a very large number of particles, so large that it is hopeless to track the energy or velocity of each particle. Instead, we hope to describe the average properties of the system using statistical assumptions and statistical probability distributions.

The probability of observing a state of given energy can be written in terms of the Boltzmann exponential probability distribution

$$\exp\left(-\frac{E}{k_B T}\right)$$

Suppose E is the kinetic energy of a particle with velocity $\mathbf{v} = u_x\hat{\mathbf{x}} + u_y\hat{\mathbf{y}} + u_z\hat{\mathbf{z}}$. We can express the kinetic energy in vector notation as

$$E = \frac{1}{2}m\,\mathbf{v}\cdot\mathbf{v} = \frac{1}{2}m(u_x^2 + u_y^2 + u_z^2)$$

Inserting this energy into the Boltzmann probability distribution leads to the Maxwell-Boltzmann distribution

$$p(u_x, u_y, u_z) = \mathcal{C}\exp\left[-\frac{m}{2k_B T}(u_x^2 + u_y^2 + u_z^2)\right]$$

If we demand that the distribution is normalized such that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(u_x, u_y, u_z)du_x du_y du_z = 1$$

then

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathcal{C}\exp\left[-\frac{m}{2k_B T}(u_x^2 + u_y^2 + u_z^2)\right] du_x du_y du_z = 1$$

We note that this triple integral is separable and can be represented as a product of three one-dimensional integrals[16]

$$\mathcal{C}\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{m}{2k_B T}u_x^2}\,du_x \times \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{m}{2k_B T}u_y^2}\,du_y \times \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{m}{2k_B T}u_z^2}du_z$$
$$= \mathcal{C}\sqrt{\frac{2\pi k_B T}{m}} \times \sqrt{\frac{2\pi k_B T}{m}} \times \sqrt{\frac{2\pi k_B T}{m}} = \mathcal{C}\left(\frac{2\pi k_B T}{m}\right)^{\frac{3}{2}} = 1$$

[16] We use the identity

$$\int_{\infty}^{\infty} \mathrm{e}^{-\alpha x^2}\,dx = \sqrt{\frac{\pi}{\alpha}}$$

where $x = u_x$ and $\alpha = \frac{m}{2k_B T}$.

BRO0914

Mathematical methods for molecular science : theor
By: Straub, John Edward

As a result, the normalization constant $C = \left(\frac{m}{2\pi k_B T}\right)^{\frac{3}{2}}$. The final result for the statistical distribution of velocities is

$$p(u_x, u_y, u_z) = \left(\frac{m}{2\pi k_B T}\right)^{\frac{3}{2}} \exp\left[-\frac{m}{2k_B T}(u_x^2 + u_y^2 + u_z^2)\right]$$

Note that

$$p(u_x, u_y, u_z)\ du_x du_y du_z$$

expresses the probability of finding a particle with velocity components in the range

$$[u_x, u_x + du_x] \times [u_y, u_y + du_y] \times [u_z, u_z + du_z]$$

which is a small increment of volume $du_x du_y du_z$ in the space of all possible velocities.

9.3.2 *Mean and variance of the Maxwell-Boltzmann distribution*

We can compute the mean and variance of the Maxwell-Boltzmann distribution using a similar approach. Consider first the mean of the velocity defined as

$$\begin{aligned}\overline{\mathbf{v}} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{v}\ p(u_x, u_y, u_z)\ du_x du_y du_z \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (u_x\hat{\mathbf{x}} + u_y\hat{\mathbf{y}} + u_z\hat{\mathbf{z}})\ p(u_x, u_y, u_z)\ du_x du_y du_z\end{aligned}$$

where integration is carried out in cartesian coordinates, defined in Figure 9.16. This integral is a sum of three triple integrals, each corresponding to the average of the velocity in one of the three principal directions x, y, and z depicted in Figure 9.17. As such, we express the average velocity as

$$\overline{\mathbf{v}} = \overline{u_x}\ \hat{\mathbf{x}} + \overline{u_y}\ \hat{\mathbf{y}} + \overline{u_z}\ \hat{\mathbf{z}}$$

where the averages are defined

$$\begin{aligned}\overline{u_x} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u_x\ p(u_x, u_y, u_z)\ du_x du_y du_z \\ \overline{u_y} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u_y\ p(u_x, u_y, u_z)\ du_x du_y du_z \\ \overline{u_z} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u_z\ p(u_x, u_y, u_z)\ du_x du_y du_z\end{aligned}$$

Each of the averages is a three-dimensional integral that we recognize is separable and can be evaluated as a product of three one-dimensional integrals.

The average $\overline{u_x}$ can be expressed as

$$\begin{aligned}\overline{u_x} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u_x\ p(u_x, u_y, u_z)\ du_x du_y du_z \\ &= \left(\frac{m}{2\pi k_B T}\right)^{\frac{3}{2}} \int_{-\infty}^{\infty} u_x e^{-\frac{mu_x^2}{2k_B T}}\ du_x \times \int_{-\infty}^{\infty} e^{-\frac{mu_y^2}{2k_B T}}\ du_y \times \int_{-\infty}^{\infty} e^{-\frac{mu_z^2}{2k_B T}} du_z\end{aligned}$$

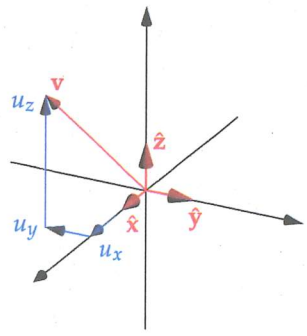


Figure 9.16: The decomposition of the velocity vector $\mathbf{v} = u_x\ \hat{\mathbf{x}} + u_y\ \hat{\mathbf{y}} + u_z\ \hat{\mathbf{z}}$ in terms of its cartesian components.

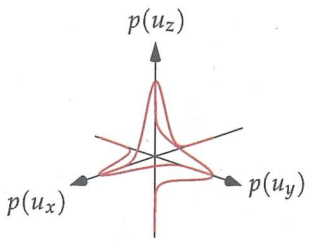


Figure 9.17: Projections of the three one-dimensional Maxwell-Boltzmann velocity probability distributions $p(u_x)$, $p(u_y)$, and $p(u_z)$. The product of these three one-dimensional distributions forms the separable three-dimensional velocity probability distribution $p(u_x, u_y, u_z)$.

Evaluating the integrals, we find

$$\begin{aligned}\overline{u_x} &= \left(\frac{m}{2\pi k_B T}\right)^{\frac{3}{2}} \int_{-\infty}^{\infty} u_x \mathrm{e}^{-\frac{m}{2k_B T}u_x^2}\, du_x \times \sqrt{\frac{2\pi k_B T}{m}} \times \sqrt{\frac{2\pi k_B T}{m}} \\ &= \left(\frac{m}{2\pi k_B T}\right)^{\frac{1}{2}} \int_{-\infty}^{\infty} u_x \mathrm{e}^{-\frac{m}{2k_B T}u_x^2}\, du_x\end{aligned}$$

where we have evaluated the integrals over u_y and u_z as we did in normalizing the distribution. We are left with an integral over u_x in which the integrand is an odd function of u_x. As such, we find

$$\overline{u_x} = \int_{-\infty}^{\infty} u_x \mathrm{e}^{-\frac{m}{2k_B T}u_x^2}\, du_x = 0$$

Repeating the process for the averages of u_y and u_z leads to $\overline{u_y} = \overline{u_z} = 0$. The final result is that the mean velocity $\overline{\mathbf{v}} = 0$.[17]

[17] This result makes physical sense. It is equally likely that the velocity will be oriented in the positive or negative direction in a given cartesian dimension. This makes the average velocity in each cartesian dimension zero.

What about the variance of the velocity? We can define the variance in terms of the mean and mean square values of $\mathbf{v}$ as

$$\sigma_v^2 = \overline{|\mathbf{v}|^2 - |\overline{\mathbf{v}}|^2} = \left(\overline{u_x^2} + \overline{u_y^2} + \overline{u_z^2}\right) - \left(\overline{u_x}^2 + \overline{u_y}^2 + \overline{u_z}^2\right)$$

Grouping terms we find

$$\begin{aligned}\sigma_v^2 &= \left(\overline{u_x^2} - \overline{u_x}^2\right) + \left(\overline{u_y^2} - \overline{u_y}^2\right) + \left(\overline{u_z^2} - \overline{u_z}^2\right) \\ &= \sigma_{u_x}^2 + \sigma_{u_y}^2 + \sigma_{u_z}^2\end{aligned}$$

The total variance in the distribution of velocities is the sum of the variances in the components of the velocities u_x, u_y, and u_z. Therefore, to determine σ_v^2 we should determine $\sigma_{u_x}^2$, $\sigma_{u_y}^2$, and $\sigma_{u_z}^2$. Noting that $\sigma_{u_x}^2 = \overline{u_x^2} - \overline{u_x}^2 = \overline{u_x^2}$, as we determined that $\overline{u_x} = 0$, we find

$$\begin{aligned}\overline{u_x^2} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u_x^2\, p(u_x, u_y, u_z)\, du_x du_y du_z \\ &= \left(\frac{m}{2\pi k_B T}\right)^{\frac{3}{2}} \int_{-\infty}^{\infty} u_x^2\, \mathrm{e}^{-\frac{mu_x^2}{2k_B T}} du_x \times \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{mu_y^2}{2k_B T}} du_y \times \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{mu_z^2}{2k_B T}} du_z \\ &= \left(\frac{m}{2\pi k_B T}\right)^{\frac{3}{2}} \int_{-\infty}^{\infty} u_x^2\, \mathrm{e}^{-\frac{m}{2k_B T}u_x^2} du_x \times \sqrt{\frac{2\pi k_B T}{m}} \times \sqrt{\frac{2\pi k_B T}{m}} \\ &= \left(\frac{m}{2\pi k_B T}\right)^{\frac{1}{2}} \int_{-\infty}^{\infty} u_x^2\, \mathrm{e}^{-\frac{m}{2k_B T}u_x^2}\, du_x\end{aligned}$$

The remaining integral can be performed by integration by parts or by using our differentiation with respect to a parameter trick explored in the complements to Chapter 7. The result is[18]

$$\overline{u_x^2} = \frac{k_B T}{m}\sqrt{\frac{2\pi k_B T}{m}}\sqrt{\frac{m}{2\pi k_B T}} = \frac{k_B T}{m}$$

[18] We use the identity

$$\begin{aligned}\int_{-\infty}^{\infty} x^2\, \mathrm{e}^{-\alpha x^2}\, dx &= -\frac{d}{d\alpha}\int_{-\infty}^{\infty} \mathrm{e}^{-\alpha x^2}\, dx \\ &= -\frac{d}{d\alpha}\sqrt{\frac{\pi}{\alpha}} = \sqrt{\frac{\pi}{4\alpha^3}}\end{aligned}$$

where $x = u_x$ and $\alpha = m/2k_B T$.

By symmetry, we recognize the result will be the same for $\overline{u_y^2}$ and $\overline{u_z^2}$. With

BRO0914
Mathematical methods for molecular science : theor
By: Straub, John Edward

this result, we can write

$$\sigma_v^2 = \sigma_{u_x}^2 + \sigma_{u_y}^2 + \sigma_{u_z}^2 = \frac{k_B T}{m} + \frac{k_B T}{m} + \frac{k_B T}{m} = \frac{3k_B T}{m}$$

for the variance in the velocity in three dimensions. These examples demonstrate the challenge of multiple integration over complicated functions resulting from models in the physical sciences. It is not that the integrals require great invention. The integral identities used in the derivation are familiar. The principal challenge comes from managing the complicated algebraic expressions without error.

9.3.3 *Properties of the electron density probability distribution*

The state of an electron in a one-electron atom is described in terms of a *wave function*. The square of the modulus of that function describes the electron's spatial distribution (or shape) in three-dimensional space.

Consider the form of the wave function for four possible states of an electron in a one-electron atom. These functions are most naturally expressed in spherical polar coordinates. For the electron in the lowest energy state, referred to as the 1*s* *ground state*, the wave function is

$$\psi_{1s}(r,\theta,\varphi) = N_1 \, \mathrm{e}^{-r/a_0}$$

where N_1 is a normalization constant, r is the distance of the electron from the
nucleus, and a_0 is known as the *Bohr radius*.

The wave functions for the electron in the four excited states are

$$\psi_{2s}(r,\theta,\varphi) = N_2 \left(2 - \frac{r}{a_0}\right) \mathrm{e}^{-r/2a_0}$$

$$\psi_{2p_x}(r,\theta,\varphi) = N_2 \left(\frac{r}{a_0}\right) \mathrm{e}^{-r/2a_0} \sin\theta\cos\varphi$$

$$\psi_{2p_y}(r,\theta,\varphi) = N_2 \left(\frac{r}{a_0}\right) \mathrm{e}^{-r/2a_0} \sin\theta\sin\varphi$$

$$\psi_{2p_z}(r,\theta,\varphi) = N_2 \left(\frac{r}{a_0}\right) \mathrm{e}^{-r/2a_0} \cos\theta$$

where N_2 is a normalization constant. Figure 9.18 depicts the radial dependence of the $\psi_{1s}(r,\theta,\varphi)$ and $\psi_{2s}(r,\theta,\varphi)$ wave functions.

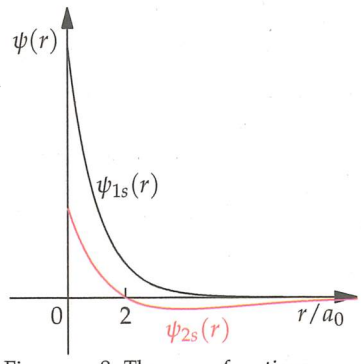


Figure 9.18: The wave functions $\psi_{1s}(r)$ (black line) and $\psi_{2s}(r)$ (red line) as a function of the distance expressed in reduced units r/a_0.

We can determine the normalization constants N_1 and N_2 by taking the square of the wave function and integrating the resulting electron distribution over all space. Consider the ψ_{1s} wave function where the square of the wave function is

$$|\psi_{1s}(r,\theta,\varphi)|^2 = \left(N_1 \, \mathrm{e}^{-r/a_0}\right)^2 = N_1^2 \, \mathrm{e}^{-2r/a_0}$$

The normalization constant can be determined from the definition

$$\int_0^\infty \int_0^\pi \int_0^{2\pi} |\psi_{1s}(r,\theta,\varphi)|^2 \; r^2 \sin\theta \, d\varphi d\theta dr = 1$$

where we chose to work in spherical polar coordinates reflecting the fact that

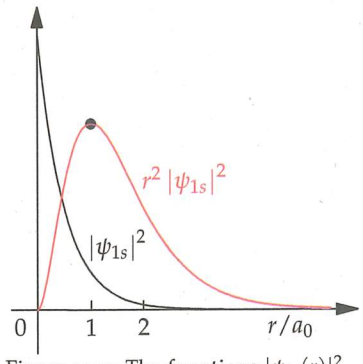


Figure 9.19: The functions $|\psi_{1s}(r)|^2$ (black line) and $r^2|\psi_{1s}(r)|^2$ (red line) over $r \in [0, 6a_0]$. The peak of the distribution occurs at the Bohr radius $r = a_0$ (black dot). The volume element $dV = r^2 \sin\theta \, dr d\theta d\varphi$ contributes a factor of r^2 to the overall radial dependence of the distribution.

the wave functions are natural functions of r, θ, and φ. It follows that

$$\int_0^\infty \int_0^\pi \int_0^{2\pi} N_1^2 \, e^{-2r/a_0} \, r^2 \sin\theta \, d\varphi d\theta dr = 1$$

The volume element for integration in spherical polar coordinates $dV = r^2 \sin\theta dr d\theta d\varphi$ contributes a factor of r^2 to the radial dependence of the integrand (see Figure 9.19).

This three-dimensional integral is separable in spherical polar coordinates and can be expressed in terms of the product of three one-dimensional integrals:

$$N_1^2 \int_0^\infty e^{-2r/a_0} \, r^2 \, dr \times \int_0^\pi \sin\theta \, d\theta \times \int_0^{2\pi} d\varphi = 1$$

We have encountered each of these integrals before. Integration over θ and φ results in a familiar factor of 4π:

$$\int_0^\pi \sin\theta \, d\theta \times \int_0^{2\pi} d\varphi = 2 \times 2\pi = 4\pi$$

This means that the original triple integral can be simplified to

$$4\pi N_1^2 \int_0^\infty e^{-2r/a_0} \, r^2 \, dr = 1$$

where the r^2 term scales $|\psi(r)|^2$ in proportion to the growing area to be integrated over as r increases (see Figure 9.19). The final integral over an exponential function can be evaluated using integration by parts or differentiation with respect to a parameter explored in the complements to Chapter 7. The result is

$$\int_0^\infty x^2 e^{-ax} dx = \frac{2}{a^3}$$

where $x = r$ and $a = 2/a_0$. The final result is

$$4\pi N_1^2 \left(\frac{a_0^3}{4}\right) = N_1^2 \pi a_0^3 = 1$$

so that $N_1 = (\pi a_0^3)^{-1/2}$. A similar calculation can be used to show that $N_2 = (32\pi a_0^3)^{-1/2}$ for each of these four excited states. Examples of the three-dimensional forms of the wave functions noted above are provided in Figure 9.20 and Figure 9.21.

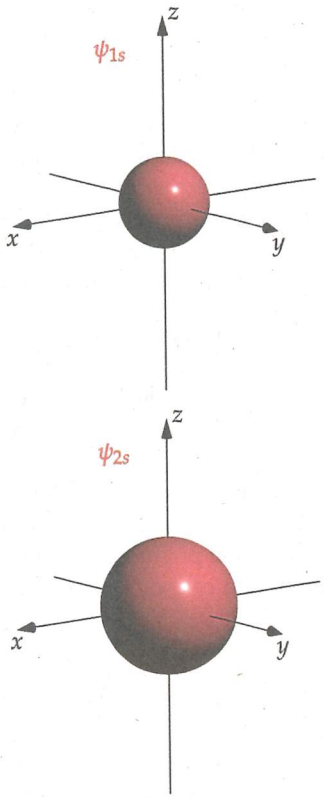


Figure 9.20: Surface of the modulus squared of the ψ_{1s} and ψ_{2s} wave functions for a one-electron atom.

9.3.4 *Mean and variance of the electron density probability distribution*

Let's compute the mean and variance of the position of the electron in the 1s ground state. The mean position of the electron is defined as the position r averaged over the electron density distribution $|\psi_{1s}(r,\theta,\varphi)|^2$ and is written

$$\begin{aligned}\bar{r} &= \int_0^\infty \int_0^\pi \int_0^{2\pi} r \, |\psi_{1s}(r,\theta,\varphi)|^2 \, r^2 \sin\theta \, d\varphi d\theta dr \\ &= \frac{1}{\pi a_0^3} \int_0^\infty \int_0^\pi \int_0^{2\pi} r^3 \, e^{-2r/a_0} \sin\theta \, d\varphi d\theta dr\end{aligned}$$

Mathematical methods for molecular science : theor
By: Straub, John Edward

This three-dimensional integral over all space is separable and can be written as a product of three one-dimensional integrals:

$$\overline{r} = \frac{1}{\pi a_0^3}\int_0^\infty r^3\, e^{-2r/a_0}\, dr \times \int_0^\pi \sin\theta\, d\theta \times \int_0^{2\pi} d\varphi$$
$$= 4\pi\frac{1}{\pi a_0^3}\int_0^\infty r^3\, e^{-2r/a_0}\, dr$$

The integral over r can be performed using integration by parts, but only with many steps. The most straightforward way to determine the integral is to use the identity

$$\int_0^\infty x^3 e^{-ax} dx = -\frac{d^3}{da^3}\int_0^\infty e^{-ax} dx = -\frac{d^3}{da^3}\frac{1}{a} = \frac{6}{a^4}$$

Applying this identity to our integral of interest with $x = r$ and $a = 2/a_0$ results in

$$\overline{r} = 4\pi\frac{1}{\pi a_0^3}\int_0^\infty r^3\, e^{-2r/a_0}\, dr = 4\frac{1}{a_0^3}\left(\frac{6a_0^4}{16}\right) = \frac{3}{2}a_0$$

We can use the same approach to determine the variance in the position of the electron, defined as

$$\sigma_r^2 = \overline{\overline{r^2} - \overline{r}^2}$$

We determine $\overline{r^2}$ to be

$$\overline{r^2} = \int_0^\infty\int_0^\pi\int_0^{2\pi} r^2\, |\psi_{1s}(r,\theta,\varphi)|^2\, r^2 \sin\theta\, d\varphi d\theta dr$$
$$= \frac{1}{\pi a_0^3}\int_0^\infty\int_0^\pi\int_0^{2\pi} r^4\, e^{-2r/a_0} \sin\theta\, d\varphi d\theta dr$$
$$= 4\pi\frac{1}{\pi a_0^3}\int_0^\infty r^4\, e^{-2r/a_0}\, dr$$

Using the integral identity

$$\int_0^\infty x^4 e^{-ax} dx = \frac{d^4}{da^4}\int_0^\infty e^{-ax} dx = \frac{d^4}{da^4}\frac{1}{a} = \frac{24}{a^5}$$

where $x = r$ and $a = 2/a_0$, leads to

$$\overline{r^2} = 4\pi\frac{1}{\pi a_0^3}\int_0^\infty r^4\, e^{-2r/a_0}\, dr = 4\frac{1}{a_0^3}\left(\frac{24a_0^5}{32}\right) = 3a_0^2$$

The final result for the variance in the position of the electron is

$$\sigma_r^2 = \overline{\overline{r^2} - \overline{r}^2} = 3a_0^2 - \left(\frac{3a_0}{2}\right)^2 = \frac{3}{4}a_0^2$$

We find that an apparently complicated set of three-dimensional integrals can be reduced to a series of one-dimensional integrations of familiar functions.

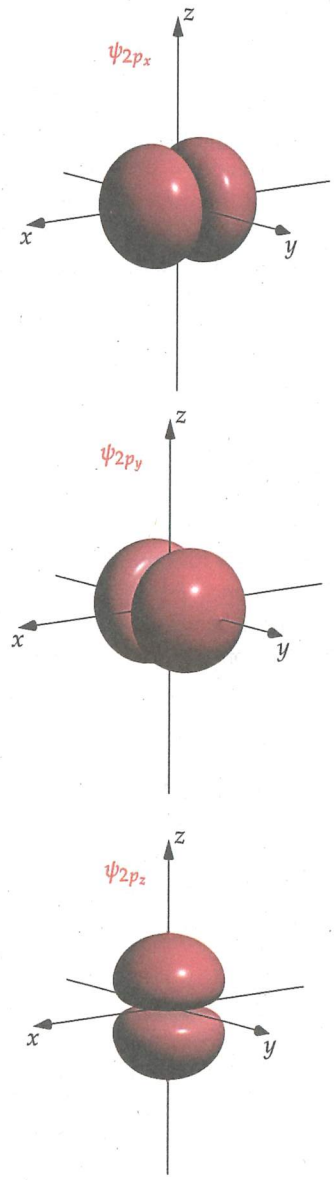


Figure 9.21: Surface of the modulus squared of the $\psi_{2p_x}(r,\theta,\varphi)$, $\psi_{2p_y}(r,\theta,\varphi)$, and $\psi_{2p_z}(r,\theta,\varphi)$ wave functions for a one-electron atom.

A9 Connecting the gaussian and binomial probability distributions

Let's examine the binomial probability distribution when the number of outcomes is $N = 20$. The distribution of probabilities is shown in Figure 9.22. Note that the mean of the distribution is $\overline{n} = \frac{N}{2} = 10$ and the variance is $\sigma_n^2 = \frac{N}{4} = 5$. Suppose we use these values of the mean and variance to parameterize a gaussian distribution. We find

$$p(n;N) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left[-\frac{(n-\overline{n})^2}{2\sigma_n^2}\right]$$

so that

$$p(n;N) = \frac{1}{\sqrt{\pi N/2}} \exp\left[-\frac{(n-N/2)^2}{N/2}\right] \tag{9.11}$$

The result is shown in Figure 9.22, where the continuous gaussian probability distribution function is compared with the discrete binomial probability distribution.

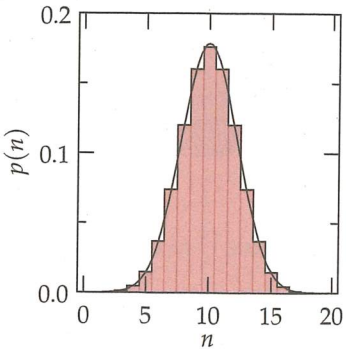


Figure 9.22: The binomial distribution for the relative probability of 21 possible outcomes resulting from $N = 20$ flips of a coin. Shown for comparison is the gaussian distribution.

For the binomial distribution, we found that the ratio of the width of the distribution σ_n relative to the mean $\overline{n}$ varies as

$$\frac{\sigma_n}{\overline{n}} = \frac{\sqrt{N}}{N} = \frac{1}{\sqrt{N}}$$

As N increases the mean increases according to $\overline{n} = \frac{N}{2}$ and standard deviation increases according to $\sigma_n = \frac{\sqrt{N}}{2}$. As N increases, the distribution appears to be more and more narrow. Another way to interpret this is to think of σ_n as the uncertainty in the measurement of the mean $\overline{n}$, where we might report our measurement as

$$\overline{n} \pm \sigma_n$$

The relative uncertainty, σ_n, in the measurement of the mean, $\overline{n}$, decreases with increasing N according to $N^{-1/2}$. For the distribution in Figure 9.2 with $N = 3$, we find $\frac{\sigma_n}{\overline{n}} = 0.577$, for the distribution in Figure 9.3 with $N = 6$, we find $\frac{\sigma_n}{\overline{n}} = 0.408$, and for the distribution in Figure 9.13 with $N = 20$, we find $\frac{\sigma_n}{\overline{n}} = 0.224$. As N increases, the ratio $\frac{\sigma_n}{\overline{n}}$ decreases, making the variance a smaller and smaller fraction of the mean. This demonstrates the increasing certainty in our measurement of the mean with increasing N.

This point is made more clearly in Figure 9.23 in which the gaussian approximation of the binomial distribution provided by Equation 9.11 is used to compare $Np(n;N)$ for varying $N = 10$, 50, and 150. Both the mean and standard deviation increase with N. The mean increases according to

$$\overline{n} \propto N$$

while the standard deviation increases more slowly according to

$$\sigma \propto \sqrt{N}$$

BRO0914

Mathematical methods for molecular science : theor
By: Straub, John Edward

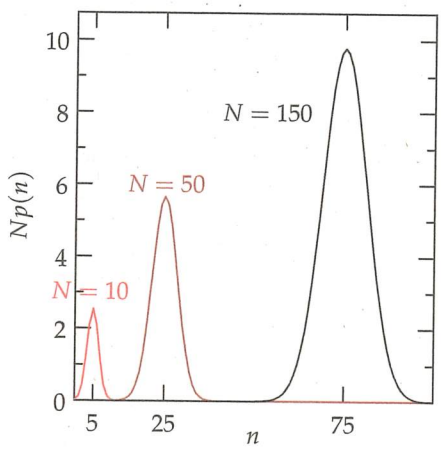


Figure 9.23: The gaussian approximation of the binomial distribution $p(n;N)$ for $N = 10, 50$, and 150.

As N increases, the probability becomes more narrowly distributed about the mean. This relative narrowing is best exposed in Figure 9.24 in which the distributions are plotted as a function of the normalized index $\frac{n}{N}$ for $N = 10, 50$, and 150. As N increases, the mean increases according to $\overline{n} = \frac{N}{2}$ and standard deviation increases according to $\sigma_n = \frac{\sqrt{N}}{2}$.

Finally, let's consider the random sampling of an *independent random variable*. We assume that when we sample the variable once, the outcome of the next sampling is independent of the prior result. We make a series of N observations from which we compute the mean. As N is increased, the normalized sum of the measured averages tends to a gaussian distribution. Remarkably this is true even if the variables themselves (such as the coin flips) are not distributed according to a gaussian distribution. This important result is known as the *central limit theorem*.[19] Let's return to the coin flip. Suppose we flip a coin many times. As the number of flips is increased, the probability of observing a certain number of heads in a series of flips will approach a gaussian probability distribution also referred to as the *normal distribution*. This remarkable result explains the ubiquitous use of the normal distribution in modeling phenomena in the physical and social sciences.

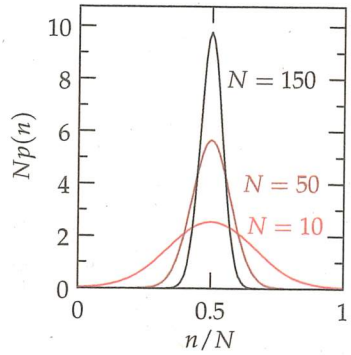


Figure 9.24: The gaussian approximation of the binomial distribution $p(n;N)$ for N, including 10, 50, and 150.

B9 Uniform distributions of independent random variables

Flipping a coin provides a means of generating independent *random numbers*. For example, suppose we take a head to be 0 and a tail to be 1. The outcome of $N = 20$ random coin flips can be written

00110 11110 11101 00101

forming a series of independent random variables. In this case there are 8 heads and 12 tails, an outcome with probability $p(n;N) = p(8;20) = 0.120$, or 12 percent. Other physical methods for generating random variables include choosing numbered balls from a bin or measuring decay events from a source of radioactive matter.

[19] This observation and its application to the study of coin flips was first reported by French mathematician *Abraham de Moivre* (1667-1754).

Now suppose we want to generate 10,000 independent random variables. We won't want to do that by flipping a coin! The challenge of generating independent random variables using mathematics alone is not at all trivial and has captured the imagination of mathematicians for decades. Consider the *middle square method*.[20] One takes a large odd integer, squares it, and then selects the middle set of numbers as the random variable.

For example, we start with the *random seed* 579,203, square it to form 335,476,115,209, and select the middle digits to arrive at 476,115. Repeating the process leads to 685,493. Repeating again yields 900,653 and 175,826 and 914,782. And so on. We can visualize this process as

$$
\begin{gathered}
(579,203)^2 \\
335,476,115,209 \\
(476,115)^2 \\
226,685,493,225 \\
(685,493)^2 \\
469,900,653,049 \\
(900,653)^2 \\
811,175,826,409 \\
(175,826)^2 \\
030,914,782,276 \\
914,782
\end{gathered}
$$

Note that the last operation produced an 11-digit number. We created a 12-digit number by padding the left side with a 0.

This method is not perfect. It has a relatively short *recurrence time* after which the series repeats itself. It can also converge to zero. As such, we say that the series is formed of *pseudorandom numbers*. If we wish to have the generated series of pseudorandom numbers occur over the interval $[0,1]$, we simply divide by one million.

[20] This method was developed by Hungarian-American mathematician and physicist *John von Neumann* (1903-1957). Von Neumann invented and built the first modern computer, developed *game theory* into a powerful tool in the social sciences, and made critical contributions to the development of the first atomic bomb.

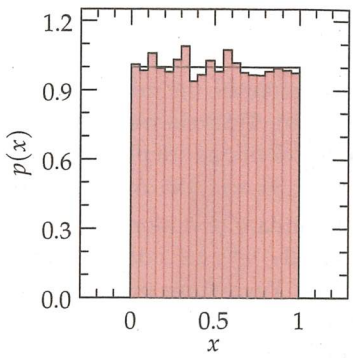


Figure 9.25: A distribution derived from 10,000 independent uniform random variables $x \in [0,1]$ (shaded red). The black straight line is the exact uniform distribution.

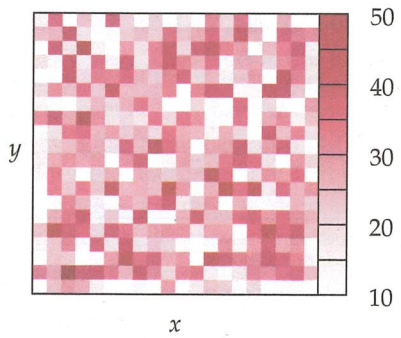


Figure 9.26: A two-dimensional probability distribution $p(x,y)$ for a sample of 10,000 independent uniform random variables x and y (shaded red). The color bar shows the number of observations of a pair (x,y) in a given region.

Suppose we have a process for generating random numbers $x \in [0,1]$ from a *uniform probability distribution*. The result of sampling 10,000 independent random numbers $x \in [0,1]$ is shown in Figure 9.25 and compared with a uniform distribution. The randomly sampled numbers approximate the uniform distribution over the range of $x \in [0,1]$. The computed mean value $\overline{x} = 0.497$ is slightly smaller than the exact value $\overline{x} = 0.5$, and the computed

BRO0914

Mathematical methods for molecular science : theor
By: Straub, John Edward

standard deviation $\sigma_x = 0.2883$ is slightly smaller than the exact value $\sigma_x = 12^{-\frac{1}{2}} = 0.2887\ldots$ for $N = 10,000$.

We can use the same method to generate a two-dimensional uniform probability distribution of independent random variables $x \in [0,1]$ and $y \in [0,1]$. To do this we can use the middle square method to generate x and repeat the process to generate y. The result for sampling 10,000 independent random variables from a uniform distribution in x and y is shown in Figure 9.26.

Note the significant variation in the number of observations in a given increment $dx\,dy$ across the distribution, even for a sampling of 10,000 numbers. As the number of dimensions increases, the number of samples must increase to achieve the same level of convergence. These examples convey the challenges associated with generating random samples, and the substantial data set required to represent an adequately converged random distribution.

C9 Gaussian distributions of independent random variables

The middle square method provides a way to generate independent random variables corresponding to a uniform probability distribution. How can we sample independent random variables from a gaussian distribution? It turns out that there is a clever method for transforming independent random variables sampled from a uniform distribution to independent random variables sampled from a gaussian distribution.

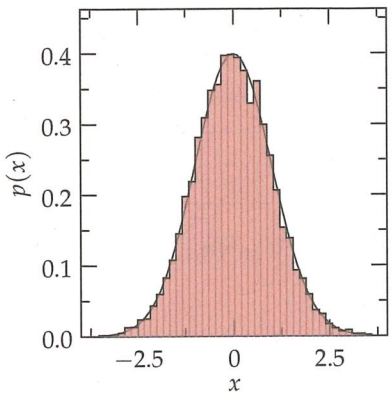


Figure 9.27: A distribution derived from 10,000 independent gaussian random variables (shaded red). The black line is the fit to a gaussian probability distribution function.

Suppose we generate independent random numbers uniformly over the interval $x \in [0,1]$. We generate two such numbers x_1 and x_2. We can transform these numbers from two independent random variables sampled from a uniform distribution, x_1 and x_2, into two independent random variables sampled from a gaussian distribution, z_1 and z_2, using the relations

$$z_1 = \sqrt{-2\ln x_1}\cos(2\pi x_2)$$
$$z_2 = \sqrt{-2\ln x_1}\sin(2\pi x_2)$$

This is known as the *Box-Mueller method*.[21]

[21] This method was originally presented by American mathematician and philosopher *Norbert Wiener* (1894-1964). Wiener was a famous child prodigy who made seminal contributions to the fields of cybernetics, robotics, and control systems.

Take two independent pseudorandom variables sampled from a uniform distribution in the interval $[0, 1]$:

$$x_1 = 0.579203 \qquad x_2 = 0.476115$$

Using the Box-Mueller method they can be transformed to two independent random variables sampled from a gaussian distribution:

$$z_1 = \sqrt{-2\ln(0.579203)}\cos\left[2\pi(0.476115)\right] = -0.680981$$
$$z_2 = \sqrt{-2\ln(0.579203)}\sin\left[2\pi(0.476115)\right] = 0.102972$$

Figure 9.27 shows the result of a sampling of 10,000 independent random variables compared with a gaussian probability distribution. We see that the Box-Mueller method provides a simple and effective way to generate independent random gaussian variables.

It is also possible to extend this method to sample a two-dimensional gaussian probability distribution in the variables x and y. Figure 9.28 shows a two-dimensional histogram over the xy-plane, which depicts the frequency of sampling two independent gaussian random variables, x and y, for a sample of 50,000 points. For a large sampling of points the distribution will be well approximated by a two-dimensional gaussian probability distribution function.

The ability to sample independent random variables from a gaussian distribution is useful to model physical processes such as the random motion of atoms in an ideal gas or the molecular diffusion of molecules in a liquid. It also provides a means of modeling statistical error. Ultimately, our success in generating independent gaussian random variables relies on our ability to effectively generate uniform random numbers.

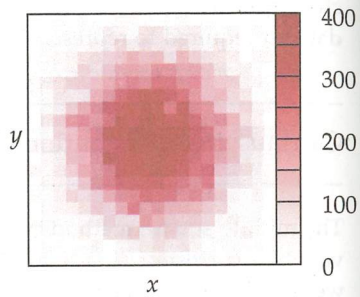


Figure 9.28: A two-dimensional probability distribution $p(x, y)$ for a sample of 50,000 independent gaussian random variables x and y (shaded red). The color bar shows the number of observations of a pair (x, y) in a given region.

D9 Three definitions of Pythagorean means

There are several ways to define the mean value of a series of numbers. The most familiar is the *arithmetic mean*, defined as

$$A_M = \frac{1}{n}\left(x_1 + \ldots + x_n\right)$$

This mean is useful for averaging *reaction times*. The arithmetic mean of a and b is $A_M = \frac{1}{2}(a + b)$.

However, we can also consider the *geometric mean*, defined as

$$G_M = \sqrt[n]{x_1 \times \ldots \times x_n}$$

The geometric mean of a and b is $G_M = \sqrt{ab}$. In contrast, the *harmonic mean* is defined as

$$H_M = n\left(\frac{1}{x_1} + \ldots + \frac{1}{x_n}\right)^{-1}$$

This mean is useful for averaging *reaction rates*. The harmonic mean of a and b is $H_M = 2(\frac{1}{a} + \frac{1}{b})^{-1} = 2\,\frac{ab}{a+b}$. The arithmetic mean, geometric mean, and

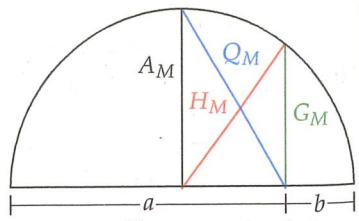


Figure 9.29: The geometric relationship between the arithmetic mean, A_M, geometric mean, G_M, harmonic mean, H_M, and quadratic mean, Q_M.

BRO0914

Mathematical methods for molecular science : theor
By: Straub, John Edward

harmonic mean are known as *Pythagorean means*.

In addition, we have the *quadratic mean*, also known as the *root mean square* value, and defined as

$$Q_M = \frac{1}{n}\sqrt{x_1^2 + \ldots + x_n^2}$$

This mean is useful in computing the root mean square speed. The quadratic mean of a and b is $Q_M = \frac{1}{2}\sqrt{a^2 + b^2}$.

The geometric relationships between these four definitions of the mean are displayed in Figure 9.29 for the case of $a = 0.8$ and $b = 1 - a = 0.2$, for which $A_M = 0.5$, $G_M = 0.4$, $H_M = 0.32$, and $Q_M = 0.41$. Note that only the case $a = b$ leads to the same mean value $A_M = G_M = H_M = Q_M$ for all four definitions of the mean.

E9 Propagation of error through total differentials and Taylor series

Suppose we have a function $f(x, y)$ of two independent variables x and y. Let's see how the error in our estimate of the function f is related to the known error in the averages of the variables x and y.

We define the mean value of the function, $\overline{f}$, and the mean values of the variables, $\overline{x}$ and $\overline{y}$. We further define the deviation of the observed values from the mean values as the errors

$$\varepsilon_x = x - \overline{x} \qquad \varepsilon_y = y - \overline{y} \qquad \varepsilon_f = f - \overline{f}$$

Since $f(x, y)$ is a function of x and y, we can ask how the error in our estimate of the function, ε_f, is related to the error in our measurements of the variables, ε_x and ε_y. We start from the Taylor series

$$\begin{aligned} f(x,y) = \overline{f} &+ \frac{\partial f}{\partial x}\Big|_{\overline{x},\overline{y}} (x - \overline{x}) + \frac{\partial f}{\partial y}\Big|_{\overline{x},\overline{y}} (y - \overline{y}) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}\Big|_{\overline{x},\overline{y}} (x - \overline{x})^2 \\ &+ \frac{1}{2}\frac{\partial^2 f}{\partial y^2}\Big|_{\overline{x},\overline{y}} (y - \overline{y})^2 + \frac{\partial^2 f}{\partial y \partial x}\Big|_{\overline{x},\overline{y}} (x - \overline{x})(y - \overline{y}) + \ldots \end{aligned}$$

which we can rewrite as

$$\begin{aligned} \varepsilon_f = f(x,y) - \overline{f} &= \frac{\partial f}{\partial x}\Big|_{\overline{x},\overline{y}} \varepsilon_x + \frac{\partial f}{\partial y}\Big|_{\overline{x},\overline{y}} \varepsilon_y + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}\Big|_{\overline{x},\overline{y}} \varepsilon_x^2 \\ &+ \frac{1}{2}\frac{\partial^2 f}{\partial y^2}\Big|_{\overline{x},\overline{y}} \varepsilon_y^2 + \frac{\partial^2 f}{\partial y \partial x}\Big|_{\overline{x},\overline{y}} \varepsilon_x \varepsilon_y + \ldots \end{aligned}$$

If the errors ε_x and ε_y are small enough, we can truncate the series and ignore higher order terms to find

$$\varepsilon_f \approx \frac{\partial f}{\partial x}\Big|_{\overline{x},\overline{y}} \varepsilon_x + \frac{\partial f}{\partial y}\Big|_{\overline{x},\overline{y}} \varepsilon_y$$

If the errors ε_x and ε_y are large, we need to retain higher order terms in the Taylor series to determine an accurate measure of the error in our estimate of f. We want to relate the mean square error in our estimate of the function

$f(x,y)$, defined as

$$\overline{\varepsilon_f^2} = \overline{(f-\overline{f})^2}$$

to the mean square error in our measurements of the variables x and y, defined $\overline{\varepsilon_x^2}$ and $\overline{\varepsilon_y^2}$.[22] Using the Taylor series approximation above, we find

$$\begin{aligned}\varepsilon_f^2 &\approx \left(\frac{\partial f}{\partial x}\Big|_{\overline{x},\overline{y}} \varepsilon_x + \frac{\partial f}{\partial y}\Big|_{\overline{x},\overline{y}} \varepsilon_y\right)^2 \\ &= \left(\frac{\partial f}{\partial x}\right)^2_{\overline{x},\overline{y}} \varepsilon_x^2 + \left(\frac{\partial f}{\partial y}\right)^2_{\overline{x},\overline{y}} \varepsilon_y^2 + 2\left(\frac{\partial f}{\partial x}\frac{\partial f}{\partial y}\right)_{\overline{x},\overline{y}} \varepsilon_x\varepsilon_y\end{aligned}$$

Now we must average over each side of the equation.

On the left-hand side, ε_f^2 becomes $\overline{\varepsilon_f^2}$. On the right-hand side, the partial derivatives are constant coefficients. We average the squared error so that ε_x^2 and ε_y^2 become $\overline{\varepsilon_x^2}$ and $\overline{\varepsilon_y^2}$. Finally, we note that since x and y are independent variables, $\overline{\varepsilon_x\varepsilon_y} = 0$. This leads to our final result[23]

$$\overline{\varepsilon_f^2} \approx \left(\frac{\partial f}{\partial x}\right)^2_{\overline{x},\overline{y}} \overline{\varepsilon_x^2} + \left(\frac{\partial f}{\partial y}\right)^2_{\overline{x},\overline{y}} \overline{\varepsilon_y^2}$$

We interpret the mean square error in our estimate of f as the variance

$$\sigma_f^2 = \overline{\varepsilon_f^2} = \overline{(f-\overline{f})^2}$$

Similarly $\sigma_x^2 = \overline{\varepsilon_x^2}$ and $\sigma_y^2 = \overline{\varepsilon_y^2}$. We can now write our final result relating the mean square error in our measurement of the variables x and y to the mean square error in our estimate of the function f as

$$\sigma_f^2 \approx \left(\frac{\partial f}{\partial x}\right)^2_{\overline{x},\overline{y}} \sigma_x^2 + \left(\frac{\partial f}{\partial y}\right)^2_{\overline{x},\overline{y}} \sigma_y^2 \tag{9.12}$$

Let's apply this result to the measurement of the pressure of a gas. Suppose we measure the temperature, T, and volume, V, of the gas. Using the ideal gas law, the pressure of the gas can be written as a function of the temperature and volume:

$$p(T,V) = \frac{RT}{V}$$

Let's see how the error in our estimate of p is related to our error in the measurements of T and V. Taking $p \equiv f$, $T \equiv x$, and $V \equiv y$, the resulting expression for the mean square error in our estimate of the pressure σ_p^2 is

$$\sigma_p^2 \approx \left(\frac{\partial p}{\partial T}\right)^2_{\overline{T},\overline{V}} \sigma_T^2 + \left(\frac{\partial p}{\partial V}\right)^2_{\overline{T},\overline{V}} \sigma_V^2 \tag{9.13}$$

where σ_T^2 and σ_V^2 are the mean square errors in the temperature and volume, respectively. This distribution is depicted in Figure 9.30.

Suppose there is 20% random error in $\overline{T}$ and $\overline{V}$. Interpreting the random error as the root-mean square error or standard deviation, we can write

$$\sigma_T = 0.2\,\overline{T} \qquad \sigma_V = 0.2\,\overline{V}$$

[22] Note that $\overline{\varepsilon_f} = \overline{(f-\overline{f})} = \overline{f} - \overline{f} = 0$. To measure the error σ_f we must average the square of the error and take the square root

$$\sigma_f = \sqrt{\overline{\varepsilon_f^2}} = \sqrt{\overline{(f-\overline{f})^2}}$$

[23] Suppose we have a function of one variable $f(x)$. We define the mean values $\overline{f}$ and $\overline{x}$ and associated errors $\varepsilon_f = f - \overline{f}$ and $\varepsilon_x = x - \overline{x}$. Suppose we take N measurements, each uncorrelated in error. We expect that $\overline{\varepsilon_f^2} = N f_x^2\, \overline{\varepsilon_x^2}$, which we can also write as $\sigma_f^2 = N\sigma_x^2$. As such

$$\sigma_f \propto \sqrt{N}\sigma_x.$$

BRO0914

Mathematical methods for molecular science : theor
By: Straub, John Edward

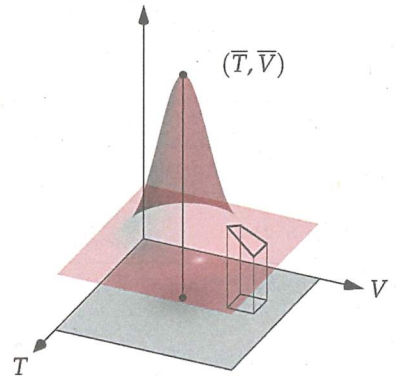


Figure 9.30: A two-dimensional probability distribution for the variables T and V. The distribution is centered around the average $(\overline{T}, \overline{V})$. The width of the distribution represents the standard deviations σ_T and σ_V in the T and V dimensions, respectively. The probability of observing a value of $T \in [T, T + dT]$ and $V \in [V, V + dV]$ is given by the volume of the vertical column over the area $dT\, dV$.

so that $\sigma_T^2 = (0.2\overline{T})^2$ and $\sigma_V^2 = (0.2\overline{V})^2$. We further know that

$$\frac{\partial p}{\partial T} = \frac{R}{V} \qquad \frac{\partial p}{\partial V} = -\frac{RT}{V^2}$$

Combining these results with Equation 9.13, we find

$$\begin{aligned}\sigma_p^2 &\approx \left(\frac{R}{\overline{V}}\right)^2 \times (0.2\overline{T})^2 + \left(-\frac{R\overline{T}}{\overline{V}^2}\right)^2 \times (0.2\overline{V})^2 \\ &= 0.08 \left(\frac{R\overline{T}}{\overline{V}}\right)^2 = 0.08\, p^2\end{aligned}$$

Let's look at one more example involving the average of small differences between large numbers. Consider the function

$$f(x, y) = x - y$$

where $\overline{x} \pm \sigma_x = 30 \pm 1$ and $\overline{y} \pm \sigma_y = 29 \pm 1$. We can return to our earlier result for the mean square error, defined as

$$\sigma_f^2 \approx \left(\frac{\partial f}{\partial x}\right)^2_{\overline{x},\overline{y}} \sigma_x^2 + \left(\frac{\partial f}{\partial y}\right)^2_{\overline{x},\overline{y}} \sigma_y^2$$

where

$$\frac{\partial f}{\partial x} = 1 \qquad \frac{\partial f}{\partial y} = -1$$

so that $\sigma_f^2 \approx \sigma_x^2 + \sigma_y^2 = 1 + 1 = 2$. It follows that

$$\sigma_f = \sqrt{2} = 1.41$$

From our original expression for $f(x, y)$, we find $\overline{f} = \overline{x} - \overline{y} = 30 - 29 = 1$ so that $\overline{f} \pm \sigma_f = 1 \pm 1.41$. The error is larger than the difference itself! This example demonstrates the importance of considering not only the relative error in variables, but how that relative error translates into the total error in the function itself.

F9 End-of-chapter problems

> Life's most important questions are, for the most part, nothing but probability problems.
>
> Pierre-Simon Laplace

Warm-ups

9.1 Calculate the probability of getting (a) exactly 5 heads in 10 tosses of a fair coin, (b) exactly 1 head in 5 tosses of a fair coin, and (c) 5 heads in the order of THHTHTTTHH.

HINT: A coin toss has two possible outcomes, head or tail. The binomial distribution gives the probability of observing n heads and $N-n$ tails in N flips of a fair coin as

$$p(n;N) = \frac{N!}{(N-n)!\,n!}\left(\frac{1}{2}\right)^N$$

where the binomial coefficient

$$\binom{N}{n} = \frac{N!}{(N-n)!\,n!}$$

is the number of ways that n heads and $N-n$ tails can be observed in N flips out of 2^N possible outcomes.

9.2 Determine the relative error in *Stirling's approximation*

$$\ln N! \cong N \ln N - N$$

in calculations of $\ln N!$ for $N = 10, 50$, and 100. The relative error is defined as the difference between the estimated value and the exact value, divided by the exact value.

9.3 In the card game *bridge* you can distribute 52 cards having 4 suits (hearts, diamonds, clubs, and spades) of 13 cards each. A bridge hand consists of 13 cards. What is the total number of possible bridge hands?

9.4 Find the number of different permutations for the letters in the words Euler, Laplace, and Lagrange.

HINT: An event has k possible outcomes. The multinomial coefficient gives the number of ways of observing k possible outcomes, defined as a set of $\{n_1, n_2, n_3, \ldots, n_k\}$ where n_l is the number of times the lth outcome was observed in N trials

$$\binom{N}{n_1, n_2, n_3, \cdots, n_k} = \frac{(n_1+n_2+n_3+\cdots n_k)!}{n_1!n_2!n_3!\cdots n_k!} = \frac{N!}{n_1!n_2!n_3!\cdots n_k!}$$

where

$$\sum_{l=1}^{k} n_l = N$$

9.5 Consider the simple uniform distribution

$$p(x) = \begin{cases} C & 0 \le x \le 5 \\ 0 & \text{otherwise} \end{cases}$$

BRO0914

Mathematical methods for molecular science : theor
By: Straub, John Edward

(a) If $p(x)$ is normalized, what is the normalization constant C?

(b) What is the mean $\overline{x}$?

(c) What is the standard deviation σ_x defined in terms of the variance $\sigma_x^2 = \overline{(x-\overline{x})^2}$?

9.6 Consider the exponential distribution

$$p(r)\,dr = Ce^{-\lambda r}dr \qquad 0 \le r \le \infty$$

(a) If $p(r)$ is normalized, what is the normalization constant C?

(b) What is the mean $\overline{r}$?

(c) What is standard deviation σ_r defined in terms of the variance $\sigma_r^2 = \overline{(r-\overline{r})^2}$

Homework exercises

9.7 Consider the probability distribution of the one-dimensional *particle in a box*

$$p(x)\,dx = |\psi_n(x)|^2\,dx = \frac{2}{L}\sin^2\left(\frac{n\pi x}{L}\right)dx \qquad n = 1,2,3,\ldots$$

where $0 \le x \le L$.

(a) Show that $p(x)$ is normalized.

(b) Show that the mean is

$$\overline{x} = \frac{L}{2}$$

(c) Show that the variance is

$$\overline{(x-\overline{x})^2} = \frac{L^2}{12}\left(1 - \frac{6}{n^2\pi^2}\right).$$

9.8 Consider the distribution of energies of an ideal gas

$$p(\varepsilon)\,d\varepsilon = \frac{2\pi}{(\pi k_B T)^{3/2}}\,\varepsilon^{1/2}\exp\left(-\frac{\varepsilon}{k_B T}\right)d\varepsilon$$

where $0 \le \varepsilon < \infty$.

(a) Show that $p(\varepsilon)$ is normalized.

(b) Determine the mean $\overline{\varepsilon}$. Compare your result to the average kinetic energy of an ideal gas particle

$$\frac{1}{2}m\overline{v^2} = \frac{3}{2}k_B T$$

(c) Show that the variance is

$$\sigma_\varepsilon^2 = \overline{(\varepsilon-\overline{\varepsilon})^2} = \frac{3}{2}(k_B T)^2$$

9.9 The mean value $\overline{n}$ and variance $\sigma_n^2 = \overline{(n-\overline{n})^2}$ of a binomial distribution are defined by Equations 9.9 and 9.10 to be

$$\overline{n} = \frac{N}{2} \qquad \sigma_n^2 = \overline{(n-\overline{n})^2} = \frac{N}{4}$$

when each outcome is assumed to have probability $\frac{1}{2}$. Consider the general case in which one outcome has probability p and the other outcome has probability $1-p=q$. The binomial probability of observing n events each having probability p and $N-n$ events each having probability $1-p=q$ takes the form

$$p(n;N) = \frac{N!}{n!(N-n)!}\, p^n\, (1-p)^{N-n} = \frac{N!}{n!(N-n)!}\, p^n\, q^{N-n}$$

(a) Show that the mean value is given by

$$\overline{n} = \sum_{n=0}^{N} np(n;N) = Np$$

which reduces to the result $\overline{n} = N/2$ for the case $p = q = \frac{1}{2}$.

(b) Show that the variance is given by

$$\sigma_n^2 = \overline{(n-\overline{n})^2} = Np(1-p) = Npq$$

which reduces to the result $\sigma_n^2 = \overline{(n-\overline{n})^2} = N/4$ for the case $p = q = \frac{1}{2}$.

9.10 Find the mean value

$$\overline{r} = 4\left(\frac{1}{a_0}\right)^3 \int_0^\infty \mathrm{e}^{-2r/a_0}\, r^3\, dr$$

for an electron in the 1s state of the hydrogen atom.

9.11 The probability distribution of a quantum mechanical particle in a three-dimensional rectangular box of volume $V = L_x L_y L_z$ can be written as

$$\begin{aligned} I &= \iiint |\psi(x,y,z)|^2\, dx\, dy\, dz \\ &= \int_0^{L_z}\int_0^{L_y}\int_0^{L_x} C\sin^2\left(\frac{n_x \pi x}{L_x}\right)\sin^2\left(\frac{n_y \pi y}{L_y}\right)\sin^2\left(\frac{n_z \pi z}{L_z}\right) dx\, dy\, dz = 1 \end{aligned}$$

where $n_x = 1,2,3,\cdots$, $n_y = 1,2,3,\cdots$, and $n_z = 1,2,3,\cdots$.

(a) Carry out the multiple integral to determine the normalization constant C.

(b) Determine the average position of the particle in the box

$$\overline{\mathbf{r}} = \iiint \mathbf{r}\, |\psi(x,y,z)|^2\, dx\, dy\, dz$$

where $\mathbf{r} = x\,\hat{\mathbf{x}} + y\,\hat{\mathbf{y}} + z\,\hat{\mathbf{z}}$.

9.12 Consider the Maxwell distribution of velocities defined as

$$p(v)dv = C\exp\left[-\frac{\beta m}{2}\left(u_x^2 + u_y^2 + u_z^2\right)\right] du_x\, du_y\, du_z$$

where $\beta = \dfrac{1}{k_B T}$, the velocity $\boldsymbol{v} = u_x\,\mathbf{i} + u_y\,\mathbf{j} + u_z\,\mathbf{k}$, and the speed u is defined as the magnitude of the velocity

$$u = |\mathbf{v}| = \sqrt{u_x^2 + u_y^2 + u_z^2}$$

BRO0914

Mathematical methods for molecular science : theor
By: Straub, John Edward

(a) Carry out the following integral

$$I = \iiint p(v)dv = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(u_x, u_y, u_z)\, du_x\, du_y\, du_z = 1$$

to determine the normalization constant C.

(b) Show that the distribution in (a) may be written in spherical polar coordinates as

$$p(v)dv = C \exp\left(-\frac{\beta m}{2}u^2\right) u^2 \sin\theta\, du\, d\theta\, d\varphi$$

where $u^2 = u_x^2 + u_y^2 + u_z^2$.

(c) Determine the average speed defined by the integral

$$I = \iiint u\, p(v)dv = \int_0^{\infty}\int_0^{\pi}\int_0^{2\pi} C\, u \exp\left(-\frac{\beta m}{2}u^2\right) u^2 \sin\theta\, d\varphi\, d\theta\, du$$

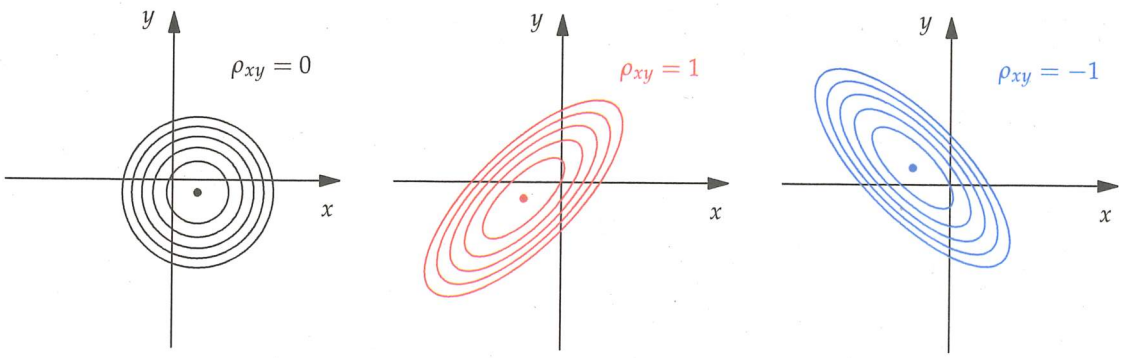
(d) Determine the mean square speed defined by the integral

$$I = \iiint u^2 p(v)dv = \int_0^{\infty}\int_0^{\pi}\int_0^{2\pi} C\, u^2 \exp\left(-\frac{\beta m}{2}u^2\right) u^2 \sin\theta\, d\varphi\, d\theta\, du$$

9.13 Consider a bivariate (two variable) probability distribution $p(x, y)$ with variances $\sigma_x^2 = \overline{(x-\overline{x})^2}$ and $\sigma_y^2 = \overline{(y-\overline{y})^2}$. The cross correlation between x and y can be measured using

$$\rho_{xy} = \frac{1}{\sigma_x \sigma_y}\overline{(x-\overline{x})(y-\overline{y})}$$

where ρ_{xy} is defined as the correlation coefficient. The distribution is shown in the figure below for three values of the correlation coefficient ρ_{xy}.



Assume that x and y are independent random variables with probability distribution given by

$$p(x, y)\, dx\, dy = \frac{\alpha}{\pi} e^{-\alpha\left[(x-\overline{x})^2 + (y-\overline{y})^2\right]}\, dx\, dy$$

Show that $\rho_{xy} = 0$.

9.14 Consider the bivariate normal probability distribution

$$p(x,y)\,dx\,dy = \frac{1}{2\pi\sqrt{1-\rho_{xy}^2}}\exp\left[-\frac{1}{2(1-\rho_{xy}^2)}(x^2+y^2-2\rho_{xy}\,xy)\right]dx\,dy$$

The distribution is shown in the figure above for three values of the correlation coefficient ρ_{xy}.

(a) Show that the *marginal probability* distributions take the form

$$p(x) = \int_{-\infty}^{\infty} p(x,y)\,dy = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$$

and

$$p(y) = \int_{-\infty}^{\infty} p(x,y)\,dx = \frac{1}{\sqrt{2\pi}}e^{-y^2/2}$$

by integrating over x (or y) leading to a normal distribution in y (or x).

(b) Perform the integral

$$I = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\,dx\,dy$$

and demonstate that the distribution is normalized.

9.15* The volume of a cylindrical capillary tube is given by the expression $V = \pi r^2 h$, where r is the radius of the tube and h is the length of the tube. If the radius of the capillary tube is found to be 0.030 cm with a error of ± 0.002 cm and the length of the tube is found to be 4.0 cm with a error of ± 0.1 cm, what is the volume of the capillary tube and its error? What measurement must be made to a higher precision to decrease the error in the volume?

9.16* The molar mass of a vapor is determined by filling a bulb of known volume with the vapor at a known temperature and pressure and measuring the mass of the bulb. The vapor is assumed to be an ideal gas so that molar mass, M, is given by

$$M = \frac{mRT}{pV}$$

where m is the mass of the vapor, R is the gas constant, T is the absolute temperature, p is the pressure, and V is the volume. Measurements are summarized as $m = 1.0339 \pm 0.0007$ g, $T = 274.0 \pm 0.5$ K, $p = 1.050 \pm 0.001$ bar, $V = 0.1993 \pm 0.0001$ liters, and $R = 0.08314\,L \cdot \text{bar} \cdot \text{mol}^{-1} \cdot \text{K}^{-1}$ (no error as the value is taken to be exact). Determine the error in the molar mass M.

9.17* Suppose you measure independent variables x and y. You have a function $f(x,y)$ that depends on those variables. Average quantities are defined $\overline{x}$, $\overline{y}$, and $\overline{f}$. We can further define the error in x, y, and $f(x,y)$ as

$$\varepsilon_x = x - \overline{x} \qquad \varepsilon_y = y - \overline{y} \qquad \varepsilon_f = f - \overline{f}$$

(a) Use a Taylor series expansion to express the measured error in $f(x,y)$, ε_f, as a function of errors ε_x and ε_y in x and y. HINT: Expand $f(x,y)$ about the point $(\overline{x},\overline{y})$ to first order in $x-\overline{x}$ and $y-\overline{y}$.

(b) Derive a formula for the mean square error $\overline{\varepsilon_f^2}$ as a function of $\overline{\varepsilon_x^2}$ and $\overline{\varepsilon_y^2}$.

(c) Consider the ideal gas law with $f \equiv p = \dfrac{RT}{V}$, $x \equiv T$, and $y \equiv V$. If T and V have 10% errors, how large is the error in p?

(d) You add N measurements with uncorrelated error in x, ε_x. How will the error in ε_f increase with increasing N?