

Chapter 3

Gaussian Distributions

3.1 Introduction

The simplest and most symmetrical problem in statistical mechanics uses a variable, like the classical momentum, which runs over all values in the interval between minus infinity and infinity. In this chapter, we use the symbol X for such a variable and $X_j (j = 1, 2, \dots, N)$ for N such variables. We might expect that such a variable has so much symmetry that it has the potential for producing situations which are both particularly tractable and also particularly robust. By robust one means that the basic problems or their solutions remain unchanged under a wide class of transformations. We start here by considering the probabilistic behavior with one Gaussian variable, X , and move on to the consideration of many such variables.

In one sense the discussion of this chapter is a natural outgrowth of the last one. There we were interested in the comparison between the structure of problems of the statistical mechanics of one particle and the ones with many non-interacting particles. With Gaussian variables, if the probability distribution is proportional to the exponential of a quadratic form in X (or the X_j 's) then the problem may once more be solved exactly as if it were a problem of non-interacting systems.

3.2 One Variable

A Gaussian probability distribution for a single random variable X is one in which the probability of finding X to lie between x and $x + dx$ has the form¹:

$$\frac{\exp[-\beta(x - \langle X \rangle)^2/2]}{\sqrt{2\pi/\beta}} dx. \quad (3.1)$$

This distribution contains two parameters, $\langle X \rangle$, which measures the center, or mean of the distribution and β , which measures its squared width. This squared width is called the

¹The word Gaussian is also used to describe an integral in which the integrand is the exponential of a quadratic form, i.e. $\int_{-\infty}^{\infty} dx \exp(-ax^2 + bx + c)$.

variance. As we have already seen, β is the inverse of the variance in the distribution, so that

$$\beta^{-1} = \langle (X - \langle X \rangle)^2 \rangle. \quad (3.2)$$

Any probability distribution in one variable, x , which varies over the entire real line and which is of the form of an exponential of a function quadratic in x

$$\rho \sim \exp(-ax^2 + bx + c)$$

may be cast into the structure shown on the right hand side of Eq. (3.1). (See Problem (3.1).)

Our work in describing statistical mechanics has relied quite heavily upon understanding the behavior of exponential functions of the basic variables. We continue along this track by analyzing the so-called *generating function*, $\langle \exp(iqX) \rangle$, with i being the square root of -1 and q being real. In this context, a generating function is the name for an average which is used to generate (i.e. *compute*) other averages. For the Gaussian case, our generating function is the integral

$$\langle \exp(iqX) \rangle = \int_{-\infty}^{\infty} dx \frac{\exp(-\beta(x - \langle X \rangle)^2/2)}{\sqrt{2\pi/\beta}} \exp(iqx)$$

which is easily evaluated to get the result:

For any Gaussian probability distribution with variance β^{-1} , the average of an exponential takes the form

$$\langle \exp(iqX) \rangle = \exp\left(iq\langle X \rangle - \frac{q^2}{2\beta}\right). \quad (3.3)$$

Conversely, if some variable, X , has a probability distribution for which

$$\langle \exp(iqX) \rangle = \exp\left[iqa - \frac{q^2}{2\beta}\right] \quad (3.4)$$

for all values of q , then the probability distribution for X must be Gaussian and a and β respectively have the interpretation of the mean and the inverse variance of the Gaussian.

These two statements look like very special results. Why should they be interesting or important?

One reason is that the Gaussian distribution exhibits a remarkable tenacity or stability, or what we call a peculiar robustness. Imagine that the variable X is of the form $X = aY + b$ where a and b are constants and Y is a Gaussian variable. Equation (3.3) then has the consequence that X is also a Gaussian random variable. Thus a linear function of a Gaussian random variable is also a Gaussian random variable. Next consider an X of the form

$$X = aY + bZ, \quad (3.5a)$$

where a and b are constants and Y and Z are independent Gaussian random variables. By independent it is meant that the fluctuations in the two variables are uncorrelated or *independent*. The mathematical definition of independence use the joint probability $\rho(z, y)dz dy$. This expression gives the probability for finding Z between z and $z + dz$ and, at the same time, finding Y between y and $y + dy$. Independence is that statement that the joint probability distribution $\rho(z, y)$ is a product of the individual probability distributions for Z and Y .

Thus it follows at once from Eq. (3.3) and the meaning of the word independent that X is also a Gaussian random variable. If you add up many independent Gaussian variables you end up with a new Gaussian variable. Thus, if

$$X = \sum_{j=1}^N a_j Y_j, \quad (3.5b)$$

where the a_j are constants and the Y_j are independent Gaussian random variables then the resulting X is certainly a Gaussian random variable. Consequently, a sum of very many Gaussian physical effects (represented say by the Y 's) is a net Gaussian effect.

3.3 Many Gaussian Variables

To describe the effect of many Gaussian Variables, we generalize Eqs. (3.5a) and (3.5b): $X_1, X_2, \dots, X_j, \dots, X_N$ are said to be a set of *Gaussian variables* if *the expectation value of an exponential formed from a linear combination of these variables is an exponential of a quadratic form in the coefficients*. The words are a mouthful, the equation is simple. For any vector \mathbf{q} with components $q_j (j = 1, 2, \dots, N)$ we have

$$\left\langle \exp \left(i \sum_{j=1}^N q_j X_j \right) \right\rangle = \exp \left(i \sum_{j=1}^N q_j \langle X_j \rangle - \sum_{j,k=1}^N q_j G_{jk} \frac{q_k}{2} \right). \quad (3.6)$$

Thus, the quadratic form defines this many-variable Gaussian.² Naturally, we shall usually write Eq. (3.6) in matrix-vector notation as

$$\langle \exp(i\mathbf{q} \cdot \mathbf{X}) \rangle = \exp \left(i\mathbf{q} \cdot \langle \mathbf{X} \rangle - \mathbf{q} \cdot G \cdot \frac{\mathbf{q}}{2} \right). \quad (3.7)$$

Of course $\langle \mathbf{X} \rangle$ means the average of the vector \mathbf{X} and G is related to a variance or correlation matrix for these variables. In fact, a second order power series expansion (in \mathbf{q}) of Eq. (3.7) implies:

$$\langle (X_j - \langle X_j \rangle)(X_k - \langle X_k \rangle) \rangle = G_{jk}. \quad (3.8)$$

²Analytic continuation enables us to extend this results to complex values of the q 's.

For the definition (3.6) to make sense, we require that G be a positive definite symmetric matrix.³ Often, G is called a green function.

What probability distribution could give rise to averages of the form (3.6)? The reader will not be surprised to hear that the defining probability distribution must be an exponential of a quadratic form. Specifically, I ask you to prove in a homework exercise that if

$$\rho(x_1, x_2, \dots, x_N) dx_1 dx_2 \cdots dx_N$$

is defined to be the probability that X_1 have a value between x_1 and $x_1 + dx_1$ while X_2 has a value between x_2 and $x_2 + dx_2, \dots$ then from (3.6) it follows that:

$$\rho(x_1, x_2, \dots, x_N) = (2\pi)^{-N/2} (\det C)^{1/2} e^{-[(x - \langle X \rangle) C (x - \langle X \rangle)]/2} \quad (3.9)$$

Here C is a symmetric matrix which is the inverse of G , namely

$$\sum_{m=1}^N C_{jm} G_{mk} = \sum_{m=1}^N G_{jm} C_{mk} = \delta_{j,k} \quad (3.10)$$

or in matrix form

$$CG = GC = 1. \quad (3.11)$$

The determinant of C has a value which is the product of all the eigenvalues of C .

The key results in obtaining the solution (3.10) are Eqs. (3.6) and (3.9). That solution makes the many-variable Gaussian case look just like the one-variable case. Imagine doing a change of variables from the X 's to the appropriate linear combination of X 's which serves to diagonalize the matrix C . Then G is simultaneously diagonalized. After this replacement the probability density is an exponential of a sum of terms, each one quadratic in one of the X -variables. Then, Eq. (3.6) gives the total probability as a product of probabilities for the individual linear combinations. Thereby the entire problem is reduced to a combination of subproblems for the individual and non-correlated linear combinations.

We have defined Gaussian correlations by using a generating function approach — and G appears in the solution for the generating function. Conversely, Eq. (3.9) shows that the matrix inverse of G , called C , appears in the basic problem definition. Thus to go from problem to solution, one need only invert a matrix.

³Any anti-symmetric part of G will contribute nothing to the right hand side of Eq. (3.7). Hence we do not lose any generality by demanding that G is a symmetric matrix. Any symmetric real matrix is Hermitian. One can describe it by listing its eigenvalues, all of which must be real. If they are all positive then the matrix is said to be positive definite. However, if one eigenvalue is negative, then one can set up a Y as a linear combination of X_j corresponding to the eigen direction. A negative eigenvalue of G corresponds to a negative value of $\langle (Y - \langle Y \rangle)^2 \rangle$. This negativity is quite impossible for real Y . Consequently, only positive eigenvalues are possible in a formulation based upon Eq. (3.8).

An alternate formulation is possible, based upon Gaussian integrals. In that case the positivity is a condition for the convergence of the integrals.

3.4 Lattice Green Function

Suppose we have a lattice with lattice sites \mathbf{r}_n , and one variable, X_n , for each site, then we can define a particularly simple and important problem by giving the coefficient matrix

$$C_{n,m} = \begin{cases} 1 & \text{if } \mathbf{r}_n = \mathbf{r}_m \\ -K & \text{if } \mathbf{r}_n \text{ and } \mathbf{r}_m \text{ are nearest neighbors.} \end{cases} \quad (3.12)$$

In this problem definition the on-site interaction serves to normalize the Gaussian variables while the coupling K serves to define the strength of the interaction between different lattice sites. As we shall see this simple and exactly solvable problem is in fact closely related to several different situations involving phase transitions. The Gaussian problem itself undergoes a kind of phase transition at a point at which one of the eigenvalues of C approaches zero. When that happens, the correlation matrix G goes to infinity, and very large correlations tend to develop in the system. Some thermodynamic derivatives for the system become very large, and the system shows every sign of doing something interesting. We shall explore this interesting behavior in considerable detail below.

In the case defined by Eq. (3.12), and in many other situations describing a problem with a high degree of symmetry,⁴ one can relate the matrix and its inverse by using some sort of Fourier transform. In such situations, one can often obtain a full solution for the inverse.

For example if the lattice sites are set to be integers on a simple cubic lattice in dimension d , with a nearest neighbor separation, a :

$$\mathbf{r}_n = a(n_1, n_2, \dots, n_\mu, \dots, n_d), \quad (3.13)$$

then the natural representation of the C 's and G 's is as integrals over wave vector, each component going from $-\pi/a$ to π/a . Thus

$$C_{n,m} = \left(\frac{a}{2\pi}\right)^d \int_{-\pi/a}^{\pi/a} dq_1 \cdots \int_{-\pi/a}^{\pi/a} dq_\mu \cdots \int_{-\pi/a}^{\pi/a} dq_d C(\mathbf{q}) \exp[i\mathbf{q} \cdot (\mathbf{r}_n - \mathbf{r}_m)], \quad (3.14a)$$

$$G_{n,m} = \left(\frac{a}{2\pi}\right)^d \int_{-\pi/a}^{\pi/a} dq_1 \cdots \int_{-\pi/a}^{\pi/a} dq_\mu \cdots \int_{-\pi/a}^{\pi/a} dq_d G(\mathbf{q}) \exp[i\mathbf{q} \cdot (\mathbf{r}_n - \mathbf{r}_m)]. \quad (3.14b)$$

We say that \mathbf{q} 's which lie within the range of integration in expression (3.14a) lie in the first Brillouin zone. We then write this same multiple integral in an abbreviated notation as, for example,

$$G_{n,m} = \left(\frac{a}{2\pi}\right)^s \int_{-\pi/a}^{\pi/a} d^d \mathbf{q} G(\mathbf{q}) \exp[i\mathbf{q} \cdot (\mathbf{r}_n - \mathbf{r}_m)]. \quad (3.15)$$

⁴Here, of course, the symmetry is an invariance is under spatial translations.

Note that the variables of integration must sit in the right zone. In view of the defining properties of Fourier transforms we have

$$\left(\frac{a}{2\pi}\right)^d \int_{-\pi/a}^{\pi/a} d^d \mathbf{q} \exp[i\mathbf{q} \cdot \mathbf{r}_n] = \delta_{\mathbf{r}_n, \mathbf{0}}, \quad (3.16a)$$

whenever \mathbf{r}_n is a difference between two lattice vectors. The inverse transform is derived from the relation

$$\sum_n \exp[i\mathbf{q} \cdot \mathbf{r}_n] = \delta(\mathbf{q}) \left(\frac{a}{2\pi}\right)^{-d}, \quad (3.16b)$$

whenever \mathbf{q} is in the first Brillouin zone. Here, $\delta(\mathbf{q})$ is a abbreviation for the product of delta functions

$$\delta(\mathbf{q}) = \prod_{\mu=1}^d \delta(q_\mu). \quad (3.17)$$

It is easy to calculate G . Working from Eqs. (3.12), (3.16a) and (3.16b)), we find that the Gaussian probability is defined by the nearest neighbor interaction which has the Fourier transform

$$C(\mathbf{q}) = 1 - 2K \sum_{\mu=1}^d \cos(q_\mu a). \quad (3.18)$$

As a consequence of Eq. (3.10) the Green function, $G(q)$ has a Fourier transform which obeys $C(\mathbf{q})G(\mathbf{q}) = 1$, so that

$$G(\mathbf{q}) = \frac{1}{1 - 2K \sum_{\mu=1}^d \cos(q_\mu a)}. \quad (3.19)$$

Thus the final result is the coordinate space expression

$$G_{n,m} = \left(\frac{a}{2\pi}\right)^d \int_{-\pi/a}^{\pi/a} d^d \mathbf{q} \frac{\exp[i\mathbf{q} \cdot (\mathbf{r}_n - \mathbf{r}_m)]}{1 - 2K \sum_{\mu=1}^d \cos(q_\mu a)}. \quad (3.20)$$

The quantities $G_{n,m}$ and $C_{n,m}$ can be considered to be matrices in some big space labeled by all the possible values of n . The transformation to Fourier variables \mathbf{q} is essentially a diagonalization of these matrices to a situation in which the diagonal element is labeled by \mathbf{q} and the diagonal elements are respectively $G(\mathbf{q})$ and $C(\mathbf{q})$. We require that these diagonal elements be positive in order that the whole calculation make any sense whatsoever. If K is positive, the most dangerous place is $\mathbf{q} = 0$. For this value of \mathbf{q} we have the minimum value of C , namely

$$C(0) = 1 - 2Kd. \quad (3.21)$$

Thus we might expect a kind of change in behavior as K approaches $1/(2d)$ from below. We shall investigate this point in considerable detail in several of the chapters below.

3.5 Gaussian Random Functions

In some of our later work, we shall need a simple generalization of the set of Gaussian random variables $\{X_j\}$. This generalization, $X(t)$ which is a random function of one parameter, t , defined in direct analogy to Eq. (3.6). We usually think of t as defining time and of $X(t)$ as a random process which goes on through time. Basically our definition says that any linear superposition of the $X(t)$ is a Gaussian random variable of the usual kind. This assumption leads us to define the Gaussian nature of $X(t)$ by using a generalization of Eq. (3.6). Specifically, define,

$$Q = \int_{-\infty}^{\infty} dt q(t)X(t), \quad (3.22)$$

using any ordinary function of time, $q(t)$. Then define the average of an exponential of Q by the statement

$$\left\langle \exp \left(i \int dt q(t)X(t) \right) \right\rangle = \exp \left[i \int dt q(t)\langle X(t) \rangle - \int ds dt q(s)G(s,t)q(t)/2 \right]. \quad (3.23a)$$

In this definition, $\langle X(t) \rangle$ is the average of the random function while the Green function G is the correlation function for $\delta X(t) = X(t) - \langle X(t) \rangle$ defined by

$$\langle \delta X(t) \delta X(s) \rangle = G(s,t). \quad (3.23b)$$

All this just generalizes Eq. (3.6) and what follows from the case of a set of random variables, X_j , parameterized by the discrete index j to a set $X(t)$ parameterized by a continuous variable t .

3.6 Central Limit Theorem

It is important that we have simple solutions for Gaussian problems. But we should also know how Gaussian behaviors can arise in the first place. The latter is explained by what is called the central limit theorem: I shall not state this theorem in its full generality here, but merely give one special case of it. Imagine that we had a set of Y_j 's, which are independent but identically distributed random variables, each with a finite variance. These might arise for example, as the kinetic energies of a group of essentially identical particles. The mathematical statistician calls such a set of variable *iid*, which means identically and independently distributed. Now form an X as in Eq. (3.5b). Imagine that all the a_j 's were equal to one another. Let N be very large. Then, according to the central limit theorem $\delta X = X - \langle X \rangle$ has a behavior in which all averages of powers of δX are correctly represented by a Gaussian distribution. In rough words,

Whenever some effect represented by the random variable, X , may be represented as the sum of many independent small effects, then the predominant behavior of X is Gaussian. Even more roughly: very complex multi-component causations produce Gaussian outcomes.

Thus, in the absence of other information, when behaviors are complex, one tends to assume Gaussian behavior. This assumption guides the analysis of error in experiments in the physical sciences and the analysis of most of the fluctuations observed in the biological or social sciences.

There is a sense in which this central limit theorem is true, and another sense in which it is false.⁵ Under the conditions stated, whenever the probability distribution function $\rho(x)$ is large it is correctly described by a Gaussian distribution. However, there are very unlikely large-deviation events in which $|(X - \langle X \rangle)|$ is much larger than the square root of the variance. This occurrence is very unlikely. So it may not be very serious that the Gaussian distribution usually miss-estimates these unlikely events. In particular, it is often true that very large deviations occur with a higher probability than the Gaussian distribution would predict.

3.7 Distribution of Energies

We carry through a specific calculation in which we shall see both the central limit theorem and also the non-Gaussian behavior of unlikely events. First let us set up the problem. Imagine that we have a total energy composed of a sum of very many Gaussian components. Thus we might have

$$H = \sum_{j=1}^N \frac{p_j^2}{2m}, \quad (3.24)$$

where the p_j 's are Gaussian variables, each with variance $m\beta^{-1}$. Physically, our problem is one in which an energy, H is composed of a sum of many quadratic kinetic energy terms. (Note that we cannot directly apply Eq. (3.4) since H is not a sum of Gaussian variables but their *squares*. Thus H is not a Gaussian random variable.)

In the remainder of this section, we shall set $m = 1$, since carrying around many factors of the mass is not very instructive.

We wish to calculate the probability, $\rho(E)$ that the value of H is E . To start off, imagine that you wished to calculate the probability that the value of H were less than E . The probability in question is $\langle \eta(E - H) \rangle$ where η is called the Heaviside function and is defined by

$$\eta(E - H) = \begin{cases} 1 & \text{for } E > H, \\ 0 & \text{otherwise.} \end{cases} \quad (3.25)$$

The derivative of the Heaviside function with respect to E is $\delta(H - E)$. In addition, the derivative of the probability that H were less than E is proportional to the probability that

⁵All theorems are true of course. Under the stated conditions for the theorem the results follow. However that does not mean that the inferences we draw are necessarily correct. Not all distributions can even be approximately described as Gaussian.

H is exactly E . In this way we reach the conclusion⁶

Whenever H is a random variable which has a continuum of possible values then whenever the probability function is sufficiently smooth, the probability that H will take on a value between E and $E + dE$ is $\rho(E)dE$ where the probability density $\rho(E)$ is given by

$$\rho(E, N) = \langle \delta(H - E) \rangle. \quad (3.26)$$

The N in Eq. (3.26) is a reminder that we have N quadratic terms in the Hamiltonian so that the specific probability distribution for the N p -values is:

$$\rho(p_1, p_2, \dots, p_N) = \frac{\exp\left(-\sum_{j=1}^N \beta p_j^2 / 2\right)}{Z(\beta)}. \quad (3.27a)$$

Because each p_j is statistically independent of each other momentum, the partition function has a product form

$$Z(\beta) = z(\beta)^N \quad (3.27b)$$

(see Eq. (2.14)). We have already calculated the partition function for a single quadratic degree of freedom. The result is

$$z(\beta) = \int dp e^{-\beta p^2 / 2} = \sqrt{\frac{2\pi}{\beta}}. \quad (3.27c)$$

Thus we can use Eq. (3.26) to calculate the probability of observing an energy E as

$$\rho(E, N) = \frac{1}{z^N} \int dp_1 dp_2 \dots dp_N e^{-\beta E} \delta(H - E) \quad (3.28)$$

To do the integral effectively, think of the N momenta as one long momentum vector $\mathbf{p} = (p_1, p_2, \dots, dp_N)$, so that the Hamiltonian is just proportional to square of the vector, $H = p^2/2$, and the integral in (3.28) can be written as an N -dimensional integral as:

$$dp_1 dp_2 \dots dp_N = d^N \mathbf{p} = d\Omega_N p^{N-1} dp.$$

In the last form of writing the N -dimensional integral is expressed as an $N - 1$ dimensional angular integral, $d\Omega_N$, times a radial integration element $p^{N-1} dp$. (The latter form is required to give the entire integral the right dimensions.)

We can now write the probability as:

$$\rho(E, N) = \frac{1}{z^N} \int d\Omega_N \int p^{N-1} dp e^{-\beta E} \delta\left(\frac{p^2}{2} - E\right).$$

⁶In a system with quantized energy levels $\delta(H - E)$ takes on the values infinity and zero depending upon whether or not E is an energy level. Since the energy levels tend to be very close to one another, one can average over a small range of E . After the average, the function $\rho(E, N)$ will tend to be quite smooth.

The first integral gives the ‘solid’ angle of an N -dimensional sphere; the second is a simple delta function integral. The net result is that the probability of observing an energy E for a system of N particles known to have a temperature β^{-1} is

$$\rho(E, N) = \frac{1}{[z(\beta)]^N} \Omega_N (2E)^{(N-2)/2} e^{-\beta E}. \quad (3.29)$$

We check this formula by demanding that the integral over all E of the result (3.29) be unity. Thus, we find

$$1 = \frac{1}{(2\pi)^{N/2}} \Omega_N \int_0^\infty dE \beta (2\beta E)^{(N-2)/2} e^{-\beta E}.$$

We thus get an expression for the solid angle of an N -dimensional sphere,

$$\Omega_N = 2(\pi)^{N/2} \frac{1}{\int_0^\infty dE (E)^{(N-2)/2} e^{-E}}.$$

The integral is the well-known definition of the gamma function, which is also expressed in terms of the factorial function:

$$\Gamma(n) = \int_0^\infty dx x^{n-1} e^{-x} = (n-1)!.$$

Thus our final result for the solid angle is

$$\Omega_N = \frac{2(\pi)^{N/2}}{\Gamma(N/2)}. \quad (3.30)$$

Thus, Eq. (3.29) gives

$$\rho(E, N) = \frac{1}{E \Gamma(N/2)} (\beta E)^{N/2} e^{-\beta E}. \quad (3.31)$$

3.8 Large Deviations

We are finally in a position to begin talking about the behavior of large deviations. In Fig. 3.1 we plot the probability of observing an energy value E in a system of N quadratic degrees of freedom. We have drawn this out for the first few even values of N . In general the logarithm of probability is of the form.

$$\left(\frac{N}{2} \ln \beta E \right) - \beta E. \quad (3.32)$$

The first term in Eq. (3.32) is like an entropy term since it describes the amount of phase space which is available at a given value of E . This term strongly increases with E . The second term describes an energy factor which demands that the probability fall off quite strongly for large E . These two terms together describe a situation in which phase space

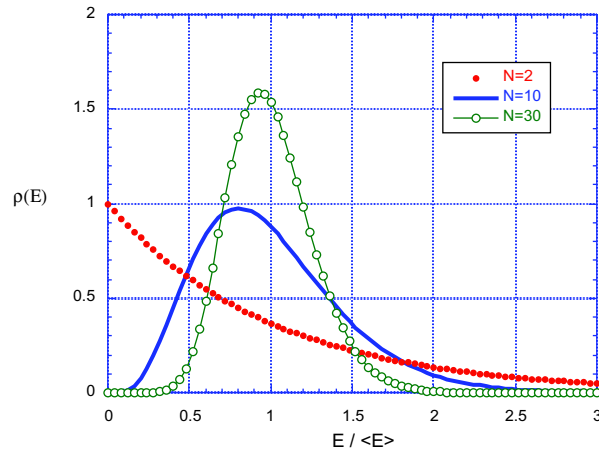


Fig. 3.1. Probability distributions for the energy for three different values of N .

increases with E while the energy term, $-\beta E$, which gives an even more rapidly decreasing probability.

The average value of E is defined by taking the integral of the probability of Eq. (3.31) time E . As we know from the consideration of many Gaussian variables the average energy is given by $N/(2\beta)$. A brief calculation shows that this is also the energy at which the right hand side of (3.32) reaches its peak. So as N gets larger, the average energy grows with N . The peak also has a variance, i.e. $\langle(E - \langle E \rangle)^2\rangle$. This variance is dimensionally like an energy squared. If it grew as fast as $\langle E \rangle^2$, it would grow proportionally to N^2 . But in our discussion of intensive and extensive quantities, we said that the variance was extensive. Therefore, it must grow as N . As N increases, the peak in the distribution gets bigger and bigger. The width of the distribution only grows as $N^{1/2}$, so that the ratio of width to average gets smaller and smaller. The decrease of this relative width has as its consequence that the energy distribution gets sharper and sharper.

Next we ask about whether a Gaussian distribution does represent the behavior of the extreme values in this situation.

But before we go to work, we should ask about why that question is interesting. One answer is that the distribution of extreme events is of considerable practical and theoretic interest. For example, if we have a usual breeding population of 100 members of a species in which births are 90% female, it would be interesting to know what are the odds that one can produce a generation of all females. If you have a company which has reserves which fluctuate, how often will the reserves go negative? If you have a dam which holds back most of the floods, how often will you get a flood so much bigger than the usual to burst the dam? How often will a natural fluctuation in the weather produce a temperature of 120 degrees Fahrenheit in Chicago? If you have a group of atoms in contact with their environment, how often will a fluctuation cool them enough so that they will effectively freeze? In one model, glasses move so slowly because there are large clumps of material in correlated motion. This model might then be used to calculate the extremely small probability that the glass will

produce a clump of a given size. All interesting questions! All involve calculations of large deviations from some ‘normal’ behavior.

So now turn back to the distribution function given in Eq. (3.31). For large N we expect that probabilities can get very large so we take the natural logarithm of the probability. We know that the mean energy is $E = N\beta^{-1}/2$ so we use as our independent variable the ratio of energy to mean energy defined as

$$R = \frac{2\beta E}{N}. \quad (3.33)$$

Then Eq. (3.31) can be written as

$$\ln[\rho(E, N+2)\beta^{-1}] = \frac{N}{2} \ln \frac{RN}{2} - \ln \frac{N}{2}! - \frac{NR}{2}.$$

To get a simple result for the entire expression, we should simplify the factorial. For large values of its argument one can get an excellent approximation for $M!$ using the Sterling approximation

$$\ln M! = M \ln M - M + \ln \sqrt{\frac{M}{2\pi}} + \dots. \quad (3.34)$$

Each term is much smaller than the previous. If we include the first two we include all effects of order $N \ln N$ or N in the logarithm of the probability and the expression turns into the simple statement.

$$\ln[\rho(E, N+2)\beta^{-1}] \cong \frac{N}{2}(1 - R + \ln R). \quad (3.35)$$

Notice that this expression has many simple features. It reaches a maximum at $R = 1$. Thus, the most probable value of the energy is its mean value. Near this point, we have an expansion of the form

$$\ln[\rho(E, N+2)\beta^{-1}] = -\frac{N(R-1)^2}{4} + \frac{N(R-1)^3}{6} + \dots. \quad (3.36)$$

The first term is the Gaussian approximation to the probability and the second is a non-Gaussian correction. For large N , the likely events all have $|(R-1)|$ of the order of or less than $N^{-1/2}$. Thus the deviation from the mean is usually small. In that case, the correction term in Eq. (3.36) is of order $N^{-1/2}$ compared with the first term. However, if we look for a sufficiently unlikely event, for example one in which $|(R-1)|$ is of order unity, then the non-Gaussian behavior can substantially modify the probability. In these circumstances the non-Gaussian term multiplies the probability by a factor of order of some power of $\exp N$. For large N that is a large factor.

Nonetheless in this example, the Gaussian approximation is really quite good. In Eq. (3.36) for large N the first term can get to be an extraordinarily large negative number before the second one counts for anything. For example, for a million degrees of freedom

($N = 10^6$) if we want to see an effect with a likelihood of e^{-100} we would then choose an energy only 2 per cent above the most likely energy, i.e. $R = 1.02$. Then the next correction provided by the second term would be an order one effect changing the 100 in the exponent to something like 99. We could hardly hope to see a change like that.

In contrast, the Gaussian approximation turns out to be quite poor in many situations involving fluid flow. For example one can study the behavior of a contaminant carried around by the motion of some fluid. If the fluid motion is chaotic, the contaminant gets spread in a very uneven fashion around the system. To see the unevenness one measures the probability $P(\Delta\rho, r)$ that two points in the fluid separated by a distance r will have a difference in density $P(\Delta\rho)$. This probability density is given empirically by an exponential distribution

$$P(\Delta\rho) = \frac{1}{2\sigma} e^{-|\Delta\rho|/\sigma}, \quad (3.37)$$

where σ is the standard deviation of $\Delta\rho$. Contrast this with usual Gaussian form

$$P(\Delta\rho) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\Delta\rho)^2}{2\sigma^2}\right], \quad (3.38)$$

which has been the usual guess in statistical problems since the time of Galton.⁷ Chaotic and turbulent systems often show exponential behaviors, like (3.37). Improbable (very bad!!) events are much more likely with the exponential form than with the Gaussian form (3.38). For example, an event with $\Delta\rho = 6\sigma$ has a chance of 10^{-9} of occurring in the Gaussian case, while with the exponential the chance is 0.0025. The Gaussian estimate is clearly quite dangerously wrong in this situation.

3.9 On Almost Gaussian Integrals

We have already seen how statistical mechanics gives us very highly peaked probability distributions. Often this is a result of the same kind of phenomenon which we saw in the probability distribution for the energy: The probability includes one factor (often an energy factor) which is dying off very rapidly while another factor (often a number of states or an entropy factor) grows quite vividly. When the growth rate equals the decay rate, the probability reaches a quite sharp maximum.

The simplest integrals of this type have the structure

$$Y(N) = \int_a^b dx \exp(N\Gamma(x)), \quad (3.39)$$

where N is a number which goes to infinity and $\Gamma(x)$ is a smooth function which reaches a maximum at a single point x_0 within the interior of the range of integration. Take this

⁷Sir Francis Galton (1822–1911) pioneered the use of statistics in the social sciences.

simplest case and let $a < x_0 < b$. Now expand about x_0 . The expansion is one in which the linear term in $x - x_0$ must vanish because x_0 is an extremal, and thus takes the form

$$\Gamma(x) \cong \Gamma(x_0) + \frac{1}{2}\Gamma''(x_0)(x - x_0)^2 + \dots \quad (3.40)$$

Here, we need $\Gamma''(x_0)$ to be negative so that we have a maximum in the integrand. Assuming that this is so, we substitute (3.40) into (3.39) and extend the range of integration in x to $[-\infty, \infty]$. The integral is easily performed to give the result:

$$Y = \sqrt{\frac{2\pi}{-N\Gamma''(x_0)}} \exp[N\Gamma(x_0)]. \quad (3.41)$$

This method of integration is called the method of steepest descent or saddle-point integration. These names describe what one does when the vanishing slope appears someplace off the real axis in the complex plane. Notice that the leading N -dependence occurs in the exponent. The multiplicative factor of $1/\sqrt{N}$ is a relatively small correction.

3.10 Three Versions of Gaussian Problems

Social scientists, economists, and students of finance make very considerable use of Gaussian variables. In fact, there are at least three different ways that Gaussians are used in physics and other statistical applications.

(1) The first is the obvious situation in which the physical variables, X_j , have a Gaussian distribution. Then, in this case, averages of the X 's are given by formulas like

$$\langle (X_j - \langle X_j \rangle)(X_k - \langle X_k \rangle) \rangle = G_{jk} \quad (3.42)$$

and its higher order analog

$$\begin{aligned} & \langle (X_j - \langle X_j \rangle)(X_k - \langle X_k \rangle)(X_m - \langle X_m \rangle)(X_n - \langle X_n \rangle) \rangle \\ & = G_{jk}G_{mn} + G_{jm}G_{kn} + G_{jn}G_{km}. \end{aligned} \quad (3.43)$$

In this situation, we say the X 's are *normally distributed*.

(2) Another very common situation is one in which the basic variables, which we now call Z_j are each the exponential of a linear combination of Gaussian variables, i.e.

$$\ln Z_j = \sum_k a_{j,k} X_k. \quad (3.44)$$

Here the a 's form a matrix of coefficients. In this case, we say that the Z 's have a log-normal distribution. Using Eq. (3.6) we can then form the average of Z_j as

$$\langle Z_j \rangle = \exp \left(\langle Y_j \rangle + \frac{1}{2} \langle (Y_j - \langle Y_j \rangle)^2 \rangle \right), \quad (3.45)$$

where Y_j is an abbreviation for the composite Gaussian variable defined by the right hand side of Eq. (3.44). In this case, we say that the Z 's are log-normally distributed.

(3) In physics we also use complex exponentials of Gaussian variables, as for example,

$$\Phi = e^{i\theta}. \quad (3.46)$$

Since expression (3.46) is invariant under the change $\theta \rightarrow (\theta + 2\pi)$, the variable θ may be thought of as an angle. The Gaussian structure then gives the average of Φ in a form which is very much the same as in the log normal case, except for a crucial sign difference. Equation (3.6) implies:

$$\langle \Phi \rangle = \exp \left(i\langle \theta \rangle - \frac{1}{2} \langle (\theta - \langle \theta \rangle)^2 \rangle \right). \quad (3.47)$$

We shall refer to a variable like θ as a phase variable.

We shall, at various times, use all three of these kinds of variables of Gaussian theory.

Homework

Problem 3.1 (Generating Function). Equation (3.3) is followed by the line ‘Conversely if (3.3) is true, the distribution of X must be Gaussian, with variance β^{-1} ’. Prove this statement.

Problem 3.2 (Two Gaussian Variables). Consider a Hamiltonian involving two Gaussian variables, X and Y . Start from the statement that the average formed by these two variables is of the form

$$\langle \exp(aX + bY) \rangle = \exp[a^2 + b^2 - ab]$$

for any a and b . What is the probability distribution for X alone, for Y alone, for X and Y together?

Problem 3.3 (Breeding). If we have a usual breeding population of 100 members of a species in which births are 90% female, it would be interesting to know what are the odds that one can produce a generation of all females. What are they?

Problem 3.4 (Sterling Approximation). Statistical physics makes considerable use of an approximation for the factorial function called the Sterling approximation. See Eq. (3.34) An alternative form is that for large N :

$$N! \cong N^N e^{-N} \sqrt{2\pi N}.$$

- (a) Make a table for $N = 1, 2, \dots, 10$ to see how good this approximation is and to figure out the relative sizes of the successive terms.

- (b) There is a simple integral formula for $N!$, namely,

$$N! = \int_0^{\infty} dx x^N e^{-x}.$$

One can derive this formula by using

$$\frac{d}{dx} e^{-x} = -e^{-x}$$

and integrating by parts N times. Show how this derivation works.

- (c) Evaluate this integral by the saddle point method and in this way derive the Sterling approximation.

Problem 3.5 (Entropy of Dilute Gas). Consider a dilute gas of N identical atoms of energy E confined in a box of volume Ω . Let N be much larger than one.

- (a) Calculate the entropy of this gas.
- (b) Try to explain this answer using the classical limit of quantum mechanics and trying to estimate the volume in phase space occupied by such a gas.

Problem 3.6 (Higher Order Correlations). Derive Eq. (3.43). If X is a Gaussian variable with average $\langle X \rangle$ and variance β , what is the value of $\langle X^p \rangle$ for $p = 3, 4, 5$ and 6 ?