2. Probability Concepts

In the preceding chapter, we introduced probability notions without any definitions. In order to formulate essential concepts more precisely, it is necessary to have some more precise expression of these concepts. The intention of this chapter is to provide some background, and to present a number of essential results. It is not a thorough outline of mathematical probability, for which the reader is referred to standard mathematical texts such as those by *Feller* [2.1] and *Papoulis* [2.2].

2.1 Events, and Sets of Events

It is convenient to use a notation which is as general as possible in order to describe those occurrences to which we might wish to assign probabilities. For example, we may wish to talk about a situation in which there are 6.4×10^{14} molecules in a certain region of space; or a situation in which a Brownian particle is at a certain point $\boldsymbol{x}$ in space; or possibly there are 10 mice and 3 owls in a certain region of a forest.

These occurrences are all examples of practical realisations of *events*. More abstractly, an event is simply a member of a certain space, which in the cases most practically occuring can be characterised by a vector of integers

$$\boldsymbol{n} = (n_1, n_2, n_3 \ldots) \tag{2.1.1}$$

or a vector of real numbers

$$\boldsymbol{x} = (x_1, x_2, x_3 \ldots). \tag{2.1.2}$$

The dimension of the vector is arbitary.

It is convenient to use the language of set theory, introduce the concept of a *set of events*, and use the notation

$$\omega \in A \tag{2.1.3}$$

to indicate that the event ω is one of events contained in A. For example, one may consider the set $A(25)$ of events in the ecological population in which there are no more than 25 animals present; clearly the event $\bar{\omega}$ that there are 3 mice, a tiger, and no other animals present satisfies

$$\bar{\omega} \in A(25). \tag{2.1.4}$$

More significantly, suppose we define the set of events $A(\boldsymbol{r}, \Delta V)$ that a molecule is within a volume element ΔV centred on a point $\boldsymbol{r}$. In this case, the practical significance of working in terms of sets of events becomes clear, because we should normally be able to determine whether or not a molecule is within a neighbourhood ΔV of $\boldsymbol{r}$, but to determine whether the particle is exactly at $\boldsymbol{r}$ is impossible. Thus, if we define the event $\omega(\boldsymbol{y})$ that the molecule is at point $\boldsymbol{y}$, it makes sense to ask whether

$$\omega(\boldsymbol{y}) \in A(\boldsymbol{r}, \Delta V) \tag{2.1.5}$$

and to assign a certain probability to the *set* $A(\boldsymbol{r}, \Delta V)$, which is to be interpreted as the probability of the occurrence of (2.1.5)

2.2 Probabilities

Most people have an intuitive conception of a probability, based on their own experience. However, a precise formulation of intuitive concepts is fraught with difficulties, and it has been found most convenient to axiomatise probability theory as an essentially abstract science, in which a probability measure $P(A)$ is *assigned* to every set A, in the space of events, including

$$\text{the set of all events: } \Omega \tag{2.2.1}$$

$$\text{the set of no events: } \varnothing\,; \tag{2.2.2}$$

in order to define probability, we need our sets of events to form a closed system (known by mathematicians as a *σ-algebra*) under the set theoretic operations of union and intersection.

2.2.1 Probability Axioms

We introduce the probability of A, $P(A)$, as a function of A satisfying the following *probability axioms*:

$$\text{(i) } P(A) \geqslant 0 \quad \text{for all } A\,, \tag{2.2.3}$$

$$\text{(ii) } P(\Omega) = 1\,, \tag{2.2.4}$$

(iii) if A_i ($i=1, 2, 3, \ldots$) is a countable (but possibly infinite) collection of *nonoverlapping sets*, i.e., such that

$$A_i \cap A_j = \varnothing \quad \text{for all} \quad i \neq j, \tag{2.2.5}$$

then

$$P\left(\bigcup_i A\right) = \sum_i P(A_i)\,. \tag{2.2.6}$$

These are all the axioms needed. Consequentially, however, we have:

(iv) if $\bar{A}$ is the complement of A, i.e., the set of all events not contained in A, then

$$P(\bar{A}) = 1 - P(A), \tag{2.2.7}$$

(v) $P(\varnothing) = 0.$ (2.2.8)

2.2.2 The Meaning of $P(A)$

There is no way of making probability theory correspond to reality without requiring a certain degree of intuition. The probability $P(A)$, as axiomatised above, is the intuitive probability that an "*arbitrary*" event ω, i.e., an event ω "*chosen at random*", will satisfy $\omega \in A$. Or more explicitly, if we choose an event "*at random*" from Ω N times, the relative frequency that the particular event chosen will satisfy $\omega \in A$ approaches $P(A)$ as the number of times, N, we choose the event, approaches infinity. The number of choices N can be visualised as being done one after the other ("*independent*" tosses of one die) or at the same time (N dice are thrown at the same time "*independently*"). All definitions of this kind must be intuitive, as we can see by the way undefined terms ("*arbitrary*", "*at random*", "*independent*") keep turning up. By eliminating what we now think of as intuitive ideas and axiomatising probability, *Kolomogorov* [2.3] cleared the road for a rigorous development of mathematical probability. But the circular definition problems posed by wanting an intuitive understanding remain. The simplest way of looking at axiomatic probability is as a formal method of manipulating probabilities using the axioms. In order to apply the theory, the probability space must be defined *and* the probability measure P assigned. These are *a priori probabilities*, which are simply assumed. Examples of such a priori probabilities abound in applied disciplines. For example, in equilibrium statistical mechanics one assigns equal probabilities to equal volumes of phase space. Einstein's reasoning in Brownian motion assigned a probability $\phi(\Delta)$ to the probability of a "push" Δ from a position x at time t.

The task of applying probability is (i) to assume some set of *a priori* probabilities which seem reasonable and to deduce results from this and from the structure of the probability space, (ii) to measure experimental results with some apparatus which is constructed to measure quantities in accordance with these a priori probabilities.

The structure of the probability space is very important, especially when the space of events is compounded by the additional concept of time. This extension makes the effective probability space infinite-dimensional, since we can construct events such as "the particle was at points $\boldsymbol{x}_n$ at times t_n, $n = 0, 1, 2, \ldots, \infty$".

2.2.3 The Meaning of the Axioms

Any intuitive concept of probability gives rise to nonnegative probabilities, and the probability that an arbitrary event is contained in the set of all events must be 1 no matter what our definition of the word arbitrary. Hence, axioms (i) and (ii) are understandable. The heart of the matter lies in axiom (iii). Suppose we are dealing with only 2 sets A and B, and $A \cap B = \varnothing$. This means there are *no* events con-

tained in both A and B. Therefore, the probability that $\omega \in A \cup B$ is the probability that *either* $\omega \in A$ or $\omega \in B$. Intuitive considerations tell us this probability is the sum of the individual probabilities, i.e.,

$$P(A \cup B) \equiv P\{(\omega \in A) \text{ or } (\omega \in B)\} = P(A) + P(B) \tag{2.2.9}$$

(notice this is not a proof—merely an explanation).

The extension now to any finite number of nonoverlapping sets is obvious, but the extension only to any *countable* number of nonoverlapping sets requires some comment.

This extension must be made restrictive because of the existence of sets labelled by a continuous index, for example, $\boldsymbol{x}$, the position in space. The probability of a molecule being in the set whose only element in $\boldsymbol{x}$ is zero; but the probability of being in a region R of finite volume is nonzero. The region R is a union of sets of the form $\{\boldsymbol{x}\}$—but not a *countable* union. Thus axiom (iii) is not applicable and the probability of being in R is *not* equal to the sum of the probabilities of being in $\{\boldsymbol{x}\}$.

2.2.4 Random Variables

The concept of a random variable is a notational convenience which is central to this book. Suppose we have an abstract probability space whose events can be written $\boldsymbol{x}$. Then we can introduce the random variable $F(\boldsymbol{x})$ which is a function of $\boldsymbol{x}$, which takes on certain values for each $\boldsymbol{x}$. In particular, the identity function of $\boldsymbol{x}$, written $\boldsymbol{X}(\boldsymbol{x})$ is of interest; it is given by

$$\boldsymbol{X}(\boldsymbol{x}) = \boldsymbol{x}. \tag{2.2.10}$$

We shall normally use capitals in this book to denote random variables and small letters $\boldsymbol{x}$ to denote their values whenever it is necessary to make a distinction.

Very often, we have some quite different underlying probability space Ω with values ω, and talk about $\boldsymbol{X}(\omega)$ which is some function of ω, and then omit explicit mention of ω. This can be for either of two reasons:

i) we specify the events by the values of $\boldsymbol{x}$ anyway, i.e., we identify $\boldsymbol{x}$ and ω;
ii) the underlying events ω are too complicated to describe, or sometimes, even to know.

For example, in the case of the position of a molecule in a liquid, we really should interpret each ω as being capable of specifying all the positions, momenta, and orientations of each molecule in that volume of liquid; but this is simply too difficult to write down, and often unnecessary.

One great advantage of introducing the concept of a random variable is the simplicity with which one may handle functions of random variables, e.g., $\boldsymbol{X}^2$, $\sin(\boldsymbol{a} \cdot \boldsymbol{X})$, etc, and compute means and distributions of these. Further, by defining stochastic differential equations, one can also quite simply talk about time development of random variables in a way which is quite analogous to the classical description by means of differential equations of nonprobabilistic systems.

2.3 Joint and Conditional Probabilities: Independence

2.3.1 Joint Probabilities

We explained in Sect. 2.2.3 how the occurrence of mutually exclusive events is related to the concept of nonintersecting sets. We now consider the concept $P(A \cap B)$, where $A \cap B$ is nonempty. An event ω which satisfies $\omega \in A$ will only satisfy $\omega \in A \cap B$ if $\omega \in B$ as well.

$$\text{Thus, } P(A \cap B) = P\{(\omega \in A) \text{ and } (\omega \in B)\} \tag{2.3.1}$$

and $P(A \cap B)$ is called the *joint probability* that the event ω is contained in both classes, or, alternatively, that both the events $\omega \in A$ and $\omega \in B$ occur. Joint probabilities occur naturally in the context of this book in two ways:

i) When the event is specified by a vector, e.g., m mice and n tigers. The probability of this event is the joint probability of [m mice (and any number of tigers)] and [n tigers (and any number of mice)]. All vector specifications are implicitly joint probabilities in this sense.

ii) *When more than one time is considered:* what is the probability that (at time t_1 there are m_1 tigers and n_1 mice) and (at time t_2 there are m_2 tigers and n_2 mice). To consider such a probability, we have effectively created out of the events at time t_1 and events at time t_2, *joint events* involving one event at each time. In essence, there is no difference between these two cases except for the fundamental dynamical role of time.

2.3.2 Conditional Probabilities

We may specify conditions on the events we are interested in and consider only these, e.g., the probability of 21 buffaloes given that we know there are 100 lions. What does this mean? Clearly, we will be interested only in those events contained in the set $B =$ {all events where exactly 100 lions occur}. This means that we to define conditional probabilities, which are defined only on the collection of all sets contained in B. we define the conditional probability as

$$P(A \mid B) = P(A \cap B)/P(B) \tag{2.3.2}$$

and this satisfies our intuitive conception that the conditional probability that $\omega \in A$ (given that we know $\omega \in B$), is given by dividing the probability of joint occurrence by the probability ($\omega \in B$).

We can define in both directions, i.e., we have

$$P(A \cap B) = P(A \mid B)P(B) = P(B \mid A)P(A). \tag{2.3.3}$$

There is no particular conceptual difference between, say, the probability of {(21 buffaloes) given (100 lions)} and the reversed concept. However, when two times

are involved, we do see a difference. For example, the probability that a particle is at position x_1 at time t_1, given that it was at x_2 at the *previous* time t_2, is a very natural thing to consider; indeed, it will turn out to be a central concept in this book. The converse sounds strange, i.e., the probability that a particle is at position x_1 at time t_1, given that it will be at position x_2 at a later time t_2. It smacks of clairvoyance—we cannot conceive of any natural way in which we would wish to consider it, although it is, in principle, a quantity very similar to the "natural" conditional probability, in which the condition precedes the events under consideration.

The natural definition has already occurred in this book, for example, the $\phi(\Delta)d\Delta$ of Einstein (Sect. 1.2.1.) is the probability that a particle at x at time t will be in the range $[x + \Delta, x + \Delta + d\Delta]$ at time $t + \tau$, and similarly in the other examples. Our intuition tells us as it told Einstein (as can be seen by reading the extract from his paper) that this kind of conditional probability is directly related to the time development of a probabilistic system.

2.3.3 Relationship Between Joint Probabilities of Different Orders

Suppose we have a collection of sets B_t such that

$$B_t \cap B_j = \varnothing \tag{2.3.4}$$

$$\bigcup_t B_t = \Omega \tag{2.3.5}$$

so that the sets divide up the space Ω into nonoverlapping subsets.
Then

$$\bigcup_t (A \cap B_t) = A \cap (\bigcup_t B_t) = A \cap \Omega = A \tag{2.3.6}$$

Using now the probability axiom (iii), we see that $A \cap B_t$ satisfy the conditions on the A_t used there, so that

$$\sum_t P(A \cap B_t) = P[\bigcup_t (A \cup B_t)] \tag{2.3.7}$$

$$= P(A) \tag{2.3.8}$$

and thus

$$\sum_t P(A \mid B_t)P(B_t) = P(A) \tag{2.3.9}$$

Thus, summing over all mutually exclusive possibilities of B in the joint probability eliminates that variable.

Hence, in general,

$$\sum_t P(A_t \cap B_j \cap C_k \ldots) = P(B_j \cap C_k \cap \ldots). \tag{2.3.10}$$

The result (2.3.9) has very significant consequences in the development of the theory of stochastic processes, which depends heavily on joint probabilities.

2.3.4 Independence

We need a probabilistic way of specifying what we mean by independent events. Two sets of events A and B should represent independent sets of events if the specification that a particular event is contained in B has no influence on the probability of that event belonging to A. Thus, the conditional probability $P(A \mid B)$ should be independent of B, and hence

$$P(A \cap B) = P(A)P(B) \tag{2.3.11}$$

In the case of several events, we need a somewhat stronger specification. The events $(\omega \in A_i)$ $(i = 1,2,\ldots, n)$ will be considered to be independent if for any subset $(i_1, i_2, \ldots, i_k)$ of the set $(1,2, \ldots, n)$,

$$P(A_{i_1} \cap A_{i_2} \ldots A_{i_k}) = P(A_{i_1})P(A_{i_2}) \ldots P(A_{i_k}) . \tag{2.3.12}$$

It is important to require factorisation for all possible combinations, as in (2.3.12). For example, for three sets A_i, it is quite conceivable that

$$P(A_i \cap A_j) = P(A_i)P(A_j) \tag{2.3.13}$$

for all different i and j, but also that

$$A_1 \cap A_2 = A_2 \cap A_3 = A_3 \cap A_1 . \qquad \text{(see Fig 2.1)}$$

This requires

$$P(A_1 \cap A_2 \cap A_3) = P(A_2 \cap A_3 \cap A_3) = P(A_2 \cap A_3) = P(A_2)P(A_3) \neq P(A_1)P(A_2)P(A_3). \tag{2.3.14}$$

We can see that the occurrence of $\omega \in A_2$ and $\omega \in A_3$ necessarily implies the occurrence of $\omega \in A_1$. In this sense the events are obviously not independent.

Random variables $X_1, X_2, X_3, \ldots$, will be said to be independent random variables, if for all sets of the form $A_i = (x$ such that $a_i \leqslant x \leqslant b_i)$ the events $X_1 \in A_1$,

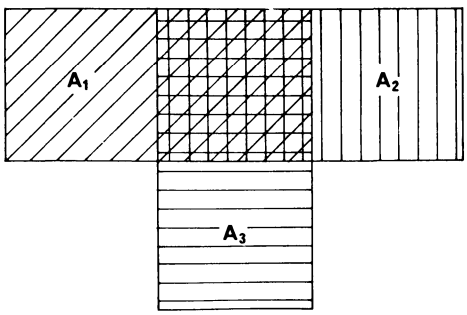

Fig. 2.1. Illustration of statistical independence in pairs, but not in threes. In the three sets $A_j \cap A_i$ is, in all cases, the central region. By appropriate choice of probabilities, we can arrange

$$P(A_i \cap A_j) = P(A_i)P(A_j)$$

$X_2 \in A_2$, $X_3 \in A_3$, ... are independent events. This will mean that all values of the X_t are assumed independently of those of the remaining X_t.

2.4 Mean Values and Probability Density

The mean value of a random variable $R(\omega)$ in which the basic events ω are countably specifiable is given by

$$\langle R \rangle = \sum_{\omega} P(\omega) R(\omega) , \tag{2.4.1}$$

where $P(\omega)$ means the probability of the set containing only the single event ω. In the case of a continuous variable, the probability axioms above enable us to define a probability density function $p(\omega)$ such that if $A(\omega_0, d\omega_0)$ is the set

$$(\omega_0 \leqslant \omega < \omega_0 + d\omega_0) , \tag{2.4.2}$$

then

$$p(\omega_0) d\omega_0 = P[A(\omega_0, d\omega_0)] \tag{2.4.3}$$

$$\equiv p(\omega_0, d\omega_0) . \tag{2.4.4}$$

The last is a notation often used by mathematicians. Details of how this is done have been nicely explained by *Feller* [2.1]. In this case,

$$\langle R \rangle = \int_{\omega \in \Omega} d\omega \, R(\omega) p(\omega) . \tag{2.4.5}$$

One can often (as mentioned in Sect. 2.2.4) use R itself to specify the event, so we will often write

$$\langle R \rangle = \int dR \, R p(R) . \tag{2.4.6}$$

Obviously, $p(R)$ is not the same function of R as $p(\omega)$ is of ω–more precisely

$$p(R_0) dR_0 = P[R_0 < R < R_0 + dR_0)] . \tag{2.4.7}$$

2.4.1 Determination of Probability Density by Means of Arbitrary Functions

Suppose for every function $f(R)$ we know

$$\langle f(R) \rangle = \int dR \, f(R) \, p(R) , \tag{2.4.8}$$

then we know $p(R)$. The proof follows by choosing

$$f(R) = 1 \qquad R_0 \leqslant R < R_0 + dR_0$$

$$= 0 \qquad \text{otherwise.}$$

Because the expectation of an arbitrary function is sometimes a little easier to work with than a density, this relation will be used occasionally in this book.

2.4.2 Sets of Probability Zero

If a density $p(R)$ exists, the probability that R is in the interval $(R_0, R_0 + dR)$ goes to zero with dR. Hence, the probability that R has *exactly* the value R_0 is zero; and similarly for any other value.

Thus, in such a case, there are sets $S(R_t)$, each containing only one point R_t, which have zero probability. From probability axiom (iii), any countable union of such sets, i.e., any set containing only a countable number of points (e.g., all rational numbers) has probability zero. In general, all equalities in probability theory are at best only "*almost certainly true*", i.e., they may be untrue on sets of probability zero. Alternatively, one says, for example,

$$X = Y \text{ (with probability 1)} \tag{2.4.9}$$

which is by no means the same as saying that

$$X(R) = Y(R) \text{ for all } R. \tag{2.4.10}$$

Of course, (if the theory is to have any connection with reality) events with probability zero do not occur.

In particular, notice that our previous result if inspected carefully, only implies that we know $p(R)$ only with probability 1, given that we know $\langle f(R) \rangle$ for all $f(R)$.

2.5 Mean Values

The question of what to measure in a probabilistic system is nontrivial. In practice, one measures either a set of individual values of a random variable (the number of animals of a certain kind in a certain region at certain points in time; the electric current passing through a given circuit element in each of a large number of replicas of that circuit, etc.) or alternatively, the measuring procedure may implicitly construct an average of some kind. For example, to measure an electric current, we may measure the electric charge transferred and divide by the time taken—this gives a measure of the average number of electrons transferred per unit time. It is important to note the essential difference in this case, that it will not normally be possible to measure anything other than a few selected averages and thus, higher moments (for example) will be unavailable.

In contrast, when we measure individual events (as in counting animals), we can then construct averages of the observables by the obvious method

$$\bar{X}_N = \frac{1}{N} \sum_{n=1}^{N} X(n) . \tag{2.5.1}$$

The quantities $X(n)$ are the individual observed values of the quantity X. We expect

that as the number of samples N becomes very large, the quantity $\bar{X}_N$ approaches the mean $\langle X \rangle$ and that, in fact,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f[X(n)] = \lim_{N\to\infty} \overline{f(X)}_N = \langle f(X) \rangle \tag{2.5.2}$$

and such a procedure will determine the probability density function $p(x)$ of X if we carry out this procedure for all functions f. The validity of this procedure depends on the degree of independence of the successive measurements and is dealt with in Sect. 2.5.2.

In the case where only averages themselves are directly determined by the measuring method, it will not normally be possible to measure $X(n)$ and therefore, it will not, in general, be possible to determine $\overline{f(X)}_N$. All that will be available will be $f(\bar{X}_N)$—quite a different thing unless f is linear. We can often find situations in which measurable quantities are related (by means of some theory) to mean values of certain functions, but to hope to measure, for example, the mean value of an arbitrary function of the number of electrons in a conductor is quite hopeless. The mean number—yes, and indeed even the mean square number, but the measuring methods available are not direct. We do *not* enumerate the individual numbers of electrons at different times and hence arbitrary functions are not attainable.

2.5.1 Moments, Correlations, and Covariances

Quantities of interest are given by the *moments* $\langle X^n \rangle$ since these are often easily calculated. However, probability densities must always vanish as $x \to \pm \infty$, so we see that higher moments tell us only about the properties of unlikely large values of X. In practice we find that the most important quantities are related to the first and second moments. In particular, for a single variable X, *the variance* defined by

$$\text{var}\{X\} \equiv \{\sigma[X]\}^2 \equiv \langle [X - \langle X \rangle]^2 \rangle \,, \tag{2.5.3}$$

and as is well known, the *variance* var $\{X\}$ or its square root the *standard deviation* $\sigma[X]$, is a measure of the degree to which the values of X deviate from the mean value $\langle X \rangle$.

In the case of several variables, we define the *covariance matrix* as

$$\langle X_i, X_j \rangle \equiv \langle (X_i - \langle X_i \rangle)(X_j - \langle X_j \rangle) \rangle \equiv \langle X_i X_j \rangle - \langle X_i \rangle \langle X_j \rangle \,. \tag{2.5.4}$$

Obviously,

$$\langle X_i, X_i \rangle = \text{var}\ \{X_i\} \,. \tag{2.5.5}$$

If the variables are independent *in pairs*, the covariance matrix is diagonal.

2.5.2 The Law of Large Numbers

As an application of the previous concepts, let us investigate the following model of measurement. We assume that we measure the same quantity N times, obtaining

sample values of the random variable $X(n)$ $(n = 1, 2, \ldots, N)$. Since these are all measurements of the same quantity at successive times, we assume that for every n, $X(n)$ has the same probability distribution but we do not assume the $X(n)$ to be independent. However, provided the covariance matrix $\langle X(n), X(m)\rangle$ vanishes sufficiently rapidly as $|n - m| \to \infty$, then defining

$$\bar{X}_N = \frac{1}{N}\sum_{n=1}^{N} X(n)\,, \tag{2.5.6}$$

we shall show

$$\lim_{N\to\infty} \bar{X}_N = \langle X\rangle\,. \tag{2.5.7}$$

It is clear that

$$\langle \bar{X}_N\rangle = \langle X\rangle\,. \tag{2.5.8}$$

We now calculate the variance of $\bar{X}_N$ and show that as $N \to \infty$ it vanishes under certain conditions:

$$\langle \bar{X}_N\bar{X}_N\rangle - \langle \bar{X}_N\rangle^2 = \frac{1}{N^2}\sum_{n,m=1}^{N} \langle X_n, X_m\rangle\,. \tag{2.5.9}$$

Provided $\langle X_n, X_m\rangle$ falls off sufficiently rapidly as $|n - m| \to \infty$, we find

$$\lim_{N\to\infty} (\mathrm{var}\{\bar{X}_N\}) = 0 \tag{2.5.10}$$

so that $\lim_{N\to\infty} \bar{X}_N$ is a deterministic variable equal to $\langle X\rangle$.

Two models of $\langle X_n, X_m\rangle$ can be chosen.

a) $$\langle X_n, X_m\rangle \sim K\lambda^{|m-n|} \qquad (\lambda < 1) \tag{2.5.11}$$

for which one finds

$$\mathrm{var}\{\bar{X}_N\} = \frac{2K}{N^2}\left(\frac{\lambda^{N+2} - N(\lambda - 1) - \lambda}{(\lambda - 1)^2}\right) - \frac{K}{N} \to 0\,. \tag{2.5.12}$$

b) $$\langle X_n, X_m\rangle \sim |n - m|^{-1} \qquad (n \neq m) \tag{2.5.13}$$

and one finds approximately

$$\mathrm{var}\{\bar{X}_N\} \sim \frac{2}{N}\log N - \frac{1}{N} \to 0\,. \tag{2.5.14}$$

In both these cases, $\mathrm{var}\{X_N\} \to 0$. The rate of convergence is very different. In-

terpreting n, m as the times at which the measurement is carried out, one sees than even very slowly decaying correlations are permissible. The law of large numbers comes in many forms, which are nicely summarised by *Papoulis* [2.2]. The central limit theorem is an even more precise result in which the limiting distribution function of $\bar{X}_N - \langle X \rangle$ is determined (see Sect. 2.8.2).

2.6 Characteristic Function

One would like a condition where the variables are independent, not just in pairs. To this end (and others) we define the characteristic function.

If $\boldsymbol{s}$ is the vector $(s_1, s_2, \ldots, s_n)$, and $\boldsymbol{X}$ the vector of random variables $(X_1, X_2, \ldots, X_n)$, then the characteristic function (or moment generating function) is defined by

$$\phi(\boldsymbol{s}) = \langle \exp(\mathrm{i}\boldsymbol{s} \cdot \boldsymbol{X}) \rangle = \int d\boldsymbol{x}\, p(\boldsymbol{x}) \exp(\mathrm{i}\boldsymbol{s} \cdot \boldsymbol{x}) . \tag{2.6.1}$$

The characteristic function has the following properties [Ref. 2.1, Chap. XV]

i) $\phi(\mathbf{0}) = 1$

ii) $\phi|(\boldsymbol{s})| \leqslant 1$

iii) $\phi(\boldsymbol{s})$ is a uniformly continuous function of its arguments for all finite real $\boldsymbol{s}$ [2.5].

iv) If the *moments* $\langle \prod_i X_i^{m_i} \rangle$ exist, then

$$\langle \prod_i X_i^{m_i} \rangle = \left[\prod_i \left(-\mathrm{i} \frac{\partial}{\partial s_i} \right)^{m_i} \phi(\boldsymbol{s}) \right]_{\boldsymbol{s}=\mathbf{0}} . \tag{2.6.2}$$

v) A sequence of probability densities converges to limiting probability density if and only if the corresponding characteristic functions converge to the corresponding characteristic function of the limiting probability density.

vi) Fourier inversion formula

$$p(\boldsymbol{x}) = (2\pi)^{-n} \int d\boldsymbol{s}\, \phi(\boldsymbol{s}) \exp(-\mathrm{i}\boldsymbol{x} \cdot \boldsymbol{s}) \tag{2.6.3}$$

Because of this inversion formula, $\phi(\boldsymbol{s})$ determines $p(\boldsymbol{x})$ with probability 1. Hence, the characteristic function does truly *characterise* the probability density.

vii) Independent random variables: from the definition of independent random variables in Sect. 2.3.4, it follows that the variables $X_1, X_2 \ldots$ are independent if and only if

$$p(x_1, x_2, \ldots, x_n) = p_1(x_1) p_2(x_2) \ldots p_n(x_n), \tag{2.6.4}$$

in which case,