Chapter 3

Probability theory

3.1 Probability

Most physical, chemical or biological systems accessible to experimental observations consists of many particles: atoms, molecules etc. In trying to understand their properties, we cannot hope to follow the motion of all the constituent particles in full detail. The outcome of an experiment is not completely predictable from the limited data available about the initial conditions of the system, even when the physical laws governing the system are known. Normally the best one can do is to predict what will happen most of the time, and estimate the frequency deviations from this expectation will occur and what their magnitudes might be. The mathematical tool which has been developed to make such predictions is the theory of probability.

The mathematical setting for probability theory is the following. There is a space Ω, called the *sample space*. The set Ω can be any collection of elements. The elements or points of the sample space, designated by ω, are called the *elementary events*. Subsets of Ω containing more than one point are called *compound events*, or just *events*. The idea is that the ω's are the possible results of experiments and Ω is the set of all possible results of experiments. For example, if the experiment is throwing of a die, Ω consists of six points, which may be taken as the integers, $1, \ldots, 6$. In this case Ω has $2^6 = 64$ subsets, counting the empty set $\emptyset$ and the entire space Ω.

To each of the subsets $E \in \Omega$, there is assigned a probability $P(E)$ called the probability of E. This is supposed to represent the liklihood that the event E will occur in a given experiment or trial. How are these numbers to be assigned? Do probabilities reflect the degree of certainty or beliefe that an event has occured or will occur, or do they represent the relative frequency with which events occur if the experiment is repeated sufficiently many times? Most physical scientists adopt the relative frequency interpretation i.e. if an experiment is repeated N times and the event E is observed to happen in N_E of these then

$$P(E) = \frac{N_E}{N}, \quad N \to \infty$$

Another possibility is to assume equal a priori probability for the elementary events, i.e. if the space Ω consists of N elements then

$$P(\omega) = \frac{1}{N}$$

from which $P(E)$ can be obtained for any E.

The mathematical theory of probability is concerned with relations between the probabilities of various sets, and the properties of various functions associated with these probabilities. Let E_1 and E_2 be sets. The set of points common to both E_1 and E_2 is denoted by $E_1 \cap E_2 = E_1E_2$. It represents the simultaneous occurence of both E_1 and E_2. The set $E_1 \cup E_2 = E_1 + E_2$ denote the set of points belonging to E_1 or to E_2 or to both. This corresponds to the occurence of either E_1 or E_2 or both. The set of points in E_1 but not in E_2 is denoted by $E_1 - E_2$. This corresponds to the occurence of E_1 and non-occurence of E_2. The complement of a set E denoted by $\overline{E}$ is the set of points not in $E : \overline{E} = \Omega - E$. This corresponds to the nonoccurence of E. If $E_1 \cap E_2 = E_1E_2 = \emptyset$ we say that the events E_1 and E_2 are mutually exclusive (or disjoint).

Example 3.1 *Consider the tossing of a fair coin. The elementary events are ω_H and ω_T where ω_H and ω_T are the elementary events that the coin lands with the head up and the tail up, respectively. These events are mutually exclusive, i.e. $\omega_H \cap \omega_T = \emptyset$.*

We can now state the axioms at the foundation of probability theory.

Definition 3.1 *A probability function is a real scalar function P defined on the subsets of a space Ω with the following properties:*

1. *For every $E \subseteq, \Omega$, $P(E) \geq 0$.*
2. *$P(\Omega) = 1$.*
3. *For sets $E_i, i \in I$ of any mutually exclusive events $E_i \subseteq \Omega$ where the set I is countable , $P(\cup_i E_i) = \sum_i P(E_i)$.*

In this definition, Properties 1–3 imply that, for every $E \subseteq \Omega$, $0 \leq P(E) \leq 1$. In particular, $P(E) = 0$ if $E = \emptyset$. For this reason, event $\emptyset$ is called impossible. Because of Property 2, space Ω is also called the certain event.

We notice that P is a measure on Ω with finite total measure, i.e. P is a set function on the subsets of Ω. There are certain technical requirements on the sets E_i, they must form what is called a $\sigma-$field or $\sigma-$ring. This requirement is necessary since there may be sets for which the probability is not defined.

From the axioms of probability theory one may derive a number of useful auxiliary formulae. For any two sets A and B

$$A = A \cap \Omega = A \cap (B + \overline{B}) = A \cap B + A \cap \overline{B}$$

and since $(A \cap B) \cap (A \cap \overline{B}) = \emptyset$ this implies

$$P(A) = P(AB) + P(A\overline{B})$$

Also

$$P(A + B) = P(A) + P(B) - P(AB)$$

With these results one can solve many problems dealing with finite sample spaces, i.e. when Ω contains only a finite number of points. However finite sample spaces will not suffice for the development of the theory of Brownian motion.

3.2 Conditional probability and independence.

So far we have been concerned with the probability of occurence of certain events without regard to what other events may have occured. We therefore introduce the concept of *conditional probability* as:

Definition 3.2 *For any two events $E_1, E_2 \subseteq \Omega$ such that $P(E_1) \neq 0$,*

$$P(E_2|E_1) = P(E_1 \cap E_2)/P(E_1)$$

is called the conditional probability of event E_2 under the condition that event E_1 occurs.

A conditional probability is a genuine probability, it satisfies all of the axioms in the previous section. The sample space is however different than Ω.

Example 3.2 *As an elementary example suppose you throw a fair dice. The probability that the upper face bears more than three dots is $1/2$. The probability that the face bears more than three dots if it is known that the number of dots is even is $2/3$. In the first case the sample space is $\{1, 2, \ldots, 6\}$; in the second case it is $\{2, 4, 6\}$.*

We now turn to the concept of independent events. Qualitatively we say that two events are independent if the occurence of one has no influence on the occurence of the other. The events E_1 and E_2 are independent if

$$P(E_1|E_2) = P(E_1), \quad P(E_2) \neq 0. \tag{3.1}$$

Alternatively

$$P(E_1 \cap E_2) = P(E_1)P(E_2). \tag{3.2}$$

The concept of conditional probability and independence can be extended to more than two events. Consider three events E_1, E_2 and E_3. The conditional probability of E_3 given E_1 and E_2 i.e. given $E_1 \cap E_2 = E_1E_2$ is defined by

$$P(E_1E_2E_3) = P(E_3|E_1E_2)P(E_1E_2).$$

The three events are independent if

$$P(E_1E_2E_3) = P(E_1)P(E_2)P(E_3)$$

and, in addition, if the events are pairwise independent i.e.

$$P(E_jE_k) = P(E_j)P(E_k) \quad j, k = 1 - 3, j \neq k.$$

3.3 Random variables and probability distributions

In elementary examples it is in general possible to describe the events of the sample space in detail. In more complicated, and in particular more realistic, situations this is awkward, or just not possible. We do however often know the results of some numerical measurement carried our on the system.

Example 3.3 *For example, suppose we have a box of electrical resistors, all nominally of the sam resistance. Because of the vagaries of the manufactoring process, the actual resistancees will differ from each other and from the nominal resistance. if a resistor is picked from the box at random, simultaneously a resistance value is also picked at random from the set of resistance values of the resistans. The resistance is a real valued function defined on the set of resistors. Its value can be determined by experiment. Thus we have numerically valued function defined on the sample space.*

Random or stochastic variables are variables whose values are determined by the outcome of experiments. The most we can know about a random variable is the probability that a particular value of it will be realized in an experiment.

Any real or complex single valued function X of elementary events ω defined on a sample space Ω is called a random variable. Random variables can also be vector valued. We write the functional dependence as

$$X(\omega) = x$$

Random variables are point functions on Ω such that the inverse mapping of every interval in $\mathcal{R}$ corresponds to an event in Ω. In a given experiment a random variable may have any one of a number of values. Therefore one must distinguish a random varaible X from its possible realizations $\{x_i\}$.

Since we often do not know very much about the detailed structure of the sample space, it is usually not possible to specify the form of a random variable in an analytic way. But there is a convenient way to characterize random variables analytically. We first discuss this for the case that the random variable is a real valued function. Let us define a function $F_X(x)$ by

$$F_X(x) = P(X \leq x) \tag{3.3}$$

The notation $P(X \leq x)$ means the probability that the random variable X has a value less than or equal to the number x. It is defined as the probability of the set of events $\{\omega\}$ such that $X(\omega) \leq x$,

$$P(X \leq x) = \Pr\{\omega \in \Omega, X(\omega) \leq x\}$$

Clearly F_X has the following properties

i) $F(-\infty) = 0$,

ii) $F(\infty) = 1$,

iii) F is a non-decreasing function of x

$F_X(x)$ is called the cumulative distribution function. The following theorem about the structure of cumulative distribution functions has been proved:

Theorem 3.1 *Every cumulatative distribution function can be written as the sum of three parts: an absolutely continous part $F_{\rm ac}$, an atomic part $F_{\rm a}$, and a singular but continous part $F_{\rm s}$. This means that*

$$F_{\rm ac}(x) = \int_{-\infty}^{x} {\rm d}y\, \varrho(y)$$

where $\varrho(x) \geq 0$ is some integrable function. The atomic part has jumps at a denumerable number of points $\{x_k\}$ and is constant between the points. It can be written as a sum of Dirac delta functions

$$F_{\mathrm{a}}(x) = \int_{-\infty}^{x} \mathrm{d}y \sum_k p_k \delta(x - x_k)$$

The singular continous part F_{s} is continous everywhere but has its points of increase concentrated on set of measure zero probability.

The function $\varrho(y)$ is a probability density where $\varrho(y)\mathrm{d}y$ = prob. to find X in the interval $(y, y + \mathrm{d}y)$. For a cumulative distribution with an absolutely continous part and an atomic part we have the density

$$\varrho_X(x) = \varrho(x) + \sum_k p_k \delta(x - x_k) \tag{3.4}$$

Since $F_X(\infty) = 1$ we must have

$$\int_{-\infty}^{\infty} \mathrm{d}y\, \varrho(y) + \sum_k p_k = 1 \tag{3.5}$$

We note that

$$\varrho_X(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_X(x) \tag{3.6}$$

Often we wish to find the probability density not for the stochastic variable X, but for some new stochastic variable

$$Y = H(X)$$

where $H(x)$ is a known function of X. The probability density $\varrho_Y(y)$ for the stochastic variable Y is defined as

$$\varrho_Y(y) = \int_{-\infty}^{\infty} \mathrm{d}x\, \delta\left(y - H(x)\right) \varrho_X(x)$$

where $\delta(y - H(x))$ is the Dirac delta function.

Then with

$$u = H(x), \quad \mathrm{d}u = H'(x)\mathrm{d}x$$

we have

$$\varrho_Y(y) = \int_{H(-\infty)}^{H(\infty)} \mathrm{d}u\, \delta(y - u) \frac{\varrho_X(H^{-1}(u))}{H'(H^{-1}(u))} = \frac{1}{H'\left[H^{-1}(y)\right]} \varrho_X\left[H^{-1}(y)\right] = \frac{\varrho_X(x)}{H'(x)}. \tag{3.7}$$

3.4 Expectations

The distribution function of a random variable contains all possible information about the variable, and may be difficult to compute. In practice we have only limited information of the distrubution such as the central tendency and its spread etc.

The n'th moment of X is defined as

$$\langle X^n \rangle = \int_{-\infty}^{\infty} \mathrm{d}x x^n \varrho_X(x) \tag{3.8}$$

The first moment $\mu = \langle X \rangle$ is the mean value of X. Here every possible value of X is weigted by the probability of its occurence and the results added. Thus x-values of high probability are more important than those of low probability in determining the expectation. Consequently $\langle X \rangle$ usually gives an indication where the highest probability values of X are likely to be.

There are two other measures of central tendency that are sometimes used: the median and the mode. The median is that value of x for which $F_X(x) \geq 1/2$, i.e. half of the cases will have values greater than x, half less. The mode is the value of x for which $\varrho(x)$ is a maximum. There may be more than one mode for a given distribution; such distributions are called multimodal.

A useful characterization of the width or spread of a distribution would be the mean of the distribution from the mean $\langle |X - \langle X \rangle| \rangle$. However this quantity is difficult to handle analytically, and a more useful measure of the width is the variance defined by:

$$\sigma^2 = \langle (X - \langle X \rangle)^2 \rangle = \int \mathrm{d}x\, x^2 \varrho_X(x) - \left[\int \mathrm{d}x x \varrho_X(x) \right]^2 \tag{3.9}$$

This quantuty is always positive and vanishes only when everything is concentrated at the mean. The *standard deviation* is defined as σ, the square root of the variance.

The central moments are generally defined by

$$\mu_j = \langle (X - \langle X \rangle)^j \rangle \tag{3.10}$$

Moments of order higher than the second are often difficult to obtain.

3.5 Characteristic function of a random variable

A useful function for studying a distribution function is its Fourier transform

$$f_X(k) = \langle \mathrm{e}^{\mathrm{i}kX} \rangle = \int_{-\infty}^{\infty} \mathrm{d}x\, \mathrm{e}^{\mathrm{i}kx} \varrho_X(x) \tag{3.11}$$

This function is called the *characteristic function*, and it is uniquely defined by the probability distribution function. Because ϱ_X is a probability density and hence nonnegative and integrable from $-\infty$ to ∞, $f_X(k)$ always exists. Furtermore $\varrho_X(x)$ is determined uniquely by its characteristic function. The inversion formula of (3.11) reads

$$\varrho_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}k\, \mathrm{e}^{-\mathrm{i}kx} f_X(k) \tag{3.12}$$

For the atomic parts of ϱ_X we use the formulae

$$\begin{aligned} \int_{-\infty}^{\infty} \mathrm{d}x\, \mathrm{e}^{\mathrm{i}kx} \delta(x - x_0) &= \mathrm{e}^{\mathrm{i}kx_0} \\ \int_{-\infty}^{\infty} \mathrm{d}k\, \mathrm{e}^{\mathrm{i}kx_0} \mathrm{e}^{-\mathrm{i}kx} &= 2\pi\delta(x - x_0) \end{aligned}$$

Characteristic functions are continous functions of k and have the property that

$$f_X(0) = 1; \quad |f_X)k)| \leq 1; \quad f_X(-k) = f_X^*(k) \tag{3.13}$$

The product of two characteristic functions is always a characteristic function.

We can write (3.11) as a Taylor series

$$f_X(k) = \sum_{n=0}^{\infty} \frac{(\mathrm{i}k)^n}{n!} \langle X^n \rangle \tag{3.14}$$

i.e.

$$\langle X^n \rangle = (-\mathrm{i})^n \left(\frac{\mathrm{d}^n}{\mathrm{d}k^n} f_X(k) \right)_{k=0} \tag{3.15}$$

It is often useful to write the characteristic function $f_X(k)$ in terms of cumulants c_n. Let

$$g(k) = \ln f(k) = \sum_{n=1}^{\infty} \frac{(\mathrm{i}k)^n}{n!} c_n \tag{3.16}$$

Then

$$c_n = -\mathrm{i} \left(\frac{\mathrm{d}^n}{\mathrm{d}k^n} g(k) \right)_{k=0} \tag{3.17}$$

This yields

$$\begin{aligned} c_1 &= -\mathrm{i} \left(\frac{f'(k)}{f(k)} \right)_{k=0} = \langle X \rangle \\ c_2 &= -\left(\frac{\mathrm{d}^2}{\mathrm{d}k^2} g(k) \right)_{k=0} = -\left[\frac{f''(k)}{f(k)} - \frac{f'(k)^2}{f^2(k)} \right]_{k=0} = \langle X^2 \rangle - \langle X \rangle^2 = \sigma^2 \end{aligned} \tag{3.18}$$

Example 3.4 Gaussian distribution. *For the Gaussian distribution we find*

$$\begin{aligned} f_X(k) &= \left(\frac{1}{2\pi\sigma^2} \right)^{1/2} \int_{-\infty}^{\infty} \mathrm{d}x \mathrm{e}^{\mathrm{i}kx} \mathrm{e}^{-(x-\mu)^2/2\sigma^2} = \mathrm{e}^{\mathrm{i}k\mu} \left(\frac{1}{2\pi\sigma^2} \right)^{1/2} \int_{-\infty}^{\infty} \mathrm{d}z \mathrm{e}^{\mathrm{i}kz} \mathrm{e}^{-z^2/2\sigma^2} \\ &= \mathrm{e}^{\mathrm{i}k\mu - k^2/2\sigma^2} \left(\frac{1}{2\pi\sigma^2} \right)^{1/2} \int_{-\infty}^{\infty} \mathrm{d}z \exp\left[-\frac{1}{2} \left(z/\sigma - \mathrm{i}k\sigma \right)^2 \right] = \mathrm{e}^{\mathrm{i}k\mu - k^2/2\sigma^2} \end{aligned}$$

This is the characteristic function for a Gaussian random variable. When the Gaussian variable is centerd, $f_X(k)$ assumes the form $\exp(-k^2\sigma^2/2)$.

Example 3.5 Poisson distribution. *The characteristic function for the Poisson distribution is given by*

$$f_X(k) = \langle \mathrm{e}^{\mathrm{i}kX} \rangle = \sum_{n=0}^{\infty} \mathrm{e}^{\mathrm{i}kn} \frac{\lambda^n}{n!} \mathrm{e}^{-\lambda} = \mathrm{e}^{-\lambda} \sum_{n=0}^{\infty} \frac{1}{n!} \left(\lambda \mathrm{e}^{\mathrm{i}k} \right)^n = \mathrm{e}^{\lambda\left(\mathrm{e}^{\mathrm{i}k} - 1\right)}$$

3.6 Distributions of several random variables

We consider now a system which involves more than one random variable. Suppose that the random variables $X_1, X_2, \ldots, X_n$ are defined as functions on the same space, as happens, for example if each X_i is the three dimensional position vector $\boldsymbol{r}$ of a gas of molecule in a container. The functional form now reads

$$\begin{pmatrix} X_1(\omega) \\ \vdots \\ X_n(\omega) \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \tag{3.19}$$

The joint distribution function for the stochastic variables $X_1, X_2, \ldots, X_n$ are defined by

$$F_{X_1X_2\cdots X_n}(x_1, x_2, \ldots, x_n) = \Pr\{X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n\} \tag{3.20}$$

where $\{X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n\} = \{X_1 \le x_1\} \cap \{X_2 \le x_2\} \cdots \cap \{X_n \le x_n\}$ and the probability is taken for the values of elementary events ω that satisfy the inequalities in (3.20). In other words $F_{X_1,\cdots,X_n}(x_1, \ldots, x_n)$ is the probability that simultaneously the stochastic variables X_i have values in the intervals $\{-\infty < X_i < x_i\}, i = 1, \ldots, n$. If the random variables are continous it will be possible to write

$$F_{X_1X_2\cdots X_n}(x_1, x_2, \ldots, x_n) = \int_{-\infty}^{x_1} \mathrm{d}y_1 \cdots \int_{-\infty}^{x_n} \mathrm{d}y_n \, \varrho_{X_1,\cdots,X_n}(y_1, \ldots, y_n) \tag{3.21}$$

and $\varrho(x_1, \ldots, x_n$ is called the joint probability density function, and is defined as

$$\varrho_{X_1,\cdots,X_n}(x_1, \ldots, x_n) = \frac{\partial^n}{\partial x_1, \cdots \partial x_n} F_{X_1X_2\cdots X_n}(x_1, x_2, \ldots, x_n) \tag{3.22}$$

For the reduced distribution function $F_{X_1X_2\cdots X_{n-1}}(x_1, x_2, \ldots, x_{n-1})$ consistency requires that

$$\begin{aligned} & F_{X_1X_2\cdots X_{n-1}}(x_1, x_2, \ldots, x_{n-1}) = F_{X_1X_2\cdots X_n}(x_1, x_2, \ldots, x_{n-1}, \infty) \\ = & \int_{-\infty}^{x_1} \mathrm{d}y_1 \cdots \int_{-\infty}^{x_{n-1}} \mathrm{d}y_{n-1} \int_{-\infty}^{\infty} \mathrm{d}y_n \, \varrho_{X_1,\cdots,X_n}(y_1, \ldots, y_n) \end{aligned} \tag{3.23}$$

The joint characteristic function is defined as

$$f_{X_1\cdots X_n}(\boldsymbol{k}) = f_{X_1\cdots X_n}(k_1, \ldots, k_n) = \langle \exp(\mathrm{i}\boldsymbol{k} \cdot \boldsymbol{X}) \rangle = \langle \exp \mathrm{i}\,(k_1X_1 + k_2X_2 + \cdots + k_nX_n) \rangle \tag{3.24}$$

Here mean values are defined by

$$\langle g(X_1, \ldots, X_n) \rangle = \int \mathrm{d}\boldsymbol{x} \, g(\boldsymbol{x}) \, \varrho(\boldsymbol{x}) \tag{3.25}$$

where $\int \mathrm{d}\boldsymbol{x} = \int_{-\infty}^{\infty} \mathrm{d}x_1 \cdots \int_{-\infty}^{\infty} \mathrm{d}x_n$.

When the random variables are independent we have from (3.2) and (**??**)

$$\begin{aligned} & \Pr\{\omega : X_1(\omega) \le x_1, X_2(\omega) \le x_2, \ldots, X_n(\omega) \le x_n\} \\ = & \Pr\{\omega : X_1(\omega) \le x_1\} \Pr\{\omega : X_2(\omega) \le x_2\} \cdots \Pr\{\omega : X_n(\omega) \le x_n\} \end{aligned} \tag{3.26}$$

This implies that for independent variables

$$F_{X_1X_2\cdots X_{n-1}}(x_1, x_2, \ldots, x_{n-1}) = \prod_{i=1}^{n} F_{X_i}(x_i) \tag{3.27}$$

The corresponding density is

$$\varrho_{X_1\cdots X_n}(x_1,\ldots,x_n)=\prod_{-=1}^{n}\varrho_{X_i}(x_i) \tag{3.28}$$

and from (3.24)

$$f_{X_1\cdots X_n}(k_1,\ldots,k_n)=\prod_{i=1}^{n}f_{X_i}(k_i) \tag{3.29}$$

Multidimensional random variables can be analyzed in terms of moments as before, and the most important of these are the expectation vector and the variance matrix. The expectation vector $\mu_X=(\mu_1,\mu_2,\ldots,\mu_n)$ of the random variable $X=(X_1,X_2,\ldots,X_n)$ is

$$\mu_X=\int \mathrm{d}\boldsymbol{x}\,\boldsymbol{x}\,\rho_X(\boldsymbol{x}) \tag{3.30}$$

The variance matrix $\boldsymbol{C}$ of n random variables $X_1,X_2,\ldots,X_n$ is defined as the $n\times n$ matrix whise ij-element is

$$C_{ij}=\langle(X_i-\mu_i)(X_j-\mu_j)\rangle=\int \mathrm{d}\boldsymbol{x}\,(x_i-\mu_i)(x_j-\mu_j)\,\varrho_X(\boldsymbol{x}) \tag{3.31}$$

For any random variables, the variance matrix C is symmetric and so-called non-negative definite, i.e. all of its egenvalues are non-negative. Every diagonal entry of matrix C is non-negative and is called the variance of the i th entry. The nondiagonal element C_{ij} is called the covariance of X_i and X_j. The random variables X_i and X_j, $i\neq j$, are called uncorrelated if and only if C_{ij} is zero.

Let X_1 and X_2 be two continous random variables with probability density functions $\varrho_{X_1}(x_1)$ and $\varrho_{X_2}(x_2)$ respectively and let $\varrho_{X_1X_2}(x_1,x_2)$ be their joint density function. Suppose given $X_1=x_1$, $\varrho_{X_2|X_1}(x_2|x_1)\mathrm{d}x_2$ is the probability that one finds X_2 in the interval $(x_2,x_2+\mathrm{d}x_2)$. Then $\varrho_{X_2|X_1}$ is called the conditional probability density. Now

$$\varrho_{X_1X_2}(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2=\varrho_{X_2|X_1}(x_2|x_1)\mathrm{d}x_2\times\varrho_{X_1}(x_1)\mathrm{d}x_1 \tag{3.32}$$

and therefore

$$\varrho_{X_2|X_1}(x_2|x_1)=\frac{\varrho_{X_1X_2}(x_1,x_2)}{\varrho_{X_1}(x_1)} \tag{3.33}$$

provided $\varrho_{X_1}(x_1)\neq 0$. In general we can define conditional probabilities for arbitrary variables as

$$\varrho_{X_{k+1}\cdots X_n|X_1\cdots X_k}(x_{k+1},\ldots,x_n|x_1,\ldots,x_k)=\frac{\varrho_{X_1\cdots X_n}(x_1,\ldots,x_n)}{\varrho_{X_1\cdots X_k}(x_1,\ldots,x_k)} \tag{3.34}$$

3.7 Joint normal distribution

One of the most important distribution is the multivariate Gaussian distribution. Let us consider n Gaussian variables $X_1,X_2,\ldots,X_n$, let $\langle X_i\rangle=\mu_i$ and let $\boldsymbol{C}$ denote the variance matrix defined above

$$C_{ij}=\langle(X_i-\mu_i)(X_j-\mu_j)\rangle$$

We write $\boldsymbol{x}-\mu$ for the column vector with elements $x_i-\mu_i, i=1,\ldots,n$ and $(\boldsymbol{x}-\mu)^T$ for the corresponding row vector with these elements. Assuming that the determinant of the variance matrix does not vanish we define the joint normal distribution of $X_1, X_2, \ldots, X_n$ as that with the probability density function

$$\varrho_{X_1\cdots X_n}(x_1,x_2,\ldots,x_n)=\left[\frac{\det(\boldsymbol{G})}{(2\pi)^n}\right]^{1/2}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\mu)^T\boldsymbol{G}(\boldsymbol{x}-\mu)\right) \tag{3.35}$$

where $\boldsymbol{G}=\boldsymbol{C}^{-1}$ is the inverse of the variance matrix. Here

$$(\boldsymbol{x}-\mu)^T\boldsymbol{G}(\boldsymbol{x}-\mu)=\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i-\mu_i)G_{ij}(x_j-\mu_j)$$

Since $\boldsymbol{C}$ is a real symmetric matrix $\boldsymbol{G}$ is also real and symmetric. It can therefore be diagonalized using an orthogonal transformation $\boldsymbol{Q}$ where

$$\boldsymbol{Q}^{-1}=\boldsymbol{Q}^T \quad \boldsymbol{Q}^{-1}\boldsymbol{Q}=1$$

Then we have

$$\boldsymbol{Q}\boldsymbol{G}\boldsymbol{Q}^T=\boldsymbol{D}$$

where $\boldsymbol{D}$ is a diagonal matrix with the eigenvalues g_i of G on the diagonal. Let

$$\boldsymbol{u}=\boldsymbol{Q}(\boldsymbol{x}-\mu)$$

Since $\det(\boldsymbol{Q})=\det(\boldsymbol{Q}^T)=1$ the Jacobian J of this transformation is one and we have

$$\begin{aligned}\int \mathrm{d}\boldsymbol{x}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\mu)^T\boldsymbol{G}(\boldsymbol{x}-\mu)\right)&=\int \mathrm{d}\boldsymbol{u}\exp\left(-\frac{1}{2}\boldsymbol{u}^T\boldsymbol{Q}\boldsymbol{G}\boldsymbol{Q}^T\boldsymbol{u}\right)\\ =\quad\int \mathrm{d}\boldsymbol{u}\exp\left(-\frac{1}{2}\sum_{i=1}^{n}g_iu_i^2\right)&=\sqrt{\frac{(2\pi)^n}{g_1\cdots g_n}}=\sqrt{\frac{(2\pi)^n}{\det(\boldsymbol{G})}}\end{aligned}$$

which show that the distribution is properly normalized.

The characteristic function is

$$\begin{aligned}f_{X_1\cdots X_n}(k_1,\ldots,k_n)&=\sqrt{\frac{\det(\boldsymbol{G})}{(2\pi)^n}}\int \mathrm{d}\boldsymbol{x}\,\mathrm{e}^{\mathrm{i}\boldsymbol{k}^T\boldsymbol{x}}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\mu)^T\boldsymbol{G}(\boldsymbol{x}-\mu)\right)\\ =\quad \mathrm{e}^{\mathrm{i}\boldsymbol{k}^T\mu}\sqrt{\frac{\det(\boldsymbol{G})}{(2\pi)^n}}\int \mathrm{d}\boldsymbol{u}\exp\left(\mathrm{i}\boldsymbol{k}^T\boldsymbol{Q}^T\boldsymbol{u}-\frac{1}{2}\sum_{i=1}^{n}g_iu_i^2\right)\\ =\quad \mathrm{e}^{\mathrm{i}\boldsymbol{k}^T\mu}\sqrt{\frac{1}{(2\pi)^n}}\int \mathrm{d}\boldsymbol{u}\exp\left(\mathrm{i}\boldsymbol{k}^T\boldsymbol{Q}^T\boldsymbol{D}^{-1/2}\boldsymbol{u}-\frac{1}{2}\sum_{i=1}^{n}u_i^2\right)\end{aligned}$$

Now let $\boldsymbol{v}=\boldsymbol{D}^{-1/2}\boldsymbol{Q}\boldsymbol{k}$ and $\boldsymbol{v}^T=\boldsymbol{k}^T\boldsymbol{Q}^T\boldsymbol{D}^{-1/2}$, then

$$\begin{aligned}f_{X_1\cdots X_n}(k_1,\ldots,k_n)&=\exp\left(\mathrm{i}\boldsymbol{k}^T\mu-\frac{1}{2}\boldsymbol{v}^T\boldsymbol{v}\right)\sqrt{\frac{1}{(2\pi)^n}}\int \mathrm{d}\boldsymbol{u}\exp\left(-\frac{1}{2}\sum_{i=1}^{n}(u_i-\mathrm{i}v_i)^2\right)\\ &=\exp\left(\mathrm{i}\boldsymbol{k}^T\mu-\frac{1}{2}\boldsymbol{v}^T\boldsymbol{v}\right)=\exp\left(\mathrm{i}\boldsymbol{k}^T\mu-\frac{1}{2}\boldsymbol{k}^T\boldsymbol{G}^{-1}\boldsymbol{k}\right)\end{aligned}$$

The last step follows since

$$\boldsymbol{v}^T\boldsymbol{v} = \boldsymbol{k}^T\boldsymbol{Q}^T\boldsymbol{D}^{-1/2}\boldsymbol{D}^{-1/2}\boldsymbol{Q}\boldsymbol{k} = \boldsymbol{k}^T\boldsymbol{Q}^T\boldsymbol{D}^{-1}\boldsymbol{Q}\boldsymbol{k} = \boldsymbol{k}^T\boldsymbol{G}^{-1}\boldsymbol{k} = \boldsymbol{k}^T\boldsymbol{C}\boldsymbol{k}$$

Therefore

$$f_{X_1\cdots X_n}(k_1,\ldots,k_n) = \exp\left(\mathrm{i}\boldsymbol{k}^T\mu - \frac{1}{2}\boldsymbol{k}^T\boldsymbol{C}\boldsymbol{k}\right) \tag{3.36}$$

From this we can directly find the moments

$$\langle X_i\rangle = -\mathrm{i}\left(\frac{\partial f}{\partial k_i}\right)_{k=0} = \mu_i$$

and

$$\langle (X_i-\mu_i)(X_j-\mu_j)\rangle = -\left(\frac{\partial^2 f}{\partial k_i \partial k_j}\right)_{k=0} = C_{ij}$$

3.8 Central limit theorem and law of large numbers

We now consider the central limit theorem for a set of independent random variables $X_1, X_2, \ldots, X_n$ Suppose that the set is identically distributed, each variable having a finite mean $\mu < \infty$ and variance $\sigma^2 < \infty$. The characteristic function of X_r is

$$f_{X_r}(k) = \langle \mathrm{e}^{\mathrm{i}kX_r}\rangle = 1 + \mathrm{i}k\mu - \frac{1}{2}k^2(\sigma^2+\mu^2) + (\mathrm{i}k)^3\lambda(k) \tag{3.37}$$

where $\lambda(k) < \infty, k \to 0$. For the centered variable $Y_r = X_r - \mu$ we similarly find

$$f_{Y_r}(k) = 1 - \frac{1}{2}k^2\sigma^2 + (\mathrm{i}k)^3\lambda_1(k) \tag{3.38}$$

For the scaled variable αY_r we have

$$f_{\alpha Y_r}(k) = \langle \mathrm{e}^{\mathrm{i}k\alpha Y_r}\rangle = f_{Y_r}(\alpha k) = 1 - \frac{1}{2}(\alpha k)^2\sigma^2 + (\mathrm{i}\alpha k)^3\lambda_1(\alpha k) \tag{3.39}$$

Consider now the variable

$$Y = \sum_{i=1}^{n}\frac{Y_i}{\sqrt{n}} \tag{3.40}$$

which is scaled with $\alpha = 1/\sqrt{n}$. Then

$$\begin{aligned} f_Y(k) &= \langle \mathrm{e}^{\mathrm{i}kY}\rangle = \langle \exp\left[\mathrm{i}k\sum_{i=1}^{n}\frac{Y_i}{\sqrt{n}}\right]\rangle = \prod_{i=1}^{n} f_{\frac{1}{\sqrt{n}}Y_i}(k) \\ &= \left[1 - \frac{k^2}{2n}\sigma^2 + \left(\frac{\mathrm{i}k}{\sqrt{n}}\right)^3\lambda_1\left(\frac{k}{\sqrt{n}}\right)\right]^n = \left[1 - \frac{k^2\left[\sigma^2 + \frac{\mathrm{i}2k}{\sqrt{n}}\lambda_1\left(\frac{k}{\sqrt{n}}\right)\right]}{2n}\right]^n \\ &\to \exp\left(-\frac{1}{2}k^2\left[\sigma^2 + \frac{\mathrm{i}2k}{\sqrt{n}}\lambda_1\left(\frac{k}{\sqrt{n}}\right)\right]\right) = \mathrm{e}^{-\frac{1}{2}k^2\sigma^2} \quad n\to\infty \end{aligned} \tag{3.41}$$

This is the characteristic function of a Gaussian distribution with mean value zero and variance σ^2. The random variable

$$Y = \sum_{i=1}^{n} \frac{(X_i - \langle X_i \rangle)}{\sqrt{n}}$$

therefore becomes a Gaussian variable as $n \to \infty$. This is a formulation of the *central limit theorem.*

The law of large numbers applies to n independent experiments and may be stated as follows:

> *If an avent A has a propability p of occuring then the fraction of outcomes of A approaches p in the limit* $n \to \infty$.

The proof has two steps. The first involves the derivation of the Chebycheff inequality. With $\mu = \langle X \rangle$, the variance of a stochastic variable X is

$$\sigma_X^2 = \int_{-\infty}^{\infty} \mathrm{d}x (x-\mu)^2 \varrho_X(x)$$

If we now delete the range of variables x for which $|x - \mu| \leq \epsilon$ we can write

$$\begin{aligned} \sigma_X^2 &\geq \int_{-\infty}^{\mu-\epsilon} \mathrm{d}x (x-\mu)^2 \varrho_X(x) + \int_{\mu+\epsilon}^{\infty} \mathrm{d}x (x-\mu)^2 \varrho_X(x) \\ &\geq \epsilon^2 \left[\int_{-\infty}^{\mu-\epsilon} \mathrm{d}x \varrho_X(x) + \int_{\mu+\epsilon}^{\infty} \mathrm{d}x \varrho_X(x) \right] = \epsilon^2 \mathrm{Pr}\left\{ |X - \mu| \geq \epsilon \right\} \end{aligned}$$

This gives the Chebycheff inequality

$$\mathrm{Pr}\left\{ |X - \mu| \geq \epsilon \right\} \leq \frac{\sigma_X^2}{\epsilon^2}$$

Let's now consider n independent measurements of the stochastic variable X. Let Y_n be the mean value of the outcomes

$$Y_n = \frac{1}{n} (X_1 + \cdots + X_n)$$

then $\langle Y_n \rangle = \mu$ and from Chebycheff's inequality

$$\mathrm{Pr}\left\{ |Y_n - \mu| \geq \epsilon \right\} \leq \frac{\sigma_{Y_n}}{\epsilon^2}$$

But for independent events

$$\sigma_{Y_n}^2 = \frac{1}{n} \sigma_X^2$$

i.e.

$$\mathrm{Pr}\left\{ |Y_n - \mu| \geq \epsilon \right\} \leq \frac{\sigma_X^2}{n\epsilon^2} \to 0 \quad n \to \infty$$

3.9 Infinitely divisible and stable distributions

The concept of an infinitely divisible distribution function is important to all limit theorems. The limiting distribution can only belong to this class of distributions. A stochastic variable Y is infinitely divisible if for any integer n it can be represented by a sum

$$Y = X_1 + X_2 + \cdots + X_n$$

of identically distributed stochastic variables X_i, $(i = 1, \ldots, n)$. The distribution function $\varrho_Y(y)$ is infinitely divisible if and only if for any n its characteristic function $f_Y(k)$ is the n'th power of some characteristic function $f_X(k)$. Thus

$$f_Y(k) = [f_X(k)]^n\,; \quad f_X(k) = [f_Y(k)]^{1/n}$$

Example 3.6 Gaussian distribution. *The characteristic function for a Gaussian distribution is*

$$f_Y(k) = \mathrm{e}^{\mathrm{i}k\mu - k^2\sigma_Y^2/2}$$

The n'th root is

$$f_X(k) = [f_Y(k)]^{1/n} = \exp\left(\mathrm{i}k\frac{\mu}{n} - \frac{1}{2}k^2\frac{\sigma_Y^2}{n}\right)$$

which is a Gaussian random variable with mean μ/n and variance σ_Y^2/n.

Example 3.7 Poisson distribution. *The density for a stochastic variable with mean $a + \lambda h$ and variance $h^2\lambda$ is*

$$\varrho_Y(y) = \sum_{m=0}^{\infty} \frac{\lambda^m}{m!}\mathrm{e}^{-\lambda}\delta(y - a - mh)$$

The characteristic function is

$$f_Y(k) = \sum_{m=0}^{\infty} \frac{\lambda^m}{m!}\mathrm{e}^{-\lambda}\mathrm{e}^{\mathrm{i}k(a-mh)} = \exp\left(\mathrm{i}ak + \lambda\left(\mathrm{e}^{\mathrm{i}kh} - 1\right)\right)$$

The n'th root is

$$f_X(k) = \exp\left(\mathrm{i}k\frac{a}{n} + \frac{\lambda}{n}\left(\mathrm{e}^{\mathrm{i}kh} - 1\right)\right)$$

which is again a Poisson distribution with density

$$\varrho_Y(y) = \sum_{m=0}^{\infty} \frac{(\lambda/n)^m}{m!}\mathrm{e}^{-\lambda}\delta\left(y - \frac{a}{n} - mh\right)$$

Example 3.8 Cauchy distribution. *The Cauchy distribution has density*

$$\varrho_Y(y) = \frac{1}{\pi}\frac{a}{a^2 + (y-b)^2}$$

with $\langle Y\rangle = b$ *but* $\sigma_Y^2 = \infty$. *The characteristic function is*

$$\begin{aligned} f_Y(k) &= \frac{1}{\pi}\int_{-\infty}^{\infty} \mathrm{d}y \mathrm{e}^{\mathrm{i}ky}\frac{a}{a^2+(y-b)^2} = \frac{\mathrm{e}^{\mathrm{i}kb}}{\pi}\int_{-\infty}^{\infty}\mathrm{d}u\mathrm{e}^{\mathrm{i}ku}\frac{a}{a^2+u^2} \\ &= \frac{\mathrm{e}^{\mathrm{i}kb}}{\pi}\int_{-\infty}^{\infty}\mathrm{d}u\mathrm{e}^{\mathrm{i}ku}\frac{a}{(a+\mathrm{i}u)(a-\mathrm{i}u)} = \begin{cases} \mathrm{e}^{\mathrm{i}kb-ka} & k>0 \\ \mathrm{e}^{\mathrm{i}kbka} & k<0 \end{cases} = \mathrm{e}^{\mathrm{i}kb-|k|a} \end{aligned}$$

The n'th root is

$$f_X(k) = \exp\left(\mathrm{i}k\frac{b}{n} - |k|\frac{a}{n}\right)$$

which is again a Cauchy distribution with density

$$\varrho_X(x) = \frac{1}{\pi}\frac{a/n}{(a/n)^2+(y-b/n)^2}$$

Stable distributions

The theory of Brownian motion relies on the central limit theorem. The theorem states that the sum of n independent and identically distributed random variables

$$S_n = \sum_{i=1}^{n} X_i$$

obeys a Gaussian distribution in the limit $n\to\infty$, provided the first and second moments of X_i are finite. These restrictions are so mild that many distributions belong to the domain of attraction of the Gaussian.

However, an exception is the Cauchy distribution

$$\varrho(x) = \frac{1}{\pi}\frac{1}{a^2+x^2} \tag{3.42}$$

whose second moment is infinite. The Cauchy distribution occurs in many physical situations, in the Ornstein-Zernike theory of critical opalescence and in the lifetime broadening of spectral lines, for instance. Therefore the question arise of whether it could also emerge as a limiting distribution for S_n, and what the limiting distribution would look like if the random variables were distributed according to (3.42).

The Cauchy distribution is just one example of a whole class of distributions which posess long, inverse-power-law tails

$$\varrho(x) \simeq \frac{1}{|x|^{1+\alpha}}, \quad 0<\alpha<2, \quad |x|\to\infty$$

These broad tails preclude the convergence to the Gaussian for $n\to\infty$, but not the existence of a limiting distribution.

The form and properties of these more general limiting distributions were worked out in the 1930's by P. Lévy, A. Khintchine and others. They are today called Lévy or stable distributions.

Consider a set of random variables $\{X_k\}_{k=1}^n$ which are independent and identically distributed according to

$$\Pr\{x<X_k<x+\mathrm{d}x\} = \varrho(x)\mathrm{d}x, \quad k=1,\ldots,n$$

Then one can ask the following questions:

i) Is it possible to find real constants a_n and b_n so that the distribution of the normalized sum

$$S_n = \frac{1}{b_n}\sum_{k=1}^{n} X_k - a_n \tag{3.43}$$

converges to a limiting distribution if n tends to infinity, i.e.

$$\Pr\{x < S_n < x + \mathrm{d}x\}\ n \overrightarrow{\to} \infty\ L(x)\mathrm{d}x$$

ii) What are the forms and the properties of all possible limiting distributions?

iii) When does the probability density $\varrho(x)$ belong to the domain of attraction of a specific $L(x)$?

The answer to these questions requires the definition of a *stable distribution.*

Definition 3.3 *A probability density is called stable if it is invariant under convolution, i.e. if there are constants a and b such that*

$$\varrho(ax+b) = \int_{-\infty}^{\infty} \mathrm{d}y\, \varrho(a_1(x-y)+b_2)\varrho(a_2 y + b_2) \tag{3.44}$$

for all real constants $a_1 > 0, b_1, a_2 > 0, b_2$.

If the probability distribution for the variable X is $\varrho_X(x)$ then $Y = aX + b$ has the density

$$\begin{aligned}\varrho_Y(y)\mathrm{d}y &= \Pr\{y < Y < y + \mathrm{d}y\} = \Pr\{y < aX + b < y + \mathrm{d}y\} \\ &= \Pr\{(y-b)/a < X < (y-b)/a + \mathrm{d}y/a\} = \varrho_X\left(\frac{y-b}{a}\right)\frac{\mathrm{d}y}{a}\end{aligned}$$

i.e.

$$\varrho_Y(y) = \frac{1}{a}\varrho_X\left(\frac{y-b}{a}\right)$$

For the characterictic function we have

$$f_Y(k) = \int_{-\infty}^{\infty} \mathrm{d}y\, \mathrm{e}^{\mathrm{i}ky}\varrho_y(y) = \frac{1}{a}\int_{-\infty}^{\infty} \mathrm{d}y\, \mathrm{e}^{\mathrm{i}ky}\varrho_X\left(\frac{y-b}{a}\right) = \int_{-\infty}^{\infty} \mathrm{d}u \mathrm{e}^{\mathrm{i}k(au+b)}\varrho_X(u) = \mathrm{e}^{\mathrm{i}kb} f_X(ka)$$

Therefore the convolution in (3.44) implies

$$\mathrm{e}^{\mathrm{i}kb} f_X(ka) = \mathrm{e}^{\mathrm{i}kb_1} f_X(ka_1)\mathrm{e}^{\mathrm{i}kb_2} f_X(ka_2)$$

Example 3.9 Gaussian distribution. *A Gaussian distribution satisfies this relation and is therefore stable. Since for a Gaussian variable*

$$f_X(k) = \mathrm{e}^{\mathrm{i}k\mu - k^2\sigma^2/2}$$

we have

$$\mathrm{e}^{\mathrm{i}kb}\mathrm{e}^{\mathrm{i}ka\mu-(ka)^2\sigma^2/2} = \mathrm{e}^{\mathrm{i}kb_1}\mathrm{e}^{\mathrm{i}ka_1\mu-(ka_1)^2\sigma^2/2}\mathrm{e}^{\mathrm{i}kb_2}\mathrm{e}^{\mathrm{i}ka_2\mu-(ka_2)^2\sigma^2/2}$$

The constants a and b must therefore be determined such that

$$\begin{aligned} b + a\mu &= b_1 + a_1\mu + b_2 + a_2\mu \\ a^2 &= a_1^2 + a_2^2 \end{aligned}$$

so that

$$\begin{aligned} a &= \sqrt{a_1^2 + a_2^2} \\ b &= b_1 + b_2 + \mu(a_1 + a_2 - \sqrt{a_1^2 + a_2^2}) \end{aligned}$$

Similarly one can show that the Cauchy distribution is stable, i.e. the convolution of two Cauchy densities is again a Cauchy density. The general case solved by Lévy and Khnitchin is summarized in the following theorem:

Theorem 3.2 *(***Canonical representation***.) A probability density $L_{\alpha,\beta}(x)$ is stable iff the logarithn of its characteristic function*

$$L_{\alpha,\beta}(k) = \langle \mathrm{e}^{\mathrm{i}kX} \rangle = \int_{-\infty}^{\infty} \mathrm{d}x\, \mathrm{e}^{\mathrm{i}kx} L_{\alpha,\beta}(x)$$

reads

$$\ln L_{\alpha,\beta}(k) = \mathrm{i}\gamma k - c|k|^{\alpha}\left(1 + \mathrm{i}\beta\frac{k}{|k|}\omega(k,\alpha)\right)$$

where γ, c, α and β are real constants taking the values: γ arbitrary, $c \geq 0$

$$0 < \alpha < 2, \quad -1 \leq \beta \leq 1$$

and the function $\omega(k,\alpha)$ is given by

$$\omega(k,\alpha)\emptyset \begin{cases} \tan\left(\frac{\pi\alpha}{2}\right) & \alpha \neq 1 \\ \frac{2}{\pi}\ln|k| & \alpha = 1 \end{cases}$$

The constants γ and c are scale factors. Replacing $x - \gamma$ with $c^{1/\alpha}x$ shifts the origin and rescales the abscissa, but does not alter the function $L_{\alpha,\beta}$ (unless $\alpha = 1, \beta \neq 0$). In contrast α and β define the shape and the properties of $L_{\alpha,\beta}(x)$.

The parameter α characterizes the large x behaviour of $L_{\alpha,\beta}$ and determines which moments exist:

i) $0 < \alpha < 2$: Each stable distribution behaves as

$$L_{\alpha,\beta}(x) \simeq \frac{1}{|x|^{1+\alpha}}, \quad x \to \pm\infty$$

and

$$\langle |x|^{\delta} \rangle = \int_{-\infty}^{\infty} \mathrm{d}x\, |x|^{\delta} L_{\alpha,\beta}(x) < \infty, \quad \text{if} \quad 0 < \delta < \alpha$$

In particular, the latter property implies that the variance does not exist if $\alpha < 2$ and that both mean value and variance do not exist if $\alpha < 1$.

ii) $\alpha = 2$: $L_{\alpha,\beta}(x)$ is independent of β since $\omega(k,2) = 0$, and L is therefore Gaussian.

iii) $\beta = 0$: $L_{\alpha,\beta}(x)$ is an even function of x.

iv) $\beta = \pm 1$: $L_{\alpha,\beta}(x)$ exhibits a pronounced asymmetry for some choices of α. For instance if $0 < \alpha < 1$ its support lies in the intervals $(-\infty, \gamma]$ for $\beta = 1$ and $[\gamma, \infty)$ for $\beta = -1$.

The next theorem answers the question which probability densities $\varrho(x)$ has a $L_{\alpha,\beta}$ as limit distribution.

Theorem 3.3 *The probability density $\varrho(x)$ belongs to the domain of attraction of a stable density $L_{\alpha,\beta}(x)$ with characteristic exponent $\alpha\,(0 < \alpha < 2)$ if and only if*

$$\varrho(x) \simeq \frac{\alpha a^\alpha c_\pm}{|x|^{1+\alpha}}, \quad x \to \pm\infty \tag{3.45}$$

where $c_+ \geq 0, c_- \geq 0$ and $a > 0$ are constants.

These constants are directly related to the prefactor c and the asymmetry parameter β by

$$\begin{aligned} c &= \begin{cases} \frac{\pi(c_+ + c_-)}{2\alpha\Gamma(\alpha)\sin\frac{\pi\alpha}{2}} & \alpha \neq 1 \\ \frac{\pi}{2}(c_+ + c_-) & \alpha = 1 \end{cases} \\ \beta &= \begin{cases} \frac{(c_- - c_+)}{c_+ + c_-)} & \alpha \neq 1 \\ \frac{(c_+ - c_-)}{c_+ + c_-)} & \alpha = 1 \end{cases} \end{aligned}$$

Furthermore if $\varrho(x)$ belongs to the domain of attraction of a stable distribution its absolute moments of order δ exist for $\delta < \alpha$:

$$\langle |x|^\delta \rangle = \int_{-\infty}^{\infty} \mathrm{d}x |x|^\delta \varrho(x) = \begin{cases} < \infty & 0 \leq \delta < \alpha \\ \infty & \delta > \alpha \end{cases}$$

and the normalization constant in (3.43) which characterizes the typical scaling behaviour of S_n is given by

$$b_n = a n^{1/\alpha}$$

so that

$$\lim_{n\to\infty} \Pr\left\{ x < \frac{1}{a n^{1/\alpha}} \sum_{k=1}^{n} X_k - a_n < x + \mathrm{d}x \right\} = L_{\alpha,\beta}(x)\mathrm{d}x$$

where a is the same constant as in (3.45) and

$$\begin{aligned} 0 < \alpha < 1: &\quad a_n = 0 \\ 1 < \alpha < 2: &\quad a_n b_n = n\langle X \rangle \end{aligned}$$

Examples

Example 3.10 *Waiting time distribution with long tail. For a random walker with a distribution of jump lengths $f(\boldsymbol{r})$ and waiting times between jumps $\psi(t)$ the probability density $p(\boldsymbol{r}, t)$ for being at position $\boldsymbol{r}$ at time t is given by (2.72) i.e.*

$$p(\boldsymbol{q}, z) = \frac{1 - \hat{\psi}(z)}{z} \frac{1}{1 - \hat{\psi}(z) f(\boldsymbol{q})} \tag{3.46}$$

When $\psi(t)$ and $f(\boldsymbol{r})$ both have finite mean values and variance this leads to the diffusion result in the limit $\boldsymbol{q} \to 0, z \to 0$.

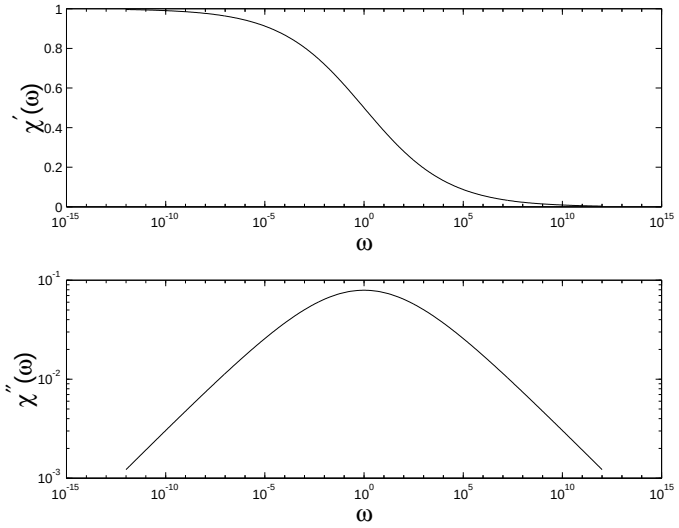

Figure 3.1: *Plot of the real and imaginary parts of the Cole-Cole function.*

Assume now that $\psi(t)$ has a long tail

$$\psi(t) = \frac{\alpha A}{\Gamma(1-\alpha)} \frac{1}{t^{1+\alpha}}, \quad 0 < \alpha < 1, \quad t \to \infty$$

so that $\langle t \rangle = \infty$. Such long waiting times may occure in complex systems with many low lying metastable states. During relaxation towards equilibrium the system tries to find the lowest minimum of the free energy. A particle can be trapped in a low lying state for a long time until it eventually gets sufficient energy to jump over a barrier and finds another metastable state and so on until equilibrium is reached.

For the Laplace transform

$$\hat{\psi}(z) = \int_0^\infty \mathrm{d}t \, \mathrm{e}^{-zt} \psi(t)$$

there is now no expansion in a Taylor series with powers z, z^2 etc. However

$$\frac{\mathrm{d}\hat{\psi}(z)}{\mathrm{d}z} = -\int_0^\infty \mathrm{d}t \, \mathrm{e}^{-zt} t \psi(t)$$

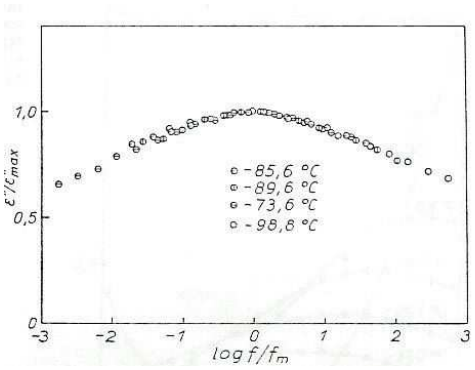

Figure 3.2: *The absorption curve of polyethylene adiphate for various temperatures. The curves are superimposed on a master curve when scaled with the peak value.*

The factor t multiplying $\psi(t)$ implies that long times are dominating the integral and we can approximate ψ with its behaviour for long times

$$\frac{\mathrm{d}\hat{\psi}(z)}{\mathrm{d}z} \approx -\frac{\alpha A}{\Gamma(1-\alpha)} \int_0^\infty \mathrm{d}t\, \mathrm{e}^{-zt} \frac{1}{t^\alpha} = -\frac{\alpha A}{\Gamma(1-\alpha)} z^{\alpha-1} \int_0^\infty \mathrm{d}u\, \mathrm{e}^{-u} u^{-\alpha} = -\alpha A z^{\alpha-1}$$

where we used the integral representation of the gamma-function

$$\Gamma(x) = \int_0^\infty \mathrm{d}u\, \mathrm{e}^{-u} u^{x-1}$$

Since $\hat{\psi}(0) = 1$ we find

$$\hat{\psi}(z) = 1 - Az^\alpha + \ldots$$

Introducing this into (3.46) gives

$$p(\boldsymbol{q}, z) = \frac{Az^{\alpha-1}}{az^\alpha f(\boldsymbol{q}) + (1 - f(\boldsymbol{q}))} \tag{3.47}$$

The corresponding susceptibility or response function is

$$\chi(\boldsymbol{q}, z) = f^{-1}(\boldsymbol{q}) - zp(\boldsymbol{q}, z) = \frac{(1 - f(\boldsymbol{q}))/f(\boldsymbol{q})}{Az^a f(\boldsymbol{q}) + (1 - f(\boldsymbol{q})}$$

and has a so called Cole-Cole form. Such spectra are often seen in experiments on complex systems.

We can invert (3.47) in time if we write

$$p(\boldsymbol{q}, z) = \frac{1}{zf(\boldsymbol{q})} \left\{ 1 + \frac{1 - f(\boldsymbol{q})}{Az^\alpha f(\boldsymbol{q})} \right\}^{-1} = \frac{1}{zf(\boldsymbol{q})} \sum_{n=0}^{\infty} (-1)^n \left[\frac{1 - f(\boldsymbol{q})}{Az^\alpha f(\boldsymbol{q})} \right]^n$$

But

$$\int_0^\infty \mathrm{d}t \mathrm{e}^{-zt} t^{n\alpha} = \frac{1}{z^{1+n\alpha}} \int_0^\infty \mathrm{d}u\, \mathrm{e}^{-u} u^{n\alpha} = \frac{\Gamma(1 + n\alpha)}{z^{1+n\alpha}}$$

and so

$$p(\boldsymbol{q},t) = \frac{1}{f(\boldsymbol{q})}\sum_{n=0}^{\infty}(-1)^n\left[\frac{1-f(\boldsymbol{q})}{Af(\boldsymbol{q})}\right]^n\frac{t^{n\alpha}}{\Gamma(1+n\alpha)} = \frac{1}{f(\boldsymbol{q})}M_\alpha\left(\frac{1-f(\boldsymbol{q})}{Af(\boldsymbol{q})}t^\alpha\right)$$

where $M_\alpha(t)$ is the Mittag-Leffler function

$$M_\alpha(t) = \sum_{n=0}^{\infty}(-1)^n\frac{t^n}{\Gamma(1+n\alpha)}$$

Notice that

$$M_1(t) = \mathrm{e}^{-t}.$$

Compared with an exponential the Mittag-Leffler function decays slower due to the factor $\alpha < 1$ in the gamma-function $\Gamma(1+n\alpha)$. Another origin of slowing down is the fact that time enters via t^α.

In fig 3.1 we show the real and imaginary parts with $z = \mathrm{i}\omega$ of the Cole-Cole distribution. A characteristic property of the imaginary part $\chi''(\omega)$ is the symmetris shape when plotted versus $\log\omega$. This distribution is abundant in many complex stystems. An example is shown in figure 3.2 which show the imaginary part of the dielectric constant for polyester (polyethylene adipate). The various curves for different temperatures scale when plotted as function $\epsilon''(\omega)/\epsilon''_{\max}$. This master curve can be described by a Cole-Cole law.

The Cole-Cole law is used to interpret data in such diverse fields as physics, biology, geophysics etc.

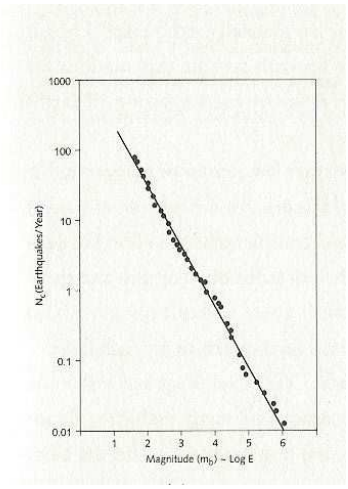

Figure 3.3: *Distribution of eartquake magnitudes in the New Madrid zone in the southeastern United States during the period 1974-1983. The points show the number of earthquakes with magnitude larger than a given magnitude m. The straight line indicates a power law distribution of earthquakes. From Johnston, A. C. and Nava, S. J. Geophys. Res.* **90** *6737 (1985).*

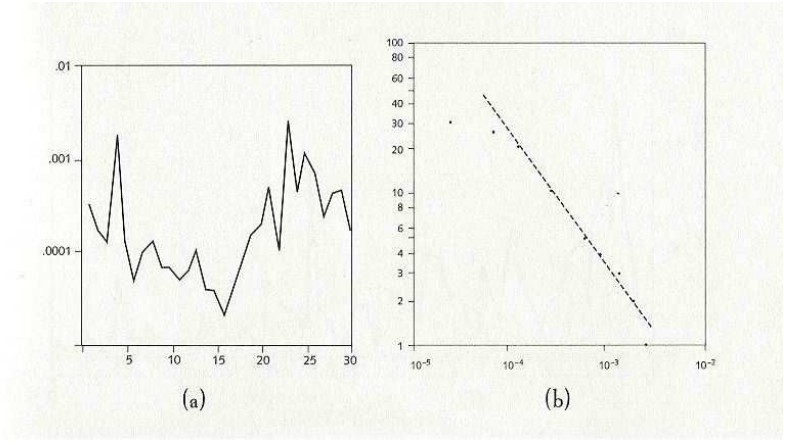

Figure 3.4: *a) Monthly variations of cotton prices during a period of 30 months. b) The curve shows the number of months where the relative variation exceeded a given fraction. Note the smooth transition from small variations to large variations. The straight line indicates a power low. Other commodities follow a similar pattern. From Mandelbrot, B. J. Business Univ. Chicago* **36** *307 (1963).*

Because of their composite nature, complex systems often exhibit distributions with long tails. There are many examples from variuos areas: physics, chemistry, biology, geoscience and social sciences. Some examples of distributions with long tails where the first or second moments may not exist are shown below.

The number of earthquakes of a given magniytude follows a simple distribution function known as the Gutenberg-Richter law. It turns out that every time there are about 1000 earthquakes of say magnitude 4 on the Richter scale, there are 100 earthquakes of magnitude 5, 10 of magnitude 6, and so on. This law is illustrated in fig. 3.3 which show how many earthquakes there were of each magnitude in a region of the southeastern United States known as the New Madrid earthquake zone durin ghte period 1974-1983. The Gutenberg-Richter law manifest itself as a straight line in this plot.

In economics, an empirical pattern similar to the Gutenberg-Richter law holds. Mandelbrot pointed out in 1966 that the probability of having small and large variations on prices of stocks, cotton and other commodities follows a very simple pattern, namely the Lévy distribution. Mandelbrot had collected data for the variation of cotton prices from month to month over several years. He then counted how often the monthly variationwas between 10 and 20 percent, how often the variation was between 5 and 10 per cent, and so on, and plotted the results on a logarithmic plot (fig. 3.4). The distribution of price changes follows approximately a straight line a power law. The price variations are "scale free" with no typical size of the variations, just as earthquakes do not have a typical characteristic size. Mandelbrot also studied several different commodities, and found that they all followed a similar pattern.

In his book *Human Behaviour and the Principle of Least Effort* Zipf have made a number of striking observations of some simple regularities in systems of human origin. Figure 3.5 shows how many cities in the world around 1920 had more than a given number

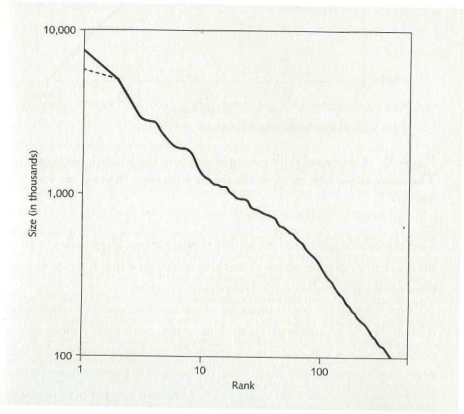

Figure 3.5: *Ranking of cities around the year 1920. The curves shows the number of cities in which the population exceeds a given size or, equivalently, the relative ranking of cities versus their population. From Zipf, G. K. Human Behaviour and the Principle of least Effort. Cambridge, Addison-Wesley, 1949.*

of inhabitants. There were a couple of cities larger than 8 million, ten larger than 1 million, and 100 larger than 200 000. The curve is roughly a straght line on a logarithmic plot. Zipf made similar plots for many geographical areas and found the same behaviour.

Zipf also counte dhow often a given word was used in a piece of literature, such as James Joye's *Ulysses* or a collection of American newspapers. The tnth most frequently used word (the word of rank 10) appeared 2653 times. The twnetieth most used word appeared 1311 times. The 20000th most frequent word was used only once. Figure 3.6 shows the frequency of words used in the English language versus their ranking. The word of rank 1, *the*, is used with a frequencey of 9 percent. The word of rank 10 ,*I*, has the frequency of 1 percent, the word of rank 100, *say*, is used with a frequency of 0.1 percent, and so on. Again a remarkable staright line emerges. It does not matter whether the data are taken from newspapers, the Bible or *Ulysses* - the curve is the same. The regularity expressed by the straight lines in the logarithmic plot of rank versus frequancy with slope near unity is referred to as Zopf's law.

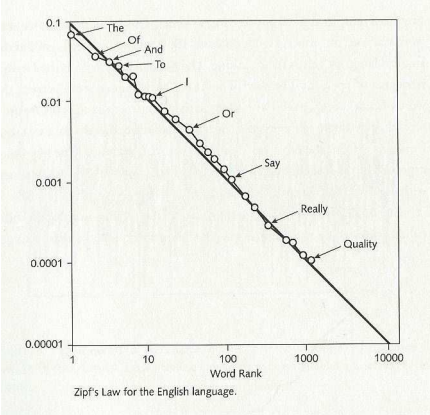

Figure 3.6: *Ranking of words in the English language. The curve shows how many words appear with more than a given frequency. From Zipf, G. K. Human Behaviour and the Principle of least Effort. Cambridge, Addison-Wesley, 1949.*