Probability

Riccardo Fantoni*
Università di Trieste, Dipartimento di Fisica, strada Costiera 11, 34151 Grignano (Trieste), Italy
(Dated: March 31, 2026)

Starting from the lecture notes for the graduate course PHYSICS 498C "Non-Equilibrium Statistical Mechanics" held by by Klaus Schulten on Spring 1998 at Loomis in Urbana/Champaign, IL, USA.

We review some basic probability concepts and use them in the theory of measure of a physical quantity either in the laboratory or on a computer (through the Monte Carlo method) and in the theory of stochastic (Markovian) processes used to treat stochastic differential equations and their deterministic counterparts like the Einstein or the more general Smoluchowski diffusion equation. We then put on the same footing the Brownian dynamics of a stochastic process with the thermal equilibrium of a many body statistical system and show how they can both be studied through the method of path integral Monte Carlo.

CONTENTS

* riccardo.fantoni@scuola.istruzione.it

I. INTRODUCTION

In the *theory of measure* (in physics), sampling is the selection of a subset of measurements from within a statistical population to estimate characteristics of the whole population. The subset, called a statistical *sample* (or sample, for short), is meant to reflect the whole population, and statisticians attempt to collect samples that are representative of the population. Sampling has lower costs and faster data collection compared to recording data from the entire population (in many cases, collecting the whole population is impossible, like getting sizes of all stars in the universe), and thus, it can provide insights in cases where it is unfeasible to measure an entire population.

One usually makes three hypotheses:

- The ensemble of the experimental measures (which is a small part of all the measurements that can be made) is a sample of all possible measures.
- Postulates the existence of a distribution of the results of all measures that can be done; this is called the *parent distribution*.
- From the sample it is necessary to approximate the *mean value* and the *variance* of the parent distribution.

The more fine are the experimental instruments the better one discovers that the physical quantities being measured fluctuates. When one arrives at the atomic structure the uncertainty principle states that the result of the measure is unavoidably governed by probabilistic laws. Then we can postulate that: "it does not exist the *true value* of a measurement but only the mean value of the parent distribution". This is then the value that is more reasonable to accept to represent the physical quantity. We will then call the *sampling* of a measurement the gathering of a sequence of samples statistically relevant and comparable so that one can increase the accuracy of the measure by a *statistical* or *probabilistic* analysis of the sequence.

One then want to define the *probability* of a certain *casual* or *random* or *aleatory* phenomenon; the term "aleatory" comes from the latin word "alea" which means dice. A *casual trial* is an experiment which does not have an a priori certain result. The calculus of probabilities has been developed in France, in the seventh century, to solve some dice problems posed by a player, the cavalier de Méré, to Blaise Pascal. We still have the letters exchanged on this argument in 1654 between Pascal and the mathematician Pierre de Fermat. The first systematic treatises on probability, that have been published in the two following centuries, are due to the mathematicians Jakob Bernoulli, Pierre-Simon Laplace and Carl Friedrich Gauss.

The *sampling space* or *space of the events* is the set Ω of all the possible outcomes of a casual trial. An *event* is a subset of the sampling space, $E \subseteq \Omega$. An *elementary event* is a subset of Ω which has only one element. We say that the event E is *realized* if the outcome of the experiment is an element of E otherwise we say that it did not realize.

There are two *objectivist* different definition for the probability of an event and one *subjectivist*. The objectivist definitions are:

- The *classical* or *theoretical* or *a priori* definition of probability is as follows. If Ω is a sampling space made of a finite number of elementary events, all with the same probability of realization, one defines the probability of event $E \subseteq \Omega$ the ratio between the the number k of elements of E and the number n of elements of Ω.

$$P\{E\} = \frac{k}{n}. \tag{1.1}$$

- The *frequentist* or *empiric* or *statistical* definition is as follows. If an event E is realized k times in n repeated trials under the same conditions and for n sufficiently large, the relative frequency $f = k/n$ can be considered an estimate of $p\{E\}$.

The subjectivist definition assign numbers between 0 and 1 per subjective probability, that is, as a "degree of belief" that an experimenter assigns to the realization of an event E constrained to the *coherence principle*, i.e. the experimenter has to be ready to pay an amount equal to $P\{E\}$ in case the event does not realize.

All three definitions presents some shortcomings. The classic definition requires that the space of events be *finite* and that all elementary events be *equally probable*. The frequentist definition requires the *repeatability* of the experiment under the *same conditions* many times. The subjectivist definition can lead different experimenters to assign different probabilities to the same event.

These considerations show how the problem of the formulation of a unifying *definition* for the probability of an event is not an easy one. The problem has been solved only recently in 1933 by the Russian mathematician Kolmogorov. He approached the problem in an *axiomatic* way. In his approach the concept of probability becomes a *primitive concept* (like the point, the line, the plane in Euclidean geometry) which is not *explicitly* defined but *implicitly* described by the axioms founding the theory of probability. The axioms are three. Given a sampling space Ω, the probability $P\{E\}$ of an event $E \subseteq \Omega$ is a real number such that

- $0 \leq P\{E\} \leq 1$;
- $P\{\Omega\} = 1$;
- if A and B are two *mutually exclusive* events, i.e. $P\{A \text{ and } B\} = 0$, then $P\{A \text{ or } B\} = P\{A\} + P\{B\}$.

In the axiomatic approach when one has to assign a probability in a concrete case he is free to follow the most proper approach (classical, frequentist, or subjectivist) with the only requirement not to violate the axioms. In the next section we will use this Kolmogorov approach. As usual we have the following equivalences between logical symbols and operations between sets: "and"$\equiv \cap$ the interception, "or"$\equiv \cup$ the union, "not E"$\equiv \overline{E}$ the complementary. So that for example $P\{\text{not } E\} = 1 - P\{E\}$.

In this set of notes we merge three containers, namely, the lecture notes for the physics graduate course of "Laboratory I" held by prof. Liana Martinelli at the Physics Department of the University of Pisa in 1990 [1], the book on "Monte Carlo Methods" by Melvin Kalos and Paula Whitlock [2], and the lecture notes for the graduate course of "Nonequilibrium Statistical Mechanics in Biophysics" held by prof. Klaus Schulten at the Physics Department of the University of Illinois at Urbana-Champaign in spring 1998 [3]. The material from these last lecture notes start from Section XIII. We try to give proofs to all the arguments and results described.

In Sections II-XII we give a bit of mathematical properties of probability describing the Monte Carlo integration method in Section X. In Section VII we present various methods to sample random variables describing the M(RT)^2 algorithm used to sample a generic one, in the last subsection. In Section XII we discuss some features of a Monte Carlo evaluation of a finite-dimensional integral; we compare Monte Carlo method with quadrature methods of integration and present the importance sampling method. In Section XIII we describe the theory of Markov stochastic processes and introduce three famous ones: the Wiener process, the Ornstein-Uhlenbeck process, and the White noise process. In Section XIV we introduce a first order stochastic differential equation as a prototype of the Langevin equation and in Section XV we introduce Ito and Stratonovich that are able to reduce a stochastic differential equation to a deterministic differential equation. We will then reduce the Langevin stochastic equation to the deterministic Fokker-Planck equation. In Section XVI we will see how the Fokker-Planck for the Langevin equation in absence of a force field and in the strong friction regime reduces to Einstein diffusion equation, i.e. free diffusion, and will solve it in various geometries. In Section XVII we will see how the Fokker-Planck for the Langevin equation in presence of a

force field and in the strong friction regime reduces to Smoluchowski diffusion equation, i.e. Brownian diffusion, and will solve it in a linear potential in various geometries and in a harmonic oscillator in one dimension. In Section XVIII we introduce the adjoint Smoluchowski equation useful to calculate correlation functions. In Section XIX we put on the same footing the Brownian dynamics of a stochastic process, i.e. the Fokker-Planck equation for the conditional probability distribution, with the thermal equilibrium of a many body statistical system, i.e. the Bloch equation for the thermal density matrix, and show how they can both be studied through the method of path integral Monte Carlo. In particular we will show how the Monte Carlo method can be used to justify the path integration of the Fokker-Planck equation through its underlying Markov stochastic process.

II. RANDOM EVENTS

Given an *elementary event* with a countable set of discrete random outcomes, $E_1, E_2, \ldots, E_n, \ldots$, there is associated with each possible outcome E_k a number called a *probability*, p_k, which can lie between 0 and 1, $0 \le p_k \le 1$.

If the kth outcome never occurs, $p_k = 0$; if it is sure to occur, $p_k = 1$. Conversely, if $p_k = 0$, we say that the event almost surely does not occur; and if $p_k = 1$, the event almost surely occurs. Another notation for the probability of event E_k is $P\{E_k\} = p_k$.

The following are some simple properties of the probability of events:

1. $P\{E_i \text{ and/or } E_j\} \le p_i + p_j$.
2. E_i and E_j are said to be *mutually exclusive* events if and only if the occurrence of E_i implies that E_j does not occur, $E_i \implies \bar{E}_j$ (not E_j), and vice versa. If E_i and E_j are mutually exclusive, $P\{E_i \text{ and } E_j\} = 0, P\{E_i \text{ or } E_j\} = p_i + p_j$.
3. A whole class of events can be mutually exclusive for all i and j. When the class is *exhaustive*, that is, all possible events have been enumerated, $P\{\text{some } E_i\} = \sum_i p_i = 1$.

   In the following, we consider a compound experiment consisting of just two elementary events. For clarity, we imagine the first to have outcomes $\{E_i\}$ with probability p_{1i} and the second to have outcomes $\{F_j\}$ and probabilities p_{2j}, respectively. Each of p_{1j} and p_{2j} obeys statements 1, 2, and 3 above. An outcome of such a composite event is a pair (E_i, F_j).
4. The probability of the specific outcome (E_i, F_j) is p_{ij}, called the *joint probability* for E_i and F_j.
5. $p_{ij} = p_{1i} \cdot p_{2j}$, if and only if events E_i and F_j are *independent*.
6. Suppose E_i and F_j are not independent; then the joint probability can be written as

$$p_{ij} = \left(\sum_k p_{ik}\right)\left(\frac{p_{ij}}{\sum_k p_{ik}}\right) = p(i)\left(\frac{p_{ij}}{\sum_k p_{ik}}\right), \tag{2.1}$$

   $p(i)$ defines a new number called the *marginal probability* for event E_i, that is, the probability that E_i does, in fact, occur, whatever the second event may be. Therefore,

$$\sum_i p(i) = \sum_i \sum_k p_{ik} = 1 \tag{2.2}$$

   The marginal distribution for the second event, F_j, can similarly be written as $p(j) = \sum_k p_{kj}$.
7. The second factor of Eq. (2.1) is the *conditional probability*

$$p(j|i) = \frac{p_{ij}}{\sum_k p_{ik}} \tag{2.3}$$

   and is the probability for event F_j occurring, given that event E_i has occurred. The probability for some F_j should be 1, and indeed

$$\sum_j p(j|i) = \sum_j \frac{p_{ij}}{\sum_k p_{ik}} = 1 \quad \text{for every } i. \tag{2.4}$$

All joint probabilities can be factored into a marginal distribution and a conditional probability. This scheme can be generalized to treat the joint occurrence of three or more elementary events.

III. RANDOM VARIABLES

In the following discussion we shall assume that for every elementary outcome E_i, there is an associated real number x_i. A random selection, X, of one of the possible values $x_1, x_2, \ldots$ is called a random variable. The probability that the value x_i is chosen is given by $p_i = P\{X = x_i\}$. The *expectation* of this random variable X, that is, the stochastic mean value, is defined as

$$E(X) \equiv \sum_i P\{X = x_i\} x_i = \sum_i p_i x_i = \mu. \tag{3.1}$$

μ is called the *expected* or *mean value*. It is common in physics to write this as $\langle X \rangle$, and we shall often use that notation.

Consider some real-valued function $g(x_i) = g_i$, where the x_i correspond to a countable set of elementary events with probabilities p_i. If X is a random variable, then $g(X)$ is also a random variable. The expectation of $g(X)$ is defined as

$$E(g(X)) = \langle g(X) \rangle = \sum_i p_i g(x_i). \tag{3.2}$$

From the definition of the expected value of a function, we have the property that

$$\langle \text{constant} \rangle = \text{constant}, \tag{3.3}$$

and that for any constants λ_1, λ_2 and two functions g_1, g_2,

$$\langle \lambda_1 g_1(X) + \lambda_2 g_2(X) \rangle = \lambda_1 \langle g_1(X) \rangle + \lambda_2 \langle g_2(X) \rangle. \tag{3.4}$$

An important application of expected values is to the powers of X. The nth moment of X is defined as the expectation of the nth power of X,

$$\langle X^n \rangle \equiv \sum_i p_i x_i^n, \tag{3.5}$$

The *central moments* of X are given by

$$\langle (X - \mu)^n \rangle \equiv \sum_i p_i (x_i - \langle X \rangle)^n. \tag{3.6}$$

The second central moment has particular significance,

$$\langle (X - \mu)^2 \rangle = \langle (X - \langle X \rangle)^2 \rangle = \sum_i p_i (x_i - \langle X \rangle)^2 = \sum_i p_1 x_i^2 - \langle X \rangle^2 = \langle X^2 \rangle - \langle X \rangle^2, \tag{3.7}$$

and is called the *variance* of X or $\text{var}\{X\}$. The square root of the variance is a measure of the dispersion of the random variable. It is referred to as the *standard deviation* and sometimes the *standard error*. The variance of a function of the random variable, $g(X)$, can be determined as

$$\text{var}\{g(X)\} = \langle (g(X) - \langle g(X) \rangle)^2 \rangle = \langle g^2(X) \rangle - \langle g(X) \rangle^2. \tag{3.8}$$

Consider two random variables, X and Y. They are are not in general independent. Two random variables are said to be independent if they derive from independent events. As we have seen in Eq. (3.4), the expectation of a linear combination is the linear combination of the expectations. This result does not require that X and Y be independent. The effect of statistical dependence will be seen in the variance of a linear combination of the two variables,

$$\text{var}\{\lambda_1 X + \lambda_2 Y\} = \lambda_1^2 \text{var}\{X\} + \lambda_2^2 \text{var}\{Y\} + 2[\lambda_1 \lambda_2 \langle XY \rangle - \lambda_1 \lambda_2 \langle X \rangle \langle Y \rangle]. \tag{3.9}$$

The expectation of the product is

$$\langle XY \rangle = \sum_{i,j} p_{ij} x_i y_j. \tag{3.10}$$

If X and Y are independent, $p_{ij} = p_{1i} \cdot p_{2j}$ and

$$\langle XY \rangle = \sum_i p_{1i} x_i \sum_j p_{2j} y_j = \langle X \rangle \langle Y \rangle. \tag{3.11}$$

Then Eq. (3.9) becomes simply

$$\text{var}\{\lambda_1 X + \lambda_2 Y\} = \lambda_1^2 \text{var}\{X\} + \lambda_2^2 \text{var}\{Y\}. \tag{3.12}$$

When X and Y are not necessarily independent, we introduce a new quantity, the *covariance*, which is a measure of the degree of independence of the two random variables X and Y:

$$\text{cov}\{X, Y\} = \langle XY \rangle - \langle X \rangle \langle Y \rangle. \tag{3.13}$$

The covariance equals 0 when X and Y are independent and $\text{cov}\{X, X\} = \text{var}\{X\}$.

Note that zero covariance does not by itself imply independence of the random variables. The following simple example illustrates that even functional dependence can still yield a zero covariance. Let X be a random variable that may be -1, 0, or 1 with equal probabilities, and define $Y = X^2$. Obviously, $\langle X \rangle = 0, \langle XY \rangle = \langle X^3 \rangle = 0$ so $\text{cov}\{XY\} = 0$.

The covariance can have either a positive or negative value. Another quantity derived from the covariance is the correlation coefficient,

$$\rho(X, Y) = \frac{\text{cov}\{X, Y\}}{\sqrt{\text{var}\{X\}\text{var}\{Y\}}}, \tag{3.14}$$

so that $-1 \leq \rho(X, Y) \leq 1$. Since the covariance can be positive or negative, the variance of a linear combination of two dependent random variables can be greater or less than the variance if the variables were independent.

A. The Binomial Distribution

Consider two events E_0 and E_1 that are mutually exclusive and exhaustive:

$$P\{E_1\} = p, \quad x_1 = 1, \tag{3.15}$$

$$P\{E_0\} = 1 - p, \quad x_0 = 0. \tag{3.16}$$

Let X_i be the random variable that is the ith outcome of a series of N such events. The expectations of X_i and its square become

$$E(X_i) = p \cdot 1 + (1 - p) \cdot 0 = p, \tag{3.17}$$

$$E(X_i^2) = p, \tag{3.18}$$

and the variance is then

$$\text{var}\{X_i\} = p - p^2 = p(1 - p). \tag{3.19}$$

Each outcome is either 0 or 1, and we set X to be the sum of the N outcomes $\{X_i\}$,

$$X = \sum_{i=1}^{N} X_i. \tag{3.20}$$

The probability that $X = n$ is the probability that n of the X_i were 1 and $N - n$ were 0. That is,

$$P\{X = n\} = \binom{N}{n} p^n (1 - p)^{N-n} \quad n = 0, 1, \ldots, N. \tag{3.21}$$

This is the *binomial distribution*. $\binom{N}{n} = N!/n!(N - n)!$ is the binomial coefficient, which counts the number of different ways in which the n E_1 and the $(N - n)$ E_0 may occur. The expected value of X is

$$\langle X \rangle = \sum_{n=0}^{N} n \binom{N}{n} p^n (1 - p)^{N-n} = Np. \tag{3.22}$$

This may be verified by direct computation of the sum or by noting that the expected value of X is the sum of the expected value of all the X_i. The variance of X is easily determined; since the X_i are independent, the result in Eq. (3.12) may be employed and

$$\langle (X - Np)^2 \rangle = \sum_{i=1}^{N} \text{var}\{X_i\} = \sum_{i=1}^{N} p(1 - p) = Np(1 - p). \tag{3.23}$$

B. The Geometric Distribution

Suppose we carry out a certain experiment repeatedly and independently, where there are only two outcomes: failure or success. If the outcome is a failure, the experiment will be repeated; otherwise, we stop the procedure. Let the random variable X of interest be the number of times we repeat the experiment until the first success. Let q be the probability of failure in one experiment and $p = 1 - q$ be the probability of success. Then

$$P\{X = n\} = q^{n-1}p, \quad n = 1, 2, \ldots. \tag{3.24}$$

The average number of experiments that will be carried out is

$$\langle X \rangle = \sum_{n=1}^{\infty} nq^{n-1}p = \frac{p}{(1-q)^2} = \frac{1}{p}. \tag{3.25}$$

The variance of X can be calculated as

$$\mathrm{var}\{X\} = \langle X^2 \rangle - \langle X \rangle^2 = \left(\frac{2}{p^2} - \frac{1}{p}\right) - \frac{1}{p^2}. \tag{3.26}$$

As an example, in particle transport problems, the number of collisions one particle makes follows this distribution if the medium is infinite, homogeneous, and if the relative probability of different outcomes is constant.

C. The Poisson Distribution

A random variable X is said to follow a *Poisson distribution* when

$$P\{X = n\} = \frac{\lambda^n}{n!}e^{-\lambda}, \quad n = 0, 1, \ldots, \tag{3.27}$$

where λ is a parameter of the distribution. It is easy to show that

$$\langle X \rangle = \lambda, \tag{3.28}$$

$$\mathrm{var}\{X\} = \lambda. \tag{3.29}$$

This distribution is fundamental in the theory of probability and stochastic processes. It is of great use in applications such as research into queuing service systems and similar discrete event problems.

IV. CONTINUOUS RANDOM VARIABLES

In the previous sections, we have assumed that the random events belonged to a discrete, countable set. Probabilities can be associated with continuous variables as well, giving rise to distribution functions. Such distributions are present both in nature and in artificial stochastic processes. As an example, consider the scattering of a photon by an atom. The angle at which the photon is scattered has values that are continuous between 0° and 180° with some angular intervals occurring more often than others.

Given that x is a real number, $-\infty < x < \infty$, a distribution function (or *cumulative* distribution function) may be defined as

$$F(x) \equiv P\{\text{a random selection of } X \text{ gives a value less than } x\} = P\{X \leq x\}. \tag{4.1}$$

Suppose $x_1 < x_2$. Then $x_1 < X \leq x_2$, and $X \leq x_1$ are mutually exclusive and exhaustive for events in which $X \leq x_2$. Thus

$$P\{x_1 < X \leq x_2\} + P\{X \leq x_1\} = P\{X \leq x_2\} \tag{4.2}$$

and

$$1 \geq P\{x_1 < X \leq x_2\} = P\{X \leq x_2\} - P\{X \leq x_1\} \geq 0. \tag{4.3}$$

Therefore, $F(x)$ is a nondecreasing function of its argument. We conclude that $F(-\infty) = 0$ and $F(\infty) = 1$. Furthermore, the difference $P\{X \leq x_2\} - P\{X \leq x_1\}$ may be written as $F(x_2) - F(x_1)$. The distribution function may have intervals on which it is differentiable; in these intervals, the *probability distribution function* (pdf) may be defined as

$$f(x) \equiv \frac{dF(x)}{dx} \geq 0. \tag{4.4}$$

If $F(x)$ is not continuous, discrete values of the distribution function are singled out at the discontinuities. For example, imagine that $F(x)$ is piecewise constant everywhere except at a countable number of places; the distribution now describes a discrete set of random variables. Formally, we may use the Dirac delta function to write

$$f(x) = \sum_i \delta(x - x_i) \cdot p_i, \tag{4.5}$$

where p_i is the jump of the distribution function at x_i. This emphasizes the fact that $f(x)$ need not be bounded.

V. EXPECTATIONS OF CONTINUOUS RANDOM VARIABLES

Let $f(x)$ be the pdf of a continuous random variable x. It has the normalization property

$$\int_{-\infty}^{\infty} f(x)\, dx = F(\infty) = 1. \tag{5.1}$$

The mean value of x is defined as

$$E(X) \equiv \int_{-\infty}^{\infty} x\, dF(x) = \int_{-\infty}^{\infty} x f(x)\, dx. \tag{5.2}$$

The expected value of any function of the random variable is defined as

$$E(g(X)) \equiv \int_{-\infty}^{\infty} g(x) f(x)\, dx \tag{5.3}$$

Let us consider a few representative and interesting distributions. For each case we will give $F(x)$, $f(x) = F'(x)$, the mean and variance:

- Uniform

$$F(x) = \begin{cases} 0 & x < 0, \\ x & 0 \le x \le a, \\ 1 & x > a. \end{cases} \qquad f(x) = \begin{cases} 0 & x < 0, x > a, \\ 1/a & 0 < x < a. \end{cases} \qquad \langle x \rangle = a/2 \quad \mathrm{var}(x) = a/12 \tag{5.4}$$

- Exponential

$$F(x) = \begin{cases} 0 & x < 0, \\ 1 - e^{-\lambda x} & x \ge 0. \end{cases} \qquad f(x) = \begin{cases} 0 & x < 0, \\ \lambda e^{-\lambda x} & x \ge 0. \end{cases} \qquad \langle x \rangle = 1/\lambda \quad \mathrm{var}(x) = 1/\lambda^2 \tag{5.5}$$

- Normal $\mathcal{N}(\mu, \sigma^2)$

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\, dt \quad f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad \langle x \rangle = \mu \quad \mathrm{var}(x) = \sigma^2 \tag{5.6}$$

- $\chi^2(n)$

$$F(x) = P(k/2, x/2) \quad f(x) = \begin{cases} 0 & x < 0, \\ \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2} & x \ge 0. \end{cases} \qquad \langle x \rangle = n \quad \mathrm{var}(x) = 2n \tag{5.7}$$

  where $\Gamma(s)$ is the gamma function and $P(s,t)$ is the regularized gamma function.

- Cauchy (Lorentz)

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{x}{a}\right) \quad f(x) = \frac{1}{\pi} \frac{a}{a^2 + x^2} \qquad \langle x \rangle = ? \quad \mathrm{var}(x) = \infty \tag{5.8}$$

Note that for the uniform distribution we have $x = F(x) = P\{X \le x\}$ so that if ξ is a uniformly distributed random number, then $a \ge \xi$ occurs with probability a. This property will be used ubiquitously in Section XI where random numbers uniformly distributed in the interval $(0, 1)$ are used to sample other random variables.

The importance of the normal distribution will become clear in Section IX. Random variables distributed according to $\mathscr{N}(1, 0)$ are said *standard normal* or *unit normal*.

The importance of the χ^2 distribution will become clear in Section IX A. If x follows a $\chi^2(n)$ distribution then for big values of n ($n \gtrsim 30$) the distribution for the variable $\sqrt{2x} - \sqrt{2n-1}$ tends to a normal $\mathscr{N}(0, 1)$ distribution.

A less well-behaved example is provided by the Cauchy or Lorentz function. The mean value of a random variable sampled from a Cauchy distribution is

$$\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{xa}{a^2 + x^2}\, dx. \tag{5.9}$$

To evaluate this improper integral using elementary calculus, the infinite endpoints of integration are replaced by finite quantities b and b', and the behavior of the integrand as b and b' approach infinity is considered. The integrand clearly diverges unless $b = b'$, and then the mean value is 0. This suggests that the mean value of a series of random variables may be undefined unless the variables are chosen in some special way. The variance of a random variable sampled from a Cauchy distribution is infinity since the integral

$$\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{x^2 a}{a^2 + x^2}\, dx. \tag{5.10}$$

diverges no matter how it is evaluated. In spite of the infinite variance, the Cauchy distribution can be sampled and used as needed.

Note that for each distribution there is a length scale, called variously $a, 1/\lambda$, or σ. As the scale becomes small, the normalization of the pdf grows large, inversely as the length scale, so as to ensure $\int f(x)\, dx = 1$. For those distributions with a standard deviation σ, the width is proportional to σ. In the Cauchy distribution, a is a measure of the width.

VI. BIVARIATE CONTINUOUS RANDOM DISTRIBUTIONS

A joint probability may be defined for continuous distributions,

$$F(x, y) \equiv P\{X \le x, Y \le y\}. \tag{6.1}$$

$F(x, y)$ is termed a *bivariate* distribution function. The associated bivariate probability distribution function is

$$f(x, y) \equiv \frac{\partial^2 F(x, y)}{\partial x \partial y} \tag{6.2}$$

and the expected value of any function of random variables X, Y is

$$E(g(X, Y)) = \langle g(X, Y)\rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y) f(x, y)\, dxdy. \tag{6.3}$$

The covariance and correlation coefficient, $\mathrm{cov}\{X, Y\}$ and $\rho(X, Y)$ for continuous random variables are defined as in the discrete case, replacing sums by integrals.

If X and Y are correlated, it is useful to write the joint pdf as

$$f(x, y) = \frac{f(x, y)}{\int_{-\infty}^{\infty} f(x, y)\, dy}\int_{-\infty}^{\infty} f(x, y)\, dy. \tag{6.4}$$

We have

$$P\{X \le x\} = \int_{-\infty}^{x}\int_{-\infty}^{\infty} f(t, y)\, dydt = \int_{-\infty}^{x} m(t)\, dt, \tag{6.5}$$

where

$$m(x) \equiv \int_{-\infty}^{\infty} f(x, y)\, dy, \tag{6.6}$$

is called the *marginal* probability distribution function for x. The first factor in Eq. (6.4) is the conditional probability (2.3); that is, given an X, a Y may be chosen from

$$f(y|x) = \frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,y)\,dy} = \frac{f(x,y)}{m(x)}. \tag{6.7}$$

That is, when the marginal and conditional functions can be determined, sampling a bivariate distribution requires simply sampling two univariate distributions.

The relationship in Eq. (6.4) is easily generalized to handle more than two correlated random variables, and sampling the multivariate distribution will then involve sampling the sequence of univariate distributions so defined. What happens to the marginal pdf and the conditional probability when X and Y are independent (i.e. $f(x,y) = f_1(x)f_2(y)$) is left for the reader.

The expectation of the conditional probability, called the conditional expectation of Y, for fixed X is

$$E(Y|X) \equiv \int_{-\infty}^{\infty} yf(y|x)\,dy = \frac{\int_{-\infty}^{\infty} yf(x,y)\,dy}{\int_{-\infty}^{\infty} f(x,y)\,dy} = \frac{\int_{-\infty}^{\infty} yf(x,y)\,dy}{m(x)} \tag{6.8}$$

The conditional expectation $E(Y|X)$ is a function of the random variable X and is itself a random variable. The expectation of $E(Y|X)$ is

$$E(E(Y|X)) = \int_{-\infty}^{\infty} E(Y|X)m(x)\,dx. \tag{6.9}$$

Upon substituting in the definition for $E(Y|X)$

$$E(E(Y|X)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)\,dydx = E(Y). \tag{6.10}$$

A more general result for a function $g(X,Y)$ is

$$E(E(g(X,Y)|X)) = E(g(X,Y)). \tag{6.11}$$

VII. SUM OF RANDOM VARIABLES

Suppose that the random variables $X_1, X_2, \ldots, X_j, \ldots$ are all drawn at random, but not necessarily independently, from the probability distribution function $f(x)$. Let g_i be possibly different functions of X_i and λ_i real numbers. Define the function G by

$$G \equiv \sum_{i=1}^{N} \lambda_i g_i(X_i). \tag{7.1}$$

The expected value of G is

$$E(G) = \langle G\rangle = E\left(\sum_{i=1}^{N} \lambda_i g_i(X_i)\right) = \sum_{i=1}^{N} \lambda_i \left\langle g_i(X_i)\right\rangle, \tag{7.2}$$

since the expectation is a linear operation. If all the X_j are independent, then the variance of G becomes

$$\mathrm{var}\{G\} = \sum_{i=1}^{N} \lambda_i^2 \mathrm{var}\{g_i(X_i)\}. \tag{7.3}$$

Let $\lambda_i = 1/N$ and all the $g_i(x)$ be identical and equal to $g(x)$; then the expected value of G becomes

$$E(G) = E\left(\frac{1}{N}\sum_{i=1}^{N} g(X_i)\right) = \frac{1}{N}\sum_{i=1}^{N} \langle g(X_i)\rangle = \langle g(X)\rangle. \tag{7.4}$$

The function G, which is the arithmetic average of the $g(x)$, has the same mean as $g(x)$. G is said to be an *estimator* of $\langle g(X)\rangle$.

The variance of G in Eq. (7.3) becomes

$$\mathrm{var}\{G\} = \frac{1}{N^2}\sum_{i=1}^{N}\mathrm{var}\{g_i(X_i)\} = \frac{1}{N}\mathrm{var}\{g(X)\}. \tag{7.5}$$

That is, as N, the number of samples of X, increases, the variance of the mean value of G decreases as $1/N$. This result leads to the central idea of *Monte Carlo* evaluation of integrals; that is, an integral may be estimated by a sum

$$E\left(\frac{1}{N}\sum_{i=1}^{N} g(X_i)\right) = \langle g(X)\rangle = \sum_{i=1}^{N} g(X_i)P\{x < X_i \le x + dx\} = \int_{-\infty}^{\infty} g(x)f(x)\,dx, \tag{7.6}$$

where in the last equality we used the definition (4.4). To use relation (7.6), select a series of random variables, X_i, from $f(x)$; evaluate $g(x)$ for each X_i. The arithmetic mean of all the values of $g(X_i)$ is an estimate of the integral, and the variance of this estimate decreases as the number of terms increases.

VIII. DISTRIBUTION OF THE MEAN OF A RANDOM VARIABLE

In the discussion that follows on estimating integrals, it is assumed that the variance of the random variable always exists.

The most general result of the kind we need is the "*law of large numbers*" of probability theory. Suppose the random variables $X_1, X_2, \ldots, X_N$ are independent and all drawn from the same distribution. These are called *independent, identically distributed* (or i.i.d.) random variables. Then the expectation of each X is μ. As $N \to \infty$, the average value of the $\{X_i\}$, $\bar{X}_N = \sum_{i=1}^{N} X_i/N$, converges to μ almost surely [4]

$$P\left\{\lim_{N\to\infty} \bar{X}_N = \mu\right\} = 1. \tag{8.1}$$

The implication of the law of large numbers is that the mean of N sampled random variables converges (in probability) to its expected value. To estimate the speed of convergence, we need stronger assumptions. The most important way of strengthening the hypothesis is to assume that the variance exists, which we do in the following.

A. Chebyshev inequality

Markov inequality states that for any non-negative real-valued random variable Y and any positive number a, we have $P\{|Y| \ge a\} \le E(|Y|)/a$. This can for example be proven observing that

$$E(Y) = \int_{-\infty}^{\infty} yf(y)\,dy = \int_{0}^{\infty} yf(y)\,dy \ge \int_{a}^{\infty} yf(y)\,dy \ge a\int_{a}^{\infty} f(y)\,dy = aP\{Y \ge a\}. \tag{8.2}$$

Chebyshev inequality states that for any random variable X with mean value μ and finite non-zero variance σ^2 then for any real number $k > 0$

$$P\{|X - \mu| \ge k\sigma\} \le \frac{1}{k^2}. \tag{8.3}$$

One way to prove Chebyshev inequality is to apply Markov inequality to the random variable $Y = (X - \mu)^2$ with $a = (k\sigma)^2$ as follows

$$P\{|X - \mu| \ge k\sigma\} = P\{(X - \mu)^2 \ge k^2\sigma^2\} \le \frac{E((X - \mu)^2)}{k^2\sigma^2} = \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}. \tag{8.4}$$

This inequality could be called the *first fundamental theorem of Monte Carlo* for it gives an estimation of the chances of generating a large deviation in a Monte Carlo calculation. For definiteness, let $k = 10$. Then, using Eq. (7.5), the inequality becomes

$$P\left\{(G - \langle G\rangle)^2 \ge \frac{100}{N}\mathrm{var}\{g\}\right\} \le \frac{1}{100}. \tag{8.5}$$

Since by making N big enough, the variance of G becomes as small as one likes, the probability of getting a large deviation relative to k between the estimate of the integral and the actual value becomes very small. For large sample

size (large N), the range of values of G that will be observed with some fixed probability will be contained in a region of decreasing size near $\langle g\rangle$. This is the heart of the Monte Carlo method for evaluating integrals.

A much stronger statement than the Chebychev inequality about the range of values of G that can be observed is given by the *central limit theorem* of probability. For any fixed value of N, there is a pdf that describes the values of G that occur in the course of a Monte Carlo calculation [1]. As $N \to \infty$, however, the central limit theorem shows that there is a specific limit distribution for the observed values of G, namely, the normal distribution (see Section V). Recalling the previous section VII where $G_N = \sum_{i=1}^{N} g(X_i)/N$ set

$$t_N = \frac{G_N - \langle G_N\rangle}{\sqrt{\mathrm{var}\{G_N\}}} = \frac{\sqrt{N}(G_N - \langle g\rangle)}{\sqrt{\mathrm{var}\{g\}}}, \tag{8.6}$$

then for $t = \lim_{N\to\infty} t_N$

$$f(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}. \tag{8.7}$$

That is, the observed G_N is within one *standard error* (i.e. $\sigma/\sqrt{N}$) of $\langle g\rangle$ 68.3% of the time, within two standard errors 95.4% of the time, and within three standard errors 99.7% of the time. The result (8.7) will be proven in Section VIII C.

The central limit theorem is very powerful in that it gives a specific distribution for the values of G_N, but it applies only asymptotically. How large N must be before the central limit theorem applies depends on the problem [2].

Without the central limit theorem, there is in general only the much weaker upper bound of the Chebychev inequality to suggest how much the observed G_N deviates from the actual mean. Of course, in specific cases, studies can be made of the distribution of the estimator. Much Monte Carlo is done assuming that the theorem has been satisfied no matter what the sample size; reported errors must be considered optimistic in such cases.

The variance used in the discussion given above may itself be estimated using independent values of $g(X_i)$ in the following way

$$\left\langle \frac{1}{N}\sum_{i=1}^{N} g^2(X_i) - \left[\frac{1}{N}\sum_{i=1}^{N} g(X_i)\right]^2 \right\rangle = \langle g^2\rangle - \frac{1}{N^2}\left\langle \frac{1}{N}\sum_{i=1}^{N} g^2(X_i) + \frac{1}{N}\sum_{i\neq j=1}^{N} g(X_i)g(X_j)\right\rangle. \tag{8.8}$$

Using the independence of $g(X_i)$ and $g(X_j)$ in evaluating $\langle g(X_i)g(X_j)\rangle$, we find the right-hand side equal to

$$\left(1 - \frac{1}{N}\right)\langle g^2\rangle - \frac{N(N-1)}{N^2}\langle g\rangle^2 = \frac{N-1}{N}\mathrm{var}\{g\}. \tag{8.9}$$

Thus an estimator for σ^2 is

$$\sigma^2 \approx \frac{N}{N-1}\left\{\frac{1}{N}\sum_{i=1}^{N} g^2(X_i) - \left[\frac{1}{N}\sum_{i=1}^{N} g(X_i)\right]^2\right\}. \tag{8.10}$$

An estimator of the variance of the estimated mean is given by

$$\mathrm{var}\{G_N\} \approx \frac{1}{N-1}\left\{\frac{1}{N}\sum_{i=1}^{N} g^2(X_i) - \left[\frac{1}{N}\sum_{i=1}^{N} g(X_i)\right]^2\right\}. \tag{8.11}$$

We will now open an interlude on the properties of the characteristic function that will be used in Section VIII C to prove the central limit theorem.

---

[1] For fixed N, if G was calculated M times, each time with a different sequence of i.i.d. random variables, the set $\{G_j\}, j = 1, \ldots, M$ has a specific distribution function.

[2] If for a particular problem the third central moment μ_3 (see Eq. (8.15)) of g exists, then the central limit theorem will be substantially satisfied when $|\mu_3| \ll \sigma^3\sqrt{N}$.

B. Characteristic Functions

The *characteristic function* is a way to describe a random variable X drawn from a pdf $f(x)$. The characteristic function is defined as

$$\varphi_X(t) \equiv E\left(e^{itX}\right) = \int_{-\infty}^{\infty} e^{itx} f(x)\, dx, \tag{8.12}$$

as the Fourier transform of the pdf. It is a function of $t \in \mathbb{R}$ and determines the behavior and properties of the probability distribution of X.

Characteristic functions can also be used to find *moments* of a random variable. Provided that the nth moment exists, the characteristic function can be differentiated n times

$$E\left(X^n\right) = i^{-n}\left[\frac{d^n}{dt^n}\varphi_X(t)\right]_{t=0} = i^{-n}\varphi_X^{(n)}(0), \tag{8.13}$$

A Taylor expansion of the characteristic function (8.12) around $(it) = 0$ gives

$$\varphi_X(t) = \sum_{n=0}^{\infty} E\left(X^n\right)\frac{(it)^n}{n!}. \tag{8.14}$$

Sometimes central moments (3.6)

$$\mu_n = E\left((X - E(X))^n\right), \tag{8.15}$$

are used in preference to ordinary moments, computed in terms of deviations from the mean instead of from zero, because the higher-order central moments relate only to the spread and shape of the distribution, rather than also to its location.

For example, suppose X follows a Gaussian distribution i.e. $X \sim \mathcal{N}(\mu, \sigma^2)$. Then

$$\varphi_X(t) = e^{\mu it - \frac{1}{2}\sigma^2 t^2} \tag{8.16}$$

and

$$E\left(X\right) = i^{-1}\left[\frac{d}{dt}\varphi_X(t)\right]_{t=0} = i^{-1}\left[(i\mu - \sigma^2 t)\varphi_X(t)\right]_{t=0} = \mu, \tag{8.17}$$

$$E\left(X^2\right) = \mu^2 + \sigma^2. \tag{8.18}$$

The logarithm of a characteristic function is a *cumulant generating function*, which is useful for finding *cumulants* κ_n

$$\ln[\varphi_X(t)] = \sum_{n=1}^{\infty} \kappa_n \frac{(it)^n}{n!}. \tag{8.19}$$

The first cumulant is the mean, the second cumulant is the variance, and the third cumulant is the same as the third central moment. But fourth and higher-order cumulants are not equal to central moments. In some cases theoretical treatments of problems in terms of cumulants are simpler than those using moments. In particular, when two or more random variables are statistically independent, the nth-order cumulant of their sum is equal to the sum of their nth-order cumulants. As well, the third and higher-order cumulants of a normal distribution are zero, and it is the only distribution with this property [5]. See also Section XIII C.

C. The Central Limit Theorem

Theorem VIII.1 (Central Limit Theorem)**.** *Suppose $X_1, X_2, X_3 \ldots$ is a sequence of i.i.d. random variables with $E(X_i) = \mu$ and $var\{X_i\} = \sigma^2 < \infty$. Then, as n approaches infinity, the random variables $\sqrt{n}(\bar{X}_n - \mu)$, with $\bar{X}_n = \sum_{i=1}^{n} X_i/n$ converge in distribution to a normal $\mathcal{N}(0, \sigma^2)$:*

$$\sqrt{n}\left(\bar{X}_n - \mu\right) \longrightarrow \mathcal{N}\left(0, \sigma^2\right). \tag{8.20}$$

The *central limit theorem* has a proof which uses characteristic functions.

Proof. Assume $X_1, X_2, X_3 \ldots$ are i.i.d. random variables, each with mean μ and finite variance σ^2. We want to prove that the sum $X_1 + \cdots + X_n$ has mean $n\mu$ and variance $n\sigma^2$. Consider the random variable

$$Z_n = \frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n\sigma^2}} = \sum_{i=1}^{n} \frac{X_i - \mu}{\sqrt{n\sigma^2}} = \sum_{i=1}^{n} \frac{1}{\sqrt{n}} Y_i, \tag{8.21}$$

where in the last step we defined the new random variables $Y_i = \frac{X_i - \mu}{\sigma}$, each with zero mean and unit variance $\mathrm{var}\{Y\} = 1$. The characteristic function of Z_n is given by

$$\begin{aligned} \varphi_{Z_n}(t) = \varphi_{\sum_{i=1}^n \frac{1}{\sqrt{n}} Y_i}(t) &= \varphi_{Y_1}\left(\frac{t}{\sqrt{n}}\right) \varphi_{Y_2}\left(\frac{t}{\sqrt{n}}\right) \cdots \varphi_{Y_n}\left(\frac{t}{\sqrt{n}}\right) \\ &= \left[\varphi_{Y_1}\left(\frac{t}{\sqrt{n}}\right)\right]^n, \end{aligned} \tag{8.22}$$

where in the last step we used the fact that all of the Y_i are identically distributed. The characteristic function of Y_1 is, by Taylor theorem,

$$\varphi_{Y_1}\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right), \quad \left(\frac{t}{\sqrt{n}}\right) \to 0 \tag{8.23}$$

where $o(t^2/n)$ is "little o notation" for some function of t that goes to zero more rapidly than t^2/n. By the limit of the exponential function $e^x = \lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n$, the characteristic function of Z_n equals

$$\varphi_{Z_n}(t) = \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n \to e^{-\frac{1}{2}t^2}, \quad n \to \infty. \tag{8.24}$$

All of the higher order terms vanish in the limit $n \to \infty$. The right hand side equals the characteristic function of a standard normal distribution $\mathcal{N}(0,1)$, which implies through *Lévy continuity theorem* [6] that the distribution of Z_n will approach $\mathcal{N}(0,1)$ as $n \to \infty$. Therefore, the sample average

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n} \tag{8.25}$$

is such that

$$\frac{\sqrt{n}}{\sigma}\left(\bar{X}_n - \mu\right) = Z_n \tag{8.26}$$

converges to the normal distribution $\mathcal{N}(0,1)$, from which the central limit theorem follows. □

IX. DISTRIBUTION OF THE SUM OF INDEPENDENT RANDOM VARIABLES

Let X be chosen from $f_1(x)$ and Y independently chosen from $f_2(y)$. If the sum $Z = X + Y$ is formed, what is the probability distribution function for Z? The distribution function is defined as

$$F_3(z) = P\{Z \le z\} = P\{X + Y \le z\}. \tag{9.1}$$

Since X and Y are independent, their joint probability distribution function is

$$f(x,y) = f_1(x) f_2(y). \tag{9.2}$$

The variables x and y can be considered to form a point in the xy plane. What fraction of the time does the point (x,y) lie below the line $X + Y = Z$? The (cumulative) distribution function is

$$F_3(z) = \int_{x+y\le z} \int f_1(x) f_2(y)\, dxdy = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_2(y) f_1(x)\, dydx = \int_{-\infty}^{\infty} F_2(z-x) f_1(x)\, dx. \tag{9.3}$$

Differentiating with respect to z, one obtains for the pdf of z

$$f_3(z) = \int_{-\infty}^{\infty} f_2(z-x) f_1(x)\, dx. \tag{9.4}$$

This is a convolution and Fourier transforms can be used to evaluate it. Recalling the definition of the characteristic function (8.12) we then find

$$\varphi_3(t) = \varphi_1(t)\varphi_2(t), \tag{9.5}$$

or the characteristic function of a sum is the product of the characteristic functions of the terms of the sum [3]. Clearly, induction gives the same result for a sum of n variables. The characteristic function may be inverted (by a Fourier transform) to give the pdf to which it corresponds. If n identical functions constitute the sum, this result can be used to prove the central limit theorem as shown in Section VIII C.

When $F(x)$ is the normal distribution $\mathscr{N}(0,1)$ of Section V, the characteristic function is $\exp[-t^2/2]$. If a normal distribution is sampled n times, the sum is also distributed normally. The characteristic function is then $\exp[-nt^2/2]$, which when inverted gives the normal distribution $\mathscr{N}(0,\sqrt{n})$. A similar conclusion follows for the Cauchy distribution; the characteristic function is $\exp[-|t|]$, and after n samples, the characteristic function of the sum is $\exp[-n|t|]$. The distribution of the sum of n Cauchy variables has a Cauchy distribution and the "width" of this Cauchy distribution increases as n. As a final example, consider the exponential distribution $\lambda\exp[-\lambda t]$; its characteristic function is $1/(1-it/\lambda)$ Therefore, a sum of exponential random variables will not be distributed exponentially.

Note that the width of the distribution of the sum of n random variables increases with n. This is not contradictory with the earlier result on the mean of n random variables. It is not difficult to show that if the characteristic distribution for x is $\varphi(t)$, then the characteristic function x/n is $\varphi(t/n)$. From this it follows that, if n variables are drawn from the normal distribution $\mathscr{N}(0,1)$, then the mean has characteristic function $\exp[-(n/2)(t/n)^2] = \exp[-t^2/2n]$. This may be inverted to give the distribution function $\mathscr{N}(0,1/\sqrt{n})$ for the mean. The latter shrinks with n as predicted by the central limit theorem. Indeed, we see that here the limiting behavior holds exactly for any n.

In the same way, the characteristic function for the mean of n Cauchy variables is $\exp[-n|t|/n] = \exp[-|t|]$. Thus the distribution of the mean is the same as for a single variable. Again the limit distribution for the mean is exact for any n, but now the distribution does not change its width. This does not, of course, satisfy the central limit theorem because the Cauchy distribution does not satisfy the requirement of a finite variance.

In the next section we will derive the χ^2 distribution already defined in Eq. (5.7) of Section V.

A. Sum of the Square of Independent Gaussian Variables

If $\{Z_1,\ldots,Z_n\}$ are independent, standard normal random variables, $\mathscr{N}(0,1)$, then the sum of their squares,

$$X = \sum_{i=1}^{n} Z_i^2, \tag{9.6}$$

is distributed according to the χ^2 distribution with n degrees of freedom presented in Section V. This is usually denoted as $X \sim \chi^2(n)$.

This can easily be proven writing the probability distribution for X as

$$f(x)\,dx = \int_{\mathscr{V}} \frac{e^{-(z_1^2+z_2^2+\cdots+z_n^2)/2}}{(2\pi)^{n/2}}\,dz_1\,dz_2\cdots dz_n, \tag{9.7}$$

where $\mathscr{V}$ is the elemental shell volume determined by the $(n-1)$-dimensional surface in n-space for which (9.6) holds. It can be seen that this surface is the surface of an n-dimensional ball or, alternatively, an k-sphere where $k=n-1$ with radius $r=\sqrt{x}$, and that the term in the exponent is simply expressed in terms of x. Since it is a constant, it may be removed from inside the integral.

$$f(x)\,dx = \frac{e^{-x/2}}{(2\pi)^{n/2}}\int_{\mathscr{V}} dz_1\,dz_2\cdots dz_n, \tag{9.8}$$

The integral is now simply the surface area A of the $(n-1)$-sphere times the infinitesimal thickness of the sphere which is

$$dr = \frac{dx}{2x^{1/2}}. \tag{9.9}$$

---

[3] Note that if the two random variables are not independent, this statement is generally not true.

The area of a $(n-1)$-sphere is

$$A = \frac{2r^{n-1}\pi^{n/2}}{\Gamma(n/2)}. \tag{9.10}$$

Substituting, realizing that $\Gamma(z+1) = z\Gamma(z)$, and cancelling terms yields

$$f(x)\,dx = \frac{e^{-x/2}}{(2\pi)^{n/2}} A\,dr = \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}\,dx. \tag{9.11}$$

This was anticipated in Section V.

X. MONTE CARLO INTEGRATION

We here summarize the main results gathered so far about the Monte Carlo integration.

If $X_1, X_2, \ldots, X_n, \ldots$ are i.i.d. random variables with probability distribution function $f(x)$ (x does not necessarily have to be in $\mathbb{R}$, i.e. be on the real line), then for a function $g(x)$, an estimator is

$$G_N = \frac{1}{N}\sum_{i=1}^{N} g(X_i),$$
$$\langle G_N \rangle = \int_{-\infty}^{\infty} g(x) f(x)\,dx,$$

and

$$\mathrm{var}\{G_N\} = \frac{1}{N}\mathrm{var}\{g\}.$$

As $N \to \infty$ and if the variance exists, the distribution of possible values of G_N narrows about the mean as $1/\sqrt{N}$; or the probability of finding a G_N some fixed distance away from $\langle G_N \rangle$ becomes smaller. We have defined an estimator as a useful approximation to a quantity of interest which may be derived from a Monte Carlo calculation.

In the development of the Monte Carlo method so far, it has been assumed that the random variables are drawn from a continuous distribution function and that they are used to approximate an integral. Similar procedures can be employed to perform sums by Monte Carlo.

The basic random variable used in Monte Carlo has been set by historical convention to be distributed uniformly between 0 and 1, $f(x) = 1, 0 \le x \le 1$. When a variable is generated using `RAND`(u) or a similar function, it is a *pseudorandom* variable; that is, it was generated by a deterministic algorithm. Since it is possible to use truly random variables in any calculation [4], why are pseudorandom variables used? An absolute requirement in debugging a computer code is the ability to repeat a particular run of the program. If truly random numbers were used, an identical calculation could not be repeated and the recurrence of an error would be left to chance. It has been suggested that pseudorandom numbers be used to debug a program and that in the actual exercising of the program truly random numbers be used. This method also suffers from the inability to repeat a particular calculation. If after many hours of computer time an error should occur in a code (subtle logical errors occur in many Monte Carlo codes), it is of utmost importance to be able to repeat the error at will as an aid to debugging. It is also very useful to be able to repeat a calculation when changes are made or when the program is moved to a different computer. Furthermore, lengthy calculations may require the generation of an enormous number of "random variables" at great rates, and deterministic algorithms can be created to fill this need. Finally, there are situations in which the same sequence of random variables must be generated, for example, to change a parameter but retain the correlation with a previous calculation. In that case, the use of pseudorandom sequences offers a great advantage.

The procedure detailed above may be easily generalized. To evaluate the L-dimensional integral over the hypervolume $\mathscr{V}$

$$\int \cdots \int_{\mathscr{V}} g(x_1, x_2, \ldots, x_L)\,dx_1 dx_2 \cdots dx_L,$$

---

[4] True random number generators generate these numbers by tapping into unpredictable physical phenomena like hardware thermal noise, radioactive decay, atmospheric noise, quantum phenomena, chaotic systems, ..., rather than using deterministic algorithms (pseudo-random)

L uniform random variables could be sampled, the function $g(x_1, x_2, \ldots, x_L)$ calculated, and the whole process repeated N times. The arithmetic mean of the function values gives an unbiased estimate of the integral. That one can work in many dimensions is a characteristic of Monte Carlo quadrature in contrast to discrete numerical quadrature, and is a property that can be exploited to great advantage in many applications. A Monte Carlo quadrature becomes more and more efficient than the usual quadrature methods as the number of dimensions increases. It can be proven that in general the Monte Carlo integration becomes more efficient than a grid based integration as soon as the number of dimensions of the integral is bigger than twice the order of the quadrature integration scheme. This is shown in Section XII.

XI. SAMPLING RANDOM VARIABLES

We have sketched how a Monte Carlo calculation is a numerical stochastic process. The next step consists in designing and carrying out such a process so that answers to interesting questions may be obtained. In doing so, it is usually required that random variables be drawn from distribution functions that define the process. For example, to evaluate the integral $\int f(x)g(x)\,dx$, values of X must be drawn from $f(x)$ and the average value of $g(x)$ over a set of such X calculated. The process of "sampling X from $f(x)$" as it is ordinarily called, is therefore an essential technical matter. It is the purpose of this section to introduce the reader to the methods required. It will be beyond our scope to give a complete review or survey of methods or of known algorithms. Our treatment may serve to illustrate important principles, to exercise ideas of probability, but above all to demonstrate that sampling any $f(x)$ can in fact be carried out. At the same time, some specific techniques and results will be presented. First, we must define what we mean by sampling. Consider some space Ω_0 and $x \in \Omega_0$, together with a probability distribution function [5] $f(x)$, where

$$\int_{\Omega_0} f(x)\,dx = 1. \tag{11.1}$$

A sampling procedure is an algorithm that can produce a sequence of values of X ("random variables") $X_1, X_2, \ldots$ such that for any $\Omega \in \Omega_0$

$$P\{X_k \in \Omega\} = \int_{\Omega} f(x)\,dx \leq 1. \tag{11.2}$$

For a one-dimensional distribution defined on $(0, 1)$, this means that

$$P\{X_k \in (a, b)\} = \int_a^b f(x)\,dx \leq 1, \quad 0 < a < b < 1. \tag{11.3}$$

Informally, for small values of $b - a = dx$,

$$P\{X_k \in dx\} = f(x)\,dx. \tag{11.4}$$

It will be possible to do this only by already having a sequence of some basic random variables. It has become conventional to start with random variables that are independent and uniformly distributed on $(0, 1)$. We shall denote these by $\xi_1, \xi_2, \ldots$, and assume that they can be generated by a computer procedure called `RAND`(u). Such routines are widely available and usually giving satisfactory imitations of truly random variables. These are called *pseudorandom numbers*. It is important to note that here satisfactory means that the results are adequate in a particular context. No general method has ever been proved acceptable in any but the most elementary calculations, and well-known computer manufacturers have supplied bad pseudorandom generators. It is unfortunately necessary to test such generators both intrinsically and in the context of a specific class of applications. However, in the discussion that follows, an indefinite supply of uniform pseudorandom variables is assumed to exist.

A. Transformation of Random Variables

Suppose that X is a random variable with cumulative distribution function $F_X(x)$ and pdf,

$$f_X(x) = \frac{dF_X}{dx}, \tag{11.5}$$

[5] The case in which the space is discrete, and the case of mixed discrete and continuous variables need only a slight change of notation.

and that the random variable $Y = y(X)$ is a continuous nondecreasing function of x. What is $F_Y(y)$? The variable X and the function $y(X)$ map into each other $y(X) \leq y(x)$ iff $X \leq x$, so the probabilities become

$$P\{y(X) = Y \leq y(x)\} = P\{X \leq x\}, \tag{11.6}$$

or

$$F_Y(y) = F_X(x). \tag{11.7}$$

The relationship between the probability distribution functions may be determined by differentiating this equation

$$f_Y(y)\frac{dy}{dx} = f_X(x). \tag{11.8}$$

Suppose that $y(X)$ is a nonincreasing function of X; then

$$P\{y(X) \leq y(x)\} = P\{X \geq x\} = 1 - P\{X < x\}, \tag{11.9}$$

The cumulative distribution function for Y is, therefore,

$$F_Y(y) = 1 - F_X(x), \tag{11.10}$$

and

$$f_Y(y)\frac{dy}{dx} = -f_X(x). \tag{11.11}$$

The probabilities in this equation are nonnegative since dy/dx is negative. The relationship between the pdf's of X and Y for both cases can be combined in one equation as

$$f_Y(y)\left|\frac{dy}{dx}\right| = f_X(x). \tag{11.12}$$

This equation is also written as

$$|f_X(x)\,dx| = |f_Y(y)\,dy|, \tag{11.13}$$

reflecting the fact that all the values of X in dx map into values of Y in dy.

We now consider the problem of finding an algorithm to sample a specified function. This is usually the form in which the problem is posed. For univariate distributions, there is a general *inversion* technique that may be justified as follows. Let $Y = y(X)$ be an increasing function of X. The cumulative distribution function of Y may be determined from Equation (11.7). If $X = \xi$ is uniform, its cumulative distribution function is

$$F_\xi(\xi) = \begin{cases} 0 & \xi \leq 0, \\ \xi & 0 < \xi < 1, \\ 1 & \xi \geq 1. \end{cases} \tag{11.14}$$

Therefore, on $(0, 1)$ the cumulative distribution function for Y is determined by solving the equation

$$F_Y(y) = \xi \quad \text{for } Y. \tag{11.15}$$

We now give some simple examples.

Sampling an Exponential

Suppose the pdf for Y is given by

$$f_Y(y) = \lambda e^{-\lambda y}, 0 < y < \infty. \tag{11.16}$$

Then from Equations (11.15) and (11.16)

$$F_Y(y) = \int_0^y \lambda e^{-\lambda u}\,du = 1 - e^{-\lambda y} = \xi \tag{11.17}$$

which yields

$$Y = -\frac{1}{\lambda}\ln(1 - \xi), \tag{11.18}$$

an increasing function of ξ. The expression in this equation is computationally equivalent [6] to $-(1/\lambda)\ln(\xi)$. This is true since, if ξ is uniform on $(0,1)$, then $1-\xi$ is also uniform on $(0,1)$. The decision about which form to use depends both on whether $f_Y(y)$ is increasing or decreasing and on convenience.

Sampling $\frac{2}{\pi}\frac{1}{1+y^2}$

As another example, let

$$f_Y(y) = \frac{2}{\pi}\frac{1}{1+y^2}, \quad 0 < y < \infty, \tag{11.19}$$

then the cumulative distribution function is

$$F_Y(y) = \int_0^y \frac{2}{\pi}\frac{1}{1+u^2}\,du = \frac{2}{\pi}\arctan y = \xi. \tag{11.20}$$

Solving this equation for Y yields

$$Y = \tan\left(\frac{\pi}{2}\xi\right). \tag{11.21}$$

Sampling $re^{-r^2/2}$

A useful pdf is

$$f_R(r) = re^{-r^2/2}, \quad 0 < r < \infty, \tag{11.22}$$

whose cumulative distribution function is given by

$$F_R(r) = \int_0^r ue^{-u^2/2}\,du = 1 - e^{-r^2/2} = \xi. \tag{11.23}$$

A random variable that is distributed as in Eq. (11.22) is then

$$R = \sqrt{-2\ln(1-\xi)}, \tag{11.24}$$

which is computationally equivalent to

$$\sqrt{-2\ln(\xi)}. \tag{11.25}$$

Sampling a Gaussian (Box-Muller method)

It is frequently necessary to sample a Gaussian,

$$\varphi(y|0,1) = \frac{1}{\sqrt{2\pi}}e^{-y^2/2}, \quad -\infty < y < \infty, \tag{11.26}$$

in Monte Carlo calculations. In practice, it is easier to sample two independent Gaussian random variables, Y_1 and Y_2, together, than to sample a single variable

$$f(y_1,y_2) = \varphi(y_1 0,1)\varphi(y_2|0,1) = \frac{1}{2\pi}e^{-(y_1^2+y_2^2)/2}. \tag{11.27}$$

Now, the equation can be transformed into the independent polar coordinates R and Φ by the transformation

$$Y_1 = R\cos\Phi, \tag{11.28}$$
$$Y_2 = R\sin\Phi, \tag{11.29}$$

---

[6] By computationally equivalent we do not mean that the value of Y is the same in both cases. Rather, the distributions are the same and both give statistically equivalent results when used in a Monte Carlo calculation.

and rewritten as

$$f(y_1, y_2)\, dy_1 dy_2 = \left(e^{-r^2/2} r\, dr\right)\left(\frac{1}{2\pi}\, d\phi\right). \tag{11.30}$$

The angle Φ is distributed uniformly on $(0, 2\pi)$ and may be sampled by

$$\Phi = 2\pi\xi_2. \tag{11.31}$$

The probability distribution function for R is the same as that introduced in Eq. (11.22), so R can be sampled as in Eq. (11.25). The two independent Gaussian random variables become

$$Y_1 = \sqrt{-2\ln\xi_1}\cos(2\pi\xi_2), \tag{11.32}$$

$$Y_2 = \sqrt{-2\ln\xi_1}\sin(2\pi\xi_2), \tag{11.33}$$

this is known as the *Box–Muller method* (though it was invented by Wiener [7]). A good function may be programmed using the Box–Muller method; however, as written the method is convenient but slow. One advantage is that it permits the sampling of a gaussian random variable in one coded expression.

Sampling a Gaussian approximately

An approximate Gaussian random variable may also be generated by invoking the central limit theorem. By sampling N uniform random variables $\xi_1, \xi_2, \ldots, \xi_N$ and forming the sum

$$Y = \sqrt{\frac{12}{N}}\left(\sum_{k=1}^{N}\xi_k - \frac{N}{2}\right), \tag{11.34}$$

a variable with mean 0 and variance 1 is generated. The central limit theorem asserts that this will be nearly gaussian for large N (See Eq. (5.4 and Section VIII C). A value of $N = 12$ appears to be sufficiently large for many purposes and avoids the evaluation of the factor $\sqrt{12/N}$.

B. Numerical Transformation

In the preceding examples, an expression for sampling a random variable was derived by applying

$$F_T(y) = \xi. \tag{11.35}$$

This equation can always be used if $y(\xi)$ is an increasing function, but in practice solving it may require a slow iteration procedure. For example, a transcendental equation must be solved anew for each value of ξ.

An attempt to generate a Gaussian random variable by the equation $\varphi(x|0,1) = \xi$ requires solving the expression for the error function,

$$\frac{1}{\sqrt{2\pi}}\int_0^x e^{-u^2/2}\, du = \mathrm{erf}(x) = \xi, \tag{11.36}$$

for X. The method of setting a uniform random variable equal to the cumulative distribution function is useful only if the resulting equation is economically solvable. Finally, it may be that $f(y)$ is known only numerically (e.g. from experimental data).

C. Sampling Discrete Distributions

Let ξ be a random variable uniformly distributed in $(0, 1)$. Using Eq. (4.3), we have that

$$P\{0 \le x_1 < \xi \le x_2 \le 1\} = F_\xi(x_2) - F_\xi(x_1) = x_2 - x_1. \tag{11.37}$$

The chance that ξ lies in an interval $[x_1, x_2]$ of $(0, 1)$ is equal to the length of the interval. Suppose we have a class of events E_k with probabilities f_k and we wish to sample one at random. We may generate a uniform variable ξ and, if it lies in an interval of length f_k on $(0, 1)$, assign event k to that trial. Better, since $\sum_k f_k = 1$, it is possible to

take the interval $(0,1)$ and exhaust it by dividing it into segments each of which has a length equal to some f_l. The interval into which a ξ falls determines the identity of the event.

A uniform random variable is generated, and the smallest l is found for which the sum of the f_k is greater than the random number; that is,

$$\sum_{k=0}^{l-1} f_k < \xi \leq \sum_{k=0}^{l} f_k. \tag{11.38}$$

(When $l = 0$, the sum is defined to equal 0.) Whenever $0 < \xi < f_1$, event 1 takes place; if $f_1 < \xi < f_1 + f_2$, event 2 takes place; and so on.

Thus, if we must choose between equally likely events, we may consider the first to be selected if some $\xi < 1/2$ otherwise, the second is selected. To select three events with probabilities $1/2, 1/4$, and $1/4$, we choose the first if $\xi < 1/2$, the second if $\xi < 3/4$, and the third otherwise.

Suppose we must choose among K equally likely events,

$$f_k = 1/K \quad k = 1, 2, ..., K. \tag{11.39}$$

The sums in Eq. (11.38) are formed and reduced to

$$\frac{l-1}{K} < \xi \leq \frac{l}{K}, \tag{11.40}$$

The appropriate value of l for a particular ξ is then

$$l = [\xi K] + 1, \tag{11.41}$$

where $[u]$ indicates truncation to the largest integer less than u.

In searching for an index l satisfying Eq. (11.38), a binary search is strongly recommended when the total number of intervals is large and if $\sum f_k$ does not converge fast. If a serial search is to be used and the index can be arranged at our disposal, then the index with the largest probability should be put in the first place, and so on, to reduce the average time of searching.

D. Composition of Random Variables

We have indicated that transforming or inversion of random variables may lead to unpleasantly complicated equations to solve numerically. Another technique for generating random variables having a required distribution is to take two or more different (usually independent) random variables drawn from known distributions and combine them in interesting ways. The Box–Muller method of Eqs. (11.32) and (11.33) is in fact an example of sampling by composition. The simplest example is to simply add two independent random variables.

Sampling the Sum of Two Uniform Random Variables

Let ξ_1 and ξ_2 be uniform on $(0,1)$ and $Z = \xi_1 + \xi_2$; then

$$F_Z(z) = P\{\xi_1 + \xi_2 < z\} \tag{11.42}$$

is the area under the line $Z = \xi_1 + \xi_2$ within the unit square $(0,1) \times (0,1)$ in the plane (ξ_1, ξ_2). Consider the case where $\xi_1 + \xi_2 < 1$. Geometrically, $F_Z(z)$ is seen to be the area of the triangle with sides equaling z

$$F_Z(z) = \frac{1}{2} z^2. \tag{11.43}$$

For $\xi_1 + \xi_2 > 1$, the cumulative distribution function is

$$F_Z(z) = 1 - \frac{1}{2}(2 - z)^2. \tag{11.44}$$

The corresponding pdf 's are

$$f_Z(z) = \begin{cases} z & 0 < z < 1, \\ 2 - z & 1 \leq z < 2, \end{cases} \tag{11.45}$$

It is left as an exercise for the reader to find the pdf when two uniform random variables are multiplied.

E. Rejection Techniques

There is a kind of composition method that leads to very general techniques for sampling any probability distribution. It has one new feature, namely, that a trial value for a random variable is selected and proposed. This value is subjected to one or more tests (involving one or more other random variables) and it may be accepted, that is, used as needed, or rejected. If it is rejected, the cycle of choosing and testing a trial value is repeated until an acceptance takes place. An important property of the method is that the normalization of the distribution need not be known explicitly to carry out the sampling.

A disadvantage is that it may have low *efficiency*, that is, many values are rejected before one is accepted. It is an interesting technical challenge to devise efficient rejection techniques for varieties of distributions.

F. Multivariate Distributions

Except for the bivariate normal distribution, which was introduced as a technical device, our treatment of sampling has been devoted to random variables in one dimension. Multivariate distributions are also important since Monte Carlo is at its best in treating many-dimensional problems. One may use the ideas of marginal and conditional distributions (Eq. (6.4) and the discussion that follows) to reduce multivariate to univariate sampling.

Sampling a Brownian Bridge

In a number of interesting applications (e.g. [8]), it becomes necessary to sample a "Brownian bridge" which may be defined as follows. For some integer M and real numbers τ and x_i, $0 \le i \le M$, let

$$B(x_0, x_1, \ldots, x_M, \tau) = \sqrt{2\pi M\tau}\, e^{(x_M - x_0)^2/2M\tau} \prod_{i=1}^{M} \frac{e^{-(x_{i-1}-x_i)^2/2\tau}}{\sqrt{2\pi\tau}} \tag{11.46}$$

and

$$\int \cdots \int B(x_0, x_1, \ldots, x_M, \tau)\, dx_1 dx_2 \ldots dx_{M-1} = 1. \tag{11.47}$$

Suppose that x_0 and x_M are fixed and it is required to sample the variables $x_1, x_2, \ldots, x_{M-1}$. To accomplish this, we compute a marginal distribution for (say) x_j, $0 < j < M$, by integrating over $x_1, x_2, \ldots, x_{j-1}, x_{j+1}, \ldots, x_{M-1}$. The identity

$$\int \frac{e^{-(x-y)^2/2\sigma_1^2}}{\sqrt{2\pi\sigma_1^2}} \frac{e^{-(y-z)^2/2\sigma_2^2}}{\sqrt{2\pi\sigma_2^2}}\, dy = \frac{e^{-(x-z)^2/2(\sigma_1^2+\sigma_2^2)}}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}}, \tag{11.48}$$

may be used recursively to find

$$\begin{aligned} &\int \cdots \int B(x_0, x_1, \ldots, x_M, \tau)\, dx_1 dx_2 \ldots dx_{j-1} dx_{j+1 \ldots} dx_{M-1} \\ &= \sqrt{2\pi M\tau}\, e^{(x_M - x_0)^2/2M\tau} \frac{e^{-(x_0-x_j)^2/2j\tau}}{\sqrt{2\pi j\tau}} \frac{e^{-(x_j-x_M)^2/2(M-j)\tau}}{\sqrt{2\pi(M-j)\tau}} \\ &= \frac{e^{-(x_j-\bar{x})^2/2\sigma_j^2}}{\sqrt{2\pi\sigma_j^2}} \end{aligned} \tag{11.49}$$

where

$$\bar{x} = \frac{jx_M + (M-j)x_0}{M}, \tag{11.50}$$

$$\sigma_j^2 = \frac{j(M-j)\tau}{M}. \tag{11.51}$$

Thus, the marginal distribution for x_j given x_0 and x_M is a Gaussian or normal distribution with mean $\bar{x}$ and variance σ_j^2, and may be sampled in any of the ways discussed previously in Section XI A Then given x_j and x_0, one may sample x_k, $0 < k < j$, using a Gaussian distribution with mean $(kx_j + (j-k)x_0)/j$ and variance $k(j-k)\tau/j$. Similarly, given

x_j and x_M one may sample x_l, $j < l < M$, from a Gaussian distribution with mean $((l-j)x_M + (M-l)x_j)/(M-j)$ and variance $(l-j)(M-l)\tau/(M-j)$.

One technique that is often used is to let M be 2^n and use recursive *bisection* in which j is first set to 2^{n-1} and the intervals are successively halved.

The generalization to more dimensions follows a similar approach. The variable $\bar{x}$ becomes the mean of two vectors, and the x_j is sampled from the appropriate multivariate Gaussian distributions.

G. The M(RT)^2 Algorithm

The last sampling method we shall discuss is an advanced sampling technique first described in a paper by Metropolis [7] et al. [9], M(RT)^2. The method is related to rejection techniques of Section XI E since it involves explicitly proposing a tentative value that may be rejected and because the normalization of the sampled function is irrelevant, we never need to know it.

The M(RT)^2 algorithm is very simple and powerful; it can be used to sample essentially any distribution function regardless of analytic complexity in any number of dimensions. Complementary disadvantages are that sampling is correct only asymptotically and that successive variables produced are correlated, often very strongly. This means that the evaluation of integrals normally produces positive correlations in the values of the integrand, with consequent increase in variance for a fixed number of steps as compared with independent samples. Also the method is not well suited to sampling distributions with parameters that change frequently.

The method was motivated by an analogy with the behavior of systems in statistical mechanics that approach an equilibrium whose statistical properties are independent of the kinetics of the system.

By *system*, we mean here simply a point x in a space (typically in $\mathbb{R}^d$) that may be thought of as a possible description of a physical problem. By *kinetics*, we mean a stochastic transition that governs the evolution of the system: a probability distribution function $K(X|Y)$ that ensures the evolution of a system known to be at Y will bring it near X next. K may be thought of as a model of the physical process by which a system changes or as a mathematical abstraction. In a Monte Carlo calculation, it plays the role of a sampling distribution. As we shall discuss in detail, one condition for a system to evolve toward equilibrium and stay there is, quite simply, that the system be on the average as likely to move into a specific neighborhood of X from a neighborhood of Y as to move exactly in the reverse direction. If the probability distribution for observing the system near X in equilibrium is $f(X)$, then the kinetics must satisfy

$$K(X|Y)f(Y) = K(Y|X)f(X). \tag{11.52}$$

This relation is called *detailed balance*. $K(X|Y)f(Y)$ is the probability of moving from Y to X expressed as the a priori chance of finding the system near Y [i.e. $f(Y)$] times the conditional probability [i.e. $K(X|Y)$] that it will move to X from Y.

In treating a physical system, one usually assumes that $K(X|Y)$ is known, and one has the task of finding $f(X)$. The M(RT)^2 algorithm (as in much of Monte Carlo) reverses this: one has the task of finding a convenient and correct kinetics that will equilibrate the system so that the given $f(X)$ turns out to be the chance of observing the system near X.

This turns out to be extremely easy given the elegant device suggested by M(RT)^2. Transitions are *proposed* from, say, Y to X' using essentially *any* distribution $T(X'|Y)$. Then on comparing $f(X')$ with $f(Y)$ and taking into account T as well, the system is either moved to X' (move "accepted") or returned to Y (move "rejected"). Acceptance of the move occurs with probability $A(X'|Y)$, which must be calculated so as to satisfy detailed balance.

We then have

$$K(X|Y) = A(X|Y)T(X|Y). \tag{11.53}$$

Detailed balance requires

$$A(X|Y)T(X|Y)f(Y) = A(Y|X)T(Y|X)f(X). \tag{11.54}$$

We expect that the ratio

$$\frac{T(Y|X)f(X)}{T(X|Y)f(Y)} \tag{11.55}$$

---

[7] In the literature, it is often referred to as the Metropolis algorithm.

will play a significant role in determining A.

Given a pdf $f(X)$, where X is a many-dimensional vector, the M(RT)^2 technique establishes a random walk whose steps are designed so that when repeated again and again, the asymptotic distribution of the X is $f(X)$. Suppose that $X_1, X_2, X_3, \ldots, X_n$ are the steps in a random walk. Each of the X is a random variable and has an associated probability $\phi_1(X), \phi_2(X), \phi_3(X), \ldots, \phi_n(X)$, where $\phi_1(X)$ can be any distribution for X. The $\phi_n(X)$ have the property that asymptotically

$$\lim_{n\to\infty} \phi_n(X) = f(X). \tag{11.56}$$

At each step, in the random walk, there is a transition distribution $T(X|Y)$, that is, the probability distribution function for a trial move to X from Y. The $T(X|Y)$ is normalized such that

$$\int T(X|Y)\, dX = 1, \tag{11.57}$$

for all values of Y. A quantity $q(X|Y)$ is defined as

$$q(X|Y) = \frac{T(Y|X)f(X)}{T(X|Y)f(Y)} \geq 0, \tag{11.58}$$

where we explicitly assume that it is possible to move from X to Y if one can move from Y to X and vice versa. From $q(X|Y)$, the probability of accepting a move can be calculated; one frequently used possibility is

$$A(X|Y) = \min\{1, q(X|Y)\}. \tag{11.59}$$

The algorithm can now be described concretely. At step n of the random walk, the value of X is X_n; a possible next value for X, X'_{n+1}, is sampled from $T(X'_{n+1}|X_n)$, and the probability of accepting X'_{n+1} is computed. If $q(X'_{n+1}|X_n) > 1$ then $A(X'_{n+1}|X_n) = 1$; if $q(X'_{n+1}|Xn) < 1$, then $A(X'_{n+1}|X_n) = q(X'_{n+1}|X_n)$, where

$$q(X'_{n+1}|X_n) = \frac{T(X_n|X'_{n+1})f(X'_{n+1})}{T(X'_{n+1}|X_n)f(X_n)}. \tag{11.60}$$

With probability $A(X'_{n+1}|X_n)$, we set $X_{n+1} = X'_{n+1}$; Otherwise,we set $X_{n+1} = X_n$. That is, if $A(X'_{n+1}|X_n) > \xi$, then $X_{n+1} = X'_{n+1}$; otherwise, $X_{n+l} = X_n$. For $q(X'_{n+1}|X_n) > 1$, X_{n+1} will always equal X'_{n+1}. This procedure contains an element of rejection; however, if an X'_{n+1} is not accepted, we use the previous value rather than sample a new value.

As the random walk proceeds, a recursive relationship develops between succeeding $\phi_n(X)$. Let $\phi_N(X)$ be the distribution of values of X_n; what is the distribution ϕ_{n+1} for the values of X_{n+1}? There are two contributions to the distribution of the X_{n+1}: the probability of entering into the vicinity dX of X when we successfully move from X_n and the probability that once we are at X, we will stay at X. If we start out at some value Y contained in dY, the probability of moving from the neighborhood of Y to the neighborhood of X is $T(X|Y)\phi_n(Y)\, dY$. The probability of successfully moving from Y to X is $A(X|Y)T(X|Y)\phi_n(Y)\, dY$, so the net probability of successfully moving from any point Y to a neighborhood of X becomes

$$\int A(X|Y)T(X|Y)\phi_n(Y)\, dY. \tag{11.61}$$

In a similar manner, the net probability that a move away from X is not accepted is

$$\int [1 - A(Y|X)]T(Y|X)\, dY, \tag{11.62}$$

where $T(Y|X)$ is the probability of moving from X to Y and $[1 - A(Y|X)]$ is the probability that the move was not accepted. Upon multiplying Eq. (11.62) by $\phi_n(X)$, the probability that we were at X, the relationship for $\phi_{n+1}(X)$ becomes

$$\phi_{n+1}(X) = \int A(X|Y)T(X|Y)\phi_n(Y)\, dY + \phi_n(X) \int [1 - A(Y|X)]T(Y|X)\, dY. \tag{11.63}$$

The random walk generates a recursion relationship for the distribution functions.

Earlier, we asserted that the asymptotic distribution sampled in the random walk would be $f(X)$. According to a theorem in Feller [10], if a random walk defines a system that is *ergodic*, then an asymptotic pdf exists and is unique if

$$\phi_n(X) = f(X) \implies \phi_{n+1}(X) = f(X), \tag{11.64}$$

that is, if $f(X)$ is a stationary point of the recursion. Systems defined by random walks can be partitioned into several categories. If, in a random walk, the probability of returning to a neighborhood about X is 0, then the system is called a *null system* and the expected recurrence time is infinite. An example would be a one-dimensional system where X_{n+1} is constrained to be greater than X_n. A system where the random walk will return to the neighborhood of X every T steps is called *periodic*. An ergodic system is one in which the random walk may return to the neighborhood of X, but does not do so periodically; it is neither null nor periodic.

The system generated by the M(RT)^2 sampling method is ergodic, but the proof will be omitted [11]. The system obeys detailed balance, which guarantees that, if we can move from X to Y, we can move from Y to X and the expected number of moves is the same in each direction when the asymptotic behavior is reached, i.e. Eq. (11.54) holds. In this equation the left-hand side is the net number of moves from Y to X and the right-hand side is the net number of moves from X to Y. Equation (11.54) is easily proved by using $q(X|Y)q(Y|X) = 1$, which follows from the definition of $q(X|Y)$. Suppose that $q(X|Y) > 1$; then $A(X|Y) = 1$. $A(Y|X) = q(Y|X)$ since $q(Y|X) < 1$. Substituting these into Equation (11.54), we derive

$$T(X|Y)f(Y) = q(Y|X)T(Y|X)f(X) = \frac{T(X|Y)f(Y)}{T(Y|X)f(X)}T(Y|X)f(X). \tag{11.65}$$

The same answer would have resulted if we had chosen $q(Y|X) > 1$, so the algorithm developed in Eq. (11.58) and (11.59) satisfies detailed balance. If we set $\phi_n(X) = f(X)$ in Eq. (11.63), the resulting equation is

$$\phi_{n+1}(X) = \int A(X|Y)T(X|Y)\phi_n(Y)\,dY + \int [1 - A(Y|X)]T(Y|X)f(X)\,dY. \tag{11.66}$$

The first integral cancels the negative portion of the second integral by detailed balance, and we are left with

$$\phi_{n+1}(X) = \int T(Y|X)f(X)\,dY = f(X). \tag{11.67}$$

since $\int T(Y|X)\,dY = 1$. Therefore, $f(X)$ is guaranteed to be the asymptotic distribution of the random walk.

The form for the probability of accepting a move is not limited to that given in Eq. (11.59). Another relation that has been used is

$$A'(X|Y) = \frac{q(X|Y)}{1 + q(X|Y)} \tag{11.68}$$

and

$$A'(Y|X) = \frac{q(Y|X)}{1 + q(Y|X)} = \frac{1}{q(X|Y) + 1}. \tag{11.69}$$

A random walk whose probability of accepting a move is governed by Eq. (11.69) also exhibits detailed balance:

$$A'(X|Y)T(X|Y)f(Y) = A'(Y|X)T(Y|X)f(X). \tag{11.70}$$

Upon substituting Eq. (11.69) for A' on both sides of the equation we find

$$\frac{q(X|Y)}{1 + q(X|Y)}T(X|Y)f(Y) = \frac{1}{1 + q(X|Y)}T(Y|X)f(X), \tag{11.71}$$

which reduces to

$$\frac{T(Y|X)f(X)}{T(X|Y)f(Y)}T(X|Y)f(Y) = T(Y|X)f(X). \tag{11.72}$$

Either form for $A(X|Y)$, Eq. (11.59) or (11.69), may be used in the M(RT)^2 method. The former has been shown to give more rapid convergence in certain cases [12].

In much of the literature of statistical physics, when a Monte Carlo calculation is mentioned what is meant is an application of the method of Metropolis et al. The great utility of the M(RT)^2 method is that it enables us to sample very complicated many-dimensional probability distribution functions in a simple, straightforward way. Unfortunately, the method does have major drawbacks of which the user must be aware. We are guaranteed to sample $f(X)$, but only asymptotically; therefore, we must throw away L steps of the random walk until the steps are being sampled from $f(X)$. Furthermore, L is very difficult to estimate in advance. Normally, substantial trial and error is used to estimate an appropriate value. The number of steps discarded may be minimized by selecting a $\phi_1(X)$ that is as close an approximation to $f(X)$ as possible. In addition, by making $T(X|Y)$ approximate $f(X)$, rapid convergence and

small correlation are obtained. Note that were it possible to sample $T(X|Y) = f(X)$ exactly, then (cf. Eqs. (11.58) and (11.59)) all moves are accepted and the samples are independent. In that case, of course, one would not resort to the M(RT)^2 random walk, but it is clear that approximations to this limiting case may be fruitful. In a Monte Carlo calculation, we are often trying to evaluate quantities of the form

$$G = \frac{\int g(X)f(X)\,dX}{\int f(X)\,dX}. \tag{11.73}$$

For example, if we are trying to simulate the equilibrium properties of a many-body system, G might be the energy or the radial distribution function (the probability that pairs of particles are found at varying separations). In practice, for each quantity of interest, a different number of steps in the random walk may have to be discarded since the asymptotic limit of the system is reached at varying rates. The averaging over the steps in the random walk begins only after the L steps have been thrown away, that is,

$$G = \frac{\int g(X)f(X)\,dX}{\int f(X)\,dX} = \sum_{n=L}^{L+N-1} \frac{g(X_n)}{N}. \tag{11.74}$$

The successive X_n in the random walk are not independent and in most circumstances there is positive correlation. The variance of the calculated G will then be larger than if the steps were independent.

In the most usual applications, $T(X|Y)$ is taken to be uniform over a small domain (e.g. a square or cube of side s) in the space of X or in a subspace thereof.

The "lore" for using the M(RT)^2 method recommends that the average acceptance probability should be approximately 50% to avoid: too high acceptance ratios when little new information is added at each Monte Carlo step or too low acceptance ratios when many Monte Carlo steps are needed to gain some new information. Moreover one should take account of the computer time used in the sampling computation of $f(X)$.

XII. MONTE CARLO EVALUATION OF FINITE-DIMENSIONAL INTEGRALS

In this section, we explore the ideas that underlie Monte Carlo quadrature somewhat more systematically. If an integral having the form

$$G = \int_\Omega g(x)f(x)\,dx, \tag{12.1}$$

where

$$f(x) \geq 0, \quad \int_\Omega f(x)\,dx = 1, \tag{12.2}$$

must be evaluated, then the following game of chance may be used to make numerical estimates. We draw a set of variables $X_1, X_2, \ldots, X_N$ from $f(x)$ (i.e. we "sample" the probability distribution function $f(x)$ in the sense defined in Section VII) and form the arithmetic mean

$$G_N = \frac{1}{N}\sum_i g(X_i). \tag{12.3}$$

The quantity G_N is an estimator for G and the fundamental theorem of Monte Carlo guarantees that

$$\langle G_N \rangle = G, \tag{12.4}$$

if the integral (12.1) exists. Since G_N estimates G, we can write

$$G_N = G + error. \tag{12.5}$$

If the variance exists, the error appearing in the last statement is a random variable whose mean is 0 and whose width is characterized for large N by

$$|\text{error}| = \epsilon \sim \frac{\sigma_1}{\sqrt{N}} \tag{12.6}$$

where

$$\sigma_1^2 = \int_\Omega g^2(x)f(x)\,dx - G^2. \tag{12.7}$$

The error estimate may be inverted to show the number of samples needed to yield a desired error, ϵ,

$$N \sim \frac{\sigma_1^2}{\epsilon^2}. \tag{12.8}$$

The integral in Equation (12.1) could also be evaluated by numerical quadrature. Let us assume that the domain of integration is an n-dimensional unit hypercube; then a numerical integration procedure can be written:

$$G \sim \sum_i w_i g(X_i) f(X_i), \tag{12.9}$$

where X_i is a lattice of points that fills the unit hypercube and w_i is a series of quadrature weights. The error associated with this quadrature is bounded by

$$\epsilon \leq c^k, \tag{12.10}$$

where h measures the size of the interval separating the individual X_i. The constants c and k depend on the actual numerical integration method used, and k normally increases with more accurate rules. The bound on the error in Eq. (12.10) is not a statistical variable, but is an absolute number. The actual error, however, is usually predictable to some degree.

If we assume that the time necessary for a computation, T_c, will be proportional to the total number of points used, then

$$T_c \propto N = N_0 \left(\frac{1}{h}\right)^n, \tag{12.11}$$

where N_0 is a constant of the order of 1 and n is the number of dimensions. Eq. (12.10) can be rewritten as

$$h \geq \left(\frac{\epsilon}{c}\right)^{1/k}, \tag{12.12}$$

and Eq. (12.11) becomes

$$T_c \propto N_0 \left(\frac{c}{\epsilon}\right)^{n/k} = t_0 \epsilon^{-n/k}. \tag{12.13}$$

The greater the accuracy demanded in a calculation, the greater the computational time will be.

In a Monte Carlo calculation, the total computation time is the product of the time t_1 for an individual sampling of times the total number of points;

$$T_c = t_1 N. \tag{12.14}$$

From Eq. (12.8) this may be rewritten as

$$T_c = \frac{t_1 \sigma_1^2}{\epsilon^2} = \frac{t_1'}{\epsilon^2}, \tag{12.15}$$

the exponent of ϵ is the same in any number of dimensions. For large n, it is difficult to find a k in Eqs. (12.10) and (12.13) such that $n/k < 2$, so asymptotically, as $n \to \infty$, a Monte Carlo calculation is more advantageous than a numerical integration of Eq. (12.1). The Monte Carlo calculation will take less total time for the same value of ϵ. This assumes that the two error estimates can be directly compared.

In spite of the apparently slow convergence ($\sim 1/\sqrt{N}$) of the error of Monte Carlo quadrature, it is in fact more efficient computationally than finite difference methods in dimensions higher than $6 - 10$.

Two different Monte Carlo evaluations of an integral can have differing variances. The quantity

$$Q_1 = t_1 \sigma_1^2, \tag{12.16}$$

is a measure of the quality (efficiency) of a Monte Carlo calculation. The decision on which Monte Carlo algorithm to use in a large computation can be based on the values of Q_1 extracted from some trial calculations. A common phenomenon is for t_1 to increase as σ_1 decreases through a more elaborate Monte Carlo algorithm. The question is then whether the decrease in σ^2 will more than compensate for the increase in time. It will, if Q_1 decreases.

Three major classes of techniques are used to reduce the variance in Monte Carlo quadrature:

1. *Importance sampling* can be introduced into the calculation to increase the likelihood of sampling variables where the function is large or rapidly varying.

2. The expected value of a random variable can be used rather than the variable itself. This substitution never increases variance and, many times, will substantially reduce it.

3. Correlations between succeeding samples may be exploited to advantage. In *control variates*, an easily evaluated approximation to the integrand is used to reduce the variance. If successive random variables are negatively correlated, the variance will be smaller than if they were independent. The technique called *antithetic variates* exploits the reduction in variance that results when negatively correlated samples are deliberately produced and grouped together.

We then see how the Monte Carlo integration method is ideally suited to evaluate functional integrals which can be discretized into highly multi-dimensional integrals as shown in Ref. [8].

A. Importance Sampling

Suppose we have an n-dimensional integral

$$G = \int_\Omega g(x) f(x)\, dx, \tag{12.17}$$

that we wish to evaluate. The function $f(x)$ is not necessarily the best pdf to use in the Monte Carlo calculation even though it appears in the integrand. A different pdf, $\tilde{f}(x)$, can be introduced into the integral as follows

$$G = \int_\Omega \left[\frac{g(x)f(x)}{\tilde{f}(x)}\right] \tilde{f}(x)\, dx, \tag{12.18}$$

where

$$\tilde{f}(x) \geq 0, \quad \int_\Omega tilde f(x)\, dx = 1, \tag{12.19}$$

and $g(x)f(x)/\tilde{f}(x) < \infty$ except perhaps on a countable set of points. The variance of G when $\tilde{f}(x)$ is used becomes

$$\mathrm{var}\{G\}_{\tilde{f}} = \int_\Omega \left[\frac{g(x)f(x)}{\tilde{f}(x)}\right]^2 \tilde{f}(x) - G^2\, dx. \tag{12.20}$$

G^2 being fixed, we want the $\tilde{f}(x)$ that will minimize the quantity $\int [g^2(x)f^2(x)/\tilde{f}(x)]\, dx$. Of course, the integral is minimized by choosing $\tilde{}\,\tilde{f}(x)$ as large as we like, but we have the additional constraint expressed by Eq. (12.19). The function $\tilde{}\,\tilde{f}(x)$ that satisfies the criteria given above may be deduced by using a Lagrange multiplier λ. In this method, we wish to find $\tilde{f}(x)$ such that

$$L[\tilde{f}] = \int_\Omega \frac{g^2(x)f^2(x)}{\tilde{f}(x)}\, dx + \lambda^2 \int_\Omega \tilde{f}(x)\, dx, \tag{12.21}$$

is minimized. We consider small variations of $\tilde{f}(x)$ on the quantity in the functional L and set the variation of L equal to zero

$$\frac{\delta L[\tilde{f}]}{\delta \tilde{f}} = 0. \tag{12.22}$$

Performing the functional differentiation yields

$$-\left[\frac{g(x)f(x)}{\tilde{f}(x)}\right]^2 + \lambda^2 = 0, \tag{12.23}$$

or

$$\tilde{f} = \lambda |g(x)f(x)|. \tag{12.24}$$

The value of λ may be found by requiring that $\int \tilde{f}(x)\, dx = 1$. If $g(x) \geq 0$, then $\tilde{f}(x) = \lambda g(x)f(x)$ and $\lambda = 1/G$, so

$$\tilde{f}(x) = \frac{g(x)f(x)}{G}. \tag{12.25}$$

Then a Monte Carlo algorithm to evaluate the integral would be to sample a series of X_i from $\tilde{}\tilde{f}(x)$ and construct the sum

$$\tilde{G}_N = \frac{1}{N}\sum_{i=1}^{N}\frac{g(X_i)f(X_i)}{\tilde{f}(X_i)} = \frac{1}{N}\sum_{i=1}^{N}\frac{g(X_i)f(X_i)}{g(X_i)f(X_i)/G} = \frac{1}{N}\sum_{i=1}^{N}G = G. \tag{12.26}$$

If we already know the correct answer G, the Monte Carlo calculation will certainly give it back with zero variance! This clearly corresponds to the minimum variance calculation. Although we cannot in practice use the $\tilde{}\ \tilde{f}(x)$ prescribed by Eq. (12.25) we expect that "similar" functions will reduce the variance. One important criterion is that $g(X_i)f(X_i)/\tilde{f}(X_i)$ be bounded from above.

XIII. MARKOV PROCESSES

The playground for random variables are *stochastic processes* of which the *Markov process* plays a privileged role since it allows a more complete theory. We will here talk interchangeably about random variable or *noise*.

A. Stochastic Processes

We are now embarking on an essential aspect of our description, namely, how stochastic aspects of noise $\eta(t)$ are properly accounted for. Obviously, a particular realization of the time-dependent process $\eta(t)$ does not provide much information. Rather, one needs to consider the probability of observing a certain sequence of noise values $\eta_1, \eta_2, \ldots$ at times $t_1, t_2, \ldots$. The essential information is entailed in the conditional probabilities (2.3)

$$p(\eta_1, t_1; \eta_2, t_2; \ldots | \eta_0, t_0; \eta_{-1}, t_{-1}; \ldots), \tag{13.1}$$

when the process is assumed to generate noise at fixed times t_i, $t_i < t_j$ for $i < j$. Here $p(|)$ is the probability that the random variable $\eta(t)$ assumes the values $\eta_1, \eta_2, \ldots$ at times $t_1, t_2, \ldots$, if it had previously assumed the values $\eta_0, \eta_{-1}, \ldots$ at times $t_0, t_{-1}, \ldots$.

B. The Markov case

An important class of random processes are the so-called *Markov processes* for which the conditional probabilities depend only on η_0 and t_0 and not on earlier occurrences of noise values. In this case holds

$$p(\eta_1, t_1; \eta_2, t_2; \ldots | \eta_0, t_0; \eta_{-1}, t_{-1}; \ldots) = p(\eta_1, t_1; \eta_2, t_2; \ldots | \eta_0, t_0). \tag{13.2}$$

This property allows one to factorize $p(|)$ into a sequence of consecutive conditional probabilities.

$$\begin{aligned} p(\eta_1, t_1; \eta_2, t_2; \ldots | \eta_0, t_0; \eta_{-1}, t_{-1}; \ldots) &= p(\eta_2, t_2; \eta_3, t_3; \ldots | \eta_1, t_1) p(\eta_1, t_1 | \eta_0, t_0) \\ &= p(\eta_3, t_3; \eta_4, t_4; \ldots | \eta_2, t_2) p(\eta_2, t_2 | \eta_1, t_1) p(\eta_1, t_1 | \eta_0, t_0) \\ &\vdots \end{aligned}$$

The unconditional probability for the realization of $\eta_1, \eta_2, \ldots$ at times $t_1, t_2, \ldots$ is

$$p(\eta_1, t_1; \eta_2, t_2; \ldots) = \sum_{\eta_0} p(\eta_0, t_0) p(\eta_1, t_1 | \eta_0, t_0) p(\eta_2, t_2 | \eta_1, t_1) \cdots, \tag{13.3}$$

where $p(\eta_0, t_0)$ is the unconditional probability for the appearance of η_0 at time t_0. One can conclude from Eq. (13.3) that a knowledge of $p(\eta_0, t_0)$ and $p(\eta_i, t_i | \eta_{i-1}, t_{i-1})$ is sufficient for a complete characterization of a Markov process.

Before we proceed with three important examples of Markov processes we will take a short detour and give a quick introduction on mathematical tools that will be useful in handling probability distributions like $p(\eta_0, t_0)$ and $p(\eta_i, t_i | \eta_{i-1}, t_{i-1})$.

C. Characteristics of Probability Distributions

For simplicity we now deal with a one-dimensional random variable η with values on the complete real axis. In probability theory the Fourier transform $G(s,t)$ of $p(\eta,t)$ is referred to as the *characteristic function* of $p(\eta,t)$ (See also Section VIII B)

$$G(s,t) = \int_{-\infty}^{\infty} d\bar{\eta}\, p(\bar{\eta},t) e^{i\bar{\eta}s} \tag{13.4}$$

Since the Fourier transform can be inverted to yield $p(\eta,t)$

$$p(\eta,t) = \int_{-\infty}^{\infty} \frac{ds}{2\pi}\, G(s,t) e^{-is\eta}, \tag{13.5}$$

$G(s,t)$ contains all information on $p(\eta,t)$.

The characteristic function can be interpreted as an average of $e^{is\eta(t)}$, and denoted by

$$G(s,t) = \left\langle e^{is\eta(t)} \right\rangle. \tag{13.6}$$

Eq. (13.6) prompts one to consider the Taylor expansion of (13.6) for (is) around 0:

$$G(s,t) = \sum_{n=0}^{\infty} \langle \eta^n(t) \rangle \frac{(is)^n}{n!}, \tag{13.7}$$

where

$$\langle \eta^n(t) \rangle = \int d\eta\, \eta^n p(\eta,t). \tag{13.8}$$

are the so-called *moments* of $p(\eta,t)$. One can conclude from (13.5), (13.7), and (13.8) that the moments $\langle \eta^n(t) \rangle$ completely characterize $p(\eta,t)$.

The moments $\langle \eta^n(t) \rangle$ can be gathered in a statistical analysis as averages of powers of the stochastic variable $\eta(t)$. Obviously, it is of interest to employ averages which characterize a distribution $p(\eta,t)$ as succinctly as possible, i.e., through the smallest number of averages. Unfortunately moments $\langle \eta^n(t) \rangle$ of all orders of n contain significant information about $p(\eta,t)$. There is another, similar, but more useful scheme to describe probability distributions $p(\eta,t)$; the *cumulants* $\langle\langle \eta^n(t) \rangle\rangle$. As moments are generated by the characteristic function $G(s,t)$, cumulants are generated by the logarithm of the characteristic function $\ln[G(s,t)]$

$$\ln[G(s,t)] = \sum_{n=1}^{\infty} \langle\langle \eta^n(t) \rangle\rangle \frac{(is)^n}{n!}. \tag{13.9}$$

Cumulants can be expressed in terms of $\langle \eta^n(t) \rangle$ by taking the logarithm of Eq. (13.7) and comparing the result with (13.9). The first three cumulants are

$$\left\langle\left\langle \eta^1(t) \right\rangle\right\rangle = \left\langle \eta^1(t) \right\rangle, \tag{13.10}$$

$$\left\langle\left\langle \eta^2(t) \right\rangle\right\rangle = \left\langle \eta^2(t) \right\rangle - \left\langle \eta^1(t) \right\rangle^2, \tag{13.11}$$

$$\left\langle\left\langle \eta^3(t) \right\rangle\right\rangle = \left\langle \eta^3(t) \right\rangle - 3 \left\langle \eta^2(t) \right\rangle \left\langle \eta^1(t) \right\rangle + 2 \left\langle \eta^1(t) \right\rangle^3. \tag{13.12}$$

These expressions reveal that the first cumulant is equal to the average of the stochastic variable and the second cumulant is equal to the variance (3.8). The higher orders of cumulants contain less information about $p(\eta,t)$ than lower ones. In fact it can be shown, that in the frequently arising case of probabilities described by Gaussian distributions (the corresponding random processes are called Gaussian) all, but the first and second-order cumulants vanish. For non-Gaussian distributions, though, all cumulants are non-zero as stated in the theorem of Marcienkiewicz [5]). Nevertheless, cumulants give a more succinct description of $p(\eta,t)$ than moments do, dramatically so in case of Gaussian processes. This is not the only benefit as we will see considering scenarios with more than one random variable $\eta(t)$.

Cumulants are particularly useful if one has to consider the sum of statistically independent random values, for example the sum $\sigma = \eta_0 + \eta_1$. In this case one can easily prove that the characteristic function for the two variables η_0 and η_1 is the product of the two characteristic functions for each variable (see Section VI). So its logarithm is the sum of the logarithms of the two characteristic functions. Therefore a Taylor expansion shows that $\langle\langle \eta_0^{n_0} \eta_1^{n_1} \rangle\rangle = 0$ for

all $n_0, n_1 \geq 1$. Thus $\langle\langle(\eta_0+\eta_1)^n\rangle\rangle = \langle\langle\eta_0^n\rangle\rangle + \langle\langle\eta_1^n\rangle\rangle$. This result implies that cumulants of any order are simply added if one accumulates the corresponding statistically independent random variables, hence the name "cumulant". For an arbitrary number of statistically independent random variables η_j or even continuously many $\eta(t)$ one can write

$$\left\langle\left\langle\left(\sum_j \eta_j\right)^n\right\rangle\right\rangle = \sum_j \langle\langle\eta_j^n\rangle\rangle\,, \tag{13.13}$$

$$\left\langle\left\langle\left(\int dt\,\eta(t)\right)^n\right\rangle\right\rangle = \int dt\,\langle\langle\eta^n(t)\rangle\rangle\,, \tag{13.14}$$

properties which will be utilized below.

We will now furnish concrete, analytic expressions for the probabilities characterizing three important Markov processes. We begin with the so-called Wiener process.

D. Wiener Process

The *Wiener process*, described by $\omega(t)$ for $t > 0$, is characterized by the probability distributions

$$p(\omega_0, t_0) = \frac{1}{\sqrt{4\pi D t_0}} e^{-\omega_0^2/4Dt_0}, \tag{13.15}$$

$$p(\omega_1, t_1|\omega_0, t_0) = \frac{1}{\sqrt{4\pi D\Delta t}} e^{-(\Delta\omega)^2/4D\Delta t}, \tag{13.16}$$

$$\Delta\omega = \omega_1 - \omega_0, \tag{13.17}$$

$$\Delta t = t_1 - t_0. \tag{13.18}$$

The probabilities are parameterized through the constant D, referred to as the diffusion constant, since the probability distributions $p(\omega_0, t_0)$ and $p(\omega_1, t_1|\omega_0, t_0)$ are solutions of the diffusion equation (16.13) discussed extensively below. The Wiener process is homogeneous in time and space, which implies that the conditional transition probability $p(\omega_1, t_1|\omega_0, t_0)$ depends only on the relative variables $\Delta\omega$ and Δt. Put differently, the probability $p(\Delta\omega, \Delta t)$ for an increment $\Delta\omega$ to occur is independent of the current state of the Wiener process $\omega(t)$. The probability $p(\Delta\omega, \Delta t) = p(\omega_0 + \Delta\omega, t_0 + \Delta t|\omega_0, t_0)$ is given by Eq. (13.16).

Characteristic Functions, Moments, Correlation Functions and Cumulants for the Wiener process

In case of the Wiener process simple expressions can be provided for the characteristic function, moments and cumulants. Combining (13.16) and (13.5) one obtains for the characteristic function

$$G(s,t) = e^{-Dts^2}. \tag{13.19}$$

A Taylor expansion allows one to identify the moments [8]

$$\langle\omega^n(t)\rangle = \begin{cases} 0 & \text{for odd } n, \\ (n-1)!!(2Dt)^{n/2} & \text{otherwise.} \end{cases} \tag{13.20}$$

The definition (13.9) and (13.19) leads to the expression for the cumulants

$$\langle\langle\omega^n(t)\rangle\rangle = \begin{cases} 2Dt & \text{for } n = 2, \\ 0 & \text{otherwise.} \end{cases} \tag{13.21}$$

For the two-dimensional characteristic functions one can derive, using (13.16) and (13.6)

$$G(s_0, t_0; s_1, t_1) = e^{-D(s_0^2 t_0 + s_1^2 t_1 + 2s_0 s_1 \min(t_0, t_1))}, \tag{13.22}$$

---

[8] The double factorial $n!!$ for positive $n \in \mathbb{N}$ denotes the product $n(n-2)(n-4)\cdots 1$ for odd n and $n(n-2)(n-4)\cdots 2$ for even n.

From this follow the correlation functions

$$\langle\omega^{n_1}(t_1)\omega^{n_0}(t_0)\rangle = \begin{cases} 0 & \text{for odd } n_0+n_1, \\ 2D\min(t_0,t_1) & \text{for } n_0=n_1=1, \\ 12D^2t_0\min(t_0,t_1) & \text{for } n_0=1, n_1=3, \\ 4D^2[t_0t_1+2\min^2(t_0,t_1)] & \text{for } n_0=2, n_1=2, \\ \dots & \end{cases} \tag{13.23}$$

and, using the definition (13.9), the cumulants

$$\langle\!\langle\omega^{n_1}(t_1)\omega^{n_0}(t_0)\rangle\!\rangle = \begin{cases} 2D\min(t_0,t_1) & \text{for } n_0=n_1=1, \\ 0 & \text{otherwise for } n_0,n_1\neq 0. \end{cases} \tag{13.24}$$

The Wiener Process as the Continuum Limit of a Random Walk on a Lattice

The Wiener process is closely related to a random walk on a one-dimensional lattice with lattice constant a. A n-step walk on a lattice is performed in discrete times steps $t_j = j\tau$, with $j = 0, 1, 2, \dots, n$. The walk may start at an arbitrary lattice site x_0. One can choose this starting position as the origin of the coordinate system so that one can set $x_0 = 0$. The lattice sites are then located at $x_i = ia, i \in \mathbb{Z}$. At each time step the random walker moves with equal probability to the neighboring right or left lattice site. Thus, after the first step with $t = \tau$ one will find the random walker at $x = \pm a$, i.e. at site $x_{\pm 1}$ with probability $P(\pm a, \tau) = 1/2$. For a two-step walk the following paths are possible:

- path 1: two steps to the left,
- path 2: one step to the left and then one step to the right,
- path 3: one step to the right and then one step to the left,
- path 4: two steps to the right.

Each path has a probability of 1/4, a factor 1/2 for each step. Paths 2 and 3 both terminate at lattice site x_0. The probability to find a random walker after two step at position $x_0 = 0$ is therefore $P(0, 2\tau) = 1/2$. The probabilities for lattice sites $x_{\pm 2}$ reached via path 1 and 4 respectively are simply $P(\pm 2a, 2\tau) = 1/4$.

For an n-step walk one can proceed like this summing over all possible paths that terminate at a given lattice site x_i. Such a summation yields the probability $P(ia, n\tau)$. However, to do so effectively a more elegant mathematical description is appropriate. We denote a step to the right by an operator $\mathscr{R}$, and a step to the left by an operator $\mathscr{L}$. Consequently a single step of a random walker is given by $\frac{1}{2}(\mathscr{L}+\mathscr{R})$, the factor $\frac{1}{2}$ denoting the probability for each direction. To obtain a n-step walk the above operator $\frac{1}{2}(\mathscr{L}+\mathscr{R})$ has to be iterated n times. For a two-step walk one gets $\frac{1}{4}(\mathscr{L}+\mathscr{R})\circ(\mathscr{L}+\mathscr{R})$. Expanding this expression results in $\frac{1}{4}(\mathscr{L}^2+\mathscr{L}\circ\mathscr{R}+\mathscr{R}\circ\mathscr{L}+\mathscr{R}^2)$. Since a step to the right and then to the left amounts to the same as a step first to the left and then to the right, it is safe to assume that $\mathscr{R}$ and $\mathscr{L}$ commute. Hence one can write $\frac{1}{4}\mathscr{L}^2+\frac{1}{2}\mathscr{L}\circ\mathscr{R}+\frac{1}{4}\mathscr{R}^2$. As the operator expression $\mathscr{L}^p\circ\mathscr{R}^q$ stands for p steps to the left and q steps to the right one can deduce that $\mathscr{L}^p\circ\mathscr{R}^q$ represents the lattice site x_{q-p}. The coefficients are the corresponding probabilities $P((q-p)a, (q+p)\tau)$. The algebraic approach above proves useful, since one can utilize the well known binomial formula

$$\left[\frac{1}{2}(\mathscr{L}+\mathscr{R})\right]^n = \left(\frac{1}{2}\right)^n \sum_{k=0}^{n}\binom{n}{k}\mathscr{L}^k\circ\mathscr{R}^{n-k}, \tag{13.25}$$

and obtains as coefficients of x_i the probabilities

$$P(ia, n\tau) = \frac{1}{2^n}\binom{n}{\frac{n+i}{2}}, \tag{13.26}$$

which can be expressed as

$$P(x,t) = \frac{1}{2^{t/\tau}}\binom{t/\tau}{\frac{t}{2\tau}+\frac{x}{2a}}. \tag{13.27}$$

The moments of the discrete probability distribution $P(x,t)$ are

$$\langle x^n(t)\rangle = \sum_{x=-\infty}^{\infty} x^n P(x,t) = \begin{cases} 0 & \text{for odd } n, \\ a^2\frac{t}{\tau} & \text{for } n=2, \\ a^4\frac{t}{\tau}\left(3\frac{t}{\tau}-2\right) & \text{for } n=4, \\ a^6\frac{t}{\tau}\left(15\left(\frac{t}{\tau}\right)^2-30\frac{t}{\tau}+16\right) & \text{for } n=6, \\ \dots & . \end{cases} \tag{13.28}$$

We now want to demonstrate that in the continuum limit a random walk reproduces a Wiener process. For this purpose we show that the unconditional probability distributions of both processes match. We do not consider conditional probabilities $p(x_1,t_1|x_0,t_0)$ as they equal unconditional probabilities $p(x_1-x_0,t_1-t_0)$ in both cases; in a Wiener process as well as in a random walk. To turn the discrete probability distribution (13.27) into a continuous probability density distribution one considers adjacent bins centered on every lattice site that may be occupied by a random walker. Note, that only every second lattice site can be reached after a particular number of steps. Thus, these adjacent bins have a base length of $2a$ by which we have to divide $P(x,t)$ to obtain the probability density distribution $p(x,t)$ in these bins

$$p(x,t)\,dx = \frac{1}{2a}\frac{1}{2^{t/\tau}}\binom{t/\tau}{\frac{t}{2\tau}+\frac{x}{2a}}\,dx. \tag{13.29}$$

We then rescale the lattice constant a and the length τ of the time intervals to obtain a continuous description in time and space. However, τ and a need to be rescaled differently, since the spatial extension of the probability distribution $p(x,t)$, characterized by its standard deviation

$$\sqrt{\langle\langle x^2(t)\rangle\rangle} = a\sqrt{\frac{t}{\tau}} \tag{13.30}$$

is not proportional to t, but to $\sqrt{t}$. This is a profound fact and a common feature for all processes accumulating uncorrelated values of random variables in time. Thus, to conserve the temporal-spatial proportions of the Wiener process one rescales the time step τ by a factor ϵ and the lattice constant a by a factor $\sqrt{\epsilon}$

$$\tau \to \epsilon\tau \quad \text{and} \quad a \to \sqrt{\epsilon}a. \tag{13.31}$$

A continuous description of the binomial density distribution (13.29) is then approached by taking the limit $\epsilon \to 0$. In this limit the number of steps $n = \frac{t}{\epsilon\tau}$ in the random walk goes to infinity and one observes the following identity derived in Appendix A

$$\begin{aligned} p(x,t)\,dx &= \frac{1}{2\epsilon a}2^{-\frac{t}{\epsilon\tau}}\binom{\frac{t}{\epsilon\tau}}{\frac{t}{2\epsilon\tau}+\frac{x}{2\sqrt{\epsilon}a}}\,dx \\ &= \sqrt{\frac{n\tau}{4a^2t}}2^{-n}\binom{n}{\frac{n}{2}+\frac{x}{a}\sqrt{\frac{n\tau}{4t}}}\,dx \\ &= \sqrt{\frac{\tau}{2\pi a^2 t}}e^{-\frac{x^2\tau}{2a^2t}}\,dx\left(1+O\left(\frac{1}{n}\right)\right). \end{aligned} \tag{13.32}$$

The fraction τ/a^2 is invariant under rescaling (13.31) and, hence, this quantity remains in the continuous description (13.32) of the probability density distribution $p(x,t)$. Comparing equations (13.32) and (13.16) one identifies $D = a^2/2\tau$. The relation between random step length a and time unit τ obviously determines the rate of diffusion embodied in the diffusion constant D: the larger the steps a and the more rapidly these are performed, i.e., the smaller τ, the quicker the diffusion process and the faster the broadening of the probability density distribution $p(x,t)$. According to (13.30) this broadening is then $\sqrt{2Dt}$ as expected for a diffusion process.

A Wiener Process can be Integrated, but not Differentiated

We want to demonstrate that the path of a Wiener process cannot be differentiated. For this purpose we consider the differential defined through the limit

$$\frac{d\omega(t)}{dt} \equiv \lim_{\Delta t\to 0}\frac{\omega(t+\Delta t)-\omega(t)}{\Delta t} = \lim_{\Delta t\to 0}\frac{\Delta\omega(t)}{\Delta t}. \tag{13.33}$$

What is the probability for the above limit to render a finite absolute value for the derivative smaller or equal an arbitrary constant v? For this to be the case $|\Delta\omega(t)|$ has to be smaller or equal $v\Delta t$. The probability for that is

$$\int_{-v\Delta t}^{v\Delta t} d(\Delta\omega) p(\Delta\omega, \Delta t) = \frac{1}{\sqrt{4\pi D\Delta t}} \int_{-v\Delta t}^{v\Delta t} d(\Delta\omega) e^{(\Delta\omega)^2/4D\Delta t} = \mathrm{erf}\left[\sqrt{\frac{\Delta t}{D}\frac{v}{2}}\right]. \tag{13.34}$$

The above expression vanishes for $\Delta t \to 0$. Hence, taking the differential as proposed in Eq. (13.33) we would almost never obtain a finite value for the derivative. This implies that the velocity corresponding to a Wiener process is almost always plus or minus infinity.

Nonetheless one can study the behavior of $\dot{x}(t) = dx(t)/dt$ for the random walk on the lattice with the pdf of Eq. (13.27). And in this case one can show that the integral equation

$$x(t) = \int^t ds\, \dot{x}(s), \tag{13.35}$$

is well defined in the $\epsilon \to 0$ limit.

Two questions come to mind. First, do stochastic equations like $\dot{x}(t) = dx(t)/dt$ make any sense? Second, is $\dot{x}(t)$ unique or are there other stochastic processes that sum up to $x(t)$?

The first question is quickly answered. Stochastic differential equations are only well defined by the integral equations they imply. Even then the integrals in these integral equations have to be handled carefully as will be shown below.

We answer the second question by simply introducing another stochastic processes, the Ornstein-Uhlenbeck process, the integral over which also yields the Wiener process. Nevertheless, all processes that yield the Wiener process by integration over time do exhibit certain common properties that are used to define one encompassing, idealized Markov processes, the so-called Gaussian white noise. This process may be viewed as the time-derivative of the Wiener process. Gaussian white noise will be our third example of a stochastic process.

E. Ornstein-Uhlenbeck Process

Our second example for a Markov process is the *Ornstein-Uhlenbeck process*. The Ornstein-Uhlenbeck process, describing a random variable $v(t)$, is defined through the probabilities

$$p(v_0, t_0) = \frac{1}{\sqrt{\pi\gamma\sigma^2}} e^{-v_0^2/\gamma\sigma^2}, \tag{13.36}$$

$$p(v_1, t_1 | v_0, t_0) = \frac{1}{\sqrt{\pi S}} e^{-\left(v_1 - v_0 e^{-\gamma\Delta t}\right)^2/S}, \tag{13.37}$$

$$S = \gamma\sigma^2\left(1 - e^{-2\gamma\Delta t}\right), \tag{13.38}$$

$$\Delta t = |t_1 - t_0|. \tag{13.39}$$

The probabilities are characterized through two parameters σ and γ. Their significance will be explained further below. The process is homogeneous in time, since (13.37) depends solely on t, but is not homogeneous in v. Furthermore, the Ornstein-Uhlenbeck Process is stationary, i.e., $p(v_0, t_0)$ does not change in time.

Characteristic Functions, Moments, Correlation Functions and Cumulants for the Ornstein-Uhlenbeck process

The characteristic function, associated with the unconditional probability distribution $p(v_0, t_0)$ in (13.36) is also independent of time and given by

$$G(s) = e^{-\gamma(\sigma s/2)^2}. \tag{13.40}$$

The associated moments and cumulants are

$$\langle v^n(t)\rangle = \begin{cases} 0 & \text{for odd } n, \\ (n-1)!!(\gamma\sigma^2/2)^{n/2} & \text{otherwise.} \end{cases} \tag{13.41}$$

and

$$\langle\langle v^n(t)\rangle\rangle = \begin{cases} \gamma\sigma^2/2 & \text{for } n = 2, \\ 0 & \text{otherwise.} \end{cases} \tag{13.42}$$

The characteristic function for the conditional probability (13.37) is

$$G(s_0,t_0;s_1,t_1)=e^{-\gamma\sigma^2\left(s_0^2+s_1^2+2s_0s_1e^{-\gamma|t_1-t_0|}\right)/4}, \tag{13.43}$$

The corresponding correlation functions, defined according to (13.8) are

$$\langle v^{n_1}(t_1)v^{n_0}(t_0)\rangle=\begin{cases}0 & \text{for odd } n_0+n_1,\\ \frac{1}{2}\gamma\sigma^2e^{-\gamma|t_1-t_0|} & \text{for } n_0=n_1=1,\\ \frac{3}{4}\gamma^2\sigma^4e^{-\gamma|t_1-t_0|} & \text{for } n_0=1,n_1=3,\\ \frac{1}{4}\gamma^2\sigma^4\left(1+2e^{-\gamma|t_1-t_0|}\right) & \text{for } n_0=2,n_1=2,\\ \dots & .\end{cases} \tag{13.44}$$

This implies that the correlation of $v(t_1)$ and $v(t_0)$ decays exponentially. As for the Wiener process, the most compact description of the unconditional probability is given by the cumulants

$$\langle\!\langle v^{n_1}(t_1)v^{n_0}(t_0)\rangle\!\rangle=\begin{cases}\frac{1}{2}\gamma\sigma^2e^{-\gamma|t_1-t_0|} & \text{for } n_0=n_1=1,\\ 0 & \text{otherwise for } n_0,n_1\neq 0.\end{cases} \tag{13.45}$$

The Wiener process as the integral of the Ornstein-Uhlenbeck process

We want to demonstrate now that integration of the Ornstein-Uhlenbeck process $v(t)$ yields the Wiener process under a particular limiting procedure.

Let us consider the formal relationship

$$\tilde{\omega}(t)=\int_0^t ds\, v(s), \tag{13.46}$$

where $\tilde{\omega}(t)$ is another random process. Let us relate the cumulants (13.9) of the $\tilde{\omega}$ process and the ones (13.42),(13.45) for the Ornstein-Uhlenbeck process, v, according to

$$\langle\!\langle\tilde{\omega}(t)\tilde{\omega}(t')\rangle\!\rangle=\left\langle\!\left\langle\int_0^t ds\, v(s)\int_0^{t'} ds'\, v(s')\right\rangle\!\right\rangle=\int_9^t ds\int_0^{t'} ds'\,\langle\!\langle v(s)v(s')\rangle\!\rangle\,, \tag{13.47}$$

assuming $t\geq t'$. By means of (13.45) follows

$$\begin{aligned}\langle\!\langle\tilde{\omega}(t)\tilde{\omega}(t')\rangle\!\rangle&=\frac{1}{2}\gamma\sigma^2\left[\int_0^{t'} ds'\int_0^{s'} ds\, e^{-\gamma(s'-s)}+\int_0^{t'} ds\int_0^{s} ds'\, e^{-\gamma(s-s')}+\int_{t'}^{t} ds\int_0^{t'} ds'\, e^{-\gamma(s-s')}\right]\\ &=\sigma^2t'+\frac{\sigma^2}{2\gamma}\left[-1+e^{-\gamma t'}+e^{-\gamma t}-e^{-\gamma(t-t')}\right].\end{aligned} \tag{13.48}$$

For times long compared to the time scale of velocity relaxation γ^{-1} one reproduces Eq. (13.24) (we don't treat explicitly the case $t\leq t'$) where $D=\sigma^2/2$. The relationship between the Ornstein-Uhlenbeck and Wiener processes defined through (13.46) holds for all cumulants, not just for the cumulants of second order. We only had to prove the relation (13.48), since all other cumulants of both processes are simply 0. This allows one to state that

$$\omega(t)=\lim_{\gamma\to\infty}\int_0^t dd\, v(s). \tag{13.49}$$

With respect to their probability distributions the Ornstein-Uhlenbeck process $v(t)$ and the velocity of a random walker $\dot{x}(t)$ are different stochastic processes. However, in the limit $\gamma\to\infty$ and $\epsilon\to 0$ the following moment and correlation function turn out to be the same for both processes, if $D=\sigma^2/4=a^2/2\tau$

$$\langle v(t)\rangle=\langle\dot{x}(t)\rangle=0, \tag{13.50}$$

$$\lim_{\gamma\to\infty}\langle v(t_1)v(t_0)\rangle=\lim_{\epsilon\to 0}\langle\dot{x}(t_1)\dot{x}(t_0)\rangle=2D\delta(t_1-t_0). \tag{13.51}$$

Hence, one uses these properties to define an idealized Markov process, the so-called *Gaussian white noise*.

F. White Noise Process

An important idealized stochastic process is the so-called "*Gaussian white noise*". This process, denoted by $\xi(t)$, is not characterized through conditional and unconditional probabilities, but through the following statistical moment and correlation function

$$\langle \xi(t) \rangle = 0, \tag{13.52}$$
$$\langle \xi(t_1)\xi(t_0) \rangle = \zeta^2 \delta(t_1 - t_0). \tag{13.53}$$

The attribute *Gaussian* implies that all cumulants higher than of second order are 0

$$\langle\langle \xi^{n_1}(t_1)\xi^{n_0}(t_0) \rangle\rangle = \begin{cases} \zeta^2\delta(t_1 - t_0) & \text{for } n_0, n_1 = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{13.54}$$

The reason why this process is termed *white* is connected with its correlation function (13.53), the Fourier transform of which is constant, i.e., entails all frequencies with equal amplitude just as white radiation. The importance of the process $\xi(t)$ stems from the fact that many other stochastic processes are described through stochastic differential equations with a (white) noise term $\xi(t)$. We will show this for the Wiener process below and for the Ornstein-Uhlenbeck processes later in the notes.

As hinted by the examples in this section we can show that the integral of Gaussian white noise happens to be a Wiener process. We prove this in the same fashion as above by deriving the cumulants for $\int dt\ \xi(t)$. Again the task is simplified by the fact that only one cumulant is non-zero, namely,

$$\begin{aligned} \left\langle\left\langle \int_0^{t_1} ds_1\, \xi(s_1) \int_0^{t_0} ds_0\, \xi(s_0) \right\rangle\right\rangle &= \int_0^{t_1} ds_1 \int_0^{t_0} ds_0\, \langle\langle \xi(t_1)\xi(t_0) \rangle\rangle \\ &= \int_0^{t_1} ds_1 \int_0^{t_0} ds_0\, \zeta^2\delta(t_1 - t_0) \\ &= \zeta^2 \min(t_1, t_0). \end{aligned} \tag{13.55}$$

We demonstrated, thereby, the important relationship between white noise $\xi(t)$ and the Wiener process $\omega(t)$

$$\omega(t) = \int_0^t ds\, \xi(s), \tag{13.56}$$

for $2D = \zeta^2 = 1$.

XIV. STOCHASTIC DIFFERENTIAL EQUATIONS

In this section we assume that the constituents of matter can be described classically. We are interested in reaction processes occurring in the bulk, either in physiological liquids, membranes or proteins. The atomic motion of these materials is described by the *Newtonian* equation of motion

$$m_i \frac{d^2}{dt^2} r_i = -\frac{\partial}{\partial r_i} V(r_1, \ldots, r_N), \tag{14.1}$$

where r_i $(i = 1, 2, \ldots, N)$ describes the position of the i-th atom. The number N of atoms is, of course, so large that solutions of Eq. (14.1) for macroscopic systems are impossible. In microscopic systems like proteins the number of atoms ranges between 10^3 to 10^5, i.e., even in this case the solution is extremely time consuming.

However, most often only a few of the degrees of freedom are involved in a particular biochemical reaction and warrant an explicit theoretical description or observation. For example, in the case of transport one is solely interested in the position of the center of mass of a molecule. It is well known that molecular transport in condensed media can be described by phenomenological equations much simpler than Eq. (14.1), e.g., by the Einstein diffusion equation. The same holds true for reaction processes in condensed media. In this case one likes to focus onto the reaction coordinate, e.g., on a torsional angle.

A. Langevin Equation

In fact, there exist successful descriptions of a small subset of degrees of freedom by means of Newtonian equations of motion with effective force fields and added frictional as well as (time dependent) fluctuating forces. Let us assume we like to consider motion along a small subset of the whole coordinate space defined by the coordinates $q_1, q_2, \ldots, q_M$ for $M \ll N$. The equations which model the dynamics in this subspace are then of the following *Langevin* type

$$\mu_j \frac{d^2}{dt^2} q_j = -\frac{\partial}{\partial q_j} W(q_1, \ldots, q_M) - \gamma_j \frac{d}{dt} q_j + \sigma_j \xi_j(t) \quad j = 1, 2, \ldots, M. \tag{14.2}$$

The first term on the right hand side of this equation describes the force

field derived from an effective potential $W(q_1, \ldots, q_M)$, the second term describes the velocity ($\dot{q}_j$) dependent frictional forces, and the third term the fluctuating forces $\xi_j(t)$ with coupling constants σ_j. $W(q_1, \ldots, q_M)$ includes the effect of the thermal motion of the remaining $n-$M degrees of freedom on the motion along the coordinates $q_1, q_2, \ldots, q_M$.

Equations of type (14.2) will be studied in detail further below. We will not "derive" these equations from the Newtonian equations (14.1) of the bulk material, but rather show by comparison of the predictions of Eq. (14.1) and Eq. (14.2) to what extent the suggested phenomenological descriptions apply. To do so and also to study further the consequences of Eq. (14.2) we need to investigate systematically the solutions of stochastic differential equations.

B. First order stochastic differential equations

Let $\boldsymbol{x}$, $\boldsymbol{A}$, and $\boldsymbol{\eta}$ be 2 dimensional column vectors and $\boldsymbol{B}$ a 2×2 matrix. We consider stochastic differential equations in the form of a first order differential equation

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{A}[\boldsymbol{x}(t), t] + \boldsymbol{B}[\boldsymbol{x}(t), t] \cdot \boldsymbol{\eta}(t), \tag{14.3}$$

subject to the initial condition

$$\boldsymbol{x}(0) = \boldsymbol{x}_0. \tag{14.4}$$

In this equation $\boldsymbol{A}[\boldsymbol{x}(t), t]$ represents the so-called drift term and $\boldsymbol{B}[\boldsymbol{x}(t), t] \cdot \boldsymbol{\eta}(t)$ the noise term which will be properly characterized further below. Without the noise term, the resulting equation

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{A}[\boldsymbol{x}(t), t], \tag{14.5}$$

describes a deterministic drift of particles along the field $\boldsymbol{A}[\boldsymbol{x}(t), t]$.

Equations like (14.5) can actually describe a wide variety of phenomena, like chemical kinetics or the

ring of neurons. Since such systems are often subject to random perturbations, noise is added to the deterministic equations to yield associated stochastic differential equations. In such cases as well as in the case of classical Brownian particles, the noise term $\boldsymbol{B}[\boldsymbol{x}(t), t] \cdot \boldsymbol{\eta}(t)$ needs to be specified on the basis of the underlying origins of noise. We already introduced above several mathematical models of noise and will consider the issue of constructing suitable noise terms throughout this notes. For this purpose, one often adopts a heuristic approach, analyzing the noise from observation or from a numerical simulation and selecting a noise model with matching characteristics. These characteristics are introduced below.

Before we consider characteristics of the noise term $\boldsymbol{\eta}(t)$ in (14.3) we like to demonstrate that the one-dimensional Langevin equation (14.2) of a classical particle, written here in the form

$$\mu \ddot{q} = f(q) - \gamma \dot{q} + \sigma \xi(t) \tag{14.6}$$

is a special case of (14.3). In fact, defining $\boldsymbol{x} \in \mathbb{R}^2$ with components $x_1 = \mu \dot{q}$, $x_2 = \mu q$ reproduces Eq. (14.6) if one defines

$$\boldsymbol{A}[\boldsymbol{x}(t), t] = \begin{pmatrix} f(x_2/\mu) - \gamma x_1/\mu \\ x_1 \end{pmatrix}, \quad \boldsymbol{B}[\boldsymbol{x}(t), t] = \begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix}, \quad \boldsymbol{\eta}(t) = \begin{pmatrix} \xi(t) \\ 0 \end{pmatrix}. \tag{14.7}$$

The noise term represents a stochastic process. We consider only the factor $\boldsymbol{\eta}(t)$ which describes the essential time dependence of the noise source in the different degrees of freedom. The matrix $\boldsymbol{B}[\boldsymbol{x}(t), t]$ describes the amplitude and the correlation of noise between the different degrees of freedom.

XV. FOKKER-PLANCK EQUATION

Aim of this section is to determine the (deterministic) partial differential equation satisfied by the conditional probability distribution for the stochastic process $q(t)$ described by the Langevin equation. This is called the *Fokker-Planck equation* [13].

A. Ito calculus

We return to the stochastic differential equation (14.3), which we will now express as an integral equation. We will model the noise term $\boldsymbol{\eta}(t)$ by a tupel of normalized Gaussian white noise processes $\boldsymbol{\xi}(t)$ with $\zeta = 1$

$$\boldsymbol{x}(t) = \int_0^t ds\, \boldsymbol{A}[\boldsymbol{x}(s), s] + \int_0^t ds\, \boldsymbol{B}[\boldsymbol{x}(s), s] \cdot \boldsymbol{\eta}(s), \tag{15.1}$$

Since $\boldsymbol{x}(t)$ is continuous, the first integral on the right hand side is well defined, e.g., in the sense of a Riemann integral. However, the second integral poses problems. Let us consider the simple one-dimensional case with an arbitrary function or stochastic process $G[t]$

$$I(t) = \int_0^t ds\, \xi(s) G[s]. \tag{15.2}$$

One can rewrite this integral in terms of a normalized Wiener process $\omega(t)$ with $D = 1/2$. One subsitutes $d\omega(s)$ for $ds\, \xi(s)$ thanks to Eq. (13.56), to find

$$I(t) = \int_0^t d\omega(s)\, G[s]. \tag{15.3}$$

The kind of *Riemann-Stieltjes* integral (15.3) can be approximated by the sums $I_n^{(\alpha)}(t)$ which evaluate the integral at n discrete times steps $t_j = j\, t/n$ as follows

$$I^{(\alpha)} = \underset{n\to\infty}{\text{ms-lim}}\, I_n^{(\alpha)}(t), \tag{15.4}$$

$$I_n^{(\alpha)}(t) = \sum_{j=1}^{n} G[(1-\alpha)t_{j-1} + \alpha t_j][\omega(t_j) - \omega(t_{j-1})]. \tag{15.5}$$

Two remarks about Eqs. (15.4) and (15.5) are due. First, one has to specify the meaning of approximation, since one is dealing with random variables. To approximate the integral $I^{(\alpha)}$ one takes the so-called *mean square limit*. Such a limit has to satisfy the following condition of convergence

$$X = \underset{n\to\infty}{\text{ms-lim}} X_n \quad \text{if} \quad \lim_{n\to\infty} \left\langle (X_n - X)^2 \right\rangle = 0 \tag{15.6}$$

Second, note that the sum $I_n^{(\alpha)}(t)$ is parameterized by α. This allows one to choose the position where to evaluated $G[t]$ within the time step intervals $[t_{j-1}, t_j]$. In the limit $n \to \infty$ as the intervals $[t_{j-1}, t_j]$ become infinitely small, non-stochastic integrals become independent of α, not so $I_n^{(\alpha)}(t)$!

The dependence on α can be made plausible. Consider the case $G(t) = \omega(t)$. The mean squared difference of the Wiener process $\omega(t)$ at the left and the right side of the interval $[t_{j-1}, t_j]$ is given by the standard deviation of the difference $\Delta\omega(t_j) = \omega(t_j) - \omega(t_{j-1})$

$$\sqrt{\langle [\omega(t_j) - \omega(t_{j-1})]^2 \rangle} = \sqrt{t_j - t_{j-1}}, \tag{15.7}$$

where $t_j - t_{j-1} = \Delta t_j$. The difference in summing over all values on the left side of the intervals as opposed to the right side is then on average given by

$$\sum_{j=1}^{n} \underbrace{\sqrt{t_j - t_{j-1}}}_{O\left(\sqrt{1/n}\right)} \underbrace{[\omega(t_j) - \omega(t_{j-1})]}_{O\left(\sqrt{1/n}\right)}. \tag{15.8}$$

If we sum over all terms we obtain n-times an expression of order $O(1/n)$ and consequently a finite value; a finite difference!

If we compare this observation with ordinary calculus we see the essential discrepancy. Consider the integral $\int_0^t ds\, f(s)$. The difference between evaluating the left and right side of interval $[t_{j-1}, t_j]$ is given by $f'(t_{j-1})(t_j - t_{j-1})$. Again the difference of summing over all values $f(t)$ on the left side of the intervals as opposed to the right side is

$$\sum_{j=1}^{n} \underbrace{f'(t_{j-1})(t_j - t_{j-1})}_{O(1/n)} \underbrace{(t_j - t_{j-1})}_{O(1/n)}. \tag{15.9}$$

This sum is of order $O(1/n)$ and approaches 0 for $n \to \infty$. It is consequently irrelevant which side of the interval $[t_{j-1}, t_j]$ we choose to evaluate $f(t)$. The underlying cause for the α-dependence of stochastic integrals is evident. It is the $\sqrt{1/n}$-scaling property of stochastic processes already discussed in Section VIII C! The α- dependence is here to stay.

B. Ito rules

We have seen that it is not admissible to neglect the α-dependence. Nevertheless it is possible to develop a consistent calculus by assuming a fixed value for parameter α. There are two popular approaches, each with distinct benefits and disadvantages

$$\alpha = 0 \quad \textit{Ito calculus}, \tag{15.10}$$

$$\alpha = \frac{1}{2} \quad \textit{Stratonovich calculus}. \tag{15.11}$$

The Stratonovich calculus with $\alpha = 1/2$ exhibits the same integration rules as ordinary calculus. It also models processes with finite correlation time correctly. However, the rules for the Ito calculus are easier to derive. In many instances corresponding derivations are impossible in the Stratonovich case. Hence we begin with an introduction to Ito calculus. Later, in Section XV D we will compare the Ito and Stratonovich approaches. In any case, we have to keep in mind, that the choice of $\alpha = 0$ or $\alpha = 1/2$ is not arbitrary and has to be justified when modeling physical processes with stochastic differential and corresponding integral equations. For now we set $\alpha = 0$.

The foundation of Ito calculus are the four *Ito rules*

$$d\omega_i(t) d\omega_j(t) = \delta_{ij} dt, \tag{15.12}$$

$$[d\omega(t)]^N = 0 \quad \text{for } N > 2, \tag{15.13}$$

$$[d\omega(t)]^N dt = 0 \quad \text{for } N \geq 1, \tag{15.14}$$

$$[dt]^N = 0 \quad \text{for } N > 1. \tag{15.15}$$

As with distributions, like the Dirac delta function $\delta(x)$, these rules (15.12)-(15.15) have to be seen in the context of integration. Furthermore the integration has to be over so-called non-anticipating functions or processes $G[t]$. This will become clear as we prove rule (15.12) for the one-dimensional case.

Rules (15.12) and (15.13) have to be read as

$$\int_0^t [d\omega(s)]^N G[s] = \underset{n\to\infty}{\text{ms-lim}} \sum_{j=1}^{n} G[t_{j-1}][\Delta\omega(t_j)]^N = \begin{cases} \int_0^t ds\, G[s] & \text{for } N = 2, \\ 0 & \text{for } N > 2. \end{cases} \tag{15.16}$$

for a non-anticipating function $G[t]$ that is statistically independent of $[\omega(s) - \omega(t)]$ for any $s > t$

$$\langle G[t][\omega(s) - \omega(t)] \rangle = 0 \quad \text{for } t < s. \tag{15.17}$$

To prove rule (15.12) we have to show that the following mean square limit vanishes

$$\begin{aligned} I &= \left\langle \left[\int_0^t [d\omega(s)]^2 G[s] - \int_0^t ds\, G[s] \right]^2 \right\rangle \\ &= \underset{n\to\infty}{\text{ms-lim}} \left\langle \left[\sum_{j=1}^{n} [\Delta\omega^2(t_j) - \Delta t_j] G[t_{j-1}] \right]^2 \right\rangle \\ &= \underset{n\to\infty}{\text{ms-lim}} \left\langle \sum_{j=1}^{n} \underbrace{G^2[t_{j-1}]}\, \underbrace{[\Delta\omega^2(t_j) - \Delta t_j]^2} \right\rangle \\ &\quad + \underset{n\to\infty}{\text{ms-lim}} \left\langle \sum_{i\neq j=1}^{n} \underbrace{G[t_{i-1}] G[t_{j-1}] [\Delta\omega^2(t_j) - \Delta t_j]}\, \underbrace{[\Delta\omega^2(t_i) - \Delta t_i]} \right\rangle. \end{aligned} \tag{15.18}$$

Each of the above underbraced terms is statistically independent of the other underbraced factor. Here the non-anticipation property (15.17) of $G[t]$ comes into play! We obtain

$$\begin{aligned} I &= \underset{n\to\infty}{\text{ms-lim}} \sum_{j=1}^{n} \left\langle G^2[t_{j-1}] \right\rangle \underbrace{\left\langle [\Delta\omega^2(t_j) - \Delta t_j]^2 \right\rangle}_{=2[\Delta t_j]^2 \text{ due to (13.20)}} \\ &\quad + \underset{n\to\infty}{\text{ms-lim}} \sum_{i\neq j=1}^{n} \left\langle G[t_{i-1}] G[t_{j-1}] [\Delta\omega^2(t_i) - \Delta t_i] \right\rangle \underbrace{\left\langle [\Delta\omega^2(t_j) - \Delta t_j] \right\rangle}_{=0 \text{ due to (13.20)}} \\ &= \underset{n\to\infty}{\text{ms-lim}}\, 2 \sum_{j=1}^{n} \left\langle G^2[t_{j-1}] \right\rangle [\Delta t_j]^2. \end{aligned} \tag{15.19}$$

As $[\Delta t_j]^2$ is of order $O(1/n^2)$ and as long as $G[s]$ is a bounded function, the above sum vanishes as $n \to \infty$. Thus, we have proven Ito first rule (15.12). All the other rules are shown in a similar manner.

C. Ito formula

Combining the one dimensional version of the stochastic differential equation (14.3) and Ito rules we can derive another important equation, the so-called *Ito formula*. Let $f[x(t)]$ be an arbitrary function of a process $x(t)$ that satisfies the one-dimensional stochastic differential equation

$$dx(t) = (a[x(t), t] + b[x(t), t]\xi(t))\, dt = a[x(t), t]\, dt + b[x(t), t]\, d\omega(t). \tag{15.20}$$

To determine the change of $f[x(t)]$ with respect to dx and dt we perform a Taylor expansion

$$df[x(t)] = f[x(t) + dx(t)] - f[x(t)] = f'[x(t)]\, dx(t) + \frac{1}{2} f''[x(t)]\, dx^2(t) + O(dx^3(t)). \tag{15.21}$$

substituting equation (15.20) for dx we can write

$$df[x(t)] = f'[x(t)]a[x(t), t]\, dt + f'[x(t)]b[x(t), t]\, d\omega(t) + \frac{1}{2} f''[x(t)]\, (b[x(t), t]\, d\omega(t))^2 + \dots \tag{15.22}$$

We can neglect higher orders of $d\omega(t)$ and dt due to Ito rules (15.13)-(15.15). We can also substitute $d\omega^2(t)$ by dt due to (15.12) and obtain

$$df[x(t)] = f'[x(t)]a[x(t), t]\, dt + f'[x(t)]b[x(t), t]\, d\omega(t) + \frac{1}{2} f''[x(t)]\, (b[x(t), t])^2\, dt. \tag{15.23}$$

The resulting Ito formula, now in more than one-dimension, reads

$$df[\boldsymbol{x}(t)] = \sum_i A_i\, \partial_i f[\boldsymbol{x}(t)]\, dt + \sum_{i,j} B_{ij}\, \partial_i f[\boldsymbol{x}(t)]\, d\omega_j(t) + \frac{1}{2} \sum_{i,j,k} B_{ik} B_{jk}\, \partial_i \partial_j f[\boldsymbol{x}(t)]\, dt, \tag{15.24}$$

where $\partial_i = \partial/\partial x_i$ is a partial derivative. This formula is most helpful when one has to find a relation between the stochastic differential equation (14.3) of $\boldsymbol{x}(t)$ and a distribution function $f[\boldsymbol{x}(t)]$. We will now utilize Ito formula to derive the Fokker-Planck equation.

Again we consider the stochastic differential equation (14.3) with a noise term characterized through white noise

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{A}[\boldsymbol{x}(t), t] + \boldsymbol{B}[\boldsymbol{x}(t), t] \cdot \boldsymbol{\eta}(t), \tag{15.25}$$

assuming $\boldsymbol{\eta}(t) = \boldsymbol{\xi}(t)$ satisfying to Eqs. (13.52) and (13.53). For the sake of Ito calculus one has to assume that coefficient $\boldsymbol{B}[x(t), t]$ is a non-anticipating function. With this in mind we neglect the arguments $\boldsymbol{x}(t)$ and t of $\boldsymbol{A}$ and $\boldsymbol{B}$ for easier reading in the rest of this section.

We start from the average of Ito formula (15.24) and use the properties of white noise (13.52) and (13.53)

$$\langle df[\boldsymbol{x}(t)] \rangle = \sum_i \langle A_i\, \partial_i f[\boldsymbol{x}(t)]\, dt \rangle + \sum_{i,j} \langle B_{ij}\, \partial_i f[\boldsymbol{x}(t)]\, d\omega_j(t) \rangle + \frac{1}{2} \sum_{i,j,k} \langle B_{ik} B_{jk}\, \partial_i \partial_j f[\boldsymbol{x}(t)]\, dt \rangle\,. \tag{15.26}$$

The second sum on the right hand side vanishes, since $\boldsymbol{B}[\boldsymbol{x}(t),t]$ and $\partial_i f[\boldsymbol{x}(t)]$ are non-anticipating functions and therefore statistically independent of $d\omega_j(t)$, and because of equation (13.52) considering that $d\omega_j(t) = \xi_j(t)\,dt$

$$\langle B_{ij}\,\partial_i f[\boldsymbol{x}(t)]\,d\omega_j(t)\rangle = \langle B_{ij}\,\partial_i f[\boldsymbol{x}(t)]\rangle \underbrace{\langle \xi_j(t)\rangle}_{=0}\,dt = 0 \tag{15.27}$$

One is left with equation

$$\left\langle \dot{f}[\boldsymbol{x}(t)]\right\rangle = \sum_i \langle A_i\,\partial_i f[\boldsymbol{x}(t)]\rangle + \frac{1}{2}\sum_{i,j,k}\langle B_{ik}B_{jk}\,\partial_i\partial_j f[\boldsymbol{x}(t)]\rangle\,. \tag{15.28}$$

According to definition (13.3) $\langle f[x(t)]\rangle$ can be expressed as

$$\langle f[\boldsymbol{x}(t)]\rangle = \int d\boldsymbol{x}\, f[\boldsymbol{x}]p(\boldsymbol{x},t|\boldsymbol{x}_0,t_0). \tag{15.29}$$

The reader should note that the initial value of $\langle f[\boldsymbol{x}(t)]\rangle$ is $f[\boldsymbol{x}_0]$ in accordance with the initial condition assumed in Eq. (14.4). Applying the time derivative to the right hand side of (15.29) and comparing with (15.28) yields

$$\int d\boldsymbol{x} f[\boldsymbol{x}]\partial_t p(\boldsymbol{x},t|\boldsymbol{x}_0,t_0) = \int d\boldsymbol{x}\left(\sum_i A_i\,\partial_i f[\boldsymbol{x}(t)] + \frac{1}{2}\sum_{i,j}\left[\boldsymbol{B}\cdot\boldsymbol{B}^T\right]_{ij}\partial_i\partial_j f[\boldsymbol{x}(t)]\right)p(\boldsymbol{x},t|\boldsymbol{x}_0,t_0), \tag{15.30}$$

where $\partial_t = \partial/\partial t$ is a partial derivative and the superscript T denotes the transposed matrix. Partial integration assuming a volume Ω with a surface $\partial\Omega$ allows one to change the order of the partial differential operators respect to x_i. For example, the first sum becomes

$$\int_\Omega d\boldsymbol{x}\sum_i A_i\,(\partial_i f[\boldsymbol{x}])\,p(\boldsymbol{x},t|\boldsymbol{x}_0,t_0) = -\int_\Omega d\boldsymbol{x}\, f[\boldsymbol{x}]\sum_i \partial_i\left[A_i p(\boldsymbol{x},t|\boldsymbol{x}_0,t_0)\right] + \int_{\partial\Omega} d\boldsymbol{a}\cdot\boldsymbol{A}f[\boldsymbol{x}]p(\boldsymbol{x},t|\boldsymbol{x}_0,t_0),$$

where $d\boldsymbol{a}$ is the elementary surface area on $\partial\Omega$ and Gauss theorem has been used. Assuming a $p(\boldsymbol{x},t|\boldsymbol{x}_0,t_0)$ of finite spatial extent, such that it vanishes on the boundary $\partial\Omega$, we can neglect the surface term. Applying the same calculation twice to the second term in (15.30) leads to

$$\int d\boldsymbol{x}\, f[\boldsymbol{x}]\partial_t p(\boldsymbol{x},t|\boldsymbol{x}_0,t_0) = \int d\boldsymbol{x}\, f[\boldsymbol{x}]\left(-\sum_i\partial_i\left[A_i p(\boldsymbol{x},t|\boldsymbol{x}_0,t_0)\right] + \frac{1}{2}\sum_{i,j}\partial_i\partial_j\left[\left[\boldsymbol{B}\cdot\boldsymbol{B}^T\right]_{ij}p(\boldsymbol{x},t|\boldsymbol{x}_0,t_0)\right]\right).$$

Since $f[\boldsymbol{x}(t)]$ is arbitrary we can conclude

$$\partial_t p(\boldsymbol{x},t|\boldsymbol{x}_0,t_0) = -\sum_i\partial_i\left[A_i p(\boldsymbol{x},t|\boldsymbol{x}_0,t_0)\right] + \frac{1}{2}\sum_{i,j}\partial_i\partial_j\left[\left[\boldsymbol{B}\cdot\boldsymbol{B}^T\right]_{ij}p(\boldsymbol{x},t|\boldsymbol{x}_0,t_0)\right]. \tag{15.31}$$

This is the celebrated Fokker-Planck equation which describes the time evolution of the probability that the stochastic process determined by (14.3) assumes the value $\boldsymbol{x}$ at time t when it had assumed the value $\boldsymbol{x}_0$ at time t_0.

Note, that the above Fokker-Planck equation holds for the stochastic differential equation (14.3) only within the framework of Ito calculus. The relation between the stochastic differential equation (14.3) and the Fokker-Planck equation (15.31) is slightly different when Stratonovitch calculus is applied as will be shown in the next section.

D. Stratonovich calculus

We take a quick look at Stratonovich calculus mentioned in Section XV B. We want to clarify the α-dependence of our results in Sections XV B and XV C. For this purpose it is sufficient to focus on processes satisfying the stochastic differential equation (14.3). It is possible to show that a solution $x(t)$ of the stochastic differential equation

$$\text{Ito } \alpha = 0\text{:}\quad \dot{\boldsymbol{x}}(t) = \boldsymbol{A}[\boldsymbol{x}(t),t] + \boldsymbol{B}[\boldsymbol{x}(t),t]\cdot\boldsymbol{\eta}(t), \tag{15.32}$$

solves a stochastic differential equation of the same form with different coefficients, this time however according to Stratonovich calculus

$$\text{Stratonovich } \alpha = \tfrac{1}{2}\text{:}\quad \dot{\boldsymbol{x}}(t) = \overset{\mathscr{S}}{\boldsymbol{A}}[\boldsymbol{x}(t),t] + \overset{\mathscr{S}}{\boldsymbol{B}}[\boldsymbol{x}(t),t]\cdot\boldsymbol{\eta}(t). \tag{15.33}$$

We give a derivation for $\overset{\mathscr{S}}{\boldsymbol{A}}$ and $\overset{\mathscr{S}}{\boldsymbol{B}}$ in the one-dimensional case as the lower case coefficients $\overset{\mathscr{S}}{a}$ and $\overset{\mathscr{S}}{b}$ indicate. As a first step we solve the integral corresponding to equation (15.33)

$$x(t) = x(t_0) + \int_{t_0}^{t} ds\, \overset{\mathscr{S}}{a}[x(s), s] + \oint_{t_0}^{t} d\omega(s)\, \overset{\mathscr{S}}{b}[x(s), s]. \tag{15.34}$$

The $\mathscr{S}$ on the second integral sign denotes a Stratonovich integral which has to be solved like a Riemann-Stieltjes integral as in Eq. (15.5) with $\alpha = 1/2$. The last term of Eq. (15.33) is the only one that differs from Ito calculus and thus it is the only term that needs to be investigated. One can rewrite the last term as an Ito integral. We do so neglecting the mean square limit notation in the defintion of a Stratonovich integral and write

$$\begin{aligned}\oint_{t_0}^{t} d\omega(s)\, \overset{\mathscr{S}}{b}[x(s), s] &\simeq \sum_i [\omega(t_i) - \omega(t_{i-1})]\, \overset{\mathscr{S}}{b}[x(\tau), \tau] \\ &= \sum_i [\omega(t_i) - \omega(\tau)]\, \overset{\mathscr{S}}{b}[x(\tau), \tau] + \sum_i [\omega(\tau) - \omega(t_{i-1})]\, \overset{\mathscr{S}}{b}[x(\tau), \tau],\end{aligned} \tag{15.35}$$

with $\tau \equiv (t_i + t_{i-1})/2$ and $\overset{\mathscr{S}}{b}[x(\tau), \tau]$ can be approximated by extrapolation starting with $\overset{\mathscr{S}}{b}[x(t_{i-1}), t_{i-1}]$ at the left side of the interval $[t_{i-1}, t_i]$.

$$\begin{aligned}\overset{\mathscr{S}}{b}[x(\tau), \tau] = \overset{\mathscr{S}}{b}[x(t_{i-1}), t_{i-1}] &+ \left(\partial_x \overset{\mathscr{S}}{b}[x(t_{i-1}), t_{i-1}]\right)(x(\tau) - x(t_{i-1})) \\ &+ \left(\partial_t \overset{\mathscr{S}}{b}[x(t_{i-1}), t_{i-1}]\right)(\tau - t_{i-1}) \\ &+ \frac{1}{2}\left(\partial_x^2 \overset{\mathscr{S}}{b}[x(t_{i-1}), t_{i-1}]\right)(x(\tau) - x(t_{i-1}))^2 + \ldots\end{aligned} \tag{15.36}$$

Since $x(t)$ is a solution of Ito stochastic equation (15.32) one can apply (15.32) to determine the infinitesimal displacement $x(\tau) - x(t_{i-1})$.

$$x(\tau) - x(t_{i-1}) = a[x(t_{i-1}, t_{i-1})](\tau - t_{i-1}) + b[x(t_{i-1}, t_{i-1})](\omega(\tau) - \omega(t_{i-1})). \tag{15.37}$$

Filling equation (15.37) into (15.36) and applying Ito rules (15.12)-(15.15) to the infinitesimal displacements $dt = (\tau - t_{i-1})$ and $d\omega(t) = \omega(\tau) - \omega(t_{i-1})$ one obtains

$$\begin{aligned}\overset{\mathscr{S}}{b}[x(\tau), \tau] = \overset{\mathscr{S}}{b}[x(t_{i-1}), t_{i-1}] &+ a[x(t_{i-1}), t_{i-1}]\left(\partial_x \overset{\mathscr{S}}{b}[x(t_{i-1}), t_{i-1}]\right)(\tau - t_{i-1}) \\ &+ b[x(t_{i-1}), t_{i-1}]\left(\partial_x \overset{\mathscr{S}}{b}[x(t_{i-1}), t_{i-1}]\right)(\omega(\tau) - \omega(t_{i-1})) \\ &+ \left(\partial_t \overset{\mathscr{S}}{b}[x(t_{i-1}), t_{i-1}]\right)(\tau - t_{i-1}) \\ &+ \frac{1}{2} b^2[x(t_{i-1}), t_{i-1}]\left(\partial_x^2 \overset{\mathscr{S}}{b}[x(t_{i-1}), t_{i-1}]\right)(\tau - t_{i-1}) + \ldots\end{aligned} \tag{15.38}$$

Substituting the above result (15.38) into the second sum of Eq. (15.35) one finds

$$\begin{aligned}
\mathop{\mathscr{I}\!\!\!\!\!\!\int}_{t_0}^{t} d\omega(s)\, \overset{\mathscr{S}}{b}[x(s),s] \simeq & \sum_i [\omega(t_i)-\omega(\tau)]\, \overset{\mathscr{S}}{b}[x(\tau),\tau] \\
& + \sum_i [\omega(\tau)-\omega(t_{i-1})]\, \overset{\mathscr{S}}{b}[x(t_{i-1}),t_{i-1}] \\
& + \sum_i \underbrace{[\omega(\tau)-\omega(t_{i-1})]\,[\tau-t_{i-1}]}_{=0 \text{ due to (15.14)}} a[x(t_{i-1}),t_{i-1}] \left(\partial_x \overset{\mathscr{S}}{b}[x(t_{i-1}),t_{i-1}]\right) \\
& + \sum_i \underbrace{[\omega(\tau)-\omega(t_{i-1})]^2}_{=(\tau-t_{i-1}) \text{ due to (15.12)}} b[x(t_{i-1}),t_{i-1}] \left(\partial_x \overset{\mathscr{S}}{b}[x(t_{i-1}),t_{i-1}]\right) \\
& + \sum_i \underbrace{[\omega(\tau)-\omega(t_{i-1})]\,[\tau-t_{i-1}]}_{=0 \text{ due to (15.14)}} \left(\partial_t \overset{\mathscr{S}}{b}[x(t_{i-1}),t_{i-1}]\right) \\
& + \frac{1}{2}\sum_i \underbrace{[\omega(\tau)-\omega(t_{i-1})]\,[\tau-t_{i-1}]}_{=0 \text{ due to (15.14)}} b^2[x(t_{i-1}),t_{i-1}] \left(\partial_x^2 \overset{\mathscr{S}}{b}[x(t_{i-1}),t_{i-1}]\right). \qquad (15.39)
\end{aligned}$$

The first two terms on the right hand side of Eq. (15.39) make up a sum that approximates an Ito integral with time steps just half the size. In the fourth term one can replace $(\tau - t_{i-1})$ by $\frac{1}{2}(t_i - t_{i-1})$. The result, again written for the multi-dimensional case, is

$$\mathop{\mathscr{I}\!\!\!\!\!\!\int}_{t_0}^{t} d\boldsymbol{\omega}(s)\cdot \overset{\mathscr{S}}{\boldsymbol{B}}{}^{T}[x(s),s] = \int_{t_0}^{t} d\boldsymbol{\omega}(s)\cdot \overset{\mathscr{S}}{\boldsymbol{B}}{}^{T}[x(s),s] + \frac{1}{2}\int_{t_0}^{t} ds \sum_{i,j} B_{ij}[\boldsymbol{x}(s),s]\partial_i \overset{\mathscr{S}}{B}{}^{T}_{jk}[\boldsymbol{x}(s),s]. \qquad (15.40)$$

Note, that the above connection (15.40) between Ito and Stratonovich integrals only holds for $\boldsymbol{x}(t)$ satisfying Ito stochastic differential equation (15.32) or Stratonovich's stochastic differential equation (15.33). There is no general connection between Ito and Stratonovich integrals!

Substituting (15.40) into Stratonovich integral equation (15.34) and comparing the coefficients with the integral solving Ito stochastic differential equation (15.32) we obtain the following relations

$$A_k = \overset{\mathscr{S}}{A}_k + \frac{1}{2}\sum_{i,j} \overset{\mathscr{S}}{B}_{ij}\partial_i \overset{\mathscr{S}}{B}_{kj}, \qquad (15.41)$$

$$B_{jk} = \overset{\mathscr{S}}{B}_{jk}, \qquad (15.42)$$

and conversely

$$\overset{\mathscr{S}}{A}_k = A_k - \frac{1}{2}\sum_{i,j} B_{ij}\partial_i B_{kj}, \qquad (15.43)$$

$$\overset{\mathscr{S}}{B}_{jk} = B_{jk}. \qquad (15.44)$$

We see that a difference between Ito and Stratonovich calculus only occurs, if $\boldsymbol{B}$ depends on $\boldsymbol{x}(t)$, that is if $\partial_i B_{kj} \neq 0$.

To conclude this section we write down the Fokker-Planck equation in Stratonovich terms. One simply substitutes the coefficients $\boldsymbol{A}$ and $\boldsymbol{B}$ according to equations (15.41) and (15.42), and applies the product rule for differential operations to simplify the expression

$$\partial_t p(\boldsymbol{x},t|\boldsymbol{x}_0,t_0) = -\sum_i \partial_i \left[\overset{\mathscr{S}}{A}_i p(\boldsymbol{x},t|\boldsymbol{x}_0,t_0)\right] + \frac{1}{2}\sum_{i,j,k} \partial_i \overset{\mathscr{S}}{B}_{ik}\left[\partial_j \overset{\mathscr{S}}{B}_{jk} p(\boldsymbol{x},t|\boldsymbol{x}_0,t_0)\right]. \qquad (15.45)$$

XVI. EINSTEIN DIFFUSION EQUATION

In this section we want to consider the theory of the Fokker-Planck equation for molecules moving under the influence of random forces in force-free environments. Examples are molecules involved in Brownian motion in a fluid.

Obviously, this situation applies to many chemical and biochemical system and, therefore, is of great general interest. Actually, we will assume that the fluids considered are viscous in the sense that we will neglect the effects of inertia. The resulting description, referred to as *Brownian motion* in the limit of strong friction, applies to molecular systems except if one considers very brief time intervals of a picosecond or less. The general case of Brownian motion for arbitrary friction will be covered further below.

A. Derivation and Boundary Conditions

Particles moving in a liquid without forces acting on the particles, other than forces due to random collisions with liquid molecules, are governed by the Langevin equation

$$m\ddot{\boldsymbol{r}} = -\gamma\dot{\boldsymbol{r}} + \sigma\xi(t). \tag{16.1}$$

In the limit of strong friction holds

$$|\gamma\dot{\boldsymbol{r}}| \gg |m\ddot{\boldsymbol{r}}| \tag{16.2}$$

and Eq. (16.1) becomes

$$\gamma\dot{\boldsymbol{r}} = \sigma\xi(t). \tag{16.3}$$

To this stochastic differential equation corresponds the Fokker-Planck equation (c.f. Eqs. (14.7), (15.25), and (15.31))

$$\partial_t p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0) = \nabla^2 \frac{\sigma^2}{2\gamma^2} p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0), \tag{16.4}$$

with $\nabla^2 = \boldsymbol{\nabla}\cdot\boldsymbol{\nabla}$ the Laplacian operator. We assume in this section that σ and γ are spatially independent such that we can write

$$\partial_t p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0) = \frac{\sigma^2}{2\gamma^2} \nabla^2 p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0). \tag{16.5}$$

This is the celebrated *Einstein diffusion equation* which describes microscopic transport of material and heat.

In order to show that the Einstein diffusion equation (16.5) reproduces the well-known diffusive behavior of particles we consider the mean square displacement of a particle described by this equation, i.e.,

$$\left\langle (\boldsymbol{r}(t) - \boldsymbol{r}(t_0))^2 \right\rangle = \int_{\Omega_\infty} d^3\boldsymbol{r}\, (\boldsymbol{r} - \boldsymbol{r}_0)^2 p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0). \tag{16.6}$$

Integration over Eq. (16.5) yields

$$\frac{d}{dt}\left\langle (\boldsymbol{r}(t) - \boldsymbol{r}(t_0))^2 \right\rangle = \frac{\sigma^2}{2\gamma^2} \int_{\Omega_\infty} d^3\boldsymbol{r}\, (\boldsymbol{r} - \boldsymbol{r}_0)^2 \nabla^2 p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0). \tag{16.7}$$

Applying Green theorem for two functions $u(\boldsymbol{r})$ and $v(\boldsymbol{r})$

$$\int_{\Omega_\infty} d^3\boldsymbol{r}\, (u\nabla^2 v - v\nabla^2 u) = \int_{\partial\Omega_\infty} d\boldsymbol{a}\cdot(u\boldsymbol{\nabla}v - v\boldsymbol{\nabla}u), \tag{16.8}$$

for an infinite volume Ω and considering the fact that $p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0)$ must vanish at infinity we obtain

$$\frac{d}{dt}\left\langle (\boldsymbol{r}(t) - \boldsymbol{r}(t_0))^2 \right\rangle = \frac{\sigma^2}{2\gamma^2} \int_{\Omega_\infty} d^3\boldsymbol{r}\, p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0) \nabla^2 (\boldsymbol{r} - \boldsymbol{r}_0)^2. \tag{16.9}$$

With $\nabla^2(\boldsymbol{r} - \boldsymbol{r}_0)^2 = 6$ this is

$$\frac{d}{dt}\left\langle (\boldsymbol{r}(t) - \boldsymbol{r}(t_0))^2 \right\rangle = 6\frac{\sigma^2}{2\gamma^2} \int_{\Omega_\infty} d^3\boldsymbol{r}\, p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0). \tag{16.10}$$

For a reaction free case, when the conditional distribution is a normalized probability distribution, we can conclude

$$\left\langle (\boldsymbol{r}(t) - \boldsymbol{r}(t_0))^2 \right\rangle = 6\frac{\sigma^2}{2\gamma^2} t. \tag{16.11}$$

For diffusing particles one expects for this quantity a behavior $6Dt$ where D is the diffusion coefficient. Hence, the calculated dependence describes a diffusion process with diffusion coefficient

$$D = \frac{\sigma^2}{2\gamma^2}. \tag{16.12}$$

One can write the Einstein diffusion equation accordingly

$$\partial_t p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0) = D\nabla^2 p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0). \tag{16.13}$$

We have stated before that the Wiener process describes a diffusing particle as well. In fact, the three-dimensional generalization of (13.16)

$$p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0) = \frac{1}{\sqrt{4\pi(t-t_0)}} e^{-(\boldsymbol{r}-\boldsymbol{r}_0)^2/4D(t-t_0)}, \tag{16.14}$$

is the solution of (16.13) for the initial and boundary conditions

$$p(\boldsymbol{r}, t \to t_0|\boldsymbol{r}_0, t_0) = \delta(\boldsymbol{r}-\boldsymbol{r}_0), \quad p(|\boldsymbol{r}| \to \infty, t|\boldsymbol{r}_0, t_0) = 0. \tag{16.15}$$

One refers to the solution (16.14) as the *Green function*. The Green function is only uniquely defined if one specifies spatial boundary conditions on the surface $\partial\Omega$ surrounding the diffusion space. Once the Green function is available one can obtain the solution $p(\boldsymbol{r}, t)$ for any initial condition, e.g. for $p(\boldsymbol{r}, t \to t_0) = f(r)$

$$p(\boldsymbol{r}, t) = \int_{\Omega_\infty} d^3\boldsymbol{r}_0\, p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0) f(\boldsymbol{r}_0). \tag{16.16}$$

We will show below that one can also express the observables of the system in terms of the Green function. We will also introduce Green functions for different spatial boundary conditions. Once a Green function happens to be known, it is invaluable. However, because the Green function entails complete information about the time evolution of a system it is correspondingly difficult to obtain and its usefulness is confined often to formal manipulations. In this regard we will make extensive use of Green functions later on. The system described by the Einstein diffusion equation (16.13) may either be closed at the surface of the diffusion space or open, i.e., $\partial\Omega$ either may be impenetrable for particles or may allow passage of particles. In the latter case $\partial\Omega$ describes a reactive surface. These properties of Ω are specified through the boundary conditions on $\partial\Omega$. In order to formulate these boundary conditions we consider the flux of particles through consideration of the total number of particles diffusing in Ω defined through

$$N_\Omega(t|\boldsymbol{r}_0, t_0) = \int_\Omega d^3\boldsymbol{r}\, p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0). \tag{16.17}$$

In light of Einstein diffusion equation (16.13) we can then define a flux of particles as

$$\boldsymbol{j}(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0) = D\boldsymbol{\nabla} p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0). \tag{16.18}$$

This is also known as *Fick law* and it allows to express the change of the total number of particles in case the flux does not vanish at the surface $\partial\Omega$ of the diffusion space Ω, according to

$$\partial_t N_\Omega(t|\boldsymbol{r}_0, t_0) = \int_\Omega d^3\boldsymbol{r}\, \nabla^2 p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0) = \int_{\partial\Omega} d\boldsymbol{a} \cdot \boldsymbol{j}(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0), \tag{16.19}$$

where in the second equality Gauss theorem has been used.

We will refer to

$$\boldsymbol{\mathcal{J}}(\boldsymbol{r}) = D(\boldsymbol{r})\boldsymbol{\nabla}, \tag{16.20}$$

as the flux operator. This operator, when acting on a solution of the Einstein diffusion equation, yields the local flux of particles (probability) in the system.

The flux operator $\boldsymbol{\mathcal{J}}(r)$ governs the spatial boundary conditions since it allows one to measure particle (probability) exchange at the surface of the diffusion space. There are three types of boundary conditions possible. These types can be enforced simultaneously in disconnected areas of the surface $\partial\Omega$. Let us denote by $\partial\Omega_1, \partial\Omega_2$ two disconnected parts of $\partial\Omega$ such that $\partial\Omega = \partial\Omega_1 \cup \partial\Omega_2$. An example is a volume Ω lying between a sphere of radius R_1 ($\partial\Omega_1$) and of radius R_2 ($\partial\Omega_2$). The separation of the surfaces $\partial\Omega_i$ with different boundary conditions is necessary in order to assure that a continuous solution of the diffusion equation exists. Such solution cannot exist if it has to satisfy in an infinitesimal neighborhood entailing $\partial\Omega$ two different boundary conditions.

The first type of boundary condition is specified by

$$\hat{\boldsymbol{a}}(\boldsymbol{r}) \cdot \boldsymbol{\mathcal{J}}(\boldsymbol{r}) p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0) = 0, \quad \boldsymbol{r} \in \partial\Omega_i, \tag{16.21}$$

which obviously implies that particles do not cross the boundary, i.e., are reflected. Here $\hat{\boldsymbol{a}}(\boldsymbol{r})$ denotes a unit vector normal to the surface $\partial\Omega_i$ at $\boldsymbol{r}$. We will refer to (16.21) as the *reflection boundary condition*.

The second type of boundary condition is

$$p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0) = 0, \quad \boldsymbol{r} \in \partial\Omega_i. \tag{16.22}$$

This condition implies that all particles arriving at the surface $\partial\Omega_i$ are taken away such that the probability on $\partial\Omega_i$ vanishes. This boundary condition describes a reactive surface with the highest degree of reactivity possible, i.e., that every particle on $\partial\Omega_i$ reacts. We will refer to (16.22) as the *reaction boundary condition*.

The third type of boundary condition,

$$\hat{\boldsymbol{a}}(\boldsymbol{r}) \cdot \boldsymbol{\mathcal{J}}(\boldsymbol{r}) p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0) = w p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0), \quad \boldsymbol{r} \in \partial\Omega_i, \tag{16.23}$$

describes the case of *intermediate reactivity* at the boundary. The reactivity is measured by the parameter w. For $w = 0$ in (16.23) $\partial\Omega_i$ corresponds to a non-reactive, i.e., reflective boundary. For $w \to \infty$ the condition (16.23) can only be satisfied for $p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0) = 0$, i.e., every particle impinging onto $\partial\Omega_i$ is consumed in this case. We will refer to (16.23) as the *radiation boundary condition*.

In the following we want to investigate some exemplary instances of the Einstein diffusion equation for which analytic solutions are available.

B. Free Diffusion in One-Dimensional Half Space

As a first example we consider a particle diffusing freely in a one-dimensional half-space $x > 0$. This situation is governed by the Einstein diffusion problem in one dimension

$$\partial_t p(x, t|x_0, t_0) = D\partial_x^2 p(x, t|x_0, t_0), \tag{16.24}$$

$$p(x, t \to t_0|x_0, t_0) = \delta(x - x_0) \quad \text{initial condition}, \tag{16.25}$$

$$p(x \to \infty, t|x_0, t_0) = 0 \quad \text{boundary condition at } x = \infty, \tag{16.26}$$

$$\begin{cases} \partial_x p(x, t|x_0, t_0) = 0 & \text{reflecting wall at } x = 0, \\ p(x, t|x_0, t_0) = 0 & \text{absorbing wall at } x = 0, \end{cases} \quad \text{boundary condition at } x = 0. \tag{16.27}$$

One-Dimensional Half-Space with Reflecting Wall

For the *reflective* wall at $x = 0$ one finds for $x \geq 0$

$$p(x, t|x_0, t_0) = \frac{1}{\sqrt{4\pi D(t - t_0)}} \left(e^{-(x-x_0)^2/4D(t-t_0)} + e^{-(x+x_0)^2/4D(t-t_0)} \right), \tag{16.28}$$

This solution bears a simple interpretation. The first term of this solution describes a diffusion process which is unaware of the presence of the wall at $x = 0$. In fact, the term extends with nonvanishing values into the unavailable half-space $x \leq 0$. This "loss" of probability is corrected by the second term which, with its tail for $x \geq 0$, balances the missing probability. In fact, the $x \geq 0$ tail of the second term is exactly the mirror image of the"missing" $x \leq 0$ tail of the first term. One can envision that the second term reflects at $x = 0$ that fraction of the first term which describes a freely diffusing particle without the wall.

One-Dimensional Half-Space with Absorbing Wall

For the *absorbing* wall at $x = 0$ one finds for $x \geq 0$

$$p(x, t|x_0, t_0) = \frac{1}{\sqrt{4\pi D(t - t_0)}} \left(e^{-(x-x_0)^2/4D(t-t_0)} - e^{-(x+x_0)^2/4D(t-t_0)} \right) \tag{16.29}$$

In this case, because of particle removal by the wall at $x = 0$ the total number of particles is not conserved. The particle number corresponding to the Greens function $p(x, t|x_0, t_0)$ is

$$N(t|x_0, t_0) = \int_0^\infty dx\, p(x, t|x_0, t_0). \tag{16.30}$$

Introducing the integration variable $y = x/\sqrt{4D(t - t_0)}$ Eq. (16.30) can be rewritten

$$\begin{aligned} N(t|x_0, t_0) &= \frac{1}{\sqrt{\pi}} \int_0^\infty dy\, e^{-(y-y_0)^2} - \frac{1}{\sqrt{\pi}} \int_0^\infty dy\, e^{-(y+y_0)^2} \\ &= \frac{1}{\sqrt{\pi}} \int_{-y_0}^\infty dy\, e^{-y^2} - \frac{1}{\sqrt{\pi}} \int_{y_0}^\infty dy\, e^{-y^2} \\ &= \frac{1}{\sqrt{\pi}} \int_{-y_0}^{y_0} dy\, e^{-y^2} \\ &= \frac{2}{\sqrt{\pi}} \int_0^{y_0} dy\, e^{-y^2} \\ &= \mathrm{erf}(y_0) \\ &= \mathrm{erf}\left(\frac{x_0}{\sqrt{4D(t - t_0)}} \right). \end{aligned} \tag{16.31}$$

The particle number decays to zero asymptotically. In fact

$$N(t|x_0, t_0) \sim \frac{x_0}{\sqrt{\pi D(t - t_0)}} \quad \text{for } t \to \infty. \tag{16.32}$$

This decay is actually a consequence of the *ergodic theorem* which states that one-dimensional Brownian motion with certainty will visit every point of the space, i.e., also the absorbing wall. We will see below that for three-dimensional Brownian motion not all particles, even after arbitrary long time, will encounter a reactive boundary of finite size.

The rate of particle decay, according to (16.31), is

$$\partial_t N(t|x_0, t_0) = -\frac{1}{\sqrt{4\pi D(t - t_0)}(t - t_0)} e^{-x_0^2/4D(t-t_0)}. \tag{16.33}$$

An alternative route to determine the decay rate follows from (16.19) which reads for the case considered here,

$$\partial_t N(t|x_0, t_0) = -D\partial_x p(x, t|x_0, t_0)|_{x=0}. \tag{16.34}$$

Evaluation of this expression yields the same result as Eq. (16.33). This illustrates how useful the relationship (16.19) can be.

C. Free Diffusion in One-Dimensional Finite Interval

We consider now a particle diffusing freely in a finite, one-dimensional interval $\Omega = [0, a]$. The boundaries of Ω at $x = 0$ and a are assumed to be reflective. The diffusion coefficient D is assumed to be constant. The conditional distribution function $p(x, t|x_0, t_0)$ obeys the diffusion equation

$$\partial_t p(x, t|x_0, t_0) = D\partial_x^2 p(x, t|x_0, t_0), \tag{16.35}$$

$$p(x, t \to t_0|x_0, t_0) = \delta(x - x_0) \quad \text{initial condition}, \tag{16.36}$$

$$\partial_x p(x, t|x_0, t_0) = 0 \quad \text{reflecting wall at } x = 0, a. \tag{16.37}$$

In order to solve this diffusion problem we expand $p(x, t|x_0, t_0)$ in terms of eigenfunctions of the diffusion operator

$$\mathscr{L} = D\partial_x^2, \tag{16.38}$$

where we restrict the function space to those functions which obey (16.37). The corresponding functions are

$$v_n(x) = A_n \cos\left(\frac{n\pi}{a} x \right) \quad n = 0, 1, 2, \ldots. \tag{16.39}$$

In fact for these functions for $n = 0, 1, 2, \ldots$ holds

$$\mathscr{L} v_n(x) = \lambda_n v_n(x), \tag{16.40}$$

$$\lambda_n = -D \left(\frac{n\pi}{a} \right)^2. \tag{16.41}$$

We can then define the scalar product in the function space

$$\langle g|f\rangle_\Omega = \int_0^a dx\, g(x) f(x), \tag{16.42}$$

Then the eigenfunctions (16.39) will form an orthonormal set, $\langle v_n|v_m\rangle_\Omega = \delta_{nm}$, provided that A_n are chosen as

$$A_n = \begin{cases} \sqrt{1/a} & n = 0, \\ \sqrt{2/a} & n = 1, 2, \ldots. \end{cases} \tag{16.43}$$

Without proof we note that the functions v_n, defined in (16.39), form a complete basis for the function space considered. We can, hence, readily expand $p(x, t|x_0, t_0)$ in terms of v_n

$$p(x, t|x_0, t_0) = \sum_{n=0}^{\infty} \alpha_n(t|x_0, t_0) v_n(x). \tag{16.44}$$

Inserting this expansion into the diffusion equation (16.35) and taking the scalar product $\langle v_m|$ leads to

$$\partial_t \alpha_m(t|x_0, t_0) = \lambda_m \alpha_m(t|x_0, t_0), \tag{16.45}$$

from which we conclude

$$\alpha_m(t|x_0, t_0) = e^{\lambda_m (t-t_0)} \beta_m(x_0, t_0). \tag{16.46}$$

Here, $\beta_m(x_0, t_0)$ are time-independent constants which are determined by the initial condition (16.36)

$$\sum_{n=0}^{\infty} \beta_n(x_0, t_0) v_n(x) = \delta(x - x_0). \tag{16.47}$$

Taking again the scalar product $\langle v_m|$ results in

$$\beta_m(t|x_0, t_0) = v_m(x_0). \tag{16.48}$$

Summarizing we find

$$p(x, t|x_0, t_0) = \sum_{n=0}^{\infty} e^{\lambda_n (t-t_0)} v_n(x_0) v_n(x). \tag{16.49}$$

Let us assume now that the system considered is actually distributed initially according to a distribution $f(x)$ for which we assume $\langle 1|f\rangle_\Omega = 1$. The distribution $p(x, t)$, at later times, is then

$$p(x, t) = \int_0^a dx_0\, p(x, t|x_0, t_0) f(x_0). \tag{16.50}$$

Employing the expansion (16.49) this can be written

$$p(x, t) = \sum_{n=0}^{\infty} e^{\lambda_n (t-t_0)} v_n(x) \int_0^a dx_0\, v_n(x_0) f(x_0). \tag{16.51}$$

We consider now the behavior of $p(x, t)$ at long times. One expects that the system ultimately assumes a homogeneous distribution in Ω, i.e., that $p(x, t)$ relaxes as $\lim_{t\to\infty} p(x, t) = 1/a$. This asymptotic behavior, indeed, follows from (16.51). We note from (16.41)

$$\lim_{t\to\infty} e^{\lambda_n (t-t_0)} = \begin{cases} 1 & n = 0, \\ 0 & n = 1, 2, \ldots. \end{cases} \tag{16.52}$$

From (16.39) and (16.43) follows $v_0(x) = 1/\sqrt{a}$ and, hence,

$$\lim_{t\to\infty} p(x, t) = \frac{1}{a} \int_0^a dx_0\, f(x_0) = \frac{1}{a}, \tag{16.53}$$

where the last equality is due to the fact that f is a probability distribution. The solution presented here provides in a nutshell the typical properties of solutions of the more general Smoluchowski diffusion equation accounting for the presence of a force field which will be provided in Section XVII.

D. Free Diffusion around a Spherical Object

Likely the most useful example of a diffusion process stems from a situation encountered in a chemical reaction when a molecule diffuses around a target and either reacts with it or vanishes out of its vicinity. We consider the idealized situation that the target is stationary. Also we assume that the target is spherical (radius a) and reactions can arise anywwhere on its surface with equal likelyhood. Furthermore, we assume that the diffusing particles are distributed initially at a distance $r_0 > a$ from the center of the target with all directions being equally likely. In effect we describe an ensemble of reacting molecules and targets which undergo their reaction diffusion processes independently of each other.

The probability of finding the molecule at a distance r at time t is then described by a spherically symmetric distribution $p(r,t|r_0,t_0)$ since neither the initial condition nor the reaction-diffusion condition show any orientational preference. The ensemble of reacting molecules is then described by the diffusion equation

$$\partial_t p(r,t|r_0,t_0) = D\nabla^2 p(r,t|r_0,t_0), \tag{16.54}$$

$$p(r,t\to t_0|r_0,t_0) = \frac{1}{4\pi r_0^2}\delta(r-r_0) \quad \text{initial condition}, \tag{16.55}$$

$$p(r\to\infty,t|r_0,t_0) = 0 \quad \text{boundary condition at } r=\infty, \tag{16.56}$$

$$D\partial_r p(r,t|r_0,t_0) = wp(r,t|r_0,t_0) \quad \text{boundary condition at } r=a. \tag{16.57}$$

As pointed out above, w controls the likelyhood of encounters with the target to be reactive: $w=0$ corresponds to an unreactive surface, $w\to\infty$ to a surface for which every collision leads to reaction and, hence, to a diminishing of $p(r,t|r_0,t_0)$. The boundary condition for arbitrary w values adds significantly to the complexity of the solution, i.e., the following derivation would be simpler if the limits $w=0$ or $w\to\infty$ would be considered. However, a closed expression for the general case can be provided and, in view of the frequent applicability of the example we prefer the general solution.

We first notice that the Laplace operator ∇^2, expressed in spherical coordinates (r,θ,ϕ), reads

$$\nabla^2 = \frac{1}{r^2}\left[\partial_r(r^2\partial_r) + \frac{1}{\sin^2\theta}\partial_\phi^2 + \frac{1}{\sin\theta}\partial_\theta(\sin\theta\partial_\theta)\right]. \tag{16.58}$$

Since the distribution function $p(r,t|r_0,t_0)$ is spherically symmetric, i.e., depends solely on r and not on θ and ϕ, one can drop, for all practical purposes, the respective derivatives. Employing furthermore the identity

$$\frac{1}{r^2}\partial_r(r^2\partial_r f(r)) = \frac{1}{r}\partial_r^2(rf(r)). \tag{16.59}$$

one can restate the diffusion equation (16.54)

$$\partial_t rp(r,t|r_0,t_0) = D\partial_r^2 rp(r,t|r_0,t_0). \tag{16.60}$$

For the solution of (16.54)-(16.57) we partition

$$p(r,t|r_0,t_0) = u(r,t|r_0,t_0) + v(r,t|r_0,t_0), \tag{16.61}$$

$$u(r,t\to t_0|r_0,t_0) = \frac{1}{4\pi r_0^2}\delta(r-r_0), \tag{16.62}$$

$$v(r,t\to t_0|r_0,t_0) = 0. \tag{16.63}$$

The functions $u(r,t|r_0,t_0)$ and $v(r,t|r_0,t_0)$ are chosen to obey individually the radial diffusion equation (16.60) and, together, the boundary conditions (16.56) and (16.57). We fist construct $u(r,t|r_0,t_0)$ without regard to the boundary condition at $r=a$ and construct then $v(r,t|r_0,t_0)$ such that the proper boundary condition is obeyed.

The solution for u can be read from the previous Section XVI B

$$ru(r,t|r_0,t_0) = \frac{1}{4\pi r_0}\frac{1}{\sqrt{4\pi D(t-t_0)}}e^{-(r-r_0)^2/4D(t-t_0)}, \tag{16.64}$$

where we will worry about the boundary condition at $r=a$ later on.

We want to determine now the solution $v(r,t|r_0,t_0)$ which must satisfy

$$\partial_t rv(r,t|r_0,t_0) = D\partial_r^2 rv(r,t|r_0,t_0), \tag{16.65}$$

$$rv(r,t\to t_0|r_0,t_0) = 0. \tag{16.66}$$

Any solution of these homogeneous linear equations can be multiplied by an arbitrary constant C. This freedom allows one to modify $v(r,t|r_0,t_0)$ such that $u(r,t|r_0,t_0)+Cv(r,t|r_0,t_0)$ obeys the desired boundary condition (16.57) at $r=a$.

To construct a solution of (16.65) and (16.65) we consider the Laplace transformation

$$\tilde{v}(r,s|r_0,t_0)=\int_0^\infty d\tau\, e^{-s\tau}v(r,t_0+\tau|r_0,t_0). \tag{16.67}$$

Applying the Lapace transform to (16.65) and integrating by parts yields for the left hand side

$$-rv(r,t_0|r_0,t_0)+sr\tilde{v}(r,s|r_0,t_0). \tag{16.68}$$

The first term vanishes, according to (16.65), and one obtains

$$\frac{s}{D}r\tilde{v}(r,s|r_0,t_0)=\partial_r^2\tilde{v}(r,s|r_0,t_0). \tag{16.69}$$

The solution with respect to boundary condition (16.56) is

$$r\tilde{v}(r,s|r_0,t_0)=C(s|r_0)e^{-r\sqrt{s/D}}, \tag{16.70}$$

where $C(s|r_0)$ is an arbitrary constant which will be utilized to satisfy the boundary condition (16.57). In case of the function $r\tilde{p}(r,s|r_0,t_0)$ the extra factor r modifies the boundary condition. One can readily verify, using

$$D\partial_r r\tilde{p}(r,s|r_0,t_0)=D\tilde{p}(r,s|r_0,t_0)+rD\partial_r\tilde{p}(r,s|r_0,t_0), \tag{16.71}$$

and replacing at $r=a$ the last term by the right hand side of (16.57),

$$\partial_r r\tilde{p}(r,s|r_0,t_0)=\frac{wa+D}{Da}a\tilde{p}(a,s|r_0,t_0). \tag{16.72}$$

One can derive the Laplace transform of $u(r,t_0|r_0,t_0)$ using the identity

$$\int_0^\infty d\tau\, e^{-s\tau}\frac{1}{4\pi r_0}\frac{1}{\sqrt{4\pi D\tau}}e^{-(r-r_0)^2/4D\tau}=\frac{1}{4\pi r_0}\frac{1}{\sqrt{4Ds}}e^{-|r-r_0|\sqrt{s/D}}, \tag{16.73}$$

and one obtains for $r\tilde{p}(r,s|r_0,t_0)$

$$r\tilde{p}(r,s|r_0,t_0)=\frac{1}{4\pi r_0}\frac{1}{\sqrt{4Ds}}e^{-|r-r_0|\sqrt{s/D)}}+C(s|r_0)e^{-r\sqrt{s/D}}. \tag{16.74}$$

Solving for C in the boundary condition (16.72) gives

$$C(s|r_0)=\frac{\sqrt{s/D}-(wa+D)/Da}{\sqrt{s/D}+(wa+D)/Da}\frac{1}{4\pi r_0}\frac{1}{\sqrt{4Ds}}e^{-(r_0-2a)\sqrt{s/D)}}. \tag{16.75}$$

Application of the inverse Laplace transformation to $r\tilde{p}(r,s|r_0,t_0)$ leads to the final result

$$\begin{aligned}p(r,t|r_0,t_0)=&\frac{1}{4\pi rr_0}\frac{1}{\sqrt{4\pi D(t-t_0)}}\left(e^{-(r-r_0)^2/4D(t-t_0)}+e^{-(r+r_0-2a)^2/4D(t-t_0)}\right)\\&-\frac{1}{4\pi rr_0}\alpha e^{\alpha^2D(t-t_0)+\alpha(r+r_0-2a)}\mathrm{erfc}\left(\alpha\sqrt{D(t-t_0)}+\frac{r+r_0-2a}{\sqrt{4D(t-t_0)}}\right).\end{aligned}$$

where $\mathrm{erfc}(z)$ is the complimentary error function

$$\mathrm{erfc}(z)=\frac{2}{\sqrt{\pi}}\int_z^\infty dx\, e^{-x^2}, \tag{16.76}$$

and

$$\alpha=\frac{wa+D}{Da}=\begin{cases}1/a & \text{reflective boundary,}\\ \infty & \text{absorptive boundary,}\end{cases} \tag{16.77}$$

where in the last equality we used the properties of the intermediate reactivity boundary condition (16.23).

One can then find the *reaction rate* at $r = a$

$$K(t|r_0, t_0) = 4\pi a^2 D \partial_r p(r, t|r_0, t_0)|_{r=a} = 4\pi a^2 w p(a, t|r_0, t_0), \tag{16.78}$$

and the *fraction of particle reacted* at the boundary $r = a$

$$N_{\text{react}}(t|r_0, t_0) = \int_{t_0}^{t} du\, K(u|r_0, t_0). \tag{16.79}$$

In particular one finds

$$\lim_{t\to\infty} N_{\text{react}}(t|r_0, t_0) = \frac{a}{r_0} - \frac{1}{r_0\alpha}. \tag{16.80}$$

Even for $w, \alpha \to \infty$ this fraction is less than one in accordance with the ergodic behavior of particles diffusing in three-dimensional space. In order to overcome the a/r_0 limit on the overall reaction yield one can introduce long range interactions which effectively increase the reaction radius a.

E. Rotational Diffusion

We want to describe the dielectric relaxation. The electric polarization of liquids originates from the dipole moments of the individual liquid molecules. The contribution of an individual molecule to the polarization in the z-direction is

$$P_3 = P_0 \cos\theta. \tag{16.81}$$

We consider the relaxation of the dipole moment assuming that the rotational diffusion of the dipole moments can be described as diffusion on the unit sphere. The diffusion on a unit sphere is described by the three-dimensional diffusion equation

$$\partial_t p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0) = D\nabla^2 p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0) \tag{16.82}$$

for the condition $r = |\boldsymbol{r}| = 1$ and $r_0 = |\boldsymbol{r}_0| = 1$. In order to obey this condition one employs the Laplace operator ∇^2 in terms of spherical coordinates (r, θ, ϕ) as given in (16.58) and sets $r = 1$, dropping also derivatives with respect to r. This yields the rotational diffusion equation

$$\partial_t p(\Omega, t|\Omega_0, t_0) = \tau^{-1}\left[\frac{1}{\sin^2\theta}\partial_\phi^2 + \frac{1}{\sin\theta}\partial_\theta(\sin\theta\partial_\theta)\right] p(\Omega, t|\Omega_0, t_0), \tag{16.83}$$

$$p(\Omega, t \to t_0|\Omega_0, t_0) = \delta(\Omega - \Omega_0) \quad \text{initial condition}, \tag{16.84}$$

where $\Omega = (\theta, \phi)$ is the solid angle with $d\Omega = d(\cos\theta)d\phi$ the elementary solid angle and $\delta(\Omega) = \delta(\cos\theta)\delta(\phi)$. We have also introduced, instead of the diffusion constant, the rate constant τ^{-1} since the replacement $r \to 1$ altered the units in the diffusion equation; τ has the unit of time. In the present case the diffusion space has no boundary; however, we need to postulate that the distribution and its derivatives are continuous on the sphere.

One way of ascertaining the continuity property is to expand the distribution in terms of spherical harmonics $Y_{lm}(\Omega)$ which obey the proper continuity, i.e.,

$$p(\Omega, t|\Omega_0, t_0) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} A_{lm}(t|\Omega_0, t_0) Y_{lm}(\Omega). \tag{16.85}$$

In addition, one can exploit the eigenfunction property

$$\left[\frac{1}{\sin^2\theta}\partial_\phi^2 + \frac{1}{\sin\theta}\partial_\theta(\sin\theta\partial_\theta)\right] Y_{lm}(\Omega) = -l(l+1)Y_{lm}(\Omega). \tag{16.86}$$

Inserting Eq. (16.85) into the diffusion equation (16.83) and using (16.86) results in

$$\sum_{l=0}^{\infty}\sum_{m=-l}^{l} \partial_t A_{lm}(t|\Omega_0, t_0) Y_{lm}(\Omega) = -\sum_{l=0}^{\infty}\sum_{m=-l}^{l} l(l+1)\tau^{-1} A_{lm}(t|\Omega_0, t_0) Y_{lm}(\Omega). \tag{16.87}$$

The orthonormality of the spherical harmonics

$$\int d\Omega\, Y^*_{l'm'}(\Omega) Y_{lm}(\Omega) = \delta_{l'l}\delta_{m'm}, \tag{16.88}$$

leads to

$$\partial_t A_{lm}(t|\Omega_0,t_0) = -l(l+1)\tau^{-1}A_{lm}(t|\Omega_0,t_0), \tag{16.89}$$

with solution

$$A_{lm}(t|\Omega_0,t_0) = e^{-l(l+1)(t-t_0)/\tau}a_{lm}(\Omega_0), \tag{16.90}$$

and

$$p(\Omega,t|\Omega_0,t_0) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} e^{-l(l+1)(t-t_0)/\tau}a_{lm}(\Omega_0)Y_{lm}(\Omega). \tag{16.91}$$

The coefficients a_{lm} are determined from the initial condition (16.84) and the completeness relationship of spherical harmonics states

$$\sum_{l=0}^{\infty}\sum_{m=-l}^{l} Y^*_{lm}(\Omega_0)Y_{lm}(\Omega) = \delta(\Omega-\Omega_0), \tag{16.92}$$

so that from Eq. (16.91) with $t \to t_0$ follows

$$a_{lm}(\Omega_0) = Y^*_{lm}(\Omega_0), \tag{16.93}$$

and, hence,

$$p(\Omega,t|\Omega_0,t_0) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} e^{-l(l+1)(t-t_0)/\tau}Y^*_{lm}(\Omega_0)Y_{lm}(\Omega). \tag{16.94}$$

It is interesting to consider the asymptotic, i.e., the $t \to \infty$, behavior of this solution. All exponential terms will vanish, except the term with $l = 0$. Hence, the distribution approaches asymptotically the limit

$$\lim_{t\to\infty} p(\Omega,t|\Omega_0,t_0) = Y^2_{00}(\Omega) = \frac{1}{4\pi}. \tag{16.95}$$

This result corresponds to the homogeneous, normalized distribution on the sphere, a result which one may have expected all along. One refers to this distribution as the equilibrium distribution denoted by

$$p_0(\Omega) = \frac{1}{4\pi}. \tag{16.96}$$

The equilibrium average of the polarization expressed in (16.81) is

$$\langle P_3\rangle = \int d\Omega\, P_0\cos(\theta)p_0(\Omega) = 0. \tag{16.97}$$

Another quantity of interest is the so-called equilibrium correlation function

$$\langle P_3(t)P_3(t_0)\rangle = P_0^2\int d\Omega\int d\Omega_0\cos(\theta)\cos(\theta_0)p(\Omega,t|\Omega_0,t_0)p_0(\Omega_0). \tag{16.98}$$

Using

$$Y_{10}(\Omega) = \sqrt{\frac{3}{4\pi}}\cos\theta, \tag{16.99}$$

one finds

$$\langle P_3(t)P_3(t_0)\rangle = \frac{1}{3}P_0^2\sum_{m=-l}^{l} e^{-l(l+1)(t-t_0)/\tau}|C_{10,lm}|^2, \tag{16.100}$$

where

$$C_{10,lm} = \int d\Omega\, Y^*_{10}(\Omega)Y_{lm}(\Omega) = \delta_{l1}\delta_{m0}, \tag{16.101}$$

due to the orthonormality condition of the spherical harmonics (16.88). And therefore

$$\langle P_3(t)P_3(t_0)\rangle = \frac{1}{3}P_0^2 e^{-2(t-t_0)/\tau}. \tag{16.102}$$

XVII. SMOLUCHOWSKI DIFFUSION EQUATION

We want to apply now our derivation to the case of a Brownian particle in a force field $\boldsymbol{F}(\boldsymbol{r})$. The corresponding Langevin equation is

$$m\ddot{\boldsymbol{r}} = -\gamma\dot{\boldsymbol{r}} + \boldsymbol{F}(\boldsymbol{r}) + \sigma\xi(t), \tag{17.1}$$

for scalar friction γ constant and amplitude σ of the fluctuating force. We will assume in this section the limit of strong friction. In this limit the magnitude of the frictional force $\gamma\dot{\boldsymbol{r}}$ is much larger than the magnitude of the force of inertia $m\ddot{\boldsymbol{r}}$, i.e.,

$$|\gamma\dot{\boldsymbol{r}}| \gg |m\ddot{\boldsymbol{r}}| \tag{17.2}$$

and therefore Eq. (17.1) becomes

$$\gamma\dot{\boldsymbol{r}} = \boldsymbol{F}(\boldsymbol{r}) + \sigma\xi(t). \tag{17.3}$$

To this stochastic differential equation corresponds the Fokker-Planck equation (c.f. Eqs. (14.7), (15.25), and (15.31))

$$\partial_t p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0) = \left(\nabla^2 \frac{\sigma^2}{2\gamma^2} - \boldsymbol{\nabla}\cdot\frac{\boldsymbol{F}(\boldsymbol{r})}{\gamma}\right) p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0). \tag{17.4}$$

In case that the force field can be related to a scalar potential, i.e., in case $\boldsymbol{F}(\boldsymbol{r}) = -\boldsymbol{\nabla}U(\boldsymbol{r})$, one expects that the Boltzmann distribution $\exp[-U(\boldsymbol{r})/k_BT]$ is a stationary, i.e., time-independent, solution and that, in fact, the system asymptotically approaches this solution. This expectation should be confined to force fields of the stated kind, i.e., to force fields for which holds $\boldsymbol{\nabla}\times\boldsymbol{F} = 0$. Fokker-Planck equations with more general force fields will be considered further below.

A. Derivation of the Smoluchoswki Diffusion Equation for Potential Fields

It turns out that the expectation that the Boltzmann distribution is a stationary solution of the Smoluchowski equation has to be introduced as a postulate rather than a consequence of (17.4). Defining the parameters $D = \sigma^2/2\gamma^2$ (cf. Eq. (16.12)) and $\beta = 1/k_BT$ the *postulate* of the stationary behavior of the Boltzmann equation is

$$\left(\boldsymbol{\nabla}\cdot\boldsymbol{\nabla}D(\boldsymbol{r}) - \boldsymbol{\nabla}\cdot\frac{\boldsymbol{F}(\boldsymbol{r})}{\gamma(\boldsymbol{r})}\right) e^{-\beta U(\boldsymbol{r})} = 0. \tag{17.5}$$

We have included here the possibility that the coefficients σ and γ defining the fluctuating and dissipative forces are spatially dependent. In the following we will not explicitly state the dependence on the spatial coordinates $\boldsymbol{r}$ anymore.

Actually, the postulate (17.5) of the stationarity of the Boltzmann distribution is not sufficient to obtain an equation with the appropriate behavior at thermal equilibrium. Actually, one needs to require the more stringent postulate that at equilibrium there does not exist a net flux of particles (or of probability) in the system. This should hold true when the system asymptotically comes to rest as long as there are no particles generated or destroyed, e.g., through chemical reactions. We need to establish the expression for the flux before we can investigate the ramifications of the indicated postulate.

An expression for the flux can be obtained in a vein similar to that adopted in the case of free diffusion (cf. Eq. (16.18)). We note that (17.4) can be written

$$\partial_t p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0) = \boldsymbol{\nabla}\cdot\left(\boldsymbol{\nabla}D - \frac{\boldsymbol{F}(\boldsymbol{r})}{\gamma}\right) p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0). \tag{17.6}$$

We can then define a flux of particles as

$$\boldsymbol{j}(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0) = \left(\boldsymbol{\nabla}D - \frac{\boldsymbol{F}(\boldsymbol{r})}{\gamma}\right) p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0). \tag{17.7}$$

Since particles are neither generated nor destroyed inside the volume Ω, we must interpret $\boldsymbol{j}(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0)$ as a particle flux at the boundary $\partial\Omega$. Since the volume and its boundary are arbitrary, the interpretation of $\boldsymbol{j}(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0)$ as given by (17.7) as a flux should hold everywhere in Ω.

We can now consider the ramifications of the postulate that at equilibrium the flux vanishes. Applying (17.7) to the Boltzmann distribution $p_B(\boldsymbol{r}) = N\exp[-\beta U(\boldsymbol{r})]$, for some appropriate normalization factor N, yields the equilibrium flux

$$\boldsymbol{j}_B(\boldsymbol{r}) = \left(\boldsymbol{\nabla} D - \frac{\boldsymbol{F}(\boldsymbol{r})}{\gamma}\right) N e^{-\beta U(\boldsymbol{r})}. \tag{17.8}$$

With this definition the postulate discussed above is $\boldsymbol{j}_B = 0$ which can be written as

$$e^{-\beta U(\boldsymbol{r})}\left(D\beta\boldsymbol{F}(\boldsymbol{r}) + \boldsymbol{\nabla} D - \frac{\boldsymbol{F}(\boldsymbol{r})}{\gamma}\right) = 0. \tag{17.9}$$

From this follows

$$\boldsymbol{\nabla} D = \boldsymbol{F}(\boldsymbol{r})(\gamma^{-1} - D\beta), \tag{17.10}$$

an identity which is known as the so-called *fluctuation-dissipation theorem*.

The fluctuation-dissipation theorem is better known for the case of spatially independent D in which case follows $D\beta\gamma = 1$, i.e., with the definitions above

$$\sigma^2 = 2k_B T\gamma. \tag{17.11}$$

This equation implies a relationship between the amplitude σ of the fluctuating forces and the amplitude γ of the dissipative (frictional) forces in the Langevin equation (17.1), hence, the name fluctuation-dissipation theorem. The theorem states that the amplitudes of fluctuating and dissipative forces need to obey a temperature-dependent relationship in order for a system to attain thermodynamic equilibrium. There exist more general formulations of this theorem which we will discuss further below in connection with response and correlation functions.

In its form (17.10) the fluctuation-dissipation theorem allows us to reformulate the Fokker-Planck equation above. For any function $f(\boldsymbol{r})$ holds with (17.10)

$$\boldsymbol{\nabla}\cdot\boldsymbol{\nabla} Df = \boldsymbol{\nabla}\cdot D\boldsymbol{\nabla} f + \boldsymbol{\nabla}\cdot f\boldsymbol{\nabla} D = \boldsymbol{\nabla}\cdot D\boldsymbol{\nabla} f + \boldsymbol{\nabla}\cdot\boldsymbol{F}(\gamma^{-1} - D\beta)f. \tag{17.12}$$

From this follows finally for the Fokker-Planck equation (17.4)

$$\partial_t p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0) = \boldsymbol{\nabla}\cdot D\left(\boldsymbol{\nabla} - \beta\boldsymbol{F}(\boldsymbol{r})\right) p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0). \tag{17.13}$$

One refers to this equation as the *Smoluchowski equation*.

The Smoluchowski equation (17.13), in the case $\boldsymbol{F}(\boldsymbol{r}) = -\boldsymbol{\nabla} U(\boldsymbol{r})$, can be written in the convenient form

$$\partial_t p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0) = \boldsymbol{\nabla}\cdot De^{-\beta U(\boldsymbol{r})}\boldsymbol{\nabla} e^{\beta U(\boldsymbol{r})} p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0). \tag{17.14}$$

This form shows immediately that p_B is a stationary solution. The form also provides a new expression for the flux $\boldsymbol{j}$, namely,

$$\boldsymbol{j}(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0) = D(\boldsymbol{\nabla} - \beta\boldsymbol{F}(\boldsymbol{r}))p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0) = De^{-\beta U(\boldsymbol{r})}\boldsymbol{\nabla} e^{\beta U(\boldsymbol{r})} p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0), \tag{17.15}$$

and the flux operator

$$\boldsymbol{\mathcal{J}}(\boldsymbol{r}) = D(\boldsymbol{\nabla} - \beta\boldsymbol{F}(\boldsymbol{r})). \tag{17.16}$$

This operator, when acting on a solution of the Smoluchowski equation, yields the local flux of particles (probability) in the system. As for the free diffusion case of Section XVI A we can think of the three spatial boundary conditions (16.21), (16.22), and (16.23).

B. One-Dimensional Diffuson in a Linear Potential

We consider now diffusion in a linear potential

$$U(x) = cx, \tag{17.17}$$

with a position-independent diffusion coefficient D. This system is described by the Smoluchowski equation (17.13)

$$\partial_t p(x, t|x_0, t_0) = D\left(\partial_x^2 + \beta c\partial_x\right) p(x, t|x_0, t_0). \tag{17.18}$$

This will be the first instance that we study of a system in which diffusing particles are acted on by a non-vanishing force. The techniques to solve the Smoluchowski equation in the present case will be particular for the simple force field here considered.

Infinite Space

We consider first the situation that the particles diffusing under the influence of the potential (17.17) have available the infinite space $\Omega_\infty =]-\infty, \infty[$. In this case hold the boundary conditions

$$\lim_{x \to \pm\infty} p(x,t|x_0,t_0) = 0. \tag{17.19}$$

The initial condition is as usual

$$p(x,t_0|x_0,t_0) = \delta(x - x_0). \tag{17.20}$$

In order to solve (17.18), (17.19), and (17.20) we introduce the following time dependent spatial coordinates

$$\tau = Dt,\ b = \beta c,\ y = x + b\tau,\ y_0 = x_0 + b\tau_0, \tag{17.21}$$

and express the solution $p(x,\tau||x_0,\tau_0) = q(y,\tau|y_0,\tau_0)$. Introducing this into (17.18) yields

$$\partial_\tau q(y,\tau|y_0,\tau_0) = \partial_y^2 q(y,\tau|y_0,\tau_0). \tag{17.22}$$

This equation has the same form as the Einstein equation for freely diffusing particles for which the solution in case of the diffusion space Ω_∞ is

$$q(y,\tau|y_0,\tau_0) = \frac{1}{\sqrt{4\pi(\tau - \tau_0)}} e^{-(y-y_0)^2/4(\tau-\tau_0)}. \tag{17.23}$$

Expressing the solution in terms of the original coordinates and constants yields

$$p(x,t|x_0,t_0) = \frac{1}{\sqrt{4\pi D(t - t_0)}} e^{-(x-x_0+D\beta c(t-t_0))^2/4D(t-t_0)}. \tag{17.24}$$

This solution is identical to the distribution of freely diffusing particles, except that the center of the distribution drifts down the potential gradient with a velocity $D\beta c$.

The above result (17.24) can be used for a stochastic processes with exponential growth by performing a simple substitution. Comparing the Fokker-Planck equation (15.31) and the Langevin equation (14.3) with the Smoluchowski equation (17.18) of the previous example one can easily derive within Ito calculus the corresponding stochastic differential equation

$$\partial_t x(t) = -D\beta c + \sqrt{D}\xi(t), \tag{17.25}$$

or equivalently

$$dx = -D\beta c\, dt + \sqrt{D}\, d\omega. \tag{17.26}$$

This equation displays the mechanism that generates the stochastic trajectories within a linear potential (17.17) . The increment dx of a trajectory $x(t)$ is given by the drift term $-D\beta c\, dt$, which is determined by the force c of the linear potential and the friction $\gamma = D\beta$

. Furthermore the increment dx is subject to Gaussian noise $d\omega$ scaled by $\sqrt{D}$.

We now consider a transformation of the spatial variable x. Let $x \to y = e^x$. This substitution and the resulting differential $dy/y = dx$ render the stochastic differential equation

$$dy = -D\beta c y\, dt + \sqrt{D} y\, d\omega. \tag{17.27}$$

This equation describes a different stochastic process $y(t)$. Just considering the first term on the right hand side of (17.27), one sees that $y(t)$ is subject to exponential growth or decay depending on the sign of c. Neglecting the second term on the right hand side of (17.27) one obtains the deterministic trajectory

$$y(t) = y(0)e^{-D\beta ct}. \tag{17.28}$$

This dynamics is typical for growth or decay processes in physics, biology or economics. Furthermore, $y(t)$ is subject to Gaussian noise $d\omega$ scaled by $y\sqrt{D}$. The random fluctuation are consequently proportional to y, which is the case when the growth rate and not just the increment are subject to stochastic fluctuations.

Since (17.26) and (17.27) are connected via the simple mapping $y = e^x$ we can readily state the solution of equation (17.27) by substituting $\ln y$ for x in (17.24)

$$p(y,t|y_0,t_0) = p(x(y),t|x(y_0),t_0)\frac{dx}{dy} = \frac{1}{\sqrt{4\pi D(t - t_0)y}} e^{-(\ln(y/y_0)+D\beta c(t-t_0))^2/4D(t-t_0)}. \tag{17.29}$$

Half Space

We consider now diffusion in a half-space $\Omega_\infty = [0, \infty[$ under the influence of a linear potential (17.17) with a reflective boundary at $x = 0$

$$D(\partial_x + \beta c)p(x, t|x_0, t_0) = 0 \quad \text{at } x = 0, \tag{17.30}$$

$$\lim_{x\to\infty} p(x, t|x_0, t_0) = 0. \tag{17.31}$$

The initial condition is as usual

$$p(x, t_0|x_0, t_0) = \delta(x - x_0). \tag{17.32}$$

The solution has been determined by Smoluchowski [14] and can be stated in the form

$$p(x, t_0|x_0, 0) = \sum_{j=1}^{3} p_j(x, t_0|x_0, 0), \tag{17.33}$$

$$p_1(x, t_0|x_0, 0) = \frac{1}{\sqrt{4\pi Dt}} e^{-(x-x_0+\beta cDt)^2/4Dt}, \tag{17.34}$$

$$p_2(x, t_0|x_0, 0) = \frac{1}{\sqrt{4\pi Dt}} e^{\beta cx_0-(x+x_0+\beta cDt)^2/4Dt}, \tag{17.35}$$

$$p_3(x, t_0|x_0, 0) = \frac{\beta c}{2} e^{-\beta cx} \text{erfc}\left[(x + x_0 - \beta cDt)/\sqrt{4Dt}\right]. \tag{17.36}$$

C. One-Dimensional Diffusion in a Harmonic Potential

We consider now diffusion in a harmonic potential

$$U(x) = \frac{1}{2} f x^2, \tag{17.37}$$

which is simple enough to yield an analytic solution of the corresponding Smoluchowski equation

$$\partial_t p(x, t|x_0, t_0) = D\left(\partial_x^2 + \beta f \partial_x x\right) p(x, t|x_0, t_0). \tag{17.38}$$

We assume presently a constant diffusion coefficient D. The particle can diffuse in the infinite space $\Omega_\infty =]-\infty, \infty[$. However, the potential confines the motion to a finite area such that the probability distribution vanishes exponentially for $x \to \pm\infty$ as expressed through the boundary condition

$$\lim_{x\to\pm\infty} x^n p(x, t|x_0, t_0) = 0 \quad \forall n \in \mathbb{N}. \tag{17.39}$$

The initial condition is as usual

$$p(x, t_0|x_0, t_0) = \delta(x - x_0). \tag{17.40}$$

In thermal equilibrium, particles will be distributed according to the Boltzmann distribution

$$p_0(x) = \sqrt{\frac{\beta f}{2\pi}} e^{-\beta f x^2/2}. \tag{17.41}$$

We expect that the solution for the initial condition (17.40) will asymptotically decay towards (17.41).

The mean square deviation from the average position of the particle at equilibrium, i.e., from $\langle x\rangle = 0$, is

$$\delta^2 = \int_{-\infty}^{\infty} dx\, (x - \langle x\rangle)^2 p_0(x) = \frac{1}{2\kappa^2}, \tag{17.42}$$

$$\kappa = \sqrt{\frac{\beta f}{2}}. \tag{17.43}$$

where we used result $I_1(1) = \sqrt{\pi}/2$ in Eq. (C19). Or $\delta = 1/\sqrt{\beta f}$. We then introduce

$$\xi = x/\sqrt{2}\delta, \ \tau = t/\tilde{\tau}, \tag{17.44}$$

with $\tilde{\tau} = 2\delta^2/D$. Then the Smoluchowski equation for

$$q(\xi, \tau|\xi_0, \tau_0) = \sqrt{2}\delta p(x, t_0|x_0, t_0), \tag{17.45}$$

reads

$$\partial_\tau q(\xi, \tau|\xi_0, \tau_0) = (\partial_\xi^2 + 2\partial_\xi \xi) q(\xi, \tau|\xi_0, \tau_0), \tag{17.46}$$

$$q(\xi, \tau \to \tau_0|\xi_0, \tau_0) = \delta(\xi - \xi_0) \quad \text{initial condition}, \tag{17.47}$$

$$\lim_{\xi \to \pm\infty} \xi^n q(\xi, \tau|\xi_0, \tau_0) = 0 \quad \text{boundary condition } \forall n \in \mathbb{N}. \tag{17.48}$$

In the following we choose $\tau_0 = 0$. In order to solve (17.46)-(17.48) we seek to transform the Smoluchowski equation to the free diffusion equation through the choice of the time-dependent position variable

$$y = \xi e^{2\tau}, \; y_0 = \xi_0, \tag{17.49}$$

replacing

$$q(\xi, \tau|\xi_0, 0) = v(y, \tau|y_0, 0). \tag{17.50}$$

We note that this definition results in a time-dependent normalization of $v(y, \tau|y_0, 0)$, namely,

$$1 = \int_{-\infty}^{\infty} d\xi \, q(\xi, \tau|\xi_0, 0) = e^{-2\tau} \int_{-\infty}^{\infty} dy \, v(y, \tau|y_0, 0). \tag{17.51}$$

Using

$$\partial_\xi = \frac{\partial y}{\partial \xi} \partial y = e^{2\tau} \partial_y, \tag{17.52}$$

$$\partial_\xi^2 = e^{4\tau} \partial_y^2, \tag{17.53}$$

$$\begin{aligned} \partial_\tau q &= \partial_\tau v + \frac{\partial y}{\partial \tau} \partial_y v \\ &= \partial_\tau v + 2y \partial_y v, \end{aligned} \tag{17.54}$$

we find the following Smoluchowski equation for v

$$\partial_\tau v(y, \tau|y_0, 0) = e^{4\tau} \partial_y^2 v(y, \tau|y_0, 0) + 2v(y, \tau|y_0, 0). \tag{17.55}$$

In order to deal with a properly normalized distribution we define further

$$v(y, \tau|y_0, 0) = e^{2\tau} w(y, \tau|y_0, 0). \tag{17.56}$$

The Smoluchowski equation for w becomes then

$$\partial_\tau w(y, \tau|y_0, 0) = e^{4\tau} \partial_y^2 w(y, \tau|y_0, 0), \tag{17.57}$$

which, indeed, has the form of a free diffusion equation, albeit with a time-dependent diffusion coefficient $\tilde{D}(\tau) = e^{4\tau}$.

It turns out that the solution of a diffusion equation with time-dependent diffusion coefficient

$$\partial_\tau w(y, \tau|y_0, 0) = \tilde{D}(\tau) \partial_y^2 w(y, \tau|y_0, 0), \tag{17.58}$$

$$w(y, \tau \to 0|y_0, 0) = \delta(y - y_0) \quad \text{initial condition}, \tag{17.59}$$

is a straightforward generalization of the corresponding solution of the free diffusion equation, namely,

$$w(y, \tau|y_0, 0) = \left(4\pi \int_0^\tau d\tau' \, \tilde{D}(\tau') \right)^{-1/2} e^{-(y-y_0)^2/4 \int_0^\tau d\tau' \, \tilde{D}(\tau')}. \tag{17.60}$$

Then the solution for w is

$$w(y, \tau|y_0, 0) = \left(4\pi \int_0^\tau d\tau' \, e^{4\tau'} \right)^{-1/2} e^{-(y-y_0)^2/4 \int_0^\tau d\tau' \, e^{4\tau'}}. \tag{17.61}$$

The corresponding solution for q is

$$q(\xi, \tau|\xi_0, 0) = \frac{1}{\sqrt{\pi \left(1 - e^{-4\tau}\right)}} e^{-\left(\xi - \xi_0 e^{-2\tau}\right)^2 / \left(1 - e^{-4\tau}\right)}. \tag{17.62}$$

We then arrive at the following solution for p

$$p(x,t|x_0,t_0) = \frac{1}{\sqrt{\pi S(t,t_0)}} e^{-\left(x - x_0 e^{-2(t-t_0)/\tilde{\tau}}\right)^2 / S(t,t_0)}, \tag{17.63}$$

$$S(t,t_0) = \frac{2}{\beta f}\left(1 - e^{-4(t-t_0)/\tilde{\tau}}\right), \tag{17.64}$$

$$\tilde{\tau} = \frac{2}{\beta f D}. \tag{17.65}$$

One notices that this distribution asymptotically, i.e., for $t \to \infty$, approaches the Boltzmann distribution (17.41). We also note that this distribution (17.63)-(17.65) is identical to the conditional probability of the Ornstein-Uhlenbeck process (13.37)-(13.39) for

$$\gamma = \frac{2}{\tilde{\tau}},\ \gamma\sigma^2 = \frac{2}{\beta f} \to \sigma^2 = \frac{\tilde{\tau}}{\beta f} = \frac{2}{(\beta f)^2 D}. \tag{17.66}$$

An alternative derivation of this *Brownian oscillator* is offered in Appendix B where an expansion in Hermite polynomials is used.

XVIII. ADJOINT SMOLUCHOWSKI DIFFUSION EQUATION

The adjoint or backward Smoluchowski equation governs the $\boldsymbol{r}_0$-dependence of the solution $p(\boldsymbol{r},t|\boldsymbol{r}_0,t_0)$ of the Smoluchowski equation also referred to as the *forward* equation. The *backward* equation complements the forward equation and it often useful to determine observables connected with the solution of the Smoluchowski equation.

A. Forward and Backward Smoluchowski Equation

The Smoluchowski equation in a diffusion domain Ω can be written

$$\partial_t p(\boldsymbol{r},t|\boldsymbol{r}_0,t_0) = \mathscr{L}(\boldsymbol{r}) p(\boldsymbol{r},t|\boldsymbol{r}_0,t_0), \tag{18.1}$$

where

$$\mathscr{L}(\boldsymbol{r}) = \boldsymbol{\nabla}\cdot D(\boldsymbol{\nabla} - \beta \boldsymbol{F}(\boldsymbol{r})). \tag{18.2}$$

For $\boldsymbol{F}(\boldsymbol{r}) = -\boldsymbol{\nabla}U(\boldsymbol{r})$ one can express

$$\mathscr{L}(\boldsymbol{r}) = \boldsymbol{\nabla}\cdot D e^{-\beta U(\boldsymbol{r})} \boldsymbol{\nabla} e^{\beta U(\boldsymbol{r})}. \tag{18.3}$$

With the Smoluchowski equation (18.1) are associated three possible spatial boundary conditions for $h(r) = p(\boldsymbol{r},t|\boldsymbol{r}_0,t_0)$ on the surface $\partial\Omega$ of the diffusion domain Ω with local normal $\hat{\boldsymbol{a}}(\boldsymbol{r})$:

$$(i) \qquad \hat{\boldsymbol{a}}(\boldsymbol{r})\cdot D(\boldsymbol{\nabla} - \beta\boldsymbol{F}(\boldsymbol{r}))h(\boldsymbol{r}) = 0, \qquad \boldsymbol{r}\in\partial\Omega \tag{18.4}$$

$$(ii) \qquad h(\boldsymbol{r}) = 0, \qquad \boldsymbol{r}\in\partial\Omega \tag{18.5}$$

$$(iii) \qquad \hat{\boldsymbol{a}}(\boldsymbol{r})\cdot D(\boldsymbol{\nabla} - \beta\boldsymbol{F}(\boldsymbol{r}))h(\boldsymbol{r}) = w(\boldsymbol{r})h(\boldsymbol{r}), \qquad \boldsymbol{r}\in\partial\Omega \tag{18.6}$$

where, in the latter equation, $w(\boldsymbol{r})$ is a continuous function which describes the effectivity of the surface $\partial\Omega$ to react locally. The Eqs. (18.1)-(18.6) allow one to determine the probability $p(\boldsymbol{r},t|\boldsymbol{r}_0,t_0)$ to find a particle at position $\boldsymbol{r}$ at time t, given that the particle started diffusion at position $\boldsymbol{r}_0$ at time t_0. It holds

$$p(\boldsymbol{r},t_0|\boldsymbol{r}_0,t_0) = \delta(\boldsymbol{r}-\boldsymbol{r}_0). \tag{18.7}$$

For the Smoluchowski equation (18.1) exists an alternative form

$$\partial_t p(\boldsymbol{r},t|\boldsymbol{r}_0,t_0) = \mathscr{L}^\dagger(\boldsymbol{r}_0) p(\boldsymbol{r},t|\boldsymbol{r}_0,t_0), \tag{18.8}$$

the so-called *adjoint* or *backward* equation, which involves a differential operator that acts on the $\boldsymbol{r}_0$-dependence of $p(\boldsymbol{r},t|\boldsymbol{r}_0,t_0)$. The latter operator $\mathscr{L}^\dagger(\boldsymbol{r}_0)$ is the adjoint of the operator $\mathscr{L}(\boldsymbol{r})$ defined in (18.2) above.

Below we will determine the operator $\mathscr{L}^\dagger(\boldsymbol{r}_0)$ as well as the boundary conditions which the solution $p(\boldsymbol{r},t|\boldsymbol{r}_0,t_0)$ of (18.8) must obey when $p(\boldsymbol{r},t|\boldsymbol{r}_0,t_0)$ obeys the boundary conditions (18.4)-(18.6) in the original, so-called *forward* Smoluchowski equation (18.1). Before proceeding with the derivation of the backward Smoluchowski equation we need to provide two key properties of the solution $p(\boldsymbol{r},t|\boldsymbol{r}_0,t_0)$ of the forward Smoluchowski equation (18.1) connected with the time translation invariance of the equation and with the Markov property of the underlying stochastic process.

Homogeneous Time

In case that the Smoluchowski operator $\mathscr{L}(\boldsymbol{r})$ governing (18.1) and given by (18.3) is time-independent, one can make the substitution $t \to \tau = t - t_0$ in (18.1). This corresponds to the substitution $t_0 \to \tau_0 = 0$. The Smoluchowski equation (18.1) reads then

$$\partial_\tau p(\boldsymbol{r}, \tau|\boldsymbol{r}_0, 0) = \mathscr{L}(\boldsymbol{r}) p(\boldsymbol{r}, \tau|\boldsymbol{r}_0, 0), \tag{18.9}$$

the solution of which is $p(\boldsymbol{r}, t - t_0|\boldsymbol{r}_0, 0)$, i.e., the solution of (18.1) for $p(\boldsymbol{r}, 0|\boldsymbol{r}_0, 0) = \delta(\boldsymbol{r} - \boldsymbol{r}_0)$. It follows

$$p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0) = p(\boldsymbol{r}, t - t_0|\boldsymbol{r}_0, 0), \tag{18.10}$$

Chapman-Kolmogorov Equation

The solution $p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0)$ of the Smoluchowski equation corresponds to the initial condition (18.7). The solution $p(\boldsymbol{r}, t)$ for an initial condition

$$p(\boldsymbol{r}, t_0) = f(\boldsymbol{r}), \tag{18.11}$$

can be expressed

$$p(\boldsymbol{r}, t) = \int_\Omega d\boldsymbol{r}_0 \, p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0) f(\boldsymbol{r}_0). \tag{18.12}$$

as can be readily verified. In fact, taking the time derivative yields

$$\partial_t p(\boldsymbol{r}, t) = \int_\Omega d\boldsymbol{r}_0 \, \partial_t p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0) f(\boldsymbol{r}_0) = \mathscr{L}(\boldsymbol{r}) \int_\Omega d\boldsymbol{r}_0 \, p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0) f(\boldsymbol{r}_0) = \mathscr{L}(\boldsymbol{r}) p(\boldsymbol{r}, t). \tag{18.13}$$

Furthermore, we note using (18.7)

$$p(\boldsymbol{r}, t_0) = \int_\Omega d\boldsymbol{r}_0 \, \delta(\boldsymbol{r} - \boldsymbol{r}_0) f(\boldsymbol{r}_0) = f(\boldsymbol{r}). \tag{18.14}$$

One can apply identity (18.12) to express $p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0)$ in terms of the probalities $p(\boldsymbol{r}, t|\boldsymbol{r}_1, t_1)$ and $p(\boldsymbol{r}_1, t_1|\boldsymbol{r}_0, t_0)$

$$p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0) \int_\Omega d\boldsymbol{r}_1 \, p(\boldsymbol{r}, t|\boldsymbol{r}_1, t_1) p(\boldsymbol{r}_1, t_1|\boldsymbol{r}_0, t_0). \tag{18.15}$$

This latter identity is referred to as the *Chapman-Kolmogorov equation*. Both (18.12) and (18.15) state that knowledge of the distribution at a single instance t, i.e., $t = t_0$ or $t = t_1$, allows one to predict the distributions at all later times. The Chapman-Kolmogorov equation reflects the Markov property of the stochastic process assumed in the derivation of the Smoluchowski equation.

We like to state finally the Chapman-Kolmogorov equation (18.15) for the special case $t_1 = t - \tau$. Employing identity (18.10) one obtains

$$p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0) \int_\Omega d\boldsymbol{r}_1 \, p(\boldsymbol{r}, \tau|\boldsymbol{r}_1, 0) p(\boldsymbol{r}_1, t - \tau|\boldsymbol{r}_0, t_0). \tag{18.16}$$

Taking the time derivative yields, using (18.1),

$$\partial_t p(\boldsymbol{r}, t|\boldsymbol{r}_0, t_0) \int_\Omega d\boldsymbol{r}_1 \, p(\boldsymbol{r}, \tau|\boldsymbol{r}_1, 0) \mathscr{L}(\boldsymbol{r}_1) p(\boldsymbol{r}_1, t - \tau|\boldsymbol{r}_0, t_0). \tag{18.17}$$

The Adjoint Smoluchowski Operator

We want to determine now the operator $\mathscr{L}^\dagger$ in (18.8). For this purpose we prove the following identity [14]

$$\int_\Omega d\boldsymbol{r} \, g(\boldsymbol{r}) \mathscr{L}(\boldsymbol{r}) h(\boldsymbol{r}) = \int_\Omega d\boldsymbol{r} \, h(\boldsymbol{r}) \mathscr{L}^\dagger(\boldsymbol{r}) g(\boldsymbol{r}) + \int_{\partial\Omega} d\boldsymbol{a} \cdot \boldsymbol{P}(g, h), \tag{18.18}$$

$$\mathscr{L}(\boldsymbol{r}) = \boldsymbol{\nabla} \cdot D\boldsymbol{\nabla} - \beta \boldsymbol{\nabla} \cdot D\boldsymbol{F}(\boldsymbol{r}), \tag{18.19}$$

$$\mathscr{L}^\dagger(\boldsymbol{r}) = \boldsymbol{\nabla} \cdot D\boldsymbol{\nabla} - \beta D\boldsymbol{F}(\boldsymbol{r}) \cdot \boldsymbol{\nabla}, \tag{18.20}$$

$$\boldsymbol{P}(g, h) = g(\boldsymbol{r}) D\boldsymbol{\nabla} h(\boldsymbol{r}) - h(\boldsymbol{r}) D\boldsymbol{\nabla} g(\boldsymbol{r}) - \beta D\boldsymbol{F}(\boldsymbol{r}) g(\boldsymbol{r}) h(\boldsymbol{r}). \tag{18.21}$$

The operator $\mathscr{L}^\dagger$ is called the *adjoint* to the operator $\mathscr{L}$, and $\boldsymbol{P}(g,h)$ is called the *concomitant* of $\mathscr{L}$.

To prove (18.18)-(18.21) we note, using $\boldsymbol{\nabla}\cdot w(\boldsymbol{r})\boldsymbol{q}(\boldsymbol{r}) = \boldsymbol{q}(\boldsymbol{r})\cdot\boldsymbol{\nabla}w(\boldsymbol{r}) + w(\boldsymbol{r})\boldsymbol{\nabla}\cdot\boldsymbol{q}(\boldsymbol{r})$

$$\begin{aligned}\boldsymbol{\nabla}\cdot(gD\boldsymbol{\nabla}h - hD\boldsymbol{\nabla}g) &= ((\boldsymbol{\nabla}g))D((\boldsymbol{\nabla}h)) + g\boldsymbol{\nabla}\cdot D\boldsymbol{\nabla}h - ((\boldsymbol{\nabla}h))D((\boldsymbol{\nabla}g)) - h\boldsymbol{\nabla}\cdot D\boldsymbol{\nabla}g\\ &= g\boldsymbol{\nabla}\cdot D\boldsymbol{\nabla}h - h\boldsymbol{\nabla}\cdot D\boldsymbol{\nabla}g,\end{aligned} \tag{18.22}$$

where the double brackets $((\ldots))$ limit the scope of the differential operators. Furthermore

$$\boldsymbol{\nabla}\cdot D\boldsymbol{F}gh = g\boldsymbol{\nabla}\cdot D\boldsymbol{F}h + hD\boldsymbol{F}\cdot\boldsymbol{\nabla}g. \tag{18.23}$$

Eqs. (18.22) and (18.23) can be combined, using (18.19)-(18.21),

$$g\mathscr{L}h = h\mathscr{L}^\dagger g + \boldsymbol{\nabla}\cdot\boldsymbol{P}(g,h), \tag{18.24}$$

from which follows (18.18).

In case

$$\boldsymbol{P}(g,h) = 0, \quad \text{for } \boldsymbol{r}\in\partial\Omega, \tag{18.25}$$

which implies a condition on the functions g and h, (18.18) corresponds to the identity

$$\langle g|\mathscr{L}(\boldsymbol{r})h\rangle_\Omega = \langle\mathscr{L}^\dagger g|h\rangle_\Omega, \tag{18.26}$$

a property which is the conventional definition of a pair of adjoint operators. We like to determine now which conditions g and h must obey for (18.25) to be true.

We assume that h obeys one of the three conditions (18.4)-(18.6) and try to determine if conditions for g on $\partial\Omega$ can be found such that (18.25) and, hence, (18.26) hold. For this purpose we write (18.25) using (18.21)

$$g(\boldsymbol{r})D[\boldsymbol{\nabla} - \beta\boldsymbol{F}(\boldsymbol{r})]h(\boldsymbol{r}) - h(\boldsymbol{r})D\boldsymbol{\nabla}g(\boldsymbol{r}) = 0, \quad \boldsymbol{r}\in\partial\Omega. \tag{18.27}$$

Then we need

$$(i') \qquad \hat{\boldsymbol{a}}(\boldsymbol{r})\cdot D\boldsymbol{\nabla}g(\boldsymbol{r}) = 0, \qquad \boldsymbol{r}\in\partial\Omega \tag{18.28}$$

$$(ii') \qquad g(\boldsymbol{r}) = 0, \qquad \boldsymbol{r}\in\partial\Omega \tag{18.29}$$

$$(iii') \qquad wg(\boldsymbol{r}) - \hat{\boldsymbol{a}}(\boldsymbol{r})\cdot D\boldsymbol{\nabla}g(\boldsymbol{r}) = 0, \qquad \boldsymbol{r}\in\partial\Omega \tag{18.30}$$

Derivation of the Adjoint Smoluchowski Equation

The Chapman-Kolmogorov equation in the form (18.17) allows one to derive the adjoint Smoluchowski equation (18.8). For this purpose we replace $\mathscr{L}(\boldsymbol{r}_0)$ in (18.17) by the adjoint operator using (18.26)

$$\begin{aligned}\partial_t p(\boldsymbol{r},t|\boldsymbol{r}_0,t_0) &= \int_\Omega d\boldsymbol{r}_1\, p(\boldsymbol{r},\tau|\boldsymbol{r}_1,0)\mathscr{L}(\boldsymbol{r}_1)p(\boldsymbol{r}_1,t-\tau|\boldsymbol{r}_0,t_0)\\ &= \int_\Omega d\boldsymbol{r}_1\, p(\boldsymbol{r}_1,t-\tau|\boldsymbol{r}_0,t_0)\mathscr{L}^\dagger(\boldsymbol{r}_1)p(\boldsymbol{r},\tau|\boldsymbol{r}_1,0).\end{aligned} \tag{18.31}$$

Taking the limit $\tau\to t-t_0$ yields, with $p(\boldsymbol{r}_1,t-\tau|\boldsymbol{r}_0,t_0)\to\delta(\boldsymbol{r}_1-\boldsymbol{r}_0)$,

$$\partial_t p(\boldsymbol{r},t|\boldsymbol{r}_0,t_0) = \mathscr{L}^\dagger(\boldsymbol{r}_0)p(\boldsymbol{r},t-t_0|\boldsymbol{r}_0,0), \tag{18.32}$$

i.e., the backward Smoluchowski equation (18.8).

The spatial boundary conditions need to be specified for $\boldsymbol{r}_0\in\partial\Omega$ and they are summarized in Eqs. (18.28)-(18.30).

We note finally that $\mathscr{L}^\dagger(\boldsymbol{r})$, given by (18.20), in case that the force $\boldsymbol{F}(\boldsymbol{r})$ is related to a potential, i.e., $\boldsymbol{F}(\boldsymbol{r}) = -\boldsymbol{\nabla}U(\boldsymbol{r})$, can be written

$$\mathscr{L}^\dagger(\boldsymbol{r}) = e^{\beta U(\boldsymbol{r})}\boldsymbol{\nabla}\cdot De^{-\beta U(\boldsymbol{r})}\boldsymbol{\nabla}. \tag{18.33}$$

This corresponds to expression (18.3) for $\mathscr{L}$.

B. Correlation Functions

Often an experimentalist prepares a system in an initial distribution $B(\boldsymbol{r})p_0(\boldsymbol{r})$ at a time t_0 and probes the spatial distribution of the system with sensitivity $A(\boldsymbol{r})$ at any time $t > t_0$. The observable is then the so called *correlation function*

$$C_{A(\boldsymbol{r}),B(\boldsymbol{r})}(t) = \int_\Omega d\boldsymbol{r} \int_\Omega d\boldsymbol{r}_0\, A(\boldsymbol{r})p(\boldsymbol{r},t|\boldsymbol{r}_0,t_0)p_0(\boldsymbol{r}_0)B(\boldsymbol{r}_0), \tag{18.34}$$

where $p(\boldsymbol{r},t|\boldsymbol{r}_0,t_0)$ obeys the backward Smoluchowski equation (18.8) with the initial condition

$$p(\boldsymbol{r},t_0|\boldsymbol{r}_0,t_0) = \delta(\boldsymbol{r}-\boldsymbol{r}_0). \tag{18.35}$$

and the adjoint spatial boundary conditions (18.28)-(18.28).

We like to provide a three examples of correlation functions. A trivial example arises in the case of $A(\boldsymbol{r}) = \delta(\boldsymbol{r}-\boldsymbol{r}')$ and $B(\boldsymbol{r}) = \delta(\boldsymbol{r}-\boldsymbol{r}'')/p_0(\boldsymbol{r})$ which yields

$$C_{A,B}(t) = p(\boldsymbol{r}',t|\boldsymbol{r}'',t_0). \tag{18.36}$$

In the case one can only observe the total number of particles, i.e. $A(\boldsymbol{r}) = 1$; and for the special case $B(\boldsymbol{r}) = \delta(\boldsymbol{r}-\boldsymbol{r}'')/p_0(\boldsymbol{r})$, the correlation function is equal to the total particle number, customarily written

$$C_{1,B}(t) = N_\Omega(t|\boldsymbol{r}',t_0) = \int_\Omega d\boldsymbol{r}\, p(\boldsymbol{r},t|\boldsymbol{r}',t_0). \tag{18.37}$$

The third correlation function we want to introduce is

$$C_{e^{i\boldsymbol{k}\cdot\boldsymbol{r}},e^{i\boldsymbol{k}\cdot\boldsymbol{r}_0}}(t) = G(\boldsymbol{k},t) = \int_\Omega d\boldsymbol{r} \int_\Omega d\boldsymbol{r}_0\, e^{i\boldsymbol{k}\cdot(\boldsymbol{r}-\boldsymbol{r}_0)}p(\boldsymbol{r},t|\boldsymbol{r}_0,0)p_0(\boldsymbol{r}_0). \tag{18.38}$$

In order to evaluate a correlation function $C_{A(\boldsymbol{r}),B(\boldsymbol{r})}(t)$ one can determine first the quantity

$$C_{A(\boldsymbol{r})}(t|\boldsymbol{r}_0) = \int_\Omega d\boldsymbol{r}\, A(\boldsymbol{r})p(\boldsymbol{r},t|\boldsymbol{r}_0,t_0), \tag{18.39}$$

and evaluate then

$$C_{A(\boldsymbol{r}),B(\boldsymbol{r})}(t) = \int_\Omega d\boldsymbol{r}_0\, C_{A(\boldsymbol{r})}(t|\boldsymbol{r}_0)p_0(\boldsymbol{r}_0)B(\boldsymbol{r}_0). \tag{18.40}$$

$C_{A(\boldsymbol{r})}(t|\boldsymbol{r}_0)$ can be obtained by carrying out the integral in (18.39) over the backward Smoluchowski equation (18.8). One obtains

$$\partial_t C_{A(\boldsymbol{r})}(t|\boldsymbol{r}_0) = \mathscr{L}^\dagger(\boldsymbol{r}_0)C_{A(\boldsymbol{r})}(t|\boldsymbol{r}_0), \tag{18.41}$$

with the initial condition

$$C_{A(\boldsymbol{r})}(t_0|\boldsymbol{r}_0) = A(\boldsymbol{r}_0), \tag{18.42}$$

and the appropriate boundary condition selected from (18.28)-(18.28).

XIX. BROWNIAN DYNAMICS VS THERMAL EQUILIBRIUM OF A MANY BODY SYSTEM

We want to put on parallel grounds the Smoluchowski equation for a stochastic process in d spatial dimensions in a force field $\boldsymbol{F}(\boldsymbol{r}) = -\boldsymbol{\nabla}U(\boldsymbol{r})$ (17.13), or equivalently the Fokker-Planck equation (15.31), with $\boldsymbol{B}[\boldsymbol{r}(t),t] = \mathbb{1}\sigma(\boldsymbol{r})/\gamma(\boldsymbol{r}) = \mathbb{1}\sqrt{2D(\boldsymbol{r})}$ and $\mathbb{1}$ the d dimensional identity matrix, $\boldsymbol{A}[\boldsymbol{r}(t),t] = \boldsymbol{F}(\boldsymbol{r})/\gamma$ and taking care of the fluctuation-dissipation theorem (17.10), $\boldsymbol{\nabla}D(\boldsymbol{r}) = \boldsymbol{F}(\boldsymbol{r})[\gamma^{-1}(\boldsymbol{r}) - D(\boldsymbol{r})\beta]$,

$$\partial_t p(\boldsymbol{r},t|\boldsymbol{r}_0,t_0) = \boldsymbol{\nabla}\cdot D\left(\boldsymbol{\nabla} - \beta\boldsymbol{F}(\boldsymbol{r})\right)p(\boldsymbol{r},t|\boldsymbol{r}_0,t_0) \tag{19.1}$$

$$= \left(-\sum_{i=1}^d \partial_i A_i + \frac{1}{2}\sum_{i,j=1}^d \partial_i\partial_j\left[\boldsymbol{B}\cdot\boldsymbol{B}^T\right]_{ij}\right)p(\boldsymbol{r},t|\boldsymbol{r}_0,t_0)., \tag{19.2}$$

$$p(\boldsymbol{r},t_0|\boldsymbol{r}_0,t_0) = \delta(\boldsymbol{r}-\boldsymbol{r}_0), \tag{19.3}$$

with the Bloch equation for the thermal density $\rho(R,\beta|R_0,0)$ matrix of a many body system in thermal equilibrium at an absolute temperature $T=1/k_B\beta$, where k_B is Boltzmann constant and $R=(\boldsymbol{r}_1,\boldsymbol{r}_2,\ldots,\boldsymbol{r}_N)$ are the positions of the N bodies in the volume Ω of the d dimensional space, namely,

$$\partial_\beta\rho(R,\beta|R_0,0)=\mathscr{H}\rho(R,\beta|R_0,0), \tag{19.4}$$

$$=\left(-\sum_{i=1}^{dN}\partial_i A_i+\frac{1}{2}\sum_{i,j=1}^{dN}\partial_i\partial_j\left[\boldsymbol{B}\cdot\boldsymbol{B}^T\right]_{ij}\right)\rho(R,\beta|R_0,0), \tag{19.5}$$

$$\rho(R,0|R_0,0)=\delta(R-R_0), \tag{19.6}$$

where

$$\mathscr{H}=\lambda\boldsymbol{\nabla}_R^2+U(R) \tag{19.7}$$

is the Hamiltonian of the quantum many bodies of mass m with $\lambda=\hbar^2/2m$ interacting with a potential $U(R)$ [8]. So that in the corresponding Fokker-Planck equation (19.5) we require

$$\boldsymbol{\nabla}_R\boldsymbol{A}[R(t),t])=U(R), \tag{19.8}$$

$$\boldsymbol{B}[R(t),t]=\mathbb{1}\sqrt{\lambda}., \tag{19.9}$$

where $\mathbb{1}$ is the identity matrix of dimension dN with d the spatial dimension.

In 1827 the botanist Robert Brown examined under his microscope pollen of plants suspended in water. He observed that the pollen, about $10^{-6}m$ in diameter, performed stochastic motions of the order of $10^{-5}m$. Even though Brown could not explain the cause of this motion any continuous stochastic process like that of a pollen is now referred to as *Brownian dynamics*.

In this section we introduce a numerical technique that generates a solution of the Fokker-Planck equation (19.1) by simulating an ensemble of stochastic processes. Due to these stochastic processes one calls this numerical method Brownian dynamics as well.

We provide two derivations of the numerical method of Brownian dynamics. In the first subsection we transform the Fokker-Planck equation (19.2) into a multi-dimensional integral which ca be evaluated using the Monte Carlo integration method introduced in Sections X and XI G. We show in the second subsection that this Monte Carlo integration is equivalent to simulating an ensemble of stochastic processes. The equivalence is shown by deriving the Brownian dynamics method a second time starting with the stochastic differential equation (15.25)

$$d\boldsymbol{r}(t)=\{\boldsymbol{A}[\boldsymbol{r}(t),t]+\boldsymbol{B}[\boldsymbol{r}(t),t]\cdot\boldsymbol{\xi}(t)\}\,dt, \tag{19.10}$$

with $d\boldsymbol{\omega}=\boldsymbol{\xi}\,dt$.

A. Path Integral Solution

Discretization is the basis of many numerical procedures. This also holds true for Brownian dynamics. The object of discretization is the continuous time axis. A computer can not represent a continuous line or function. The infinitely many points would simply not fit into a finite digital computer nor could they be processed in a finite amount of time. Hence, one needs to approximate a continuous dimension with a finite set of points. This approximation is called discretization. We solve the Fokker-Planck equation numerically by breaking the time axes parameterized by t into discrete points labeled by $t_0,t_1,t_2,\ldots$. These time points do not need to be equally spaced, but they usually are to simplify the calculations.

Let $p(\boldsymbol{r},t|\boldsymbol{r}_0,t_0)$ be the, yet unknown, solution of (19.2) for the initial condition $\boldsymbol{r}(t_0)=\boldsymbol{r}_0$. One finds a solution $p(\boldsymbol{r},t|\boldsymbol{r}_0,t_0)$ on a discrete sequence of times $t_0,t_1,t_2,\ldots$ by constructing transition probabilities

$$p(\boldsymbol{r}_{i+1},t_{i+1}|\boldsymbol{r}_i,t_i)$$

from one discrete time point t_i to the next. With these transition probabilities one can reassemble the complete solution $p(\boldsymbol{r},t|\boldsymbol{r}_0,t_0)$.

We need to first disassemble the time evolution of $p(\boldsymbol{r},t|\boldsymbol{r}_0,t_0)$ into many small time increments. For this purpose we proceed as follows. To obtain a solution $p(\boldsymbol{r}_1,t_1|\boldsymbol{r}_0,t_0)$ with an initial starting position at $\boldsymbol{r}(t_0)=\boldsymbol{r}_0$ one can first solve for $p(\boldsymbol{r}_1,t_1|\boldsymbol{r}_0,t_0)$ which describes the solution for an intermediate state at some time t_1 prior to t_2 and after t_0. The probability distribution at time t_1 may then be taken as the initial condition for a second solution of (19.2) reaching from time t_1 to time t_2. This second solution can be assembled due to the linearity of (19.2) using

$p(\boldsymbol{r}_2,t_2|\boldsymbol{r}_1,t_1)$ for the initial condition $\boldsymbol{r}(t_1)=\boldsymbol{r}_1$. Summing $p(\boldsymbol{r}_2,t_2|\boldsymbol{r}_1,t_1)$ over all possible initial positions x_1 in the domain Ω weighted with the initial probability determined through $p(\boldsymbol{r}_1,t_1|\boldsymbol{r}_0,t_0)$ and the Chapman-Kolmogorow equation (18.15)

$$p(\boldsymbol{r}_2,t_2|\boldsymbol{r}_0,t_0)=\int_\Omega d\boldsymbol{r}_1\, p(\boldsymbol{r}_2,t_2|\boldsymbol{r}_1,t_1)p(\boldsymbol{r}_1,t_1|\boldsymbol{r}_0,t_0), \tag{19.11}$$

one obtains $p(\boldsymbol{r}_2,t_2|\boldsymbol{r}_0,t_0)$. These steps may be repeated. Doing so $(M-1)$-times one obtains

$$p(\boldsymbol{r}_M,t_M|\boldsymbol{r}_0,t_0)=\int\cdots\int_\Omega\left(\prod_{i=1}^{M-1}d\boldsymbol{r}_i\,p(\boldsymbol{r}_{i+1},t_{i+1}|\boldsymbol{r}_i,t_i)\right)p(\boldsymbol{r}_1,t_1|\boldsymbol{r}_0,t_0). \tag{19.12}$$

The procedure above has divided now the time evolution of $p(\boldsymbol{r}_M,t_M|\boldsymbol{r}_0,t_0)$ into M steps over time intervals $[t_{i+1},t_i]$ where t_i, $i=1,2,\ldots,M-1$ denotes the intermediate times. In order to evaluate $p(\boldsymbol{r}_M,t_M|\boldsymbol{r}_0,t_0)$ we need to determine the transition probabilities $p(\boldsymbol{r}_{i+1},t_{i+1}|\boldsymbol{r}_i,t_i)$. The respective algorithm can exploit the possibility that one can choose the time intervals $[t_{i+1},t_i]$ very short such that certain approximations can be invoked without undue errors. In fact, for equally spaced time points t_i the length of each time segment is $\Delta t=t_{i+1}-t_i=(t-t_0)/M$. One can choose M always large enough that the time period Δt is short enough to justify the approximations introduced below or Trotter formula [15] used in Ref. [8] to treat Bloch equation (19.4) of the many body system.

Let $\Delta\boldsymbol{r}=\boldsymbol{r}_{i+1}-\boldsymbol{r}_i$ be the typical distance that a particle governed by the probability distribution $p(\boldsymbol{r},t_0+\Delta t|\boldsymbol{r}_0,t_0)$ may cover due to drift and diffusion within a time period Δt, i.e.

$$\Delta\boldsymbol{r}\sim\sqrt{\langle\langle\boldsymbol{r}^2\rangle\rangle}\sim\sqrt{\sum_j\boldsymbol{B}_{jj}^2(\boldsymbol{r}_0,t_0)\Delta t}. \tag{19.13}$$

The approximation introduced assumes that $\boldsymbol{A}(\boldsymbol{r},t)$ and $\boldsymbol{B}(\boldsymbol{r},t)$ are constant for each time period $[t_{i+1},t_i]$ and spatially independent in each range $[\boldsymbol{r}_i+\Delta\boldsymbol{r},\boldsymbol{r}_i-\Delta\boldsymbol{r}]$. One replaces than the functions $p(\boldsymbol{r}_{i+1},t_{i+1}|\boldsymbol{r}_i,t_i)$ in (19.12) by solutions of the Fokker-Planck equation (19.2) with constant coefficients $\boldsymbol{A}(\boldsymbol{r},t)$ and $\boldsymbol{B}(\boldsymbol{r},t)$. In case of boundary conditions at $\boldsymbol{r}\to\pm\infty$ the resulting expression is, according to Eqs. (17.18) and (17.24)

$$p(\boldsymbol{r}_{i+1},t_{i+1}|\boldsymbol{r}_i,t_i)=\frac{1}{\sqrt{2\pi\det{}^2\{\boldsymbol{B}(\boldsymbol{r}_i,t_i)\}\Delta t}}e^{\sum_j(\boldsymbol{r}_{i+1,j}-\boldsymbol{r}_{i,j}-\boldsymbol{A}_j(\boldsymbol{r}_i,t_i)\Delta t)^2/2\boldsymbol{B}_{jj}^2(\boldsymbol{r}_i,t_i)\Delta t}, \tag{19.14}$$

where $\boldsymbol{r}_{i,j}$ is the j-component of $\boldsymbol{r}_i$. Thus, we have solved the Fokker-Planck equation (19.2) up to an dM-dimensional integral. This integral needs to be evaluated numerically for which purpose one applies the Monte Carlo integration method of Section X.

B. Ito Calculus and Brownian Dynamics

From Eq. (19.10) over a small ime period Δt follows that

$$\boldsymbol{r}(t+\Delta t)-\boldsymbol{r}(t)=\boldsymbol{A}[\boldsymbol{r}(t),t]\Delta t+\boldsymbol{B}[\boldsymbol{r}(t),t]\cdot\Delta\boldsymbol{\omega}, \tag{19.15}$$

with $\Delta\boldsymbol{\omega}$ being a random variable with the Gaussian probability distribution (13.16) for $D=1/2$. Such a random variable can be generated with a normalized Gaussian random vector $\boldsymbol{g}$ and the mapping $\Delta\boldsymbol{\omega}=\sqrt{\Delta t}\boldsymbol{g}$ according to Eq. (13.21) and Section XI A,

$$\boldsymbol{r}(t+\Delta t)=\boldsymbol{A}[\boldsymbol{r}(t),t]\Delta t+\boldsymbol{B}[\boldsymbol{r}(t),t]\cdot\boldsymbol{g}\sqrt{\Delta t}+\boldsymbol{r}(t)\equiv\tilde{\boldsymbol{r}}(\boldsymbol{r}(t),t). \tag{19.16}$$

Then starting from $\boldsymbol{r}_0=\boldsymbol{r}(t_0)$ we generate a sample path

$$\boldsymbol{z}_{k_0,k_1,\ldots,k_M}\equiv\tilde{\boldsymbol{r}}_{k_L}(\ldots\tilde{\boldsymbol{r}}_{k_1}(\tilde{\boldsymbol{r}}_{k_0}(\boldsymbol{r}_0,t_0),t_1)\ldots,t_M), \tag{19.17}$$

and, given a sensitivity function $q(\boldsymbol{r})$, calculate

$$\begin{aligned}q(t|\boldsymbol{r}_0,t_0)&=\int_\Omega d\boldsymbol{r}\,q(\boldsymbol{r})p(\boldsymbol{r},t|\boldsymbol{r}_0,t_0)\\&=\frac{1}{L^M}\sum_{k_0,k_1,\ldots,k_M=1}^{L}q(\boldsymbol{z}_{k_0,k_1,\ldots,k_M}).\end{aligned} \tag{19.18}$$

Appendix A: Stirling approximation and asymptotic binomial distribution

In order to prove Gauß asymptotic approximation (13.32) of the binomial distribution we need Stirling formula

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \left(1 + O\left(\frac{1}{n}\right)\right). \tag{A1}$$

We set forth to prove Eq. (13.32), i.e.,

$$\sqrt{\frac{n}{2}} 2^{-n} \binom{n}{\frac{n}{2} + x\sqrt{\frac{n}{2}}} = \frac{1}{\sqrt{\pi}} e^{-x^2} \left(1 + O\left(\frac{1}{n}\right)\right). \tag{A2}$$

Applying the natural logarithm on both sides of this equation we obtain

$$\ln\left[\sqrt{\frac{n}{2}} 2^{-n} \binom{n}{\frac{n}{2} + x\sqrt{\frac{n}{2}}}\right] = -\frac{1}{2}\ln\pi - x^2 + O\left(\frac{1}{n}\right). \tag{A3}$$

We will prove this equation by transforming the left hand side step by step. First, we utilize the formula $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ for binomial coefficients.

$$\begin{aligned}
&\ln\left[\sqrt{\frac{n}{2}} 2^{-n} \binom{n}{\frac{n}{2} + x\sqrt{\frac{n}{2}}}\right] \\
&= \frac{1}{2}(\ln n - \ln 2) - n\ln 2 + \ln\left(\frac{n!}{\left(\frac{n}{2} + x\sqrt{\frac{n}{2}}\right)!\left(\frac{n}{2} - x\sqrt{\frac{n}{2}}\right)!}\right) \\
&= \frac{1}{2}\ln n - \left(n + \frac{1}{2}\right)\ln 2 + \ln n! - \ln\left[\left(\frac{n}{2} + x\sqrt{\frac{n}{2}}\right)!\right] - \ln\left[\left(\frac{n}{2} - x\sqrt{\frac{n}{2}}\right)!\right].
\end{aligned} \tag{A4}$$

Applying Stirling formula (A1) we derive furthermore

$$\begin{aligned}
&= \frac{1}{2}\ln n - \left(n + \frac{1}{2}\right)\ln 2 + \frac{1}{2}\ln(2\pi) + \left(n + \frac{1}{2}\right)\ln n - n + \\
&\quad -\frac{1}{2}\ln(2\pi) - \left(\frac{n}{2} + x\sqrt{\frac{n}{2}} + \frac{1}{2}\right)\ln\left(\frac{n}{2} + x\sqrt{\frac{n}{2}}\right) + \frac{n}{2} + x\sqrt{\frac{n}{2}} + \\
&\quad -\frac{1}{2}\ln(2\pi) - \left(\frac{n}{2} - x\sqrt{\frac{n}{2}} + \frac{1}{2}\right)\ln\left(\frac{n}{2} - x\sqrt{\frac{n}{2}}\right) + \frac{n}{2} - x\sqrt{\frac{n}{2}} \\
&= \frac{1}{2}\ln n - \left(n + \frac{1}{2}\right)\ln 2 - \frac{1}{2}\ln(2\pi) + \left(n + \frac{1}{2}\right)\ln n + \\
&\quad - \left(\frac{n}{2} + x\sqrt{\frac{n}{2}} + \frac{1}{2}\right)\ln\left(\frac{n}{2} + x\sqrt{\frac{n}{2}}\right) + \\
&\quad - \left(\frac{n}{2} - x\sqrt{\frac{n}{2}} + \frac{1}{2}\right)\ln\left(\frac{n}{2} - x\sqrt{\frac{n}{2}}\right) \\
&= (n+1)\ln n - \left(n + \frac{1}{2}\right)\ln 2 - \frac{1}{2}\ln(2\pi) + \\
&\quad - \left(\frac{n}{2} + x\sqrt{\frac{n}{2}} + \frac{1}{2}\right)\left(\ln\frac{n}{2} + \ln\left(1 + x\sqrt{\frac{2}{n}}\right)\right) + \\
&\quad - \left(\frac{n}{2} - x\sqrt{\frac{n}{2}} + \frac{1}{2}\right)\left(\ln\frac{n}{2} + \ln\left(1 - x\sqrt{\frac{2}{n}}\right)\right).
\end{aligned}$$

Performing a Taylor expansion of $\ln(1 \pm z)$ with respect to z we obtain

$$\begin{aligned}
&= (n+1)(\ln n - \ln 2) - \frac{1}{2}\ln\pi + \\
&\quad - \left(\frac{n}{2} + x\sqrt{\frac{n}{2}} + \frac{1}{2}\right)\left(\ln\frac{n}{2} + x\sqrt{\frac{2}{n}} - \frac{x^2}{n} + \frac{x^3}{3}\sqrt{\frac{8}{n^3}} + O\left(\frac{1}{n^2}\right)\right) + \\
&\quad - \left(\frac{n}{2} - x\sqrt{\frac{n}{2}} + \frac{1}{2}\right)\left(\ln\frac{n}{2} - x\sqrt{\frac{2}{n}} - \frac{x^2}{n} - \frac{x^3}{3}\sqrt{\frac{8}{n^3}} + O\left(\frac{1}{n^2}\right)\right),
\end{aligned}$$

and expanding the products up to order $O(1/n)$ we acquire the result, the right hand side of Eq. (A3),

$$\begin{aligned}
&= (n+1)(\ln n - \ln 2) - \frac{1}{2}\ln\pi + \\
&\quad -\frac{n}{2}\ln\frac{n}{2} - x\sqrt{\frac{n}{2}} + \frac{x^2}{2} - \frac{x^3}{3}\sqrt{\frac{2}{n}} - x\sqrt{\frac{n}{2}}\ln\frac{n}{2} - x^2 + \frac{x^3}{\sqrt{2n}} - \frac{1}{2}\ln\frac{n}{2} - \frac{x}{\sqrt{2n}} + O\left(\frac{1}{n}\right) + \\
&\quad -\frac{n}{2}\ln\frac{n}{2} + x\sqrt{\frac{n}{2}} + \frac{x^2}{2} + \frac{x^3}{3}\sqrt{\frac{2}{n}} + x\sqrt{\frac{n}{2}}\ln\frac{n}{2} - x^2 - \frac{x^3}{\sqrt{2n}} - \frac{1}{2}\ln\frac{n}{2} + \frac{x}{\sqrt{2n}} + O\left(\frac{1}{n}\right) \\
&= (n+1)\ln\frac{n}{2} - \frac{1}{2}\ln\pi - n\ln\frac{n}{2} - x^2 - \ln\frac{n}{2} + O\left(\frac{1}{n}\right) \\
&= -\frac{1}{2}\ln\pi - x^2 + O\left(\frac{1}{n}\right).
\end{aligned}$$

Appendix B: The Brownian Oscillator

The one-dimensional Smoluchowski equation, in case that a stationary flux-free equilibrium state $p_0(x)$ exists, can be written in the form

$$\partial_t p(x,t) = \frac{1}{\beta\gamma}\partial_x p_0(x)\partial_x \frac{1}{p_0(x)} p(x,t), \tag{B1}$$

where we employed $D = \sigma^2/2\gamma^2$ [c.f. (16.12)], the fluctuation-dissipation theorem in the form $\sigma^2 = 2\gamma/\beta$ [c.f. (17.11)], the Onsager form of the Smoluchowski equation (17.14) applied to one dimension, and $p_0(x) = Nexp[-U(x)]$. This form of the Smoluchowski equation demonstrates most clearly that it describes a stochastic system characterized through an equilibrium state $p_0(x)$ and a single constant governing the relaxation, the friction constant. The equation also assumes that the underlying stochastic process (17.3)

$$\beta\gamma\dot{x} = \partial_x \ln[p_0(x)] + \sqrt{2\beta\gamma}\xi(t), \tag{B2}$$

alters the variable x continuously and not in discrete jumps. One is inclined to invoke the Smoluchowski equation for the description of stochastic processes for which the equilibrium distribution is known. Underlying such description is the assumption that the process is governed by a single effective friction constant γ. For the sake of simplicity and in view of the typical situation that detailed information regarding the relaxation process is lacking, the Smoluchowski equation serves on well with an approximate description.

The most prevalent distribution encountered is the Gaussian distribution

$$p_0(x) = \frac{1}{\sqrt{\pi\Sigma}} e^{-(x-\langle x\rangle)^2/\Sigma}. \tag{B3}$$

The reason is the fact that many properties x are actually based on contributions from many constituents. This leads one to consider then in most cases the Smoluchowski equation for an effective quadratic potential

$$U_{\text{eff}}(x) = \frac{(x-\langle x\rangle)^2}{\beta\Sigma}. \tag{B4}$$

Due to the central limit theorem the Smoluchowski equation for a Brownian oscillator has a special significance. Accordingly, we will study the behavior of the Brownian oscillator in detail.

We consider again the diffusion in the harmonic potential already treated in Section XVII C using transformation to time-dependent coordinates. We introduce now the operator

$$\mathscr{O} = \partial_\xi^2 + 2\partial_\xi\xi, \tag{B5}$$

restricting ourselves to the functions space $\mathscr{S} = \{h(\xi)|\lim_{\xi\to\pm\infty}\xi^n h(\xi) = 0 \quad \forall n \in \mathbb{N}\}$. We define the eigenfunctions

$$\mathscr{O} f_n(\xi) = -\lambda_n f_n(\xi). \tag{B6}$$

The solution of this equation is well known

$$f_n(\xi) = c_n e^{-\xi^2} H_n(\xi), \tag{B7}$$

where H_n are the Hermite polynomials and c_n are normalization constants. The negative eigenvalues are

$$\lambda_n = 2n, \tag{B8}$$

Then we can introduce a bi-orthonormal system

$$f_n(\xi) = \frac{1}{2^n n!\sqrt{\pi}} e^{-\xi^2} H_n(\xi), \tag{B9}$$

$$g_n(\xi) = H_n(\xi), \tag{B10}$$

$$\langle g_n | f_m \rangle = \delta_{nm}. \tag{B11}$$

where the functions g_n are the eigenfunctions of

$$\mathscr{O}^\dagger = \partial_\xi^2 - 2\partial_\xi \xi, \tag{B12}$$

such that

$$\mathscr{O}^\dagger g_n(\xi) = -\lambda_n g_n(\xi), \tag{B13}$$

so that $\langle g_n | \mathscr{O} f_m \rangle = \langle \mathscr{O}^\dagger g_n | f_m \rangle$ for all n, m.

The eigenfunctions f_n form a complete basis for all functions in $\mathscr{S}$. Hence, we can expand

$$q(\xi, \tau | \xi_0, \tau_0) = \sum_{n=0}^{\infty} \alpha_n(\tau) f_n(\xi). \tag{B14}$$

Inserting this into the Smoluchowski equation (17.46) and using the bi-orthonormality property (B11) one finds

$$\dot{\alpha}_m(\tau) = -\lambda_m \alpha_m(\tau). \tag{B15}$$

The general solution of of this differential equation is

$$\alpha_m(\tau) = \beta_m e^{-\lambda_m \tau}. \tag{B16}$$

From the initial condition $q(\xi, \tau_0 | \xi_0, \tau_0) = \delta(\xi - \xi_0)$ then follows

$$\beta_m = e^{\lambda_m \tau_0} g_m(\tau_0). \tag{B17}$$

Hence, we finally obtain

$$q(\xi, \tau | \xi_0, \tau_0) = \sum_{n=0}^{\infty} e^{-\lambda_n (\tau - \tau_0)} g_n(\xi_0) f_n(\xi). \tag{B18}$$

This expression can be simplified using the generating function of a product of two Hermite polynomials

$$\sum_{n=0}^{\infty} \frac{s^n}{2^n n!\sqrt{\pi}} e^{-y^2} H_n(y) e^{-y_0^2} H_n(y_0) = \frac{1}{\sqrt{\pi(1-s^2)}} \exp\left[-\frac{1}{2}(y^2 + y_0^2)\frac{1+s^2}{1-s^2} + 2yy_0 \frac{s}{1-s^2} \right]. \tag{B19}$$

The, using $s = e^{-2(\tau - \tau_0)}$ we find

$$\begin{aligned} q(\xi, \tau | \xi_0, \tau_0) &= \frac{1}{\sqrt{\pi(1-s^2)}} \exp\left[-\frac{1}{2}(\xi^2 + \xi_0^2)\frac{1+s^2}{1-s^2} + 2\xi\xi_0 \frac{s}{1-s^2} - \frac{1}{2}\xi^2 + \frac{1}{2}\xi_0^2 \right] \\ &= \frac{1}{\sqrt{\pi(1-s^2)}} \exp\left[-\xi^2 \frac{1}{1-s^2} - \xi_0^2 \frac{s^2}{1-s^2} + 2\xi\xi_0 \frac{1}{1-s^2} \right] \\ &= \frac{1}{\sqrt{\pi(1-s^2)}} \exp\left[-\frac{(\xi - \xi_0 s)^2}{1-s^2} \right]. \end{aligned} \tag{B20}$$

One can readily recognize that this result agrees with the solution (17.62) derived in Section XVII C.

Let us now consider the solution for an initial distribution $h(\xi_0)$. The corresponding distribution $q(\xi, \tau)$ is ($\tau_0 = 0$)

$$q(\xi, \tau) = \int_{-\infty}^{\infty} d\xi_0 \, \frac{1}{\sqrt{\pi\,(1 - e^{-4\tau})}} e^{-\left(\xi - \xi_0 e^{-2\tau}\right)^2 / \left(1 - e^{-4\tau}\right)} h(\xi_0). \tag{B21}$$

It is interesting to consider the asymptotic behavior of this solution. For $\tau \to \infty$ the distribution $q(\xi, \tau)$ relaxes to

$$q(\xi) = \frac{1}{\sqrt{\pi}} e^{-\xi^2} \int_{-\infty}^{\infty} d\xi_0 \, h(\xi_0) = \frac{1}{\sqrt{\pi}} e^{-\xi^2}. \tag{B22}$$

One then recognizes that, in general, the relaxation of an initial distribution $h(\xi_0)$ to the Boltzmann distribution involves numerous relaxation times, given by the eigenvalues λ_n, even though the original Smoluchowski equation (B1) contains only a single rate constant, the friction coefficient γ.

Appendix C: Some mathematical properties

1. Error function

We have

$$\mathrm{erf}(z) \equiv \frac{2}{\sqrt{\pi}} \int_0^z dx\, e^{-x^2}, \tag{C1}$$

$$\mathrm{erfc}(z) \equiv 1 - \mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \int_z^\infty dx\, e^{-x^2}, \tag{C2}$$

$$\tag{C3}$$

so that

$$\int dx\, \mathrm{erf}(z) = z\mathrm{erf}(z) + \frac{1}{\sqrt{\pi}} e^{-z^2}, \tag{C4}$$

$$\frac{1}{2}\mathrm{erfc}(z) = 1 - \frac{1}{2}\mathrm{erfc}(-z), \tag{C5}$$

$$\mathrm{erfc}(0) = 1, \tag{C6}$$

$$\partial_z \mathrm{erfc}(z) = -\frac{2}{\sqrt{\pi}} e^{-z^2}, \tag{C7}$$

$$\lim_{z\to\infty} \mathrm{erfc}(z) = \frac{1}{\sqrt{\pi} z} e^{-z^2}, \tag{C8}$$

$$\sqrt{\pi} z\, \mathrm{erfc}(z) = 1 + O\left(\frac{1}{z^2}\right). \tag{C9}$$

2. Approximations

Stirling formula

$$\ln n! = \frac{1}{2}\ln(2\pi) + \left(n + \frac{1}{2}\right)\ln n - n + O\left(\frac{1}{n}\right). \tag{C10}$$

3. Fourier Transforms

We have

$$f(t) \to \hat{f}(\omega) = \int_{-\infty}^{\infty} dt\, e^{-it\omega} f(t), \tag{C11}$$

$$f(t) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi}\, e^{it\omega} \hat{f}(\omega) \to \hat{f}(\omega), \tag{C12}$$

$$e^{iat} \to 2\pi\delta(\omega - a), \tag{C13}$$

$$e^{-at^2} \to \sqrt{\frac{\pi}{a}} e^{-\omega^2/4a} \quad a > 0. \tag{C14}$$

4. Laplace Transforms

We have

$$f(t) \to \tilde{f}(s) = \int_0^\infty dt\, e^{-ts} f(t), \tag{C15}$$

$$\partial_t f(t) + f(0) \to s\tilde{f}(s), \tag{C16}$$

$$\frac{1}{\sqrt{\pi t}} e^{-a^2/4t} \to \frac{1}{\sqrt{s}} e^{-|a|\sqrt{s}}, \tag{C17}$$

$$\frac{1}{\sqrt{\pi t}} e^{-a^2/4t} - b e^{ab+b^2 t} \mathrm{erfc}\left(\frac{a}{\sqrt{4t}+b\sqrt{t}}\right) \to \frac{1}{b+\sqrt{s}} e^{-a\sqrt{s}}, \tag{C18}$$

5. Gaussian integrals

We have

$$I_n(\alpha) = (-1)^n \int_{-\infty}^{\infty} dx\, x^{2n} e^{-\alpha x^2/2} = \partial_\alpha^n I_0(\alpha) = \frac{\Gamma(n+1/2)}{\alpha^{n+1/2}} \quad n = 0, 1, 2, \ldots, \tag{C19}$$

$$I_0(\alpha) = \sqrt{\frac{\pi}{\alpha}}. \tag{C20}$$

AUTHOR DECLARATIONS

Conflicts of interest

None declared.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Funding

None declared.

[1] L. Martinelli and L. Baldini, *Misure ed analisi dei dati, introduzione al Laboratorio di Fisica* (Edizioni ETS, Pisa, Italy, 2004).
[2] M. H. Kalos and P. A. Whitlock, *Monte Carlo Methods* (John Wiley & Sons Inc., New York, 1986).
[3] Klaus Schulten and Theoretical Biophysics Group, *Lectures in Nonequilibrium Statistical Mechanics*, Tech. Rep. (Department of Physics and Beckman Institute, University of Illinois, 405 N. Mathews Street, Urbana, IL 61801 USA, 1998).
[4] P. Révész, *The Laws of Large Numbers* (Academic Press, New York, 1968).
[5] J. Marcienkiewicz, Sur une propriété de la loi de Gauß, Math. Z. **44**, 612 (1939).
[6] D. Williams, *Probability with Martingales* (Cambridge University Press, 1991) §18.1.
[7] G. E. P. Box and M. E. Muller, A note on the generation of random normal deviates, Annals of Mathematical Statistics **29**, 610 (1958).
[8] D. M. Ceperley, Path integrals in the theory of condensed Helium, Rev. Mod. Phys. **67**, 279 (1995).
[9] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. M. Teller, and E. Teller, Equation of state calculations by fast computing machines, J. Chem. Phys. **1087**, 21 (1953).
[10] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1 (John Wiley & Sons, New York, 1950) p. 324.

[11] W. W. Wood and F. R. Parker, Monte Carlo equation of state of molecules interacting with the Lennard-Jones potential. I. Supercritical isotherm at about twice the critical temperature, J. Chem. Phys. **27**, 720 (1957).
[12] J. P. Valleau and S. G. Whittington, A Guide to Monte Carlo for statistical mechanics: 1. Highways in Statistical Mechanics, Part A: Equilibrium Techniques, in *Modern Theoretical Chemistry*, Vol. 5, edited by B. Berne (Plenum, New York, 1976) Chap. 4.
[13] C. W. Gardiner, *Handbook of Stochastic Methods* (Springer, New York, 1983).
[14] M. V. Smoluchowski, Drei Vortrage uber Diffusion, Brownsche Bewegung und Koagulation von Kolloidteilchen, Phys. Z. **17**, 557 (1916).
[15] H. F. Trotter, On the Product of Semi-Groups of Operators, Proc. Am. Math. Soc. **10**, 545 (1959).