Chapter 1

Stochastics

Literature

1. F.J. VESELY, *Computational Physics*, WUV-Universitätsverlag, Wien (1993), ISBN 3-85114-122-9.

2. J.M. HAMMERSLEY AND D.C. HANDSCOMB, *Monte Carlo Methods*, Chapman and Hall, London (1964).

3. G.S. FISHMAN, *Monte Carlo*, Springer Series on Operations Research, Springer Verlag, New York (1996), ISBN 0.387-94527-X.

4. W.H. PRESS, B.P. FLANNERY, S.A. TEUKOLSKY, and W.T. VETTERLING, *Numerical Recipes in C*, Cambridge University Press, Cambridge (1988), ISBN 0-521-35465-X.

1.1 Introduction

It is certainly necessary to reflect on the terms *statistics* and *stochastics*. It is the purpose of *statistics* to extract essential parameters, like the mean value, the variance, etc., from data which contain a random component. In contrast, stochastic means **accidental** or **random**. Thus, statistics extracts regular patterns out of randomness while stochastics employs randomness deliberately, for instance when possible states of a model system are to be 'tested'.

One can consider random effects in model systems using different methods. For instance, in BOLTZMANN's kinetic gas theory as well as in the theory of diffusion one doesn't know about random forces. These are built in statistically by making certain assumptions on the probability density in phase space - e.g. detailed balance, molecular chaos, etc.

On the other hand, it is possible to describe the diffusion of a single particle using stochastics, *i.e.:* using an equation of motion which explicitly includes randomness. Such an equation was postulated by the mathematician P. LANGEWIN 1907 in an attempt to describe Brown's motion: the *stochastic differential equation*

$$m\frac{d^2}{dt^2}\mathbf{r}(t) = -\gamma\mathbf{v}(t) + \mathbf{S}(t), \tag{1.1}$$

or

$$\dot{\mathbf{v}}(t) = -\beta\mathbf{v}(t) + \mathbf{S}'(t). \tag{1.2}$$

Here, $-\gamma\mathbf{v}(t)$ describes the retardation due to the viscosity of the fluid the particle experiences in his motion within the fluid. $\mathbf{S}(t)$ is the *stochastic force* which has its origin in random collisions of the particle with molecules the fluid consists of. [Eq. (1.1) was derived rigorously in the 1970s.] It is the main purpose of this chapter to develop methods which will allow to treat such a stochastic force as a random variable and which will, ultimately, yield solutions of Eq. (1.1).

Of course, a solution of Eq. (1.1) is only possible using numerics and, obviously, the stochastic force has to be *sampled*. The mean value of each component of $\mathbf{S}(t)$ is zero and its variance is closely related to the viscosity γ and the temperature of the fluid. This results in the methods of **stochastic dynamics** which, obviously, are based on the sampling of some random variable. It is simply pragmatism when we identify stochastic procedures as methods which involve the use of a *random number generator.*

GIBB's formulation of statistical mechanics uses a big number of inaccurate copies of one particular system instead of studying just this one system. All these copies are members of an *ensemble* differ only little from each other and the variance is well known. In this case randomness appears only indirectly via statistical assumptions. Nevertheless, there is a stochastic method to calculate ensemble mean values: the **Monte Carlo Method**. This method generates the ensemble step by step by generating a series of random copies of the basic model system by sampling each time type and size of the variation of the previous copy. One speaks of a **random walk** in GIBB's phase space. Such a procedure depends on the use of a biased dice and it is an important task to develop random number generators which provide a sequence of numbers which possesses certain required statistical properties.

Physics has changed considerably since the development of quantum mechanics which is naturally of stochastic nature. Today, one is convinced that the fundamental processes elementary particles experience are of stochastic nature and do not depend on unknown, hidden parameters. Thus, stochastics is of elementary importance in modern physics.

1.2 Random Sampling Methods

The first component of a Monte Carlo calculation is the numerical sampling of random variables with specified probability distribution functions (PDFs). In this section we describe different techniques to generate random values of a variable x distributed in the interval $[x_{\min}, x_{\max}]$ according to a given PDF $p(x)$. We concentrate on the simple case of single-variable distributions, since random sampling from multivariate distributions can always be reduced to single variable sampling.

1.2.1 Random Number Generator

In general, random sampling algorithms are based on the use of random numbers r uniformly distributed in the interval $[0, 1]$. They are described by the PDF:

$$u(r) = \begin{cases} 1 & r \in [0,1] \\ 0 & \text{otherwise.} \end{cases} \tag{1.3}$$

This results in the cumulative distribution function

$$U(r) = \int_0^r dr'\, u(r') = \begin{cases} 0 & r < 0 \\ r & 0 \leq r < 1 \\ 1 & r \geq 1. \end{cases} \tag{1.4}$$

Among the simplest algorithms for pseudo random numbers are the *linear congruential generators* (LCG) which are based on the recursion

$$X_{i+1} = (aX_i + c) \bmod m,$$

where the integers a, c, and m are constants. These generators can further be classified into mixed $(c > 0)$ and multiplicative $(c = 0)$ types, usually denoted by LCG(a, c, m) and MLCG(a, m), respectively. An LCG generates a sequence of pseudo random integers X_1, X_2,... between 0 and $m - 1$; for a MLCG the lower bound is 1. Each X_i is then scaled into the interval $[0, 1)$. If the multiplier a is a primitive root modulo m and m is prime, the period of this generator is $m - 1$

A commonly used choice of parameters for the MLCG is the 'miracle number' $a = 16\,807 = 7^5$ and $m = 2^{31} - 1$. This yields the GGL generator (sometimes denoted by CONG or RAND0) which is given by:

$$R_n = \left(7^5\, R_{n-1}\right) \bmod \left(2^{31} - 1\right), \quad r_n = \frac{R_n}{2^{31} - 1}. \tag{1.5}$$

It produces a sequence of random numbers r_n uniformly distributed in $[0, 1)$ from a given *seed* $R_0 < 2^{31} - 1$ and odd. The generator (1.5) is known to have good random properties. However, the sequence is periodic, with a period of the order of $2^{31} - 2 \approx 2.15 \times 10^9$. This not large enough for more complex simulations to prevent re-initialization in a single run. Nevertheless, for most purposes, the random number generator provided by the operating system is good enough.

Another known problem of this generator is that D-dimensional vectors $(x_1, x_2, \ldots, x_D)$, $(x_{D+1}, x_{D+2}, \ldots, x_{2D})$, … formed by consecutive normalized random numbers $x_i \in [0, 1)$ lie on a relatively small number of parallel hyperplanes.

1.2.2 Inverse Transformation Method

The cumulative distribution function (CDF) $G(x)$ of the PDF $g(x)$

$$G(x) = \int_{x_{\min}}^{x} dx'\, g(x'), \quad g(x) \geq 0, \quad \int_{x_{\min}}^{x_{\max}} dx'\, g(x') = 1, \tag{1.6}$$

is a non-decreasing function of x and, therefore, it has an inverse function $G^{-1}(\xi)$. The transformation $\xi = G(x)$ defines a new random variable that takes values in the interval $[0, 1]$. Owing to the correspondence between the x and ξ values, the PDF of ξ, $g_\xi(\xi)$, and that of x, $g(x)$, are related by $g_\xi(\xi) d\xi = g(x) dx$. Hence

$$g_\xi(\xi) = g(x) \left(\frac{d\xi}{dx}\right)^{-1} = g(x) \left(\frac{dG(x)}{dx}\right)^{-1} \stackrel{(1.6)}{=} 1, \tag{1.7}$$

that is, ξ is uniformly distributed in the interval $[0, 1]$, *i.e.:* $g_\xi(\xi) = u(r)$ according to Eq. (1.3).

Now, it is clear that if ξ is a random number, the variable defined by $x = G^{-1}(\xi)$ is randomly distributed in the interval $[x_{\min}, x_{\max}]$ with PDF $g(x)$. This provides a practical method of generating random values of x using a generator of random numbers uniformly distributed in $[0, 1]$. The randomness of x is guaranteed by that of ξ. Note that x is the (unique) root of equation

$$\xi = \int_{\mathbf{x}_{\min}}^{x} dx'\, g(x'), \tag{1.8}$$

which will be referred to as the *sampling equation* of the variable x. This procedure for random sampling is known as the *inverse transform method*; it is particularly adequate for PDFs $g(x)$ given by simple analytical expressions such that the sampling equation (1.8) can be solved analytically.

Consider, for instance the uniform distribution in the interval $[a, b]$:

$$g(x) = u_{a,b}(x) = \frac{1}{b-a}.$$

The sampling equation (1.8) then reads

$$\xi = \frac{x-a}{b-a},$$

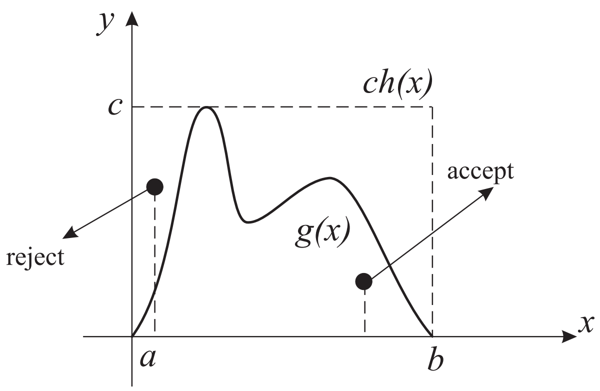


Figure 1.1: The rejection method.

which leads to the sampling formula:

$$x = a + \xi(b - a), \quad \xi \in [0, 1].$$

As another example consider the exponential distribution

$$g(s) = \frac{1}{\lambda} \exp\left\{-\frac{s}{\lambda}\right\}, \quad s > 0,$$

of the free path s of a particle between interaction events. The parameter λ represents the mean free path. In this case, the sampling equation (1.8) is easily solved to give the sampling formula

$$s = -\lambda \ln(1 - \xi) = -\lambda \ln \xi. \tag{1.9}$$

The last step follows from the fact that $1 - \xi$ is also a random number distributed equally in $(0, 1]$.

1.2.3 Rejection Methods

The inverse transform method for random sampling is based on a one-to-one correspondence between x and ξ values, which is expressed in terms of a single-valued function. There is another kind of sampling method, due to VON NEUMANN, that consists of sampling a random variable from a certain distribution [different from $g(x)$] and subjecting it to a random test to determine whether it will be accepted for use or rejected. These rejection methods lead to very general techniques for sampling from any PDF.

We want to generate random numbers following a PDF $g(x)$

$$0 \le g(x) \le 1, \quad \int_a^b dx\, g(x) = 1, \quad x \in [a, b].$$

We choose another PDF

$$0 \leq h(x) \geq 1, \qquad \int_a^b dx\, h(x) = 1,$$

which is rather easy to sample in such a way that

$$g(x) \leq c\, h(x), \quad \forall x \in [a, b],$$

with c a constant. Thus, $c\, h(x)$ is in the interval $[a, b]$ the *envelope* of the PDF $g(x)$. (See also Fig. 1.1.) Random numbers distributed according to the PDF $g(x)$ are generated according to the following procedure:

Algorithm 1 Rejection Method

begin:
Generate random variable x^T from $h(x)$
Generate random number r from $u(x)$, Eq. (1.3).
if $r > [g(x^T)/(ch(x^T))]$ **then**
  go to begin
else
  accept x^T
EXIT

Proof:

If $p(A|x^T\, \mathcal{B}) = [g(x^T)/(ch(x^T))]$ is the probability of accepting a value x^T we get:

$$\begin{aligned}
p(x|\mathcal{B}) &\propto p(x = x^T|\mathcal{B})p(A|x^T\mathcal{B}) \\
&\propto h(x^T)\frac{g(x^T)}{c\, h(x^T)} \\
&\propto g(x^T).
\end{aligned}$$

The accepted random numbers x^T follow indeed the PDF $g(x)$.

Probability of acceptance:

The probability of acceptance $P(A|\mathcal{B})$ is determined by:

$$\begin{aligned}
P(A|\mathcal{B}) &= \int dx^T\, p(Ax^T|\mathcal{B}) \\
&= \int dx^T\, p(A|x^T\mathcal{B})p(x^T|\mathcal{B}) \\
&= \int dx^T\, \frac{g(x^T)}{c\, h(x^T)} h(x^T) \\
&= \int dx^T \frac{g(x^T)}{c} \\
&= \frac{1}{c}.
\end{aligned}$$

Thus, the bigger c the worse the probability of acceptance becomes. If we apply rejection methods to d dimensions, we get:

$$P(A|\mathcal{B}) = \left(\frac{1}{c}\right)^d .$$

1.2.4 Probability Mixing

This method has its application in cases in which the PDF $f(x)$ is composed of a number of PDFs in an interval $[a, b]$:

$$f(x) = \sum_{i=1}^{N} f_i(x), \tag{1.10}$$

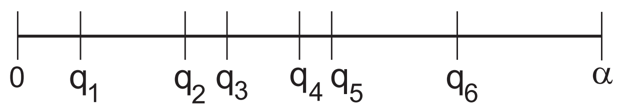
with

$$f_i(x) \geq 0, \qquad \int_{-\infty}^{\infty} dx\, f_i(x) = 1.$$

Let us define:

$$\begin{aligned} \alpha_i &= \int_a^b dx\, f_i(x), \\ q_i &= \sum_{\ell=1}^{i} \alpha_\ell, \\ \alpha &= \sum_{i=1}^{N} \alpha_i = \int_a^b dx\, f(x). \end{aligned}$$

Thus, the interval $[a, b]$ has been divided according to:



An equally distributed random number $r \in [a, b]$ can now be employed to determine the index i for which the condition

$$q_{i-1} < r < q_i$$

is fulfilled.

This procedure is quite plausible: The α_i give the area under the PDF $f_i(x)$ within the interval $[a, b]$. Thus α_i is an indicator for the importance of the PDF $f_i(x)$. This in turn gives us the probability that the random variable X is to be sampled using the PDF $f_i(x)$.

1.3 Stochastic Processes

Stochastic processes are time dependent processes which depend on randomness. Stochastic processes are used in classical thermodynamics to compensate, for instance, the fact that the exact motion of particles in a gas are unknown. Quantum processes, in contrast, are random by nature. A. EINSTEIN studied in a seminal work (1905) Brown's motion and thus started the investigation of stochastic processes in physics. N. WIENER proceeded 1922 with a systematic investigation of Brown's motion. Following N. WIENER's ideas A.N. KOLMOGOROV developed a general theory of stochastic processes. Elementary events are, in this picture, possible trajectories in phase space, *i.e.:* realizations of a process. The probability of the existence of a number of such trajectories is then a measure in the space spanned by the trajectories. Thus, the theory of stochastic processes develops quite naturally into a measure theory and integrals over subaggregates of a space spanned by functions. This was later on used particularly by FEYNMAN in his alternate access to quantum mechanics - the FEYNMAN-Integral or *path integral* in which averages over possible trajectories of quantum particles are calculated.

1.3.1 Time Series

We are interested in instances $t = 0, \Delta t, 2\Delta t, \ldots$, with $\Delta t > 0$. Each instance $t = n\Delta t$ is associated with a random variable X_n. x_n is the measured value of X_n at instance $t = n\Delta t$. An important issue in the analysis of time series is the detection of periodicities or the proof of aperiodicity. For this purpose the method of the *empirical autocorrelation coefficient* or the modern method of *spectral analysis* is employed.

The Empirical Autocorrelation Coefficient

We have a set of $N + 1$ data points (time series) $x_0, x_1, \ldots, x_N$ as the result of an experiment. We set the number $k = 1, 2, \ldots, N/2$ and investigate the two subsets

$$x_0,\, x_1,\, \ldots,\, x_{N-k}$$

and

$$x_k,\, x_{k+1},\, \ldots,\, x_N.$$

We then define the empirical autocorrelation coefficient

$$r_k = \frac{\sum\limits_{j=0}^{N-k} \left(x_j - \langle X\rangle\right)\left(y_j - \langle Y\rangle\right)}{\left[\sum\limits_{j=0}^{N-k} \left(x_j - \langle X\rangle\right)^2 \sum\limits_{j=0}^{N-k} \left(y_j - \langle Y\rangle\right)^2\right]^{1/2}}, \tag{1.11}$$

or more precisely, the k-th autocorrelation coefficient of our set of data points. (Please see also Appendix A.) Here we used $y_j = x_{j+k}$ and

$$\begin{aligned}\langle X\rangle &= \frac{1}{N-k+1}\sum_{j=0}^{N-k} x_j,\\ \langle Y\rangle &= \frac{1}{N-k+1}\sum_{j=0}^{N-k} y_j.\end{aligned}$$

Obviously, $-1 \le r_k \le 1$ and increasing values of r_k indicate a growing correlation between data points found by a time shift of $k\Delta t$.

(i) Is, for instance, $|r_k|$ particularly large for $k = m, 2m, 3m$ in comparison to other results for $|r_k|$ then there is a strong indication of a period $T = 2m\Delta t$ in the data set.

(ii) Small values of $|r_k|$ are a strong indication that there is practically no correlation between data points. We have most likely an aperiodic behavior.

Spectral Analysis of Discrete Time Series

In a first step the notion *stationary time series* needs to be defined. For this purpose we introduce a series of random variables X_n on a probability field E as

$$X_n : E \to \mathbb{R}, \qquad n = 0,\, \pm 1,\, \pm 2,\, \ldots$$

with $0 < \text{std}(X_0) < \infty$. We interpret X_n as a random variable assigned to the instance $t = n\Delta t$ and the series is stationary only if

$$\langle X_n\rangle = \langle X_0\rangle \tag{1.12}$$

$$\operatorname{cov}(X_n, X_{n+k}) = \operatorname{cov}(X_0, X_k), \qquad \forall\, n, k = 0,\, \pm 1,\, \pm 2,\, \ldots \tag{1.13}$$

Here,

$$\operatorname{cov}(X, Y) = \langle (X - \langle X\rangle)\,(Y - \langle Y\rangle)\rangle\,,$$

and, in particular,

$$\operatorname{cov}(X, X) = \text{var}(X).$$

The autocorrelation coefficient r_k is then defined as

$$r_k = \frac{\operatorname{cov}(X_n, X_{n+k})}{\text{std}(X_n)\,\text{std}(X_{n+k})}, \qquad k = 1,\, 2,\, \ldots, \tag{1.14}$$

and, thus, r_k is the autocorrelation coefficient of the random variable X_n at instance $n\Delta t$ with the random variable X_{n+k} at instance $(n+k)\Delta t$. The fact that the series is stationary in time is expressed by the fact that the expectation value $\langle X_n\rangle$ and the standard deviation $\text{std}(X_n)$ no longer depends on the instance $t = n\Delta t$. Thus, the autocorrelation coefficient r_k depends no longer on the chosen instance but only on the time interval $k\Delta t$.

If we define $R(k) = \mathrm{cov}(X_0, X_k)$, with $R(-k) = R(k)$, $\forall\, k$, we get

$$\begin{aligned} r_k &= \frac{\mathrm{cov}(X_n, X_{n+k})}{\mathrm{std}(X_n)\,\mathrm{std}(X_{n+k})} = \frac{\mathrm{cov}(X_0, X_k)}{\mathrm{std}(X_0)\,\mathrm{std}(X_k)} \\ &= [\mathrm{std}(X_k) = \mathrm{std}(X_0),\ \text{stationary}] \\ &= \frac{R(k)}{R(0)}, \end{aligned}$$

with $R(0) = \mathrm{std}(X_0)\,\mathrm{std}(X_0) = \mathrm{var}(X_0) = \mathrm{cov}(X_0, X_0)$.

Theorem 1.1 *Let $\sum\limits_{k=0}^{\infty} |R(k)| < \infty$. Then, the function*

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} R(k) e^{-ik\lambda} \tag{1.15}$$

is continuous and we have

$$R(k) = \int_{-\pi}^{\pi} d\lambda\, f(\lambda) e^{ik\lambda}, \qquad \forall\, k = 0,\, \pm 1,\, \pm 2,\, \ldots \tag{1.16}$$

This is the so-called spectral theorem.

In particular, we find for the variance:

$$\mathrm{var}(X_n) = R(0) = \int_{-\pi}^{\pi} d\lambda\, f(\lambda), \qquad \forall\, n.$$

The 2π-periodic function $f(\lambda)$ is the *spectral density* of a stationary time series. It contains information on the standard deviation $\mathrm{std}(X)$ and on the autocorrelation coefficient r_k.

Example 1: Let

$$R(\pm n) \neq 0, \quad R(k) = 0, \quad \forall\, k \neq n$$

for a fixed number $n \geq 1, \in \mathbb{N}$. Then

$$f(\lambda) = \frac{R(n)}{2\pi} \left(e^{-in\lambda} + e^{in\lambda} \right) = \frac{R(n)}{\pi} \cos(n\lambda), \quad \lambda \in \mathbb{R}.$$

Example 2: Let

$$R(0) \neq 0, \quad R(k) = 0, \quad \forall\, k \neq 0,$$

then

$$f(\lambda) = \frac{R(0)}{2\pi}, \quad \lambda \in \mathbb{R}.$$

Such an uncorrelated time series is called *white noise*.

Statistics with Discrete Time Series

(X_n) is a discrete time series which is stationary. We want to give estimates for the most important quantities on the basis of a set of data

$$x_0,\, x_1,\, \ldots,\, x_{N-1}$$

under the assumption that N is large enough.

(i) Estimate of the expectation value $\langle X_n \rangle$ for all n:

$$\mu = \frac{1}{N} \sum_{i=0}^{N-1} x_i .$$

(ii) Estimate for $R(k)$:

$$R_N(k) = \frac{1}{N-k} \sum_{j=0}^{N-k-1} (x_j - \mu)\,(x_{j+k} - \mu)\,.$$

$R_N(k)/R_N(0)$ is an estimate for the autocorrelation coefficient r_k. $R_N(0)$ itself is an estimate for the variance var(X_n).

(iii) Estimate for the spectral density:

$$f(\lambda) = \frac{1}{2N\pi} \left| \sum_{n=0}^{N-1} x_n e^{-in\lambda} \right|^2 .$$

1.3.2 MARKOV Chains - The Stochastic Matrix

J. BERNOULLI developed a scheme for experiments which was based on the idea of independent experiments. 1906 A. MARKOV studied for the first time chains of dependent events. This was the first and most simple generalization of BERNOULLI's scheme to interdependent experiments.

Basics

(i) We have a series of discrete instances

$$t = 0,\, \Delta t,\, 2\Delta t,\, 3\Delta t,\, \ldots$$

(ii) At each of these instances the system can be in the states

$$Z_1,\, Z_2,\, \ldots,\, Z_k,$$

which span the state space $\mathcal{S}$.

(iii) We define the transition probability p_{ij} from state Z_i at instance $t = n\Delta t$ to state Z_j at instance $t = (n+1)\Delta t$. Obviously, p_{ij} is the probability for the system to be in the state Z_j at $t = (n+1)\Delta t$ **if** it was in the state Z_i at $t = n\Delta t$.

> *It is typical for a* MARKOV *chain, that the transition probabilities do not depend on a particular moment in time. We assume homogeneity in time.*

(iv) Transition probabilities are organized in the so-called *transition matrix*:

$$\boldsymbol{P} = \{p_{ij}\} = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ p_{k1} & p_{k2} & \dots & p_{kk} \end{pmatrix}. \tag{1.17}$$

$\boldsymbol{P}$ is a *stochastic matrix*, *i.e.*: all elements of $\boldsymbol{P}$ obey the inequality

$$0 \le p_{ij} \le 1,$$

and all row-sums of the $k \times k$ matrix are equal to one:

$$\sum_{j=1}^{k} p_{ij} = 1, \qquad i = 1, \dots, k.$$

The transition probabilities $p_{ij}^{(k)}$ from state Z_i at $t = n\Delta t$ to a state Z_j at $t = (n+k)\Delta t$ can also be organized in a matrix

$$\boldsymbol{P}^{(k)} = \left\{p_{ij}^{(k)}\right\},$$

which, again, is a stochastic matrix. Using this definition we find the CHAPMAN-KOLMOGOROV *equation*:

$$\boldsymbol{P}^{(k+m)} = \boldsymbol{P}^{(k)}\,\boldsymbol{P}^{(m)}, \qquad \forall\, k, m = 1, 2, \dots \tag{1.18}$$

It results immediately in:

$$\boldsymbol{P}^{(k)} = \boldsymbol{P}^{k}. \tag{1.19}$$

Ergodicity

A MARKOV chain is called *ergodic* if the limits

$$p_j = \lim_{n\to\infty} p_{ij}^{(n)}, \qquad i = 1, \dots, k \tag{1.20}$$

exist and are independent of i. Moreover,

$$p_j > 0, \quad \sum_{j=0}^{k} p_j = 1.$$

Thus, we get the significant result that the MARKOV chain completely 'forgets' its state at the beginning $t = 0$. p_j is interpreted as the probability of the system to reach the state Z_j after a long time independent of its starting point. Thus, ergodic MARKOV chains can reach a stable end state independently of their starting situation.

An example: Spread of a rumor

Z_1 and Z_2 are two versions of a report, namely, Z_1: Mr. X is going to resign, and Z_2: Mr. X is not going to resign. Thus we have the following transition matrix:

$$\boldsymbol{P} = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix},$$

to the effect that:

(i) Some person receives the report Z_1. It will then pass this report on as Z_2 with a probability p and as Z_1 with a probability $(1-p)$. Consequently, the report will be modified with a probability p.

(ii) In the same way, the report Z_2 will be modified with a probability q.

Realistically $0 < p < 1$ and $0 < q < 1$ and we find

$$\lim_{n\to\infty} \boldsymbol{P}^n = \frac{1}{p+q}\begin{pmatrix} q & p \\ q & p \end{pmatrix}$$

and, consequently

$$p_1 = \frac{q}{p+q}, \qquad p_2 = \frac{p}{p+q}.$$

Finally, we find under the assumption $p = q$

$$p_1 = p_2 = \frac{1}{2},$$

with the result that after some time the public will be, with a probability of $p_1 = 1/2$, of the opinion that Mr. X will resign. This is certainly no longer in any relation to the real intentions of Mr. X.

Recurrence

The system is in the state Z_i at the time $t = 0$. w_n is the probability for the system to return to the state Z_i at the time $t = n\Delta t$ for the first time. Thus,

$$w = \sum_{n=0}^{\infty} w_n \tag{1.21}$$

is the probability for the system to return into state Z_i within a finite time span.

Theorem 1.2 *The initial state Z_i is* recurrent *if the probability w according to Eq.* (1.21) *is equal to one. If this is not the case, the state Z_i is called* transient.

In terms of transition probabilities p_{ij} the initial state Z_i is recurrent if the condition

$$\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$$

is fulfilled. Thus, if the system starts from the initial state Z_i it will return an infinite number of times to this state with probability one. On the other hand, if the initial state Z_i is transient then there will be a finite time span after which the state Z_i will never be reached again.

1.4 Random Series

1.4.1 Basics

In section 1.2 we dealt with the generation of random numbers with the particular property $r_{n,n+1} = 0$, *i.e.:* the random number generated in step n should be uncorrelated to the random number generated in step $n+1$. In this section we will deal with the generation of a sequence of random numbers which obeys a given statistical correlation. Within this context we concentrate on *random series*, random processes on a discrete time scale. A random series is a MARKOV *process* if

$$P_k(X_k \mid X_{k-1} \dots X_1, \mathcal{B}) = P_2(X_k \mid X_{k-1}, \mathcal{B}). \tag{1.22}$$

This expresses that future values of a random variable X_k are statistically determined by present events and depend only on the event immediately preceding, *i.e.:* X_{k-1}. (This property is trivial for quantum systems.) The same holds for the PDFs:

$$p_k(x_k \mid x_{k-1} \dots x_1, \mathcal{B}) = p_2(x_k \mid x_{k-1}, \mathcal{B}). \tag{1.23}$$

This is also called the MARKOV-*property*. Finally, a MARKOV series, *i.e.:* a random series with MARKOV-property, in which a random variable X can only have N discrete values is called a MARKOV *chain* as has already been discussed in Sec. 1.3.2.

The conditional probability

$$p_{\alpha\beta} = P\left(X_n = x_\beta | X_{n-1} = x_\alpha, \mathcal{B}\right)$$

is the transition probability from state x_α to state x_β and the matrix $\{p_{\alpha\beta}\} = \boldsymbol{P}$ is, again, a stochastic matrix. We introduce, furthermore, the single state probability

$$p_\alpha = P(X = x_\alpha | \mathcal{B})$$

for the random variable X to have the value x_α. One can collect these single state probabilities into the vector $\mathbf{p} = (p_\alpha)$ and this vector together with the matrix $\boldsymbol{P}$ describes completely the properties of the chain.

Thus, we find, for instance, the condition for the reversibility of a MARKOV chain as

$$p_\alpha p_{\alpha\beta} = p_\beta p_{\beta\alpha}. \tag{1.24}$$

This is the condition of *detailed balance*: p_α is the probability of realization of a state x_α and $p_{\alpha\beta}$ is the probability to find in the next step the system to be in state x_β. Reversibility means that the identical states are equally probable in reversed sequence, as it is expressed in Eq. (1.24).

Theorem 1.3 *In a stable, reversible* MARKOV *chain which is characterized by* $\mathbf{p} = (p_\alpha)$ *and* $\boldsymbol{P} = \{p_{\alpha\beta}\}$ *any value* x_α *will be realized with the relative occurrence* p_α. *This is the* Central Theorem of Monte Carlo Simulation.

A MARKOV chain is said to be *irreducible* if from each state it is possible to get to each other state: that is, for each pair $x_\alpha, x_\beta \in \mathcal{S}$, there exists an $n \geq 0$ for which $p_{\alpha\beta}^{(n)} > 0$. (This is just another way of defining ergodicity.)

For each state x_α we define the *period of* x_α (denoted d_α) to be the greatest common divisor of the numbers $n > 0$ for which $p_{\alpha\alpha}^{(n)} > 0$. If $d_\alpha = 1$, the state x_α is called *aperiodic*. It can be shown that, in an irreducible chain, all states have the same period; so we can speak of the chain having the period d. Moreover, the state space $\mathcal{S}$ can then be partitioned into subsets $\mathcal{S}_1$, $\mathcal{S}_2$,…,$\mathcal{S}_d$ around which the chain moves cyclically, *i.e.:* $p_{\alpha\beta}^{(n)} = 0$ whenever $x_\alpha \in \mathcal{S}_i$, $x_\beta \in \mathcal{S}_j$ with $j - i \neq n \pmod d$.

We now come to a fundamental topic of MARKOV chains which is the convergence to equilibrium. A probability measure $\pi = \{\pi_\alpha\}_{x_\alpha \in \mathcal{S}}$ is called a *stationary distribution* (or *invariant distribution* or *equilibrium distribution*) of the MARKOV chain in case

$$\sum_\alpha \pi_\alpha p_{\beta\alpha} = \pi_\beta, \quad \forall\, \beta. \tag{1.25}$$

A stationary probability distribution need not exist; but if it does, then a lot more follows:

Theorem 1.4 *Let* $\boldsymbol{P}$ *be the transition probability matrix of an* irreducible MARKOV *chain of period* d. *If a stationary probability distribution* π *exists, then it is unique, and* $\pi_\alpha > 0$ *for all* $x_\alpha \in \mathcal{S}$. *Moreover,*

$$\lim_{n\to\infty} p_{\alpha\beta}^{(n)} = \begin{cases} d\pi_\beta & \textit{if } x_\alpha \in \mathcal{S}_i,\ x_\beta \in \mathcal{S}_j \textit{ with } j - i = r \pmod d, \\ 0 & \textit{if } x_\alpha \in \mathcal{S}_i,\ x_\beta \in \mathcal{S}_j \textit{ with } j - i \neq r \pmod d, \end{cases}$$

for all x_α, x_β. *In particular, if* $\boldsymbol{P}$ *is aperiodic, then*

$$\lim_{n\to\infty} p_{\alpha\beta}^{(n)} = \pi_\beta.$$

This theorem shows that the MARKOV chain converges as $t \to \infty$ to the equilibrium distribution π, irrespective of the initial distribution p_α.

1.4.2 Brown's Motion

The simplest way to generate a stable, Gaussian MARKOV process is the iterative solution of the stochastic differential equation (1.1). In one dimension it is of the form:

$$\frac{\partial x(t)}{\partial t} = -\beta x(t) + S(t). \tag{1.26}$$

With $x(0) = x_0$ we find the solution:

$$x(t) = x_0 e^{-\beta t} + \int_0^t dt'\, S(t') e^{-\beta(t-t')}. \tag{1.27}$$

We concentrate on the time slice Δt and calculate the next position of the particle $x(t_{n+1})$ at $t = t_{n+1} = t_n + \Delta t$ if the position of the particle at time $t = t_n$, $x(t_n)$, is known from Eq. (1.27):

$$x(t_{n+1}) = x(t_n) e^{-\beta \Delta t} + \int_0^{\Delta t} dt'\, S(t_n + t') e^{-\beta(\Delta t - t')}.$$

A solution of this problem is only possible if the properties of the stochastic force $S(t)$ are known. Let us assume, that $S(t)$ follows a Gaussian distribution around $\langle S \rangle = 0$. Thus, the physics of the stochastic force, and, therefore, of the Brown's motion is determined by a PDF. The solution of the problem will be feasible the moment a random series is available which corresponds to a Gaussian distribution.

In a one dimensional problem we study a particle which moves along, say, the x-axis. x gives the position of the particle at the time t. We define with $P(x \in J, t'|y, t, \mathcal{B})$ the conditional probability for finding the particle within the interval J of the x-axis at time t' **if** it was in position y on the x-axis at time t. Thus, $P(x \in J, t'|y, t, \mathcal{B})$ is a transition probability. In particular, we find for Brown's motion:

$$\begin{aligned} P(x \in J, t'|y, t, \mathcal{B}) &= \int_J dx\, p(x, t'; y, t), \\ p(x, t'; y, t) &= \frac{1}{\sqrt{2\pi(t'-t)}} e^{-(x-y)^2/2(t'-t)}, \quad \forall\, t' > t. \end{aligned}$$

Here, $p(x, t'; y, t)$ is a Gaussian PDF.

The solution of our problem is found by introducing

$$Z(t_n) = \int_0^{\Delta t} dt'\, S(t_n + t') e^{\beta(\Delta t - t')}.$$

which is due to the stochastic nature of $S(t)$ a random series which can be derived from $P(x \in J, t'|y, t, \mathcal{B})$. According to our assumption, $S(t)$ is Gaussian distributed around $\langle S \rangle = 0$ and consequently, $Z(t_n)$ is Gaussian distributed around the mean value 0. We find for finite time slices

$$\langle Z \rangle = 0, \qquad \mathrm{var}(Z) = \frac{A}{2\beta}\left(1 - e^{-2\beta\Delta t}\right), \tag{1.28}$$

with $\langle S(t)S(t') \rangle = A\delta(t - t')$. A, β, and Δt are known and determine the stochastic force, the viscosity of the fluid, and the time slice.

The random series $Z(t_n)$ is generated from a Gaussian PDF and results in a numerical solution of the stochastic differential equation to describe Brown's motion. $Z(t_n)$ can be generated using the following algorithm:

Algorithm 2 Brown's Motion in One Dimension

Start at point $x(0)$. [$x(0)$ can be sampled from a Gaussian PDF.]
$n = 0$
repeat:
  $n = n + 1$
  Sample $Z(n)$ from a Gaussian PDF.
  The next position is given by:
    $x(n+1) = x(n)e^{-\beta\Delta t} + Z(n)$
go to repeat

The Diffusion Equation

There are n particles in the fluid. During the time interval τ the x-coordinates of the particles increase by an amount Δ, with Δ being different for each particle. Δ can be positive or negative. There will be a certain 'law' for Δ: the number dn of particles with a displacement within Δ and $\Delta + d\Delta$ can be expressed by

$$dn = n\, f(\Delta) d\Delta.$$

As $f(\Delta)$ is a PDF we have

$$\int_{-\infty}^{\infty} d\Delta\, f(\Delta) = 1. \tag{1.29}$$

$f(\Delta)$ is a Gaussian distribution and, thus, is significantly different from zero only for small values of Δ. Moreover, $f(\Delta)$ is symmetric:

$$f(\Delta) = f(-\Delta).$$

We are now in a position to study how the coefficient of diffusion depends on $f(\Delta)$ if we assume the number of particles ν in the unit volume to depend only on x and t. Thus, $\nu = \nu(x,t)$. We then calculate the distribution of the particles at time $t+\tau$ from their distribution at time t. Using the definition of $f(\Delta)$ it is easy to determine the number of particles which can be found at time $t+\tau$ between two planes perpendicular to the x-axis at positions x and $x+dx$:

$$\nu(x,t+\tau)dx = dx \int_{-\infty}^{\infty} d\Delta\, f(\Delta)\nu(x+\Delta,t). \tag{1.30}$$

τ is small and we expand:

$$\nu(x,t+\tau) = \nu(x,t) + \tau\frac{\partial\nu(x,t)}{\partial t}.$$

We also expand $\nu(x+\Delta,t)$ in powers of Δ:

$$\nu(x+\Delta,t) = \nu(x,t) + \Delta\frac{\partial\nu(x,t)}{\partial x} + \frac{\Delta^2}{2}\frac{\partial^2\nu(x,t)}{\partial x^2} + \cdots$$

This is used in Eq. (1.30):

$$\begin{aligned}\nu(x,t) + \tau\frac{\partial\nu(x,t)}{\partial t} &= \nu(x,t)\underbrace{\int_{-\infty}^{\infty} d\Delta\, f(\Delta)}_{=1} + \frac{\partial\nu(x,t)}{\partial x}\underbrace{\int_{-\infty}^{\infty} d\Delta\, \Delta f(\Delta)}_{=0} \\ &\quad + \frac{\partial^2\nu(x,t)}{\partial x^2}\int_{-\infty}^{\infty} d\Delta\, \frac{\Delta^2}{2} f(\Delta) + \cdots \end{aligned}\tag{1.31}$$

As $f(\Delta)$ is symmetric the second, forth, ... terms vanish on the right hand side of Eq. (1.31). Thus, Eq. (1.31) transforms into the *diffusion equation*

$$\frac{\partial}{\partial t}\nu(x,t) = D\frac{\partial^2}{\partial x^2}\nu(x,t),$$

with the *coefficient of diffusion*

$$D = \frac{1}{\tau}\int_{-\infty}^{\infty} d\Delta\, \frac{\Delta^2}{2} f(\Delta).$$

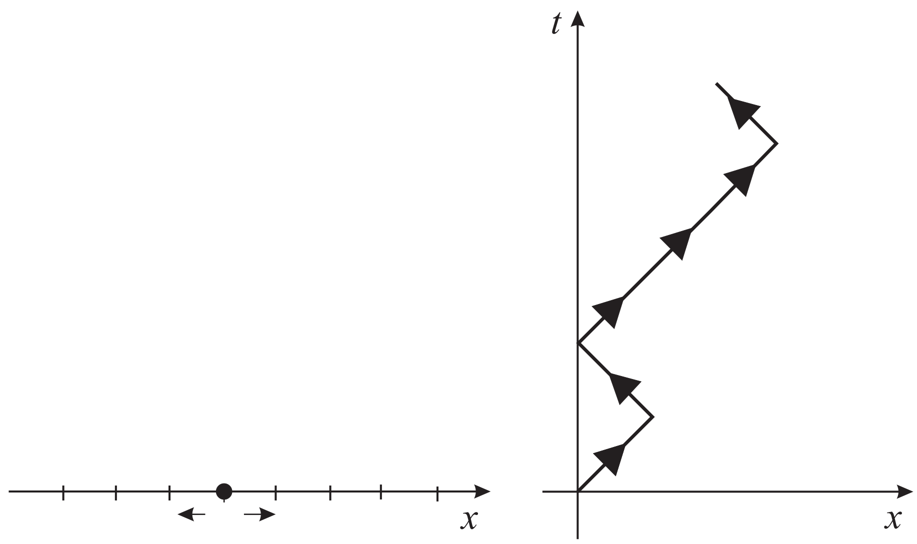


Figure 1.2: Classical random walk in one dimension.

Classical Random Walk in One Dimension

The x-axis is divided into equidistant lattice points

$$x = 0,\ \pm\Delta x,\ \pm 2\Delta x, \ldots$$

and we study the Brown's motion of the particle at discrete instances

$$t = 0,\ \Delta t,\ 2\Delta t,\ \ldots$$

At time $t = 0$ the particle is assumed to be at position $x = 0$. The particle is found at position $x = m\Delta x$ at some other instance $t = n\Delta t$. In the next time slice, $t = (n+1)\Delta t$, the particle will be found with probability $p = 1/2$ at the right hand neighbor position $x = (m+1)\Delta x$ and with the same probability at the left hand neighbor position $x = (m-1)\Delta x$ as is shown in Fig. 1.2.

Such a procedure is called *classical random walk*. It is also the most simple form of a WIENER-LÉVY process.

The WIENER-LÉVY **Process**

The parameter β in Eq. (1.2) is set equal to zero. This results in

$$\begin{aligned} x(n) &= \sum_{\nu=0}^{n-1} Z(\nu), \\ x(n+1) &= x(n) + Z(n), \end{aligned}$$

with

$$Z(n) = \int_0^{\Delta t} dt'\, S(t_n + t').$$

$Z(n)$ is a random variable sampled from a Gaussian distribution with $\langle Z \rangle = 0$ and $\mathrm{var}(Z) = A\Delta t$ according to Eq. (1.28). Obviously, x and Z are uncorrelated and we find

$$\begin{aligned}
\langle x(n) \rangle &= \sum_{\nu=0}^{n-1} \langle Z(\nu) \rangle = 0, \\
\mathrm{var}[x(n)] &= \sum_{\nu,\nu'} \underbrace{\langle Z(\nu) Z(\nu') \rangle}_{=\delta_{\nu\nu'} A\Delta t} = nA\Delta t.
\end{aligned}$$

Thus, the variance of x increases linearly with the number of steps. The random process is no longer stable. This process is Brown's random walk or an *unbiased random walk*.

Algorithm 3 Unbiased Random Walk in One Dimension

Define values A and Δt.
$n = 0$, $x(0) = 0$
repeat:
  $n = n + 1$
  Sample $Z(n)$ from Gaussian PDF, Eq. (1.28)
  $x(n+1) = x(n) + Z(n)$
  go to repeat

If we identify x with the Cartesian coordinate of a diffusing particle then $\mathrm{var}[x(n)]$ is the mean displacement of the particle after n time slices. The relation $A = 2D$ can then be used to correlate the coefficient of diffusion D to A.

The METROPOLIS Algorithm

The condition of detailed balance (1.24) does not uniquely determine the stochastic matrix $\boldsymbol{P}$. A given PDF $f(x_1, \ldots, x_N)$ can result in numerous transition matrices which obey detailed balance. Quite popular is the method discussed by N. METROPOLIS, the *asymmetric proposition*: All x_β within a certain neighborhood of x_α can be reached with the same *a priori* probability $\Pi_{\alpha\beta} = 1/Z$, with Z the number of possible x_β (including x_α). Thus, we have

$$p_{\alpha\beta} = \begin{cases} \Pi_{\alpha\beta} = 1/Z & p_\beta \geq p_\alpha \\ \Pi_{\alpha\beta} \frac{p_\beta}{p_\alpha}, & p_\beta < p_\alpha, \end{cases} \tag{1.32}$$

or in reverse:

$$p_{\beta\alpha} = \begin{cases} \Pi_{\beta\alpha} = 1/Z, & p_\alpha \geq p_\beta \\ \Pi_{\beta\alpha} \frac{p_\alpha}{p_\beta}, & p_\alpha < p_\beta. \end{cases} \tag{1.33}$$

It becomes immediately apparent that detailed balance is obeyed:

Algorithm 4 Biased Random Walk

Choose neighborhood Δx
Step n
$x(n) = x_\alpha$
sample random number r from PDF (1.3)
Generate x_β from the neighborhood of x_α:
  $x_\beta = x_\alpha + (r - 0.5)\Delta x$
This corresponds to a *a priori* probability of $\Pi_{\alpha\beta} = 1/Z$
if min$[1, p(x_\beta)/p(x_\alpha)] = 1$ **then**
  $x(n+1) = x_\beta$
else
  sample random number r from PDF (1.3)
  if $r < p(x_\beta)/p(x_\alpha)$ **then**
    $x(n+1) = x_\beta$
  else
    $x(n+1) = x_\alpha$.

(i) $p_\alpha > p_\beta$:

$$p_\alpha \frac{1}{Z} = p_\beta \frac{1}{Z}\frac{p_\alpha}{p_\beta} = p_\alpha \frac{1}{Z}$$

(ii) $p_\alpha < p_\beta$:

$$p_\alpha \frac{1}{Z}\frac{p_\beta}{p_\alpha} = p_\beta \frac{1}{Z}.$$

For $\alpha = \beta$ detailed balance is obeyed trivially.

Sometimes a symmetric proposition

$$p_{\alpha\beta} = \Pi_{\alpha\beta}\frac{p_\beta}{p_\alpha + p_\beta} \tag{1.34}$$

is used; it is also possible to apply *a priori* probabilities different from $1/Z$. The only requirement is that they are to be symmetric in α and β.

The central theorem of Monte Carlo Simulation together with METROPOLIS' asymmetric proposition are the key to another algorithm which can be used to generate random numbers according to a given probability vector **p**. This will, again, be a random walk but in contrast to the WIENER-LÉVY process a *biased random walk*. Such random numbers will not be uncorrelated, the autocorrelation coefficient will always be different from zero.

$\mathbf{p} = \{p_\alpha \,|\, \alpha = 1, 2, \ldots\}$ is the vector of probabilities for events $X = \{x_\alpha \,|\, \alpha = 1, 2, \ldots\}$ and we want to generate a random series $\{x(n)\}$ in which the relative accumulation of the event $X(n) = x_\alpha$ asymptotically reaches p_α. This results in algorithm 4. It is good practice to choose the neighborhood Δx around x_α in such a way that the rate of acceptance for new states x_β is about 50%.

One of the most important applications of the biased random walk is the simulation of a system of N particles. In this case $x(n)$ is interpreted as a vector of $3N$ coordinates and the probability $p(x_\alpha)$ is defined by the thermodynamic probability, *i.e.:* a BOLTZMANN PDF. The central theorem of Monte Carlo simulation establishes that under the condition of a correct random walk all possible microscopic arrangements of those N particles will be 'visited' according to their relative accumulation.