Chapter 9
Random Numbers and Monte-Carlo Methods

Many-body problems often involve the calculation of integrals of very high dimension which can not be treated by standard methods. For the calculation of thermodynamic averages Monte Carlo methods [94–97] *are very useful which sample the integration volume at randomly chosen points. In this chapter we discuss algorithms for the generation of pseudo-random numbers with given probability distribution which are essential for all Monte Carlo methods. We show how the efficiency of Monte Carlo integration can be improved by sampling preferentially the important configurations. Finally the famous Metropolis algorithm is applied to classical many-particle systems and nonlinear optimization problems.*

9.1 Some Basic Statistics

In the following we discuss some important concepts which are used to analyze experimental data sets [98]. Repeated measurements of some observable usually give slightly different results due to fluctuations of the observable in time and errors of the measurement process. The distribution of the measured data is described by a probability distribution, which in many cases approximates a simple mathematical form like the Gaussian normal distribution. The moments of the probability density give important information about the statistical properties, especially the mean and the standard deviation of the distribution. If the errors of different measurements are uncorrelated, the average value of a larger number of measurements is a good approximation to the "exact" value.

9.1.1 Probability Density and Cumulative Probability Distribution

Consider an observable ξ, which is measured in a real or a computer experiment. Repeated measurements give a statistical distribution of values.

P.O.J. Scherer, *Computational Physics*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-61088-7_9

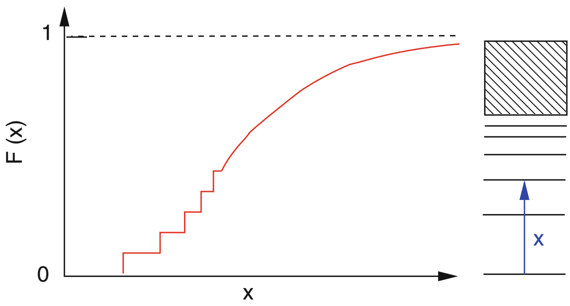

Fig. 9.1 (Cumulative probability distribution of transition energies) The figure shows schematically the distribution of transition energies for an atom which has a discrete and a continuous part

The cumulative probability distribution (Fig. 9.1) is given by the function

$$F(x) = P\{\xi \leq x\} \tag{9.1}$$

and has the following properties:

- $F(x)$ is monotonously increasing
- $F(-\infty) = 0, F(\infty) = 1$
- $F(x)$ can be discontinuous (if there are discrete values of ξ)

The probability to measure a value in the interval $x_1 < \xi \leq x_2$ is

$$P(x_1 < \xi \leq x_2) = F(x_2) - F(x_1). \tag{9.2}$$

The height of a jump gives the probability of a discrete value

$$P(\xi = x_0) = F(x_0 + 0) - F(x_0 - 0). \tag{9.3}$$

In regions where $F(x)$ is continuous, the probability density can be defined as

$$f(x_0) = F'(x_0) = \lim_{\Delta x \to 0} \frac{1}{\Delta x} P(x_0 < \xi \leq x_0 + \Delta x). \tag{9.4}$$

9.1.2 Histogram

From an experiment $F(x)$ cannot be determined directly. Instead a finite number N of values x_i are measured. By

$Z_N(x)$

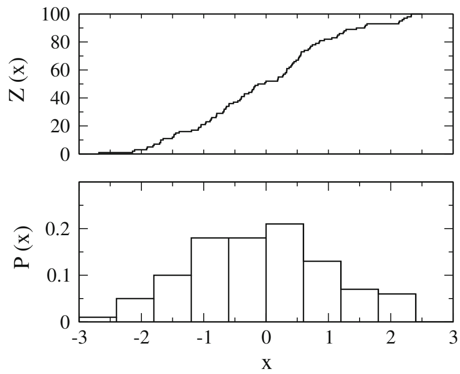

Fig. 9.2 (Histogram) The cumulative distribution of 100 Gaussian random numbers is shown together with a histogram with bin width $\Delta x = 0.6$

we denote the number of measurements with $x_i \leq x$. The cumulative probability distribution is the limit

$$F(x) = \lim_{N\to\infty} \frac{1}{N} Z_N(x). \tag{9.5}$$

A histogram (Fig. 9.2) counts the number of measured values which are in the interval $x_i < x \leq x_{i+1}$:

$$\frac{1}{N}(Z_N(x_{i+1}) - Z_N(x_i)) \approx F(x_{i+1}) - F(x_i) = P(x_i < \xi \leq x_{i+1}). \tag{9.6}$$

Contrary to $Z_N(x)$ itself, the histogram depends on the choice of the intervals.

9.1.3 Expectation Values and Moments

The expectation value of the random variable ξ is defined by

$$E[\xi] = \int_{-\infty}^{\infty} x dF(x) = \lim_{a\to-\infty, b\to\infty} \int_a^b x dF(x) \tag{9.7}$$

with the Riemann-Stieltjes-Integral [99]

$$\int_a^b x dF(x) = \lim_{N\to\infty} \sum_{i=1}^{N} x_i (F(x_i) - F(x_{i-1}))|_{x_i = a + \frac{b-a}{N} i}. \tag{9.8}$$

Higher moments are defined as

$$E[\xi^k] = \int_{-\infty}^{\infty} x^k dF(x) \tag{9.9}$$

if these integrals exist. Most important are the expectation value

$$\overline{x} = E[\xi] \tag{9.10}$$

and the variance, which results from the first two moments

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\overline{x})^2 dF = \int x^2 dF + \int \overline{x}^2 dF - 2\overline{x} \int x dF$$
$$= E[\xi^2] - (E[\xi])^2. \tag{9.11}$$

The standard deviation σ is a measure of the width of the distribution. The expectation value of a function $\varphi(x)$ is defined by

$$E[\varphi(x)] = \int_{-\infty}^{\infty} \varphi(x) dF(x). \tag{9.12}$$

For continuous $F(x)$ we have with $\mathrm{d}F(x) = f(x)\mathrm{d}x$ the ordinary integral

$$E[\xi^k] = \int_{-\infty}^{\infty} x^k f(x) dx \tag{9.13}$$

$$E[\varphi(x)] = \int_{-\infty}^{\infty} \varphi(x) f(x) dx \tag{9.14}$$

whereas for a pure step function $F(x)$ (only discrete values x_i are observed with probabilities $p(x_i) = F(x_i + 0) - F(x_i - 0)$)

$$E[\xi^k] = \sum x_i^k p(x_i) \tag{9.15}$$

$$E[\varphi(x)] = \sum \varphi(x_i) p(x_i). \tag{9.16}$$

9.1.4 Example: Fair Die

When a six-sided fair die is rolled, each of its sides shows up with the same probability of 1/6. The cumulative probability distribution $F(x)$ is a pure step function (Fig. 9.3) and

$$\overline{x} = \int_{-\infty}^{\infty} x dF = \sum_{i=1}^{6} x_i (F(x_i + 0) - F(x_i - 0)) = \frac{1}{6} \sum_{i=1}^{6} x_i = \frac{21}{6} = 3.5 \tag{9.17}$$

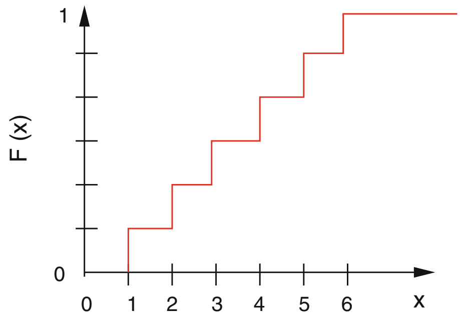

Fig. 9.3 Cumulative probability distribution of a fair die

$$\overline{x^2} = \sum_{i=1}^{6} x_i^2 (F(x_i + 0) - F(x_i - 0)) = \frac{1}{6} \sum_{i=1}^{6} x_i^2 = \frac{91}{6} = 15.1666\cdots \tag{9.18}$$

$$\sigma = \sqrt{\overline{x^2} - \overline{x}^2} = 2.9. \tag{9.19}$$

9.1.5 Normal Distribution

The Gaussian normal distribution is defined by the cumulative probability distribution

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-t^2/2} dt \tag{9.20}$$

and the probability density

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2} \tag{9.21}$$

with the properties

$$\int_{-\infty}^{\infty} \varphi(x) dx = \Phi(\infty) = 1 \tag{9.22}$$

$$\overline{x} = \int_{-\infty}^{\infty} x\varphi(x) dx = 0 \tag{9.23}$$

$$\sigma^2 = \overline{x^2} = \int_{-\infty}^{\infty} x^2 \varphi(x) dx = 1. \tag{9.24}$$

Since $\Phi(0) = \frac{1}{2}$ and with the definition

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2} dt \tag{9.25}$$

we have

$$\Phi(x) = \frac{1}{2} + \Phi_0(x) \tag{9.26}$$

which can be expressed in terms of the error function[1]

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2 dt} dt = 2\Phi_0(\sqrt{2}x) \tag{9.27}$$

as

$$\Phi_0(x) = \frac{1}{2}\text{erf}(\frac{x}{\sqrt{2}}). \tag{9.28}$$

A general Gaussian distribution with mean value $\overline{x}$ and standard deviation σ has the probability distribution

$$\varphi_{\overline{x},\sigma} = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x' - \overline{x})^2}{2\sigma^2}\right) \tag{9.29}$$

and the cumulative distribution

$$\Phi_{\overline{x},\sigma}(x) = \Phi\left(\frac{x - \overline{x}}{\sigma}\right) = \int_{-\infty}^{x} dx' \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x' - \overline{x})^2}{2\sigma^2}\right) \tag{9.30}$$

$$= \frac{1}{2}\left(1 + \text{erf}\left(\frac{x - \overline{x}}{\sigma\sqrt{2}}\right)\right). \tag{9.31}$$

9.1.6 Multivariate Distributions

Consider now two quantities which are measured simultaneously. ξ and η are the corresponding random variables. The cumulative distribution function is

$$F(x, y) = P(\xi \leq x \quad \text{and} \quad \eta \leq y). \tag{9.32}$$

---

[1] erf(x) is an intrinsic function in FORTRAN or C.

Expectation values are defined as

$$E[\varphi(x, y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(x, y) d^2 F(x, y). \tag{9.33}$$

For continuous $F(x, y)$ the probability density is

$$f(x, y) = \frac{\partial^2 F}{\partial x \partial y} \tag{9.34}$$

and the expectation value is simply

$$E[\varphi(x, y)] \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \varphi(x, y) f(x, y). \tag{9.35}$$

The moments of the distribution are the expectation values

$$M_{k,l} = E[\xi^k \eta^l]. \tag{9.36}$$

Most important are the averages

$$\overline{x} = E[\xi] \quad \overline{y} = E[\eta] \tag{9.37}$$

and the covariance matrix

$$\begin{pmatrix} E[(\xi - \overline{x})^2] & E[(\xi - \overline{x})(\eta - \overline{y})] \\ E[(\xi - \overline{x})(\eta - \overline{y})] & E[(\eta - \overline{y})^2] \end{pmatrix} = \begin{pmatrix} \overline{x^2} - \overline{x}^2 & \overline{xy} - \overline{x}\,\overline{y} \\ \overline{xy} - \overline{x}\,\overline{y} & \overline{y^2} - \overline{y}^2 \end{pmatrix}. \tag{9.38}$$

The correlation coefficient is defined as

$$\rho = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\sqrt{\left(\overline{x^2} - \overline{x}^2\right)\left(\overline{y^2} - \overline{y}^2\right)}}. \tag{9.39}$$

If there is no correlation then $\rho = 0$ and $F(x, y) = F_1(x) F_2(y)$.

9.1.7 Central Limit Theorem

Consider N independent random variables ξ_i with the same cumulative distribution function $F(x)$, for which $E[\xi] = 0$ and $E[\xi^2] = 1$. Define a new random variable

$$\eta_N = \frac{\xi_1 + \xi_2 + \cdots \xi_N}{\sqrt{N}} \tag{9.40}$$

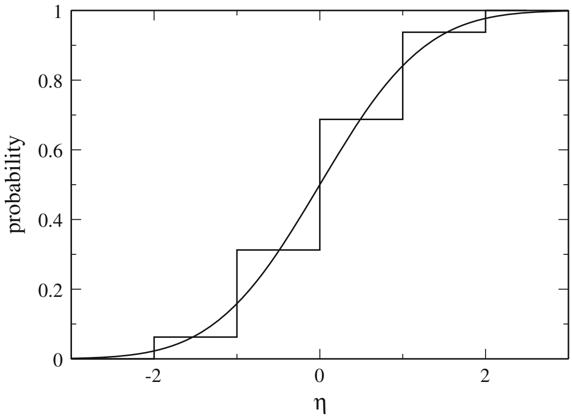

Fig. 9.4 (Central limit theorem) The cumulative distribution function of η (9.42) is shown for $N = 4$ and compared to the normal distribution (9.20)

with the cumulative distribution function $F_N(x)$. In the limit $N \to \infty$ this distribution approaches (Fig. 9.4) a cumulative normal distribution [100]

$$\lim_{N\to\infty} F_N(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-t^2/2} dt. \tag{9.41}$$

9.1.8 Example: Binomial Distribution

Toss a coin N times giving $\xi_i = 1$ (heads) or $\xi_i = -1$ (tails) with equal probability $P = \frac{1}{2}$. Then $E[\xi_i] = 0$ and $E[\xi_i^2] = 1$. The distribution of

$$\eta = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \xi_i \tag{9.42}$$

can be derived from the binomial distribution

$$1 = \left[\frac{1}{2} + \left(-\frac{1}{2}\right)\right]^N = 2^{-N} \sum_{p=0}^{N} (-1)^{N-p} \binom{N}{N-p} \tag{9.43}$$

where p counts the number of tosses with $\xi = +1$. Since

$$n = p \cdot 1 + (N-p) \cdot (-1) = 2p - N \in [-N, N] \tag{9.44}$$

the probability of finding $\eta = \frac{n}{\sqrt{N}}$ is given by the binomial coefficient

$$P(\eta = \frac{2p - N}{\sqrt{N}}) = 2^{-N} \binom{N}{N - p} \tag{9.45}$$

or

$$P(\eta = \frac{n}{\sqrt{N}}) = 2^{-N} \binom{N}{\frac{N-n}{2}}. \tag{9.46}$$

9.1.9 Average of Repeated Measurements

A quantity X is measured N times. The results $X_1 \cdots X_N$ are independent random numbers with the same distribution function $f(X_i)$. Their expectation value is the exact value $E[X_i] = \int dX_i X_i f(X_i) = X$ and the standard deviation due to measurement uncertainties is $\sigma_X = \sqrt{E[X_i^2] - X^2}$. The new random variables

$$\xi_i = \frac{X_i - X}{\sigma_X} \tag{9.47}$$

have zero mean

$$E[\xi_i] = \frac{E[X_i] - X}{\sigma_X} = 0 \tag{9.48}$$

and unit standard deviation

$$\sigma_\xi^2 = E[\xi_i^2] - E[\xi_i]^2 = E\left[\frac{X_i^2 + X^2 - 2XX_i}{\sigma_X^2}\right] = \frac{E[X_i^2] - X^2}{\sigma_X^2} = 1. \tag{9.49}$$

Hence the quantity

$$\eta = \frac{\sum_1^N \xi_i}{\sqrt{N}} = \frac{\sum_1^N X_i - NX}{\sqrt{N}\sigma_X} = \frac{\sqrt{N}}{\sigma_X}(\overline{X} - X) \tag{9.50}$$

obeys a normal distribution

$$f(\eta) = \frac{1}{\sqrt{2\pi}} e^{-\eta^2/2}. \tag{9.51}$$

From

$$f(\overline{X})d\overline{X} = f(\eta)d\eta = f(\eta(\overline{X}))\frac{\sqrt{N}}{\sigma_X}d\overline{X} \tag{9.52}$$

we obtain

$$f(\overline{X}) = \frac{\sqrt{N}}{\sqrt{2\pi}\sigma_X}\exp\left\{-\frac{N}{2\sigma_X^2}(\overline{X}-X)^2\right\}. \tag{9.53}$$

The average of N measurements obeys a Gaussian distribution around the exact value X with a reduced standard deviation of

$$\sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{N}}. \tag{9.54}$$

9.2 Random Numbers

True random numbers of high quality can be generated using physical effects like thermal noise in a diode or atmospheric noise [101]. Computers very often make use of pseudo random numbers which have comparable statistical properties but are not totally unpredictable. For cryptographic purposes sophisticated algorithms are available which are slow but cryptographically secure, e.g. the **Yarrow** [102] and **Fortuna** [103] algorithms. In computational physics, usually simpler methods are sufficient which are not cryptographically secure, but pass important statistical tests like Marsaglia's DIEHARD collection [104, 105] and TestU01 [106, 107]. Most methods use an iterated function (Sect. 22.1). A set of numbers $\mathcal{Z}$ (e.g. 32-bit integers) is mapped onto itself by an invertible function $f(r)$ and, starting from a random seed number $r_0 \in \mathcal{Z}$, the sequence

$$r_{i+1} = f(r_i) \tag{9.55}$$

is calculated to provide a series of pseudo random numbers [105]. Using 32-bit integers there are 2^{32} different numbers, hence the period cannot exceed 2^{32}. The method can be improved by taking $\mathcal{Z}$ to be the set of m-tuples of 32-bit integers $r = \{z_1, z_2 \ldots z_m\}$ and $f(r)$ a function that converts one m-tuple into another. An m-tuple of successive function values defines the iteration

$$r_i = \{z_i, z_{i-1}, \ldots z_{i-m+1}\} \tag{9.56}$$

$$r_{i+1} = \{z_{i+1}, z_i, \ldots z_{i-m+2}\} = \{f(z_i, \ldots z_{i-m+1}), z_i, \ldots z_{i-m+2}\}. \tag{9.57}$$

Table 9.1 Addition modulo 2

$0 + 0 = 0$
$1 + 0 = 1$
$0 + 1 = 1$
$1 + 1 = 0$

Using 32-bit integers, (9.57) has a maximum period of 2^{32m} (Example: for $m = 2$ and generating 10^6 numbers per second the period is 584942 years). For the initial seed, here m independent random numbers have to be provided.

The special case of a lagged RNG simply uses

$$r_{i+1} = \{z_{i+1}, z_i, \dots z_{i-m+2}\} = \{f(z_{i-m+1}), z_i, \dots z_{i-m+2}\}. \tag{9.58}$$

Popular kinds of functions $f(r)$ include linear congruent mappings, xorshift, lagged Fibonacci, multiply with carry (MWC), complimentary multiply with carry (CMWC) methods and combinations of these like the famous Mersenne Twister [108] and KISS [105] algorithms. We discuss briefly some important principles.

9.2.1 Linear Congruent Mapping (LC)

A simple algorithm, mainly of historical importance due to some well known problems [109], is the linear congruent mapping

$$r_{i+1} = (ar_i + c) \bmod b \tag{9.59}$$

with multiplier a and base b which is usually taken to be $b = 2^{32}$ for 32-bit integers since this can be implemented most easily. The maximum period is given by b.

9.2.2 Xorshift

A 32-Bit integer[2] can be viewed as a vector $\mathbf{r} = (b_0, b_1 \dots b_{31})$ of elements b_i in the field $\mathcal{F}_2 = \{0, 1\}$. Addition of two such vectors (modulo 2) can be implemented with the exclusive-or operation as can be seen from comparison with the table (Table 9.1).

An invertible linear transformation of the vector $\mathbf{r}$ can be described by multiplication with a nonsingular 32×32 matrix T

$$f(\mathbf{r}) = \mathbf{r}T. \tag{9.60}$$

[2] This method can be easily extended to 64-Bit integers.

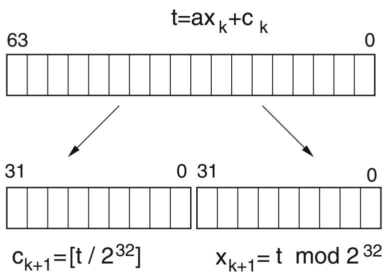

Fig. 9.5 Multiply with carry method using 64-bit integer arithmetic

To simplify the numerical calculation, Marsaglia [105] considers matrices of the special form[3]

$$T = (1 + L^a)(1 + R^b)(1 + L^c) \tag{9.61}$$

where L (R) is a matrix that produces a left (right) shift by one. For properly chosen numbers a, b, c the matrix T is of order $2^{32} - 1$ and the random numbers have the maximum possible period. There are many possible choices, one of them leads to the sequence

$$y = y \text{ xor } (y \ll 13)$$

$$y = y \text{ xor}(y \gg 17)$$

$$y = y \text{ xor } (y \ll 5). \tag{9.62}$$

9.2.3 Multiply with Carry (MWC)

This method is quite similar to the linear congruent mapping. However, instead of the constant c in (9.59) a varying carry is used.

For base $b = 2^{32}$ and multiplier $a = 698769069$ consider pairs of integers $r = [x, c]$ with $0 \le c < a$, $0 \le x < b$ excluding $[0, 0]$and $[a - 1, b - 1]$ and the iteration function[4]

$$f([x, c]) = [ax + c \bmod b, (ax + c)/b]. \tag{9.63}$$

Starting with a random seed $[x_0, c_0]$ the sequence $[x_k, c_k] = f([x_{k-1}, c_{k-1}])$ has a period of about 2^{60} [105]. If one calculates $t = ax_k + c_k$ in 64 bits, then for $b = 2^{32}$,$c_{k+1}$is given by the top 32 bits and $x_{k+1}$by the bottom 32 bits (Fig. 9.5).

[3]At least three factors are necessary for 32 and 64-Bit integers.

[4]Using integer arithmetics.

9.2.4 Complementary Multiply with Carry (CMWC)

The simple MWC method has some inherent problems which can be overcome by a slight modification. First, the base is taken to be $b = 2^{32} - 1$ and second the iteration is changed to use the $(b-1)$-complement

$$x_k = (b-1) - (ax_{k-1} + c_{k-1}) \bmod b. \tag{9.64}$$

This method can provide random numbers which pass many tests and have very large periods.

9.2.5 Random Numbers with Given Distribution

Assume we have a program that generates random numbers in the interval [0,1] like in C:

$$rand()/(double)RAND\_MAX.$$

The corresponding cumulative distribution function is

$$F_0(x) = \begin{cases} 0 & for \quad x < 0 \\ x & for \quad 0 \le x \le 1 \\ 1 & for \quad x > 1 \end{cases}. \tag{9.65}$$

Random numbers with cumulative distribution $F(x)$ can be obtained as follows:

choose a RN $r \in [0, 1]$ with $P(r \le x) = F_0(x)$
let $\xi = F^{-1}(r)$

$F(x)$ increases monotonously and therefore

$$P(\xi \le x) = P(F(\xi) \le F(x)) = P(r \le F(x)) = F_0(F(x)) \tag{9.66}$$

but since $0 \le F(x) \le 1$ we have

$$P(\xi \le x) = F(x). \tag{9.67}$$

This method of course is applicable only if F^{-1} can be expressed analytically.

9.2.6 Examples

9.2.6.1 Fair Die

A six-sided fair die can be simulated as follows:

choose a random number $r \in [0, 1]$

$$\text{Let } \xi = F^{-1}(r) = \begin{cases} 1 & \textit{for} \quad 0 \le r < \frac{1}{6} \\ 2 & \textit{for} \quad \frac{1}{6} \le r < \frac{2}{6} \\ 3 & \textit{for} \quad \frac{2}{6} \le r < \frac{3}{6} \\ 4 & \textit{for} \quad \frac{3}{6} \le r < \frac{4}{6} \\ 5 & \textit{for} \quad \frac{4}{6} \le r < \frac{5}{6} \\ 6 & \textit{for} \quad \frac{5}{6} \le r < 1 \end{cases}$$

9.2.6.2 Exponential Distribution

The cumulative distribution function

$$F(x) = 1 - e^{-x/\lambda} \tag{9.68}$$

which corresponds to the exponential probability density

$$f(x) = \frac{1}{\lambda} e^{-x/\lambda} \tag{9.69}$$

can be inverted by solving

$$r = 1 - e^{-x/\lambda} \tag{9.70}$$

for x:

choose a random number $r \in [0, 1]$
Let $x = F^{-1}(r) = -\lambda \ln(1 - r)$.

9.2.6.3 Random Points on the Unit Sphere

We consider the surface element

$$\frac{1}{4\pi} R^2 d\varphi \sin\theta d\theta. \tag{9.71}$$

Our aim is to generate points on the unit sphere (θ, φ) with the probability density

$$f(\theta,\varphi)d\varphi d\theta = \frac{1}{4\pi}d\varphi \sin\theta d\theta = -\frac{1}{4\pi}d\varphi d\cos\theta. \tag{9.72}$$

The corresponding cumulative distribution is

$$F(\theta,\varphi) = -\frac{1}{4\pi}\int_1^{\cos\theta} d\cos\theta \int_0^{\varphi} d\varphi = \frac{\varphi}{2\pi}\frac{1-\cos\theta}{2} = F_\varphi F_\theta. \tag{9.73}$$

Since this factorizes, the two angles can be determined independently:

choose a first random number $r_1 \in [0, 1]$
Let $\varphi = F_\varphi^{-1}(r_1) = 2\pi r_1$
choose a second random number $r_2 \in [0, 1]$
Let $\theta = F_\theta^{-1}(r_2) = \arccos(1 - 2r_2)$

9.2.6.4 Gaussian Distribution (Box Muller)

For a Gaussian distribution the inverse F^{-1} has no simple analytical form. The famous Box Muller method [110] is based on a 2-dimensional normal distribution with probability density

$$f(x,y) = \frac{1}{2\pi}\exp\left\{-\frac{x^2+y^2}{2}\right\} \tag{9.74}$$

which reads in polar coordinates

$$f(x,y)dxdy = f_p(\rho,\varphi)d\rho d\varphi \frac{1}{2\pi}e^{-\rho^2/2}\rho d\rho d\varphi. \tag{9.75}$$

Hence

$$f_p(\rho,\varphi) = \frac{1}{2\pi}\rho e^{-\rho^2/2} \tag{9.76}$$

and the cumulative distribution factorizes:

$$F_p(\rho,\varphi) = \frac{1}{2\pi}\varphi \cdot \int_0^{\rho} \rho' e^{-\rho'^2/2}d\rho' = \frac{\varphi}{2\pi}(1 - e^{-\rho^2}) = F_\varphi(\varphi)F_\rho(\rho). \tag{9.77}$$

The inverse of F_ρ is

$$\rho = \sqrt{-\ln(1-r)} \tag{9.78}$$

and the following algorithm generates Gaussian random numbers:

$r1 = RN \in [0, 1]$
$r2 = RN \in [0, 1]$
$\rho = \sqrt{-\ln(1 - r_1)}$
$\varphi = 2\pi r_2$
$x = \rho \cos \varphi.$

9.3 Monte-Carlo Integration

Physical problems often involve high dimensional integrals (for instance path integrals, thermodynamic averages) which cannot be evaluated by standard methods. Here Monte Carlo methods can be very useful. Let us start with a very basic example.

9.3.1 Numerical Calculation of π

The area of a unit circle ($r = 1$) is given by $r^2\pi = \pi$. Hence π can be calculated by numerical integration. We use the following algorithm:

choose N points randomly in the first quadrant, for instance N independent pairs $x, y \in [0, 1]$
Calculate $r^2 = x^2 + y^2$
Count the number of points within the circle, i.e. the number of points $Z(r^2 \leq 1)$.
$\frac{\pi}{4}$ is approximately given by $\frac{Z(r^2 \leq 1)}{N}$

The result converges rather slowly (Figs. 9.6, 9.7).

9.3.2 Calculation of an Integral

Let ξ be a random variable in the interval $[a, b]$ with the distribution

$$P(x < \xi \leq x + dx) = f(x)dx = \begin{cases} \frac{1}{b-a} & for \quad x \in [a, b] \\ 0 & else \end{cases} . \tag{9.79}$$

The expectation value of a function $g(x)$ is

$$E[g(x)] = \int_{-\infty}^{\infty} g(x)f(x)dx = \int_a^b g(x)dx \tag{9.80}$$

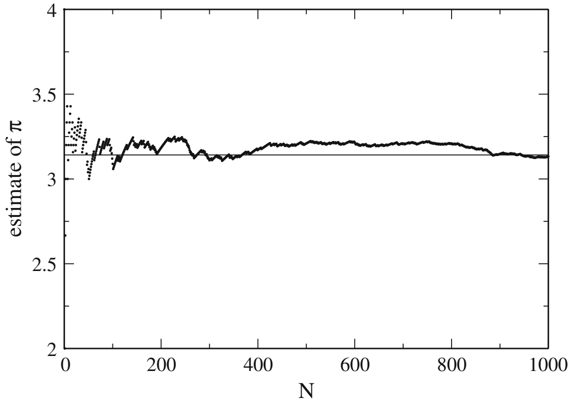

Fig. 9.6 Convergence of the numerical integration

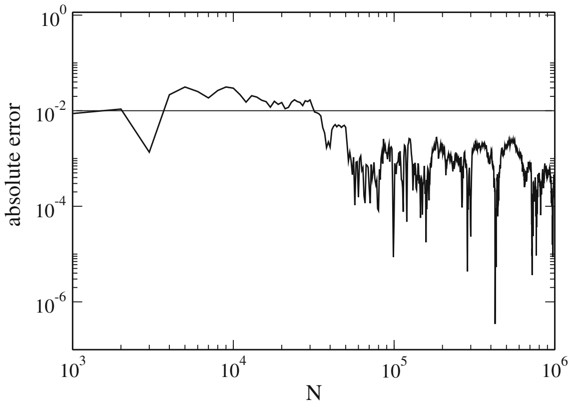

Fig. 9.7 Error of the numerical integration

hence the average of N randomly taken function values approximates the integral

$$\int_a^b g(x)dx \approx \frac{1}{N}\sum_{i=1}^{N} g(\xi_i) = \overline{g(\xi)}. \tag{9.81}$$

To estimate the error we consider the new random variable

$$\gamma = \frac{1}{N}\sum_{i=1}^{N} g(\xi). \tag{9.82}$$

Its average is

$$\overline{\gamma} = E[\gamma] = \frac{1}{N}\sum_{i=1}^{N} E[g(x)] = E[g(x)] = \int_a^b g(x)dx \tag{9.83}$$

and the variance follows from

$$\sigma_\gamma^2 = E[(\gamma - \overline{\gamma})^2] = E[(\frac{1}{N}\sum g(\xi_i) - \overline{\gamma})^2] = E[(\frac{1}{N}\sum (g(\xi_i) - \overline{\gamma}))^2] \quad (9.84)$$

$$= \frac{1}{N^2}E[\sum (g(\xi_i) - \overline{\gamma})^2] = \frac{1}{N}(\overline{g(\xi)^2} - \overline{g(\xi)}^2) = \frac{1}{N}\sigma^2_{g(\xi)}. \quad (9.85)$$

The width of the distribution and hence the uncertainty falls off as $1/\sqrt{N}$.

9.3.3 More General Random Numbers

Consider now random numbers $\xi \in [a, b]$ with arbitrary (but within $[a, b]$ not vanishing) probability density $f(x)$. The integral is approximated by

$$\frac{1}{N}\sum_{i=1}^{N}\frac{g(\xi_i)}{f(\xi_i)} = E\left[\frac{g(x)}{f(x)}\right] = \int_a^b \frac{g(x)}{f(x)}f(x)dx = \int_a^b g(x)dx. \quad (9.86)$$

The new random variable

$$\tau = \frac{1}{N}\sum_{i=1}^{N}\frac{g(\xi_i)}{f(\xi_i)} \quad (9.87)$$

according to (9.85) has a standard deviation given by

$$\sigma_\tau = \frac{1}{\sqrt{N}}\sigma(\frac{g(\xi)}{f(\xi)}) \quad (9.88)$$

which can be reduced by choosing f similar to g. Then preferentially ξ are generated in regions where the integrand is large (importance sampling).

9.3.4 Configuration Integrals

Consider a system which is described by a *ndim* dimensional configuration space $q_1 \dots q_{ndim}$ where a certain configuration has the normalized probability density

$$\varrho(q_1, \dots q_{ndim}) \quad (9.89)$$

$$\int \dots \int \varrho(q_1, \dots q_{ndim})dq^{ndim} = 1. \quad (9.90)$$

The average of an observable $A(q_1 \dots q_{ndim})$ has the form

$$< A >= \int \dots \int A(q_1 \dots q_{ndim}) \varrho(q_1, \dots q_{ndim}) dq^{ndim} \tag{9.91}$$

which will be calculated by MC integration.

Classical Thermodynamic Averages

Consider a classical N particle system with potential energy

$$V(q_1 \dots q_{3N}). \tag{9.92}$$

The probability of a certain configuration is given by its normalized Boltzmann-factor

$$\varrho(q_1 \dots q_{3N}) = \frac{\mathrm{e}^{-\beta V(q_1 \dots q_{3N})}}{\int dq^{3N} \mathrm{e}^{-\beta V(q_1 \dots q_{3N})}} \tag{9.93}$$

and the thermal average of some observable quantity $A(q_1 \dots q_{3N})$ is given by the configuration integral

$$< A >= \int A(q_1 \dots q_{ndim}) \varrho(q_1 \dots q_{3N}) dq^{3N}$$

$$= \frac{\int dq^{3N} A(q_1 \dots q_{ndim}) \mathrm{e}^{-\beta V(q_1 \dots q_{3N})}}{\int dq^{3N} \mathrm{e}^{-\beta V(q_1 \dots q_{3N})}}. \tag{9.94}$$

Variational Quantum Monte Carlo method

Consider a quantum mechanical N particle system with Hamiltonian

$$H = T + V(q_1 \dots q_{3N}). \tag{9.95}$$

According to Ritz's variational principle, the ground state energy is a lower bound to the energy expectation value of any trial wavefunction

$$E_V = \frac{< \Psi_{trial} | H | \Psi_{trial} >}{< \Psi_{trial} | \Psi_{trial} >} \geq E_0. \tag{9.96}$$

Energy and wavefunction of the ground state can be approximated by minimizing the energy of the trial wavefunction, which is rewritten in the form

$$E_V = \frac{\int \Psi^*_{trial}(q_1 \dots q_{3N}) H \Psi_{trial}(q_1 \dots q_{3N}) dq^{3N}}{\int |\Psi_{trial}(q_1 \dots q_{3N})|^2 dq^{3N}}$$

$$= \frac{\int \varrho(q_1 \dots q_{3N}) E_L(q_1 \dots q_{3N}) dq^{3N}}{\int \varrho(q_1 \dots q_{3N}) dq^{3N}} \tag{9.97}$$

with the probability density

$$\varrho(q_1 \dots q_{3N}) = |\Psi_{trial}(q_1 \dots q_{3N})|^2 \tag{9.98}$$

and the so called local energy

$$E_L = \frac{H\Psi_{trial}(q_1 \dots q_{3N})}{\Psi_{trial}(q_1 \dots q_{3N})}. \tag{9.99}$$

9.3.5 Simple Sampling

Let ξ be a random variable which is equally distributed over the range $q_{\min} \cdots q_{\max}$, i.e. a probability distribution

$$P(\xi \in [q, q + dq]) = f(q)dq \tag{9.100}$$

$$f(q) = \begin{cases} \frac{1}{q_{\max} - q_{\min}} & q \in [q_{\min}, q_{\max}] \\ 0 \text{ else} \end{cases} \tag{9.101}$$

$$\int f(q)dq = 1. \tag{9.102}$$

Repeatedly choose *ndim* random numbers $\xi_1^{(m)}, \dots \xi_{ndim}^{(m)}$ and calculate the expectation value

$$E\left(A(\xi_1 \cdots \xi_{ndim})\varrho(\xi_1, \dots \xi_{ndim})\right) = \lim_{M\to\infty} \frac{1}{M} \sum_{m=1}^{M} A(\xi_1^{(m)} \dots \xi_{ndim}^{(m)})\varrho(\xi_1^{(m)} \dots \xi_{ndim}^{(m)})$$

$$= \int A(q_1 \dots q_{ndim})\varrho(q_1 \dots q_{ndim}) f(q_1) \cdots f(q_{ndim}) dq_1 \cdots dq_{ndim}$$

$$= \frac{1}{(q_{\max} - q_{\min})^{ndim}} \int_{q_{\min}}^{q_{\max}} \cdots \int_{q_{\min}}^{q_{\max}} A(q_1 \dots q_{ndim})\varrho(q_1 \dots q_{ndim}) dq^{ndim}.$$

Hence

$$\frac{E\left(A(\xi_1 \cdots \xi_{ndim})\varrho(\xi_1, \ldots \xi_{ndim})\right)}{E\left(\varrho(\xi_1, \ldots \xi_{ndim})\right)}$$

$$= \frac{\int_{q_{\min}}^{q_{\max}} \cdots \int_{q_{\min}}^{q_{\max}} A(q_1 \ldots q_{ndim})\varrho(q_1 \ldots q_{ndim})dq^{ndim}}{\int_{q_{\min}}^{q_{\max}} \cdots \int_{q_{\min}}^{q_{\max}} \varrho(q_1 \ldots q_{ndim})dq^{ndim}} \approx < A > . \tag{9.103}$$

Each set of random numbers $\xi_1 \ldots \xi_{ndim}$ defines one sample configuration. The average over a large number M of samples gives an approximation to the average $< A >$, if the range of the q_i is sufficiently large. However, many of the samples will have small weight and contribute only little.

9.3.6 Importance Sampling

Let us try to sample preferentially the most important configurations. Choose the distribution function as

$$f(q_1 \cdots q_{ndim}) = \varrho(q_1 \ldots q_{ndim}). \tag{9.104}$$

The expectation value of A now directly approximates the configurational average

$$E\left(A(\xi_1 \cdots \xi_{ndim})\right) = \lim_{M\to\infty} \frac{1}{M} \sum_{m=1}^{M} A(\xi_1^{(m)} \ldots \xi_{ndim}^{(m)})$$

$$= \int A(q_1 \ldots q_{ndim})\varrho(q_1 \ldots q_{ndim})dq^{ndim} = < A > . \tag{9.105}$$

9.3.7 Metropolis Algorithm

The algorithm by Metropolis [111] can be used to select the necessary configurations. Starting from an initial configuration $\mathbf{q}_0 = (q_1^{(0)} \cdots q_{3N}^{(0)})$ a chain of configurations is generated. Each configuration depends only on its predecessor, hence the configurations form a Markov chain.

The transition probabilities

$$W_{i\to j} = P(\mathbf{q}_i \to \mathbf{q}_j) \tag{9.106}$$

are chosen to fulfill the condition of detailed balance (Fig. 9.8)

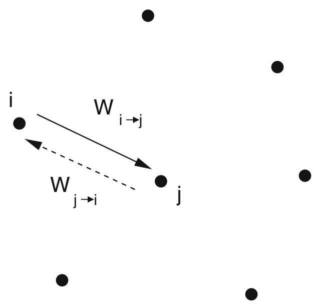

Fig. 9.8 Principle of detailed balance

$$\frac{W_{i\to j}}{W_{j\to i}} = \frac{\varrho(\mathbf{q}_j)}{\varrho(\mathbf{q}_i)}. \tag{9.107}$$

This is a sufficient condition that the configurations are generated with probabilities given by their Boltzmann factors. This can be seen from consideration of an ensemble of such Markov chains: Let $N_n(\mathbf{q}_i)$ denote the number of chains which are in the configuration $\mathbf{q}_i$ after n steps. The changes during the following step are

$$\Delta N(\mathbf{q}_i) = N_{n+1}(\mathbf{q}_i) - N_n(\mathbf{q}_i) = \sum_{\mathbf{q}_j \in conf.} N_n(\mathbf{q}_j) W_{j\to i} - N_n(\mathbf{q}_i) W_{i\to j}. \tag{9.108}$$

In equilibrium

$$N_{eq}(\mathbf{q}_i) = N_0 \varrho(\mathbf{q}_i) \tag{9.109}$$

and the changes (9.108) vanish:

$$\begin{aligned} \Delta N(\mathbf{q}_i) &= N_0 \sum_{\mathbf{q}_j} \varrho(\mathbf{q}_j) W_{j\to i} - \varrho(\mathbf{q}_i) W_{i\to j} \\ &= N_0 \sum_{\mathbf{q}_j} \varrho(\mathbf{q}_j) W_{j\to i} - \varrho(\mathbf{q}_i) \left[W_{j\to i} \frac{\varrho(\mathbf{q}_j)}{\varrho(\mathbf{q}_i)} \right] \\ &= 0. \end{aligned} \tag{9.110}$$

A solution of

$$\Delta N(\mathbf{q}_i) = \sum_{\mathbf{q}_j \in conf.} N_n(\mathbf{q}_j) W_{j\to i} - N_n(\mathbf{q}_i) W_{i\to j} = 0 \tag{9.111}$$

corresponds to a zero eigenvalue of the system of equations

$$\sum_{\mathbf{q}_j} N(\mathbf{q}_j) W_{j\to i} - N(\mathbf{q}_i) \sum_{\mathbf{q}_j} W_{i\to j} = \lambda N(\mathbf{q}_i). \tag{9.112}$$

One solution of this eigenvalue equation is given by

$$\frac{N_{eq}(\mathbf{q}_j)}{N_{eq}(\mathbf{q}_i)} = \frac{\varrho(\mathbf{q}_j)}{\varrho(\mathbf{q}_i)}. \tag{9.113}$$

However, there may be other solutions. For instance if not all configurations are connected by possible transitions and some isolated configurations are occupied initially.

Metropolis Algorithm

This famous algorithm consists of the following steps:
(a) choose a new configuration randomly (trial step) with probability

$$T(\mathbf{q}_i \to \mathbf{q}_{trial}) = T(\mathbf{q}_{trial} \to \mathbf{q}_i)$$

(b) calculate

$$R = \frac{\varrho(\mathbf{q}_{trial})}{\varrho(\mathbf{q}_i)}$$

(c) if $R \geq 1$ the trial step is accepted $\mathbf{q}_{i+1} = \mathbf{q}_{trial}$
(d) if $R < 1$ the trial step is accepted only with probability R. choose a random number $\xi \in [0, 1]$ and the next configuration according to

$$\mathbf{q}_{i+1} = \begin{cases} \mathbf{q}_{trial} & if \quad \xi < R \\ \mathbf{q}_i & if \quad \xi \geq R. \end{cases}$$

The transition probability is the product

$$W_{i \to j} = T_{i \to j} A_{i \to j} \tag{9.114}$$

of the probability $T_{i \to j}$ to select $i \to j$ as a trial step and the probability $A_{i \to j}$ to accept the trial step. Now we have

$$\begin{aligned} for\ R \geq 1 \ \to \ A_{i \to j} = 1,\ A_{j \to i} = R^{-1} \\ for\ R < 1 \ \to \ A_{i \to j} = R,\ A_{j \to i} = 1 \end{aligned} . \tag{9.115}$$

Since $T_{i \to j} = T_{j \to i}$, in both cases

$$\frac{N_{eq}(\mathbf{q}_j)}{N_{eq}(\mathbf{q}_i)} = \frac{W_{i \to j}}{W_{j \to i}} = \frac{A_{i \to j}}{A_{j \to i}} = R = \frac{\varrho(\mathbf{q}_j)}{\varrho(\mathbf{q}_i)}. \tag{9.116}$$

The size of the trial steps has to be adjusted to produce a reasonable acceptance ratio of

$$\frac{N_{accepted}}{N_{rejected}} \approx 1. \tag{9.117}$$

Multiple Walkers

To scan the relevant configurations more completely and reduce correlation between the samples, usually a large number of "walkers" is used (e.g. several hundred) which, starting from different initial conditions, represent independent Markov chains. This also offers a simple possibility for parallelization.

Problems

Problem 9.1 Central Limit Theorem

This computer experiment draws a histogram for the random variable τ, which is calculated from N random numbers $\xi_1 \cdots \xi_N$:

$$\tau = \frac{\sum_{i=1}^{N} \xi_i}{\sqrt{N}}. \tag{9.118}$$

The ξ_i are random numbers with zero mean and unit variance and can be chosen as

- $\xi_i = \pm 1$ (coin tossing)
- Gaussian random numbers

Investigate how a Gaussian distribution is approached for large N.

Problem 9.2 Nonlinear Optimization

MC methods can be used for nonlinear optimization (Traveling salesman problem, structure optimization etc.) [112]. Consider an energy function depending on many coordinates

$$E(q_1, q_2 \cdots q_N). \tag{9.119}$$

Introduce a fictitious temperature T and generate configurations with probabilities

$$P(q_1 \cdots q_N) = \frac{1}{Z} \mathrm{e}^{-E(q_1 \cdots q_N)/T}. \tag{9.120}$$

Slow cooling drives the system into a local minimum. By repeated heating and cooling other local minima can be reached (simulated annealing)

In this computer experiment we try to find the shortest path which visits each of N up to 50 given points. The fictitious Boltzmann factor for a path with total length L is

$$P(L) = e^{-L/T}. \tag{9.121}$$

Starting from an initial path $S = (i_1, i_2, \cdots i_N)$ $n < 5$ and p are chosen randomly and a new path $S' = (i_1, \cdots i_{p-1}, i_{p+n}, \cdots i_p, i_{p+n+1}, \cdots i_N)$ is generated by reverting the sub-path

$$i_p \cdots i_{p+n} \rightarrow i_{p+n} \cdots i_p.$$

Start at high temperature $T > L$ and cool down slowly.