Chapter 9

Random numbers and Monte Carlo methods

The name *Monte Carlo* comes from the city on the Mediterranean with its famous casino, and a Monte Carlo calculation implies a statistical method of studying problems based on the use of random numbers, similar to those generated in the casino games of chance.

The applications of Monte-Carlo methods are many and various, but many will come under these headings:

1. The object of study cannot be manipulated efficiently by analytical maths; a good example is the application of Monte-Carlo methods to the integration of complex functions.
2. The analyst wants to use data which describes a process: Examples include the analysis of investment portfolios and the application of oceanographic data to estimate availability of offshore loading facilities for oil tankers.

Two crucial ingredients (covered below) are needed for Monte-Carlo:

1. Probability Distribution Functions (includes Moments of a PDF and its pertinent variance)
2. Random Variables

9.1 Basic probability theory

A *random variable* X can take values x (or y, or $2\psi-1$, or 0.7).

$X = 0.7$ is an *event*, as is $X \leq 1.4$. An event ω has a *probability* $P(\omega)$ with

$$0 \leq P(\omega) \leq 1 \; . \tag{9.1}$$

The union of two events has the probability

$$P(\omega_1 \cup \omega_2) = P(\omega_1) + P(\omega_2) - P(\omega_1, \omega_2) \tag{9.2}$$

where we have used the notation $P(\omega_1, \omega_2) \equiv P(\omega_1 \cap \omega_2)$ (the property that both ω_1 *and* ω_2 occur).

Discrete probability distributions: Random variable X takes on discrete set of values; typically equidistant values The *probability distribution function* is defined as

$$F_X(n) \;\equiv\; P(X \le n) = \sum_{k \le n} p_k \;, \tag{9.3}$$

where

$$p_n \;\equiv\; P(X{=}n) \tag{9.4}$$

is the probability of the event 'X equals n'.

Continuous probability distributions: Random variable X takes on value from an interval, often $(-\infty, \infty)$. The *probability distribution function* is

$$F_X(x) \;\equiv\; P(X \le n) = \int_{-\infty}^{x} f_X(x')\, dx' \;, \tag{9.5}$$

where

$$f_X(x) \equiv \frac{dF_X(x)}{dx} = \frac{P(x < X < x{+}dx)}{dx} \tag{9.6}$$

is called *probability density function, PDF* or *probability density* or *density function*.

Note 1: For a discrete distribution, we can write

$$f_X(x) \equiv \frac{dF_X(x)}{dx} = \sum_n p_n \delta(x{-}x_n) \;, \tag{9.7}$$

where $\delta(\cdot)$ is Dirac's delta function.

Note 2: For any distribution, $F_X(-\infty) = 0$ and $F_X(\infty) = 1$.

Examples: Bernoulli distribution: X can take on values 0 or 1.

$$p_0 = p\ , \qquad p_1 = 1-p\ . \tag{9.8}$$

which implies

$$F_X(0) = p\ , \qquad F_X(1) = 1\ . \tag{9.9}$$

Uniform distribution $\mathcal{U}$: If $U \sim \mathcal{U}(0,1)$,[1] then

$$f_U(x) = \begin{cases} 0 & x < 0 \\ 1 & 0 \le x \le 1 \\ 0 & x > 1 \end{cases}\ , \tag{9.10}$$

which implies

$$F_U(x) = \begin{cases} 0 & x < 0 \\ x & 0 \le x \le 1 \\ 1 & x > 1 \end{cases}\ . \tag{9.11}$$

9.1.1 Expectation value, variance, covariance

The *expectation value* EX of X is defined as

$$EX \ \equiv \int_{-\infty}^{\infty} x\, dF_X(x)\ . \tag{9.12}$$

For a continuous distribution, this becomes

$$EX \ = \int_{-\infty}^{\infty} x\, f_X(x)\, dx\ , \tag{9.13}$$

and for a discrete one

$$EX \ = \sum_{k=-\infty}^{\infty} x_k\, p_k\ . \tag{9.14}$$

If X is a random variable, then an arbitrary function $g(X)$ is, too. So we can define

$$Eg(X) \ \equiv \int_{-\infty}^{\infty} g(x)\, dF_X(x) = \begin{cases} \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx \\ \sum_{k=-\infty}^{\infty} g(x_k)\, p_k \end{cases} \tag{9.15}$$

We can write

$$f_X(x) = \int \delta(\xi - x)\, f_X(\xi)\, d\xi = E\delta(X-x)\ . \tag{9.16}$$

[1] This is a short notation we will use for a few important distributions. It reads 'U is distributed according to the uniform distribution function with offset 0 and width 1'.

Note: The expectation value (9.12) is linear in X, thus in particular

$$E(X+Y) = EX + EY \ . \tag{9.17}$$

The *variance* $V(X)$ is defined as

$$V(X) \equiv E(X-EX)^2 \ , \tag{9.18}$$

which can be rewritten as

$$V(X) = E[X^2 - 2X\,EX + (EX)^2] = EX^2 - 2E[X\,EX] + (EX)^2 = EX^2 - (EX)^2 \ . \tag{9.19}$$

The variance of $X+Y$ is

$$\begin{aligned} V(X+Y) &= E(X-EX+Y-EY)^2 \\ &= V(X) + V(Y) + 2E(X-EX)(Y-EY) \\ &= V(X) + V(Y) + 2\,\mathrm{Kov}(X,Y) \ , \end{aligned} \tag{9.20}$$

where

$$\mathrm{Kov}(X,Y) \equiv E(X-EX)(Y-EY) = EXY - EX\,EY \tag{9.21}$$

is the *covariance* of X and Y.

The *correlation coefficient*

$$\varrho(X,Y) \equiv \frac{\mathrm{Kov}(X,Y)}{\sqrt{V(X)V(Y)}} \tag{9.22}$$

satisfies $-1 \le \varrho(X,Y) \le 1$.

If $\mathrm{Kov}(X,Y) = \varrho(X,Y) = 0$ (e.g. if X and Y are independent, see below), then

$$V(X+Y) = V(X) + V(Y) \ . \tag{9.23}$$

9.1.2 Joint and conditional probabilities

Consider 2 random variables X, Y.

$$F_{X,Y}(x,y) \equiv P(X{\le}x, Y{\le}y) \tag{9.24}$$

is the *joint* distribution function, and

$$f_{X,Y}(x,y) \equiv \frac{P(x{<}X{<}x{+}dx,\ y{<}Y{<}y{+}dy)}{dx\,dy} = \partial_x \partial_y F_{X,Y}(x,y) \tag{9.25}$$

is the joint probability density function.

If Y is irrelevant, X has the density function

$$f_X(x) = P(x<X<x+dx, Y \text{ irrelevant}) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy \tag{9.26}$$

(sometimes called the *marginal distribution*), and similar for f_Y.

Expectation values are now

$$Eg(X,Y) = \iint g(x,y) f_{X,Y}(x,y)\,dx\,dy\ . \tag{9.27}$$

If

$$f_{X,Y}(x,y) = f_X(x) f_Y(y)\ , \tag{9.28}$$

X and Y are called *independent* random variables. In that case,

$$\begin{aligned} \mathrm{Kov}(X,Y) &= EXY - EX\,EY = \iint f_{X,Y}(x,y)\,dx\,dy - EX\,EY \\ &= \int f_X(x)\,dx \int f_Y(y)\,dx - EX\,EY = 0\ . \end{aligned} \tag{9.29}$$

But $\mathrm{Kov}(X,Y) = 0$ (X and Y are *uncorrelated*) is *not* sufficient for stochastic independence. Counter example:

$$X \sim \mathcal{U}(0,1)\ , \qquad Y = 4X(1-X)\ . \tag{9.30}$$

The *conditional probability* of ω_1 under the condition ω_2 is

$$P(\omega_1|\omega_2) = \frac{P(\omega_1,\omega_2)}{P(\omega_2)}\ . \tag{9.31}$$

For example,

$$P(X\leq x|Y<2) = \frac{P(X\leq x|Y<2)}{P(Y<2)}\ . \tag{9.32}$$

If ψ_k ($k = 0, 1, \ldots$) is a complete set of mutually exclusive events, i.e.

$$\psi_k \cap \psi_l = \emptyset \text{ for } k \neq l\ , \qquad \text{and } \sum_k P(\psi_k) = 1\ , \tag{9.33}$$

then

$$\sum_k P(\omega|\psi_k)P(\psi_k) = \sum_k P(\omega,\psi_k) = P(\omega)\ , \tag{9.34}$$

i.e.

$$P(\omega) = \sum_k P(\omega|\psi_k)P(\psi_k)\ . \tag{9.35}$$

Equation (9.35) is called the *total probability theorem*.

9.1.3 Distribution of sums of random variables

Let X and Y be two random variables with probability density functions f_X and f_Y, respectively. What is the distribution of $Z \equiv X + Y$?

We can write the probability density of Z as

$$f_Z(z) = E\delta(X + Y - z) = \iint dx\, dy\, f_{X,Y}(x, y)\, \delta(x+y-z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z-x)\, dx\ . \tag{9.36}$$

If X and Y are independent, this becomes

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx\ , \tag{9.37}$$

i.e. the density function of the sum $X + Y$ is the *convolution* of the individual density functions.

9.1.4 Individual distributions

Geometric distribution:

$$p_n = p(1-p)^n\ , \qquad F_X(n) = 1 - (1-p)^{n+1} \tag{9.38}$$

Mean value and variance:

$$EX = \frac{1-p}{p}\ , \qquad V(X) = \frac{1-p}{p^2}\ . \tag{9.39}$$

Binomial distribution $\mathcal{B}(N, p)$:

$$P(n|N) = \binom{N}{n} p^n (1-p)^{N-n} \tag{9.40}$$

Mean value and variance:

$$EX = Np\ , \qquad V(X) = Np(1-p)\ . \tag{9.41}$$

Poisson distribution:

$$P(n) = \frac{\lambda^n e^{-\lambda}}{n!}\ . \tag{9.42}$$

Mean value and variance:

$$EX = \lambda\ , \qquad V(X) = \lambda\ . \tag{9.43}$$

Interpretation: radioactive decay, decay rate γ [decays/s], time interval Δt, parameter $\lambda = \alpha\Delta t$. Probability of getting n decays during the time interval is $P(n)$.

- The Poisson distribution represents and approximation to the binomial distribution for the special case where the average number of successes is much smaller than the possible number n; that is, when $\mu << n$.

Uniform distribution $\mathcal{U}(x_0, w)$:

$$F_X(x) = \begin{cases} 0 & x < x_0 \\ \dfrac{x - x_0}{w} & x_0 \le x \le x_0{+}w \\ 1 & x > x_0{+}w \end{cases} , \qquad f_X(x) = \begin{cases} 0 & x < x_0 \\ \dfrac{1}{w} & x_0 \le x \le x_0{+}w \\ 0 & x > x_0{+}w \end{cases} \tag{9.44}$$

Normal (or Gaussian) distribution $\mathcal{N}(\mu, \sigma)$:

$$F_X(x) = \frac{1 + \operatorname{erf} \dfrac{x-\mu}{\sqrt{2}\sigma}}{2} , \qquad f_X(x) = \frac{1}{2\pi\sigma^2} e^{-(x-\mu)^2/(2\sigma^2)} , \tag{9.45}$$

where $\operatorname{erf} z = (2/\sqrt{\pi}) \int_0^z e^{-\zeta^2}\, d\zeta$ is the *error function*.
Mean value and variance:

$$EX = \mu , \qquad V(X) = \sigma^2 . \tag{9.46}$$

- The Gaussian distribution represents and approximation to the binomial distribution for the special case where the number of possible different observations n becomes infinitely large and the probability of success for each is finitely large. It is also the limiting case for the Poisson distribution as μ becomes large.
- The Gaussian distribution is undoubtedly the most important in statistical analysis of data. Practically, it is useful because it seems to describe the distribution of random observations of many experiments, as well as describing the distribution obtained when we try to estimate the parameters of most other probability distributions.

Exponential distribution:

$$F_X(x) = 1 - e^{-\alpha x} , \qquad f_X(x) = \alpha e^{-\alpha x} . \tag{9.47}$$

Mean value and variance:

$$EX = \frac{1}{\alpha} , \qquad V(X) = \frac{1}{\alpha^2} . \tag{9.48}$$

Interpretation: radioactive decay, α is the decay rate; x is waiting time for first decay event.

Gamma distribution:

$$F_X(x) = \frac{\gamma(\beta, \alpha x)}{(\beta{-}1)!} \text{ (incomplete gamma function)} , \qquad f_X(x) = \frac{x^{\beta-1}\alpha^\beta e^{-\alpha x}}{(\beta{-}1)!} . \tag{9.49}$$

Mean value and variance:

$$EX = \frac{1}{\alpha} , \qquad V(X) = \frac{1}{\alpha^2} . \tag{9.50}$$

Interpretation: radioactive decay, decay rate α; x is waiting time for βth decay event.

Cauchy distribution:

$$F_X(x) = \frac{1}{2} + \frac{1}{\pi} \arctan \frac{x}{b} , \qquad f_X(x) = \frac{b}{\pi} \frac{1}{b^2 + x^2} . \tag{9.51}$$

Mean value undefined, variance ∞.

9.2 Generating random numbers with a given distribution

9.2.1 Congruential generators

Consider the following recursion:

$$x_{n+1} = 3\, x_n \mod 7 . \tag{9.52}$$

Start e.g. with $x_1 = 5$, then x_n follows the sequence $5, 1, 3, 2, 6, 4, 5, \ldots$, i.e. goes through all values (except 0, which is absorbing) before cycling. Divide x_n by 7 to get sequence $0.71, 0.14, 0.43, 0.29, 0.86, 0.57, 0.71, \ldots$.

On the other hand,

$$x_{n+1} = 3\, x_n \mod 7 . \tag{9.53}$$

has two shorter cycles: $5, 3, 6, 5, \ldots$ and $2, 4, 1, 2, \ldots$

Linear congruential method:

$$x_{n+1} = a\, x_n + c \mod m . \tag{9.54}$$

To get realizations of $Y \sim \mathcal{U}(0, 1)$, use $y_n = x_n/m$.

Important to find good values of A, c and m — and test the (pseudo) random numbers. E.g. Parker & Miller: $a = 7^5 = 16808$, $c = 0$, $m = 2^{31} - 1 = 2147483647$ (will cycle after at most 2×10^9 iterations).

9.2.2 Other distributions

Transformation method

Methods like the congruential generators yield uniformly distributed random numbers on $(0, 1)$. We often need to map these to random numbers with other distributions – so what happens to the probability density function if we transform a random variable $X \mapsto g(X)$?

Let $Y = g(X)$, and assume $g(x)$ is monotonically non-decreasing, then

$$P(X < x) = P[g(X) < g(x)] \,, \tag{9.55}$$

or

$$F_X(x) = F_Y(g(x)) = F_Y(y) \,, \tag{9.56}$$

where $y \equiv g(x)$.

Note that taking the derivative w.r.t. x, we get

$$f_X(x) = f_Y[g(x)]\, g'(x) = f_Y(y) \frac{dy}{dx} \,, \tag{9.57}$$

which can be written as

$$f_X(x)\, dx = f_Y(y)\, dy \,. \tag{9.58}$$

Going back to Eq. (9.56), we now replace X by $Y \sim \mathcal{U}(0, 1)$ and Y by X:

$$F_U(u) = F_X[g(u)] \,, \tag{9.59}$$

and since $F_U(u) = u$, we find

$$F_X(x) = u \,, \tag{9.60}$$

or

$$x = g(u) = F_X^{-1}(u) \,, \tag{9.61}$$

provided we can invert the function $F_X(\cdot)$.

We thus get the following method to construct random numbers X with the density function $f_X(x)$:

1. Integrate $f_X(x)$ to get
$$F_X(x) = \int_{-\infty}^{x} f_X(x')\, dx'$$
2. Obtain uniformly distributed random numbers u_i
3. Solve the equation
$$F_X(x_i) = u_i \,. \tag{9.62}$$
for x_i (i.e. invert F_X). The values x_i will sample the desired distribution.

Example: To construct random numbers X that are exponentially distributed,

$$f_X(x) = \alpha\, e^{-\alpha x} \qquad x \geq 0\ , \tag{9.63}$$

we find

$$F_X(x) = 1 - e^{-\alpha x} \qquad x \geq 0\ , \tag{9.64}$$

and thus

$$1 - e^{-\alpha x_i} = u_i\ , \tag{9.65}$$

which leads to

$$x_i = -\frac{1}{\alpha} \ln(1 - u_i)\ . \tag{9.66}$$

If u_i is uniformly distributed over $[0, 1]$, then $1 - u_i$ is as well, so instead of Eq. (9.67), we can use the slightly simpler

$$x_i = -\frac{1}{\alpha} \ln \tilde{u}_i\ , \tag{9.67}$$

where $\tilde{u}_i$ are uniformly distributed random numbers.

Can we use this to construct normally distributed random numbers? Only if we are willing to somehow numerically calculate the inverse error function. But there is a better method, that allows us to construct two normally distributed random variables at once.

Consider $X \sim \mathcal{N}(0, 1)$ and $Y \sim \mathcal{N}(0, 1)$, which are supposed to be independent. The joint probability density is

$$f_{X,Y}(x, y) = f_X(x)\, f_Y(y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2} = \frac{1}{2\pi} e^{-r^2/2}\ , \tag{9.68}$$

where $r = \sqrt{x^2 + y^2}$. For $R \equiv \sqrt{X^2 + Y^2}$, we find the probability distribution function

$$F_R(r) = \int\limits_{r' \leq r} f_{X,Y}(x, y)\, dx'\, dy' = \frac{1}{2\pi} \int\limits_{r' \leq r} e^{-r'^2/2}\, dx'\, dy'\ . \tag{9.69}$$

Switching to polar coordinates, we have $dx\, dy = d\varphi\, r\, dr$ and thus

$$F_R(r) = P(R \leq r) = \frac{1}{2\pi} \int_0^{2\pi} d\varphi' \int_0^r dr'\, r'\, e^{-r'^2/2} = 1 - e^{-r^2/2}\ . \tag{9.70}$$

Setting $F_R(r_i) = 1 - F_U(u_i) = 1 - u_i$ as above[2], we get

$$r_i = \sqrt{-2 \ln u_i}\ , \tag{9.71}$$

so we now know how to construct random numbers R. To get X and Y, we have to multiply R with a random phase factor,

$$X = R \cos\Phi\ , \qquad Y = R \sin\Phi\ , \tag{9.72}$$

[2]We could use $F_R(r_i) = F_U(u_i) = u_i$, but again $1 - U$ has the same characteristics as U

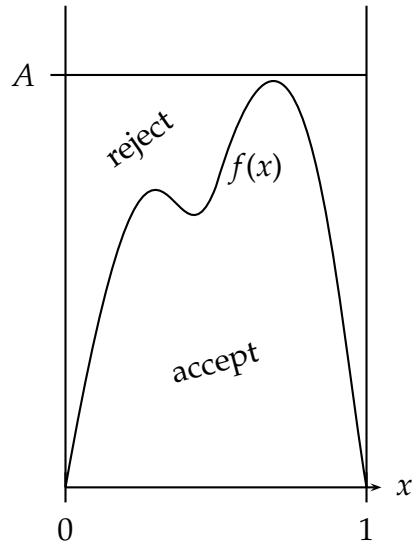
where $\Phi \sim \mathcal{U}(0, 2\pi)$. One can verify that the resulting random numbers X, Y are statistically independent and are normally distributed.

Thus, to construct pairs of normally distributed random numbers, we have the following procedure:

1. Obtain a pair (u_i, v_i) of uniformly distributed random numbers;
2. calculate
$$r_i = \sqrt{-2 \ln u_i}\,, \qquad \varphi_i = 2\pi v_i\,; \tag{9.73}$$
3. the normally distributed variables are
$$x_i = r_i \cos\varphi_i\,, \qquad y_i = r_i \sin\varphi_i\,. \tag{9.74}$$

Acceptance-rejection method

Consider a distribution with a density $f(x)$ that lives only on the interval $[0, 1]$:



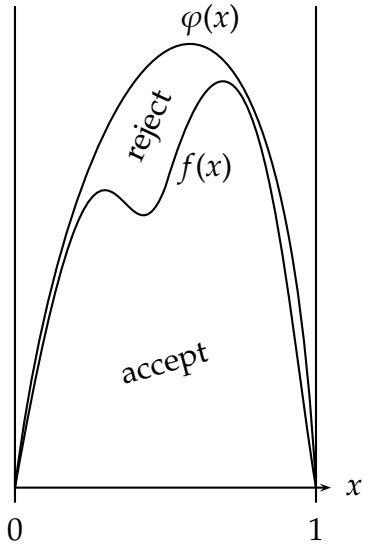
Pick a point (x_i, y_i) in that rectangle at random:

$$X \sim \mathcal{U}(0, 1)\,, \qquad Y \sim \mathcal{U}(0, A)\,. \tag{9.75}$$

If we now discard all pairs (x_i, y_i) that are above the curve $y = f(x)$, then the probability of a certain x value will be equal to $f(x)$. Formally, this becomes

$$P[x<X<x+dx \,|\, Y<f(X)] = f(x)\,dx\,. \tag{9.76}$$

If we use the method in this form, we may have to discard a lot of points, but the method can be somewhat generalized: consider a distribution with density $f(x)$ and introduce a *comparison function* $\varphi(x)$ with $f(x) \le \varphi(x)$ everywhere (in the motivation section above, $\varphi(x)$ was just a constant function).



If we can choose random points (x_i, y_i) that are below $\varphi(x)$, but have constant density in x and y, we can discard all points with $f(x_i) < y_i < \varphi(x_i)$ and find that the x coordinates of the remaining points again have probability density function $f(x)$. So how can we distribute the points evenly below the curve $\varphi(x)$?

If we use $X \sim \mathcal{U}(0,1)$ and for each x_i choose a y_i distributed according to $Y \sim \mathcal{U}[0, \varphi(x_i)]$, then we will have too many points where $\varphi(x)$ is small, and too few where it is large. We must compensate for this by having a density in X that is proportional to $\varphi(x)$. This is our original problem, but now for the comparison function $\varphi(x)$, which we can control. If we choose $\varphi(x)$ such that it can be analytically integrated and the integral can be inverted, then we can use the transformation method to get random numbers x_i distributed according to the probability density function

$$f_X(x) = \frac{\varphi(x)}{\int_{-\infty}^{\infty} \varphi(x')\,dx'} \,. \tag{9.77}$$

This leads us to the following recipe.

1. Choose a comparison function $\varphi(x)$ that satisfies

$$\varphi(x) \geq f(x) \tag{9.78}$$

everywhere, but such that it can be analytically integrated:

$$\widetilde{\Phi}(x) = \frac{\int_{-\infty}^{x} \varphi(x')\,dx'}{\int_{-\infty}^{\infty} \varphi(x')\,dx'} \tag{9.79}$$

and $\widetilde{\Phi}$ can be analytically inverted;

2. obtain a pair (u_i, v_i) of uniformly distributed random numbers;
3. calculate
$$x_i = \widetilde{\Phi}^{-1}(u_i) \,, \qquad \text{and} \qquad y_i = v_i\, \varphi(x_i) \,. \tag{9.80}$$
4. accept x_i as next random number if
$$y_i < f(x_i) \,, \tag{9.81}$$
otherwise reject it and start again.

Note1: The efficiency of our acceptance-rejection method (i.e. the fraction of accepted points) is
$$\eta = \frac{\int_{-\infty}^{\infty} f(x)\, dx}{\int_{-\infty}^{\infty} \varphi(x)\, dx} = \frac{1}{\int_{-\infty}^{\infty} \varphi(x)\, dx} \,. \tag{9.82}$$
It depends on how close the comparison function is to the function $f(x)$. The constant function which we used in the motivation above is often quite inefficient.

Note2: For many density functions, the Cauchy profile
$$\varphi(x) = \frac{A\, b}{\pi} \frac{1}{b^2 + (x - x_0)^2} \tag{9.83}$$
is a good choice as comparison function. In this case, one has
$$\widetilde{\Phi}(x) = \frac{1}{2} + \frac{1}{\pi} \arctan \frac{x - x_0}{b} \,, \tag{9.84}$$
and
$$\widetilde{\Phi}^{-1}(u) = x_0 + b \tan[\pi(u - 1/2)] \,. \tag{9.85}$$

9.2.3 Superposition method

If the density $f_X(x)$ is the linear superposition of (a finite number of or infinitely many) density functions for which we can generate random numbers (e.g. because we know the distribution functions and can invert them),
$$f_X(x) = \sum_k c_k f_{X_k}(x) \,, \qquad \text{where} \sum_k c_k = 1 \,, \tag{9.86}$$
then we can generate random numbers X as follows.

1. Obtain two uniformly distributed random numbers u, v.

2. Use u to draw a random number k with probability $P(K{=}k) = c_k$ (i.e. treat c_k as probabilities of a discrete random variable, see § 9.2.4 below)
3. Use v to generate a random number Y according to the distribution for X_k.

To see that this indeed yields random numbers with the desired distribution, we note that

$$\sum_{k=0}^{\infty} c_k \, P(X{\leq}x \,|\, K{=}k) \;=\; \sum_{k=0}^{\infty} P(K{=}k)\, P(X{\leq}x \,|\, K{=}k) \;=\; P(X{\leq}x) \tag{9.87}$$

because of the total probability theorem Eq. (9.35).

Example: To generate a random number x for the distribution with

$$f_X(x) = \frac{1}{3} + \frac{2}{3}\,2\,x\,, \qquad 0 \leq x \leq 1\,, \tag{9.88}$$

we

1. draw u and v,
2. use u to choose which distribution to choose: if $u \leq 1/3$, we choose a random number with uniform distribution, $f_X(x) = 1$, else one with linear distribution $f_X(x) = 2x$,
3. and use v to draw the final number. To draw a uniformly distributed number, we just use $x = v$. For the linear distribution, we have $F_X(x) = x^2$, thus $F_X^{-1}(y) = \sqrt{y}$, and thus $x = \sqrt{v}$.

Written even more as a recipe:

1. Draw $u, v \sim \mathcal{U}(0,1)$.
2. $$\begin{cases} x = v & \text{if } u \leq \dfrac{1}{3} \\ x = \sqrt{v} & \text{if } u > \dfrac{1}{3} \end{cases} \tag{9.89}$$

This method is very powerful, as it works for an arbitrary linear superposition.

9.2.4 Discrete distributions

If we have a discrete distribution with probabilities p_k and the distribution function

$$F_X(k) = \sum_{m=0}^{k} p_k\,, \tag{9.90}$$

the following yields the desired random numbers: Obtain a uniformly distributed number u, then

$$\text{if } u \le F_X(0) \qquad \text{choose } x = 0\,, \tag{9.91}$$
$$\text{if } F_X(0) < u \le F_X(1) \qquad \text{choose } x = 1\,, \tag{9.92}$$
$$\text{if } F_X(1) < u \le F_X(2) \qquad \text{choose } x = 2\,, \tag{9.93}$$
$$\text{if } F_X(2) < u \le F_X(3) \qquad \text{choose } x = 3\,, \tag{9.94}$$
$$\vdots \tag{9.95}$$

This works because the probability that u is in a certain interval is equal to the length of this interval, and the length of the kth interval is $F_X(k) - F_X(k-1) = p_k$.

Example: For the binomial distribution $\mathcal{B}(N,p)$ with $N = 5$, p=1/3, we have

$$p_0 = \frac{32}{243}\,, p_1 = \frac{80}{243}\,, p_2 = \frac{80}{243}\,, p_3 = \frac{40}{243}\,, p_4 = \frac{10}{243}\,, p_5 = \frac{1}{243}\,, \tag{9.96}$$

and, calculating the partial sums,

$$F_X(0) = \frac{32}{243}\,, F_X(1) = \frac{112}{243}\,, F_X(2) = \frac{192}{243}\,, F_X(3) = \frac{231}{243}\,, F_X(4) = \frac{242}{243}\,, F_X(5) = \frac{243}{243}\,. \tag{9.97}$$

Thus,

$$\text{if } u \le \frac{32}{243} \qquad \text{choose } x = 0\,, \tag{9.98}$$
$$\text{if } \frac{32}{243} < u \le \frac{112}{243} \qquad \text{choose } x = 1\,, \tag{9.99}$$
$$\text{if } \frac{112}{243} < u \le \frac{192}{243} \qquad \text{choose } x = 2\,, \tag{9.100}$$
$$\text{if } \frac{192}{243} < u \le \frac{231}{243} \qquad \text{choose } x = 3\,, \tag{9.101}$$
$$\text{if } \frac{231}{243} < u \le \frac{242}{243} \qquad \text{choose } x = 4\,, \tag{9.102}$$
$$\text{if } \frac{242}{243} < u \le \frac{243}{243} \qquad \text{choose } x = 5\,. \tag{9.103}$$

9.3 The central limit theorem

The sum $\sum X_i$ of N independent random variables has the probability density function $\bigotimes_i f_{X_i}(x)$, i.e. the convolution of the individual PDFs. As convolutions are best dealt with in Fourier space, let us introduce the Fourier transform of the probability density function,

$$\chi_X(t) \equiv \int_{-\infty}^{\infty} e^{itx} f_X(x)\, dx\,, \tag{9.104}$$

which is called the *characteristic function* of X. and can be written as

$$\chi(t) = Ee^{itX} \; . \tag{9.105}$$

The latter form works also for discrete distributions.

If χ is sufficiently smooth (i.e. if f_X decays sufficiently fast), we can expand $\ln \chi$ in a Taylor series:

$$\ln \chi(t) = \sum_{n=0}^{\infty} \kappa_n \frac{(it)^n}{n!} \; . \tag{9.106}$$

Here κ_n are called the *cumulants* of the distribution, and are related to the central moments

$$\begin{aligned}
\mu &\equiv EX \; , && (9.107)\\
\mu_2 &\equiv E(X-\mu)^2 \; , && (9.108)\\
\mu_3 &\equiv E(X-\mu)^3 \; , && (9.109)\\
\mu_4 &\equiv E(X-\mu)^4 \; , && (9.110)\\
&\dots
\end{aligned}$$

via

$$\begin{aligned}
\kappa_0 &= 0 \; , && (9.111)\\
\kappa_1 &= \mu \; , && (9.112)\\
\kappa_2 &= \mu_2 \; , && (9.113)\\
\kappa_3 &= \mu_3 \; , && (9.114)\\
\kappa_4 &= \mu_4 - 3\mu_2^2 \; , && (9.115)\\
&\dots
\end{aligned}$$

For a normal distribution $\mathcal{N}(\mu, \sigma)$, we get

$$\begin{aligned}
\chi(t) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} e^{itx} \, dx && (9.116)\\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{x^2-2\mu x+\mu^2-2\sigma^2 ixt}{2\sigma^2}} \, dx && (9.117)\\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{[x-(\mu+\sigma^2 it)]^2 x-2\mu\sigma^2 it+\sigma^2 t}{2\sigma^2}} \, dx && (9.118)\\
&= e^{it\mu-\frac{\sigma^2 t}{2}} \; , && (9.119)
\end{aligned}$$

or

$$\ln \chi(t) = \mu it - \frac{\sigma^2 t}{2} \; . \tag{9.120}$$

Hence,

$$\kappa_0 = 0 , \tag{9.121}$$
$$\kappa_1 = \mu , \tag{9.122}$$
$$\kappa_2 = \sigma^2 , \tag{9.123}$$
$$\kappa_3 = 0 , \tag{9.124}$$
$$\kappa_4 = 0 , \tag{9.125}$$
$$\dots$$

If we add two random variables, $Z = X + Y$, the resulting PDF is

$$f_Z(z) = f_X \otimes f_Y = \int_{-\infty}^{\infty} f_X(x)\, f_Y(z-x)\, dx . \tag{9.126}$$

The resulting characteristic function is simply

$$\chi_Z(t) = \chi_X(t)\, \chi_Y(t) . \tag{9.127}$$

If we divide a random variable X by a number n,

$$Y = \frac{X}{n} , \tag{9.128}$$

the PDF becomes [3]

$$f_Y(y) = n f_X(ny) , \tag{9.130}$$

and the resulting characteristic function is

$$\chi_Y(t) = \int e^{ity}\, n\, f_X(ny)\, dy = \int e^{i(t/n)x}\, f_X(x)\, dx = \chi_X(t/n) . \tag{9.131}$$

So what is the characteristic function of the arithmetic mean of n random variables

$$Y = \frac{1}{n} \sum_{k=1}^{n} X_k \, ? \tag{9.132}$$

According to Eqs. (9.127) and (9.131), we need to take the product and scale the argument:

$$\chi_Y(t) = \prod_{k=1}^{n} \chi_k(t/n) . \tag{9.133}$$

[3] This follows from

$$f_X(x)\, dx = f_Y(y)\, dy . \tag{9.129}$$

We can write this as

$$\ln \chi_Y(t) = \sum_{k=1}^{n} \ln \chi_k(t/n) \,, \tag{9.134}$$

or, for the cumulants,

$$\kappa_0^{(Y)} = 0 \,, \tag{9.135}$$

$$\kappa_1^{(Y)} = \frac{\sum \kappa_1^{(i)}}{n} \,, \tag{9.136}$$

$$\kappa_2^{(Y)} = \frac{\sum \kappa_2^{(i)}}{n^2} \,, \tag{9.137}$$

$$\kappa_3^{(Y)} = \frac{\sum \kappa_3^{(i)}}{n^3} \,, \tag{9.138}$$

$$\dots$$

In terms of the moments, this means that

$$\mu^{(Y)} = \frac{\sum \mu_1^{(i)}}{n} = \langle \mu \rangle \,, \tag{9.139}$$

$$\mu_2^{(Y)} = \frac{\sum \mu_2^{(i)}}{n^2} = \frac{\langle \mu_2 \rangle}{n} \,, \tag{9.140}$$

$$\mu_3^{(Y)} = \frac{\sum \mu_3^{(i)}}{n^3} = \frac{\langle \mu_3 \rangle}{n^2} \,, \tag{9.141}$$

$$\dots \tag{9.142}$$

If n is large, we can neglect the moments higher than 2 and find that we asymptotically get a normal distribution with

$$\mu = \langle \mu \rangle \,, \qquad \sigma = \sqrt{\frac{\langle \mu_2 \rangle}{n}} \,. \tag{9.143}$$

At the same time, the *skewness* $\gamma_1 \equiv \mu_3/\sigma^3$ and higher cumulants that the normal distribution does not have tend to zero. This statement is called *central limit theorem*.

Example 1: Averaging exponentially distributed random variables. For the exponential distribution $X \sim \mathcal{E}(\alpha)$, we have

$$\chi_X(t) = \int_{x=0}^{\infty} \alpha e^{-\alpha t} e^{itx} \, dx = \frac{1}{1 - \dfrac{it}{\alpha}} \,. \tag{9.144}$$

Thus,

$$\ln \chi_X(t) = -\ln\left(1 - \frac{it}{\alpha}\right) = \frac{it}{\alpha} + \frac{(it)^2}{2\alpha^2} + \dots \tag{9.145}$$

and

$$\kappa_1 = \frac{1}{\alpha}\,, \qquad \kappa_2 = \frac{1}{\alpha^2}\,, \qquad \kappa_3 = \frac{2}{\alpha^3}\,, \qquad \dots \tag{9.146}$$

Averaging n independent exponentially distributed random variables, we get

$$\ln \chi_Y(t) = -n \ln\left(1 - \frac{it}{\alpha n}\right) = \frac{it}{\alpha} + \frac{(it)^2}{2\alpha^2 n} + \frac{(it)^3}{3\alpha^2 n^2} + \dots \tag{9.147}$$

and thus

$$\mu = \frac{1}{\alpha}\,, \qquad \sigma = \frac{1}{\alpha\sqrt{n}}\,, \qquad \mu_3 = \frac{2}{\alpha^3 n^2}\,, \qquad \dots \tag{9.148}$$

Thus for large n, the distribution of Y approaches a normal distribution:

$$X \sim \mathcal{N}\left(\frac{1}{\alpha}, \frac{1}{\alpha\sqrt{n}}\right) \text{ for } n \to \infty \tag{9.149}$$

The skewness goes like

$$\gamma_1 = \frac{2}{\sqrt{n}} \to 0 \text{ for } n \to \infty\ , \tag{9.150}$$

implying that the distribution gets more symmetric with increasing n.

Example 2: Averaging Cauchy-distributed random variables. For the Cauchy distribution, we have

$$\chi_X(t) = \int_{x=0}^{\infty} \frac{b}{\pi} \frac{1}{b^2 + x^2} e^{itx}\, dx = e^{-b|t|}\,. \tag{9.151}$$

Thus,

$$\ln \chi_X(t) = -b|t|\,. \tag{9.152}$$

For the average Y of n Cauchy-distributed variables, we find

$$\ln \chi_Y(t) = -nb\left|\frac{t}{n}\right| = -b|t| \tag{9.153}$$

— this is *unchanged*. Conclusion: averaging Cauchy-distributed random variables is useless, as it does not change (an in particular not narrow) the distribution.

Reason: None of the cumulants/moments κ_i or μ_i exists, because $\ln \chi$ has no continuous derivatives (which in turn reflects the slow decay of $f_X(x)$ for $|x| \to \infty$). Thus, the central limit theorem does not hold for this distribution.

9.4 A basic illustration of the Monte Carlo method

Recall that in many applications of Monte Carlo, the physical process is simulated directly, and *there is no need to even write down the differential equations* that describe the behavior system. The only requirement is that the physical (or mathematical) system is described by probability distribution functions (PDFs). Once the related PDFs are known, the Monte Carloe can proceed by random sampling from the PDFs.

A trivial example of the method (found in standard MC textbooks) is the estimation of the area of a circle. It is trivial because there is a well known formula which is quick and easy to use, but this example has most of the elements of more complex applications.

Step 1

Draw a square on a piece of paper the length of whose sides are the same as the diameter of the circle.

Step 2

Draw a circle in the square such that the centre of the circle and the square are the same.

Step 3

Randomly cover the surface of the square with dots, so it looks like what is shown in Figure 9.1.

Step 4

Count all the dots, then count the ones which fall inside the circle, the area of the circle is estimated thus:

$$A_{\text{circle}} = A_{\text{square}} \times \frac{\text{N dots inside circle}}{\text{N all dots}} \tag{9.154}$$

Now do the experiment many time, and each experiment gives you an estimate of the area of the circle. The result is “averaging" (in the statistical sense) over all the values (i.e. the area of the circle) you estimated. Of course, the larger the number of dots, the greater the accuracy of each estimate, and thus of the “averaged" result.

9.5 Monte Carlo integration

Monte Carlo integration is the approximate calculation of (mostly multi-dimensional) integrals by averaging over a large sample of random numbers.

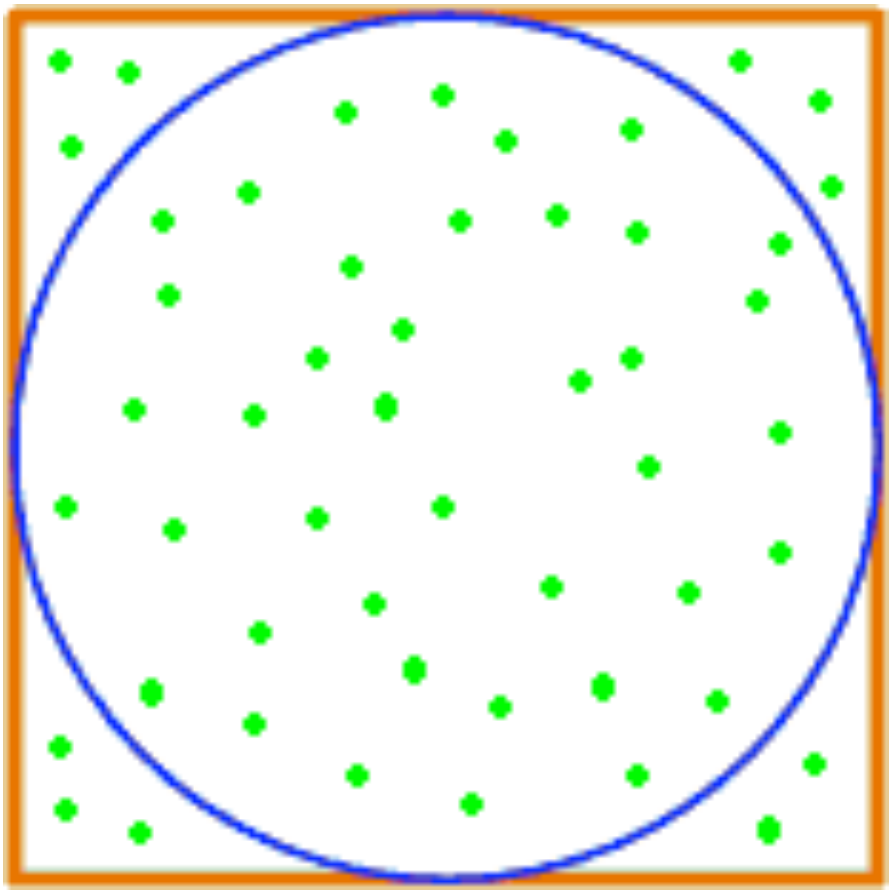


Figure 9.1: Estimation of the area of a circle by the Monte Carlo method.

Example: Let $x_i, i = 1, \ldots N$ be exponentially distributed random numbers. Then the expectation value $E \sin X$ is given by

$$E \sin X = \int_{x=0}^{\infty} \sin x \, e^{-x} \, dx \; , \tag{9.155}$$

which can be approximated by the average

$$E \sin X \approx \langle \sin x_i \rangle \equiv \frac{\sum_{i=1}^{N} \sin x_i}{N} \; . \tag{9.156}$$

The standard deviation of the error will be

$$\delta = \frac{\sigma(x_i)}{\sqrt{N-1}} = \sqrt{\frac{\left\langle \sin x_i^2 \right\rangle - \langle \sin x_i \rangle^2}{N-1}}\,, \tag{9.157}$$

thus we can write

$$\int_{x=0}^{\infty} \sin x\, e^{-x}\, dx \approx \langle \sin x_i \rangle \pm \sqrt{\frac{\left\langle \sin x_i^2 \right\rangle - \langle \sin x_i \rangle^2}{N-1}}\,. \tag{9.158}$$

Comparing with the exact value $I = 1/2$, we find values and errors as in Table 9.1.

Table 9.1: Example of Monte Carlo integration of $I \equiv \int_{x=0}^{\infty} \sin x\, e^{-x}\, dx$. The errors are for the values found in this experiment and will be different in a different realization.

N	Monte Carlo approximant I_N	error
10	0.46591	-0.017
100	0.52414	0.012
1 000	0.50023	0.00012
10 000	0.50525	0.0026
100 000	0.49846	-0.00077
1 000 000	0.49937	-0.00031
10 000 000	0.49976	-0.00012

As we can see, the error decreases quite slowly, and we know that it goes like $\delta \sim 1/N^{1/2}$. Using the trapezoid rule, we would have an error that scales like $1/N^2$, and other rules like high-order Newton–Cotes, Romberg or Gauss–Laguerre would give yet much faster convergence.

However, if we have a 10-dimensional integral, then the trapezoid rule (applied to each of the 10 directions) will have an error that scales like $1/N_x^2 = N^{1/5}$, where N is the total number of points. The error of the Monte Carlo method, on the other hand, will still go like $1/N^{1/2}$ and thus decline considerably faster. The cost will be comparable for a 4-dimensional integral. Even for higher-order methods there will be a dimensionality d starting from which Monte Carlo integration is cheaper.

Another, often more important reason for using Monte Carlo integration for multi-dimensional integrals is the fact that it is easily copes with integrals over complicated domains, while higher-order methods are essentially restricted to Cartesian products. For example to calculate the volume of the intersection of a cone with a sphere,

$$V \equiv \int_{\substack{x^2+y^2<z^2 \\ (x-1)^2+y^2+z^2<4}} dV \tag{9.159}$$

we can uniformly sample random points in a box that encloses the volume in question ($[-1,2]\times[-2,2]\times[-2,2]$ will do), and then count the points inside the volume. The fraction of positive counts will be approximately equal to the ratio of V to the volume of the whole box. This leads us to the following code:

Monte Carlo 1

```
real, dimension(3) :: a
real, dimension(2) :: xr=(/-1,2/), yr=(/-2,2/), zr=(/-2,2/)
real               :: Lx=xr(2)-xr(1), Ly=yr(2)-yr(1), Lz=zr(2)-zr(1)
integer            :: count
!
count = 0
do i=1,N
    call random_number(a)     ! [x,y,z]
    x = xr(1)+a(1)*Lx; y = yr(1)+a(2)*Ly; z = zr(1)+a(3)*Lz
    if ((x**2+yy**2 < z**2) .and. ((x-1)**2+y**2+z**2 < 4)) then
        count = count+1
    endif
enddo
print*, 'Volume ~ ', count*Lx*Ly*Lz/N
```

Calculating the integral

$$V \equiv \int_{x^2+y^2<z^2,\ (x-1)^2+y^2+z^2<4} f(\mathbf{x})\, dV \tag{9.160}$$

is almost as simple:

Monte Carlo 2

```
real, dimension(3) :: a
real, dimension(2) :: xr=(/-1,2/), yr=(/-2,2/), zr=(/-2,2/)
real               :: Lx=xr(2)-xr(1), Ly=yr(2)-yr(1), Lz=zr(2)-zr(1)
real               :: sum
!
sum = 0.
do i=1,N
    call random_number(a)     ! [x,y,z]
    x = xr(1)+a(1)*Lx; y = yr(1)+a(2)*Ly; z = zr(1)+a(3)*Lz
    if ((x**2+yy**2 < z**2) .and. ((x-1)**2+y**2+z**2 < 4)) then
        sum = sum + f(x,y,z)
    endif
enddo
print*, 'Integral ~ ', sum*Lx*Ly*Lz/N
```

Basic theorem of Monte Carlo integration:

$$\int g(\mathbf{x}) f(\mathbf{x})\, dx^d \approx \langle g(\mathbf{x}_i)\rangle \pm \sqrt{\frac{\langle g(\mathbf{x}_i)^2\rangle - \langle g(\mathbf{x}_i)\rangle^2}{N-1}}\,, \tag{9.161}$$

where x_i are d-dimensional random vectors, distributed according to the PDE $f_{\mathbf{X}}(\mathbf{x}) = f(\mathbf{x})$.

Note: One can attempt to minimize the variance of $g(x_i)$ by absorbing some of g's variability into f, thus mapping

$$f \mapsto \widetilde{f}\,, \qquad g \mapsto \widetilde{g} \tag{9.162}$$

with $\widetilde{g}\widetilde{f} = gf$. This method is called *importance sampling*, because it basically works by sampling more points where the integrand gf is larger.

In the extreme case where $\widetilde{g} = 1$, we would get $\langle g^2\rangle - \langle g\rangle^2 = 0$, and thus have no error at all. However, the resulting integral

$$\int_{\mathcal{V}} \widetilde{f}(\mathbf{x})\, dx^d = 1 \tag{9.163}$$

does not tell us anything new, and in fact we have now replaced the problem of calculating the integral $\int g(\mathbf{x}) f(\mathbf{x})\, dx^d$ by that of generating random numbers x_i that are distributed according to $\widetilde{f}(\mathbf{x}) = g(\mathbf{x}) f(\mathbf{x})$ over $\mathcal{V}$, which is a more difficult problem.

Nevertheless, a transformation of type (9.162) can sometimes dramatically reduce the statistical error.

9.5.1 Error estimates

While it is sometimes possible to directly estimate the statistical error of a Monte Carlo result, this is by no means the rule. It is however quite simple to estimate the error based on a number N of independent realizations of the Monte Carlo process. Since Monte Carlo simulations require a large number of realizations anyway, this simply means that we group our realizations in N sets.

Within each set, we determine the average $\gamma_i \equiv \langle g\rangle_i$, i=1,...,N of the quantity we are calculating, and calculate the average and standard deviation

$$\overline{\gamma} = \frac{\sum_{i=1}^{N} \gamma_i}{N}\,, \qquad \sigma^2 = \frac{\sum_{i=1}^{N} (\gamma_i - \overline{\gamma})^2}{N-1}\,. \tag{9.164}$$

The Monte Carlo estimate of $\langle g\rangle$ is then

$$\langle g\rangle = \overline{\gamma} \pm \frac{\sigma}{\sqrt{N}} \tag{9.165}$$

In practise, a value of 10..20 for N gives reasonably good results.

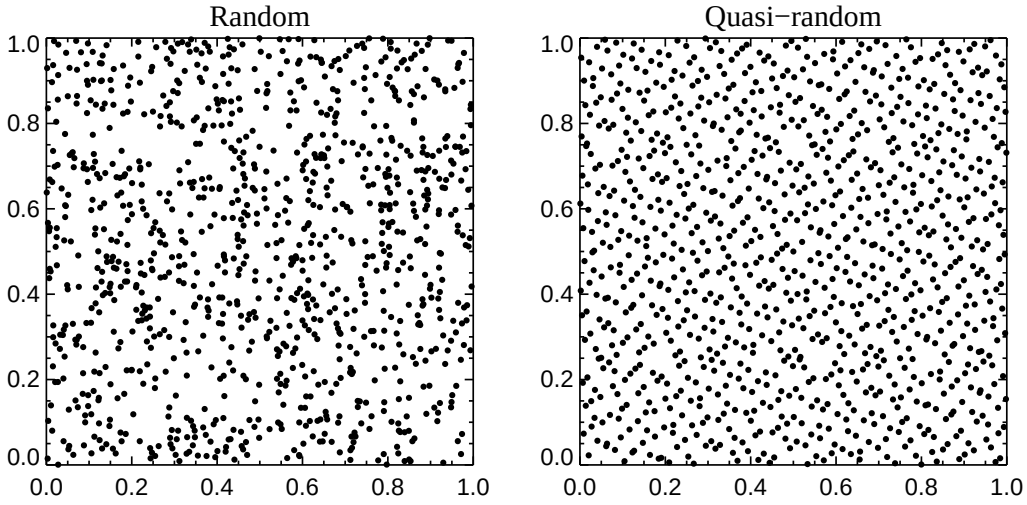


Figure 9.2: Random numbers vs. "quasi-random" numbers. The coordinates (x_i, y_i) of the points on the left are independent, uniformly distributed random numbers. The points on the right are "quasi-random" points (which are in fact completely deterministic), generated by the '`sobseq`' routine from [NR90]. In both cases, 1024 points are shown.

9.5.2 Quasi-random numbers

The fact that the error of a Monte-Carlo integral goes like $1/\sqrt{N}$ follows directly from Eq. (9.23). Using independent random numbers, we cannot obtain a better scaling.

A similar method called "Quasi-Monte Carlo method" utilizes a deterministic sequence of points. These points form "quasi-random" sequences and 'avoid each other' and thus sample the hypercube more efficiently, as can be seen in Fig. 9.2. The scaling of the error for Quasi-Monte Carlo integration is $\delta \sim \ln^d N/N$ (and thus close to $1/N$) for the integration of smooth functions over the full d-dimensional hypercube.

If we integrate over a subvolume of the hypercube, the error is be dominated by the points near the (hyper)surface of that subvolume, which 'by chance' happen to be inside (and thus counted) or outside (and thus discarded). The thickness of that layer around the hypersurface is $\sim N^{1/d}$ (this is the typical separation of points in one direction, here in the direction normal to the hypersurface), and this is equal to the fraction of points that are close to the boundary (only the direction normal to the hypersurface is selective, the other directions are irrelevant for whether a point is a boundary point).

The number of boundary points is thus $\sim N^{1/d} N = N^{(d-1)}/d$. According to the statistical 'square root law', this leads to statistical fluctuations of order $\sim N^{(d-1)/2d}$, which contribute to the integral (after dividing by N) as $1/N^{(d+1)/2d}$. This exponent is shown in Table 9.2 for different dimensionalities d. For $d = 3$, it is still markedly different from the scaling exponent 1/2 for the Monte Carlo method, thus Quasi-Monte Carlo will be worth while in 3 dimensions.

Table 9.2: Scaling exponent over dimensionality d for Quasi-Monte Carlo integration over a volume that is not a Cartesian product.

d	$-\frac{d+1}{2d}$	num. value
1	0	0.000
2	$-1/4$	-0.250
3	$-1/3$	-0.333
4	$-3/8$	-0.375
5	$-2/5$	-0.400
6	$-5/12$	-0.417
8	$-7/16$	-0.438
10	$-9/20$	-0.450
15	$-7/15$	-0.467
20	$-19/40$	-0.475

9.6 The Metropolis et al algorithm

Metropolis et al have come up with an algorithm for Monte Carlo methods that is very popular ever since. It provides a method to generate random numbers by a random process ('random walk') that are asymptotically distributed according to any given distribution. *Asymptotic* here means that the process needs some time to equilibrate before the numbers have the desired distribution. Another constraint to keep in mind is the fact that the random numbers are typically strongly correlated. For many applications this is not important, because one needs to average over many correlation 'times' anyway.

9.6.1 Random processes

A discrete *random process* is a sequence of random variables $X_1, X_2, \ldots$.

For a *Markovian random process* ('Markov chain') the distribution of X_{k+1} is completely determined by X_k (Markov chains are 'memory-less'). The dependence on the previous state is characterized by the *transition probability*

$$P(x \to x') \;=\; P(X_{n+1}{=}x' \,|X_n{=}x) \tag{9.166}$$

(for a continuous distribution[4]), or

$$P(i \to k) \;=\; P(X_{n+1}{=}k \,|X_n{=}i) \tag{9.167}$$

(for a discrete one).

[4] This is a very imprecise notation, since for a continuous distribution $P(X = x_0)$ is always zero (X has a finite probability to be in an interval, but if the interval length tends to zero, the probability does, too). But it should be clear enough how to make sense of it when needed.

Under certain, not very restrictive conditions, a Markov process leads to a stationary distribution, which is not modified by $P(i \to k)$. A straight-forward way of enforcing that the stationary distribution is $f(i)$ is to require *detailed balance*:

$$f(i)\, P(i \to k) \;=\; f(k)\, P(k \to i)\;, \tag{9.168}$$

which can also be written as

$$\frac{P(i \to k)}{P(k \to i)} \;=\; \frac{f(k)}{f(i)}\;. \tag{9.169}$$

Since both transition probabilities have to be ≤ 1, we can choose

$$P(i \to k) = \begin{cases} 1\;, & f(k) \geq f(i)\;, \\ \dfrac{f(k)}{f(i)}\;, & f(k) < f(i)\;. \end{cases} \tag{9.170}$$

"Example": In thermodynamics, we are often interested in partition sums of the form

$$Ef = \frac{\sum\limits_i f(E_i)\, e^{-\beta E_i}}{\sum\limits_i e^{-\beta E_i}} \tag{9.171}$$

(or corresponding integrals), where $\beta \equiv 1/(k_{\mathrm{B}} T)$.

To obtain a series of energies E_i with the Boltzmann distribution $p_i \sim e^{-\beta E_i}$, we can use Eq. (9.170) and get

$$P(E_i \to E_k) = \begin{cases} 1\;, & E_k \leq E_i\;, \\ e^{-\beta(E_k - E_i)}\;, & E_k > E_i\;. \end{cases} \tag{9.172}$$

To evaluate the average (9.171), we can thus proceed as follows:

1. Start with an arbitrary state i.
2. Consider switching to another state k which would lead to an energy difference $\Delta E \equiv E_k - E_i$:

$$\begin{cases} \text{if } \Delta E \leq 0 & \text{switch unconditionally;} \\ \text{if } \Delta E > 0 & \text{draw random number } u\text{, then } \begin{cases} \text{if } u < e^{-\beta \Delta E} & \text{switch;} \\ \text{if } u \geq e^{-\beta \Delta E} & \text{do not switch;} \end{cases} \end{cases} \tag{9.173}$$

   This decision procedure can be compactly expressed as:

$$\text{Switch if } u < e^{-\beta \Delta E}\;, \tag{9.174}$$

   as u will never exceed 1.

3. Repeat step 2 *many* times.
4. Evaluate Ef as $\langle f_i \rangle$ over the states i,
   (a) generously discarding the first states it took for the process to reach equilibrium;
   (b) making sure that you do count 'new' states even if they resulted from not switching — if you do not switch from a state, its probability is accordingly higher.

9.6.2 Ising model of a ferromagnet

The Ising model considers an array of spins $s_i = \pm 1$ that only interact with their nearest neighbours. The total energy is given by

$$E = -\varepsilon \sum_{\text{neighbours } (i,k)} s_i s_k \ , \tag{9.175}$$

where ε is a positive energy and summation is over all pairs of nearest neighbours (such that each combination of neighbours is counted only once). The total energy can be minimized by aligning all spins parallel with each other, which results in an energy $-\varepsilon$ per spin.

The one-dimensional Ising model shows smooth temperature dependence (the magnetization decreases, and thermal energy increases, continuously with increasing temperature).

For the one-dimensional Ising model, we find

$$E = -\varepsilon \sum_i s_i s_{i+1} \ . \tag{9.176}$$

In two or more dimensions, however, the Ising model exhibits a *phase transition*: There is a critical temperature where magnetization goes to zero and the heat capacity is infinitely large (because of latent heat).

One way of doing a Monte Carlo simulation to get the thermodynamical properties of the Ising model is to choose a spin position at random and then decide whether to flip it or not. The energy difference when flipping the ith spin from s_i to $-s_i$ is

$$\Delta E = 2\,\varepsilon\, s_i \sum_{k(i)} s_k \ , \tag{9.177}$$

where the sum is over all spins $k(i)$ that are nearest neighbours to i. In one dimension, this becomes

$$\Delta E = 2\,\varepsilon\, s_i (s_{i-1} + s_{i+1}) \ . \tag{9.178}$$

The heat capacity can be calculated from

$$c_v \propto \frac{1}{N}\left(\frac{\varepsilon}{k_{\rm B}T}\right)^2 \left(\left\langle E^2\right\rangle - \langle E\rangle^2\right) \tag{9.179}$$

and the magnetic susceptibility from

$$\chi \propto \frac{1}{N}\frac{\varepsilon}{k_{\rm B}T}\left(\left\langle M^2\right\rangle - \langle M\rangle^2\right) \ , \tag{9.180}$$

where $M \propto \sum S_i$ is the total magnetization.

Note 1: The right hand sides of Eqs. (9.179) are essentially the variances of E and M. As c_v and χ diverge at a phase transition, total energy and magnetization will be subject to huge fluctuations close to the critical temperature.

Note 2: In a finite system ($N < \infty$), the phase transition will be smoothed out. It only occurs in the limit $N \to \infty$.

Note 3: The normalization factor $1/Z \equiv 1/\sum e^{-\beta E_i}$ never needs to be evaluated.

9.6.3 Quantum Monte Carlo integration

Monte Carlo methods have various applications in quantum mechanics. One of them is the *variational quantum Monte Carlo method* for finding the ground state of many-particle problems.

For an n-particle problem, the energy of a state $\psi(\mathbf{x}) = \langle \mathbf{x}|\psi\rangle$ is

$$E = \frac{\int \psi^*(\mathbf{x})\hat{H}\psi(\mathbf{x})\, dx^{3n}}{\int \psi^*(\mathbf{x})\psi(\mathbf{x})\, dx^{3n}} \tag{9.181}$$

$$= \int \frac{\psi^*(\mathbf{x})\psi(\mathbf{x})}{\int \psi^*(\mathbf{x})\psi(\mathbf{x})\, dx^{3n}}\ \frac{1}{\psi(\mathbf{x})}\hat{H}\psi(\mathbf{x})\ dx^{3n} \tag{9.182}$$

$$= \int f(\mathbf{x})\ \frac{1}{\psi(\mathbf{x})}\hat{H}\psi(\mathbf{x})\ dx^{3n} \ , \tag{9.183}$$

where $\mathbf{x}$ is a $3n$-dimensional position vector (3 dimensions for each particle) and $\hat{H}$ is the Hamiltonian.

$$f(\mathbf{x}) \equiv \frac{\psi^*(\mathbf{x})\psi(\mathbf{x})}{\int \psi^*(\mathbf{x})\psi(\mathbf{x})\, dx^{3n}} \tag{9.184}$$

is a (joint) probability density.

Eq. (9.183) has the form (9.161) of a multidimensional Monte Carlo integral and can thus be evaluated using Monte Carlo integration.

To generate the random positions $\mathbf{X}$, we can use the Metropolis et al algorithm, starting with random coordinates $\mathbf{X}_0$ and then accepting or rejecting a change $\Delta\mathbf{X}$ in all six coordinates (alternatively, one could do one coordinate at a time).

After discarding a certain number of initial values, we calculate the average of

$$\tilde{E}_i = \frac{1}{\psi(\mathbf{x}_i)} \hat{H}\psi(\mathbf{x}_i) \,, \tag{9.185}$$

which will converge to the value of the integral (9.181).

Again, the normalization factor $1/\int |\psi(\mathbf{x})|^2 \, dx^{3n}$ is never needed, so we can work with unnormalized trial wave functions $\psi(r)$ here.

Note 1: As a rule of thumb, an acceptance ratio of ≈ 0.5 is optimal (some say 0.2 has advantages); we can tune the amplitude of $\Delta\mathbf{X}$ to achieve this. [We cannot do that for the Ising model where phase space is discrete.]

Note 2: As always with the Metropolis et al algorithm, the random positions $\mathbf{X}$ are strongly correlated. This does not affect the precision of the integral, provided we integrate over many correlation times. However, simple estimates of the statistical error will be far too optimistic, since effectively the number of independent values will not be equal to N (the number of Metropolis steps), but N/N_{corr}, where N_{corr} is the correlation 'time'.

The method described in Sec. 9.5.1 will however work, provided the size of the sets is considerably larger than the correlation 'time'; this will always be the case if you have good statistics and the number of sets is reasonably small.

Example: Consider the one-dimensional potential well

$$U(x) = \begin{cases} \infty \,, & x < 0 \,, \\ \alpha x \,, & x \geq 0 \,. \end{cases} \tag{9.186}$$

The stationary Schrödinger equation for the wave function $\psi(x)$ is

$$-\frac{\hbar^2}{2m}\psi'' + \alpha x \psi = E\psi \,, \qquad x \geq 0 \,, \tag{9.187}$$

with the boundary conditions $\psi(0) = \psi(\infty) = 0$. Here $\hbar = h/2\pi$ is Planck's constant, m the particle mass, and E is the energy of the eigenstate $\psi(x)$.

Using appropriate units, this can be written as

$$\hat{H}\psi \;\equiv\; -\psi'' + x\psi \;=\; E\psi \,, \qquad x \geq 0 \,. \tag{9.188}$$

For an arbitrary 'trial' wave function $\Phi(x)$, the energy E_0 of the ground state satisfies the inequality

$$E_0 \leq \frac{\int \Phi^*(x)\hat{H}\Phi(x)\,dx}{\int \Phi^*(x)\Phi(x)\,dx}\ , \tag{9.189}$$

and equality holds only if $\Phi(x) \propto \psi_0(x)$ is the correct eigenfunction.

Using the trial function

$$\Phi(x) = \begin{cases} Axe^{-kx}\ , & x \geq 0 \\ 0\ , & x < 0\ , \end{cases} \tag{9.190}$$

we can use the Metropolis et al algorithm to approximate the integral for any given value of k. Varying k and choosing the minimum value obtained, we get an approximation for E_0.

To apply the algorithm, we write the right-hand-side of Eq. (9.189) as

$$E = \int \frac{\Phi^*(x)\Phi(x)}{\int \Phi^*(x)\Phi(x)\,dx}\ \frac{1}{\Phi(x)}\hat{H}\Phi(x)\ dx \equiv \int f(x)g(x)\ dx\ , \tag{9.191}$$

and find that

$$f(x)\ \propto\ \Phi(x)^2\ =\ \begin{cases} x^2e^{-2kx}\ , & x \geq 0 \\ 0\ , & x < 0 \end{cases} \tag{9.192}$$

and

$$g(x)\ \equiv\ \frac{1}{x}e^{kx}\left(-\partial_x^2 + x\right)xe^{-kx}\ =\ \frac{1}{x}e^{kx}\left(-k^2x + 2k + x^2\right)e^{-kx}\ =\ -k^2 + \frac{2k}{x} + x\ . \tag{9.193}$$

We sample random positions according to $f(x)$ by doing a random walk in the following way. We draw a displacement vector Δx with components uniformly distributed[5] over the interval $L/2, L/2$, evaluate $f(x)$ at the new position $x_k = x_i + \Delta x$, and accept x_k with probability

$$p_{\text{acc}} = \min\left(1, \frac{f(x_k)}{f(x_i)}\right)\ , \tag{9.194}$$

i.e. draw a uniformly distributed random number u and

$$\text{accept } x_k \text{ if } u < \frac{f(x_k)}{f(x_i)}.$$

If the acceptance ratio r_{acc} is considerably below 0.5, we decrease the interval length L, if it is larger, we increase L until we get $r \approx 0.5$. Averaging $g(x)$ over the positions x_i (counting the same position several times if a shift was rejected), we get

$$E \approx \langle g(x_i)\rangle\ . \tag{9.195}$$

[5]We could choose other distributions, e.g. $\Delta X \sim \mathcal{N}(0, L)$, but uniformly distributed numbers are generated fastest.

Note: It turns out that distributions with non-vanishing expectation value give wrong results here. Even if $E\Delta X = 0$, the results seem to be off when the probability density function of the distribution is not symmetric. This is quite surprising and I do not understand this behaviour.

9.7 Monte Carlo in the banking business

Banks invest in high risk projects need to manage their risk. At one extreme, the performance of a bank with a single investment would be dependent on that investment, if it failed, the bank would lose money, it was a spectacular success, the bank would be highly profitable. However, by spreading its funds over several ventures, the probability of failure is reduced but the profits from the successful ones are offset by the cost of the failures.

Monte Carlo methods provide a means of modelling the behaviour of a portfolio. In the example below, a fictional bank makes between 1 and 20 investments. Historically, 50% of investments fail to create marketable products. This can be modelled with the binomial distribution, which for a given probability of success, provides an estimate of the number of success for a given number of investments. The example in Figure 9.3 shows the probability of a given number of successes, for 10 investments.

Figure 9.3: MC in banking: Probability of a given number of successes, for 10 investments.

Of those that start trading, the distribution of revenues is shown in the diagram given in Figure 9.4:

It is a perversity of nature, that the distribution of desirable outcomes are *left skewed*, i.e. the probability of a modest success is greater than that of a spectacular one, hence sales have been modelled as left skewed, whilst costs are right skewed, i.e. the probability of exceeding budgets is great, this is shown in the diagram shown in Figure 9.4:

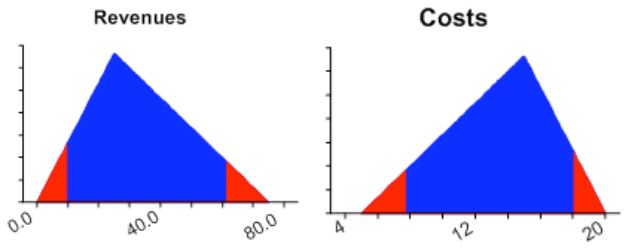


Figure 9.4: MC in banking: (left) Distribution of revenues; (right) Distribution of costs.

Using these models we can estimate the bank's ROR (Rate of Return) using the process outlined in the simplified flowchart in Figure 9.5: The value of 1,000 simulations is arbitrary, in practice the number should be appropriate to the application. For example, if an event within the process occurs infrequently, the overall number of cycles should be large enough to ensure that the results include all likely outcomes.

The results for investments in 1, 5, 10, 15 and 20 projects have been presented in the form of a line graph showing the probability that a given ROR will be exceeded. In this graph, the red line shows the probability of the ROR exceeding 0 (i.e. not making a loss). For a single investment, the probability of not making a loss is 50%, by increasing the number of projects to 20, the probability of not making a loss rises to nearly 80%. However, this reduction in risk, is offset by decreased upside (i.e. making large profits). For a single project, the probability of exceeding a 40% ROR is 42%, as the number of investments increases to 20, this figure falls to 34%.

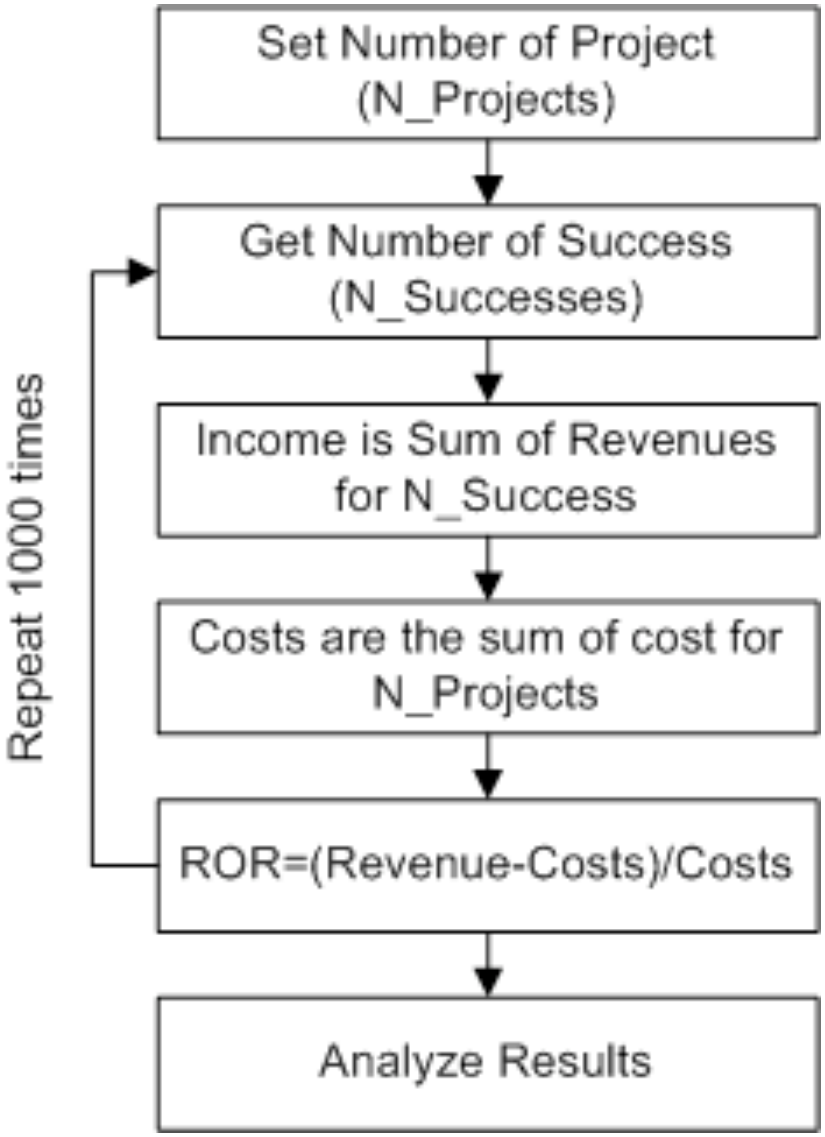


Figure 9.5: MC banking: The Rate of Return Algorithm.

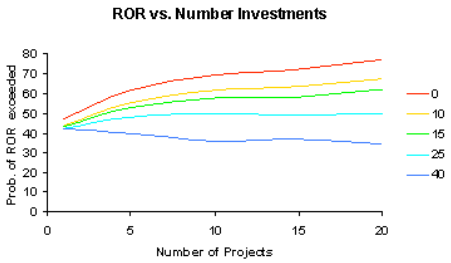


Figure 9.6: MC banking: Probability of exceeding the Rate of Return.

9.8 Appendix

9.8.1 Lab exercises

Note: It is often desirable to generate a really large number of random numbers, in order to get good statistics. If you generate all of them in one large array, you are likely to use up most or all of the system memory, which will affect not only you, but also others that are working on the system.

To see how much memory you use, you have the choice between

(a) checking from IDL with

```
print, (memory())[0]/2.^20    ; print IDL's memory usage in MB
N=400 & a=fltarr(N,N,N)
print, (memory())[0]/2.^20    ; ditto
a=0.
print, (memory())[0]/2.^20    ; ditto
```

(b) checking from the shell:

```
top
# (then press 'M' and look for the column labelled '%MEM')
```

To avoid over-using memory by accident, you can start IDL using

```
(ulimit -v 300000; idl)
```


As a long-term solution, replace the alias for idl in your '~/.bashrc' by

Question 20 *Uniformly distributed random numbers*

(a) Use the |randomu| function to generate 10 000 uniformly distributed (pseudo) random numbers u_i.

(b) Calculate the mean value and standard deviation and compare to expectation value and variance of the distribution. Does the mean get more precise if you increase the sample count?

(c) Calculate the correlation coefficient $\varrho(u_{0..499}, u_{500..999}$. What can you conclude about the quality of the random number generator?

(d) Using |histogram()| and a plot style |PSYM=10|, plot a histogram of the PDF for your random numbers. Make sure the x axis and the normalization are correct.

(e) Write a function |randomu2(seed,rand,x0,width)| that implements the uniform distribution with offset x_0 and width w. Use it to plot the histogram of the PDF for $\mathcal{U}(-3, 5)$.

(f) Plot histograms of the sum of two, three and four uniformly distributed random numbers.

Question 21 *Normal distribution*

(a) Write a function |randomnormal()| that returns an array of (pseudo) random numbers x_i that are distributed according to the normal distribution $\mathcal{N}(0, 1)$ (use the special transformation method from the lecture).

(b) Determine the average and the standard deviation and plot a histogram of the PDF.

(c) Compare these results to random numbers obtained using the built-in function |randomn()|.

Question 22 *Superposition method*

We want to generate random numbers X with the probability density function

$$f_X(x) = \frac{3}{8}(1 + x^2)\, x \qquad -1 \le x \le 1\ . \tag{9.196}$$

(a) Can we use the transformation method?

(b) Use the superposition method to write a generator for this distribution. Plot a histogram of the PDF.

(c) Now use the acceptance-rejection method with a constant comparison function to write another generator.

Question 23 *Birthday problem*

In a group of 25 people with randomly distributed (and independent) birthdays, how large is the probability that two of the birthdays are on the same day?

[Do this numerically]

Question 24 *Monte Carlo integration*

(a) Use Monte Carlo integration to calculate the volume of the unit sphere.

(b) Now generalize your code to calculate the d-hypervolume of the d-dimensional unit hypersphere for $d = 1, 2, 3, \ldots, 10$.

Question 25 *Quasi-random numbers*

Note: Quasi-ransom numbers are generated by the function `sobseq()`. To learn how to call that function, you can do `dummy=sobseq(/HELP)`.

(a) Use Quasi-Monte Carlo integration to calculate the area of the unit circle. Compare the accuracy to Monte Carlo integration.

(b) Plot the first 128 quasi-random points in two dimensions (don't forget to reset the generator). Now overplot points number 129 to 512 in another colour, and then points513 to 1024 in yet another colour.

Repeat this with uniformly distributed random numbers and compare.

(c) Verify the isotropy of quasi-random numbers on $[-1, 1] \times [-1, 1]$ by counting the numbers in annular segments

$$0.8 < r < 1 \,, \qquad \varphi_i < \varphi \leq \varphi_i + \delta\varphi$$

where you need to choose the number of points and the size $\delta\varphi$ of the angular bins appropriately.

Repeat this with uniformly distributed random numbers.

Now use the same diagnostics, but choose (quasi) random points from $[-1, 1] \times [-4, 4]$; again, do this for quasi random numbers and for uniformly distributed random numbers. Conclusion?

Question 26 *Random walk in 1 dimension*

A particle starts at $x = 0$ for $t = 0$ and then moves by Δ during the time interval δt. The values of Δ for different time steps are stochastically independent.

We expect the probability density function of the x-position X to spread according to the diffusion equation

$$\frac{\partial f}{\partial t} = D\frac{\partial^2 f}{\partial t^2} \ ,$$

which has the Green's function

$$G(x-x', t-t') = \frac{1}{\sqrt{4\pi D(t-t')}}\, e^{-\frac{(x-x')^2}{4D(t-t')}} \tag{9.197}$$

(a) Assume $\Delta = \pm 1$, where the sign is randomly chosen for each time step. Plot the probability density function after $N_t = 1000$ steps and try to fit a Gaussian (9.197) to it. What is the diffusion constant D? How can you test whether the broadening of the PDF is really described by Eq. (9.197)?

(b) Repeat part (a) for $\Delta \sim \mathcal{U}(-0.5, 0.5)$.

(c) Returning to $\Delta = \pm 1$, try to obtain the PDF of the *return time*, i.e. the time it takes for one particle to return to the origin for the first time. Will the average return time be finite?

Question 27 *One-dimensional Ising model*

Following the outline given in the lecture, program the one-dimensional Ising model

$$E = -\varepsilon \sum_{i=0}^{N-1} s_i(s_{i-1}+s_{i+1})$$

for N spins. Assume periodic boundary conditions, i.e. $s_N = s_0$, $s_{-1} = s_{N-1}$.

(a) Using $N = 20$ points, plot energy per spin E/N and magnetization per spin $M/N = \sum s_i/N$ as functions of temperature T.

(b) Plot the specific heat c_v and the magnetic susceptibility χ as functions of temperature T.

(c) If you double the number of points, do you get markedly different results?

(d) What are the options to vectorize the calculation?

Hints:

- Do not forget to discard a number of early values while the system tries to find its equilibrium. The lower the temperature, the longer equilibration will take.
- The IDL function |shift(|*vector*|,|*n*|)| takes a vector and performs a cyclic shift of its elements by n.

Question 28 *Quantum Monte Carlo*

Using the trial wave function

$$\Phi(x) = Axe^{-kx} \,, \tag{9.198}$$

find a Monte Carlo approximation E to the energy E_0 of the ground state for a particle in the potential well

$$U(x) = \begin{cases} \infty \,, & x < 0 \,, \\ x \,, & x \geq 0 \,. \end{cases} \tag{9.199}$$

(a) Find $E(k)$ for a set of values of the parameter k, at least $\{0.5, 1, 1.5, 2\}$.

(b) If you choose a normal distribution for the displacement instead of a uniform one, do the energies change? Can you explain?