

Contents

1	INTRODUCTION	1
2	SAMPLE SPACE, EVENTS AND PROBABILITIES	2
3	RANDOM VARIABLES	3
4	MEAN, VARIANCE AND COVARIANCE	11
5	CHARACTERISTIC FUNCTION	19
6	CHEBYSHEV INEQUALITY	23
7	LAW OF LARGE NUMBERS	24
8	CENTRAL LIMIT THEOREM (CLT)	25
8.1	Lévy Distributions	27
8.2	Illustration of the CLT	28
9	RANDOM NUMBERS	32
9.1	Tests for Randomness	33
9.2	Pseudo Random Number Generators	36
10	RANDOM SAMPLING TECHNIQUES	42
10.1	Inversion Technique	42
10.2	Rejection Technique	45
10.3	Sampling from a Gaussian	47
10.4	Metropolis Sampling	49
11	ANALOGUE MONTE CARLO	55
12	IMPORTANCE SAMPLING	57
12.1	Exponential Biasing	58
12.2	Spanier Technique	61
13	CLOSING REMARKS	62

1 INTRODUCTION

Monte Carlo is a powerful numerical technique that makes use of random numbers to solve a problem. I assume we all know what random numbers are. However this issue is, by no means, trivial and I shall have something to say on this later.

Historically, the first large scale Monte Carlo work carried out dates back to the middle of the twentieth century. This work pertained to studies of neutron multiplication, scattering, propagation and eventual absorption in a medium or leakage from it. Ulam, von Neumann and Fermi were the first to propose and employ the Monte Carlo method as a viable numerical technique for solving practical problems.

There were of course several isolated and perhaps not fully developed instances earlier, when Monte Carlo has been used in some form or the other. An example is the experiment performed in the middle of the nineteenth century, consisting of throwing a needle randomly on a board notched with parallel lines, and inferring the value of π from the number of times the needle intersects a line; this is known as Buffon's needle problem, see for example [1].

The Quincunx constructed by Galton [2] toward the end of the nineteenth century, consisted of balls rolling down an array of pins (which deflect the balls randomly to their left or right) and getting collected in the vertical compartments placed at the bottom. The heights of the balls in the compartments approximate the binomial distribution. This Monte Carlo experiment is a simple demonstration of the Central Limit Theorem.

In the nineteen twenties, Karl Pearson perceived the use of random numbers for solving complex problems in probability theory and statistics that were not amenable to exact solutions. Pearson encouraged L. H. C. Tippett to produce a table of random numbers to help in such studies, and a book of random sampling numbers [3] was published in the year 1927. This was followed by another publication of random numbers by R. A. Fisher and F. Yates. Pearson and his students used this method to obtain the distributions of several complex statistics.

In India, P. C. Mahalanobis [4] exploited 'random sampling' technique to solve a variety problems like the choice of optimum sampling plans in survey work, choice of optimum size and shape of plots in experimental work *etc.*, see [5].

Indeed, descriptions of several modern Monte Carlo techniques appear in a paper by Kelvin [6], written nearly hundred years ago, in the context of a discussion on the Boltzmann equation. But Kelvin was more interested in the results than in the technique, which to him was *obvious!*

Monte Carlo technique derives its name from a game very popular in Monaco. The children get together at the beach and throw pebbles at random on a square which has a circle inscribed in it. From the fraction of the pebbles that fall inside the circle, one can estimate the value of π . This way of estimating the value of π goes under the name rejection technique. A delightful variant of the game is discussed by Krauth [7]; this variant brings out in a simple way the essential principle behind the Metropolis rejection method. We shall discuss the rejection technique as well as the Metropolis rejection technique later.

In this monograph I shall define certain minimal statistical terms and invoke some important results in mathematical statistics to lay a foundation for the Monte Carlo

methods. I shall try to make the presentation as simple and as complete as possible.

2 SAMPLE SPACE, EVENTS AND PROBABILITIES

Consider an experiment, real or imagined, which leads to more than one outcome. We collect all the possible outcomes of the experiment in a set Ω and call it the **sample space**. An outcome is often denoted by the symbol ω . Certain subsets of Ω are called **events**. The class of all events is denoted by \mathcal{F} . For every pair of events $\mathcal{A}_1 \in \mathcal{F}$ and $\mathcal{A}_2 \in \mathcal{F}$, if $\mathcal{A}_1 \cup \mathcal{A}_2$, $\mathcal{A}_1 \cap \mathcal{A}_2$, and $\mathcal{A}_1 - \mathcal{A}_2$ are also events $\in \mathcal{F}$, then \mathcal{F} is called a **field**. To each event \mathcal{A} we assign a real number $0 \leq \mathcal{P}(\mathcal{A}) \leq 1$ called the **probability** of the event. If Ω consists of infinitely many outcomes, we demand that $\cup_{i=1}^{\infty} \mathcal{A}_i$ and $\cap_{i=1}^{\infty} \mathcal{A}_i$ be also events and form a **Borel field** of all possible events which includes the null event ϕ . We have $\mathcal{P}(\phi) = 0$, $\mathcal{P}(\Omega) = 1$ and $\mathcal{P}(\mathcal{A}_1 \cup \mathcal{A}_2) = \mathcal{P}(\mathcal{A}_1) + \mathcal{P}(\mathcal{A}_2) - \mathcal{P}(\mathcal{A}_1 \cap \mathcal{A}_2)$. Events \mathcal{A}_1 and \mathcal{A}_2 are **disjoint** (or **mutually exclusive**) if $\mathcal{A}_1 \cap \mathcal{A}_2 = \phi$.

Not all the subsets of Ω shall be events. One reason for this is that we may wish to assign probabilities to only some of the subsets. Another reason is of mathematical nature: we may not be able to assign probabilities to some of the subsets of Ω at all. The distinction between subsets of Ω and events and the consequent concept of Borel field are stated here simply for completeness. In applications, all *reasonably defined* subsets of Ω would be events.

How do we attach probabilities to the events ?

Following the classical definition, we say that $\mathcal{P}(\mathcal{A})$ is the ratio of the number of outcomes in the event \mathcal{A} to that in Ω , provided all *the outcomes are equally likely*. To think of it, this is the method we adopt intuitively to assign probabilities - like when we say the probability for heads in a toss of a coin is half; the probability for the number, say one, to show up in a roll of a die is one-sixth; or when we say that all micro states are equally probable while formulating statistical mechanics. An interesting problem in this context is the Bertrand paradox [8]; you are asked to find the probability for a randomly chosen chord in a circle of radius r to have a length exceeding $r\sqrt{3}$. You can get three answers $1/2$, $1/3$, and $1/4$, depending upon the experiment you design to select a chord randomly. An excellent discussion of the Bertrand paradox can be found in [9].

Assignment 1

(A) A fine needle of length $2a$ is dropped at random on a board covered with parallel lines with distance $2b$ apart. Show that the probability that the needle intersects one of the lines equals $2a/\pi b$. See page 131 of [9].

(B) Devise an experiment to select randomly a chord in a circle of radius r . What is the probability that its length exceeds $r\sqrt{3}$? See page 9-10 of [9]. What is the probability distribution of the chord length?

Alternately, we can take an operational approach. $\mathcal{P}(\mathcal{A})$ is obtained by observing the frequency of occurrence of \mathcal{A} : repeat the experiment some N times and let $N_{\mathcal{A}}$ be the number of times the event \mathcal{A} occurs. Then $N_{\mathcal{A}}/N$ in the limit of $N \rightarrow \infty$ gives the probability of the event.

As seen above, the formal study of probability theory requires three distinct notions namely the sample space Ω , the (Borel) field \mathcal{F} and the probability measure \mathcal{P} . See Papoulis [9] and Feller [10]. The physicists however, use a different but single notion, namely the **ensemble**. Consider for example a sample space that contains discrete outcomes. An ensemble is a collection whose members are the elements of the sample space but repeated as many times as would reflect their probabilities. Every member of Ω finds a place in the ensemble and every member of the ensemble is some element of the sample space. The number of times a given element (outcome) of the sample space occurs in the ensemble is such that the ratio of this number to the total number of members in the ensemble is, exactly, the probability associated with the outcome. The number of elements in the ensemble is strictly infinity.

The molecules of a gas in equilibrium is a simple example of an ensemble; the speeds of the molecules at any instant of time have Maxwell-Boltzmann distribution, denoted by $p(v)$. Each molecule can then be thought of as a member of the ensemble; the number of molecules with speeds between v_1 and v_2 , divided by the total number of molecules N in the gas is $\int_{v_1}^{v_2} p(v)dv$.

3 RANDOM VARIABLES

The next important concept in the theory of probability is the definition of a (real) **random variable**. A random variable, denoted by $X(\omega)$, is a set function: it attaches a real number x to an outcome ω . A random variable thus maps the abstract outcomes to the numbers on the real line. It stamps each outcome, so to say, with a real number.

Consider an experiment whose outcomes are **discrete** and are denoted by $\{\omega_i\}$, with i running from 1 to say N . Let $X(\omega)$ be the random variable that maps the abstract outcomes $\{\omega_i\}$ to real numbers $\{x_i\}$. Then $p_i = \mathcal{P}(\omega|X(\omega) = x_i)$ gives the probability that the random variable $X(\omega)$ takes a real value x_i . The probabilities $\{p_i\}$ obey the

conditions,

$$0 \leq p_i \leq 1, \forall i, \quad \sum_{i=1}^N p_i = 1. \quad (1)$$

Let me illustrate the above ideas through a few examples.

Tossing of a single coin

The simplest example is the tossing of a single coin. The sample space is $\Omega = \{H, T\}$, where H denotes the outcome Heads and T the Tails. There are four possible events :

$$\mathcal{F} = \{\{H\}, \{T\}, \{H, T\}, \{\phi\}\}.$$

The corresponding probabilities are $1/2$, $1/2$, 1 and 0 for a fair coin.

We can define a random variable $X(\omega)$ by attaching $+1$ to H and -1 to T , see Table (1). Then we say that the probability for the random variable X to take a value $+1$ is $\mathcal{P}[\omega|X(\omega) = 1] = 1/2$ and that for -1 is half. This defines the discrete probability distribution: $p(1) = 1/2$, $p(-1) = 1/2$, see Table (1).

Table 1: Random variable and probability for the toss of a single coin.

ω	$x = X(\omega)$	$\mathcal{P}[\omega X(\omega) = x]$
H	+1	1/2
T	-1	1/2

The physicists' ensemble containing \mathcal{N} members would be such that half of \mathcal{N} are Heads and half Tails. If the probability of Heads is $3/4$ and of Tails is $1/4$, then the corresponding ensemble would contain $3\mathcal{N}/4$ heads and $\mathcal{N}/4$ Tails.

Tossing of several coins

Consider now the tossing of N independent fair coins. Let X_1, X_2, \dots, X_N be the corresponding random variables. $X_i(H) = +1$ and $X_i(T) = -1 \forall i = 1, 2, \dots, N$. For each coin there are two possible (equally likely) outcomes. Therefore, for N coins there are 2^N possible (equally likely) outcomes. For the N -coin-tossing experiment each outcome is a distinct string of H and T , the string length being N .

Table 2: Random variable and Probability for the three - coin -tossing experiment

ω	$x = Y_3(\omega)$	$\mathcal{P} [\omega Y_3(\omega) = x]$
HHH	+3	1/8
HHT HTH THH	+1	3/8
TTH THT HTT	-1	3/8
TTT	-3	1/8

An example with $N = 3$ is shown in Table (2), where the random variable $Y_3(\omega)$ is defined by the sum: $Y_3 = X_1 + X_2 + X_3$.

The probability for each outcome (string) in the throw of N coins is thus 2^{-N} . Let us define a random variable $Y_N = X_1 + X_2 + \dots + X_N$. The probability for n heads in a toss of N coins, which is the same as the probability for the random variable Y_N to take a value $n - (N - n)$ is given by the **binomial** distribution,

$$P(n, N) \equiv \mathcal{P}[Y_N = 2n - N] = \frac{1}{2^N} \frac{N!}{n!(N - n)!}. \quad (2)$$

Fig. 1 depicts the binomial distribution for $N = 6$.

For the general case of a loaded coin, with probability for Heads as p and for the Tails as $q = 1 - p$, we have,

$$P(n, N) \equiv \mathcal{P}[Y_N = 2n - N] = \frac{N!}{n! (N - n)!} p^n q^{N-n}. \quad (3)$$

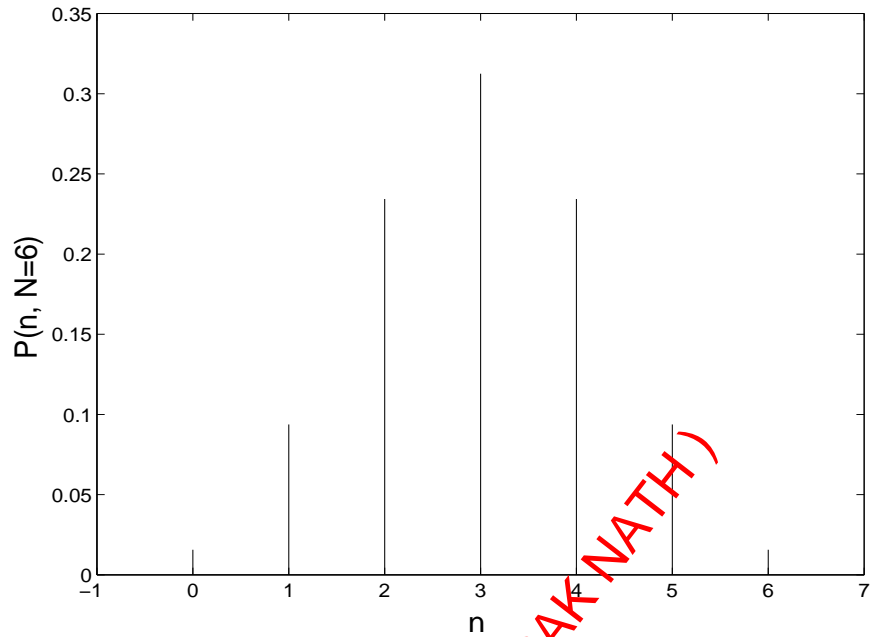


Figure 1: Binomial distribution see Eq. (2) with $N = 6$

Rolling of a fair die

Another simple example consists of rolling a die. There are six possible outcomes. The sample space is given by,

$$\Omega =$$

$$\left\{ \begin{array}{cccccc} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{array} \right\}$$

The random variable $X(\omega)$ attaches the numbers $\{x_i : i = 1, 2, \dots, 6\} \equiv \{1, 2, 3, 4, 5, 6\}$ to the outcomes $\{\omega_i : i = 1, 2, \dots, 6\}$, in the same order. If the die is not loaded we have $p(x_i) = 1/6 \forall i = 1, 2, \dots, 6$. An example of an event, which is the subset of Ω , is

$$\mathcal{A} = \left\{ \begin{array}{cccc} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{array} \right\},$$

and $\mathcal{P}(\mathcal{A}) = 1/2$. This event corresponds to the roll of an odd number. Consider the event

$$\mathcal{B} = \left\{ \begin{array}{cccc} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{array} \right\},$$

which corresponds to the roll of an even number. It is clear that the events \mathcal{A} and \mathcal{B} can not happen simultaneously; in other words a single roll of the die can not lead to both the

events \mathcal{A} and \mathcal{B} . The events \mathcal{A} and \mathcal{B} are **disjoint** (or **mutually exclusive**). If we define another event

$$\mathcal{C} = \left\{ \cdot, \cdot, \cdot, \cdot, \cdot, \cdot \right\},$$

then it is clear that \mathcal{A} and \mathcal{C} are not disjoint. We have

$$\mathcal{A} \cap \mathcal{C} = \left\{ \cdot, \cdot, \cdot, \cdot \right\}.$$

Similarly the events \mathcal{B} and \mathcal{C} are not disjoint:

$$\mathcal{B} \cap \mathcal{C} = \left\{ \cdot, \cdot \right\}.$$

Assignment 2

Consider rolling of a fair die N times. Let n_k denote the number of times the number k shows up and k runs from 1 to 6. Find an expression for the probability $P(n_1, n_2, n_3, n_4, n_5, n_6, N)$. What is the probability distribution of the random variable $n = \sum_{k=1}^6 k n_k$?

Poisson distribution

An important discrete distribution, called the **Poisson** distribution, arises in the context of random occurrences in time. In an interval $\Delta t \rightarrow 0$ positioned at any time t , there is either one occurrence with probability $\lambda \Delta t$, or no occurrence with probability $1 - \lambda \Delta t$. Here λ^{-1} is the characteristic time constant of the Poisson process.

Let $P(n, t)$ denote the probability that there are n occurrences in the interval $[0, t]$. A **master equation** can be readily written down as,

$$\begin{aligned} P(n, t) &= \lambda \Delta t P(n-1, t-\Delta t) + [1 - \lambda \Delta t] P(n, t-\Delta t), \\ P(n, t=0) &= \delta_{n,0}. \end{aligned} \tag{4}$$

To solve for $P(n, t)$, define the **generating function**,

$$\tilde{P}(z, t) = \sum_{n=0}^{\infty} z^n P(n, t). \tag{5}$$

Multiplying both sides of Eq. (4) by z^n and summing over n from 0 to ∞ , we get,

$$\begin{aligned} \tilde{P}(z, t) &= \lambda z \Delta t \tilde{P}(z, t-\Delta t) + [1 - \lambda \Delta t] \tilde{P}(z, t-\Delta t), \\ \tilde{P}(z, t=0) &= 1. \end{aligned} \tag{6}$$

In the limit $\Delta t \rightarrow 0$, we get from the above

$$\frac{\partial \tilde{P}}{\partial t} = -\lambda(1-z)\tilde{P}(z,t), \quad (7)$$

whose solution is,

$$\tilde{P}(z,t) = \exp[-\lambda(1-z)t], \quad (8)$$

since the initial condition,

$$\begin{aligned} \tilde{P}(z,t=0) &= \sum_{n=0}^{\infty} z^n P(n,t=0) \\ &= \sum_{n=0}^{\infty} z^n \delta_{n,0} \\ &= 1. \end{aligned} \quad (9)$$

Taylor expanding the right hand side of Eq. (8), we get $P(n,t)$ as the coefficient of z^n ,

$$P(n,t) = \frac{(\lambda t)^n}{n!} \exp(-\lambda t). \quad (10)$$

Fig. 2 depicts the Poisson distribution.

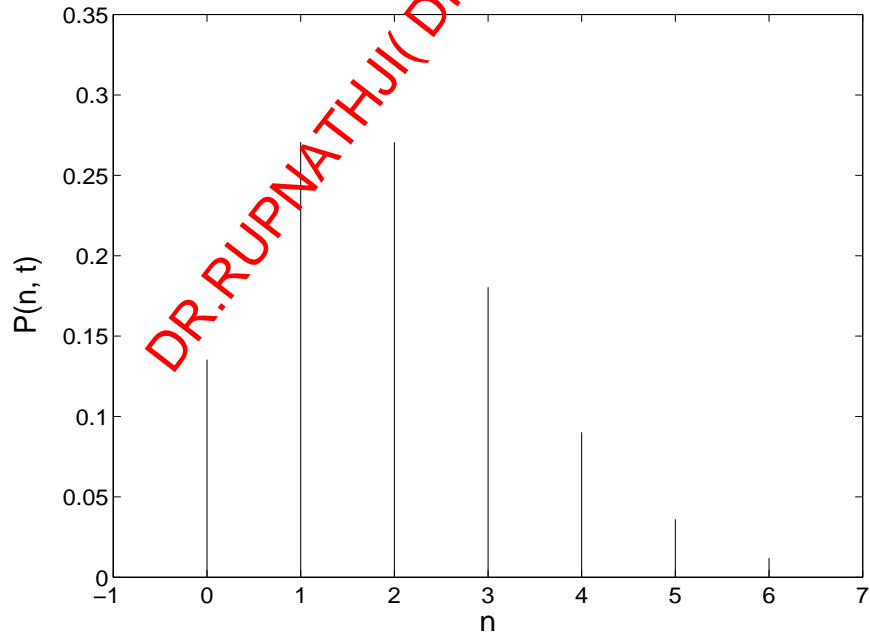


Figure 2: Poisson distribution, see Eq. (10) with $\lambda t = 2$.

Continuous distributions

What we have seen above are a few examples of discrete random variable that can take either a finite number or at most countable number of possible values. Often we have to consider random variables that can take values in an interval. The sample space is a **continuum**. Then we say that $f(x)dx$ is the probability of the event for which the random variable $X(\omega)$ takes a value between x and $x+dx$. *i.e.* $f(x)dx = \mathcal{P}[\omega|x \leq X(\omega) \leq x+dx]$. We call $f(x)$ the **probability density function** or the **probability distribution function**.

Uniform distribution

An example is the **uniform** random variable, $U(a, b)$, defined in the interval $[a, b]$. The probability density function, $f(x)$ of the random variable $U(a, b)$ is defined by $f(x)dx = dx/(b-a)$. We shall come across the random variable $U(0, 1)$ later in the context of pseudo random numbers and their generators.

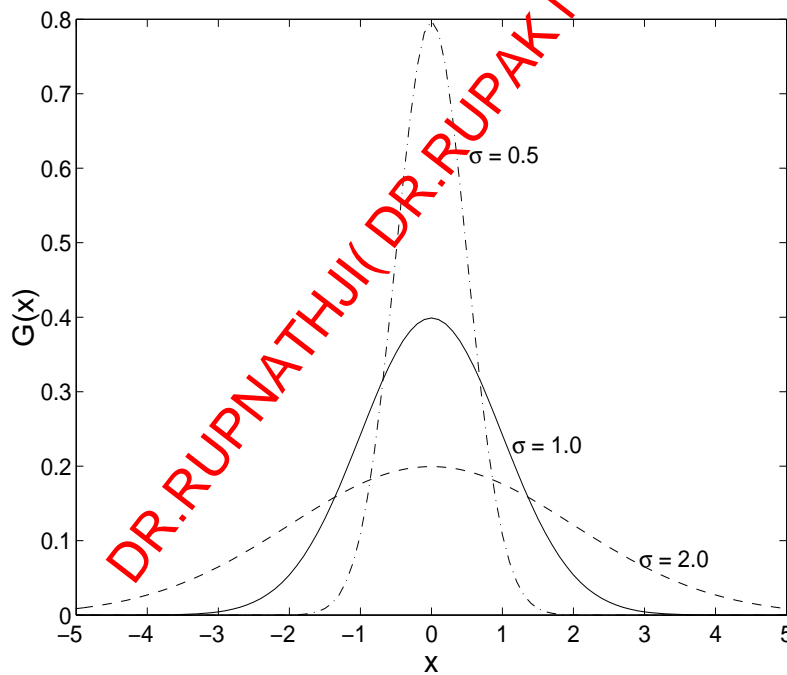


Figure 3: Gaussian distribution, see Eq. (11) with $\mu = 0$, and $\sigma = .5$ (---), 1.0 (-) and 2.0 (- -)

Gaussian distribution

A probability density function we would often come across is the **Gaussian**. This density function is of fundamental importance in several physical and mathematical applications.

It is given by,

$$G(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \quad -\infty < x < +\infty, \quad (11)$$

where μ and σ are the parameters, called the mean and standard deviation. Fig. 3 depicts the Gaussian density for $\mu = 0$ and $\sigma = 0.5, 1.0$ and 1.5 . The Gaussian density function plays a central role in the estimation of statistical errors in Monte Carlo simulation, as we would see later.

Exponential distribution:

The **exponential** density arises in the context of several physical phenomena. The time taken by a radioactive nucleus to decay is exponentially distributed. The distance a gamma ray photon travels in a medium before absorption or scattering is exponentially distributed. The exponential distribution is given by,

$$f(x) = \begin{cases} \alpha e^{-\alpha x}, & \text{for } x \geq 0 \\ 0, & \text{for } x < 0 \end{cases} \quad (12)$$

where $\alpha > 0$ is a parameter of the exponential distribution. Fig. 4 depicts the exponential distribution, with $\alpha = 1$.

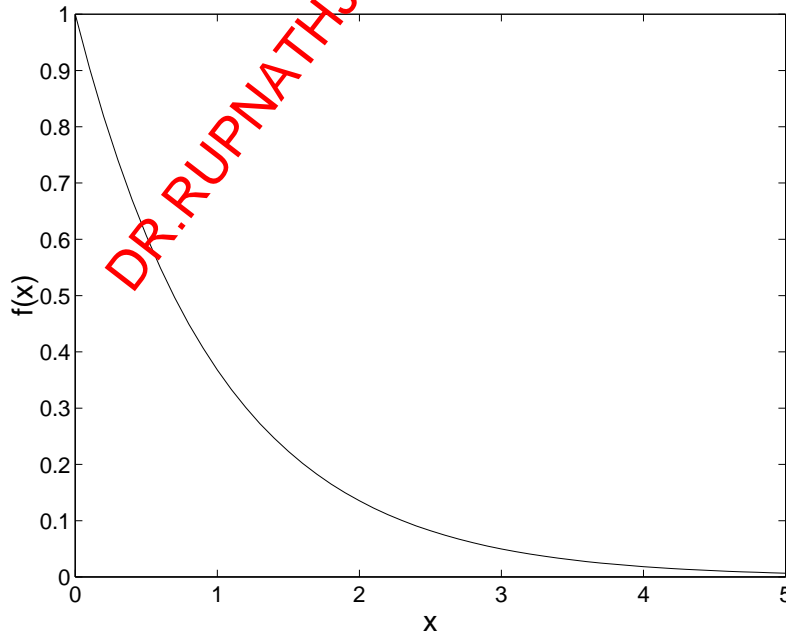


Figure 4: Exponential distribution, see Eq. (12) with $\alpha = 1$.

Cauchy distribution

An interesting distribution named after Cauchy is given by,

$$f(x) = \frac{1}{\pi} \frac{D}{D^2 + x^2}, \quad -\infty < x < +\infty, \quad (13)$$

where $D > 0$ is a scale factor. Fig. 5 depicts the **Cauchy** distribution with $D = 1$.

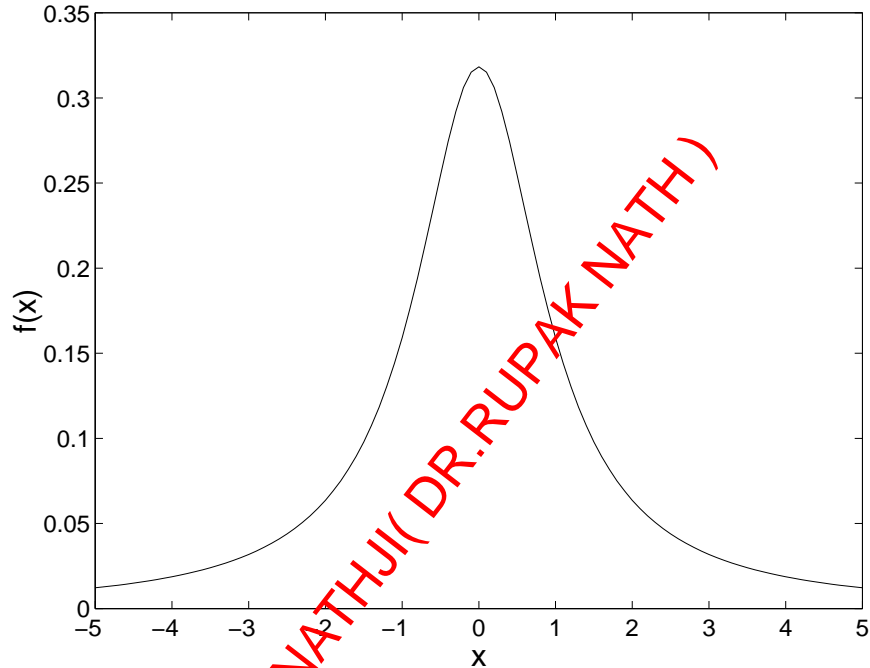


Figure 5: Cauchy distribution, see Eq. (13) with $D = 1$.

If a random variable θ is uniformly distributed in the range $-\pi/2$ to $+\pi/2$, then $x = \tan(\theta)$ follows a Cauchy distribution ($D = 1$). If X_1 and X_2 are independent Gaussian random variables each with mean zero and standard deviation σ_1 and σ_2 respectively, then the ratio $X = X_1/X_2$ is Cauchy distributed with $D = \sigma_2/\sigma_1$. The Cauchy distribution is also known in the literature as **Lorentz** distribution or **Breit-Wigner** distribution and arises in the context of resonance line shapes.

4 MEAN, VARIANCE AND COVARIANCE

Mean

The **mean** (also called the expectation value, the average, or the first moment) of a random variable $X(\omega)$, with a probability density $f(x)$ is denoted by the symbol μ and is defined

as,

$$\mu(X) = \int_{-\infty}^{+\infty} x f(x) dx. \quad (14)$$

Mean is the most important single parameter in the theory of probability; indeed in Monte Carlo calculations we would be interested in calculating the mean of some random variable, defined explicitly or implicitly. Let us repeat the experiment N times and observe that the outcomes are $\omega_1, \omega_2, \dots, \omega_N$. Let the value of the random variable $X(\omega)$ corresponding to these outcomes be x_1, x_2, \dots, x_N respectively. If N is sufficiently large then $(\sum_{k=1}^N x_k)/N$ approximately equals μ . In the limit $N \rightarrow \infty$ the arithmetic mean of the results of N experiments converges to μ .

Statistical convergence

At this point it is worthwhile discussing the meaning of **statistical convergence** *vis-a-vis* deterministic convergence that we are familiar with. A given sequence $\{A(\nu); \nu = 1, 2, \dots\}$ is said to converge to a value, say B as $\nu \rightarrow \infty$, if for (an arbitrarily small) $\delta > 0$, we can find an M such that for all $k \geq M$, $A(k)$ is *guaranteed* to be within δ of B . In the statistical context the term *guarantee* is replaced by a statement of probability, and the corresponding definition of convergence becomes: $A(\nu)$ is said to converge to B as $\nu \rightarrow \infty$, if given probability $0 < P < 1$ and (an arbitrarily small) $\delta > 0$, we can find an M such that for all $k \geq M$, the probability that $A(k)$ is within δ of B is greater than P . *This risk that convergence can only be assured with a certain probability is an inherent feature of all Monte Carlo calculations.* We shall have more to say on this issue when we take up discussions on the Chebyshev inequality, the law of large numbers and eventually the central limit theorem which help us appreciate the nature and content of Monte Carlo errors. For the moment it suffices to say that it is possible to make a quantitative probabilistic statement about how close is the arithmetic mean, $A(\nu)$ of ν experiments to the actual mean μ , for large ν ; such a statistical error estimate would depend on the number of experiments (ν) and the value of the second most important parameter in the theory of probability namely, the **variance** of the random variable, X , underlying the experiment.

Variance

Variance, σ^2 is defined as the expectation value of $(X - \mu)^2$. Formally we have,

$$\sigma^2(X) = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx. \quad (15)$$

The square root of the variance is called the **standard deviation**.

Moments

We can define, what are called the moments of the random variable. The **K^{th} moment**

is denoted by M_K , and is defined as,

$$M_K = \int_{-\infty}^{+\infty} x^K f(x) dx. \quad (16)$$

It is clear that $M_0 = 1$, which implies that $f(x)$ is normalized to unity, $M_1 = \mu$, and $\sigma^2 = M_2 - M_1^2$.

Cumulative probability distribution

The **cumulative probability density function**, denoted by $F(x)$ is defined as

$$F(x) = \mathcal{P}[\omega | X(\omega) \leq x] = \int_{-\infty}^x f(x') dx'. \quad (17)$$

$F(x)$ is a monotonic non-decreasing function of x ; $F(-\infty) = 0$ and $F(\infty) = 1$.

Sum of two random variables

Let us consider two random variables X_1 and X_2 . Let $Y_2 = X_1 + X_2$. We have,

$$\mu(Y_2) = \int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^{+\infty} dx_2 (x_1 + x_2) f(x_1, x_2), \quad (18)$$

where $f(x_1, x_2)$ denotes the **joint density** of the random variables X_1 and X_2 . Let $f_1(x)$ and $f_2(x)$ denote the probability density functions of X_1 and X_2 respectively. These are called **marginal densities** and are obtained from the joint density as follows.

$$\begin{aligned} f_1(x_1) &= \int_{-\infty}^{+\infty} dx_2 f(x_1, x_2). \\ f_2(x_2) &= \int_{-\infty}^{+\infty} dx_1 f(x_1, x_2). \end{aligned} \quad (19)$$

The integral in Eq. (18) can be evaluated and we get,

$$\begin{aligned} \mu(Y_2) &= \int_{-\infty}^{+\infty} dx_1 x_1 \int_{-\infty}^{+\infty} dx_2 f(x_1, x_2) + \\ &\quad \int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^{+\infty} dx_2 x_2 f(x_1, x_2) \\ &= \int_{-\infty}^{+\infty} dx_1 x_1 f_1(x_1) + \int_{-\infty}^{+\infty} dx_2 x_2 f_2(x_2) \\ &= \mu(X_1) + \mu(X_2). \end{aligned} \quad (20)$$

The means thus add up. The variance however does not, since it involves the square of the random variable. We have,

$$\sigma^2(Y_2) = \sigma^2(X_1) + \sigma^2(X_2) + 2 \times \text{cov}[X_1, X_2]. \quad (21)$$

The last term in the above is the **covariance** of X_1 and X_2 , and is given by

$$\text{cov}(X_1, X_2) = \int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^{+\infty} dx_2 [x_1 - \mu_1][x_2 - \mu_2] f(x_1, x_2), \quad (22)$$

where $\mu_1 = \mu(X_1)$ and $\mu_2 = \mu(X_2)$. One can define the **conditional density** of X_1 given that X_2 takes a value, say x_2 , as

$$f_c(x_1|x_2) = \frac{f(x_1, x_2)}{f_2(x_2)}. \quad (23)$$

If $f_c(x_1|x_2) = f_1(x_1)$, then we find that $f(x_1, x_2) = f_1(x_1) \times f_2(x_2)$. The random variables X_1 and X_2 are then **independent**. In that case we find from Eq. (22) that $\text{cov}(X_1, X_2) = 0$. Thus the covariance is zero if the two random variables are independent. If the covariance is positive then we say that the two random variables are positively correlated; if negative, the random variables are negatively correlated. Note that *two random variables may be uncorrelated i.e. covariance is zero) but they need not be independent; however if they are independent, they must be uncorrelated.* One usually considers the dimension - less normalized quantity called the correlation coefficient defined as,

$$C(X_1, X_2) = \frac{\text{cov}(X_1, X_2)}{\sigma(X_1)\sigma(X_2)}. \quad (24)$$

It is easily verified that $-1 \leq C(X_1, X_2) \leq 1$.

Let us calculate the mean and variance of the distributions we have seen so far:

Binomial distribution

For the toss of single fair coin,

$$\begin{aligned}\mu &= \frac{1}{2} \times (+1) + \frac{1}{2} \times (-1) = 0, \\ \sigma^2 &= \frac{1}{2} \times (+1)^2 + \frac{1}{2} \times (-1)^2 = 1.\end{aligned}\tag{25}$$

For the toss of N independent fair coins, the random variable $Y_N = (X_1 + X_2 + \dots + X_N)$ has

$$\mu_N = \frac{1}{2^N} \sum_{n=0}^N \frac{N!}{n!(N-n)!} (2n - N) = 0,\tag{26}$$

and

$$\sigma_N^2 = \frac{1}{2^N} \sum_{n=0}^N \frac{N!}{n!(N-n)!} (2n - N)^2 = N.\tag{27}$$

The random variable Y_N has a simple physical interpretation in terms of a random walk on a one-dimensional lattice. The lattice sites are on, say x axis, at unit intervals. You start at the origin. Toss a fair coin. If it shows up Heads then step to the right and go to the lattice site $x = 1$; if toss is Tails, then step to the left and go to the lattice site $x = -1$. At the new site, toss the coin again to decide whether to go the left site or the right site. After N tosses find out where you are on the axis, and your position defines the random variable Y_N . We can replace the number of coins or the tosses by time t and say $x(t)$ is the position of the random walker after time t , given it started at origin at time zero. Average of $x(t)$ is zero for all times. You do not move on the average. The variance of $x(t)$ denoted by $\sigma^2(t)$ increases linearly with time t , and the proportionality constant is often denoted by $2D$ where D is called the diffusion constant. Thus, for the random walk generated by the tossing of fair coins, the diffusion constant is one-half.

On the other hand, if we consider the random variable $\bar{Y}_N = Y_N/N$, we find its mean is zero and standard deviation is $1/\sqrt{N}$. Notice that the standard deviation of \bar{Y}_N becomes smaller and smaller as N increases.

Consider now the random variable $\bar{Y}_{\sqrt{N}} = Y_N/\sqrt{N}$; it is easily verified that its mean is zero and variance is unity, the same as that for the single toss. Thus \sqrt{N} seems to provide a natural scale for the sum of N independent and identically distributed random variables with finite variance. The reason for this would become clear in the sequel, see section 8.

For the case of a single loaded coin (with p as the probability for Heads and $q = 1 - p$ as that for the Tails) we have,

$$\mu = p \times (+1) + q \times (-1) = p - q = 2p - 1,\tag{28}$$

and

$$\sigma^2 = p \times (+1)^2 + q \times (-1)^2 - (p - q)^2 = 4p(1 - p).\tag{29}$$

For the toss of N independent and identically loaded coins, the random variable $Y_N = X_1 + X_2 + \cdots + X_N$ has,

$$\mu_N = \sum_{n=0}^N \frac{N!}{n!(N-n)!} p^n q^{N-n} (2n - N) = N(2p - 1), \quad (30)$$

and

$$\sigma_N^2 = \sum_{n=0}^N \frac{N!}{n!(N-n)!} p^n q^{N-n} (2n - N - \mu_N)^2 = 4Np(1-p). \quad (31)$$

We notice that both the mean and the variance increase linearly with N . The relevant quantity is the standard deviation relative to the mean; this quantity decreases as $1/\sqrt{N}$. The tossing of N loaded coins defines a biased random walk: there is a drift with velocity $2p - 1$ and a diffusion with a diffusion constant $D \equiv \sigma^2/2N = 2p(1-p)$.

We can also consider the random variable $\bar{Y}_N = Y_N/N$. Its mean is $2p - 1$ and is independent of N . Its standard deviation, however, decreases with N as $1/\sqrt{N}$. Consider now the natural scaling: $\bar{Y}_{\sqrt{N}} = Y_N/\sqrt{N}$; we find its mean is $(2p - 1)\sqrt{N}$ and variance $4p(1-p)$. The variance is independent of N .

Poisson distribution

The mean of the Poisson distribution is given by

$$\mu(t) = \sum_{n=0}^{\infty} n \frac{(\lambda t)^n}{n!} e^{-\lambda t} = \lambda t, \quad (32)$$

and the variance is given by,

$$\sigma^2(t) = \sum_{n=0}^{\infty} (n - \lambda t)^2 \frac{(\lambda t)^n}{n!} e^{-\lambda t} = \lambda t. \quad (33)$$

Thus, for the Poisson distribution the mean and the variance have the same magnitude.

Exponential distribution

The mean of the exponential density is given by

$$\mu = \int_0^{\infty} x \alpha \exp(-\alpha x) dx = \frac{1}{\alpha}, \quad (34)$$

and the variance is,

$$\sigma^2 = \int_0^{\infty} \left(x - \frac{1}{\alpha}\right)^2 \alpha \exp(-\alpha x) dx = \left(\frac{1}{\alpha}\right)^2. \quad (35)$$

The standard deviation of the exponential density is $\sigma = 1/\alpha$. We can associate the standard deviation with something like the expected deviation from the mean of a number,

sampled randomly from a distribution. We shall take up the issue of random sampling from a given distribution later. For the present, let us assume that we have a large set of numbers sampled independently and randomly from the exponential distribution. The expected fraction of the numbers that fall between $\mu - \sigma = 0$ and $\mu + \sigma = 2/\alpha$ can be calculated and is given by

$$\begin{aligned} \mathcal{P}(\mu - \sigma \leq x \leq \mu + \sigma) &= \int_{\mu - \sigma}^{\mu + \sigma} \alpha e^{-\alpha x} dx \\ &= 0.8647. \end{aligned} \tag{36}$$

Thus, nearly 86% of the sampled numbers are expected to lie within one sigma deviation from the mean. Of course there shall also be numbers much larger than the mean since the range of x extends up to infinity.

The distribution of the sum of N independent exponential random variables and the same scaled by N and \sqrt{N} will be considered separately later. In fact I am going to use this example to illustrate the approach to Gaussian as dictated by the Central Limit Theorem.

Cauchy distribution

Let us consider the Cauchy distribution discussed earlier. We shall try to evaluate the mean by carrying out the following integration,

$$\mu = \frac{1}{\pi} \int_{-\infty}^{+\infty} x \frac{D}{D^2 + x^2} dx. \tag{37}$$

The above integral strictly does not exist. The reason is as follows. If we carry out the integration from $-\infty$ to 0 and then from 0 to $+\infty$, these two integrals do not exist. Notice that the integrand is an odd function. Hence if we allow something like a *principal value* integration, where the limits are taken simultaneously, we can see that the integral from $-\infty$ to 0 cancels out that from 0 to $+\infty$, and we can say the mean, μ is zero, consistent with the graphical interpretation of the density, see Fig. 5. If we now try to evaluate the variance, we find,

$$\sigma^2 = \frac{1}{\pi} \int_{-\infty}^{+\infty} (x - \mu)^2 \frac{D}{D^2 + x^2} dx. \tag{38}$$

The above is an unbounded integral. Hence if we sample numbers randomly and independently from the Cauchy distribution and make an attempt to predict the extent to which these numbers fall close to the ‘mean’, we would fail. Nevertheless Cauchy distribution is a legitimate probability distribution since its integral from $-\infty$ to $+\infty$ is unity and for all values of x , the distribution function is greater than or equal to zero. But its variance is infinity and its mean calls for a ‘generous’ interpretation of the integration. Since the standard deviation for the Cauchy distribution is infinity, the width of the distribution is usually characterized by the **Full Width at Half Maximum(FWHM)** and it is $2D$.

Consider the sum $Y_N = X_1 + X_2 + \dots + X_N$ of N independent Cauchy random variables. What is the probability distribution of the random variable Y_N ? Does it tend to

Gaussian as $N \rightarrow \infty$? If not, why? What is the natural scaling for the sum Y_N ? These and related questions we shall take up later. Bear in mind that the Cauchy distribution has unbounded variance and this property will set it apart from the others which have finite variance.

Gaussian distribution

For the Gaussian distribution the two parameters μ and σ are the mean and standard deviation respectively. Let us calculate the probability that a number sampled from a Gaussian falls within one sigma interval around the mean. This can be obtained by integrating the Gaussian from $\mu - \sigma$ to $\mu + \sigma$. We get,

$$\begin{aligned} \mathcal{P}(\mu - \sigma \leq x \leq \mu + \sigma) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu-\sigma}^{\mu+\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx, \\ &= 0.6826. \end{aligned} \tag{39}$$

Thus 68% of the numbers sampled independently and randomly from a Gaussian distribution are expected to fall within one sigma interval around the mean. This interval is usually called the *one-sigma confidence interval*.

The sum Y_N of N independent Gaussian random variables is also a Gaussian with mean $N\mu$ and variance $N\sigma^2$. When you scale the sum by N , the mean is μ and the variance is σ^2/N ; on the other hand the random variable Y_N/\sqrt{N} , has mean $\mu\sqrt{N}$ and variance σ^2 . These will become clear when we take up the discussion on the Central Limit theorem later. In fact we shall see that the arithmetic mean of N independent random variables, each with finite variance, would tend to have a Gaussian distribution for large N . The standard deviation of the limiting Gaussian distribution will be proportional to the inverse of the square root of N . We shall interpret the one-sigma confidence interval of the Gaussian as the statistical error associated with the estimated mean of the N independent random variables.

Uniform distribution

For the uniform random variable $U(a, b)$, the mean is,

$$\mu = \frac{1}{b-a} \int_a^b x \, dx = \frac{a+b}{2}, \tag{40}$$

and the variance is,

$$\sigma^2 = \frac{1}{b-a} \int_a^b \left(x - \frac{a+b}{2}\right)^2 dx = \frac{(b-a)^2}{12}. \tag{41}$$

For the random variable $U(0, 1)$, the mean and variance are respectively 1/2 and 1/12.

Assignment 3

Consider N independent uniform random variables $U(0, 1)$. Let Y_N denote their sum.

(A) Find the mean and variance of (a) Y_N , (b) Y_N/N and (c) Y_N/\sqrt{N} as a function N .

(B) Derive an expression for the probability distribution of Y_N for $N = 2$.

5 CHARACTERISTIC FUNCTION

The **characteristic function** of a random variable X is defined as the expectation value of $\exp(ikX)$.

$$\Phi_X(k) = \int_{-\infty}^{+\infty} \exp(ikx) f(x) dx. \quad (42)$$

Expanding the exponential we get,

$$\Phi_X(k) = \int_{-\infty}^{+\infty} f(x) \left[1 + (ikx) + \frac{(ikx)^2}{2!} + \dots + \frac{(ikx)^n}{n!} + \dots \right] dx. \quad (43)$$

Assuming that term wise integration is valid, we find,

$$\Phi_X(k) = 1 + ikM_1 + \frac{(ik)^2}{2!} M_2 + \dots + \frac{(ik)^n}{n!} M_n \dots, \quad (44)$$

from which we get,

$$i^n M_n = \left. \frac{d^n}{dk^n} \Phi_X(k) \right|_{k=0}. \quad (45)$$

Thus we can generate all the moments from the characteristic function.

For a discrete random variable, the characteristic function is defined similarly as,

$$\Phi(k) = \sum_n \exp(ikx_n) p_n. \quad (46)$$

The logarithm of the characteristic function generates, what are called the **cumulants** or the **semi-invariants**. We have,

$$\Psi_X(k) = \ln [\Phi_X(k)] \equiv \sum_{n=1}^{\infty} \frac{(ik)^n}{n!} C_n, \quad (47)$$

where C_n is the n -th cumulant, given by,

$$i^n C_n = \left. \frac{d^n}{dk^n} \Psi_X(k) \right|_{k=0}. \quad (48)$$

How are the cumulants related to the moments ?

This can be found by considering,

$$\ln \left[1 + ikM_1 + \frac{(ik)^2}{2!}M_2 + \frac{(ik)^3}{3!}M_3 + \dots \right] = ikC_1 + \frac{(ik)^2}{2!}C_2 + \frac{(ik)^3}{3!}C_3 + \dots \quad (49)$$

Taylor-expanding the logarithm on the left hand side of the above and equating the coefficients of the same powers of k , we get,

$$C_1 = M_1 \quad (50)$$

$$C_2 = M_2 - M_1^2 = \sigma^2 \quad (51)$$

$$C_3 = M_3 - 3M_2M_1 + 2M_1^3, \quad (52)$$

$$C_4 = M_4 - 4M_3M_1 - 3M_2^2 + 12M_2M_1^2 - 6M_1^4. \quad (53)$$

The first cumulant is the same as the first moment (mean). The second cumulant is the variance.

A general and simple expression relating the moments to the cumulants can be obtained [11] as follows. We have,

$$\frac{d\Phi(k)}{dk} = \Phi(k) \frac{d\Psi(k)}{dk}, \quad (54)$$

where $\Phi(k)$, see Eq. (44) and $\Psi(k)$, see Eq. (47) are the moments and cumulants generating functions respectively; we have dropped the suffix X for convenience. We see that,

$$\begin{aligned} \frac{d^{n+1}\Phi(k)}{dk^{n+1}} &= \frac{d^n}{dk^n} \left[\Phi(k) \frac{d\Psi(k)}{dk} \right] \\ &= \sum_{m=0}^n \frac{n!}{m! (n-m)!} \frac{d^{n-m}\Phi(k)}{dk^{n-m}} \frac{d^{m+1}\Psi(k)}{dk^{m+1}}, \end{aligned} \quad (55)$$

from which it follows that,

$$\begin{aligned} C_1 &= M_1 \\ C_{n+1} &= M_{n+1} - \sum_{m=0}^{n-1} \frac{n!}{m! (n-m)!} M_{n-m} C_{m+1} \quad \forall n > 0. \end{aligned} \quad (56)$$

Let us calculate the characteristic functions of the several random variables considered so far:

Binomial distribution

The characteristic function of the random variable defined for the toss of a single fair coin is

$$\begin{aligned}\Phi_X(k) &= \frac{1}{2}e^{+ik} + \frac{1}{2}e^{-ik} \\ &= \cos(k).\end{aligned}\tag{57}$$

The characteristic function of $Y_N = X_1 + X_2 + \cdots + X_N$, defined for the toss of N independent fair coins is given by,

$$\begin{aligned}\Phi_{Y_N}(k) &= \frac{1}{2^N} \sum_{n=0}^N \frac{N!}{n!(N-n)!} e^{ik(2n-N)} \\ &= \frac{1}{2^N} \sum_{n=0}^N \frac{N!}{n!(N-n)!} (e^{ik})^n (e^{-ik})^{N-n} \\ &= \left(\frac{e^{ik} + e^{-ik}}{2} \right)^N \\ &= [\cos(k)]^N \\ &= [\Phi_X(k)]^N.\end{aligned}\tag{58}$$

Sum of independent random variables

A general result is that the characteristic function of the sum of independent random variables is given by the product of the characteristic functions of the random variables. Let X_1, X_2, \cdots, X_N be independent and not necessarily identically distributed. Let $Y_N = X_1 + X_2 + \cdots + X_N$. Formally, the characteristic function of Y_N , denoted by $\Phi_{Y_N}(k)$ is given by

$$\begin{aligned}\Phi_{Y_N}(k) &= \int dx_1 \int dx_2 \cdots \\ &\cdots \int dx_N \exp[ik(x_1 + x_2 + \cdots + x_N)] f(x_1, x_2, \cdots, x_N).\end{aligned}\tag{59}$$

Since X_1, X_2, \cdots, X_N are independent, the joint density can be written as the product of the densities of the random variables. *i.e.*,

$$f(x_1, x_2, \cdots, x_N) = f_1(x_1) f_2(x_2) \cdots f_N(x_N),\tag{60}$$

where $f_i(x_i)$ is the probability density of X_i . Therefore,

$$\begin{aligned}\Phi_{Y_N}(k) &= \prod_{i=1}^N \int dx_i \exp(ikx_i) f_i(x_i) \\ &= \prod_{i=1}^N \Phi_{X_i}(k),\end{aligned}\tag{61}$$

where $\Phi_{X_i}(k)$ denotes the characteristic function of the random variable X_i . If these N random variables are also identical, then $\Phi_{Y_N}(k) = [\Phi_X(k)]^N$.

Arithmetic mean of independent and identically distributed random variables

Let

$$\bar{Y}_N = \frac{X_1 + X_2 + \cdots + X_N}{N}, \quad (62)$$

define the arithmetic mean of N independent and identically distributed random variables. In Eq. (59), replace k by k/N ; the resulting equation defines the characteristic function of \bar{Y}_N . Thus we have,

$$\Phi_{\bar{Y}_N}(k) = \Phi_{Y_N}(k/N) = [\Phi_X(k/N)]^N. \quad (63)$$

Sum of N independent and identically distributed random variables scaled by \sqrt{N}

Later, we shall have occasions to consider the random variable $\bar{Y}_{\sqrt{N}}$, defined by,

$$\bar{Y}_{\sqrt{N}} = \frac{X_1 + X_2 + \cdots + X_N}{\sqrt{N}}, \quad (64)$$

whose characteristic function can be obtained by replacing k by k/\sqrt{N} in Eq. (59). Thus,

$$\Phi_{\bar{Y}_{\sqrt{N}}}(k) = \Phi_{Y_N}(k/\sqrt{N}) = [\Phi_X(k/\sqrt{N})]^N. \quad (65)$$

Exponential distribution

For the exponential distribution, see Eq. (12), with mean unity (*i.e.* $\alpha = 1$), the characteristic function is given by,

$$\begin{aligned} \Phi_X(k) &= \int_0^{\infty} e^{ikx-x} dx \\ &= \frac{1}{1-ik}. \end{aligned} \quad (66)$$

The sum Y_N of N exponential random variables has a characteristic function given by,

$$\Phi_{Y_N}(k) = \frac{1}{(1-ik)^N}. \quad (67)$$

The random variable \bar{Y}_N has a characteristic function given by,

$$\Phi_{\bar{Y}_N}(k) = \frac{1}{\left(1 - \frac{ik}{N}\right)^N}, \quad (68)$$

which has been obtained by replacing k by k/N in Eq. (67).

Poisson distribution

For the Poisson distribution, see Eq. (10), the characteristic function is given by,

$$\Phi(k, t) = \sum_{n=0}^{\infty} \frac{e^{ikn} (\lambda t)^n e^{-\lambda t}}{n!} = \exp \left[-\lambda t (1 - e^{ik}) \right], \quad (69)$$

which is the same as Eq. (8) if we set $z = \exp(ik)$.

Gaussian distribution

For the Gaussian distribution, the characteristic function is given by,

$$\Phi_X(k) = \exp(i\mu k - \frac{1}{2}\sigma^2 k^2) \quad (70)$$

Following the rules described above, we have,

$$\Phi_{Y_N}(k) = \exp \left[iN\mu k - \frac{1}{2}N\sigma^2 k^2 \right] \quad (71)$$

$$\Phi_{\bar{Y}_N}(k) = \exp \left[i\mu k - \frac{1}{2} \frac{\sigma^2 k^2}{N} \right] \quad (72)$$

$$\Phi_{\bar{Y}_{\sqrt{N}}}(k) = \exp \left[i\sqrt{N}\mu k - \frac{1}{2}\sigma^2 k^2 \right]. \quad (73)$$

For a Gaussian, only the first two cumulants are non-zero. identically zero. From Eq. (73) we see that the sum of N Gaussian random variables scaled by \sqrt{N} , is again a Gaussian with variance independent of N .

Assignment 4

Derive expressions for the characteristic functions of (a) the Gaussian and b) the Cauchy distributions.

6 CHEBYSHEV INEQUALITY

Let X be an arbitrary random variable with a probability density function $f(x)$, and finite variance σ^2 . The **Chebyshev inequality**, see Papoulis [9], says that,

$$\mathcal{P} \{ |X - \mu| \geq k\sigma \} \leq \frac{1}{k^2}, \quad (74)$$

for $k \geq 1$. Thus, regardless of the nature of the density function $f(x)$, the probability that the random variable X takes a value between $\mu - \epsilon$ and $\mu + \epsilon$, is greater than $1 - (\sigma^2/\epsilon^2)$, for $\epsilon \geq \sigma$. The Chebyshev inequality follows directly from the definition of the variance, as shown below.

$$\begin{aligned}
 \sigma^2 &= \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx \\
 &\geq \int_{-\infty}^{\mu - k\sigma} (x - \mu)^2 f(x) dx + \int_{\mu + k\sigma}^{+\infty} (x - \mu)^2 f(x) dx \\
 &\geq k^2 \sigma^2 \left[\int_{-\infty}^{\mu - k\sigma} f(x) dx + \int_{\mu + k\sigma}^{+\infty} f(x) dx \right] \\
 &= k^2 \sigma^2 \mathcal{P} \{ |X - \mu| \geq k\sigma \} . \tag{75}
 \end{aligned}$$

The Chebyshev inequality is simple; it is easily adapted to sums of random variables and this precisely is of concern to us in the Monte Carlo simulation. Take for example, N independent realizations of a random variable X with mean zero and variance σ^2 . Let \bar{Y}_N be the arithmetic mean of these realizations. \bar{Y}_N is a random variable. The mean of \bar{Y}_N is zero and its variance is σ^2/N . Chebyshev inequality can now be applied to the random variable \bar{Y}_N . Accordingly, a particular realization of the random variable \bar{Y}_N will lie outside the interval $(-\epsilon, +\epsilon)$ with a probability less than or equal to $\sigma^2/(N\epsilon^2)$. Thus, as ϵ becomes smaller, by choosing N adequately large we find that a realization of \bar{Y}_N can be made to be as close to the mean as we desire with a probability very close to unity.

This leads us naturally to the **laws of large numbers**. Of the several laws of large numbers, discovered over a period two hundred years, we shall see perhaps the earliest version, see Papoulis [9], which is, in a sense, already contained in the Chebyshev inequality.

7 LAW OF LARGE NUMBERS

Consider the random variables X_1, X_2, \dots, X_N which are independent and identically distributed. The common probability density has a mean μ and a *finite* variance. Let \bar{Y}_N denote the sum of the random variables divided by N . The law of large numbers says that for a given $\epsilon > 0$, as $N \rightarrow \infty$,

$$\mathcal{P} \{ |\bar{Y}_N - \mu| > \epsilon \} \rightarrow 0 .$$

It is easy to see that a realization of the random variable \bar{Y}_N is just the Monte Carlo estimate of the mean μ from a sample of size N . The law of large numbers assures us that the **sample mean** converges to the **population mean** as the sample size increases. I must emphasize that the convergence we are talking about here is in a probabilistic sense. Also notice that the law of large numbers does not make any statement about the nature of the probability density of the random variable \bar{Y}_N . It simply assures us that in the limit of $N \rightarrow \infty$, the sample mean would converge to the right answer. \bar{Y}_N is called the **consistent estimator** of μ .

The central limit theorem on the other hand, goes a step further and tells us about the nature of the probability density function of \bar{Y}_N , as we shall see below.

8 CENTRAL LIMIT THEOREM (CLT)

Let X_1, X_2, \dots, X_N be N independent and identically distributed random variables having a common Gaussian probability density with mean zero and variance σ^2 , given by,

$$G_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{x^2}{2\sigma^2}\right]. \quad (76)$$

Let $Y_N = X_1 + X_2 + \dots + X_N$. It is clear that the characteristic function of Y_N is the characteristic function of X raised to the power N . Thus, from Eq. (70),

$$\begin{aligned} \Phi_{Y_N}(k) &\equiv [\Phi_X(k)]^N \\ &= \exp\left[-\frac{1}{2}\sigma^2 k^2 N\right]. \end{aligned} \quad (77)$$

Fourier inverting the above, we find that the probability density of the random variable Y_N is Gaussian given by,

$$G_{Y_N}(x) = \frac{1}{\sigma\sqrt{2\pi N}} \exp\left[-\frac{x^2}{2N\sigma^2}\right], \quad (78)$$

with mean zero and variance $N\sigma^2$. Thus when you add Gaussian random variables, the sum is again a Gaussian random variable. Under addition, Gaussian is stable. The scaling behaviour of the distribution is evident,

$$G_{Y_N}(x) = N^{-1/2} G_X\left(\frac{x}{N^{1/2}}\right). \quad (79)$$

The variance of the sum of N independent and identically distributed Gaussian random variables increases (only) linearly with N . On the other hand the variance of \bar{Y}_N is σ^2/N , and thus decreases with N . Therefore the probability density of \bar{Y}_N is,

$$G_{\bar{Y}_N} = \frac{\sqrt{N}}{\sigma\sqrt{2\pi}} \exp\left[-\frac{Nx^2}{2\sigma^2}\right]. \quad (80)$$

The Central Limit Theorem asserts that even if the common probability density of the N random variables is not Gaussian, but some other arbitrary density with (zero mean and) finite variance, Eq. (80) is still valid but in the limit of $N \rightarrow \infty$. To see this, write the characteristic function of the common probability density as

$$\begin{aligned} \Phi_X(k) &= \int_{-\infty}^{+\infty} e^{ikx} f(x) dx \\ &= 1 - \frac{1}{2}\sigma^2 k^2 - \frac{1}{6}iM_3 k^3 + \frac{1}{24}M_4 k^4 \dots \end{aligned} \quad (81)$$

Hence the characteristic function of \bar{Y}_N is,

$$\begin{aligned}\Phi_{\bar{Y}_N}(k) &= [\Phi_X(k/N)]^N \\ &\approx \left[1 - \frac{1}{2} \frac{\sigma^2 k^2}{N^2} \left(1 + i \frac{M_3 k}{3\sigma^2} \frac{1}{N} - \frac{M_4 k^2}{12\sigma^2} \frac{1}{N^2} \cdots \right) \right]^N \\ &\sim \exp \left[-\frac{1}{2} \frac{\sigma^2 k^2}{N} \right] \quad (\text{for } N \rightarrow \infty),\end{aligned}\tag{82}$$

whose inverse is the density given by Eq. (80), see van Kampen [13].

The above can be expressed in terms of cumulant expansion. We have for the random variable X ,

$$\begin{aligned}\Phi_X(k) &= \exp[\Psi_X(k)] \\ &= \exp \left[ikC_1 + \frac{(ik)^2}{2!} C_2 + \frac{(ik)^3}{3!} C_3 + \cdots \right],\end{aligned}\tag{83}$$

where $C_1, C_2, C_3 \cdots$ are the cumulants of X . Then the characteristic function of the random variable \bar{Y}_N is given by,

$$\begin{aligned}\Phi_{\bar{Y}_N}(k) &= [\Phi_X(k/N)]^N \\ &= \exp \left[ikC_1 - \frac{1}{2} \frac{k^2 C_2}{N} \left(1 + \frac{ik C_3}{3 C_2} \frac{1}{N} - \frac{k^2 C_4}{12 C_2} \frac{1}{N^2} \cdots \right) \right].\end{aligned}\tag{84}$$

We immediately see that for $N \rightarrow \infty$

$$\Phi_{\bar{Y}_N} \sim \exp \left[ik C_1 - \frac{1}{2} \frac{k^2 C_2}{N} \right],\tag{85}$$

whose Fourier-inverse is a Gaussian with mean C_1 , and variance C_2/N .

For the random variable $\bar{Y}_{\sqrt{N}}$, defined as the sum of N random variables divided by \sqrt{N} , the characteristic function can be written as,

$$\begin{aligned}\Phi_{\bar{Y}_{\sqrt{N}}} &= [\Phi_X(k/\sqrt{N})]^N \\ &= \exp \left[ik\sqrt{N}C_1 - \frac{k^2}{2} C_2 \left\{ 1 + \frac{ik C_3}{3 C_2} \frac{1}{\sqrt{N}} - \frac{k^2 C_4}{12 C_2} \frac{1}{N} + \cdots \right\} \right] \\ &\sim \exp \left[ik\sqrt{N}C_1 - \frac{k^2 C_2}{2} \right] \quad (\text{for } N \rightarrow \infty),\end{aligned}\tag{86}$$

which upon Fourier inversion gives a Gaussian with mean $C_1\sqrt{N}$ and variance C_2 . The variance of $\bar{Y}_{\sqrt{N}}$ is independent of N .

8.1 Lévy Distributions

In the beginning of the discussion on the Central Limit Theorem, I said when you add Gaussian random variables, what you get is again a Gaussian. Thus Gaussian is *stable* under addition. This is a remarkable property. Not all distributions have this property. For example, when you add two uniform random variables, you get a random variable with triangular distribution. When two exponential random variables are added we get one with a Gamma distribution, as we would see in the next section where we shall be investigating the behaviour of the sum of several independently distributed exponential random variables. (In fact we went a step further and found that when we add N identically distributed independent random variables with finite variance, the resulting distribution tends to a Gaussian, when $N \rightarrow \infty$. This is what we called the Central Limit Theorem).

A natural question that arises in this context is: Are there any other distributions, that are stable under addition? The answer is yes and they are called the **Lévy stable distributions**, discovered by Paul Lévy [12] in the mid twenties. Unfortunately the general form of the Lévy distribution is not available. What is available is its characteristic function. Restricting ourselves to the symmetric case, the characteristic function of a Lévy distribution is given by,

$$\tilde{L}(k; \alpha) = \exp(-D|k|^\alpha), \quad (87)$$

where $D > 0$ is a scale factor, k is the transform variable corresponding to x , and α is the Lévy index. The Fourier inverse of $\tilde{L}(k; \alpha)$ is the Lévy distribution $L(x; \alpha)$. Lévy showed that for $L(x; \alpha)$ to be non negative, $0 < \alpha \leq 2$. In fact we have,

$$L(x; \alpha) \sim \begin{cases} D^{-1/\alpha}, & \text{for } x = 0; \\ D|x|^{-\alpha-1} & \text{for } |x| \rightarrow \infty; \alpha < 2 \quad ; \\ D^{-1/2} \exp(-x^2/4D) & , \forall x \text{ and } \alpha = 2 \end{cases} \quad (88)$$

The pre factor in the above can be fixed by normalization. We see that Gaussian is a special case of the Lévy distributions when $\alpha = 2$. In fact Gaussian is the only Lévy distribution with finite variance. Earlier we saw about Cauchy distribution. This is again a special case of the Lévy distribution obtained when we set $\alpha = 1$.

Consider N independent and identically distributed Lévy random variables, with the common distribution denoted by $L(x; \alpha)$. The sum Y_N is again a Lévy distribution denoted by $L_N(x; \alpha)$, obtained from $L(x; \alpha)$ by replacing D by ND . Let us scale Y_N by $N^{1/\alpha}$ and consider the random variable $Y_N/N^{1/\alpha}$. Its distribution is $L(x; \alpha)$. Thus $N^{1/\alpha}$ provides a natural scaling for Levy distributions. We have,

$$L_N(x; \alpha) = N^{-1/\alpha} L_X \left(\frac{x}{N^{1/\alpha}}; \alpha \right). \quad (89)$$

The scaling behaviour fits into a general scheme. The Gaussian corresponds to $\alpha = 2$; the natural scaling for Gaussian is thus \sqrt{N} . The Cauchy distribution corresponds to $\alpha = 1$; the natural scaling is given by N .

Thus Lévy stable law generalizes the central limit theorem: Lévy distributions are the only possible limiting distributions for the sums of independent identically distributed

random variables. The conventional Central Limit Theorem is a special case restricting each random variable to have finite variance and we get Gaussian as a limiting distribution of the sum. Lévy's more general Central Limit Theorem applies to sums of random variables with finite or infinite variance.

Stable distributions arise in a natural way when 1) a physical system evolves, the evolution being influenced by additive random factors and 2) when the result of an experiment is determined by the sum of a large number of independent random variables.

8.2 Illustration of the CLT

I shall illustrate the approach to Gaussian of the distribution of the mean, by considering exponential random variables for which all the relevant quantities can be obtained analytically. We start with the identity,

$$\int_0^{\infty} dx e^{-\alpha x} e^{ikx} = \frac{1}{\alpha - ik} \quad (90)$$

We differentiate $(N - 1)$ times, both sides of the above with respect to α and set $\alpha = 1$. We get,

$$\int_0^{\infty} dx e^{ikx} \left[\frac{e^{-x} x^{N-1}}{(N-1)!} \right] = \frac{1}{(1-ik)^N}. \quad (91)$$

We immediately identify the right hand side of the above, as the characteristic function of the sum of N independent and identically distributed exponential random variables with mean unity, see Eq. (67). In Eq. (91) above, we replace, on both sides, k by k/N , x by Nx , and dx by Ndx , to get,

$$\int_0^{\infty} dx e^{ikx} \left[\frac{e^{-Nx} N^N x^{N-1}}{(N-1)!} \right] = \frac{1}{\left(1 - i\frac{k}{N}\right)^N}. \quad (92)$$

The right hand side of the above is the characteristic function of \bar{Y}_N see Eq. (68). Thus we get an exact analytical expression for the probability density function of \bar{Y}_N , as

$$f(x) = \frac{N^N}{(N-1)!} e^{-Nx} x^{N-1}, \quad 0 \leq x < \infty, \quad (93)$$

which, for all N , is a **gamma density function**, and not a Gaussian!

The cumulants of the gamma density can be obtained by power series expansion of the logarithm of the characteristic function given by Eq. (68). We get,

$$\begin{aligned} \ln \Phi_{\bar{Y}_N}(k) &= -N \ln \left(1 - \frac{ik}{N} \right) \\ &= N \sum_{n=1}^{\infty} \frac{(ik)^n}{N^n n} \\ &= \sum_{n=1}^{\infty} \frac{(ik)^n}{n!} \left[\frac{(n-1)!}{N^{n-1}} \right], \end{aligned}$$

$$= ik + \frac{(ik)^2}{2!} \frac{1}{N} \left[1 + \frac{ik}{3} \frac{2!}{N} \cdots \right]. \quad (94)$$

The n -th cumulant is thus,

$$C_n = \frac{(n-1)!}{N^{n-1}}. \quad (95)$$

We find $C_1 = 1$ and $C_2 = 1/N$, as expected. The third cumulant is $C_3 = 2/N^2$, and it goes to zero as $N \rightarrow \infty$. In fact all the higher cumulants vanish in the asymptotic limit. (Indeed even the second cumulant goes to zero in the limit $N \rightarrow \infty$, which means that the probability density of the arithmetic mean is asymptotically a delta function, something very satisfying from the point of view of a Monte Carlo practitioner!). The important point is that the Gaussian with mean *unity* and variance $1/N$ becomes a very good approximation to the gamma density, for large N .

Replacing k by k/\sqrt{N} , x by $x\sqrt{N}$ and dx by $dx\sqrt{N}$ in Eq. (91), we get,

$$\int_0^\infty dx e^{ikx} \left[\frac{(\sqrt{N})^N}{(N-1)!} x^{N-1} e^{-\sqrt{N}x} \right] = \frac{1}{\left(1 - i\frac{k}{\sqrt{N}}\right)^N}. \quad (96)$$

The right hand side of the above equation is the characteristic function of the random variable $\bar{Y}_{\sqrt{N}} = (X_1 + X_2 + \cdots + X_N)/\sqrt{N}$. Thus, the probability density function of $\bar{Y}_{\sqrt{N}}$, is given by the gamma density,

$$f(x) = \frac{(\sqrt{N})^N}{(N-1)!} x^{N-1} \exp(-\sqrt{N}x), \quad (97)$$

whose mean is \sqrt{N} and whose variance is unity (independent of N). Fig. 6 depicts the gamma density given above for $N = 1, 10, 20, 40, 60$ and 80 along with the Gaussian of mean \sqrt{N} and variance unity. For large N we find that the Gaussian is a good fit to the gamma density, for $|x - \sqrt{N}| \ll 1$.

In the discussions above on the central limit theorem, we have assumed the random variables to be independent and identically distributed. The central limit theorem holds well under much more general conditions.

First, the random variables need not be identical. To see this, let N_1 of the N random variables have one distribution with say mean μ_1 and variance σ_1^2 . Let Y_1 denote their sum. Also let N_2 of the N random variables have another distribution with mean μ_2 and variance σ_2^2 and let their sum be denoted by Y_2 . We take the limit $N_1 \rightarrow \infty$ and $N_2 \rightarrow \infty$ such that N_1/N_2 remains a constant. Asymptotically ($N_1 \rightarrow \infty$) the random variable Y_1 tends to a Gaussian and the characteristic function is $\phi_1(k) = \exp(ikN_1\mu_1 - k^2N_1\sigma_1^2/2)$. Similarly the random variable Y_2 tends to a Gaussian and the characteristic function is $\phi_2(k) = \exp(ikN_2\mu_2 - k^2N_2\sigma_2^2/2)$. Their sum $Y = Y_1 + Y_2$ has the characteristic function $\phi(k) = \phi_1(k) \times \phi_2(k)$, since Y_1 and Y_2 are independent. Thus $\phi(k) = \exp\{ik(N_1\mu_1 + N_2\mu_2) - k^2(N_1\sigma_1^2 + N_2\sigma_2^2)/2\}$. Fourier inverse of $\phi(k)$ is a Gaussian with mean $N_1\mu_1 + N_2\mu_2$ and variance $N_1\sigma_1^2 + N_2\sigma_2^2$.

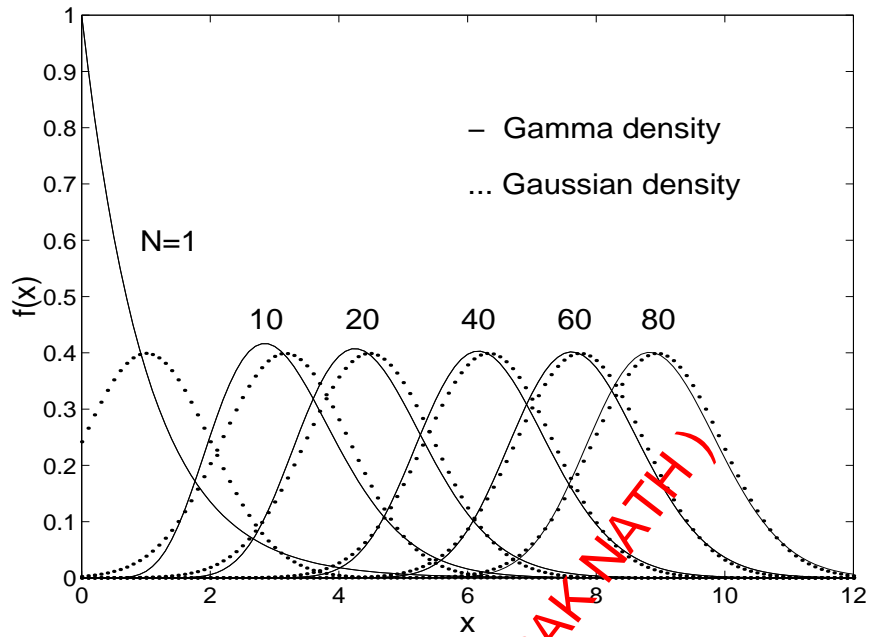


Figure 6: Gamma density and the asymptotic Gaussian

Second, the random variables can be weakly dependent, see for *e.g.* Gardiner [14]. An example is the correlated random walk, described in Assignment (6). Essentially the adjacent steps of the random walk are correlated. Despite this, asymptotically the distribution of the position of the random walk is Gaussian, see for example [15].

Assignment 5

Consider a random variable X with probability density $f(x) = \exp(-\sqrt{2}|x|)/\sqrt{2}$, for $-\infty \leq x \leq +\infty$.

(A) What is the characteristic function of X ?

Let X_1, X_2, \dots, X_N , be independent and identically distributed random variables with the common density $f(x)$ given above. Let Y_N be the sum of these. *i.e.* $Y_N = \sum_{i=1}^N X_i$.

(B) What is the characteristic function of the random variable Y_N ?

Let $\bar{Y}_N = Y_N/N$.

(C) What is the characteristic function of \bar{Y}_N ?

Let $\bar{Y}_{\sqrt{N}} = Y_N/\sqrt{N}$.

(D-a) What is the characteristic function of $\bar{Y}_{\sqrt{N}}$?

(D-b) Taylor expand the characteristic function of $\bar{Y}_{\sqrt{N}}$ in terms of moments and in terms of cumulants.

(D-c) Employing the Moment and Cumulant expansion, demonstrate the approach to Gaussian as $N \rightarrow \infty$.

(D-d) What is the mean of the asymptotic Gaussian ?

(D-e) What is its variance of the asymptotic Gaussian ?

Assignment 6

Consider one-dimensional lattice with lattice sites indexed by $(\dots - 3, -2, -1, 0, +1, +2, +3, \dots)$; a random walk starts at origin and in its first step jumps to the left site at -1 or to the right site $+1$ with equal probability. In all subsequent steps the probability for the random walk to continue in the same direction is α and reverse the direction is $1 - \alpha$. Let $x(n)$ denote the position of the random walk after n steps. Show that asymptotically ($n \rightarrow \infty$) the distribution of $x(n)$ is Gaussian. What is the mean of the limiting Gaussian? What is its variance? See [15]

Thus we find that for the central limit theorem to hold good it is adequate that each of the random variables has a finite variance and thus none of them dominates the sum excessively. They need not be identical; they can also be weakly dependent.

Therefore, whenever a large number of random variables additively determine a parameter, then the parameter tends to have a Gaussian distribution. Upon addition, the individual random variables lose their characters and the sum acquires a Gaussian distribution. It is due to this remarkable fact that Gaussian enjoys a pre-eminent position in statistics and statistical physics.

Having equipped ourselves with the above preliminaries let us now turn our attention to random and pseudo random numbers that form the backbone of all Monte Carlo

simulations.

9 RANDOM NUMBERS

What are Random numbers ?

We call a sequence of numbers *random*, if it is generated by a random physical process. How does randomness arise in a (macroscopic) physical system while its microscopic constituents obey deterministic and time reversal invariant laws ? This issue concerns the search for a microscopic basis for the second law of thermodynamics and the nature and content of entropy (randomness). I shall not discuss these issues here, and those interested should consult [16], and the references therein.

Instead we shall say that physical processes such as radioactivity decay, thermal noise in electronic devices, cosmic ray arrival times *etc.*, give rise to what we shall call a sequence of truly random numbers. There are however several problems associated with generating random numbers from physical processes. The major one concerns the elimination of the apparatus bias. See for example [17] where description of generating truly random numbers from a radioactive alpha particle source is given; also described are the bias removal techniques.

Generate once for all, a fairly long sequence of random numbers from a random physical process. Employ this in all your Monte Carlo calculations. This is a safe procedure. Indeed tables of random numbers were generated and employed in the early days of Monte Carlo practice, see for example [3]. The most comprehensive of them was published by Rand Corporation [18] in mid fifties. The table contained one million random digits.

Tables of random numbers are useful if the Monte Carlo calculations are carried out manually. However, for computer calculations, use of these tables is impractical. The reason is simple. A computer has a relatively small internal memory; it cannot hold a large table. One can keep the table of random numbers in an external storage device like a disk or tape; constant retrieval from these peripherals would considerably slow down the computations. Hence it is often desirable to generate a random number as and when required, employing a simple algorithm that takes very little time and very little memory. This means one should come up with a practical definition of randomness of a sequence of numbers.

At the outset we recognize that there is no proper and satisfactory definition of randomness. Chaitin [20] has made an attempt to capture an intuitive notion of randomness into a precise definition. Following Chaitin, consider the two sequences of binary random digits given below:

1 0 1 0 1 0 1 0 1 0 1 0 1 0
1 1 0 1 0 0 1 0 1 1 0 0 1 1

I am sure you would say that the first is not a random sequence because there is a pattern in it - a repetition of the doublet 1 and 0; the second is perhaps a random

sequence as there is no recognizable pattern. Chaitin goes on to propose that **a sequence of numbers can be called random if the smallest algorithm capable of specifying it to the computer has about the same number of bits of information as the sequence itself.**

Let us get back to the two sequences of binary numbers given above. We recognize that tossing a fair coin fourteen times independently can generate both these sequences. Coin tossing is undoubtedly a truly random process. Also the probability for the first sequence is exactly the same as that for the second and equals 2^{-14} , for a fair coin. How then can we say that the first sequence is not random while the second is? Is not a segment with a discernible pattern, part and parcel of an infinitely long sequence of truly random numbers? See [19] for an exposition of randomness along these lines.

We shall however take a practical attitude, and consider numbers generated by a deterministic algorithm. These numbers are therefore predictable and reproducible; the algorithm itself occupies very little memory. Hence by no stretch of imagination can they be called random. We shall call them **pseudo random numbers**. We shall be content with pseudo random numbers and employ them in Monte Carlo calculations. We shall reserve the name pseudo random numbers for those generated by deterministic algorithm and are supposed to be distributed independently and uniformly in the range 0 to 1. We shall denote the sequence of pseudo random numbers by $\{\xi_1, \xi_2, \dots\}$. For our purpose it is quite adequate if *one ensures that the sequence of pseudo random numbers is statistically indistinguishable from a sequence of truly random numbers*. This is a tall order! How can anybody ensure this?

9.1 Tests for Randomness

Usually we resort to what are called **tests of randomness**. A test, in a general sense, consists of constructing a function $\psi(r_1, r_2, \dots)$, where r_1, r_2, \dots , are independent variables. Calculate the value of the function for a sequence of pseudo random numbers by setting $r_i = \xi_i \forall i = 1, 2, \dots$. Compare this value with the value that ψ is expected to have if $\{r_i : i = 1, 2, \dots\}$ were truly random numbers distributed independently and uniformly in the range 0 to 1.

For example, the simplest test one can think of, is to set

$$\psi(r_1, r_2, \dots) = \frac{1}{N} \sum_{i=1}^N r_i, \quad (98)$$

which defines the average of N numbers. For a sequence of N truly random numbers we expect ψ to lie between $.5 - \epsilon$ and $.5 + \epsilon$ with a certain probability $p(\epsilon)$. Notice that for N large, from the Central Limit Theorem, ψ is Gaussian with mean 0.5 and variance $\sigma^2 = (12N)^{-1}$. If we take $\epsilon = 2\sigma = 1/\sqrt{3N}$, then $p(\epsilon)$ is the area under the Gaussian between $.5 - \epsilon$ and $.5 + \epsilon$ and is equal to 0.95. This is called the two-sigma confidence interval. Thus for a sequence of N truly random numbers, we expect its mean to be within $\pm\epsilon$ around 0.5 with .95 probability, for large N . If a sequence of N pseudo random numbers has an average that falls outside the interval $(.5 - \epsilon, .5 + \epsilon)$ then we say that it fails the test at 5% level.

Assignment 7

Carry out the above test on the random numbers generated by one of the generators in your computer. Does it pass the test at 5% level ?

Another example consists of setting,

$$\psi(r_1, r_2, \dots) \equiv C(k) = \frac{N \sum_{i=1}^N r_i r_{i+k} - \left(\sum_{i=1}^N r_i \right)^2}{N \sum_{i=1}^N r_i^2 - \left(\sum_{i=1}^N r_i \right)^2}, \quad (99)$$

with $k = 0, 1, 2, \dots, N-1$. For a sequence of truly random numbers, the function ψ above, which denotes two point auto correlation function, is expected to be unity for $k = 0$ and zero for $k \neq 0$.

Assignment 8

Calculate the auto correlation of the random numbers generated by one of the random number generators in your computer.

In practice, one employs more complicated tests, by making ψ a more complicated function of its arguments.

An example is the **run test** which is more sensitive to the correlations. The idea is to calculate the length of a run of increasing (run-up) or decreasing (run-down) size. We say the run-down length is l if we have a sequence of random numbers such that $\xi_{l+1} > \xi_l < \xi_{l-1} < \xi_{l-2} < \dots < \xi_2 < \xi_1$. We can similarly define the run-up length.

Let me illustrate the meaning of run-down length by considering a sequence of integers between 0 and 9, given below.

6	4	1	3	4	5	3	2	7	4	8					
	6	4	1	3		4	5		3	2	7		4	8	
				3			1			2			1		

The first row displays the sequence; the second depicts the same sequence with numbers separated into groups by vertical lines. Each group contains $n + 1$ numbers the first n of which are in descending order. *i.e.* these numbers are running down. The descent is broken by the $(n + 1)^{\text{th}}$ number which is greater than the n^{th} number. The third row gives the run-down length, for each group, which is n .

We can calculate the probability distribution of the run-down length as follows. Let $P(L \geq l)$ be the probability that the run-down length L is greater than or equal to l . To

calculate $P(L \geq l)$ we consider a sequence of l distinct random numbers. There are $l!$ ways of arranging these numbers. For a sequence of truly random numbers, each of these arrangements is equally likely. Of these, there is only one arrangement which has all the l numbers in the descending order. Hence $P(L \geq l) = 1/l!$. Since $P(L = l) = P(L \geq l) - P(L \geq l + 1)$, we get

$$P(l) = \frac{1}{l!} - \frac{1}{(l+1)!} \quad (100)$$

Alternately, for the probability of run-down length we have,

$$\begin{aligned} P(L \geq l) &= \int_0^1 d\xi_1 \int_0^{\xi_1} d\xi_2 \cdots \int_0^{\xi_{l-1}} d\xi_l \\ &= \int_0^1 \frac{\xi_1^{l-1}}{(l-1)!} d\xi_1 \\ &= \frac{1}{l!} \end{aligned} \quad (101)$$

In the test, we determine numerically, the distribution of the run length on the sequence of pseudo numbers and compare it with the exact distribution given by Eq. (100). Fig. 7 depicts the results of a run-up test. Description of several of the most common tests for randomness can be found in [21].

One issue becomes obvious from the above discussions. There is indeed an indefinitely large number of ways of constructing the function ψ . Hence, in principle a pseudo random number generator can never be tested thoroughly for the randomness of the sequence of random numbers it generates. What ever may be the number of tests we employ and however complicated they may be, there can be, and always shall be, surprises....surprises like the Marsaglia planes discovered in the late sixties [22] or the parallel lines discovered some three years ago by Vattulainen and co-workers [23]. We shall discuss briefly these issues a bit later.

Assignment 9

Carry out (a) run up and (b) run down tests, on the random numbers generated by one of the random number generators in your computer and compare your results with the exact analytical results. It is adequate if you consider run lengths up to six or so.

The struggle to develop better and better random number generators and the simultaneous efforts to unravel the hidden order in the pseudo random numbers generated, constitute an exciting and continuing enterprise, see for example [24]. See also Bonelli and Ruffo [25] for a recent interesting piece of work on this subject.

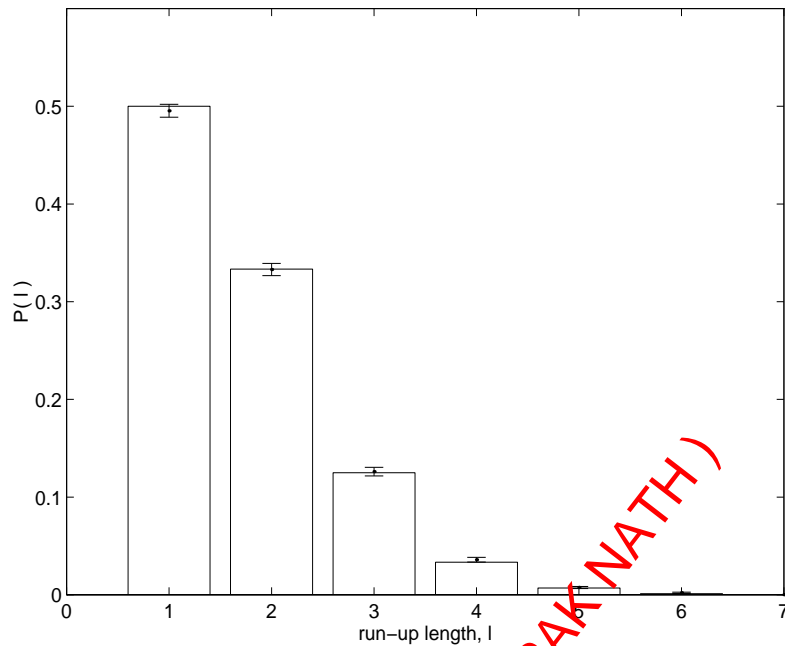


Figure 7: Results of a run-up test. The histogram plot corresponds to what is expected from a sequence of truly random numbers. See Eq. (100). The points with error bars correspond to the ones obtained with random number generator routine RAND of MATLAB version 4.1. The run-up lengths were generated from a sequence of 10,000 numbers. The statistical errors in the calculated values correspond to one-sigma confidence interval.

9.2 Pseudo Random Number Generators

The earliest pseudo random number generator was the **mid-squares** proposed by von Neumann. Start with an m digit integer ν_1 . Take the middle $m/2$ digits and call it η_1 . Square η_1 and call the resulting m digit integer as ν_2 . Take the middle $m/2$ digits of ν_2 and call it η_2 . Proceed in the same fashion and obtain a sequence of integers η_1, η_2, \dots . These are then converted to real numbers between zero and one by dividing each by $10^{m/2}$. For a properly chosen seed ν_1 , this method yields a long sequence of *apparently* good random numbers. But on a closer scrutiny, it was found that the mid-squares is not a good random number generator. I shall not spend anymore time on this method since it is not in vogue. Instead we shall turn our attention to the most widely used class of random number generators called the **linear congruential generators**, discovered by Lehmer [26]. Start with a seed R_1 and generate successive integer random numbers by

$$R_{i+1} = aR_i + b \pmod{m}, \quad (102)$$

where a , b , and m are integers. a is called the **generator** or **multiplier**; b , the **increment** and m is the **modulus**.

Equation (102) means the following. Start with an integer R_1 . This is your choice. Calculate $a \times R_1 + b$. This is an integer. Divide this by m and find the remainder. Call it R_2 . Calculate $a \times R_2 + b$; divide the result by m ; find the remainder and call it R_3 .

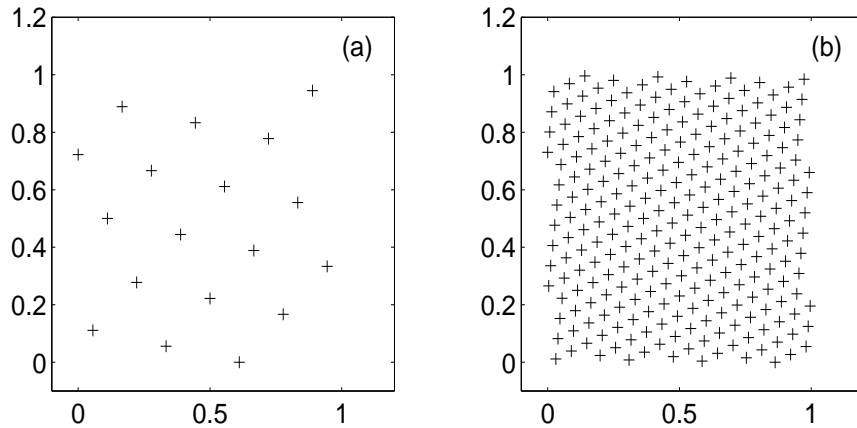


Figure 8: Random numbers of the linear congruential generator embedded in a two dimensional phase space. (a) depicts the results for the generator with $a = 7$, $b = 13$ and $m = 18$. (b) depicts the results for the generator with $a = 137$, $b = 187$ and $m = 256$. (Both these generators give full period)

Assignment 10

Construct a linear congruential generator that gives rise to a long sequence of pseudo random numbers with full period. Embed them in two or higher dimensional phase space and see if there are any patterns. Test for the randomness of the sequence.

The modulus m is usually taken as $2^{t-1} - 1$, where t is the number of bits used to store an integer and hence is machine specific. One of the t bits is used up for storing the sign of the integer. The choice $a = 7^5 = 16801$, $b = 0$ and $m = 2^{31} - 1$, for a 32 bit machine has been shown to yield good results, see [29]. The Ahren generator which specifies $a = 2^{t-2}(\sqrt{5}-1)/2$, $b = 0$, has also been shown to yield ‘good’ random numbers. The linear congruential generators have been successful and invariably most of the present day pseudo random number generators are of this class.

In the late sixties, Marsaglia [22] established unambiguously that the formation of lattice structure is an inherent feature of a linear congruential generator:

If d -tuples $(\xi_1, \xi_2, \dots, \xi_d)$, $(\xi_2, \xi_3, \dots, \xi_{d+1})$, $(\xi_3, \xi_4, \dots, \xi_{d+2})$, \dots , of the random numbers produced by linear congruential generators are viewed as points in the unit hyper cube of d dimensions, then all the points will be found to lie in a relatively small number of parallel hyper planes. Such structures occur with any near congruential generator and in any dimension.

The number of such Marsaglia planes does not exceed $(d!2^t)^{1/d}$. Besides the points

on several Marsaglia planes form regular patterns. Thus it is clear now that the existence of Marsaglia planes is undoubtedly a **serious defect** inherent in the linear congruential generators.

Then there is the other matter of the presence of correlations, hopefully weak, in the sequence of pseudo random numbers. If your Monte Carlo algorithm is sensitive to the subtle correlations present in the sequence, then you are in problem. For example Ferrenberg, Landau and Wong [30], found that the so-called high quality random number generators led to subtle but dramatic errors in algorithm that are sensitive to the correlations. See also [31].

One should perhaps conclude that the **linear congruential generators are not suitable for Monte Carlo work!** Before we come to such a harsh conclusion, let us look at the issues in perspective. If your Monte Carlo program does not require more than a small fraction (smaller the better) of the full period of the random numbers, then most likely, the presence of Marsaglia lattice structure will not affect your results. All the linear congruential generators in use today have long periods. Also one can think of devices to increase the number of Marsaglia planes. For example the Dieter-Ahren ‘solution’ to the Marsaglia problem is the following algorithm,

$$R_{i+2} = aR_{i+1} + bR_i \pmod{m}, \quad (104)$$

which requires two seeds. The modulus $m = 2^{t/d} - 1$. For a proper choice of a and b , the number of Marsaglia planes can be increased by a factor of $2^{t/d}$.

Thus over the period of thirty years we have learnt to live with the lattice defect of the linear congruential generators. However at any time you get the feeling that there is something wrong with the simulation results and you suspect that this is caused by the Marsaglia lattice structure of the linear congruential generator then you should think of employing some other random number generator, like the **inversive congruential generator (ICG)** proposed in 1986 by Eichenauer and Lehn [32] or the **explicit inversive congruential generator (EICG)**, proposed in 1993 by Eichenauer-Hermann [33]. Both these generators do not suffer from the lattice or hyper plane structure defect. I shall not get into the details of these new inversive generators, except to say that the inversive congruential generator also employs recursion just as linear congruential generators do, but the recursion is based on a nonlinear function.

More recently in the year 1995, Vattulainen and co-workers [23] found, in the context of linear congruential generators, that the successive pairs (ξ_t, ξ_{t+l}) exhibit a pattern of parallel lines on a unit square. This pattern is observed for all (ξ_t, ξ_{t+l}) with $l = 1, 2, \dots$, where the modulus $m = 2^{31} - 1$ for a 32 bit machine. The number of parallel lines strongly depends on l . For $l = (m - 1)/2$ the pairs fall on a single line, and for $l = \{(m - 1)/2\} \pm 1$, the pairs fall on so large a number of parallel lines that they can be considered as space filling. This curious phenomenon of switching from pairs falling on one or a few parallel lines to pairs falling on several lines upon tuning the parameter l is termed as transition from regular behaviour to chaotic behaviour, see [34]. Thus there seems to be a connection between pseudo random number generators and chaotic maps. **Chaos**, we know, permits only short time predictability; no long term forecasting is possible. The predictability gets lost at long times.....at times greater than the inverse of the largest Lyapunov exponent.

It is rather difficult to distinguish a chaotic behaviour from a random behaviour. Thus Chaos can provide an effective source of randomness. This is definitely meaningful idea since chaotic and random behaviours have many things common, see for example [35], where Chaos has been proposed as a source of pseudo random numbers. But then we know there is an order, strange (fractal) or otherwise, in Chaos. What exactly is the connection between the Marsaglia order found in the context of linear congruential generators and the fractal order in Chaos ? If answers to these and several related questions can be found, then perhaps we can obtain some insight into the otherwise occult art of pseudo random number generators. A safe practice is the following. Consider the situation when you are planning to use a standard Monte Carlo code on a new problem; or consider the situation when you are developing a new Monte Carlo algorithm. In either case, carefully test your Monte Carlo along with the random number generator, on several standard problems for which the results are reasonably well known. This you should do irrespective of how ‘famous’ the random generator is, and how many randomness tests it has been put through already. After all your Monte Carlo programme itself can be thought of as a new test of randomness of the pseudo random number generator.

I shall stop here the discussion on pseudo random number generators and get on to Monte Carlo. In the next section I shall discuss random sampling techniques that transform the pseudo random numbers, $\{\xi_i\}$, independent and uniformly distributed in the range 0 to 1, to $\{x_i\}$, having the desired distribution in the desired range.

10 RANDOM SAMPLING TECHNIQUES

The **random sampling techniques** help us convert a sequence of random numbers $\{\xi_i\}$, uniformly distributed in the interval $(0, 1)$ to a sequence $\{x_i\}$ having the desired density, say $f(x)$. There are several techniques that do this conversion. These are direct inversion, rejection methods, transformations, composition, sums, products, ratios, table-look-up of the cumulative density with interpolation, construction of equ-probability table, and a whole host of other techniques. In what follows we shall illustrate random sampling by considering in detail a few of these techniques.

10.1 Inversion Technique

The simplest is based on the direct **inversion** of the cumulative density function of the random variable X . We shall first consider sampling from a discrete distribution employing inversion technique. Let $\{p_i : i = 1, 2, \dots, N\}$ be the discrete probabilities for the random variable X to take values $x_i (= i)$. Fig. 9a depicts an example with $N = 5$. We first construct the cumulative probabilities $\{P_i : i = 0, N\}$ where $P_0 = 0$; $P_1 = p_1$; $P_2 = p_1 + p_2$; \dots ; $P_k = p_1 + p_2 + \dots + p_k$; $P_N = 1$. Fig. 9b depicts the cumulative probabilities as staircase of non-decreasing heights. The procedure is simple. Generate ξ , a random number uniformly distributed in the range $(0, 1)$. Find k for which $P_{k-1} \leq \xi < P_k$. Then $x_k (= k)$ is the sampled value of X . This is equivalent to the following. The value of ξ defines a point on the y axis of Fig. 9b between zero and one. Draw a line parallel to the x axis at $(0, \xi)$. Find the vertical line segment intersected by this line. The vertical line

segment is then extended downwards to cut the x axis at say x_1 . Then x_1 is the (desired) sampled value of X .

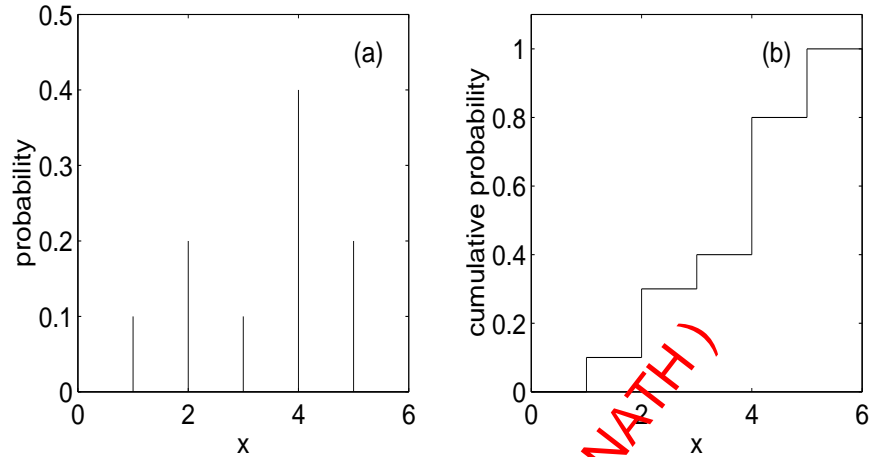


Figure 9: Discrete distribution. (a) probabilities (b) cumulative probabilities

Assignment 11

Employing inversion technique sample from (a) binomial and (b) Poisson distributions. Compare the frequency distribution you obtain with the exact.

The above can be generalized to continuous distribution. We have the cumulative probability distribution $F(x) = \int_{-\infty}^x f(x') dx'$. Note that $F(-\infty) = 0$ and $F(+\infty) = 1$. Given a random number ξ_i , we have $x_i = F^{-1}(\xi_i)$.

An example with exponential distribution is shown in Fig. 10. We have plotted in Fig. 10a the exponential distribution for $x \geq 0$. Fig. 10b depicts the cumulative distribution. Select a point on the y axis of Fig. 10b randomly between zero and one. Draw a line parallel to x axis passing through this point. Find where it intersects the curve $F(x)$. Read off the x coordinate of the intersection point, and call this x_1 . Repeat the above several times and get $\{x_i : i = 1, 2, \dots\}$. These numbers will be exponentially distributed.

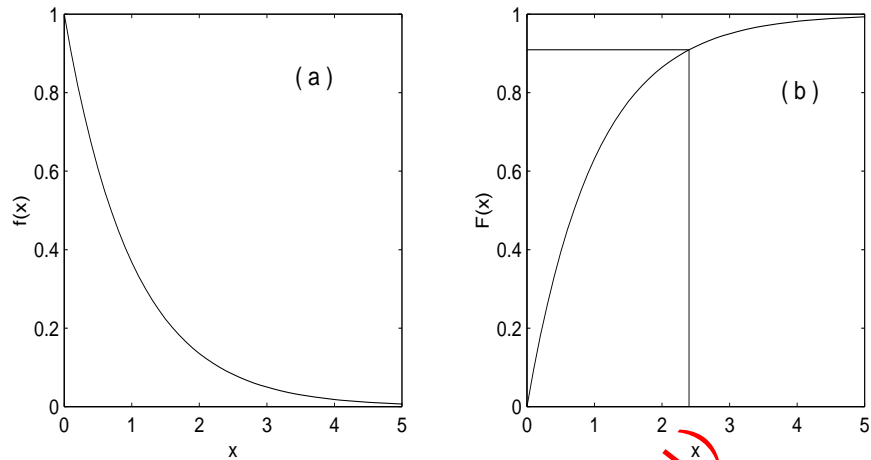


Figure 10: (a) exponential distribution (b) cumulative distribution

Assignment 12

Devise a technique to sample from the distribution $f(x) = \exp(-\sqrt{2}|x|)/\sqrt{2}$ for $-\infty \leq x \leq +\infty$. Generate $\{x_i : i = 1, 2, \dots, N\}$. Sum up these numbers and divide by \sqrt{N} . Call this y . Generate several values of y and plot their frequency distribution in the form of a histogram. Carry out this exercise with $N = 1, 2, \dots$ and demonstrate the approach to Gaussian as N becomes larger and larger.

In fact for the exponential distribution we can carry out the inversion analytically. We set $\xi = F(x)$. It is clear that the number $F(x)$ is uniformly distributed between 0 and 1. Hence the probability that it falls between $F(x)$ and $F(x) + dF(x)$ is $dF(x)$, which is equal to $f(x)dx$. Hence $x = F^{-1}(\xi)$ is distributed as per $f(x)$. For exponential distribution we have $F(x) = 1 - \exp(-x)$. Hence $x = -\log(1 - \xi)$. Since $1 - \xi$ is also uniformly distributed we can set $x = -\log(\xi)$.

Assignment 13

Generate N independent random numbers from an exponential distribution. Sum them up and divide by \sqrt{N} ; call the result y . Generate a large number of values of y and plot their frequency distribution. Plot on the same graph the corresponding gamma density and Gaussian and compare.

Assignment 14

Start with N particles, indexed by integers $i = 1, 2, \dots, N$. (e.g. $N = 1024$). Initialize $x(j) = \lambda \forall j = 1, 2, \dots, N$, where λ is the desired exponential decay constant. (e.g., $\lambda = 1$). The algorithm conserves $\sum_{i=1}^N x(i) = N\lambda$. Select independently and randomly two particles, say with indices i and j , and $i \neq j$. Let $S = x(i) + x(j)$. Split S randomly into two parts. Set $x(i)$ to one part and $x(j)$ to the other part. Repeat the above for a warm up time of say $4 \times N$ iterations. Then every subsequent time you select two particles (k and l), the corresponding $x(k)$ and $x(l)$ are two independent random numbers with exponential distribution: $\lambda \exp(-\lambda x)$ for $0 \leq x \leq \infty$.

- (a) Implement the above algorithm and generate a large number of random numbers. Plot their frequency distribution and check if they follow exponential distribution.
- (b) Prove analytically that the above algorithm leads to independent and exponentially distributed random numbers in the limit $N \rightarrow \infty$.

Analytic inversion to generate exponentially distributed random numbers is not necessarily the most robust and fast of the techniques. There are several alternate procedures for sampling from exponential distribution without involving logarithmic transformation, see for example [36]. The subject of developing ingenious algorithms for generating exponentially distributed random numbers continues to attract the attention of the Monte Carlo theorists. For example, recently Fernández and Rivero [37] have proposed a simple algorithm to generate independent and exponentially distributed random numbers. They consider a collection of N particles each having a certain quantity of energy to start with. Two distinct particles are chosen at random; their energies are added up; the sum is divided randomly into two portions and assigned to the two particles. When this procedure is repeated several times, called the warming up time, the distribution of energies amongst the particles becomes exponential. Note that this algorithm conserves the total energy. It has been shown that about one thousand particles are adequate; the warming up time is $4N$ or so. For details see [37]; see also [38].

10.2 Rejection Technique

Another useful random sampling technique is the so called **rejection technique**. The basic idea is simple. From a set of random numbers discard those that do not follow the desired distribution. What is left out must be distributed the way we want.

Let $f(x)$ be the desired distribution, defined over the range (α, β) . First select a suitable bounding function $g(x)$ such that $C \times g(x) \geq f(x)$ for all values of $\alpha \leq x \leq \beta$. Also $g(x)$ must be such that it is easy to sample from. Sample a set of $\{x_i : i = 1, 2, \dots, N\}$ independently and randomly from $g(x)$. For each value of x_i select randomly a number

between 0 and $C \times g(x_i)$. Call this set $\{y(x_i)\}$. From the set $\{x_i : i = 1, 2, \dots, N\}$ discard those for which $y(x_i) > f(x_i)$. The remaining x_i -s shall have the desired distribution $f(x)$. The efficiency of a rejection technique is the percentage of the attempts that get accepted and is given by the inverse of the area under the curve $C \times g(x)$. Usually $g(x)$ is taken as a constant for all x and C is the maximum value that $f(x)$ takes. The rejection technique would be inefficient if $C \times g(x)$ and $f(x)$ do not have similar shapes and ranges.

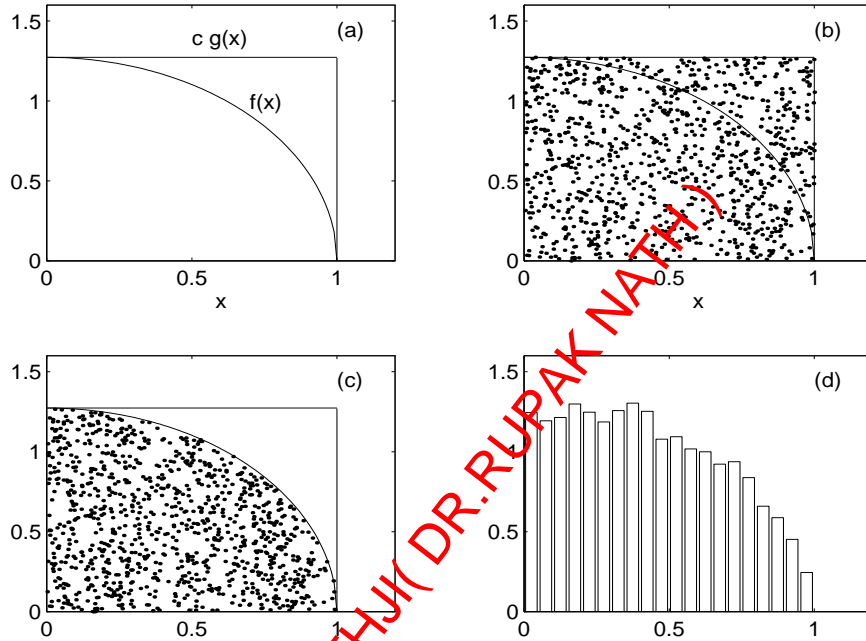


Figure 11: Illustration of the rejection technique. (a) the circular density and the $C \times$ the bounding function. (b) random points that uniformly fill up the rectangular box. (c) points that get accepted in the sampling procedure. (d) the distribution of the accepted values of x .

Let me illustrate the rejection technique by considering the circular probability density function

$$f(x) = \frac{4}{\pi} \sqrt{1-x^2} \quad \text{for } 0 \leq x \leq 1. \quad (105)$$

A convenient choice of the bounding function is $g(x) = 1 \forall 0 \leq x \leq 1$. Thus sampling from $g(x)$ is equivalent to setting $x_i = \xi_i$. The value of C is $4/\pi$. Fig. 11a depicts the density function $f(x)$ and the function $C \times g(x)$. Generate a pair of random numbers (ξ_i, ξ_j) from $U(0, 1)$. Set $x_i = \xi_i$. Calculate $f_i = f(x_i)$ and $y_i = C \times \xi_j$. If $y_i \leq f_i$, accept x_i . Repeat the above procedure several times. Fig. 11b depicts $\{(x_i, y_i); i = 1, 2, \dots, 1000\}$ and these points are seen to fill up the rectangle bounded by the lines $y = 0, y = 4/\pi, x = 0$ and $x = 1$. Fig. 11c depicts the accepted pairs $\{(x_i, y_i)\}$. Fig. 11d depicts the histogram of the distribution of the accepted values of x_i .

In the introduction, I mentioned of the ‘pebble-throwing’ game popular in the province of Monaco, from which the name Monte Carlo came. What has been described above is a straight forward adaptation of the same (Monte Carlo) game to sampling from a

(normalized) distribution. I said earlier that one can make an estimate of the value of π from the Monte Carlo game. How do we do this ?

In the rejection technique described above, we sample a point randomly from the rectangular region, see Fig. 11a. In the next step we either accept the point or reject it. Thus there are only two outcomes to the experiment. The point either falls inside the ‘circular distribution’ curve (and hence gets accepted) with a probability $p = \pi/4$ or outside (and hence gets rejected) with a probability $q = 1 - p = 1 - \pi/4$. We are essentially tossing a loaded coin; the probability of Heads is p and of Tails $q = 1 - p$. Let n denote the number of Heads in N independent throws of a loaded coin. The distribution of the random variable n is Binomial. The quantity of interest to us is n/N , whose mean and standard deviation can easily be calculated, and are given by p and $\sqrt{p(1-p)}/\sqrt{N}$ respectively. Thus we say the Monte Carlo estimate of π is $4n/N \pm (4/N^{3/2})\sqrt{n(N-n)}$, where the error term is the one-sigma confidence interval. Fig. 12 depicts the estimated value of π with the one-sigma confidence error bar, as a function of the logarithm (to the base 2) of the number of trials $N = 2^{10}, 2^{11}, \dots, 2^{18}$. The statistical convergence to the right answer as the sample size increases is seen.

Assignment 15

Employ rejection technique to sample from $f(x) = \frac{2}{\pi}\sqrt{1-x^2}$ for $-1 \leq x \leq 1$. Plot the distribution of the accepted sequence of numbers in the form of a histogram and compare with the exact. Calculate the value of π from the above experiment ? What is the statistical error ?

10.3 Sampling from a Gaussian

The Gaussian is the most important distribution in statistics and statistical physics. It is also one of the *richest*, in the sense, a very large number of techniques have been proposed to generate Gaussian random numbers. The simplest perhaps is the one that makes use of the Central Limit Theorem: take the sum of N random numbers, uniformly distributed in the range $(0, 1)$. Let Y_N denote the sum. By Central Limit Theorem, the distribution of Y_N will tend to a Gaussian in the limit $N \rightarrow \infty$. The mean of the Gaussian is $N/2$ and its variance is $N/12$. Usually one wants a standard Gaussian with mean zero and variance unity. Therefore we define the quantity q ,

$$q = \frac{Y_N - (N/2)}{\sqrt{N/12}}, \quad (106)$$

which has a Gaussian distribution of zero mean and unit variance for large N . A convenient choice is $N = 12$, which reduces q to $Y_{12} - 6$: Take twelve pseudo random numbers and add them up; subtract six from the sum and you have a mean zero variance unity Gaussian random number. This method of generating Gaussian random numbers is not exact since the tails are absent. Note that the value of q is restricted to the interval $(-6, +6)$.

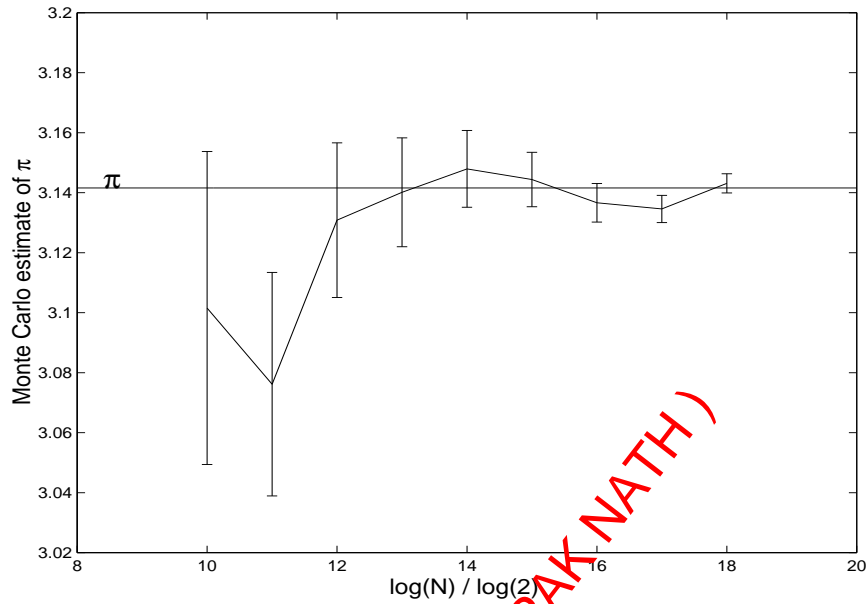


Figure 12: Monte Carlo estimate of π employing rejection technique

There is a clever technique that transforms two independent uniform random variables ξ_1 and ξ_2 to two independent Gaussian random variables q_1 and q_2 , as follows.

$$\begin{aligned} q_1 &= \sqrt{-2 \ln \xi_1} \cos(2\pi\xi_2), \\ q_2 &= \sqrt{-2 \ln \xi_1} \sin(2\pi\xi_2). \end{aligned} \quad (107)$$

This method, called **Box-Muller algorithm** [39], is easy to program. This algorithm is quite popular. However it has a few drawbacks. Box-Muller algorithm uses several multiplication, one logarithmic function, one trigonometric function and one square root function; hence it is rather slow. Besides the tails differ markedly from the true when a linear congruential generator is used.

Muller [40] gives an account of several methods of sampling from a Gaussian distribution.

Assignment 16

Generate N random numbers from $U(0, 1)$. Add them up and subtract $N/2$. Divide the result by $\sqrt{N/12}$. Call this y . Generate a large number of y and plot their frequency distribution. Take $N = 2, 3, 4, \dots$, and demonstrate the approach to Gaussian (of mean zero and variance unity).

Recently Fernández and Criado [41] have proposed a fast algorithm to generate Gaussian random numbers. Their algorithm is based on an N particles closed system interacting

two at a time, conserving energy. Start with N particles each having the same velocity unity. *i.e.* $\{v_i = 1 \forall i = 1, 2, \dots, N\}$. Pick up two particles at random; let them be i and j , and $i \neq j$. Reset their velocities as per the following iteration rule,

$$\begin{aligned} v_i(\text{new}) &= \frac{v_j(\text{old}) + v_i(\text{old})}{\sqrt{2}} \\ v_j(\text{new}) &= \frac{v_j(\text{old}) - v_i(\text{old})}{\sqrt{2}}. \end{aligned} \quad (108)$$

Repeat the above several times. After initial *warm up time* of say $4N$ iterations or so, the velocities of the pair of particles you pick up in all subsequent iterations are the desired pairs of independent Gaussian random numbers with mean zero and variance unity. Gaussian of desired mean, say μ and standard deviation, say σ can be obtained by the transformation $x = \sigma v + \mu$. Note that $\sum_{i=1}^N v_i^2 = N$ at all stages of iteration. This algorithm is found to be ten times faster than the Box-Muller algorithm. For most applications, it is adequate if ten thousand to hundred thousand particles are considered, see [41] for details.

Assignment 17

- (a) Prove analytically that the Fernández-Criado algorithm leads to independent Gaussian random numbers. Prove that the sampling is ergodic. See [41].
- (b) Generate a large number of Gaussian random numbers employing Fernández-Criado algorithm and plot their frequency distribution. Compare with the exact Gaussian.

We have implemented the algorithm proposed by Fernández-Criado and a sample result is depicted in Fig. 13. Ten thousand particles were considered. The first 40,000 iterations were considered as *warm up time*. We generated twenty thousand numbers and collected them in thirty equal bins. The statistical error (one - sigma confidence interval) on the counts in each bin were also calculated from the sample fluctuations. These are also depicted in Fig. 13.

10.4 Metropolis Sampling

An interesting technique to sample from a probability distribution is the one proposed by Metropolis and his collaborators [42], in the context of generating microstates belonging to a canonical ensemble. **Metropolis algorithm** is widely used in Monte Carlo simulation of models in statistical physics.

Here I shall illustrate the technique for sampling from an arbitrary discrete distribution $f(i)$, where the random variable takes discrete integer values i between 1 and N . We call $1, 2, \dots, N$ as states and denote the set of all states by $\Omega = \{1, 2, \dots, N\}$. Start

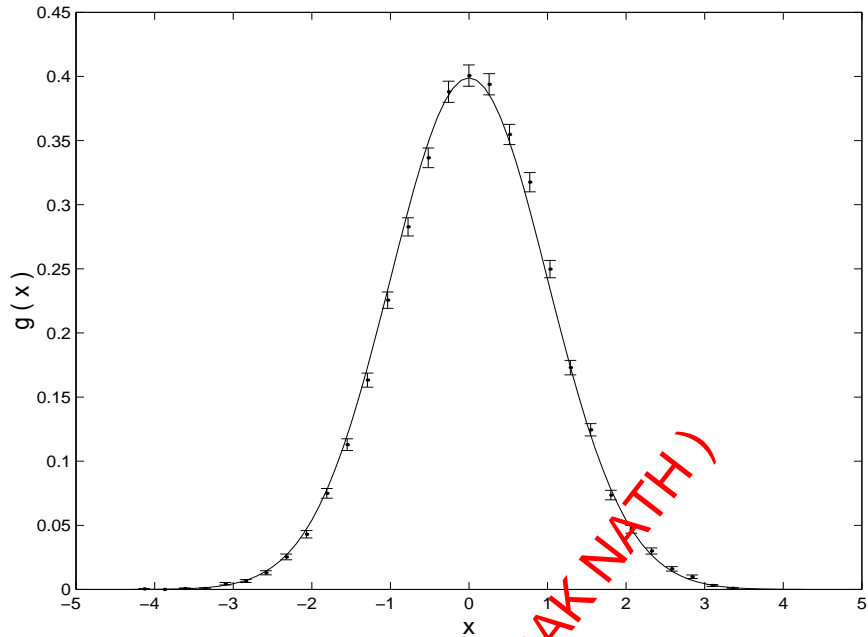


Figure 13: Sampling from a Gaussian of mean zero and variance unity employing the technique proposed by Fernández and Criado [39]. 10,000 particles were considered; warm up time was 40,000 iterations; 20,000 random numbers were generated and assembled in thirty equal bins; the error bar corresponds to one-sigma confidence in each bin. The solid curve is the exact Gaussian

with an initial arbitrary state x_0 belonging to Ω . The Metropolis sampling technique generates a **Markov chain** of states $x_1(\in \Omega), x_2(\in \Omega), \dots, x_{m-1}(\in \Omega)$. For $m \rightarrow \infty$, $\{x_{m+1}(\in \Omega), x_{m+2}(\in \Omega), \dots\}$ shall have the distribution $\{f(i) : i = 1, 2, \dots, N\}$. A Markov chain is a stochastic process whose ‘past’ has no influence on the future if its ‘present’ is specified. In other words, for a Markov Chain, we have,

$$P(x_k \leq x | x_{k-1}, x_{k-2}, \dots, x_0) = P(x_k \leq x | x_{k-1}). \quad (109)$$

We call $\{x_{m+1}, x_{m+2}, \dots\}$ the desired ensemble of states. If N_i is number of elements in the ensemble whose state index is i , then N_i divided by the total number of members in the ensemble is the probability $f(i)$. For such asymptotic convergence to the desired ensemble, it is sufficient *though not necessary* that

$$f(x_i) W(x_j \leftarrow x_i) = f(x_j) W(x_i \leftarrow x_j). \quad (110)$$

where $W(x_j \leftarrow x_i)$ is the probability of transition from state x_i to state x_j . Equation (110) is called the **detailed balance**. For conservation of probability, the transition probabilities should obey the following condition,

$$\sum_i W(i \leftarrow j) = 1 \quad \forall j. \quad (111)$$

The detailed balance condition does not specify uniquely the transition probabilities from one state to the other. A simple choice of the $N \times N$ **transition matrix** W with elements $W_{i,j} = W(i \leftarrow j)$ which is consistent with the detailed balance is given by,

$$\begin{aligned} W_{i,j} &= W_{i,j}^* \min \left(1, \frac{f(i)}{f(j)} \right) \text{ for } i \neq j, \\ W_{j,j} &= W_{j,j}^* + \sum_{i=1}^N W_{i,j}^* \max \left(0, 1 - \frac{f(i)}{f(j)} \right), \end{aligned} \quad (112)$$

where the matrix W^* with elements $W_{i,j}^*$ is an arbitrary symmetric stochastic matrix with positive elements. We call W^* the **trial** matrix. The sum of the elements in each row as well as each column of the trial matrix W^* is unity. As we shall see below, we need W^* matrix to select a trial state from the current state. Hence W^* is chosen conveniently to make the selection of the trial state simple.

Assignment 18

Verify that the transition matrix W whose elements are calculated as per the prescriptions in Eqns. (112) obeys the detailed balance given by Eq. (110).

The implementation of the Metropolis sampling procedure proceeds as follows. Let $x_i (\in \Omega)$ be the current state (or value). We select a trial state x_t , randomly and with equal probability from amongst the N states of $\Omega = \{1, 2, \dots, N\}$. This in other words means that $W_{i,j}^* = 1/N \forall i, j = 1, 2, \dots, N$. Calculate the ratio $w = f(x_t)/f(x_i)$. If $w \geq 1$, accept the trial state and set $x_{i+1} = x_t$. If $w < 1$, then generate a random number ξ uniformly distributed in the range 0 to 1. If $\xi \leq w$, then accept the trial state and set $x_{i+1} = x_t$. If $\xi > w$, reject the trial state and set $x_{i+1} = x_i$ and proceed. It may be necessary to generate several values of x , starting from an initial choice x_0 , before the string acquires the desired distribution. A good choice of x_0 is that state for which the probability is maximum.

Let us consider a simple example with three states $\Omega = \{1, 2, 3\}$ with $|f\rangle = (f_1, f_2, f_3)' = (1/4, 5/12, 1/3)'$. The matrix W^* is taken as $W_{i,j}^* = 1/3 \forall i, j$. The transition matrix constructed by the prescription, see Eq. (112), is given by,

$$W = \begin{pmatrix} 1/3 & 1/5 & 1/4 \\ 1/3 & 8/15 & 1/3 \\ 1/3 & 4/15 & 5/12 \end{pmatrix}. \quad (113)$$

As can be readily seen, the matrix W above, has the following properties.

1. $W_{i,j} > 0 \forall i, j$; This is called the **strong ergodicity** condition.
2. $\sum_{i=1}^3 W_{i,j} = 1 \forall j$. As mentioned earlier, this condition ensures the conservation of the probability. The transpose of the matrix W is usually called a stochastic matrix.

The eigenvalues of W are 1.0, 0.2, and 0.0833; the largest eigenvalue is unity and is non degenerate. The corresponding right eigenvector is $|f\rangle = (1/4, 5/12, 1/3)'$ i.e., $W|f\rangle = |f\rangle$.

Assignment 19

Construct a matrix W : $W_{i,j} > 0 \forall i, j$; and $\sum_i W_{i,j} = 1 \forall j$. Calculate its eigenvalues and the corresponding left and right eigenvectors. Operate W on an arbitrary vector several times and show that the resulting vector converges to a unique vector which is the eigenvector of the matrix W , corresponding the largest eigenvalue unity.

Thus repeated application of W on any arbitrary vector $|u\rangle$ with $\langle f|u\rangle \neq 0$, will asymptotically take the vector to $|f\rangle$. We say $|f\rangle$ is the equilibrium probability vector representing the equilibrium ensemble. Any initial ensemble represented by $|u\rangle$ with non-zero overlap with the equilibrium ensemble, will evolve to the equilibrium ensemble. The above results are true in general for any positive stochastic matrix W , by virtue of the **Peron theorem** [43].

Perons's theorem states that *a positive matrix W has an eigenvalue λ , which is real, positive and non-degenerate and which exceeds in modulus all the other eigenvalues. To this dominant eigenvalue there corresponds an eigenvector with all its elements positive.* In our example, the transpose of the positive matrix W is stochastic, i.e., $\sum_i W_{i,j} = 1$, and hence the dominant eigenvalue is unity.

Peron's theorem can be seen as follows. Consider the following left eigenvalue equation,

$$\langle L| W = \langle L| \lambda, \quad (114)$$

where $\langle L|$ is the left eigenvector corresponding to the eigenvalue λ . We can obtain the upper and lower bounds of the eigenvalue as follows.

$$\lambda L_j = \sum_{i=1}^N W_{i,j} L_i; \quad (115)$$

$$L_{min} \sum_{i=1}^N W_{i,j} \leq \lambda L_j \leq L_{max} \sum_{i=1}^N W_{i,j}, \quad (116)$$

where $L_{max} = \max\{L_k\}$ and $L_{min} = \min\{L_k\}$. Since $\sum_{i=1}^N W_{i,j} = 1$, it follows,

$$L_{min} \leq \lambda L_j \leq L_{max}. \quad (117)$$

Consider the space of all positive vectors $\{\langle L|; L_j > 0 \forall j = 1, 2, \dots, N\}$. Then

$$\frac{L_{min}}{L_j} \leq \lambda \leq \frac{L_{max}}{L_j} \quad \forall j. \quad (118)$$

The minimum value that L_{max}/L_j can take as j runs from 1 to N is unity; similarly, the maximum value that L_{min}/L_j can take is unity. Thus we get $1 \leq \lambda \leq 1$, which implies that $\lambda = 1$. Thus if the left eigenvector is positive then the corresponding eigenvalue is unity.

Consider now the space of vectors $\{\langle L|\}$ such that each vector has some of its components positive and some negative. We note that $L_{max} > 0$ and $L_{min} < 0$. The bounds can be easily worked out and are given by,

$$\max \left\{ \frac{L_{max}}{L_{min}}, \frac{L_{min}}{L_{max}} \right\} \leq \lambda \leq 1. \quad (119)$$

Thus we see that $\lambda = 1$ is the largest eigenvalue and the corresponding left eigenvector is positive. It is easily verified that $\langle U| = c(1 \ 1 \ 1 \ \dots \ 1)$, the constant vector, is the left eigenvector corresponding to the eigenvalue $\lambda = 1$. The constant c can be fixed by choosing a suitable norm.

We can easily show that the eigenvalue $\lambda = 1$ is non degenerate. To this end, assume the contrary. Let $\langle V|$ be another positive eigenvector corresponding to the eigenvalue $\lambda = 1$. Then a linear combination of $\langle U|$ and $\langle V|$ given by $\langle \eta| = \alpha \langle U| + \beta \langle V|$ is also an eigenvector. We can choose α and β such that $\langle \eta|$ has some components positive and some negative which contradicts the fact that the eigenvector is positive. This completes the proof.

Let us now consider the right eigenvector $|R\rangle$, corresponding to $\lambda = 1$. It is easily proved that $|R\rangle$ is positive. Consider the eigenvalue equation

$$(W_{1,1} - 1)R_1 + W_{1,2}R_2 + \dots + W_{1,N}R_N = 0. \quad (120)$$

Let us assume R_1 is negative. The first term in Eq. (120) is positive since $W_{1,1} < 1$. Hence at least one of the other components of $|R\rangle$ must be negative. In other words one of the elements of the set $\{R_2, R_3, \dots, R_N\}$ must be negative to render the sum in Eq. (120) zero. Without loss of generality we take R_2 to be negative. Consider the eigenvalue equation,

$$W_{2,1}R_1 + (W_{2,2} - 1)R_2 + W_{2,3}R_3 + \dots + W_{2,N}R_N = 0. \quad (121)$$

Add Eq. (120) to Eq. (121) and get,

$$(W_{1,1} + W_{2,1} - 1)R_1 + (W_{1,2} + W_{2,2} - 1)R_2 + \sum_{j=3}^N (W_{1,j} + W_{2,j})R_j = 0. \quad (122)$$

In the above, $(W_{1,1} + W_{2,1}) < 1$, $R_1 < 0$, $(W_{1,2} + W_{2,2}) < 1$ and $R_2 < 0$. Therefore we find that the first two terms in the above equation are positive. Hence at least one of the elements of the set $\{R_3, R_4, \dots, R_N\}$ must be negative to render the sum zero. Without loss of generality we take this as R_3 . Arguing along the same lines we show that if R_1 is negative then all the other components of the vector $|R\rangle$ are also negative. This in other words means that the right eigenvector corresponding to the largest eigenvalue $\lambda = 1$ is positive. We call $|R\rangle$ the equilibrium eigenvector.

Numerically the equilibrium ensemble for the three-state model is constructed as follows. Start with an arbitrary state $x_0 \in \Omega$; select a trial state $x_t \in \Omega$ with equal probability.

Note that we have taken $W_{i,j}^* = 1/3 \forall i, j$. Calculate $w = f(x_t)/f(x_0)$. Accept the trial state as the next state with a probability given by minimum of $(1, w)$. If accepted set $x_1 = x_t$; if rejected set $x_1 = x_0$; and proceed. Thus we get an ensemble of states.

Peron's theorem is special case of a more general theorem on the eigenvectors and eigenvalues of non-negative matrices proved later by Frobenius. A proof of the **Frobenius theorem** can be found in the book on matrices by Gantmacher [44]. As per this theorem, asymptotic convergence to equilibrium vector is possible even under weaker condition. Some or several elements of the transition matrix W can be zero. In other words $W_{i,j} \geq 0 \forall i, j$ and of course $\sum_I W_{i,j} = 1$. In such cases it is adequate that $(W^m)_{i,j} > 0 \forall i, j$ and for an $m < \infty$; m may depend on i and j . Physically this means that any state is accessible from any other state in finite number of steps.

This weaker condition helps us make a better choice of the trial matrix W^* . Some or several elements of this matrix can be zero. In practical terms it means that the trial state x_t can be selected randomly from a small neighbourhood of the current state x_i . Thus we set $x_t = x_i + \eta_i$ where η_i is a random integer from $-\epsilon$ to $+\epsilon$. In this approach, it may so happen that the trial value lies outside the range of x . In such a situation, keep generating new trial values until, you get one within the range. The choice of ϵ is crucial. If ϵ is too large, then the fraction of the accepted trials would be too small and the sampling poor and inefficient. On the other hand if ϵ is too small, then even though a large number of trials would get accepted, the value of x would remain close to the starting point over several trials and hence not span the entire range of x quickly and efficiently. A good criterion is to fix ϵ such that the acceptance is between 30% and 50%.

Assignment 20

Employ Metropolis sampling technique to sample from a) binomial distribution and b) Poisson distribution.

In the above we have discussed the Metropolis technique with respect to discrete distributions. The method can be readily extended to sample from continuous distribution.

Fig. 14 depicts the results of sampling from an exponential density employing Metropolis sampling.

A very large number of transformations, tricks, and formulae have been discovered for sampling from a variety of non-uniform distributions. I shall not discuss them here. These techniques are scattered over several publications on Monte Carlo and its applications to different disciplines.

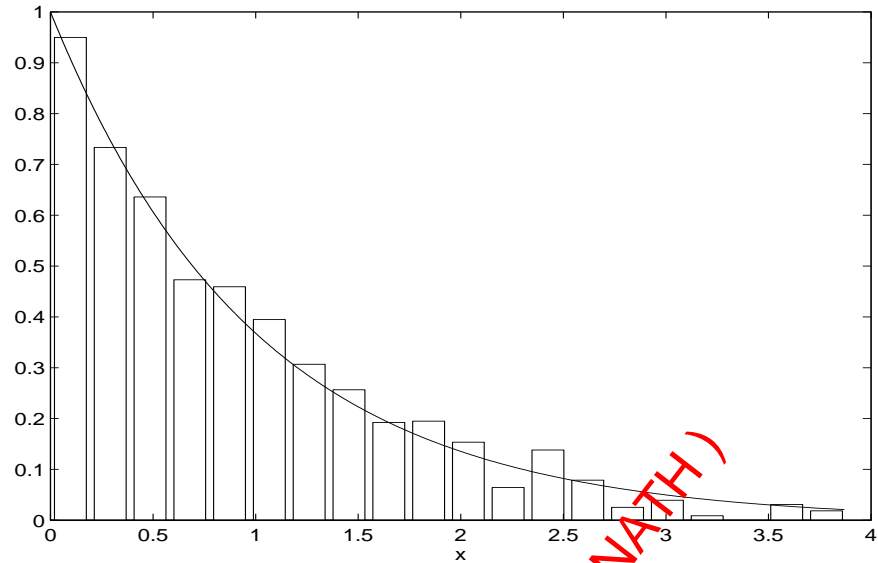


Figure 14: Sampling from an exponential density employing Metropolis algorithm. $\epsilon = 3$ gives an acceptance of 30% – 50%. The solid curve is the exact exponential distribution. The histogram represents the frequency distribution of a sample of size 10,000.

Assignment 21

Employ Metropolis technique to sample from the distribution $f(\theta) = \exp[a \cos(\theta)]$, for $0 \leq \theta \leq \pi$, where a is a parameter. The optimal value of ϵ that would give 30% to 50% rejection would depend on the value of a . Compare your results with the exact.

11 ANALOGUE MONTE CARLO

Consider evaluation of the expectation value (also known as the mean) of a function h of the random variable X . Formally we have,

$$\mu = \int_{-\infty}^{+\infty} h(x)f(x)dx, \quad (123)$$

where $f(x)$ is the probability density function of the random variable X . $h(x)$ is usually called the score function. How do we estimate μ ? Let us first consider **analogue simulation**. A Monte Carlo simulation is called analogue simulation, if it does not employ any variance reduction devices. The simulations that employ variance reduction techniques we shall term as biased simulation. In analogue Monte Carlo, we sample randomly a sequence

of $\{x_i : i = 1, 2, \dots, N\}$, from the density $f(x)$ and write,

$$\bar{h}_N = \frac{1}{N} \sum_{i=1}^N h(x_i). \quad (124)$$

In the limit $N \rightarrow \infty$, $\bar{h}_N \rightarrow \mu$. Also by the Central Limit Theorem, in the limit $N \rightarrow \infty$, the probability density of \bar{h}_N tends to a Gaussian with mean μ and variance σ^2/N , where

$$\sigma^2 = \int_{-\infty}^{+\infty} [h(x) - \mu]^2 f(x) dx. \quad (125)$$

Thus we say that **Analogue Monte Carlo** estimate of μ is given by $\bar{h}_N \pm \sigma/\sqrt{N}$, where $\pm\sigma/\sqrt{N}$ defines the one-sigma confidence interval. This means that with a probability p given by,

$$\begin{aligned} p &= \frac{\sqrt{N}}{\sigma\sqrt{2\pi}} \int_{\mu-(\sigma/\sqrt{N})}^{\mu+(\sigma/\sqrt{N})} \exp\left[-\frac{N(x-\mu)^2}{2\sigma^2}\right] dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-1}^{+1} \exp\left[-\frac{x^2}{2}\right] dx \\ &= 0.68268, \end{aligned} \quad (126)$$

we expect \bar{h}_N to lie within $\pm\sigma/\sqrt{N}$ around μ , if N is sufficiently large.

First we note that we do not know σ . Hence we approximate it by its Monte Carlo estimate S_N , given by

$$S_N^2 = \frac{1}{N} \sum_{i=1}^N h^2(x_i) - \left[\frac{1}{N} \sum_{i=1}^N h(x_i) \right]^2. \quad (127)$$

The quantity $\pm S_N/\sqrt{N}$ is called the **statistical error**. Notice that the sample size N must be large for the above estimate of the error (and of course the mean) to hold well. Sometimes it would worth the effort to test if the sample mean has acquired a normal distribution, see for example the test devised by Shapiro and Wilks [46]. Normality tests are useful in a biased Monte Carlo simulation.

The statistical error decreases as inverse of the square root of the sample size. This is rather slow - logarithmically slow. For example if we want to decrease the statistical error by a factor of two we must increase the sample size by a factor of four. The computing time increases linearly with the sample size. Hence invariably one finds that analogue Monte Carlo is practically impossible. But there is a way out.

The way out is to resort to techniques that reduce the variance without altering the mean. These are called **variance reduction techniques** and in what follows I shall describe the basic principle behind variance reduction techniques through what is called the **importance sampling**.

12 IMPORTANCE SAMPLING

Importance sampling helps us sample from the important regions of the sample space. Consider the problem described in the last section. We sample $\{x_i : i = 1, 2, \dots, N\}$ from an importance density $g(x)$ instead of the analogue density $f(x)$. To preserve the mean we define a modified score function $H(x)$ given by,

$$H(x) = h(x) \frac{f(x)}{g(x)}. \quad (128)$$

The expectation value of H is evaluated over the importance density $g(x)$; this is identically equal to the expectation value of the original score function h over the analogue density $f(x)$:

$$\begin{aligned} \mu(H) &= \int_{-\infty}^{+\infty} H(x)g(x)dx \\ &= \int_{-\infty}^{+\infty} h(x) \frac{f(x)}{g(x)} g(x) dx \\ &= \int_{-\infty}^{+\infty} h(x) f(x) dx \\ &= \mu(h). \end{aligned} \quad (129)$$

Thus we sample $\{x_i : i = 1, 2, \dots, N\}$ from $g(x)$, and calculate the ratios $\{f(x_i)/g(x_i)\}$. The biased Monte Carlo estimate of μ is given by

$$\bar{H}_N = \frac{1}{N} \sum_{i=1}^N h(x_i) \frac{f(x_i)}{g(x_i)}. \quad (130)$$

In the limit $N \rightarrow \infty$, $\bar{H}_N \rightarrow \mu$. Let us now calculate the statistical error associated with \bar{H}_N . It is adequate if we consider the second moment, since we have formally shown that the mean is preserved under importance sampling. We have

$$\begin{aligned} M_2^B(H) &= \int_{-\infty}^{+\infty} H^2(x)g(x)dx \\ &= \int_{-\infty}^{+\infty} \frac{h(x)f(x)}{g(x)} \frac{h(x)f(x)}{g(x)} g(x) dx \\ &= \int_{-\infty}^{+\infty} \left[\frac{f(x)}{g(x)} \right] h^2(x) f(x) dx. \end{aligned} \quad (131)$$

For the analogue simulation,

$$M_2^A(h) = \int_{-\infty}^{+\infty} h^2(x) f(x) dx. \quad (132)$$

Thus if we choose properly the importance function $g(x)$ and ensure that the ratio $f(x)/g(x)$ on the average is substantially less than unity, then we can make $M_2^B(H)$ to be much less than $M_2^A(h)$. This in essence is the basic principle of variance reduction techniques. Thus, sampling from an importance density helps us estimate the mean with a much better statistical accuracy for a given sample size. In fact it is due to the variance reduction techniques that Monte Carlo simulation of many problems have become possible.

12.1 Exponential Biasing

Let me illustrate the use of importance function on a simple problem where all the relevant quantities can be calculated analytically. The problem has its origin in radiation transport through thick shields. The actual problem of Monte Carlo simulation of the transport is relatively complex and a detailed discussion of this is not relevant for the purpose here. Hence, I shall be brief. Consider particles, neutrons or gammas, incident normally on the left face of a slab of material of thickness T . The particle would penetrate a distance x into the shield with a probability $\Sigma \exp(-\Sigma x)$ and enter into a collision event at a point between x and $x+dx$ with probability Σdx . Here Σ^{-1} is the mean free path (mfp). We can measure all distances in units of mfp and hence set $\Sigma = 1$. The collision can lead to absorption in which case the particle history is terminated. On the other hand if the collision is a scattering event, the history continues. In a simple-minded picture, the scattering is taken as isotropic. In one-dimensional model, this means that the particle is directed toward the right or the left face of the slab, with equal probability. The particle travels along the scattered direction and has a collision event at some distance. The history continues and is constituted by a series of alternating free flights and collisions. The history ends when the particle is absorbed inside the shield or escapes the shield. When the particle escapes through the right face it scores unity; otherwise the score is zero. The scores from a large number of particle histories are accumulated. The average of these scores is the Monte Carlo estimate of the mean transmission defined as the fraction of the incident particles that escape through right face of the shield. The sample fluctuations give an estimate of the statistical error. If the thickness of the shield exceeds 20 mfp or so, the problem becomes intractable by analogue Monte Carlo. Variance reduction techniques become imperative.

An importance sampling technique often used in this context is called the **exponential biasing**. In the simplest version of exponential biasing, the inter collision distance is sampled from the importance density $b \exp(-bx)$, where the biasing parameter b has to be optimized for minimum variance. For more details on exponential biasing see [47].

In the simple problem, we shall be interested in calculating the fraction of the incident particles that escape through the right face of the shield, without undergoing any collision whatsoever. This amounts to the following: In analogue simulation, sample a random number from the exponential density. If this exceeds T , score unity. Otherwise score zero. Collect the scores from N histories and calculate the mean and the statistical error. In the biased simulation we sample x from the importance density function $\hat{b} \exp(-\hat{b}x)$, where we have assumed that we know the value of $b = \hat{b}$, for which the variance is minimum. Let us denote this sampled value of x as x_i . If x_i exceeds T , then score $w_i = f(x_i)/g(x_i, \hat{b}) = \exp[-x_i(1 - \hat{b})]/\hat{b}$. Otherwise score zero. Accumulate the scores from a large number of histories and calculate the mean and the statistical error.

Fig. 15a depicts the probability density function $f(x)$ and the score function $h(x)$ for the case with $T = 5$ mfp. Fig. 15b depicts the importance density function $g(x, b = \hat{b})$ and the modified score function $H(x) = f(x)h(x)/g(x, \hat{b})$ for the case with $T = 5$ mfp.

For this problem, all the quantities can be calculated analytically. The expressions for

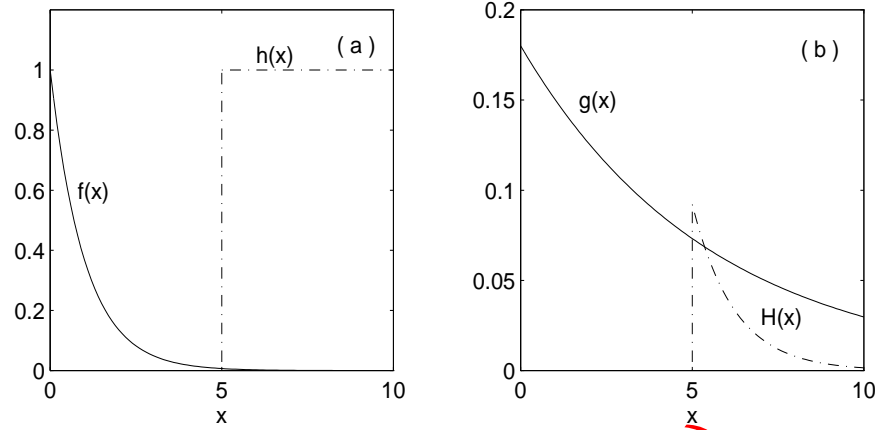


Figure 15: (a) The probability density $f(x)$ and the score function $h(x)$ for $T = 5$ mfp; (b) The importance density function $g(x)$ and the modified score function $H(x) = h(x)f(x)/g(x)$. $\hat{b} = 0.18$

the average and variance are,

$$\mu = \int_0^{\infty} h(x) \exp(-x) dx = \exp(-T), \quad (133)$$

$$\sigma^2 = (1 - \mu) \mu. \quad (134)$$

Table (6) gives the mean, relative standard deviation and the number of histories that would be required to estimate the mean in an analogue simulation, within $\pm 10\%$ of the exact.

Table 6: Analogue simulation: exact analytical results

	μ	σ/μ	N
3	4.98×10^{-2}	4.37	1.91×10^3
5	6.74×10^{-3}	1.21×10^1	1.47×10^4
10	4.54×10^{-5}	1.48×10^2	2.20×10^6
20	2.06×10^{-9}	2.20×10^4	4.85×10^{10}

We see from Table (6) that as the thickness increases the mean transmission decreases and the relative fluctuation increases. We find that the number of histories required to predict the mean within $\pm 10\%$ statistical error is over 48 billion for $T = 20$ mfp, a task which is clearly impossible on any computer.

Let us see how the use of importance sampling renders possible the impossible. Under importance sampling, the variance can be calculated analytically and is given by,

$$\sigma^2(b) = \frac{e^{-T(2-b)}}{b(2-b)} - e^{-2T}. \quad (135)$$

It is quite straight forward to calculate the value of \hat{b} for which the variance is minimum. We get,

$$\hat{b} = 1 + \frac{1}{T} - \sqrt{1 + \frac{1}{T^2}}. \quad (136)$$

Table (7) presents $T, \hat{b}, \sigma/\mu$ and the number of histories required to estimate the mean, under biased simulation within $\pm 10\%$ of the exact.

Table 7: Biased Simulation: exact analytical results

T	\hat{b}	σ/μ	N
3	.28	1.95	381
5	.18	2.55	651
10	.09	3.65	1329
20	.048	5.18	2687

We find from Table (7) that the use of importance sampling would lead to a considerable reduction of variance especially for large T . Take the case with $T = 5$ mfp. The use of importance sampling would reduce the statistical error by a factor five or so. As a consequence a sample of size 650 is adequate to estimate the mean whereas analogue simulation would require a sample of size 14,700. The results for $T = 20$ mfp are more dramatic and bring home the need and power of importance sampling. The statistical error gets reduced by a factor of 4247. We need to simulate only 2687 histories to estimate the mean within $\pm 10\%$ of the exact, as compared to 48 billion histories required under analogue simulation.

We have simulated explicitly 10000 histories under importance sampling, as follows. We sample x_i from the importance density, $\hat{b} \exp(-\hat{b}x)$, and calculate the mean of the modified score function as,

$$\bar{H}_N = \frac{1}{N} \sum_{i=1}^N H(x_i), \quad (137)$$

where,

$$H(x_i) = \begin{cases} \frac{1}{b} \exp \left[- (1 - \hat{b}) x_i \right], & \text{if } x_i \geq T, \\ 0, & \text{if } x_i < T. \end{cases} \quad (138)$$

The statistical error is calculated as,

$$\Delta \bar{H}_N = \pm \frac{1}{\sqrt{N}} \sqrt{\frac{1}{N} \sum_{i=1}^N H^2(x_i) - \bar{H}_N^2}. \quad (139)$$

Table (8) gives the estimated mean transmission \bar{H}_N , the relative statistical error, and the actual deviation of the Monte Carlo estimate from the exact transmission.

Table 8: Results of Monte Carlo simulation of 10,000 histories with importance sampling and comparison with exact analytical calculations

T	\bar{H}_N	$\frac{\Delta \bar{H}_N}{\bar{H}_N} \times 100$	$\frac{\bar{H}_N - \mu}{\mu} \times 100$
3	4.94×10^{-2}	$\pm 2.0\%$	-0.8%
5	6.76×10^{-3}	$\pm 2.6\%$	$+0.3\%$
10	4.52×10^{-5}	$\pm 3.7\%$	-0.4%
20	1.96×10^{-9}	$\pm 5.4\%$	-4.9%

We observe from Table (8) that we are able to make a good estimate of the mean transmission employing importance sampling. The corresponding results obtained by analogue simulation of 50000 histories (five times more than what we have considered for analogue simulation) are given in Table (9).

We find from Table (9) that analogue simulation of the problem with $T = 20$ mfp is impossible. On the average, we can expect one in 49 million numbers sampled from the exponential density, to have a value greater than 20. The chance of getting a score in a simulation of 50000 histories is practically nil.

12.2 Spanier Technique

In several situations, like the one discussed in section 12.1, it is possible to make a good guess of the shape of the importance function $g(x, b)$, where b is an unknown parameter to be optimized for minimum variance. Spanier [48] has proposed a technique for optimizing the parameter b , through processing of the Monte Carlo results from relatively small

Table 9: Results of analogue simulation of 50,000 histories and comparison with exact analytical results

T	\bar{h}_N	$\frac{\Delta \bar{h}_N}{\mu} \times 100$	$\frac{\bar{h}_N - \mu}{\mu} \times 100$
3	5.03×10^{-2}	$\pm 1.9\%$	+1.8%
5	6.68×10^{-3}	$\pm 5.5\%$	-0.9%
10	6.16×10^{-4}	$\pm 57.7\%$	+35.7%
20	.	.	.

samples of histories. The technique consists of expressing the second moment, for the biased problem, see Eq. (131) as,

$$M_2(b_i) = \int \frac{h(x)f(x)}{g(x, b_i)} \frac{h(x)f(x)}{g(x, \tilde{b})} g(x, \tilde{b}) dx, \quad (140)$$

where \tilde{b} is the guess value of the parameter b and $\{b_i : i = 1, 2, \dots, M\}$ are the prescribed values of b spanning its full range or a conveniently chosen interval in its range.

The implementation of **Spanier's technique** proceeds as follows. Start with a guess value \tilde{b} . Generate a set of N values of x by random sampling from the importance density $g(x, \tilde{b})$. Calculate,

$$M_2(b_j) = \frac{1}{N} \sum_{i=1}^N \frac{h(x_i)f(x_i)}{g(x_i, b_j)} \frac{h(x_i)f(x_i)}{g(x_i, \tilde{b})} \quad \text{for } j = 1, 2, \dots, M. \quad (141)$$

Thus we get M_2 at M distinct values of b . Find the value of b at which the second moment is minimum; use this as \tilde{b} in the next iteration and generate a second set of N histories. Repeat the procedure until \tilde{b} converges. N can be small for the purpose of optimizing b .

Fig. 16A presents the results for the problem with $T = 20$ mfp at three stages of iteration starting with an initial guess of $\tilde{b} = 0.1$. The number of histories generated at each stage of iteration is 500. Fig. 16B depicts the second moment as a function of b generated with the converged value of $\tilde{b} = .049$, along with the exact analytical results.

Spanier's technique is quite general and powerful. Application of this technique to radiation transport problem can be found in [48, 49]. Several modifications and improvements to this technique have been recommended and those interested can refer to [50].

13 CLOSING REMARKS

In this monograph I have made an attempt to describe the fundamental aspects of Monte Carlo theory in very general terms without reference to any particular topics like neutron

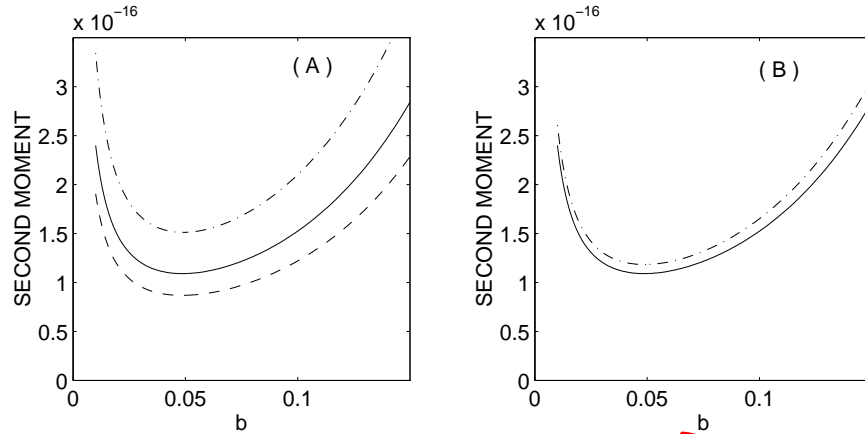


Figure 16: (A) ‘- -’ denotes the results generated by $\tilde{b} = 0.1$; the minimum of the curve occurs at $b = 0.048$; ‘-.-’ represents the results generated with $\tilde{b} = 0.048$ and the minimum occurs at $b = 0.049$. ‘-’ denotes the results generated with $\tilde{b} = 0.049$ and the minimum is found to be at $b = 0.049$. 500 histories were generated at each stage of iteration. (B) ‘-’ depicts the second moment generated by the converged value $\tilde{b} = 0.049$ and ‘-.-’ depicts the exact analytical results.

transport or problems in statistical mechanics.

An understanding of the fundamental ideas of probability theory: sample space, events, probability of events, random variables, probability distribution, moments, cumulants, characteristic function, *etc.*, is a must for appreciating the basis of Monte Carlo techniques. I have given a brief introduction to these topics.

The Central Limit Theorem and its use in estimating Monte Carlo error bars is an important topic and has been dealt with in detail, preceded by a brief discussion on the Chebyshev inequality and the law of large numbers. I have also very briefly touched upon the generalization of the Central Limit Theorem, namely the Lévy stable law.

A long sequence of pseudo random numbers constitutes the backbone of any Monte Carlo simulation. Often one finds that many Monte Carlo practitioners tend to remain ignorant of how these numbers are produced. They often consider the random number generator routine in their personal computer, the workstation or the super computer, as a black box. This tendency must go. The reason is simple. There is no random number generator which is flawless and which would yield random numbers useful for all kinds of simulations. Take for example the most popular and widely used of the random number generators. It is based on linear congruential recursion; it suffers from a very serious defect: the lattice and Marsaglia hyper plane structures that ensue when you embed the random numbers in two and higher dimensional phase space. If your problem is such that the Marsaglia lattice structure would influence your results significantly then you should seriously consider alternate means of generating pseudo random numbers. I have given a brief outline of these and related issues in this monograph.

I have also discussed some tests of randomness of the sequence of pseudo random numbers. Of course we understand that there is absolutely no guarantee that the sequence of pseudo random numbers given by the generator in your machine is adequately random

for the particular Monte Carlo application you have on hand; there can at best be only safeguards: carry out certain tests of randomness, like the uniformity test, correlation tests, run test, the tests based on embedding the sequence in phase space through construction of delay vectors or any other sophisticated tests you may think of. In fact the Monte Carlo algorithm you have developed itself can be used to test the randomness of the pseudo random number sequence.

Random sampling from non-uniform distributions constitutes the core of Monte Carlo simulations. I have dealt with inversion, rejection, and Metropolis techniques for sampling from a given distribution. The last of these, namely the Metropolis algorithm, is usually employed in simulations of problems in statistical mechanics.

Almost all Monte Carlo simulations employ, in some form or the other, techniques of variance reduction. I have tried to convey the basic principles of variance reduction through a discussion of the importance sampling device. I have illustrated the use of importance function by considering exponential biasing on a simple problem. I have also dealt briefly with the Spanier technique for optimizing the biasing parameter in importance sampling.

Let me end by saying that there are several excellent books and review articles on Monte Carlo theory and practice, which you would like to consult during your ‘Monte Carlo learning’. Some of these books that have caught my attention are listed under [5]. Let me wish you all a merry time with Monte Carlo games.

References

- [1] A. Hall, *On an experimental determination of π* , *Messeng. Math.*, **2**, 113 (1873).
- [2] Francis Galton, *Natural Inheritance* (1889), cited in M. Cundy and A. P. Rollet, *Mathematical Models*, Oxford University Press, Oxford (1961).
- [3] L. H. C. Tippett, *Random Sampling Numbers*, Cambridge University Press (1927), cited in [5].
- [4] P. C. Mahalanobis, *Dialectica* **8**, 95 (1954), cited in [5].
- [5] C. Radhakrishna Rao, *Statistics and Truth : Putting Chance to Work*, second edition, World Scientific (1997).
- [6] Lord Kelvin, *Nineteenth century clouds over the dynamical theory of heat and light*, *Phil. Mag.*(6), **2**, 1 (1901).
- [7] W. Krauth, *Introduction to Monte Carlo Algorithm*, cond-mat/9612186 (1996); see <http://lps.ens.fr/krauth>
- [8] J. Bertrand, *Calcul des Probabilités*, Gauthiers-Villars, Paris (1989).
- [9] A. Papoulis, *Probability Theory, Random Variables and Stochastic Processes*, McGraw Hill (1965).

- [10] W. Feller, *An Introduction to Probability Theory and its Applications*, Vols. I and II, John Wiley, New York (1968).
- [11] M. C. Valsakumar, Unpublished notes (1990).
- [12] P. Lévy, *Calcul des Probabilités*, Gauthier Villars, Paris (1925); P. Lévy, *Théorie de l'addition des variables aléatoires*, Gauthier Villars, Paris (1937); for a proof in English of the Lévy stable law, see, A. N. Gnedenko and V. N. Kolmogrov, *Limit Distributions for Sums of Random Variables*, Reading MA, Addison-Wesley (1954).
- [13] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*, North Holland (1981).
- [14] C. W. Gardiner, *A Hand Book of Stochastic Methods for Physics, B Chemistry and Natural sciences*, Springer (1985).
- [15] T. M. John and K. P. N. Murthy, *J. Statis. Phys.* **45**, 753 (1986).
- [16] L. Szilard, *Z. Phys.* **530**, 840 (1929); L. Brillouin, *J. Appl. Phys.*, **22**, 334 (1951); R. Landauer, *IBM J. Res. Dev.*, **5**, 183 (1960); R. J. Solomonoff, *Inf. Control*, **7**,1 (1964); A. N. Kolmogrov, *IEEE Trans. Inf. Theor.*, **1**, 3 (1965); G. J. Chaitin, *J. Assoc. Comput. Mach.*, **13**, 547 (1966). A. Wehrl, *Rev. Mod. Phys.*, **50**, 221 (1978); C. H. Bennet, *Int. J. Theor. Phys.*, **12**, 905 (1982); K. G. Denbigh and J. S. Denbigh, *Entropy in relation to incomplete knowledge*, Cambridge University Press, Cambridge, U.K. (1985); P. C. W. Davies, *Physics of Time Assymetry*, Univ. California Press, Berkeley (1985); G. J. Chaitin, *Information, Randomness and Incompleteness: papers on Algorithmic information theory*, World Scientific, Singapore (1987)(a collection of Chaitin's papers). C. H. Bennet, *IBM J. Res. Dev.*, **32**, 16 (1988); R. Landauer, *Nature*, **335**, 779 (1988); W. H. Zurek, *Phys. Rev.*, **A40**, 4731 (1989); W. H. Zurek, *Phys. Rev.* **A40**, 4731 (1989); I. Prigogine, *From Being to Becoming*, Freeman, San Francisco (1990). C. M. Caves, *Phys. Rev. Lett.*, **64**, 2111 (1990); C. M. Caves, W. G. Unruh and W. Zurek, *Phys. Rev. Lett.*, **65**, 1387 (1990); W. H. Zurek (Ed.), *Complexity, Entropy and the Physics of Information*, Vol. VIII of Santa Fe Institute Studies in the Sciences of Complexity, Addison Wesley (1990); C. H. Bennet, *Sci. Am.* **275**, 106 (1997).
- [17] N. A. Friggario and N. Clark, *Trans. Amer. Nucl. Soc.*, **22**, 283 (1975); N. A. Friggario and N. Clark, *Toward Truly Random Numbers*, Report ANL/ES-26 Part 4, Argonne National Laboratory (1978).
- [18] RAND Corporation, *A million random digits with 100000 normal deviates*, The Free Press (1955).
- [19] Ivar Eklund, *Au hasard*, Editions du Seuil (1991); English Translation: translated by Carol Volk, *Broken dice and mathematical tales of chance*, The university of Chicago Press (1993)

- [20] G. J. Chaitin, *Sci. Am.*, **232**, 47 (1975).
- [21] J. H. Ahren, U. Dieter and A. Grube, *Computing*, **6**, 121 (1970); D. Knuth, *The art of computer Programming*, Vol. 2, Reading Mass.: Addison Wesley (1969); J. C. Butcher, *Math. Comp.*, **15**, 198 (1961).
- [22] G. Marsaglia, *Proc. Nat. Acad. Sci.*, **61**, 25 (1968); G. Marsaglia, *Applications of Number Theory to Numerical Analysis*, Academic, New York (1972).
- [23] I. Vattulainen, K. Kanakaala, J. Saarinen, and T. Ala-Nissila, *Comp. Phys. Commun.*, **86**, 209 (1995).
- [24] F. Gutbrod, *New trends in pseudo random number generation*, Annual Reviews of Computational Physics VI, World Scientific (to be published); G. Marsaglia, *A Current View of Random Number Generators*, in Proc. Computer Science and Statistics, Sixteenth Symp. on Interface, Elsevier (1985); G. Marsaglia and A. Zaman, *Ann. Appl. Probability*, **1**, 462 (1991).
- [25] A. Bonelli and S. Ruffo, *Modular Transformations, Order-Chaos Transitions and Pseudo-Random Number Generation*, *Int. J. Mod. Phys.*, **C9**(4)(1998).
- [26] D. H. Lehmer, *Ann. Comp. Lab. Harvard Univ.*, **26**, 141 (1951).
- [27] M. Greenberger, *Math. Comp.*, **15**, 383 (1961); M. Greenberger, *Comm. ACM*, **8**, 177 (1965).
- [28] R. P. Chambers, *IEEE Spectrum*, **4**, 48 (1967).
- [29] W. H. Press, S. Teukolsky, W. T. Vetterling and B. P. Flannery, *Numerical Recipes*, Cambridge (1992).
- [30] A. M. Ferrenberg, D. P. Landau and Y. Joanna Wong, *Phys. Rev. Lett.*, **69**, 3382 (1992).
- [31] P. Grassberger, *Phys. Lett.*, **A181**, 43 (1993); L. N. Shchur, J. R. Heringa and H. W. J. Blöte, *Physica A* **241**, 579 (1997); I. Vattulainen, T. Ala-Nissila and K. Kankaala, *Phys. Rev. Lett.*, **73**, 2513 (1994); F. James, *Europhys. Lett.*, **24**, 24 (1993); L. N. Shchur and H. W. J. Blöte, *Phys. Rev. B* **55**, R4905 (1997); F. J. Resende and B. v. Costa, *Phys. Rev. E* **58**, 5183 (1998); R. M. D'Souza, Y. Bar-Yam, and M. Kardar, *Phys. Rev. E* **57**, 5044 (1998).
- [32] J. Eichenauer and J. Lehn, *Statist. Papers*, **27**, 315 (1986).
- [33] J. Eichenauer-Hermann, *Math. Comp.*, **60**, 375 (1993).
- [34] A de Mattis, *Order out of Chaos in Arithmetics*, in G. Maino, L. Fronzoni and M. Pettini (Eds.), *Dynamical Symmetries and Chaotic behaviour in Physical Systems*, World Scientific, Singapore (1989).

- [35] M. Lüscher, *Comput. Phys. Commun.*, **79**, 100 (1994); S.C. Pathak and S. S. Rao, *Cond-mat/931004* (1993).
- [36] J. von Neumann, *Various techniques used in connection with random digits*, U. S. National Bur. Stand. Appl. Math. Ser., No. 12, page 36, (1951); J. H. Ahren and U. Dieter, *Computer methods for sampling from the exponential and normal distribution*, *Comm. Assos. Mach.*, **15**, 873 (1972); G. Marsaglia, *Generating exponential random variables*, *Ann. Math. Stat.*, **32**, 899 (1961); J. H. Ahren, and U. Dieter, *Extensions of Forsythe's method for random sampling from normal distribution*, *Math. Comp.*, **27**, 927 (1973); G. E. Forsythe, *von Neumann's comparison method for random sampling and from the normal and other distributions*, *Math. of Comp.* **26**, 817 (1972).
- [37] J. F. Fernández and J. Rivero, *Comput. Phys.*, **10**, 83 (1996).
- [38] C. S. Wallace, *ACM Trans. Math. Software*, **22**, 119 (1996).
- [39] G. E. P. Box and M. E. Muller, *Ann. Math. Statist.*, **29**, 610 (1958).
- [40] M. E. Muller, *J. Assoc. Comp. Mach.*, **6**, 376 (1959).
- [41] J. F. Fernández and C. Criado, *Algorithm for normal random numbers*, *cond-mat/9901202* 20 Jan 1999; scheduled to appear in *Phys. Rev. E* **60**(3) (1999).
- [42] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).
- [43] O. Peron, *Math. Ann.* **64**, 1 (1907);
- [44] F. R. Gantmacher, *Applications of the theory of matrices*, (English translation: translated by J. R. Brenner), Interscience (1959).
- [45] G. Bhanot, *Rep. Prog. Phys.*, **51**, 429 (1988).
- [46] S. S. Shapiro and M. B. Wilks, *Biometrika*, **52**, 591 (1965).
- [47] F. H. Clark, *Exponential transform as an importance sampling device - a review*, Report ORNL-RSIC-14, Oak Ridge, USA (1966); H. Kahn, *Nucleonics*, **6**, 27 (1950); H. Kahn, *Nucleonics*, **6**, 60 (1950); P. S. Nagarajan, P. Sethulakshmi, and C. P. Raghavendran, *A code for neutron leakage spectra*, Report BARC/I-341, Bhabha Atomic Research Centre, Bombay, India (1975); K. P. N. Murthy, *Tracklength Biasing in Monte Carlo Radiation Transport - A Review*, in *Workshop on Monte Carlo Methods*, Bombay; Jan. 10-20 (1982). See Proceedings Published by BARC, Bombay, Ed. S. R. Dwivedi (1982)p120; K. P. N. Murthy, *Exponential Biasing - An Update*, in L. V. Krishnan, S. M. Lee and Om Pal Singh (Eds.) *Proc. Nat. Conf. Radiation Shielding and Protection (RASP-98)*, Allied (1996); L. B. Levitt, *Nucl. Sci. Eng.* **31**, 500 (1969); K. P. N. Murthy, *Pramana*, **25**, 231 (1985); K. P. N. Murthy and R. Indira, *Nuc. Sci. Eng.*, **92**, 482 (1986); K. P. N. Murthy, *Ann. Nuc. Energ.*, **7**, 389

- (1980); K. P. N. Murthy, *Ann. Nuc. Energ.*, **10**, 375 (1983); R. Indira and K. P. N. Murthy, *Ann. Nuc. Energ.*, **12**, 97 (1985); P. K. Sarkar and M. A. Prasad, *Nucl. Sci. Eng.*, **70**, 243 (1979); P. K. Sarkar and M. A. Prasad, *Nucl. Sci. Eng.*, **74**, 52 (1980); S. R. Dwivedi, *Nucl. Sci. Eng.*, **80**, 172 (1982); S. R. Dwivedi, *Ann. Nucl. Energ.*, **9**, 359 (1982); R. Indira and T. M. John, *Nucl. Sci. Eng.* **94**, 323 (1986); R. Indira, *Ann. Nucl. Energ.*, **15**, 67 (1988); R. Indira, *Ann. Nucl. Energ.*, **15**, 261 (1988).
- [48] J. Spanier, *SIAM J. Appl. Math.*, **18**, 172 (1970); J. Spanier, *A New Multistage procedure for Systematic Variance Reduction in Monte Carlo*, Report ORNL-RSIC-29, Oak Ridge, USA (1971).
- [49] K. P. N. Murthy, *Atomkernenergie*, **34**, 125 (1979); R. Indira and K. P. N. Murthy, *Monte Carlo Simulation of a Benchmark Experiment on Neutron Transport through Thick Sodium*, Report RRC-48, Reactor Research Centre, Kalpakkam, India (1981).
- [50] D. B. MacMillan, *Nuc. Sci. Eng.*, **48**, 219 (1972); R. Indira, *Ann. Nuc. Energ.*, **16**, 571 (1989).
- [51] J. M. Hammersley and D. C. Handscomb, *Monte Carlo Methods*, Chapman and Hall, London (1964); N. Metropolis and S. M. Ulam, *The Monte Carlo Method*, *J. Amer. Statist. Assoc.*, **44**, 333 (1949); M. H. Kalos and P. A. Whitlock, *Monte Carlo Methods, Vol. 1: Basics*, John Wiley, New York (1986); P. Bratley, B. L. Box, and L. E. Schrage, *A Guide to Simulation*, Springer, New York (1987); B. D. Ripley, *Stochastic Simulation*, John Wiley, New York (1987); K. Binder (Ed.), *Applications of Monte Carlo Methods in Statistical Physics*, Springer (1984); Samuel S. M. Wong, *Computational Methods in Physics and Engineering*, see Chapter 7, pp. 383-485, Prentice Hall (1992). I. M. Sobol, *The Monte Carlo Method*, Mir, Moscow (1975); F. James, *Monte Carlo Theory and Practice*, *Rep. Prog. Phys.*, **43**, 1145 (1980); E. D. Cashwell and C. J. Everett, *A Practical Manual of Monte Carlo Method for Random Walk Problems*, Pergamon, New York (1959); J. Spanier and E. M. Gelbard, *Monte Carlo Principles and Neutron Transport Problems*, Addison Wesley (1969); E. C. McGrath and D. C. Irving, *Techniques for Efficient Monte Carlo Simulation*, Vol. I, II, and III, Report. ORNL-RSIC-38, Oak Ridge National Laboratory (1975); S. R. Dwivedi (Ed.), *Proc. Workshop on Monte Carlo Methods*, BARC, Bombay (1982).