MC

Introduction to Monte Carlo Methods

1 Introduction to Monte Carlo Methods

Numerical methods that are known as Monte Carlo methods can be loosely described as statistical simulation methods, where statistical simulation is defined in quite general terms to be any method that utilizes sequences of random numbers to perform the simulation. Monte Carlo methods have been used for centuries, but only in the past several decades has the technique gained the status of a full-fledged numerical method capable of addressing the most complex applications. The name "Monte Carlo" was coined by Metropolis (inspired by Ulam's interest in poker) during the Manhattan Project of World War II, because of the similarity of statistical simulation to games of chance, and because the capital of Monaco was a center for gambling and similar pursuits. Monte Carlo is now used routinely in many diverse fields, from the simulation of complex physical phenomena such as radiation transport in the earth's atmosphere and the simulation of the esoteric subnuclear processes in high energy physics experiments, to the mundane, such as the simulation of a Bingo game or the outcome of Monty Hall's vexing offer to the contestant in "Let's Make a Deal." The analogy of Monte Carlo methods to games of chance is a good one, but the "game" is a physical system, and the outcome of the game is not a pot of money or stack of chips (unless simulated) but rather a solution to some problem. The "winner" is the scientist, who judges the value of his results on their intrinsic worth, rather than the extrinsic worth of his holdings.

Statistical simulation methods may be contrasted to conventional numerical discretization methods, which typically are applied to ordinary or partial differential equations that describe some underlying physical or mathematical system. In many applications of Monte Carlo, the physical process is simulated directly, and there is no need to even write down the differential equations that describe the behavior of the system. The only requirement is that the physical (or mathematical) system be described by probability density functions

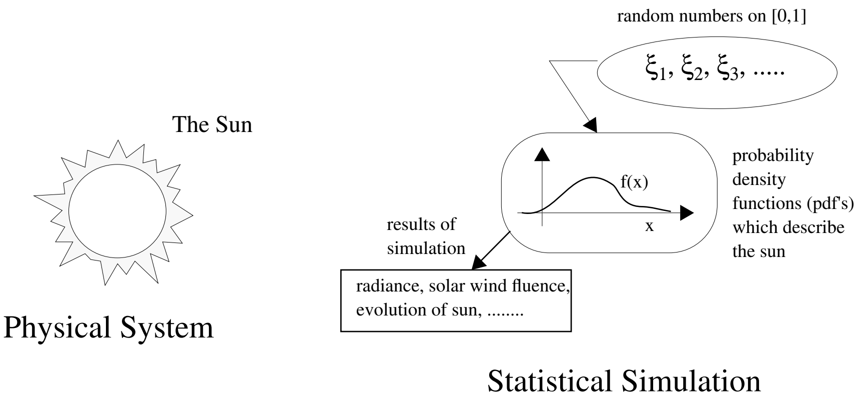

Figure 1: Monte Carlo Simulation of Physical System

(pdf's), which will be discussed in more detail later in this chapter. For now, we will assume that the behavior of a system can be described by pdf's. Once the pdf's are known, the Monte Carlo simulation can proceed by random sampling from the pdf's. Many simulations are then performed (multiple "trials" or "histories") and the desired result is taken as an average over the number of observations (which may be a single observation or perhaps millions of observations). In many practical applications, one can predict the statistical error (the "variance") in this average result, and hence an estimate of the number of Monte Carlo trials that are needed to achieve a given error.

Figure 1 illustrates the idea of Monte Carlo, or statistical, simulation as applied to an arbitrary physical system. Assuming that the evolution of the physical system can be described by probability density functions (pdf's), then the Monte Carlo simulation can proceed by sampling from these pdf's, which necessitates a fast and effective way to generate random numbers uniformly distributed on the interval [0,1]. The outcomes of these random samplings, or trials, must be accumulated or tallied in an appropriate manner to produce the desired result, but the essential characteristic of Monte Carlo is the use of random sampling techniques (and perhaps other algebra to manipulate the outcomes) to arrive at a solution of the physical problem. In contrast, a conventional numerical solution approach would start with the mathematical model of the physical system, discretizing the differential equations and then solving a set of algebraic equations for the unknown state of the system.

It should be kept in mind though that this general description of Monte Carlo methods may not directly apply to some applications. It is natural to think that Monte Carlo methods are used to simulate random, or stochastic, processes, since these can be described by pdf's. However, this coupling is actually too restrictive because many Monte Carlo applications have no apparent stochastic content, such as the evaluation of a definite integral or the inversion of a system of linear equations. However, in these cases and others, one can pose the desired solution in terms of pdf's, and while this transformation may seem artificial, this

Nuclear reactor design
Quantum chromodynamics
Radiation cancer therapy
Traffic flow
Stellar evolution
Econometrics
Dow-Jones forecasting
Oil well exploration
VLSI design

Figure 2: Monte Carlo Applications

step allows the system to be *treated* as a stochastic process for the purpose of simulation and hence Monte Carlo methods can be applied to simulate the system. Therefore, we take a broad view of the definition of Monte Carlo methods and include in the Monte Carlo rubric all methods that involve statistical simulation of some underlying system, whether or not the system represents a real physical process.

To illustrate the diversity of Monte Carlo methods, Figure 2 lists applications that have been addressed with statistical simulation techniques. As can be seen, the range of applications is enormous, from the simulation of galactic formation to quantum chromodynamics to the solution of systems of linear equations.

This wide diversity of methods is the reason that "Monte Carlo is not Monte Carlo is not Monte Carlo."

1.1 Major Components of a Monte Carlo Algorithm

Given our definition of Monte Carlo, let us now describe briefly the major components of a Monte Carlo method. These components comprise the foundation of most Monte Carlo applications, and the following sections will explore them in more detail. An understanding of these major components will provide a sound foundation for the reader to construct his or her own Monte Carlo method, although of course the physics and mathematics of the specific application are well beyond the scope of this chapter. The primary components of a Monte Carlo simulation method include the following:

- *Probability distribution functions (pdf's)* — the physical (or mathematical) system must be described by a set of pdf's.

- *Random number generator* — a source of random numbers uniformly distributed on the unit interval must be available.

- *Sampling rule* — a prescription for sampling from the specified pdf's, assuming the availability of random numbers on the unit interval, must be given.
- *Scoring (or tallying)* — the outcomes must be accumulated into overall tallies or scores for the quantities of interest.
- *Error estimation* — an estimate of the statistical error (variance) as a function of the number of trials and other quantities must be determined.
- *Variance reduction techniques* — methods for reducing the variance in the estimated solution to reduce the computational time for Monte Carlo simulation
- *Parallelization and vectorization* — algorithms to allow Monte Carlo methods to be implemented efficiently on advanced computer architectures.

The remainder of this chapter will treat each of these topics in some detail. Later chapters will describe specific applications of the Monte Carlo method, relying on material in this chapter for the generic aspects common to most, if not all, such methods. But before we actually delve into the subject of Monte Carlo methods, let us look backwards and review some of their history.

1.2 History of Monte Carlo

This section will appear in a future release.

2 An Introduction to Probability and Statistics

An essential component of a Monte Carlo simulation is the modeling of the physical process by one or more probability density functions (pdf's). By describing the process as a *pdf*, which may have its origins in experimental data or in a theoretical model describing the physics of the process, one can sample an "outcome" from the pdf, thus simulating the actual physical process. For example, the simulation of the transport of 2 MeV neutrons in a tank of water will necessitate sampling from a pdf that will yield the distance the neutron travels in the water before suffering a collision with a water molecule. This pdf is the well-known exponential distribution and is an example of a continuous pdf because the outcomes (distances to collision) are described by real numbers. The exponential distribution will be described in more detail later in this chapter. On the other hand, the simulation of roulette will require sampling from a discrete pdf that describes the probability of obtaining one of the 37 (36 outside the U.S.) numbers on a roulette wheel.

2.1 Sample Spaces, Outcomes, and Events

Let us now be more precise about our terminology and define some additional terms that lead to the concept of a pdf. Consistent with standard textbooks (e.g., [Hamming] or [DeGroot]), we will refer to the physical or mathematical process as an *experiment*, and this experiment has a number (possibly infinite) of *outcomes*, to which we will assign probabilities. The *sample space* S of the experiment is the collection of all possible outcomes s. Thus if the experiment is carried out, its outcome is assured to be in the sample space S. We will also describe one realization of the experiment as a *trial*, and by definition a trial will have an outcome s in the sample space S. The experiment may result in the occurrence of a specific *event* E_k. An event may be viewed as a consequence of the outcome (or outcomes) of the experiment. Let us now illustrate these concepts with a simple example.

Example 1 *An illustration of an experiment and the terminology used in the experiment.*

The experiment consists of one roll of a normal die (with faces labeled 1, 2, 3, 4, 5, and 6) and observing the top face of the die. The outcomes s_i are the six faces, and the sample space S consists of these six outcomes, since every realization of the experiment (i.e., each trial) results in one of these faces being the top face. (We will assume that the die will not balance on an edge or corner.) Events can then be defined in terms of the possible outcomes. Possible events that may be defined in terms of the six unique outcomes are:

- E_1: top face is an even number
- E_2: top face is larger than 4
- E_3: top face is equal to 2 (hence the event is one of the outcomes)

Disjoint events are events that cannot happen at the same time. In the above example, events E_2 and E_3 are disjoint, because a single roll of the die (as the experiment was defined) cannot lead to both events occurring. On the other hand, events E_1 and E_3 can occur at the same time, as can events E_1 and E_2.

2.2 Probability

Since this chapter is not intended to be a complete and rigorous treatment of probability, we will avoid the formal theory of probability and instead present a functional description. This should be sufficient preparation to understand the concept of a pdf, which is the goal of this section. The reader who is interested in delving deeper into the subject is encouraged to read the standard textbooks on the subject, a couple of which were named above.

To an event E_k we will assign a probability p_k, which is also denoted $P(E_k)$, or "probability of event E_k". The quantity p_k must satisfy the properties given in Figure 3 to be a legitimate probability.

Figure 3: Properties of a Valid Probability p_k

• $0 \le p_k \le 1$ • If E_k is certain to occur, $p_k = 1$. If E_k is certain not to occur, $p_k = 0$. • If events E_i and E_j are mutually exclusive, then $P(E_i \text{ and } E_j) = 0$ $P(E_i \text{ or } E_j) = p_i + p_j$ • If events E_i, $i = 1, 2, \ldots, N$, are mutually exclusive and exhaustive (one of the N events E_i is assured to occur), then $\sum_{i=1}^{N} p_i = 1$

2.2.1 Joint, Marginal, and Conditional Probabilities

We now consider an experiment that consists of two parts, and each part leads to the occurrence of specified events. Let us define events arising from the first part of the experiment by F_i with probability f_i and events from the second part by G_j with probability g_j. The combination of events F_i and G_j may be called a *composite* event, denoted by the ordered pair $E_{ij} = (F_i, G_j)$. We wish to generalize the definition of probability to apply to the composite event E_{ij}. The *joint* probability p_{ij} is defined to be the probability that the first part of the experiment led to event F_i and the second part of the experiment led to event G_j. Thus, the joint probability p_{ij} is the probability that the composite event E_{ij} occurred (i.e., the probability that both events F_i and G_j occur).

Any joint probability can be factored into the product of a marginal probability and a conditional probability:

$$p_{ij} = p(i)\, p(j|i) \tag{1}$$

where p_{ij} is the joint probability, $p(i)$ is the *marginal* probability (the probability that event F_i occurs *regardless* of event G_j), and $p(j|i)$ is the *conditional* probability (the probability that event G_j occurs *given that event F_i occurs*). Note that the marginal probability for event F_i to occur is simply the probability that the event F_i occurs, or $p(i) = f_i$. Let us now assume that there are J mutually-exclusive events G_j, $j = 1, \ldots, J$ and the following identity is evident:

$$p(i) = \sum_{k=1}^{J} p_{ik} \tag{2}$$

Using Eq. (2), we easily manipulate Eq. (1) to obtain the following expression for the joint probability

$$p_{ij} = p_{ij} \left(\frac{\sum\limits_{k=1}^{J} p_{ik}}{\sum\limits_{k=1}^{J} p_{ik}} \right) = p(i) \left(\frac{p_{ij}}{\sum\limits_{k=1}^{J} p_{ik}} \right) \tag{3}$$

Using Eq. (1), Eq. (3) leads to the following expression for the conditional probability:

$$p(j|i) = \frac{p_{ij}}{\sum\limits_{k=1}^{J} p_{ik}} \tag{4}$$

It is important to note that the joint probability p_{ij}, the marginal probability $p(i)$, and the conditional probability $p(j|i)$ are all legitimate probabilities, hence they satisfy the properties given in the box above. Finally, it is straightforward to generalize these definitions to treat a three-part experiment that has a composite event consisting of three events, or in general an n-part experiment with n events occurring.

If events F_i and G_j are *independent*, then the probability of one occurring does not affect the probability of the other occurring, therefore:

$$p_{ij} = f_i g_j \tag{5}$$

Using Eq. (4), Eq. (5) leads immediately to

$$p(j|i) = g_j \tag{6}$$

for independent events F_i and G_j. This last equation reflects the fact that the probability of event G_j occurring is independent of whether event F_i has occurred, *if events F_i and G_j are independent.*

2.2.2 Random Variables

We now define the concept of a random variable, a key definition in probability and statistics and for statistical simulation in general. We define a random variable as a real number x_i that is assigned to an event E_i. It is random because the event E_i is random, and it is variable because the assignment of the value may vary over the real axis. We will use "r.v." as an abbreviation for "random variable".

Example 2 *Associating a random variable with a roll of a die.*

Assign the number $10n$ to each face n of a die. When face n appears, the r.v. is $10n$.

Random variables are useful because they allow the quantification of random processes, and they facilitate numerical manipulations, such as the definition of mean and standard deviation, to be introduced below. For example, if one were drawing balls of different colors from a bowl, it would be difficult to envision an "average" color, although if numbers were assigned to the different colored balls, then an average could be computed. On the other hand, in many cases of real interest, there is no reasonable way to assign a real number to the outcome of the random process, such as the outcome of the interaction between a 1 eV neutron and a uranium-235 nucleus, which might lead to fission, capture, or scatter. In this case, defining an "average" interaction makes no sense, and assigning a real number to the random process does not assist us in that regard. Nevertheless, in the following discussion, we have tacitly assumed a real number x_i has been assigned to the event E_i that we know occurs with probability p_i. Thus, one can in essence say that the r.v. x_i occurs with probability p_i.

2.2.3 Expectation Value, Variance, Functions of r.v.'s

Now that we have assigned a number to the outcome of an event, we can define an "average" value for the r.v. over the possible events. This average value is called the *expectation value* for the random variable x, and has the following definition:

$$\text{expectation value (or mean)} \equiv E(x) = \bar{x} - \sum_i p_i x_i \tag{7}$$

One can define a unique, real-valued function of a r.v., which will also be a r.v. That is, given a r.v. x, then the real-valued function $g(x)$ is also a r.v. and we can define the expectation value of $g(x)$:

$$E[g(x)] = \bar{g} = \sum_i p_i \, g(x_i) \tag{8}$$

The expectation value of a linear combination of r.v.'s is simply the linear combination of their respective expectation values;

$$E[ag(x) + bh(x)] = aE[g(x)] + bE[h(x)] \tag{9}$$

The expectation value is simply the "first moment" of the r.v., meaning that one is finding the average of the r.v. itself, rather than its square or cube or square root. Thus the mean is the average value of the first moment of the r.v., and one might ask whether or not averages of the higher moments have any significance. In fact, the average of the square of the r.v. does lead to an important quantity, the *variance*, and we will now define the higher moments of a r.v. x as follows:

$$E(x^n) = \overline{x^n} \tag{10}$$

We also define "central" moments that express the variation of a r.v. about its mean, hence "corrected for the mean":

$$n^{th} \text{ central moment } = \overline{(x - \bar{x})^n} \tag{11}$$

The first central moment is zero. The second central moment is the *variance:*

$$\text{variance } \equiv \text{var}(x) \equiv \sigma^2(x) = \overline{(x - \bar{x})^2} = \sum_i p_i (x_i - \bar{x})^2 \tag{12}$$

It is straightforward to show the following important identity:

$$\sigma^2 = \overline{x^2} - \overline{x}^2 \tag{13}$$

We will also find useful the square root of the variance, which is the *standard deviation*,

$$\text{standard deviation } = \sigma(x) = [\text{var}(x)]^{1/2} \tag{14}$$

2.2.4 Variance of Linear Combination

The mean of a linear combination of r.v.'s is the linear combination of the means, as shown in Eq. (9), because the mean is a linear statistic, as is clear from Eq. (7). On the other hand, the variance is clearly not a linear statistic, since the r.v. is squared. However, we will find it necessary to consider the variance of a linear combination of r.v.'s, and it is straightforward to show the following:

$$\sigma^2(ag + bh) = a^2\sigma^2(g) + b^2\sigma^2(h) + 2ab[\overline{gh} - \bar{g}\bar{h}] \tag{15}$$

Let us consider the average value of the product of two r.v.'s:

$$E(xy) = \sum_{i,j} p_{ij}\, x_i\, y_j \tag{16}$$

Now if x and y are independent r.v.'s, then

$$p_{ij} = p_i q_j, \tag{17}$$

where q_j is the probability for the r.v. y_j to occur. But if Eq. (17) is inserted into Eq. (16), we find

$$\begin{aligned} E(xy) &= \sum_{i,j} p_{ij}\, x_i\, y_j = \sum_{i,j} p_i\, q_j\, x_i\, y_j \\ &= \sum_i p_i x_i \sum_j q_j y_j = E(x)E(y) \end{aligned} \tag{18}$$

Thus, if two r.v.'s are independent, the expectation value of their product is the product of their expectation values. Now consider the case of the variance of a linear combination of r.v.'s given in Eq. (15), and note that if the r.v.'s g and h are independent, Eq. (18) when inserted into Eq. (15) yields the following expression, *valid only when $g(x)$ and $h(x)$ are independent r.v.'s:*

$$\sigma^2[ag(x) + bh(x)] = a^2\sigma^2(g) + b^2\sigma^2(h) \tag{19}$$

2.2.5 Covariance and Correlation Coefficient

The cancellation of the last term in Eq. (15) for independent r.v.'s motivates the concept of the *covariance*.

$$\text{covariance } = \text{cov}(x, y) = \overline{xy} - \bar{x}\bar{y} \tag{20}$$

If x and y are independent, then $\text{cov}(x, y) = 0$. However, it is possible to have $\text{cov}(x, y) = 0$ even if x and y are not independent. It should be noted that the covariance can be negative.

A related quantity that arises often in statistical analysis is the *correlation coefficient*, which is a convenient measure of the degree to which two r.v.'s are correlated (or anti-correlated).

$$\text{correlation coefficient } = \rho(x,y) = \text{cov}(x,y)/[\sigma^2(x)\sigma^2(y)]^{1/2} \tag{21}$$

It is easily shown that $-1 \leq \rho(x,y) \leq 1$.

2.3 Continuous Random Variables

So far we have considered only discrete r.v.'s, that is, a specific number x_i is assigned to the event E_i, but what if the events cannot be enumerated by integers, such as the angle of scattering for an electron scattering off a gold nucleus or the time to failure for a computer chip? The above definitions for discrete r.v.'s can be easily generalized to the continuous case.

First of all, if there is a continuous range of values, such as an angle between 0 and 2π, then the probability of getting exactly a specific angle is zero, because there are an infinite number of angles to choose from, and it would be impossible to choose exactly the correct angle. For example, the probability of choosing the angle $\theta = 1.34$ radians must be zero, since there are an infinite number of alternative angles. In fact, there are an infinite number of angles between 1.33 and 1.35 radians or between 1.335 and 1.345 radians, hence the probability of a given angle must be zero. However, we can talk about the probability of a r.v. taking on a value *within* a given interval, e.g., an angle θ between 1.33 and 1.35 radians. To do this, we define a probability density function, or *pdf*.

2.3.1 Probability Density Function (pdf)

The significance of the pdf $f(x)$ is that $f(x)\,dx$ is the probability that the r.v. is in the interval $(x, x+dx)$, written as:

$$\text{prob}(x \leq x' \leq x+dx) \equiv P(x \leq x' \leq x+dx) = f(x)\,dx \tag{22}$$

This is an operational definition of $f(x)$. Since $f(x)\,dx$ is unitless (it is a probability), then $f(x)$ has units of inverse r.v. units, e.g., 1/cm or 1/s or 1/cm^2, depending on the units of x. Figure 4 shows a typical pdf $f(x)$ and illustrates the interpretation of the probability of finding the r.v. in $(x, x+dx)$ with the area under the curve $f(x)$ from x to $x+dx$.

We can also determine the probability of finding the r.v. somewhere in the finite interval $[a,b]$:

$$\text{prob}(a \leq x \leq b) \equiv P(a \leq x \leq b) = \int_a^b f(x')\,dx' \tag{23}$$

which, of course, is the area under the curve $f(x)$ from $x=a$ to $x=b$.

As with the definition of discrete probability distributions, there are some restrictions on the pdf. Since $f(x)$ is a probability density, it must be positive for all values of the r.v. x. Furthermore, the probability of finding the r.v. *somewhere* on the real axis must be unity.

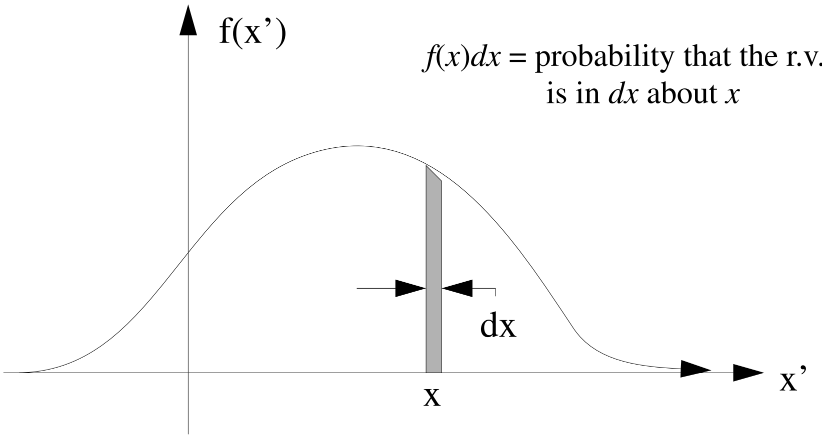

Figure 4: Typical Probability Distribution Function (pdf)

As it turns out, these two conditions are the only necessary conditions for $f(x)$ to be a legitimate pdf, and are summarized below.

$$f(x) \geq 0, \quad -\infty < x < \infty \tag{24}$$

$$\int_{-\infty}^{\infty} f(x')\, dx' = 1 \tag{25}$$

Note that these restrictions are not very stringent, and in fact allow one to apply Monte Carlo methods to solve problems that have no apparent stochasticity or randomness. By posing a particular application in terms of functions that obey these relatively mild conditions, one can treat them as pdf's and perhaps employ the powerful techniques of Monte Carlo simulation to solve the original application. We now define an important quantity, intimately related to the pdf, that is known as the cumulative distribution function, or *cdf.*

2.3.2 Cumulative Distribution Function (cdf)

The cumulative distribution function gives the probability that the r.v. x' is less than or equal to x:

$$\begin{aligned} \mathrm{CDF} &\equiv \mathrm{prob}(x' \leq x) \equiv F(x) \\ &= \int_{-\infty}^{x} f(x')\, dx' \end{aligned} \tag{26}$$

Note that since $f(x) \geq 0$, and the integral of $f(x)$ is normalized to unity, $F(x)$ obeys the following conditions:

- $F(x)$ is monotone increasing
- $F(-\infty) = 0$

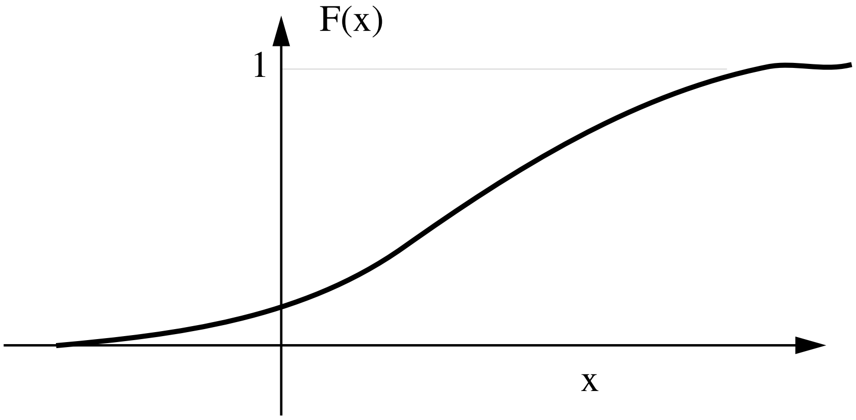

Figure 5: Representative Cumulative Distribution Function (cdf)

- $F(+\infty) = 1$

Figure 5 illustrates a representative cdf. Note the dependence of $F(x)$ as $x \to \pm\infty$. Since $F(x)$ is the indefinite integral of $f(x)$, $f(x) = F'(x)$. The cdf can also be defined for a discrete pdf; however, this will be deferred until we discuss the subject of sampling from a discrete distribution.

2.3.3 Expectation Value and Variance for Continuous pdf's

We can define the expectation value and variance for a continuous pdf, consistent with our earlier definitions for a discrete pdf:

$$E(x) \equiv \mu \equiv \bar{x} = \int_{-\infty}^{\infty} f(x')\, x'\, dx' \tag{27}$$

$$\mathrm{var}(x) \equiv \sigma^2 = \int_{-\infty}^{\infty} f(x')\, (x' - \mu)^2\, dx' \tag{28}$$

Similarly, if we define a real-valued function $g(x)$ of the r.v. x, we readily obtain the following expressions for the mean and variance of g for a continuous pdf:

$$E(g) \equiv \bar{g} = \int_{-\infty}^{\infty} f(x')\, g(x')\, dx' \tag{29}$$

$$\mathrm{var}(g) \equiv \sigma^2(g) = \int_{-\infty}^{\infty} f(x')\, [g(x') - \bar{g}]^2\, dx' \tag{30}$$

It is important to keep in mind that the quantities $\bar{x}$ and $\bar{g}$ are *true* means, *properties* of the pdf $f(x)$ and the function $g(x)$. In many cases of practical interest the true mean is not known, and the purpose of the Monte Carlo simulation will be to *estimate* the true mean.

These estimates will be denoted by a caret or hat, e.g., $\hat{x}$ and $\hat{g}$. Thus the result of a Monte Carlo simulation might be $\hat{g}$, and the hope is that this is a good approximation to the true (but unknown) quantity $\bar{g}$. This notation will be adhered to throughout this chapter on Monte Carlo methods.

2.3.4 Relationship of Discrete and Continuous pdf's

Compare these definitions for a continuous pdf with the previous definitions for the mean and variance for a discrete pdf, given in Eq. (7) and Eq. (12), respectively, and reproduced below for convenience (where the subscript "d" corresponds to "discrete"):

$$E_d(x) = \sum_{i=1}^{N} p_i x_i \tag{31}$$

$$\mathrm{var}_d(x) = \sum_{i=1}^{N} p_i (x_i - \mu)^2 \tag{32}$$

Now take the limit $N \to \infty$ to pass from the discrete to the continuous versions for these quantities:

$$\begin{aligned} \lim_{N\to\infty} E_d(x) &= \lim_{N\to\infty} \sum_{i=1}^{N} x_i p_i = \lim_{N\to\infty} \sum_{i=1}^{N} x_i \frac{p_i}{\Delta x_i} \Delta x_i \\ &= \lim_{N\to\infty} \sum_{i=1}^{N} x_i\, f_i\, \Delta x_i = \int_{-\infty}^{\infty} f(x')\, x'\, dx' \\ &= E(x) \end{aligned} \tag{33}$$

$$\begin{aligned} \lim_{N\to\infty} \mathrm{var}_d(x) &= \lim_{N\to\infty} \sum_{i=1}^{N} (x_i - \mu^2)\, p_i = \lim_{N\to\infty} \sum_{i=1}^{N} (x_i - \mu^2) \frac{p_i}{\Delta x_i} \Delta x_i \\ &= \lim_{N\to\infty} \sum_{i=1}^{N} (x_i - \mu^2)\, f_i\, \Delta x_i = \int_{-\infty}^{\infty} f(x')\,(x' - \mu^2)\, dx' \\ &= \mathrm{var}(x) \end{aligned} \tag{34}$$

2.4 Examples of Continuous pdf's

2.4.1 Exponential Distribution

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0, \ \lambda > 0 \tag{35}$$

This distribution can describe a number of physical phenomena, such as the time t for a radioactive nucleus to decay, or the time x for a component to fail, or the distance z a photon travels in the atmosphere before suffering a collision with a water molecule. The

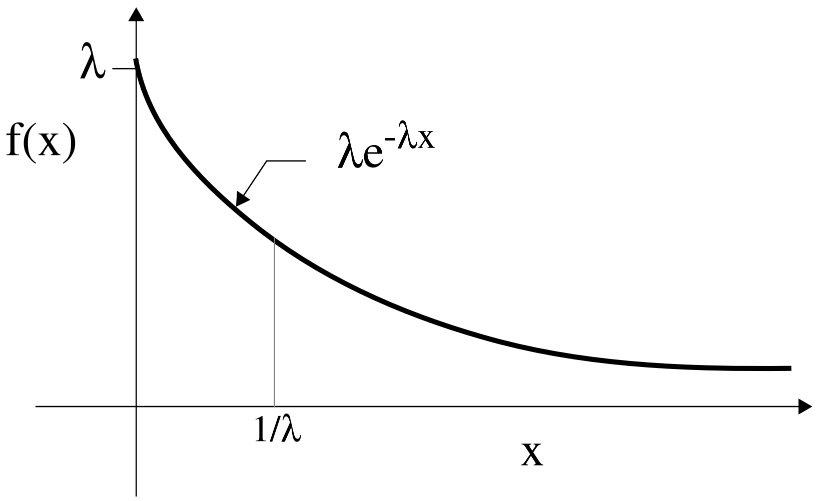

Figure 6: Exponential pdf

exponential distribution is characterized by the single parameter λ, and one can easily show that the mean and variance for the exponential distribution are given by:

$$\mu = \frac{1}{\lambda} \tag{36}$$

$$\sigma^2 = \left(\frac{1}{\lambda}\right)^2 \tag{37}$$

Figure 6 illustrates the exponential distribution. Note that the standard deviation of the exponential distribution is

$$\sigma = \frac{1}{\lambda} \tag{38}$$

Later we will learn that we can associate the standard deviation with a sort of expected deviation from the mean, meaning that for the exponential distribution, one would expect most samples x to fall within $1/\lambda$ of μ, even though the actual range of samples x is infinite. One can see this by computing the probability that a sample from the exponential distribution falls within $\sigma/2$ of the mean μ :

$$\text{prob}\left(\mu - \frac{\sigma}{2} \leq x \leq \mu + \frac{\sigma}{2}\right) = \int_{\frac{1}{2\lambda}}^{\frac{3}{2\lambda}} \lambda e^{-\lambda x}\, dx = 0.83 \tag{39}$$

Hence 83% of the samples from the exponential distribution can be expected to fall within a half of a standard deviation of the mean, although some of the samples will be far from the mean, since $0 \leq x < \infty$.

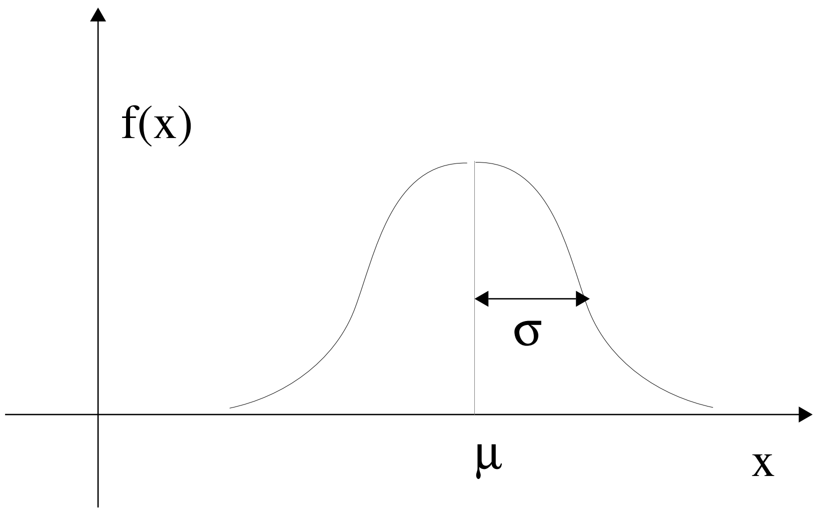

Figure 7: Gaussian (Normal) Probability Distribution Function

2.4.2 Gaussian (Normal) Distribution:

The second example is perhaps the most important pdf in probability and statistics: the *Gaussian*, or *normal*, distribution.

$$f(x) = \frac{1}{(2\pi\sigma)^{1/2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty \tag{40}$$

This is a two-parameter (σ and μ) distribution, and it can be shown that μ is the mean of the distribution and σ^2 is the variance. Figure 7 illustrates the Gaussian pdf.

Let us calculate the probability that a sample from the Gaussian distribution will fall within a single standard deviation σ of the mean μ:

$$P(\mu - \sigma \leq x \leq \mu + \sigma) = .6826 \tag{41}$$

Similarly, the probability that the sample is within two standard deviations (within "2σ") of the mean is

$$P(\mu - 2\sigma \leq x \leq \mu + 2\sigma) = .9544 \tag{42}$$

Hence 68% of the samples will, on average, fall within one σ, and over 95% of the samples will fall within two σ of the mean μ.

The Gaussian distribution will be encountered frequently in this course, not only because it is a fundamental pdf for many physical and mathematical applications, but also because it plays a central role in the estimation of errors with Monte Carlo simulation.

2.4.3 Cauchy Distribution

$$f(x) = \frac{a}{a^2 + x^2}, \quad -\infty < x < \infty \tag{43}$$

This is an interesting pdf, because strictly speaking, its mean does not exist and its variance is infinite. Given our definition of mean,

$$\mu = \int_{-\infty}^{\infty} x \frac{a}{a^2 + x^2} \, dx \tag{44}$$

we find that this integral does not exist because the separate integrals for $x > 0$ and $x < 0$ do not exist. However, if we allow a "principal value" integration, where the limits are taken simultaneously, we see that the integral for $x < 0$ will cancel the integral for $x > 0$ and the mean is zero, consistent with a graphical interpretation of this pdf, as depicted in Figure 8. However, if we try to compute the variance, we find:

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 \frac{a}{a^2 + x^2} \, dx \tag{45}$$

which is an unbounded integral. Thus if we sample from the Cauchy distribution and we attempt to predict the extent to which samples will fall "close" to the mean, we will fail. Note that the Cauchy distribution is a legitimate pdf, because it satisfies the properties of a pdf given in Eq. (24) and Eq. (25), namely,

$$\int_{-\infty}^{\infty} f(x) \, dx = \int_{-\infty}^{\infty} \frac{a}{a^2 + x^2} \, dx = 1 \tag{46}$$

$$f(x) = \frac{a}{a^2 + x^2} \geq 0, \quad \text{all } x \tag{47}$$

but its variance is infinite and its mean necessitates a more general definition of integration.

These have been examples of single random variable, or univariate, pdf's. Let us now consider bivariate pdf's, which generalize readily to multivariate pdf's (the important conceptual step is in going from one to two random variables). Bivariate distributions are needed for a number of important topics in Monte Carlo, including sampling from multidimensional pdf's and the analysis of rejection sampling.

2.4.4 Bivariate Probability Distributions

We now consider two r.v.'s x' and y', where $-\infty < x' < \infty$ and $-\infty < y' < \infty$. We ask what is the probability that the first r.v. x' falls within $[x, x + dx]$ and the second r.v. y' falls within $[y, y + dy]$, which defines the bivariate pdf $f(x, y)$:

$$f(x, y) \, dx \, dy = \text{prob}[(x \leq x' \leq x + dx) \text{ and } (y \leq y' \leq y + dy)] \tag{48}$$

Using this operational definition of $f(x, y)$, let us multiply and divide by the quantity $m(x)$, where we assume $m(x) \neq 0$,

$$m(x) = \int_{-\infty}^{\infty} f(x, y') \, dy' \tag{49}$$

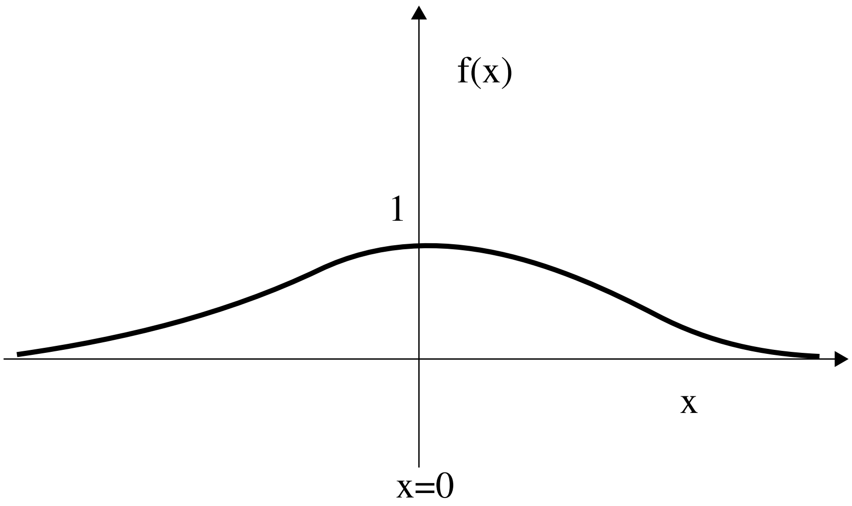

Figure 8: Cauchy Probability Distribution Function

It is readily shown that $m(x)$ satisfies the properties for a legitimate pdf given in Eq. (24) and Eq. (25), and we can interpret $m(x)$ as follows:

$$m(x)\,dx = \text{probability that } x' \text{ is in } dx \text{ about } x, \textit{ irrespective} \text{ of } y'. \tag{50}$$

The quantity $m(x)$ is known as the *marginal* probability distribution function. Now define the quantity $c(y|x)$,

$$c(y|x) = \frac{f(x,y)}{m(x)} \tag{51}$$

As with $m(x)$, it can be shown that $c(y|x)$ is a legitimate pdf and can be interpreted as follows:

$$c(y|x)\,dy = \text{probability that } y' \text{ is in } dy \text{ about } y, \textit{ assuming } x' = x. \tag{52}$$

The quantity $c(y|x)$ is called the *conditional* pdf. The constraint that $m(x) \neq 0$ simply means that the r.v.'s x' and y' are *not mutually exclusive*, meaning there is some probability that both x' and y' will occur together. Note that if x and y are independent r.v.'s, then $m(x)$ and $c(y|x)$ reduce to the univariate pdf's for x and y:

$$m(x) = f(x) \tag{53}$$

$$c(y|x) = g(y) \tag{54}$$

and therefore for independent pdf's we find that the bivariate pdf is simply the product of the two univariate pdf's:

$$f(x,y) = f(x)g(y) \tag{55}$$

2.4.5 Bivariate cdf

The cumulative distribution function for a bivariate pdf is defined in an analogous way to the univariate case:

$$\begin{aligned} F(x,y) &= \text{prob}[x' \le x \text{ and } y' \le y] \\ &= \int_{-\infty}^{x} dx' \int_{-\infty}^{y} dy' \, f(x',y') \end{aligned} \tag{56}$$

and we can express the probability that the random doublet (x', y') falls within a finite region of the x–y plane in terms of this pdf:

$$\text{prob}[a \le x \le b \text{ and } c \le y \le d] = \int_{a}^{b} dx' \int_{c}^{d} dy' \, f(x',y') \tag{57}$$

2.4.6 Sums of Random Variables

Now let us draw N samples $x_1, x_2, x_3, \ldots, x_N$ from the pdf $f(x)$ and define the following linear combination of the samples:

$$G = \sum_{n=1}^{N} \lambda_n g_n(x_n) \tag{58}$$

where the parameters λ_n are real constants and the $g_n(x)$ are real-valued functions. Since the x_n are r.v.'s, and G is a linear combination of functions of the r.v.'s, G is also a r.v. We now examine the properties of G, in particular its expectation value and variance. Referring to our earlier discussion of the mean and variance of a linear combination, expressed as Eq. (9) and Eq. (15), respectively, we find

$$E[G] \equiv \overline{G} = E\left[\sum_{n=1}^{N} \lambda_n g_n(x_n)\right] = \sum_{n=1}^{N} \lambda_n E[g_n] = \sum_{n=1}^{N} \lambda_n \overline{g_n} \tag{59}$$

$$\text{var}[G] = \text{var}\left[\sum_{n=1}^{N} \lambda_n g_n(x_n)\right] = \sum_{n=1}^{N} \lambda_n^2 \, \text{var}[g_n] \tag{60}$$

Now consider the special case where $g_n(x) = g(x)$ and $\lambda_n = 1/N$:

$$G = \frac{1}{N} \sum_{n=1}^{N} g(x_n) \tag{61}$$

Note that G is simply the *average* value of the N sampled r.v.'s. Now consider the expectation value for G, using Eq. (61):

$$\overline{G} = \frac{1}{N} E\left[\sum_{n=1}^{N} g(x_n)\right] = \frac{1}{N} \sum_{n=1}^{N} \bar{g} = \bar{g} \tag{62}$$

In other words, the *expectation value* for the average (*not* the average itself!) of N observations of the r.v. $g(x)$ is simply the expectation value for $g(x)$. This statement is not as trivial as it may seem, because we may not know $E[g]$ in general, because $E[g]$ is a property of $g(x)$ and the pdf $f(x)$. However, Eq. (62) assures us that an average of N observations of $g(x)$ will be a reasonable estimate of $E[g]$. Later, we will introduce the concept of an unbiased estimator, and suffice to say for now, that Eq. (62) proves that the simple average is an unbiased estimator for the mean. Now let us consider the variance in G, in particular its dependence on the sample size.

Considering again the case where $g_n(x) = g(x)$ and $\lambda_n = 1/N$, and using Eq. (60), the variance in the linear combination G is given by:

$$\mathrm{var}(G) = \left(\frac{1}{N}\right)^2 \sum_{n=1}^{N} \mathrm{var}[g(x_n)] = \left(\frac{1}{N}\right)^2 \sum_{n=1}^{N} \mathrm{var}(g) = \frac{1}{N}\mathrm{var}(g) \tag{63}$$

Hence the variance in the average value of N samples of $g(x)$ is a factor of N smaller than the variance in the original r.v. $g(x)$. Note that we have yet to say anything about how to estimate $\mathrm{var}(G)$, only that its value decreases as $1/N$.

This point deserves further elaboration. The quantities $E(g)$ and $\mathrm{var}(G)$ are *properties* of the pdf $f(x)$ and the real function $g(x)$. As mentioned earlier, they are known as the *true mean* and *true variance*, respectively, because they are known *a priori*, given the pdf $f(x)$ and the function $g(x)$. Then if we consider a simple average of N samples of $g(x)$, denoted G, Eq. (62) tells us that the true mean for G is equal to the true mean for $g(x)$. On the other hand, Eq. (63) tells us that the *true variance for G is $1/N$ smaller than the true variance for $g(x)$*, an important consequence for estimating errors.

Later we will show how to estimate $\mathrm{var}(G)$, an important task since in general we don't know the true mean and variance, and these terms will have to be estimated. Let us now apply this discussion to an important application of Monte Carlo methods, the evaluation of definite integrals.

2.4.7 Monte Carlo Integration (Our First Application of Monte Carlo)

We would like to evaluate the following definite integral,

$$I = \int_a^b g(x)\,dx \tag{64}$$

where we assume that $g(x)$ is real-valued on $(-\infty, \infty)$. Figure 9 depicts a typical integral to be evaluated.

The idea is to manipulate the definite integral into a form that can be solved by Monte Carlo. To do this, we define the following function on $[a, b]$,

$$f(x) = \begin{cases} 1/(b-a), & a \le x \le b \\ 0, & \text{otherwise} \end{cases} \tag{65}$$

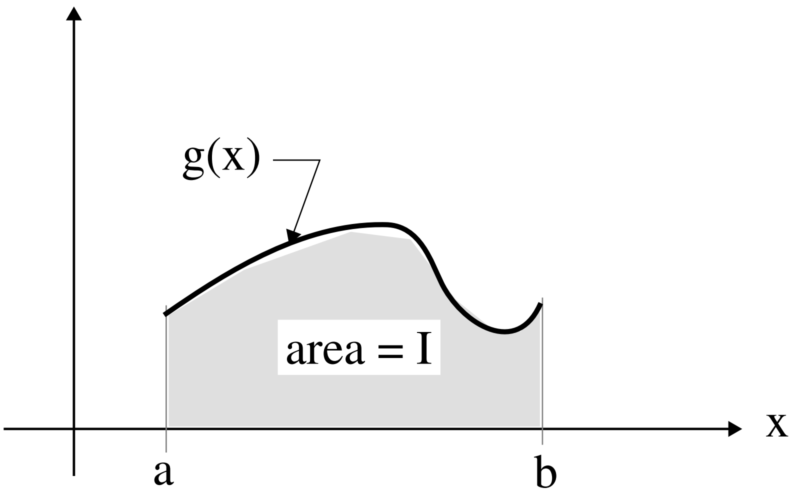

Figure 9: Monte Carlo Integration

and insert into Eq. (64) to obtain the following expression for the integral I:

$$I = \frac{1}{b-a} \int_a^b g(x)\, f(x)\, dx \tag{66}$$

Note that $f(x)$ can be viewed as a uniform pdf on the interval $[a, b]$, as depicted in Figure 9. Given that $f(x)$ is a pdf, we observe that the integral on the right hand side of Eq. (66) is simply the expectation value for $g(x)$:

$$I = \frac{1}{b-a} \int_a^b g(x)\, f(x)\, dx = \frac{1}{b-a}\bar{g} \tag{67}$$

We now draw samples x_n from the pdf $f(x)$, and for each x_n we will evaluate $g(x_n)$ and form the average G,

$$G = \frac{1}{N} \sum_{n=1}^{N} g(x_n) \tag{68}$$

But Eq. (62) states the expectation value for the average of N samples is the expectation value for $g(x)$, $\overline{G} = \bar{g}$, hence

$$I = \frac{1}{b-a}\overline{G} \approx \frac{1}{b-a} G = \frac{1}{b-a}\left(\frac{1}{N}\sum_{i=1}^{N} g(x_n)\right) \tag{69}$$

Thus we can *estimate* the true value of the integral I on $[a, b]$ by taking the average of N observations of the integrand, with the r.v. x sampled uniformly over the interval $[a, b]$. For now, this implies that the interval $[a, b]$ is finite, since an infinite interval cannot have a uniform pdf. We will see later that infinite ranges of integration can be accommodated with more sophisticated techniques.

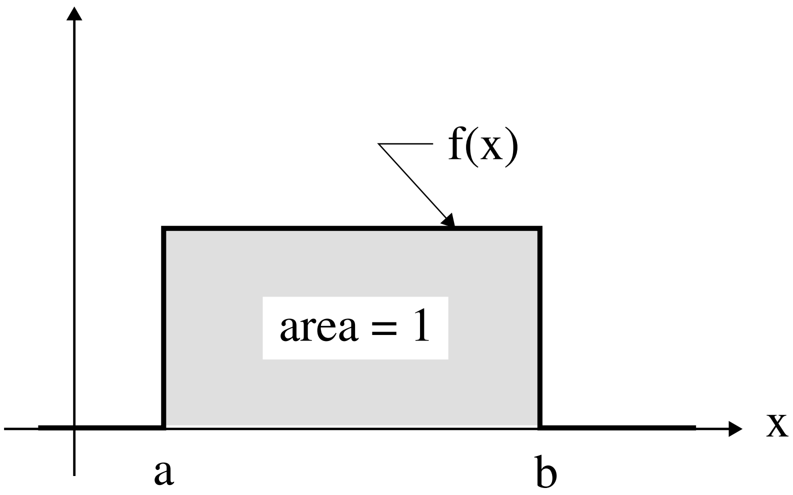

Figure 10: Uniform pdf on [a,b]

Recall that Eq. (63) related the true variance in the average G to the true variance in g,

$$\mathrm{var}(G) = \frac{1}{N}\mathrm{var}(g) \tag{70}$$

Although we do not know $\mathrm{var}(G)$, since it is a property of the pdf $f(x)$ and the real function $g(x)$, it is a constant. Furthermore, if we associate the error in our estimate of the integral I with the standard deviation, then we might expect the error in the estimate of I to decrease by the factor $N^{-1/2}$. This will be shown more rigorously later when we consider the Central Limit Theorem, but now we are arguing on the basis of the functional form of $\mathrm{var}(G)$ and a hazy correspondence of standard deviation with "error". What we are missing is a way to estimate $\mathrm{var}(G)$, as we were able to estimate $E(g)$ with G.

3 Sampling from Probability Distribution Functions

As described earlier, a Monte Carlo simulation consists of some physical or mathematical system that can be described in terms of probability distribution functions, or pdf's. These pdf's, supplemented perhaps by additional computations, describe the evolution of the overall system, whether in space, or energy, or time, or even some higher dimensional phase space. The goal of the Monte Carlo method is to simulate the physical system by random sampling from these pdf's and by performing the necessary supplementary computations needed to describe the system evolution. In essence, the physics and mathematics are replaced by random sampling of possible states from pdf's that describe the system. We now turn our attention to how one actually obtains random samples from arbitrary pdf's.

This chapter will consider sampling from both continuous and discrete pdf's. Table 1 summarizes the important properties of both types of pdf's.

Table 1: Important properties of continuous and discrete pdf's.

Property	**Continuous:** $f(x)$	**Discrete:** $\{p_i\}$
Positivity	$f(x) \geq 0$, all x	$p_i > 0$, all i
Normalization	$\int_{-\infty}^{\infty} f(x')\,dx' = 1$	$\sum_{j=1}^{N} p_j = 1$
Interpretation	$f(x)\,dx$ $\mathrm{prob}(x \leq x' \leq x + dx)$	$p_i = \mathrm{prob}(i) =$ $\mathrm{prob}(x_j = x_i)$
Mean	$\bar{x} = \int_{-\infty}^{\infty} x f(x)\,dx$	$\bar{x} = \sum_{j=1}^{N} x_j p_j$
Variance	$\sigma^2 = \int_{-\infty}^{\infty} (x - \bar{x})^2 f(x)\,dx$	$\sigma^2 = \sum_{j=1}^{N} (x_j - \bar{x})^2 p_j$

We will now discuss how to obtain a random sample x from either a continuous pdf $f(x)$ or a discrete pdf $\{p_i\}$.

3.1 Equivalent Continuous pdf's

It will be convenient to express a discrete pdf as a continuous pdf using "delta functions". This will make the ensuing discussion easier to follow and simplifies many of the manipulations for discrete pdf's. Given a discrete pdf $\{p_i\}$, let us associate event i with the discrete r.v. x_i, and then define an equivalent "continuous" pdf as follows:

$$f(x) = \sum_{i=1}^{N} p_i\,\delta(x - x_i) \tag{71}$$

Here $\delta(x - x_i)$ is the "delta" function and it satisfies the following properties:

$$\int_{-\infty}^{\infty} \delta(x - x_i)\,dx = 1 \tag{72}$$

$$\int_{-\infty}^{\infty} f(x)\,\delta(x - x_i)\,dx = f(x_i) \tag{73}$$

Using these properties, it is straightforward to show that the mean and variance of the equivalent continuous pdf, as defined in Eq. (71), are identical to the mean and variance of the original discrete pdf. Begin with the definition of the mean of the equivalent continuous pdf:

$$\bar{x} = \int_{-\infty}^{\infty} x\,f(x)\,dx = \int_{-\infty}^{\infty} x \left[\sum_{i=1}^{N} p_i\,\delta(x - x_i)\right] dx \tag{74}$$

Now take the summation outside the integral and use Eq. (73),

$$\bar{x} = \sum_{i=1}^{N} \int_{-\infty}^{\infty} x p_i \, \delta(x - x_i) \, dx = \sum_{i=1}^{N} x_i \, p_i \tag{75}$$

which is the true mean for the discrete pdf. It is left as an exercise to show that this also holds for the variance, and in general for any moment of the distribution.

Much of the material that follows holds for both discrete and continuous pdf's, and this equivalence will be useful in this discussion.

3.2 Transformation of pdf's

In order to have a complete discussion of sampling, we need to explain transformation rules for pdf's. That is, given a pdf $f(x)$, one defines a new variable $y = y(x)$, and the goal is to find the pdf $g(y)$ that describes the probability that the r.v. y occurs. For example, given the pdf $f(E)$ for the energy of the scattered neutron in an elastic scattering reaction from a nucleus of mass A, what is the pdf $g(v)$ for the speed v, where $E = \frac{1}{2}mv^2$?

First of all, we need to restrict the transformation $y = y(x)$ to be a unique transformation, because there must be a 1-to-1 relationship between x and y in order to be able to state that a given value of x corresponds unambiguously to a value of y. Given that $y(x)$ is 1-to-1, then it must either be monotone increasing or monotone decreasing, since any other behavior would result in a multiple-valued function $y(x)$.

Let us first assume that the transformation $y(x)$ is monotone increasing, which results in $dy/dx > 0$ for all x. Physically, the mathematical transformation must conserve probability, i.e., the probability of the r.v. x' occurring in dx about x must be the same as the probability of the r.v. y' occurring in dy about y, since if x occurs, the 1-to-1 relationship between x and y necessitates that y appears. But by definition of the pdf's $f(x)$ and $g(y)$,

$$\begin{aligned} f(x)\, dx &= \mathrm{prob}(x \le x' \le x + dx) \\ g(y)\, dy &= \mathrm{prob}(y \le y' \le y + dy) \end{aligned}$$

The physical transformation implies that these probabilities must be equal. Figure 11 illustrates this for an example transformation $y = y(x)$.

Equality of these differential probabilities yields

$$f(x)\, dx = g(y)\, dy \tag{76}$$

and one can then solve for $g(y)$:

$$g(y) = f(x)/[dy/dx] \tag{77}$$

This holds for the monotone increasing function $y(x)$. It is easy to show that for a monotone decreasing function $y(x)$, where $dy/dx < 0$ for all x, the fact that $g(y)$ must be positive (by definition of probability) leads to the following expression for $g(y)$:

$$g(y) = f(x)/[-dy/dx] \tag{78}$$

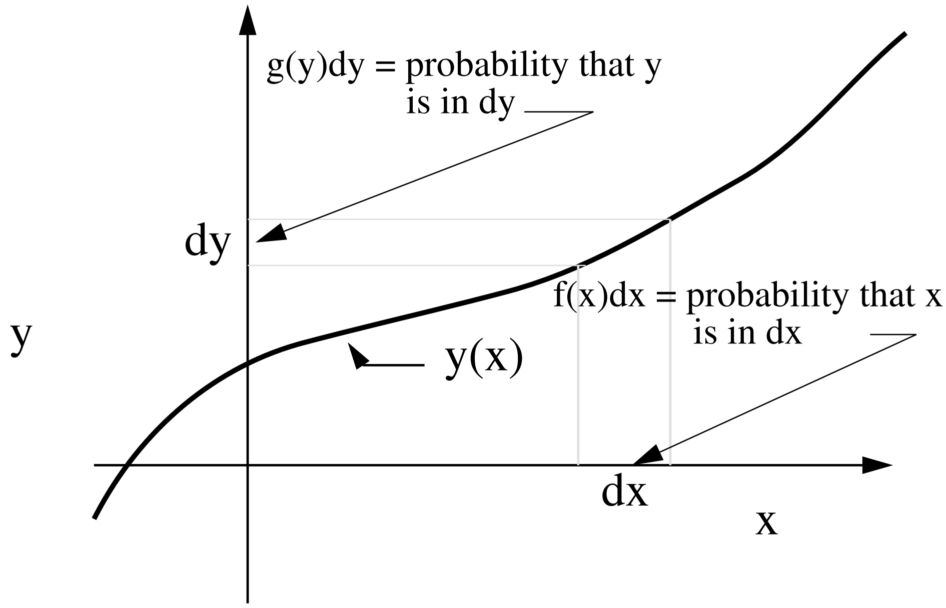

Figure 11: Transformation of pdf's

Combining the two cases leads to the following simple rule for transforming pdf's:

$$g(y) = f(x)/|dy/dx| \tag{79}$$

For multidimensional pdf's, the derivative $|dy/dx|$ is replaced by the Jacobian of the transformation, which will be described later when we discuss sampling from the Gaussian pdf.

Example 3 *An illustration of neutron elastic scattering.*

Consider the elastic scattering of neutrons of energy E_0 from a nucleus of mass A (measured in neutron masses) at rest. Define $f(E)\,dE$ as the probability that the final energy of the scattered neutron is in the energy interval dE about E, given that its initial energy was E_0. The pdf $f(E)$ is given by:

$$f(E) = \begin{cases} \dfrac{1}{(1-\alpha)E_0}, & \alpha E_0 \le E \le E_0 \\ 0, & \text{otherwise} \end{cases} \tag{80}$$

We now ask: what is the probability $g(v)\,dv$ that the neutron scatters in the speed interval dv about v, where $E = \frac{1}{2}mv^2$? Using Eq. (79), one readily finds the following expression for the pdf $g(v)$:

$$g(v) = \begin{cases} \dfrac{2v^2}{(1-\alpha)v_0}, & \sqrt{\alpha}v_0 \le v \le v_0 \\ 0, & \text{otherwise} \end{cases} \tag{81}$$

It is easy to show that $g(v)$ is a properly normalized pdf in accordance with Eq. (24).

Example 4 *An illustration of the cumulative distributtion function, or cdf.*

Perhaps the most important transformation occurs when $y(x)$ is the cumulative distribution function, or cdf:

$$y(x) = F(x) \equiv \int_{-\infty}^{\infty} f(x')\,dx' \tag{82}$$

In this case, we have $dy/dx = f(x)$, and one finds the important result that the pdf for the transformation is given by:

$$g(y) = 1, \quad 0 \le y \le 1 \tag{83}$$

In other words, the cdf is always uniformly distributed on [0,1], independently of the pdf $f(x)$! *Any value for the cdf is equally likely on the interval [0,1].* As will be seen next, this result has important ramifications for sampling from an arbitrary pdf.

3.2.1 Sampling via Inversion of the cdf

Since the r.v. x and the cdf $F(x)$ are 1-to-1, one can sample x by first sampling $y = F(x)$ and then solving for x by inverting $F(x)$, or $x = F^{-1}(y)$. But Eq. (83) tells us that the cdf is uniformly distributed on [0,1], which is denoted $U[0,1]$. Therefore, we simply use a random number generator (RNG) that generates $U[0,1]$ numbers, to generate a sample ξ from the cdf $F(x)$. Then the value of x is determined by inversion, $x = F^{-1}(\xi)$. This is depicted graphically in Figure 12. The inversion is not always possible, but in many important cases the inverse is readily obtained.

This simple yet elegant sampling rule was first suggested by von Neumann in a letter to Ulam in 1947 [Los Alamos Science, p. 135, June 1987]. It is sometimes called the "Golden Rule for Sampling". Since so much use will be made of this result throughout this chapter, we summarize below the steps for sampling by inversion of the cdf:

Step 1. Sample a random number ξ from $U[0,1]$

Step 2. Equate ξ with the cdf: $F(x) = \xi$

Step 3. Invert the cdf and solve for x: $x = F^{-1}(\xi)$

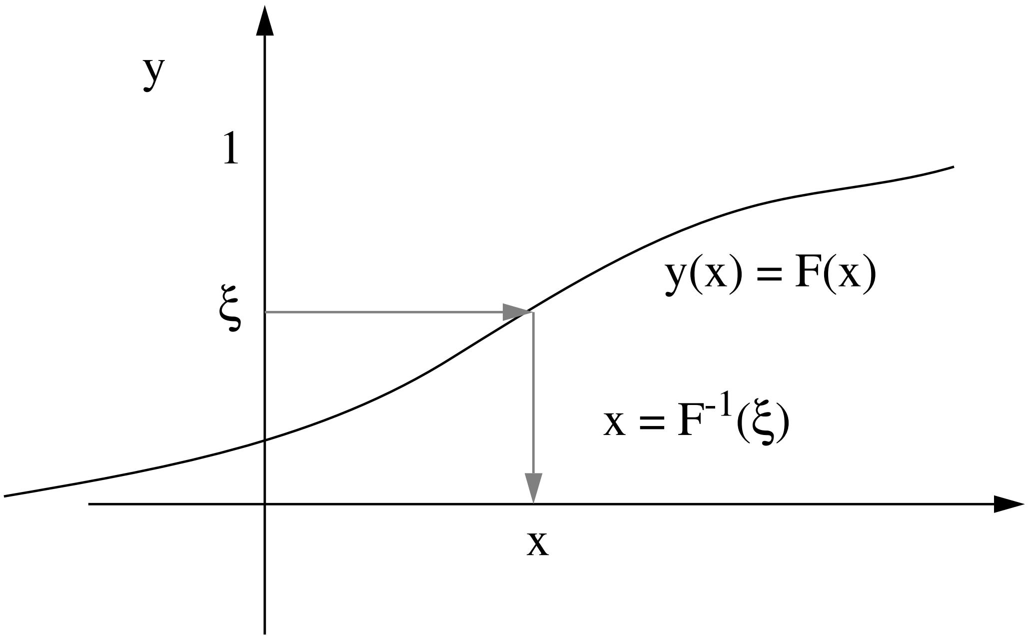

Figure 12: Sampling Using the Inverse of the cdf

Example 5 *An illustration of a uniform distribution.*

Let the r.v. x be uniformly distributed between a and b. In this case, the cdf $F(x)$ is easily found to be

$$F(x) = (x - a)/(b - a) \tag{84}$$

Now sample a random number ξ from $U[0,1]$, set it equal to $F(x)$, and solve for x:

$$x = a + (b - a)\xi \tag{85}$$

which yields a sampled point x that is uniformly distributed on the interval $[a, b]$.

Example 6 *An illustration of an exponential distribution.*

Consider the penetration of neutrons in a shield, where the pdf for the distance x to collision is described by the exponential distribution,

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0, \ \ \lambda > 0 \tag{86}$$

A distance x to collision is then determined by first sampling a value for the cdf from $U[0,1]$ and solving for x. One does not need to subtract the random number from unity, because ξ and $1-\xi$ are both uniformly distributed on [0,1], and statistically the results will be identical.

3.3 More on Sampling

The following topics will be discussed in a future release:

- Example—Gaussian distribution (Box–Muller)
- Discrete probability distribution functions
- Composition techniques
- Sum of 2 r.v.'s
- (r.v.)^n
- Sum of several pdf's
- Rejection sampling

4 Estimation of Mean and Variance

This section will appear in a future release. It will discuss the following topics:

- True mean, true variance
- Unbiased estimators
- Sample mean and variance
- Canonical tallies - implementation
- Example - Monte Carlo integration

5 Error Estimates

This section will appear in a future release. It will discuss the following topics:

- Law of large numbers
- Chebychev inequality

- Central Limit Theorem (CLT)
- Application of CLT to Monte Carlo experiments - scoring
- Standard deviation, relative standard error
- "One sigma" and "two sigma" error estimates

6 Variance Reduction

This section will appear in a future release. It will discuss the following topics:

- Zero variance (or single history) Monte Carlo
- Importance sampling
- Application - Monte Carlo integration (variational derivation)

7 Case Study: Monte Carlo Particle Transport

The analysis of particle transport problems motivated the development of the Monte Carlo method, as we noted in our earlier chapter on the history of the Monte Carlo method. While Monte Carlo methods are used in virtually all branches of science and engineering, it is still the case that the most prevalent application of Monte Carlo is for the solution of complex problems that are encountered in particle transport applications. For example, the analysis of electron transport for electron beam cancer therapy, or the analysis of photon transport in a cloudy atmosphere, or the attenuation of neutrons in a biological shield. These problems are typically characterized by the following features:

- Complex 3-D and non-Cartesian geometry (e.g., nuclear reactor plant; human body)
- Complex material configurations (e.g., semiconductor chips)
- Complicated physical phenomena due to interaction of radiation (neutrons, photons,...) with medium
- Some known source of radiation incident on (or emitted within) the geometry
- Required output is the amount of radiation, its deposition, or its effect in arbitrary regions
- It is desirable to estimate the uncertainties in the simulation
- The computational effort to carry out the simulation should be reasonable

These rather general characteristics of a typical particle transport Monte Carlo code are represented in one or more of the following modules that appear in most production Monte Carlo codes used for radiation transport analysis:

- Source module
- Boundary crossing and geometry module
- Distance to collision module
- Collision analysis module
- Scoring (tallying) module
- Estimate of variance, confidence intervals
- Variance reduction techniques

Let us now consider each of these modules in more detail.

7.1 Source Module

A specified source of radiation may either be given as (1) a *specified incident distribution* in space, energy, angle, and time or (2) as a *known source of radiation* that is emitting a specified amount of radiation as a function of time, space, energy, and angle. Although it is possible to show that these are mathematically equivalent [CdHP], the actual implementation of these "sources" into a Monte Carlo code will depend on which type of source is being examined.

In order to specify an incident flux of particles on a surface, one needs to determine the following quantities:

- position $\mathbf{r}_s = (x_s, y_s, z_s)$,
- the energy E,
- the angle $\Omega = (\Omega_x, \Omega_y, \Omega_z)$ that the incident particle is traveling, and
- the time t that the particle is incident on the surface.

For simplicity, let us assume that the incident radiation is monoenergetic at energy E_0, and it is monodirectional, traveling down the z-axis. Let us assume that the surface that is being irradiated is in the x–y plane, ranging over $x : [0, a]$ and $y : [0, b]$.

In this case, one typically knows the number of particles incident on the boundary per unit area of boundary, as a function of time, position on the boundary, energy, and angle. For example, consider a beam of monoenergetic particles incident normally and uniformly on the negative "y" surface of a "brick" of edges a, b, and c, corresponding to the three coordinate axes x, y, and z.

In this case, there are I_0 particles incident per unit area per unit time on the slab, and they are all travelling perpendicular to the surface of the slab. To start a particle in a Monte Carlo simulation, the source module would sample a position on the incoming surface of the brick.

Nuclear engineers working in reactor physics and radiation shielding areas generally employ the concept of neutron "flux" to describe the amount of radiation, while other disciplines employ a "density" or "intensity" to describe what is in essence a very similar quantity. However, since the Monte Carlo simulation is a direct analog of the physical application, how these terms relate to a real application will be apparent after a few examples.

7.2 Additional Modules

The following modules will be discussed in a future release:

- Boundary crossing and geometry module
- Distance to collision module
- Collision analysis module
- Scoring (tallying) module
- Estimate of variance, confidence intervals
- Variance reduction techniques (for particle transport applications)