Monte Carlo Methods in Engineering

Mikael Amelin

PREFACE

This compendium describes how Monte Carlo methods can be applied to simulate technical systems. The description covers background on probability theory and random number generation as well as the thoery and practice of efficient Monte Carlo simulations. The core of the compendium is based on lectures that have been given at KTH for several years; however, the presentation here also includes more explanatory texts and exercises with solutions.

I would like to give a warm thank you to colleagues and students who have helped improve the contents of this compendium by asking questions, pointing out errors and suggesting additional topics.

Stockholm
September 2013   *Mikael Amelin*

CONTENTS

INTRODUCTION

Monte Carlo methods refers to a class of methods to solve mathematical problems using random samples. A straightforward example is the computation of the expectation value of a random variable; instead of computing the expectation value according to the definition (which may involve solving complex integrals) we observe the behaviour of the random variable, i.e., we collect samples, and estimate its expactation value based on these samples. However, Monte Carlo methods may also be used for solving deterministic problems. This might seem odd at a first glance, but the idea is simply to find a random variable, the statistic properties of which is depending on the solution to the deterministic problem. An example of this would be an opinion poll. Assume that there is going to be an election. If we ignore that people may change their preferences over time and consider just one specific point of time then the share of voters who are planning to vote for a certain candidate is deterministic. However, in order to compute the true value of this share, we would have to ask all voters which candidate they favour. An alternative would be to ask a limited number of randomly chosen voters and use these samples to estimate the share of votes the candidate will obtain.

This compendium will describe how Monte Carlo methods can be used for simulation of various technical systems. The compendium includes many mathematical definitions and formulae, but it should be emphasised that this is not a mathematical textbook. The focus of the presentation will be how Monte Carlo methods can be applied to solve engineering problems; hence, the mathematics should be seen as a tool and not a topic in itself. This means that for example mathematical proofs will only be provided in order to improve the understanding of the described methods, and some mathematical details might be ignored.

1.1   Brief History

The phrase "Monte Carlo methods" was coined in the beginning of the 20th century, and refers to the famous casino in Monaco[1]—a place where random samples indeed play an important role. However, the origin of Monte Carlo methods is older than the casino.

  To be added: History of probability theory...
  To be added: Bernouille, Poisson and the law of large numbers...
  To be added: Buffon's needle
  To be added: Modern development...

1.2   Problem Definition

This entire compendium is focusing on methods for simulation of systems on one specific format. (This might seem as a large limitation, but the reader will soon see that a wide range of systems fit into this format.) An overview of this format is given in figur 1.1. The studied systems are modelled by a set of

---

1. It has been said that if Monte Carlo methods had been first explored today, they would have been referred to as "Las Vegas methods".

To be added: Figure...

Figure 1.1   Overview of the simulation problem studied in this compendium.

random input variables, which we at this point simply collect into a vector, Y. The probability distribution of these inputs must be known. We also have and a set of output variables, which we also collect into a vector, X. As these outputs are depending on the random inputs, they must be random variables as well; however, the probability distribution of the outputs is not known—in fact, the objective of simulating the system is to determine the behaviour of the outputs. Finally, we have a mathematical model of the system, which determines how the values of the outputs are calculated given the values of the input variables. We denote the mathematical model as a function g, such that $X = g(Y)$. The function g does not have to be expressed explicitly, can for example be derived from the solution to one—or even several—optimisation problems, as illustrated in the second of the following two examples:

>   **Example 1.1.** To be added: Example of a model expressed explicitly...

>   **Example 1.2.** To be added: Example of a model derived from the solution to an optimisation problem (multi-area economic dispatch problem)...

It is important to notice that the model g is deterministic! Hence, if we have two sets of input values, y_1 and y_2, producing two sets of output values, $x_1 = g(y_1)$ and $x_2 = g(y_2)$ then if $y_1 = y_2$ we will get that $x_1 = x_2$. If this property is not fulfilled, the model is missing inputs and should be reformulated, as in the following example:

>   **Example 1.3.** To be added: Example of a model which is missing an input value (power grid, where the reliability is depending on the sucess of reclosing breakers after a failure) ...

1.3   Notation

Before we start investigating the application of Monte Carlo methods to solve the simulation problem described above, it might be useful to introduce a general notation, which will be used throughout this compendium. Once the reader is familiar with this notation, it will be more straightforward to interpret the mathematical expression appearing in the following chapters.

-   **Random variables.** All random variables are denoted by upper-case Latin letters; usually just one single letter, for example Y or X, but sometimes several letters, such as *TOC* in example 1.2.

-   **Samples.** An observation of a random variable (i.e., a sample) is denoted by the lower-case of the symbol used for the random variable itself, for example y or x. In most cases, we also need an index in order to separate different samples from each other, i.e., y_i or x_i.

-   **Populations.** A population is denoted by the same upper-case Latin letter as the corresponding random variable, but using a script font, for example $\mathcal{Y}$ or $\mathcal{X}$. The value of the i:th unit in a population is denoted by the lower-case symbol used for population itself, indexed by i, for example y_i or x_i.

-   **Probability distributions.** Probability distributions are denoted by Latin f in upper or lower case (depending on interpretation[2]) and an index showing to which random variable the distribution is associated, for example f_Y or F_X. The idea of the index is to tell different probability distributions apart from each other.

-   **Statistical properties.** Key statistical properties of a probability distribution are denoted by lower-case Greek letters and an index showing to which random variable the statistical property is associated, for example μ_Y or σ_X. The idea of the index is to tell different probability distributions apart from each other.

---

2. Cf. section 2.1.

- **Estimates.** Estimates of key statistical properties for a probability distribution are denoted by upper or lower case (depending on interpretation[3]) Latin counterpart of the symbol used for the statistical property itself and an index showing to which random variable the statistical property is associated, for example M_X or s_X. The idea of the index is to tell different probability distributions apart from each other.

---

3. Cf. section 4.1.

RANDOM VARIABLES

As the idea of Monte Carlo simulation is to estimate values using random observations, it is natural that a basic understanding of probability theory is necessary. This chapter summarises the probability theory that will be used in the remainder of the compendium. The focus of the presentation will be on random variables.

2.1 Probability Distributions

Intuitively, we may understand a random variable exactly as the name suggests, i.e., a variable, the value of which is varying according to some random pattern. This pattern, which characterises the behaviour of the random variable is referred to as its probability distribution. There is an infinite number of possible probability distributions; however, some common classes of distributions have been identified and named. A brief overview can be found in appendix A.

2.1.1 Populations

The formal mathematical definition of a random variable is however slightly more complex, and will therefore not be discussed here.

For a discussion of sampling and Monte Carlo simulation, a useful interpretation of random variables is to consider a random variable to be associated with a certain *population*, which we define as follows:

> **Definition 2.1.** The random variable X corresponds to a population, X, which is a set with N members (which are referred to as "units"). Each unit has a value, x_i, which may be multi-dimensional. The values of the units in X do not have to be unique, but they should include all possible outcomes of the random variable X and the relative occurrence of a certain value should be proportional to the probability of the corresponding outcome.

> **Example 2.1.** State the population corresponding to the following random variables:

> **a)** D, which represents the result of throwing a normal six-sided dice.

> **b)** To be added...

> **Solution:**

> **a)** $\mathcal{D} = \{1, 2, 3, 4, 5, 6\}$

> **b)** To be added...

Based on this definition we can distinguish between some main categories of populations (and consequently between different categories of random variables). First we can differentiate between variables where the outcome can only belong to specific, discrete values or if the outcome can be found in continuous intervals:

> **Definition 2.2.** If the population is finite or countable infinite, the random variable is *dis-*

crete; otherwise, it is *continuous*.

To be added: Examples...

As pointed out in definition 2.1, the units in a population may have more than one value. If each unit has one value, the population directly corresponds to one random variable. However, if the units have more than one value, we may consider each value to represent a separate random variable, which then have a joint probability distribution.

> **Definition 2.3.** If each unit in the population is associated to a single value, the probability distribution is *univariate*; otherwise it is *multi-variate*.

To be added: Examples...

Finally, we can also study how the values of the units in a population is varying:

> **Definition 2.4.** If all units in univariate population have the same or almost the same value, the population is said to be *homogeneous*.

> **Definition 2.5.** If most units in univariate population have the different values, the population is said to be *heterogeneous*.

> **Definition 2.6.** If the majority of the units in a univariate population have the same value (these units are referred to as the *conformist* units), the population is said to be *duogeneous*. The remainder of the population (which is referred to as the *diverging* units) may either be homogeneous (i.e, all diverging units have the same value) or heterogeneous (i.e., the diverging units have different values).

To be added: Examples...

It may be noted that the difference between a homogeneous and a heterogeneous population may depend on the situation.

To be added: Example where the population is homogeneous for a rough estimate, whereas it can be considered heterogeneous if a precise estimate is required.

2.1.2   Other Common Definitions of Probability Distributions

To be added: Definitions of density function, frequency function, distribution function and duration curve...

2.2   Statistical Properties

The probability distribution of a random variable is a complete description of its behaviour. However, in many cases we do not need such detailed information, but would prefer some key values that describe the main characteristics of the random variable. Therefore, different statistical measures have been introduced. The most important statistical measures are defined below.

Expectation Value

The expectation value of random variable is a the mean of all possible outcomes weighted by probability:

> **Definition 2.7.** The expectation value of a random variable X is given by

$$E[X] = \frac{1}{N} \sum_{i=1}^{N} x_i \text{ (population),}$$

$$E[X] = \sum_{x \in \Omega_X} f_X(x)x \text{ (discrete random variable),}$$

$$E[X] = \int_{x \in \Omega_X} f_X(x)x\,dx \text{ (continuous random variable).}$$

As seen above, the definition varies slightly depending on whether the probability distribution is expressed as a population or using a density function. We may notice that it is the expression corre-

sponding to the weighting according to probability that varies; hence, the following parts in the definitions above fulfil the same purpose:

$$\frac{1}{N} \sum_{i=1}^{N} \leftrightarrow \sum_{x \in \Omega_X} f_X(x) \leftrightarrow \int_{x \in \Omega_X} f_X(x)\dots dx.$$

As we will see below, this pattern will appear also in the other definitions of statistical properties.

The practical interpretation of the expectation value is that if we have a set of samples of a random variable, and this set is distributed exactly according to the probability distribution, then the expectation value is the mean of those samples. We can intuitively understand that if we have a large number of samples, it is quite likely that the samples will be distributed almost according to the probability distribution of the variable; hence, the mean of a large number of samples should be approximately equal to the expectation value. (In fact, this is the very foundation of simple sampling, as we will see in chapter 4).

Although most random variables have a well-defined expectation value, one should be aware that there is no guarantee that this is the case. This might seem strange, and is best understood by an example:

> **Example 2.2 (The S:t Petersburg paradox).** Consider a game where a player pays a fixed fee to participate. The player then tosses a coin until a head appears. If a head appears in the j:th trial, the payout of the game is 2^j.
>
> To be added: Expected payout of the game...

Variance and Standard Deviation

The variance of a random variable describes how much a random variable is varying around the expectation value.

To be added: Figure

> **Definition 2.8.** The variance of a random variable X is given by
>
> $$Var[X] = E[(X - E[X])^2] = E[X^2] - (E[X])^2 \text{ (general definition)},$$
>
> $$Var[X] = \frac{1}{N} \sum_{i=1}^{N} (x_i - E[X])^2 \text{ (population)},$$
>
> $$Var[X] = \sum_{x \in \Omega_X} f_X(x)(x - E[X])^2 \text{ (discrete random variable)},$$
>
> $$Var[X] = \int_{x \in \Omega_X} f_X(x)(x - E[X])^2 dx \text{ (continuous random variable)}.$$

A disadvantage with variance is that the unit of $Var[X]$ is the square of the unit of X. For example, if X is the unserved load expressed in MWh/h then the variance of X is expressed in $(\text{MWh/h})^2$. In many cases, it is more convenient to have a measure of the variation that is directly comparable to the variable itself and the expectation value. Therefore, the notion of standard deviation has been introduced:

> **Definition 2.9.** The standard deviation of a random variable X is given by
>
> $$\sigma_X = \sqrt{Var[X]}.$$

Covariance and Correlation Coefficient

The covariance describes how different random variables in a multivariate distribution are interacting with each other:

> **Definition 2.10.** The covariance of two random variables X and Y is given by
>
> $$Cov[X, Y] = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y].$$

Switching place between X and Y in definition 2.10 does not change the result, i.e.,

$$Cov[X, Y] = Cov[Y, X]. \tag{2.1}$$

We can also observe that the covariance between a variable and itself is equal to the variance of the variable, i.e.,

$$Cov[X, X] = Var[X]. \tag{2.2}$$

To be added: Discussion of covariance matrix...

Definition 2.11. A covariance matrix shows the covariances between all random variables in a multivariate probability distribution:

$$\Sigma_X = \begin{bmatrix} Var[X_1] & Cov[X_1, X_2] & \dots Cov[X_1, X_k] \\ Cov[X_2, X_1] & Var[X_2] & \\ \vdots & & \ddots \\ Cov[X_k, X_1] & & Var[X_k] \end{bmatrix}.$$

To be added: Discussion of correlation coefficient...

Definition 2.12. The correlation factor of two random variables X and Y is given by

$$\rho_{X, Y} = \frac{Cov[X, Y]}{\sqrt{Var[X]Var[Y]}}.$$

To be added: Discussion of independent random variables...

Definition 2.13. X and Y are independent random variables if it holds for each x and y that

$$f_{X, Y}(x, y) = f_X(x)f_Y(y) \Leftrightarrow F_{X, Y}(x, y) = F_X(x)F_Y(y).$$

Theorem 2.14. If X and Y are independent then

$$E[XY] = E[X]E[Y].$$

Corollary 2.15. If X and Y are independent then they are also uncorralated.

It should be noted that the reverse of corollary 2.15 does not hold, i.e., if two variables are uncorrelated we cannot conclude that they are independent! This is because the correlation is only a measure of the linear dependence of two random variables; hence, if there is a purely non-linear relation between two random variables, they will still be uncorrelated. Compare the following two examples:

Example 2.3. Assume that Y has the frequency function

$$f_X(x) = \begin{cases} 1/5 & \text{if } x = 0, 1, 2, 3, 4, \\ 0 & \text{otherwise.} \end{cases}$$

and assume that $X = Y^2$. Are X and Y uncorrelated?

Solution: We start by computing expectation values according to definition 2.7:

$$E[XY] = \frac{1}{5}(0 + 1^2 \cdot 1 + 2^2 \cdot 2 + 3^2 \cdot 3 + 4^2 \cdot 4) = 20,$$

$$E[X] = \frac{1}{5}(0 + 1^2 + 2^2 + 3^2 + 4^2) = 6,$$

$$E[X] = \frac{1}{5}(0 + 1 + 2 + 3 + 4) = 2.$$

The covariance of X and Y can now be computed according to definition 2.10:

$$Cov[X, Y] = E[XY] - E[X]E[Y] = 20 + 6 \cdot 2 = 32.$$

As the covariance is larger than zero, we can conclude that X and Y are positively correlated.

Example 2.4. Assume that Y has the frequency function

$$f_X(x) = \begin{cases} 1/5 & \text{if } x = -2, -1, 0, 1, 2, \\ 0 & \text{otherwise.} \end{cases}$$

and assume that $X = Y^2$. Are X and Y uncorrelated?

Solution: We start by computing expectation values according to definition 2.7:

$$E[XY] = \frac{1}{5}((-2)^2 \cdot (-2) + (-1)^2 \cdot (-1) + 0 + 1^2 \cdot 1 + 2^2 \cdot 2) = 0,$$

$$E[X] = \frac{1}{5}((-2)^2 + (-1)^2 + 0 + 1^2 + 2^2) = 2,$$

$$E[X] = \frac{1}{5}((-2) + (-1) + 0 + 1 + 2) = 0.$$

The covariance of X and Y can now be computed according to definition 2.10:

$$Cov[X, Y] = E[XY] - E[X]E[Y] = 0 + 2 \cdot 0 = 0.$$

As the covariance is zero, we can conclude that X and Y are uncorrelated.

2.3   Arithmetics of Random Variables

To be added: Basic rules for arithmetic operations on random variables...

Exercises

2.1   Are the following random variables heterogeneous or duogeneous?

a) $f_X(x) = \begin{cases} 0.1 & \text{if } 10 \leq x \leq 20, \\ 0 & \text{otherwise.} \end{cases}$

b) $f_X(x) = \begin{cases} 1/11 & \text{if } x = 10, 11, \ldots, 20, \\ 0 & \text{otherwise.} \end{cases}$

c) $f_X(x) = \begin{cases} 0.6 & \text{if } x = 10, \\ 0.4 & \text{if } x = 20, \\ 0 & \text{otherwise.} \end{cases}$

d) $f_X(x) = \begin{cases} 0.6 & \text{if } x = 10, \\ 0.04 & \text{if } 10 < x \leq 20, \\ 0 & \text{otherwise.} \end{cases}$

e) $f_X(x) = \begin{cases} 0.3 & \text{if } 10 \leq x \leq 12, \\ 0.05 & \text{if } 12 < x \leq 20, \\ 0 & \text{otherwise.} \end{cases}$

More exercises to be added...

RANDOM NUMBERS

The idea of a Monte Carlo simulation is to estimate the behaviour of the simulated system using random samples. When sampling a physical system, the randomness of the samples will be generated by the system itself. However, when simulating a system using a mathematical model, it will be necessary to generate random input values. This chapter will present methods how to do this. The presentation will start with a description how to generate $U(0, 1)$-distributed random numbers and then it is explained how random numbers of any other distribution can be obtained by transformation of random numbers from a $U(0, 1)$-distribution.

3.1   Pseudo-random Numbers

One possibility to provide random numbers in a computer simulation would be to use some kind of hardware device. Such a device could be designed to generate truly random values, but there would also be an important disadvantage, namely that we would lack control of the produced random numbers. As a consequence, running the same simulation twice would generally not produce the same set of samples, which can be a problem especially when testing a simulation method or model. For example, assume that we run a simulation and we detect a few scenarios where the mathematical model produces erroneous output models. Once the errors have been corrected, it would be practical to be able to run the same scenarios again in order to verify that the problem has been solved.

Therefore, random number generation in computers are based on special mathematical functions or algorithms, which given one or more initial values (referred to as *seeds*) produces a sequence of numbers between 0 and 1. This sequence is in reality deterministic, which means that if the same seed is used, it will produce the same sequence (which makes simulations repeatable). Since these functions are not truly random, they are called *pseudo-random number generators*. However, a properly designed function will generate a sequence that has properties as close as possible to that of a true sequence of independent $U(0, 1)$-distributed random numbers.

In practice, we do not need to worry about designing good pseudo-random number generators, as such functions are readily available in almost all high-level programming languages. In fact, it is preferable to use built-in pseudo-random number generators rather than programming one of your own, as the built-in functions should have been carefully designed to provide appropriate statistical properties.

Nevertheless, it can be of interest to get an idea of the principles for generation of pseudo-random numbers. As an example, we will study the widely used *linear congruential generator*. The sequence of numbers for this pseudo-random number generator is computed using the following formulae:

$$X_{i+1} = (aX_i + c) \bmod m, \tag{3.1a}$$

$$U_i = \frac{X_i}{m}, \tag{3.1b}$$

where

$$X_1 = \text{the seed,}$$

X_i = internal state i,
a = an integer multiplier $(0 < a < m)$,
c = an integer increment $(0 < c < m)$,
m = an integer divisor $(0 < m)$,
U_i = pseudo-random number i.

The modulo operator in (3.1a) returns the remainder of when dividing $(aX_i + c)$ by m. The result of this operation is an integer in the interval $0, \ldots, m - 1$. This means that the linear congruential generator can at most produce m possible values before the sequence starts repeating itself. In order to make the sequence as similar to a uniform distribution as possible, we would like the sequence to be as long as possible. This can be achieved if the following rules are considered when choosing the constants a, c and m:

- The greatest common divisor of c and m should be 1.

- $a - 1$ should be divisable by all prime factors of m.

- $a - 1$ should be a multiple of 4 if m is a multiple of 4.

The application of linear congruential generators is demonstrated in the following two examples:

Example 3.1. What sequence of numbers is generated by a linear congruential generator where $a = 5$, $c = 3$ and $m = 8$ for the seed $X_1 = 1$?

Solution: We can notice that the constants are fulfil the requirements above. The computations when starting with $X_1 = 1$ and applying (3.1a) are shown in table 3.1 below. We can see that the sequence will repeat itself after eight values, which is the maximal length of the sequence when $m = 8$.

Table 3.1 Computation of the random number sequence in example 3.1.

i	1	2	3	4	5	6	7	8
X_i	1	0	3	2	5	4	7	6
U_i	0.125	0.000	0.375	0.250	0.625	0.500	0.875	0.750
$5X_i + 3$	8	3	18	13	28	15	38	33
X_{i+1}	0	3	2	5	4	7	6	1

Example 3.2. What sequence of numbers is generated by a linear congruential generator where $a = 5$, $c = 6$ and $m = 8$ for the seed $X_1 = 1$?

Solution: This time we do not fulfil the first requirement, as both c and m can be divided by 2. The computations when starting with $X_1 = 1$ and applying (3.1a) are shown in table 3.2 below. This time we only get four values before the sequence starts repeating itself, and for the remainder of the sequence we will only see the internal states 3, 5 and 7..

Table 3.2 Computation of the random number sequence in example 3.2.

i	1	2	3	4
X_i	1	3	5	7
U_i	0.125	0.375	0.625	0.750
$5X_i + 6$	11	21	31	41
X_{i+1}	3	5	7	3

3.2 Transformation of Random Numbers

As described in the previous section, random inputs for Monte Carlo simulations can be created by standard pseudorandom number generators available in almost all programming languages. However, these generators produce $U(0, 1)$-distributed random numbers and unfortunately it is rarely so that all inputs of the system to be simulated have that probability distribution. Hence, we need to be able to convert $U(0, 1)$-distributed random numbers to the actual probability distributions of the inputs. There

are several methods that can be applied to perform this transformation. In this section, some methods that are particularly efficient for Monte Carlo simulations are described.

3.2.1 Independent Random Numbers

To be added: Introduction...

The Inverse Transform Method

To be added: Introduction...

> **Theorem 3.1. (Inverse Transform Method)** If U is a $U(0, 1)$-distributed random number then Y is a distributed according to the distribution function $F_Y(x)$ if Y is calculated according to $Y = F_Y^{-1}(U)$.

To be added: Comments and examples...

Random Numbers from Finite Populations

To be added: Introduction...
   To be added: Search algortihm...
   To be added: Comments... (Notice that the states must be sorted if complementary random numbers will be applied to an input!)
   To be added: Examples...

Normally Distributed Random Numbers

To be added: Introduction...

> **Theorem 3.2. (Approximate Inverse Transform Method)** If U is a $U(0, 1)$-distributed random number then Y is a $N(0, 1)$-distributed random number, if Y is calculated according to

$$Q = \begin{cases} U & \text{if } 0 \leq U \leq 0.5, \\ 1 - U & \text{if } 0.5 < U \leq 1, \end{cases}$$

$$t = \sqrt{-2\ln Q},$$

$$c_0 = 2.515517, \qquad c_1 = 0.802853, \qquad c_2 = 0.010328,$$

$$d_1 = 1.432788, \qquad d_2 = 0.189269, \qquad d_3 = 0.001308,$$

$$z = t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}$$

and finally

$$Y = \begin{cases} -z & \text{if } 0 \leq U < 0.5, \\ 0 & \text{if } U = 0.5, \\ z & \text{if } 0.5 < U \leq 1. \end{cases}$$

To be added: Comments and exampels...

3.2.2 Correlated Random Numbers

A Monte Carlo simulation can have inputs that are correlated. It should however be noted that generation of correlated random numbers is not as straightforward as generation of independent random numbers.

General Method

To be added...

Random Numbers from Finite Populations

To be added...

Normally Distributed Random Numbers

To be added...

Approximative Method

To be added...

Exercises

To be added...

SIMPLE SAMPLING

The most straightforward approach to Monte Carlo simulation is to collect samples completely at random. This sampling strategy is referred to as *simple sampling* and a Monte Carlo simulation based on simple sampling might be referred to as a *crude Monte Carlo simulation*. Although we will see in the next chapter that it is in fact possible to get better results using other sampling methods, it is necessary to understand simple sampling before other methods are studied. Moreover, in some cases simple sampling will be the most efficient simulation method, either due to the properties of the simulated study or because the time to develop a more advanced simulation method is longer than the time to obtain a reasonable result from simple sampling.

4.1 Estimating Expectation Values

The idea of a Monte Carlo simulation is to collect random samples and based on these observations try to compute one or more values, for example a statistical property of a random variable or a deterministic value such as the mathematical constant π in Buffon's needle experiment. However, as the result is depending on random observations, it is inevitable that we introduce a random error—the result of a Monte Carlo simulation will most likely not be exactly equal to the true value. In that sense, the result of a Monte Carlo simulation is just an estimate, which we would like to be as close as possible to the true value.

It is important to notice that since the result of a Monte Carlo simulation is a function of several random samples, the result must also be a random variable. This is reflected in the notation for estimates that will be used in this compendium. If we are discussing general properties of an estimate calculated using one or another simulation method, we will denote the estimate by an upper-case Latin letter corresponding to the statistical property that we are trying to estimate. For example, the expectation value of the random variable X is denoted μ_X, which means that we choose the symbol M_X for the random variable representing the estimate of μ_X. However, if we are actually computing an estimate then we are in fact studying an outcome of M_X, and we will then use the symbol m_X, i.e., the lower-case Latin counterpart of μ_X.

Accuracy and Precision

In everyday language, we may use the word "accuracy" to describe if an estimate is close or not to the true value. However, in science, accuracy should be separated from "precision". In this section, the meaning and difference between these two notions will be described, and we will at the same time discuss desirable properties of an estimate.

What we want from an estimate is that it should be as close as possible to the true value. Each simulation method that we create will generate estimates from a certain probability distribution; therefore, we may study this probability distribution of the estimate, in order to determine if it is likely or not that we will get a result close to the true value. It can then be seen that there are two different ways that an estimate can be "close" to the true value.

Let us start by examining the difference between the expectation value of the estimate, $\mathrm{E}[M_X]$, and the
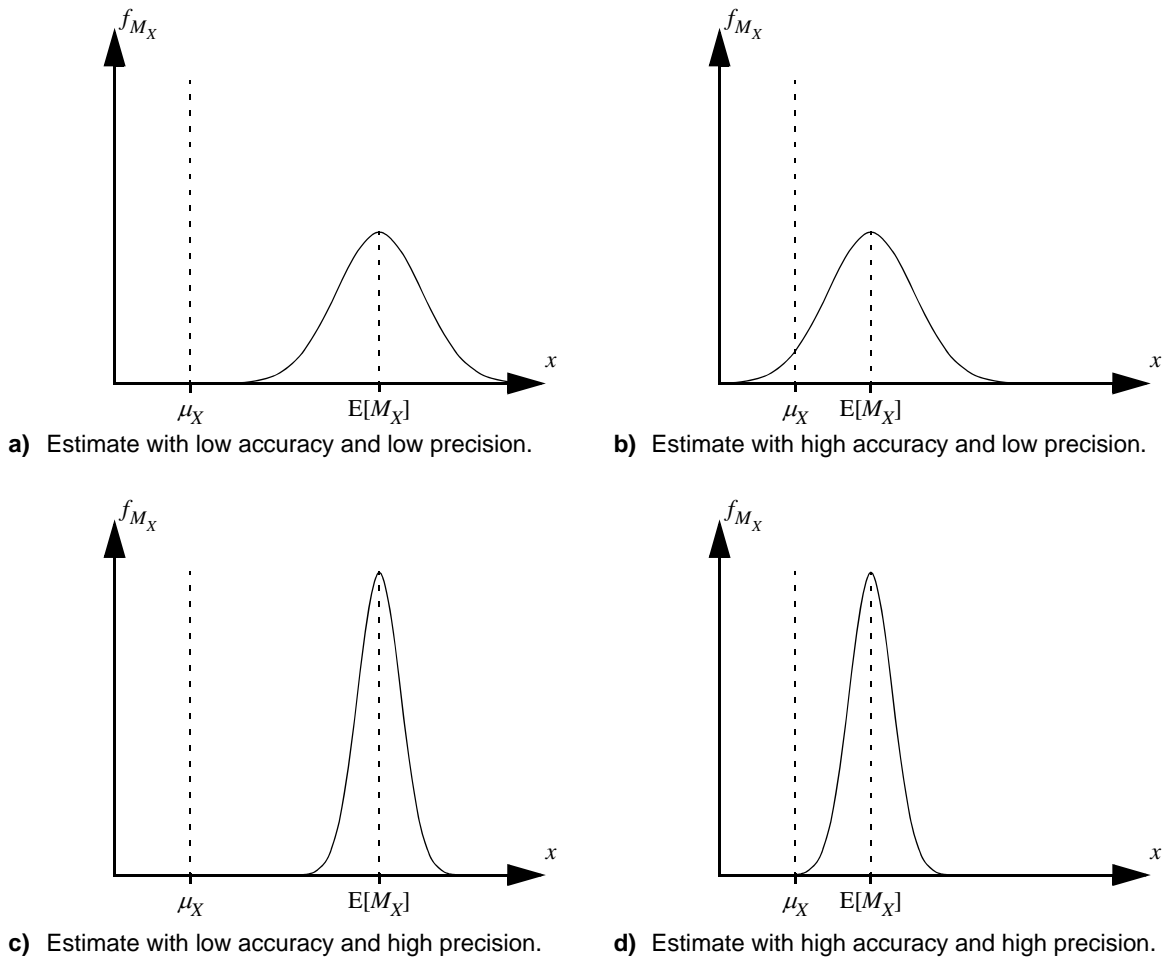
a) Estimate with low accuracy and low precision.

b) Estimate with high accuracy and low precision.

c) Estimate with low accuracy and high precision.

d) Estimate with high accuracy and high precision.

Figure 4.1   Illustration of the meaning of "accuracy" and "precision" respectively.

true value, μ_X. This difference is referred to as the *accuracy* of the estimate. Examples of estimates with low and high accuracy are shown in figures 4.1a and 4.1b. We can see that a high accuracy means that the difference between $E[M_X]$ and μ_X is small; consequently, we have a better probability that the resulting estimate from a simulation is close to the true value compared to a simulation with lower accuracy. Since the accuracy is a constant for a given simulation method (remember that each simulation method will result in its own probability distribution of the estimate) it represents the *systematical error* in a simulation. The ideal is of course if we can set up a simulation with no systematical error, i.e., that $E[M_X] = \mu_X$. In this case, we say that we have an *unbiased estimate*.

Now, we may continue by examining the variance of the estimate, $\mathrm{Var}[M_X]$. If we go back to the definition, the variance is the expected quadratic deviation from the expectation value; hence, the variance describes how much a random variable will vary around its expectation value. A low variance means that it is likely that we only have a small deviation between a single observation of the random variable and its expectation value, whereas a high variance means that it is more likely to have large deviations. The variance of the estimate is referred to as the *precision* of the estimate. The precision is inversely proportional to the variance, i.e., a large variance of the estimate means that we have a low precision, whereas a low variance yields a high precision (cf. figures 4.1c and 4.1d). The precision represents the random error of a Monte Carlo simulation. Unlike systematical errors, this is something that cannot be avoided when sampling. What we would like to is therefore to keep the random error as small as possible.

The Law of Large Numbers

To be added: Introduction...

> **Theorem 4.1. (Law of Large Numbers):** If $x_1, \ldots, x_n$ are independent observations of the random variable X then

$$m_X = \frac{1}{n} \sum_{i=1}^{n} x_i$$

is an unbiased estimate of $E[X]$.

Proof: M_X is an unbiased estimate of $E[X]$ if $E[M_X] = E[X]$. Let T_i denote the number of times that unit i appears in the samples; hence, the outcome of T_i is an integer between 0 and n. The estimate of the expectation value can then be expressed as

$$M_X = \frac{1}{n} \sum_{i=1}^{N} T_i \mathcal{X}_i.$$

The number of successful trials when n trials are performed and the probability of success is p in each trial is Binomial-distributed, i.e, T_i is $B(n,\ 1/N)$-distributed.

$$E[M_X] = E\left[\frac{1}{n} \sum_{i=1}^{N} T_i \mathcal{X}_i\right] = \sum_{i=1}^{N} \frac{1}{n} E[T_i] \mathcal{X}_i = \{\text{the expectation value of a } B(n, p)\text{-distribution is}$$

$$n{\cdot}p\} = \sum_{i=1}^{N} \frac{1}{n}\frac{n}{N} \mathcal{X}_i = \frac{1}{N} \sum_{i=1}^{N} \mathcal{X}_i = E[X]. \ \blacksquare$$

Notice the similarity between the definition of expectation value (definition 2.7) and the formula for the estimate of the expectation value (theorem 4.1 above):

$$\mu_X = \frac{1}{N} \sum_{i=1}^{N} x_i \leftrightarrow m_X = \frac{1}{n} \sum_{i=1}^{n} x_i. \tag{4.1}$$

When calculating the expectation value analytically, we enumerate all units in the population and compute the mean value, whereas in simple sampling we enumerate all selected samples and compute the mean value.

Replacement

To be added: Definition of replacement...
To be added: Discussion of the precision of the estimated expectation value...
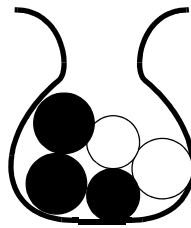
Figure 4.2   Simple urn experiment.

Example 4.1. Consider an urn with five balls. The balls are either white or black. (They also have different sizes, but that we will ignore in this example.) Introduce a random variable X representing the colour of a randomly chosen ball from the urn, such that

$$X = \begin{cases} 0 & \text{if the ball is white,} \\ 1 & \text{if the ball is black.} \end{cases}$$

Examine the probability distribution of M_X if five samples are collected with and without replacement respectively.

Solution: To be added: Define frequency function, compute μ_X, explain figure 4.3 (and add the probability of each state during the sample process)...

Theorem 4.2. In simple sampling, the variance of the estimated expectation value is

$$\text{Var}[M_X] = \frac{\text{Var}[X]}{n} \cdot \frac{N-n}{N}.$$

Proof: To be added?

The factor $(N - n)/N$ is called fpc (*finite population correction*). For infinite populations we get

$$\text{Var}[M_X] = \frac{\text{Var}[X]}{n}. \tag{4.2}$$

To be added: Interpretation of the theorem & examples...

4.2   Other Estimates

To be added: Introduction...

Estimating Variance

To be added: Introduction...

From the comparison in (4.1) one might guess that an estimate of the variance can be obtained by

$$\frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_X)^2. \tag{4.3}$$

However, this estimate requires that we know the true expectation value, and if this was the case we would also be able to analytically compute the variance. Hence, (4.3) cannot be applied in practice. But what about replacing the true expectation value by the estimated expectation value, i.e.,

$$\frac{1}{n} \sum_{i=1}^{n} (x_i - m_X)^2. \tag{4.4}$$

It turns out that the (4.4) is indeed an estimate of $\text{Var}[X]$, although, the estimate is biased.

To be added: Explanation why (4.4) is biased?

Theorem 4.3. If $x_1, \dots, x_n$ are independent observations of the random variable X then

$$s_X^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - m_X)^2$$

is an unbiased estimate of $\text{Var}[X]$.

The proof of theorem 4.3 is not trivial and will therefore will be omitted from this presentation. As a substitute, we may verify the theorem as well as (4.3) and (4.4) by an example:

Example 4.2. Consider the urn experiment from example 4.1. Examine the probability distribution of the estimates (4.3), (4.4) and theorem 4.3 respectively if five samples are collected (with replacement).

Solution: The true variance of the urn experiment is given by

$$\text{Var}[X] = \sum_x f_X(x)(x - \mu_X)^2 = 0.4 \cdot (0 - 0.6)^2 + 0.6 \cdot (1 - 0.6)^2 = 0.24.$$

As the order in which the samples appear is of no interest for neither the estimated mean nor the estimated variance, we can distinguish between six possible outcomes, as listed in table 4.1 below. The estimates of mean and variance are then computed for each combination of collected samples. The probability for a certain result can be computed using the
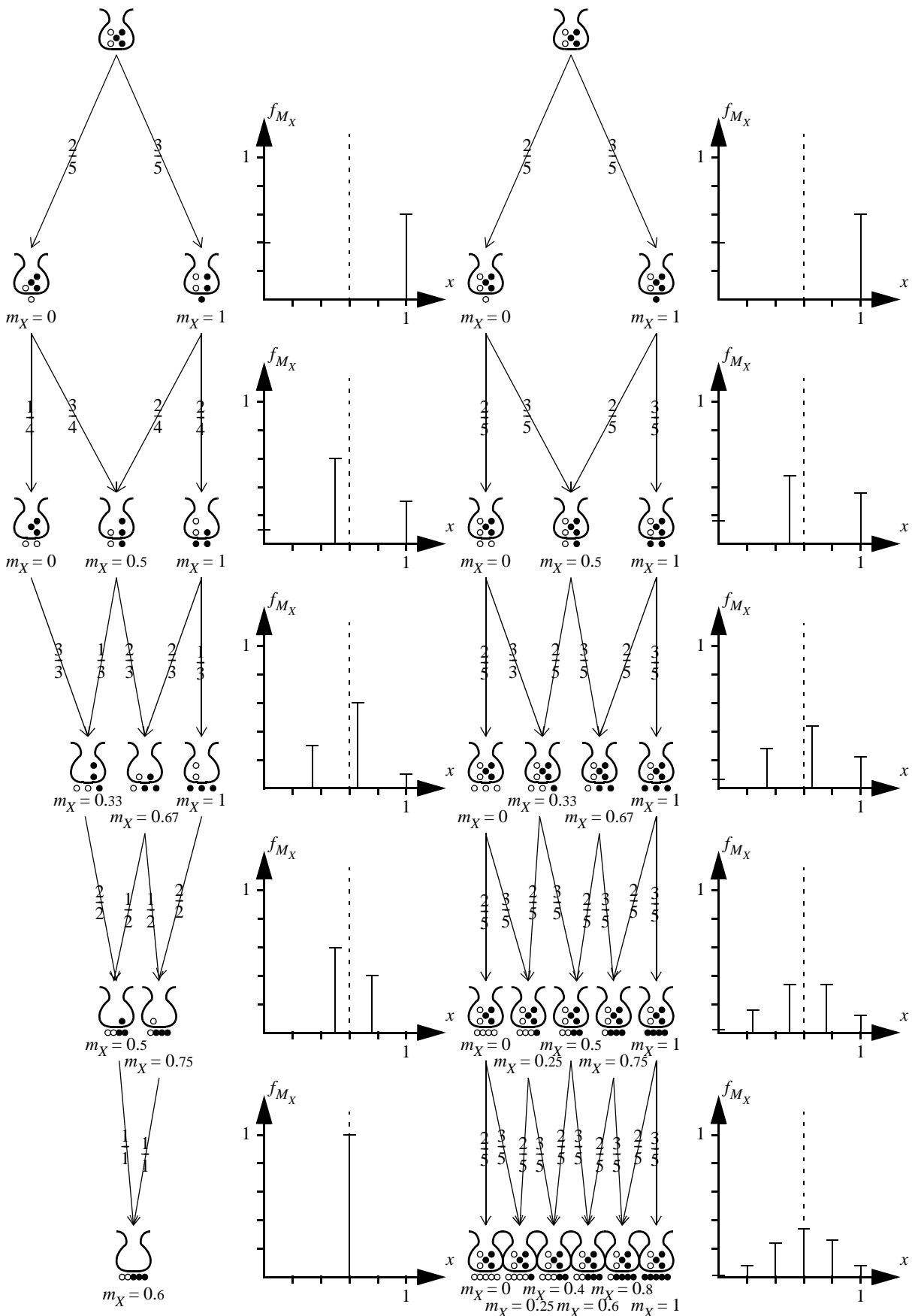
Figure 4.3   Probability distribution of the estimated expectation value in the urn experiment. Sampling   without replacement is shown to the left and sampling with replacement is shown to the right.

event tree in figure 4.3.

Table 4.1 Probability distribution of the estimated variance in the urn experiment.

Collected samples	○○○○○	○○○○●	○○○●●	○○●●●	○●●●●	●●●●●
Probability	$\frac{32}{3\ 125}$	$\frac{240}{3\ 125}$	$\frac{720}{3\ 125}$	$\frac{1\ 080}{3\ 125}$	$\frac{810}{3\ 125}$	$\frac{243}{3\ 125}$
Estimated mean, m_X	0	0.2	0.4	0.6	0.8	1
Estimated variance $\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu_X)^2$	0.36	0.32	0.28	0.24	0.2	0.16
$\frac{1}{n}\sum_{i=1}^{n}(x_i-m_X)^2$	0	0.16	0.24	0.24	0.16	0
$\frac{1}{n-1}\sum_{i=1}^{n}(x_i-m_X)^2$	0	0.2	0.3	0.3	0.2	0

From the results in table 4.1 we can compute the expectation value of the estimated variance for the three different methods:

$$\mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu_X)^2\right] = \frac{32}{3\ 125}0.36 + \frac{240}{3\ 125}0.32 + \frac{720}{3\ 125}0.28 + \frac{1\ 080}{3\ 125}0.24 + \frac{810}{3\ 125}0.2 + \frac{243}{3\ 125}0.16 = 0.24.$$

$$\mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-m_X)^2\right] = \frac{32}{3\ 125}0 + \frac{240}{3\ 125}0.16 + \frac{720}{3\ 125}0.24 + \frac{1\ 080}{3\ 125}0.24 + \frac{810}{3\ 125}0.16 + \frac{243}{3\ 125}0 = 0.192.$$

$$\mathrm{E}\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i-m_X)^2\right] = \frac{32}{3\ 125}0 + \frac{240}{3\ 125}0.2 + \frac{720}{3\ 125}0.2 + \frac{1\ 080}{3\ 125}0.3 + \frac{810}{3\ 125}0.3 + \frac{243}{3\ 125}0 = 0.24.$$

These results confirm that (4.3) and theorem 4.3 are unbiased estimates (because the expectation value of the estimate is equal to the true value), whereas (4.4) tends to underestimate the variance.

To be added: Show how S_X^2 can be computed using sums of x_i and x_i^2.

To be added: Estimating covariance...

Estimating Probability Distributions

To be added...

Estimating Covariance

To be added...

4.3 Application

To be added: Introduction...

Simulation Procedure

To be added: Block diagram and discussion...

Stopping Rules

To be added: Discussion of how many samples should be analysed in a Monte Carlo simulation...

Confidence Intervals

To be added: Confidence intervals...

Chapter 5
VARIANCE REDUCTION TECHNIQUES

It was shown in the previous chapter that the variance of an estimated expectation value, $Var[M_X]$, is related to the precision of the simulation; a low variance means that it is more likely that the result will be accurate, whereas a high variance means that there is a larger risk that the result will be inaccurate. We also learned that $Var[M_X]$ is depending on the probability distribution of the samples variable (which we cannot affect) and the number of samples (which we do control). However, until now we have studied simple sampling, where the samples are selected completely at random (i.e., each unit in the population has the same probability of being selected). Interestingly, if we manipulate the selection of samples, the variance of the estimate can be lower than for simple sampling. Such methods to improve the precision of a Monte Carlo simulation are referred to as *variance reduction techniques*. This chapter will describe six variance reduction techniques.

5.1   Complementary Random Numbers

The idea behind complementary random numbers is to reduce the influence of random fluctuations, which always appear in sampling, by creating a negative correlation between samples. In practice, this means that the generation of random numbers is manipulated in such a way that the probability of an even spread over the whole population increases.

5.1.1 Principle

Assume that the expectation value $E[X] = \mu_X$ has been estimated in two separate simulations, i.e., we have two estimates M_{X1} and M_{X2} such that

$$E[M_{X1}] = E[M_{X2}] = \mu_X. \tag{5.1}$$

It is not surprising that the mean of the two estimates is also an estimate of μ_X, and we can easily verify that this is the case, because

$$E\left[\frac{M_{X1} + M_{X2}}{2}\right] = \frac{1}{2}(E[M_{X1}] + E[M_{X2}]) = \frac{1}{2}(\mu_X + \mu_X) = \mu_X. \tag{5.2}$$

As the expectation value of the mean estimate is equal to μ_X, the mean estimate is in itself an unbiased estimate of μ_X in accordance to the discussion in section 4.1. The interesting question is now how precise this mean estimate is compared to simple sampling with the same total number of samples. From (4.2) we have that simple sampling with in total n samples results in

$$Var[M_X] = \frac{Var[X]}{n} = \frac{\sigma_X^2}{n}. \tag{5.3}$$

Now assume that the two estimates M_{X1} and M_{X2} each include $n/2$ samples (which means that the total number of samples is still n) then variance of the mean estimate is given by

$$Var\left[\frac{M_{X1} + M_{X2}}{2}\right] = \frac{1}{4}Var[M_{X1}] + \frac{1}{4}Var[M_{X2}] + \frac{1}{4} \cdot 2Cov(M_{X1}, M_{X2}). \tag{5.4}$$

If M_{X1} and M_{X2} both are independent estimates obtained with simple sampling then we get $Var[M_{X1}] = Var[M_{X2}] = \sigma_X^2$ and $Cov[M_{X1}, M_{X2}] = 0$. Hence, (5.4) yields

$$Var\left[\frac{M_{X1} + M_{X2}}{2}\right] = \frac{1}{4}(Var[M_{X1}] + Var[M_{X1}]) = \frac{Var[M_{X1}]}{2} = \frac{\sigma_X^2}{n}. \tag{5.5}$$

By comparing (5.3) and (5.5) we see that the precision is the same, which is also what we should expect—running one simulation of n independent samples should be the same thing as combining the results of two independent simulations of $n/2$ independent samples each.

However, the interesting part is that if the two estimates are *not* independent but negatively correlated then (5.4) will result in a variance that is lower than the variance for simple sampling with the same amount of samples. The questions is then how we should proceed in order to find estimates that are negatively correlated. A straightforward method is to use *complementary random numbers* when we generate scenarios for a Monte Carlo simulation.

- We start with a random number from a random number generator, which we have seen in section 3.1, corresponds to a $U(0, 1)$-distributed random variable. If U is a value from the random number generator then we define $U^* = 1 - U$ as the complementary random number of U. It is easy to verify that U and U^* are negatively correlated with $\rho_{U, U*} = -1$.

- Then we transform both the original random number and the complementary random number into the probability distribution of the input. If this is done using a the inverse transform method then we get that $Y = F_Y^{-1}(U)$ and $Y^* = F_Y^{-1}(U^*)$. These values will also be negatively correlated, but the transformation may weaken the correlation, i.e., we get $\rho_{Y, Y*} \geq -1$. This also holds for the other transformation methods presented in section 3.2.

- Next we compute the value of the output for both original and the complementary input value, i.e, $X = g(Y)$ and $X^* = g(Y^*)$. If the simulated system is such that there is a correlation between the input and output values then the original and complementary output values will also be negatively correlated, but again the correlation might be weakened, i.e., we get $\rho_{X, X*} \geq \rho_{Y, Y*}$.

- Finally, we let M_{X1} be an estimate based on n samples of X, whereas M_{X2} is obtained from sampling the corresponding values of X^*. Obviously, we will now have two estimates that are negatively correlated.

5.1.2 Application

To be added: Introduction...

Multiple Inputs

To be added: Discussion on how to manage systems with multiple inputs...

Auxiliary Inputs

Complementary random numbers result in a variance reduction if there is a negative correlation between the original and complementary result variables (X and X^*). The challenge when setting up a simulation is however that we cannot directly control the values of the result variables, but only the values of the inputs (Y and Y^*)—complementary random numbers will be efficient if the negative correlation between Y and Y^* also results in a negative correlation between $X = g(Y)$ and $X^* = g(Y^*)$. This may not be true for all inputs to a system, and we should then not apply complementary random numbers to these inputs.

Sometimes it might even not be possible to find any input variable that fulfils the criterion above. This does however not necessarily mean that complementary random numbers are not applicable to this system, as there might be a possibility to introduce an auxiliary input that does fulfil the criterion. An auxiliary input, Y_E, is defined as an arbitrary function of the actual inputs to the system, i.e.,

$$Y_E = h(Y_1, ..., Y_J). \tag{5.6}$$

A straightforward choice of the arbitrary function h would be to use the sum of some of the inputs.

Example 5.1 (auxiliary inputs for a small multi-area power system). Consider a power system divided in three areas. The inputs to a simulation of this system is the available generation capacity in each unit, $\overline{G}_1$, …, $\overline{G}_4$ and the load in each area, D_1, D_2, D_3. The objective of the simulation is to study the total operation cost and the reliability of supply. Assume that complementary random numbers should be used to improve the efficiency of the simulation. To which inputs should we apply complementary random numbers?

Solution: Each individual input will only have a weak impact on the outputs. For example, if one unit is unavailable we cannot predict that the system will have a higher operation cost or that there will be load shedding, because the other units can be available or the load can be low. However, we can introduce two auxiliary inputs, namely the total available generation capacity, $\overline{G}^{\mathrm{tot}} = \overline{G}_1 + \overline{G}_2 + \overline{G}_3 + \overline{G}_4$, and total demand, $D^{\mathrm{tot}} = D_1 + D_2 + D_3$. A negative correlation between two values of the auxiliary inputs will also result in a negative correlation between at least one of the result variables, as illustrated in table 5.1.

Table 5.1 Correlations between auxiliary inputs and result variables in example 5.1.

Total available generation capacity, $\overline{G}^{\mathrm{tot}}$	Total demand, D^{tot}	Total operation cost, *TOC*	Loss of load, *LOLO*
Low	Low	Low	Not likely
High	Low	Low	Very unlikely
Low	High	High	Quite likely
High	High	High	Not likely

In order to correctly simulate a system considering auxiliary inputs, we need to have the probability distribution of the auxiliary inputs. Although this may require some computational efforts, it should not be a problem as the probability distribution of all the original inputs is known. Moreover, we must make sure that the random values of the auxiliary inputs are consistent with the random values of the original inputs. As the auxiliary inputs are a function of some original inputs, there will clearly be a correlation between auxiliary and original inputs. We could therefore identify their joint probability distribution and randomise values according to this distribution.
To be added: Further discussion (challenge of generating correlated numbers, cf. section 3.2.2).

Example 5.2 (joint probability distribution of available generation capacity). Data for the generating units in example 5.1 are shown in table 5.2. Identify the joint probability distribution for the total generation capacity and the generation capacity in each unit. Moreover, generate a scenario using the value 0.17 from a $U(0, 1)$-distribution as well as a complementary scenario.

Table 5.2 Data for the power system in example 5.2.

Unit, g	Area, n	Installed capacity, $\overline{G}_g$ [MW]	Availability [%]
1	1	500	95
2	2	500	95
3	2	800	90
4	3	1 000	90

Solution: As each unit can either be available or unavailable, we will in total have sixteen possible states for the available generation capacity of all units. For each of these states we can compute the total available generation capacity and the probability of that state. The result of this enumeration, sorted according to increasing values of $\overline{G}^{\mathrm{tot}}$ are listed in table 5.3.

To be added: Comments... (for example, why is there a difference between state 2 and 3)

To be added: Figure showing the distribution function and the result of applying the

Table 5.3 Joint probability distribution of the available generation capacity in example 5.2.

State, i	Total available generation capacity, $\overline{G}^{\,\mathrm{tot}}$ [MW]	Available generation capacity in unit 1, $\overline{G}_1$ [MW]	Available generation capacity in unit 2, $\overline{G}_2$ [MW]	Available generation capacity in unit 3, $\overline{G}_3$ [MW]	Available generation capacity in unit 4, $\overline{G}_4$ [MW]	Probability, $f_{\overline{G}}$
1	0	0	0	0	0	$0.05 \cdot 0.05 \cdot 0.1 \cdot 0.1 = 0.000025$
2	500	500	0	0	0	$0.95 \cdot 0.05 \cdot 0.1 \cdot 0.1 = 0.000475$
3	500	0	500	0	0	$0.05 \cdot 0.95 \cdot 0.1 \cdot 0.1 = 0.000475$
4	800	0	0	800	0	$0.05 \cdot 0.05 \cdot 0.9 \cdot 0.1 = 0.000225$
5	1 000	0	0	0	1 000	$0.05 \cdot 0.05 \cdot 0.1 \cdot 0.9 = 0.000225$
6	1 000	500	500	0	0	$0.95 \cdot 0.95 \cdot 0.1 \cdot 0.1 = 0.009000$
7	1 300	500	0	800	0	$0.95 \cdot 0.05 \cdot 0.9 \cdot 0.1 = 0.004275$
8	1 300	0	500	800	0	$0.05 \cdot 0.95 \cdot 0.9 \cdot 0.1 = 0.004275$
9	1 500	500	0	0	1 000	$0.95 \cdot 0.05 \cdot 0.1 \cdot 0.9 = 0.004275$
10	1 500	0	500	0	1 000	$0.05 \cdot 0.95 \cdot 0.1 \cdot 0.9 = 0.004275$
11	1 800	0	0	800	1 000	$0.05 \cdot 0.05 \cdot 0.9 \cdot 0.9 = 0.002025$
12	1 800	500	500	800	0	$0.95 \cdot 0.95 \cdot 0.9 \cdot 0.1 = 0.081225$
13	2 000	500	500	0	1 000	$0.95 \cdot 0.95 \cdot 0.1 \cdot 0.9 = 0.081225$
14	2 300	500	0	800	1 000	$0.95 \cdot 0.05 \cdot 0.9 \cdot 0.9 = 0.038475$
15	2 300	0	500	800	1 000	$0.05 \cdot 0.95 \cdot 0.9 \cdot 0.9 = 0.038475$
16	2 800	500	500	800	1 000	$0.95 \cdot 0.95 \cdot 0.9 \cdot 0.9 = 0.731025$

inverse transform method using $U = 0{,}17$ and $U^* = 0{,}83$.

Simulation Procedure

To be added: Overview, equations, block diagram and example...

5.2 Dagger Sampling

This variance reduction technique is based on a similar principle as complementary random numbers. Dagger sampling is however limited to two-state probability distributions.

To be added: Further comments?

5.2.1 Principle

Consider a two-state random variable Y with the frequency function

$$f_Y(x) = \begin{cases} 1 - p & \text{if } x = a, \\ p & \text{if } x = b, \\ 0 & \text{otherwise,} \end{cases} \tag{5.7}$$

where $p < 0.5$. This probability distribution is clearly fulfilling the criteria of a duogeneous population (cf. definition 2.6), with the value a being the conformist units and b the diverging units.

In dagger sampling, random values of Y are not generated by the inverse transform method, but using a dagger transform:

> **Theorem 5.1. (Dagger Transform)** If U is a $U(0, 1)$-distributed random number then Y is a distributed according to the frequency function (5.7) if Y is calculated according to
>
> $$F_{Yj}^{\ddagger}(x) = \begin{cases} a & \text{if } (j-1)p \leq x < jp, \\ b & \text{otherwise,} \end{cases} \quad \text{for } j = 1, \ldots, S,$$
>
> where S is the largest integer such that $S \leq 1/p$.

The value S in theorem 5.1 is referred to as the dagger cycle length. It should be noted that one random number from the pseudorandom number generator is used to generate S values of Y.

   Example 5.3. To be added...

   We have already seen in section 5.1.1 that a negative correlation between input values can result in a variance reduction compared to simple sampling. We can also easily see that there is a negative correlation between the random values in a dagger cycle, because the diverging unit can never appear more than once in a dagger cycle. This means that if we know that the j:th value was equal to the diverging unit then we know that all the other $S - 1$ values are equal to the conformist unit; hence, the j:th value and the other values are varying in opposite directions, which is characteristic for a negative correlation.

   We can also verify the negative correlation by going back to the mathematical definition. Let us start by investigating the product of two values in a dagger cycle, i.e., $Y_j Y_k$. There are only two possible values of this product, *aa* or *ab*, since the diverging unit cannot appear more than once in the dagger cycle. We can also observe that for all dagger transforms there will only be two intervals there either the j:th or the k:th value of dagger cycle are equal to the diverging unit b; hence, the probability for this result is $2p$. This means that the expectation value of the product is given by

$$E[Y_j Y_k] = 2pab + (1 - 2p)aa. \tag{5.8}$$

Moreover, we can compute the expectation value of each value as

$$E[Y_j] = E[Y_k] = (1 - p)a + pb. \tag{5.9}$$

The covariance between Y_j and Y_k can now be computed according to definition 2.10:

$$Cov[Y_j, Y_k] = E[Y_j Y_k] - E[Y_j]E[Y_k] = pab + (1 - 2p)aa - ((1 - p)a + pb)^2 = -p^2(a + b)^2 < 0. \tag{5.10}$$

   To be added: Example with figure...

5.2.2 Application

To be added: Introduction...

Multiple Inputs

To be added: Discussion on how to manage inputs with different dagger cycle lengths...

Simulation Procedure

To be added: Overview, equations, block diagram and example...

5.3 Control Variates

To be added: Introduction...

5.3.1 Principle

5.3.2 Application

To be added: Introduction...

Finding a Simplified Model

To be added: Discussion on how to create suitable simplified models...

Simulation Procedure

To be added: Impact on coefficient of variation depending on how the control variate method is implemented...

To be added: Overview, equations, block diagram and example...

5.4 Correlated Sampling

To be added: Introduction...

5.4.1 Principle

5.4.2 Application

To be added: Introduction...
To be added: Overview, equations, block diagram and example...

5.5 Importance Sampling

To be added: Introduction...

5.5.1 Principle

5.5.2 Application

To be added: Introduction...

Multiple Inputs

To be added: Discussion on how to manage systems with multiple inputs...
   To be added: Example showing the computation of the weight factor for a simulationn with multiple inputs...

Multiple Outputs

To be added: Discussion on how to manage systems with multiple outputs...

Finding an Importance Sampling Function

To be added: Discussion on how to choose the importance sampling function using a simplified model...

Systematical Errors

To be added: Discussion on how inappropriate importance sampling functions can introduce a systematical error, and how this sometimes can be acceptable...

Simulation Procedure

To be added: Overview, equations, block diagram and example...

5.6 Stratified Sampling

To be added: Introduction...

5.6.1 Principle

To be added: Description...

   Example 5.4 (statistical properties of the strata). Consider a six-sided dice, where two

sides are labelled "0" and the remaining four sides are labelled "1". Compare the expectation value and variance of the entire population compared to the properties of two strata (one for the two units with the value 0, and another stratum for the remaining units).

Solution: The population in this example can be described as $X = \{0, 0, 1, 1, 1, 1\}$, i.e., one unit per side of the dice. However, in order to reduce the length of the computations, we can duplicates from the population, i.e., let us consider $X = \{0, 1, 1\}$. The two strata are then defined by the subsets $X_1 = \{0\}$ and $X_2 = \{1, 1\}$. The expectation values and variances can be computed directly according to the definitions:

$E[X] = (0 + 1 + 1)/3 = 2/3,$

$Var[X] = ((0 - 2/3)^2 + (1 - 2/3)^2 + (1 - 2/3)^2)/3 = 2/9,$

$E[X_1] = 0,$

$Var[X_1] = 0,$

$E[X_2] = (1 + 1)/2 = 1,$

$Var[X_2] = 0.$

5.6.2 Application

To be added: Introduction...

Sample Distribution

To be added: Discussion on how to distribute samples between strata...

The Cardinal Error

To be added: Discussion on how the practical application of the Neyman allocation may introduce a systematical error when sampling duogeneous populations...

The Cum $\sqrt{f}$ Rule

To be added: Discussion on how to design strata using a simplified model...

The Strata Tree

To be added: Discussion on how to design strata using classification of input scenarios...

Simulation Procedure

To be added: Overview, equations (including random number generation), block diagram and example...

EFFICIENT MONTE CARLO SIMULATIONS

The previous chapters have presented the mathematics of Monte Carlo simulation as well as some practical solutions to implement different methods when simulating a technical system. In this concluding chapter, all those pieces are brought together in a discussion on how to design an efficient Monte Carlo simulation.

6.1 Mathematical Model

To be added: Discussion on important steps when formulating the mathematical model...

6.2 Choice of Simulation Method

To be added: Summary of the information necessary to efficiently apply different variance reduction technique.

To be added: Discussion on how variance reduction techniques can be combined.

6.3 Testing

To be added: Discussion on how to test and verify the results of a Monte Carlo simulation...

PROBABILITY DISTRIBUTIONS

This appendix provides an overview of some important probability distributions.
  To be added: Reference to other sources.

SOLUTIONS TO EXERCISES

2.1

a) All values between 10 and 20 are equally probable; thus, this is a heterogeneous population.

b) The values 10, 11, …, 20 are equally probable; thus, this is a heterogeneous population.

c) Only the values 10 and 20 are possible; however, the value 10 has a probability of 60%. Hence, this random variable corresponds to a population where 60% of the units have the value 10 and the remainder of the units correspond to the value 20. Hence, this is a duogeneous population, where the units with the value 10 are the conformist units and the other units are the diverging units.

d) All values between 10 and 20 are possible; however, the value 10 has a probability of 60%. Hence, this random variable corresponds to a population where 60% of the units have the value 10 and the remainder of the units correspond to all values larger than 10 and smaller or equal to 20. Hence, this is a duogeneous population, where the units with the value 10 are the conformist units and the other units are the diverging units.

e) All values between 10 and 20 are possible and although values between 10 and 12 are more probable there is no particular value that is dominating, which means that we cannot identify a set of conformist units. Hence, this is a heterogeneous population.