# THE ART OF SIMULATION

K. D. TOCHER

# THE ART OF SIMULATION

K. D. TOCHER
B.Sc., Ph.D., D.Sc.

THE ENGLISH UNIVERSITIES PRESS LTD
102 NEWGATE STREET
LONDON · EC1

First Printed 1963



PRINTED AND BOUND IN ENGLAND
FOR THE ENGLISH UNIVERSITIES PRESS LTD
BY HAZELL WATSON AND VINEY LTD, AYLESBURY, BUCKS

# THE ART OF SIMULATION

1923

ELECTRICAL ENGINEERING SERIES

General Editor

ARTHUR PORTER,

PH.D., M.I.E.E.

*Professor of Industrial Engineering, University of Toronto;*
*Formerly Dean of Engineering,*
*University of Saskatchewan, Canada;*
*Professor of Light Electrical Engineering,*
*Imperial College of Science and Technology, London*

---

ANALYTICAL TECHNIQUES
FOR NON-LINEAR CONTROL SYSTEMS

J. C. WEST,

D.SC., A.M.I.E.E., A.M.A.I.E.E.

AN INTRODUCTION TO
LINEAR NETWORK ANALYSIS

P. S. FARAGO,

DR.PHIL., F.INST.P., F.R.S.E.

HIGH FREQUENCY APPLICATION OF FERRITES

J. ROBERTS,

B.SC. (ENG.), A.C.G.I., A.M.I.E.E.

EXAMPLES IN
ADVANCED ELECTRICAL ENGINEERING

J. F. W. BELL,

B.SC., PH.D., A.M.I.E.E.

# GENERAL EDITOR'S FOREWORD

by

ARTHUR PORTER, PH.D., M.I.E.E.

*Professor of Industrial Engineering, University of Toronto;*
*Formerly Dean of Engineering, University of Saskatchewan, Canada;*
*Professor of Light Electrical Engineering,*
*Imperial College of Science and Technology, London*

This volume is one of a series of texts in electrical engineering, and related subjects, the main object of which is to present advances in these subjects on the one hand, and to re-evaluate basic principles, in the light of recent developments, on the other. The books will be of special interest to research workers, postgraduates, and advanced undergraduates in electrical science and applied physics.

The publication of such a series is regarded as timely because of the increasing realisation that the average three or four years' engineering degree course is inadequate for the basic training of research engineers. To correct this situation there has been a notable increase, during the past few years, of postgraduate courses in all branches of engineering in the Universities and Colleges of Advanced Technology of the United Kingdom, the British Commonwealth, and the United States. Strong official support for these courses is being given by the Department of Scientific and Industrial Research, London, and their importance cannot be over-emphasised. But such specialised courses require suitable texts and many of them are seriously handicapped in this respect. This is not surprising in view of the proliferation of science and technology and the recognition that new languages—one of which is the language of high-speed digital computers—have been born within the last decade. It is essential, moreover, that the authors of texts at advanced levels should be actively engaged in research work because only the active participant working at the frontiers of his subject is in a position to assess the importance of new concepts and of the mathematical and experimental techniques involved. I believe that this new series of electrical engineering texts will fulfil a real need and that it will be welcomed as a medium for re-evaluating fundamental engineering principles in the light of recent advances with special emphasis on the needs of postgraduate students.

The present volume deals with the study of industrial operations and processes using large-scale data-processing and computer systems as simulators. It is perhaps the first book to be published on the subject. During recent years considerable advances have been made in the development of models of technological and sociological systems, especially those embodying information feedback, but a major difficulty has been the handling of the extensive computational work. Dr. Tocher has made notable contributions to both statistical mathematics, and to the art of simulation, especially in areas involving a deep understanding of probability theory. This book will be particularly welcomed by operational research workers and by research engineers interested in the optimisation of processes. The methods of operational research, especially those involving the application of digital computers in simulation roles, are proving of considerable value in the application of the scientific method to industrial and applied scientific problems. This book provides an authoritative introduction to these topics.

# CONTENTS

CHAPTER 1

# INTRODUCTION

The subject-matter of this book stems from three origins and contains matters of interest to three groups of scientific workers.

The first and most respectable origin lies in the theory of mathematical statistics. Before and for some time after the founding of the Royal Statistical Society in 1834, the subject of statistics consisted of the collection and display in numerical and graphical form of facts and figures from the fields of economics, actuarial science and allied descriptive sciences. One of the most useful forms of display was the histogram or frequency chart and the transformation of statistics began when it was realised that the occurrence of such diagrams could be explained by invoking the theory of probability.

The idea of a probability distribution had been established as a useful concept by mathematicians and has been studied by Laplace and Gauss, to name two of the most celebrated mathematicians interested. The idea that frequency charts could be explained as a practical consequence of the laws of probability applied to everyday matters seized the imagination of the pioneers of mathematical statistics. Since a probability distribution is by its nature, in most instances, composed of an infinite number of items, and frequency charts by their nature are composed of a finite number of items, these latter had to be thought of as *samples* from an underlying theoretical probability distribution. The problem then raised itself at how to describe a probability distribution given only a sample from it. The mathematical difficulties of this seemed immense and such steps as were taken needed experimental verification to give the early workers confidence. Thus was born the sampling experiment. A close approximation to a probability distribution was created, samples were taken, combined and transformed in suitable ways and the resulting frequency chart of sampled values compared with the predictions of theory.

Although mathematical techniques have developed to levels of sophistication that would astonish the early workers, the value of sampling experiments in mathematical statistics still remains. The advent of automatic digital computers to perform the laborious calculations necessary has revitalised this as a possible approach to the solution of problems still beyond the reach of analysis.

The second origin lay in the demands of applied mathematicians for methods of solving problems involving partial differential equations. These equations constantly arise in the mathematical formulation of numerous physical processes. Analytical solutions were found for a wide range of formal problems but practical problems involving complex or irregular boundary conditions could not yield to theoretical attack. The corresponding situations had arisen in ordinary differential equations and had been overcome by developing numerical techniques for solving such equations. The nature of the problem for partial differential equations made an attack on these lines far more difficult.

A typical problem was the solution of the so-called diffusion equation, which arises, as its name implies, in the diffusion of gases as well as in the conduction of heat in a medium and many other physical systems. The characteristic of many of these systems was that the actual mechanism for the movement of the gas (or heat) involved a large number of particles behaving in a partly regular and partly irregular manner. Averaging over the particles enabled the random element to be eliminated and a deterministic description to be given.

Now the theory of probability had studied formal models closely allied to a system of particles moving in this partly regular, partly random manner and had developed mathematical techniques for dealing with these problems which were studied under the name of random walks. The mathematical analysis of these problems gave rise to partial differential equations of the same type as the diffusion problem. Thus was born the idea of solving experimentally the diffusion and allied equations by random walks. The idea lay dormant for many years and was resurrected by Von Neumann and Ulam under the stress of the technological demands of the Second World War.

These men conceived of the extension of the principle to seek solutions to difficult mathematical problems arising from deterministic problems by finding analogous problems in probability leading in their analysis to formally identical mathematical equations and then solving the probability problems practically by sampling experiments. By this time, the ultimate stochastic nature of the physical phenomena often present in the original problem was often forgotten or completely ignored.

One of the simplest and most powerful applications of this idea, which they christened 'Monte Carlo', was the evaluation of a multi-dimensional integral. Consider the simplest case of evaluating the area of a bounded region. Surround the region with a square, scaling to make its side of unit length. Take a point in the area at random. The probability that this lies in the region of area $A$ is simply $A$. Take a large number of points at random; the proportion of these lying in the area is an estimate of $A$.

The whole idea can be generalised to higher dimensions and the result remains true. The great advantage of this method lies in the fact that for higher dimensions, conventional numerical techniques will require an enormous number of points to get any answer at all. The Monte Carlo method can get an answer with any number of points although of course the accuracy falls off as the number decreases.

In fact, the accuracy is proportional to $\sqrt{n}$, $n$ being the number of points. Conventional methods give an accuracy proportional to $n^{1/d}$ where $d$ is the number of dimensions.

The third origin lay in the new science of Operational Research. This set itself the tasks of applying scientific method in everyday life—to military problems during the last war and to problems of industry after it. In common with most sciences, it developed by building models of the systems it studied and using these models to give insight and sometimes quantitative information about these systems.

The outstanding difference between the subject-matter of conventional science and operational research lay in the greater variability of the phenomena studied. It was vital to bring regularity back into the description of these phenomena—to find a usable description of the variability. This was achieved, as had been done for economics and actuarial science in the previous century, by using probability theory.

The models studied were primarily probabilistic and the theoretical difficulties of these models were immense. A special branch of probability theory arose under the title of Queue Theory to deal with some of these problems, but the irregular nature of the boundary conditions and restrictions in the real problems could rarely be incorporated into the models that mathematical methods could solve.

Once again the scientist turned to an experimental technique. Now the problem was phrased in a language description of the real world involving queues, stocks, machines and operating instructions. The subject of simulation, as it was called, took on an identity of its own.

Yet this was nothing new. To a statistician the problem was nothing more than to find the sampling distribution of an intricately and irregularly defined statistic and because of the intricate nature of the definition, to do this by a sampling procedure.

The common element in all these approaches to our subject is probability theory. In these circumstances, it is hardly necessary to stress that an understanding of at least the elements of the theory are essential to an understanding of the principles and practice of sampling.

Any attempt to give a self-contained account of the theory necessary as part of the book would be self-defeating. It could be nothing more than a few formal definitions and readers ignorant of the subject might be tempted to read further with only such a brief introduction. Therefore it is assumed that the reader has the necessary grounding in elementary probability theory.

Likewise, the elements of the theory of statistics are assumed. The commonest distributions such as Normal, Student's, $\chi^2$, binomial, exponential, are all discussed without a long introduction and the $\chi^2$-test of goodness of fit is not explained, as this must surely be well known to all scientific workers.

Since the basic process involved is the drawing of samples from different distributions, this receives first study. The problem is reduced to drawing a sample from a uniform distribution (or drawing an integer from a range) 'at random'.

The methods of solving *this* problem by both machinery and mathematical methods are discussed.

This is followed by an account of the problems in determining a frequency distribution by sampling.

The possibility of using sampling methods as *practical* methods of acquiring knowledge rests on the possibility of taking really large samples to reduce the effect of the inherent variability. This in turn is possible only because of the existence of automatic digital computers.

It is assumed that one of these powerful devices is available and this raises the problem of describing how to use these machines for sampling. Again, this book is not the place to attempt to give details of the problems of instructing any particular machine to perform the required calculation.

However, the idea of a flow diagram as a means of describing the calculating procedure in a form suitable for transcription into a machine code is explained and forms a powerful means of reducing the volume of description of computing procedures.

Probability distributions are described by means of a set of parameters and the central problem of sampling is to specify the value of such parameters given a sample from the distribution. This is the problem of estimation, which receives full treatment. An estimate must be as accurate as possible, and methods of sampling and estimation that increase the accuracy of the estimate without a corresponding increase in the volume of computing are described.

We are then ready to tackle the problems of simulation. First simple queues are dealt with, gradually increasing in difficulty until a general case is dealt with.

More complex systems are then treated and a general technique evolved for dealing with any system.

Finally the question of the design of such large-scale experiments is considered.

CHAPTER 2

# SAMPLING FROM A DISTRIBUTION

## 2.1 Introduction

We have seen in the preceding chapter that the basic problem for simulation consists in sampling from statistical distributions, and this chapter will describe a variety of methods of performing this basic operation.

There are two main types of distributions from which samples are required: those in which the statistical variable takes a continuous range of values giving rise to a continuous probability density function, and those in which the statistical variable can take a discrete number of values. The normal distribution is typical of the first class and the binomial distribution (the number of successes in a fixed number of similar independent trials) is an example of the latter.

However, for practical purposes, we must work with approximations to continuous variables and have to be satisfied with samples from a grouped distribution represented by a histogram. If the different groups are given names, then the problem reduces to selecting a name for the appropriate frequency and consequently there is no distinction in practice between the two types of distribution.

Although it is possible to discuss the meaning of probability at great length, the earlier remarks indicate that our purposes are met by a frequency concept. Thus our problem is to devise a process of selection from the distribution so that the results of the repetition of this process will give rise to a frequency distribution of sampled values that matches the frequency distribution required in our simulation. It is possible for there to be quite intricate probability rules relating successive sample values, but the most important case that will be considered in detail first is where these samples are independent of one another in the sense that the probability distribution for any given item from the sample is independent of the preceding sample values.

To effect the random selection, some source of randomness is required, and to give a general solution to the problem we want to use a common source for all types of distribution. The most appropriate source of such randomness is from random numbers. These may be supplied either by tables, specially constructed by means of randomising machines, or by the use of mathematical devices to generate long sequences of numbers that have similar properties to random numbers, the so-called pseudo-random numbers. These techniques will each be pursued in later chapters.

We consider first the elementary methods of selecting from a frequency distribution and show in a series of steps how this can be improved and mechanised.

## 2.2 The 'Top Hat' Method

The obvious method to obtain a sample from a frequency distribution has often been used in early historical examples of simulation. It consists of representing each of the possible values of the statistical variable by means of a set of discs. The number of discs given any particular value is proportional to the frequency with which it occurs in the frequency distribution. The total number of discs used is limited in an upward direction by the problems of the physical space occupied by the discs and in a lower direction by the accuracy with which the frequencies must be approximated. The collection of discs is well shuffled in a hat or other receptacle and a disc drawn at random. The values on successive discs after drawing (each being replaced after it has been read) constitute the sample.

As an example, suppose we wish to draw a sample from a normal distribution with mean 18 units and standard deviation 4 units. The first step is to decide on the width of grouping that will be tolerable. This is settled by the application and we suppose a class width of 2 units is chosen, and the class boundaries are taken symmetrically about the mean.

The next step is to determine the probability of this random variable falling in the intervals $-1$ to 1, 1 to 3, 3 to 5, . . ., 15 to 17, . . . and these are obtained from a table of the normal distribution function. This is given in Table 1.

If we decide to represent the frequencies to three decimal places, we will require 1,000 discs. The frequency of negative values is negligible. Thus we number 28 discs with the number 10 representing the class interval from 9–11, 2 with 6 representing the interval 5–7, etc. In practice, the rounding has to be adjusted to give the sum of the frequencies exactly 1,000 or a few discs rejected. From the table given one disc is rejected.

This example raises a problem of dealing with the tails of distribu-

TABLE 1

| *Class* | *Frequency* | *No. of discs (total 1,000)* |
|---|---|---|
| −1 | 0·0000 | 0 |
| 1– 3 | 0·0001 | 0 |
| 3– 5 | 0·0005 | 1 |
| 5– 7 | 0·0024 | 2 |
| 7– 9 | 0·0092 | 9 |
| 9–11 | 0·0279 | 28 |
| 11–13 | 0·0655 | 65 |
| 13–15 | 0·1210 | 121 |
| 15–17 | 0·1747 | 175 |
| 17–19 | 0·1974 | 197 |
| 19–21 | 0·1747 | 175 |
| 21–23 | 0·1210 | 121 |
| 23–25 | 0·0655 | 65 |
| 25–27 | 0·0279 | 28 |
| 27–29 | 0·0092 | 9 |
| 29–31 | 0·0024 | 2 |
| 31–33 | 0·0005 | 1 |
| 33–35 | 0·0001 | 0 |
| ≧35 | 0·0000 | 0 |
| | | 999 |

tions which are unbounded. In order to represent such distributions correctly, we need an infinite number of discs so that the extreme tails with their correspondingly extreme small probabilities can be represented at all. This lays a practical limit on the length of tail approximated.

However, this difficulty can be overcome in the following way. Let the top group of the distribution represent the frequency with which an item should lie beyond the last bounded group. If a disc representing this unbounded group is obtained, then a further sample is taken from a distribution that gives the proportion of the items having a given value conditional upon the item being in the upper tail. Thus to accommodate the possibility of sample values in the range greater than 33 the following Table 2 of relative frequencies is constructed from more accurate tables of the normal integral.

The primary labelling of discs is adjusted to allow for the extra unbounded class, and a secondary labelling made according to the second distribution. If the first sampling gives rise to an item belonging to the upper tail, another disc is selected at random and the second value