

ANL/NDM--92

DE89 008789

NUCLEAR DATA UNCERTAINTIES - I*

BASIC CONCEPTS OF PROBABILITY

by

Donald L. Smith

December 1988

PROBABILITY THEORY. Events and event spaces. Borel fields. Boolean algebra. Random variables. Intuitive probability. Axiomatic probability. Calculus of probability. Conditional probability and independence. Probability distributions. Binomial and multinomial probability. Poisson probability. Normal probability. Bayes' theorem and applications.

Engineering Physics Division
ARGONNE NATIONAL LABORATORY
9700 South Cass Avenue
Argonne, Illinois 60439
U.S.A.

*This work supported by the U.S. Department of Energy, Energy Research Programs, under contract W-31-109-Eng-38.

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

NUCLEAR DATA AND MEASUREMENTS SERIES

The Nuclear Data and Measurements Series presents results of studies in the field of microscopic nuclear data. The primary objective is the dissemination of information in the comprehensive form required for nuclear technology applications. This Series is devoted to: a) measured microscopic nuclear parameters, b) experimental techniques and facilities employed in measurements, c) the analysis, correlation and interpretation of nuclear data, and d) the evaluation of nuclear data. Contributions to this Series are reviewed to assure technical competence and, unless otherwise stated, the contents can be formally referenced. This Series does not supplant formal journal publication but it does provide the more extensive information required for technological applications (e.g., tabulated numerical data) in a timely manner.

Copies on microfiche can be obtained by contacting:

NTIS Energy Distribution Center
P.O. Box 1300
Oak Ridge, TN 37831
U.S.A.
Comm. Phone: (615) 576-1301
FTS: 626-1301

INFORMATION ABOUT OTHER ISSUES IN THE ANL/NDM SERIES

A list of titles and authors for reports ANL/NDM-1 through ANL/NDM-50 can be obtained by referring to any report of this Series numbered ANL/NDM-51 through ANL/NDM-76, while those for ANL/NDM-51 through ANL/NDM-76 are listed in any report numbered ANL/NDM-77 through ANL/NDM-101. Requests for a complete list of titles or for copies of previous reports should be directed to:

Section Secretary
Applied Nuclear Physics Section
Engineering Physics Division
Building 316
Argonne National Laboratory
9700 South Cass Avenue
Argonne, Illinois 60439
U.S.A.

ANL/NDM-76 Alan B. Smith and Peter T. Guenther, *Scattering of Fast Neutrons from Elemental Molybdenum*, November 1982.

ANL/NDM-77 Donald L. Smith, *A Least-squares Method for Deriving Reaction Differential Cross Section Information from Measurements Performed in Diverse Neutron Fields*, November 1982.

ANL/NDM-78 A.B. Smith, P.T. Guenther and J.F. Whalen, *Fast-neutron Total and Elastic-scattering Cross Sections of Elemental Indium*, November 1982.

ANL/NDM-79 C. Budtz-Joergensen, P. Guenther, A. Smith and J. Whalen, *Few-MeV Neutrons Incident on Yttrium*, June 1983.

ANL/NDM-80 W.P. Poenitz and J.F. Whalen, *Neutron Total Cross Section Measurements in the Energy Region from 47 keV to 20 MeV*, July 1983.

ANL/NDM-81 D.L. Smith and P.T. Guenther, *Covariances for Neutron Cross Sections Calculated Using a Regional Model Based on Elemental-model Fits to Experimental Data*, November 1983.

ANL/NDM-82 D.L. Smith, *Reaction Differential Cross Sections from the Least-squares Unfolding of Ratio Data Measured in Diverse Neutrons Fields*, January 1984.

ANL/NDM-83 J.W. Meadows, *The Fission Cross Sections of Some Thorium, Uranium, Neptunium and Plutonium Isotopes Relative to ^{235}U* , October 1983.

ANL/NDM-84 W.P. Poenitz and J.W. Meadows, *^{235}U and ^{239}Pu Sample-mass Determinations and Intercomparisons*, November 1983.

ANL/NDM-85 D.L. Smith, J.W. Meadows and I. Kanno, *Measurement of the $^{51}\text{V}(n,p)^{51}\text{Ti}$ Reaction Cross Section from Threshold to 9.3 MeV*, June 1984.

ANL/NDM-86 I. Kanno, J.W. Meadows and D.L. Smith, *Energy-differential Cross-section Measurement for the $^{51}\text{V}(n,\alpha)^{48}\text{Sc}$ Reaction*, July 1984.

ANL/NDM-87 D.L. Smith, J.W. Meadows, M.M. Bretscher and S.A. Cox, *Cross-section Measurement for the $^7\text{Li}(n,n't)^4\text{He}$ Reaction at 14.74 MeV*, September 1984.

ANL/NDM-88 A.B. Smith, D.L. Smith and R.J. Howerton, *An Evaluated Nuclear Data File for Niobium*, March 1985.

ANL/NDM-89 Bernard P. Evain, Donald L. Smith and Paul Lucchese, *Compilation and Evaluation of 14-MeV Neutron-activation Cross Sections for Nuclear Technology Applications: Set I*, April 1985.

ANL/NDM-90 D.L. Smith, J.W. Meadows and P.T. Guenther, *Fast-neutron-spectrum Measurements for the Thick-target $^9\text{Be}(d,n)^{10}\text{B}$ Reaction at $E_d = 7$ MeV*, April 1985.

ANL/NDM-91 A.B. Smith, P.T. Guenther and R.D. Lawson, *On the Energy Dependence of the Optical Model of Neutron Scattering from Niobium*, May 1985.

ANL/NDM-92 Donald L. Smith, *Nuclear Data Uncertainties - I: Basic Concepts of Probability*, December 1988. [Present report--see notice at the end of this list.]

ANL/NDM-93 D.L. Smith, J.W. Meadows and M.M. Bretscher, *Integral Cross-section Measurements for $^7\text{Li}(n,n't)^4\text{He}$, $^{27}\text{Al}(n,p)^{27}\text{Mg}$, $^{27}\text{Al}(n,\alpha)^{24}\text{Na}$, $^{58}\text{Ni}(n,p)^{58}\text{Co}$ and $^{60}\text{Ni}(n,p)^{60}\text{Co}$ Relative to ^{238}U Neutron Fission in the Thick-target $^9\text{Be}(d,n)^{10}\text{B}$ Spectrum at $E_d = 7$ MeV*, October 1985.

ANL/NDM-94 A.B. Smith, D.L. Smith, P. Rousset, R.D. Lawson and R.J. Howerton, *Evaluated Neutronic Data File for Yttrium*, January 1986.

ANL/NDM-95 Donald L. Smith and James W. Meadows, *A Facility for High-intensity Neutron Irradiations Using Thick-target Sources at the Argonne Fast-neutron Generator*, May 1986.

ANL/NDM-96 M. Sugimoto, A.B. Smith and P.T. Guenther, *Ratio of the Prompt-fission-neutron Spectrum of Plutonium-239 to that of Uranium-235*, September 1986.

ANL/NDM-97 J.W. Meadows, *The Fission Cross Sections of ^{230}Th , ^{232}Th , ^{233}U , ^{234}U , ^{236}U , ^{238}U , ^{237}Np , ^{239}Pu and ^{242}Pu Relative ^{235}U at 14.74 MeV Neutron Energy*, December 1986.

ANL/NDM-98 J.W. Meadows, *The Fission Cross Section Ratios and Error Analysis for Ten Thorium, Uranium, Neptunium and Plutonium Isotopes at 14.74-MeV Neutron Energy*, March 1987.

ANL/NDM-99 Donald L. Smith, *Some Comments on the Effects of Long-range Correlations in Covariance Matrices for Nuclear Data*, March 1987.

ANL/NDM-100 A.B. Smith, P.T. Guenther and R.D. Lawson, *The Energy Dependence of the Optical-model Potential for Fast-neutron Scattering from Bismuth*, May 1987.

ANL/NDM-101 A.B. Smith, P.T. Guenther, J.F. Whalen and R.D. Lawson, *Cobalt: Fast Neutrons and Physical Models*, July 1987.

ANL/NDM-102 D. L. Smith, *Investigation of the Influence of the Neutron Spectrum in Determinations of Integral Neutron Cross-Section Ratios*, November 1987.

ANL/NDM-103 A.B. Smith, P.T. Guenther and B. Micklich, *Spectrum of Neutrons Emitted From a Thick Beryllium Target Bombarded With 7 MeV Deuterons*, January 1988.

ANL/NDM-104 L.P. Geraldo and D.L. Smith, *Some Thoughts on Positive Definiteness in the Consideration of Nuclear Data Covariance Matrices*, January 1988.

ANL/NDM-105 A.B. Smith, D.L. Smith, P.T. Guenther, J.W. Meadows, R.D. Lawson, R.J. Howerton, and T. Djemil, *Neutronic Evaluated Nuclear-Data File for Vanadium*, May 1988.

ANL/NDM-106 A.B. Smith, P.T. Guenther, and R.D. Lawson, *Fast-Neutron Elastic Scattering from Elemental Vanadium*, March 1988.

NOTICE: The work for the present report was completed in April 1985. The documentation was written and edited over the intervening period and is proffered herewith. This report is the first in a series on the subject: "Nuclear Data Uncertainties."

TABLE OF CONTENTS

	<u>Page</u>
PREFACE	ix
ABSTRACT.	1
1. INTRODUCTION.	2
1.1 Resource Material	2
1.2 Procedure	2
1.3 Content.	3
2. EVENTS AND EVENT SPACES	5
2.1 Intuitive Approach	5
2.2 Theoretical Approach	12
2.3 Event Counting	29
3. RANDOM VARIABLES AND THEIR FUNCTIONS.	43
3.1 Concept of Randomness	43
3.2 Random Variables.	45
3.3 Random-Variable Functions.	50
4. BASIC CONCEPTS OF PROBABILITY	57
4.1 Intuitive Approach	57
4.2 Axiomatic Approach	65
4.3 Probability Calculation	72
5. CONDITIONAL PROBABILITY AND INDEPENDENCE	78
5.1 Intuitive Approach	78
5.2 Theoretical Approach	84
6. PROBABILITY DISTRIBUTIONS	94
7. SOME IMPORTANT PROBABILITY LAWS GOVERNING THE FREQUENCY OF OCCURRENCE OF RANDOM EVENTS.	105
7.1 Binomial and Multinomial Probability Laws	105
7.2 Poisson Probability Law	120
7.3 Normal (Gaussian) Probability Law	122
8. BAYES' THEOREM AND APPLICATIONS	155
ACKNOWLEDGEMENTS	181

TABLE OF CONTENTS (cont'd)

	<u>Page</u>
APPENDIX I: Notation	183
APPENDIX II: Important Formulas	188
APPENDIX III: A Guide to Additional Reading for Selected Topics.	191
BIBLIOGRAPHY	197
INDEX.	201

PREFACE

Nuclear data research is conducted for the purpose of producing the nuclear information which is needed for those technologies that are based upon fundamental nuclear processes. Examples of these technologies are the production of energy via fission and fusion reactors. This field of endeavor has developed because the needs are of a quantitative nature, and meeting them often requires investigative effort which differs considerably in style from that expended in the conduct of basic nuclear research.

Since there is a strong emphasis on quantitative matters, it follows that investigators in this field should be very concerned with the accuracy of their work. Indeed, workers in this field have been quite aware of the need for good accuracy from the beginning, but it is rather surprising that there has been very little disciplined effort directed toward its proper assessment until recent times. About a decade or so ago, nuclear data evaluators began to explicitly focus on this issue, and steps were taken to define and deal with the problems. Evaluators began by addressing the matter of improving those methods which they use in data evaluation. Soon, however, it became clear that this effort would be of limited value unless data measurers improved the methods they employ to analyze their data and report their results, especially with regard to the matter of errors. Today, most experimentalists in the field are aware of this concern, and many appear to be making serious attempts to deal with it.

There has been considerable confusion regarding procedures for estimating nuclear data errors, and for including them in data analysis and evaluation manipulations. This had led to disagreements and even to some obvious mistakes. In fairness to workers in the field, it is not a trivial issue. Nuclear data studies are very involved, and the results reported are generally abstract quantities whose derivation from measured data is complex and indirect. Part of the blame does rest, however, with the nuclear data researchers themselves. The assessment of accuracy in research endeavors requires that certain rigorous mathematical methods be used, and many investigators do not seem to appreciate this to the extent that they should.

Similar methods have been used in several other fields. An example of one such field is medical research, especially with regard to the development and testing of new drugs and medical procedures. The possibilities for adverse social impact in this area can be very intimidating. A situation where this was clearly demonstrated is the well-known case of the drug Thalidomide which was found several years ago to be linked to the occurrence of a number of very serious birth defects. Current standards of medical research effectively dictate that as much effort be devoted to assessing the reliability and potential side effects of a treatment or medication as normally is involved in its discovery and development.

The accident at the Three-Mile-Island nuclear electric generating station in the U.S.A. apparently has produced more significant adverse economic consequences than actual social consequences; nevertheless, the general public continues to regard this event more from the point of view of its social impact. The subsequent, more serious accident at the Chernobyl plant in the Soviet Union has apparently produced significant consequences for the local populations and it has solidified world-wide concern over the matter of nuclear reactor safety. The nuclear industry will not be permitted to proceed in the development of this energy source unless it adheres to the highest conceivable standards of quality and safety, and can convince the public that it is doing so. Nuclear data researchers must recognize and heed this unmistakable message from society, and therefore willingly accept direct social responsibility for the quality of their work. There is an indisputable link between the quality of nuclear data and the safety and economic viability of nuclear energy sources.

Nuclear data researchers who are motivated to learn about the mathematical tools needed to deal with this subject can, in principle, garner what they need from the literature. Several documents have been prepared to help busy workers gain some proficiency in this area (e.g., Smi 81 and Man 81). These pragmatic references are very useful, but since they were produced as explicit "how-to-do-it" guides, they do not alone offer potential readers the possibility for acquiring much depth of understanding of the subject. I do not believe that researchers in this field need to pursue all the details of mathematical rigor which might be appropriate for an advanced academic course, but they ought to acquire a better understanding of the principles which form a basis for the methods which they will use than can be gained from reading only "cookbook" treatments. It would certainly be desirable if investigators in the field of nuclear data could achieve an adequate perspective of this subject through the study of material which had been prepared with their interests in mind. Unfortunately, there does not seem to be a great deal of suitable material available in the literature. It is my intent to address this deficiency by means of this report, and others to follow in this series.

I believe that the subject of errors (or uncertainties, as many individuals prefer to call them) cannot be addressed to the exclusion of related technical topics. The terms "probability," "statistics," and "error" appear together in a variety of contexts. The distinctions between them are apparently not widely understood by nuclear data researchers. A major objective of this project is to try to clarify these matters for the benefit of this community of investigators. I think that it is worthwhile to make a qualitative attempt in this regard at an early stage in the development of this topic so that the reader will understand the need to investigate certain topics which he might otherwise believe are irrelevant to his main interests. It is known that measurements do not lead to unique outcomes. Repeated attempts at the same experiment tend to produce varying results. The outcome of any experiment is, in the final analysis, governed to a large extent by probability. Error is the term which designates the dispersion or spread of possible results one can expect to observe from experimentation. Probability cannot be measured directly, so one is forced to estimate the essential features of probability distributions from the

analysis of accumulated observable results. This activity constitutes what is referred to as statistics. There are benefits to be gained from studying these concepts in a unified fashion. Recipes for performing certain routine analyses are seen to be related, in terms of the underlying theory. It is then much easier to remember them. It is suggested, therefore, that the interested reader should invest some effort and time learning the foundations of the subject in order to save time and avoid confusion later on when called upon to apply this knowledge under a variety of practical circumstances.

I have chosen to proceed in gradual stages toward my objective of providing a broad exposition of this subject. The present report constitutes the first step. I begin by discussing various aspects of probability theory, since it is basic to all other areas. My treatment of this subject is not intended to be a comprehensive treatise or review. It covers only what appear to me to be the more important concepts, presented from a point of view which should be familiar to nuclear data researchers. I make no claim of originality for the included material, but I have interpreted the well-known concepts as I see fit, consistent with the needs and interests of the community to which this work is dedicated. Probability theory draws upon material from a number of other subdisciplines of mathematics. In many instances, I have tacitly assumed that the reader will be familiar with certain facts and concepts, so they go unmentioned. In other instances, I allude to requisite or supportive material, but I have generally tried to avoid delving explicitly into sticky details in order to ensure that the reader can more easily follow the essential ideas of the present topic without becoming distracted along the way.

The approach utilized in this report is explicitly tutorial, and the style of presentation is informal and narrative by choice. Consequently, this material should be of value to students of applied nuclear science. The treatment is not particularly directed toward individuals who are already experienced in this area. They are likely to be quite familiar with most of the fundamentals addressed here and no doubt will have concerns and specific interests which are generally more advanced and technical than those which are covered in this series. Experimentally inclined investigators will likely find this material of greater relevance to their work than their theoretical colleagues. Nevertheless, there are potentially important theoretical applications for this material which are not as well exploited as they might be. Therefore, I believe that nuclear theorists could profit from a study of this work. Mathematicians are likely to be skeptical of the generally intuitive approach, the incompleteness, and the lack of emphasis on rigor in this treatment. To them I offer my apologies. This work was not conceived to serve that audience.

It is envisaged that this series will ultimately consist of five distinct volumes, each of which will be made available to the nuclear data community as a report in the Argonne National Laboratory ANL/NDM report series when the work is completed. The five general topic areas which are to be addressed in these separate reports are: i) probability theory, ii)

properties of probability functions and the nature of error, iii) statistical theory, iv) the least-squares method and related topics, and v) uncertainty analysis in modern nuclear data applications. Since the intent of this project is to provide a service to the nuclear data community, I would appreciate receiving from the readers any constructive comments, suggestions, criticisms, and corrections of errors which are deemed to be appropriate, in the interest of improving future revisions or reprints of this material which might be forthcoming.

Donald L. Smith
Argonne, Illinois
December 1988

NUCLEAR DATA UNCERTAINTIES - I*

BASIC CONCEPTS OF PROBABILITY

by

Donald L. Smith
Engineering Physics Division
Argonne National Laboratory
9700 South Cass Avenue
Argonne, Illinois 60439
U.S.A.

ABSTRACT

Some basic concepts of probability theory are presented from a nuclear-data perspective, in order to provide a foundation for thorough understanding of the role of uncertainties in nuclear data research. Topics included in this report are: events, event spaces, calculus of events, randomness, random variables, random-variable distributions, intuitive and axiomatic probability, calculus of probability, conditional probability and independence, probability distributions, binomial and multinomial probability, Poisson and interval probability, normal probability, the relationships existing between these probability laws, and Bayes' theorem. This treatment emphasizes the practical application of basic mathematical concepts to nuclear data research, and it includes numerous simple examples.

*This work supported by the U.S. Department of Energy, Energy Research Programs, under contract W-31-109-Eng-38.

1. INTRODUCTION

The motivation for undertaking the preparation of this report, and my approach toward this task, are discussed in the Preface and will not be repeated again. The basic intent of this work is provision of an exposition on certain mathematical concepts which I believe are essential to the understanding of nuclear data errors. This introduction is limited to a discussion of the resource material used to prepare this report, the procedure I have followed in writing the report, and the general content of the report.

1.1 Resource Material

First, I would like to indicate which published sources of information have influenced this work. Since there are so many works available on probability and statistics, my particular selections were, for the most part, arbitrary. I used two particular sources (Bee 58, Par 60) in undergraduate courses I had taken as a student a number of years ago. Four other sources which are available in my personal library (Eva 55, Fis 63, Bev 69, Hil 52) were also found to be useful for this project. The technical library at Argonne National Laboratory provided an additional assortment of sources which have proven to be convenient references (Bur 68, Fel 50, Bas 66, Ash 70, Tuc 67, Zeh 70, Bro 60, Bra 70, Coo 25, Cra 70, Mar 71, Ney 50, Fre 62, Men 67). The reader is referred to the Bibliography for a complete list of these sources. I had no particular reasons for limiting the number of sources considered, other than a shortage of the time required to study them. Since I found a great deal of redundancy in the content of these sources, it seemed that this particular sample of documents amply addressed most of the elementary concepts. This report in no way purports to review the subject. As a rule, the indicated sources contain little, if any, original material. In practice, I have resorted to using them only for background information and not as explicit reservoirs of material for this report. The manner of presentation, and the examples included here, are of my own choosing. Consequently, there are but a few explicit references to these published works in the pages of the present report. It is intended that this report, and later members of this series, serve as reasonably adequate and self-contained treatments of the subject rather than merely as guides to the literature. For the convenience of the reader who wishes to pursue this topic in greater depth, I have suggested a few sources for further study in Appendix III.

1.2 Procedure

I am well aware of the fact that most readers for whom this work is intended desire to learn how to estimate and propagate experimental errors and how to further incorporate error information in evaluations, curve fitting and other operations involving nuclear data. The many useful analytical methods which this mathematical discipline provides are strongly interrelated and, in fact, can be derived from a few rather basic concepts.

The reader who focuses on the methods and fails to recognize and appreciate the unity and simplicity which the underlying theory provides, will do himself a disservice. For this reason, there is a strong emphasis in this work on unifying concepts. However, in recognition of the probable needs of the readers of this work, useful procedures are clearly indicated and illustrated by examples along the way. Important formulas are summarized in Appendix II, and they are cross referenced to the appropriate sections in the text where they are introduced and discussed. To save space, I have avoided numbering many of those expressions or equations to which there are no references elsewhere in the text. As mentioned above, a guide to additional reading on selected topics is provided in Appendix III for the convenience of the reader. Finally, a topical index is provided at the end of the report.

Selection of adequate notation, and its consistent use, is always challenging for any work of a mathematical nature. I have attempted to be reasonably consistent in the use of notation, but there is no expectation that every reader will be satisfied. The notation employed in this report is summarized in Appendix I for easy reference. Important terms are underlined at the point in the text where they are either first used or are formally defined. Elsewhere, they appear without underlining. Numbered equations, figures, tables, and examples are labeled in accordance with the section in which they are introduced and in the order of their appearance therein.

Treatments of this subject which have been prepared by mathematicians---and, surprisingly enough, many of those written by and for physicists---tend to illustrate various statistical concepts by means of examples involving dice, balls and urns, playing cards, and other, similar paraphernalia associated with games of chance. The origins of probability theory derive from attempts by mathematicians to deal with games of chance, and traditions die slowly. Here, I choose to break with this tradition and use examples which, in one way or another, pertain to nuclear research. The examples are generally over-simplified to better illustrate the concepts at hand, but they will hopefully stimulate the reader to extrapolate to more realistic situations of particular interest or concern to his own work.

Assimilation of the material in this report generally requires prior exposure to a number of mathematical procedures which ought to be familiar to students at the undergraduate level. These include algebra, analytical geometry, trigonometry, the calculus, and elementary theory of matrices. In addition, it is assumed that the reader possesses a basic understanding of physics and of nuclear phenomena, including aspects which are commonly encountered in the nuclear data field. Most graduate students and professionals in the nuclear science and nuclear engineering fields will find that they are sufficiently familiar with these topics for present purposes.

1.3 Content

We begin this report by examining events and event spaces. Elementary events and compound events are considered. Event algebra, usually referred to as Boolean algebra, is introduced. Some basic combinatorial theorems and their roles in the calculus of events are discussed. Next, the concept of randomness is introduced, and random variables and their functions are

considered. Probability is defined and interpreted in terms of functional operations on event spaces. Probability theory is approached from both an intuitive and an axiomatic viewpoint, but little attempt is made to trace the history of these ideas or to discuss the various controversies which rage to this day within the community of mathematicians. Some basic combinatorial properties of probability are introduced, and the calculation of probability is considered. Following this is a discussion of other important notions from probability theory, including the concepts of conditional probability and independence. Some features of probability distributions are introduced, including that of multivariate distributions and marginal probability distributions. Next, several important probability distributions that are encountered in the field of nuclear data are considered. These are laws which govern the frequency of occurrence of random events. Emphasis is on binomial and multinomial probability, Poisson and interval probability, and normal probability. Relationships between these individual probability laws are also discussed. Finally, considerable attention is devoted to Bayes' theorem, its interpretation, and related topics, since this theorem, and the methods which it has spawned, are very important in the area of nuclear data evaluation.

This first report does not attempt to address all the important aspects of probability theory that apply to the field of nuclear data. Discussions on many of these, e.g., on the various fundamental properties of probability distribution functions and how they relate to errors and error propagation, are deferred to future reports. It is generally believed that certain matters related to nuclear data uncertainties cannot be addressed entirely within the framework of statistical theory. An example is the difficult question of how to deal with systematic error in a consistent fashion. The reader is encouraged to peruse the very readable and thoughtful essay on this subject by Youdin (You 61). Some advanced methods for dealing with these problems have been proposed (e.g., Per 82), but these ideas are at the frontier of applied mathematical science, and they are difficult to comprehend without a strong foundation in group-theoretical methods. Furthermore, they are somewhat controversial and largely untested. For this reason, I have chosen to embed the present treatment solidly within the familiar, traditional framework of conventional statistics. However, the reader should remain fully cognizant of the fact there are nuclear-data uncertainty problems which appear to fall beyond the scope of elementary random variable theory and, therefore, will not be addressed to any great extent in this report series.

2. EVENTS AND EVENT SPACES

Probability theory is that branch of mathematical science which deals with the properties of a certain class of functions that operate on spaces whose elements are denoted as events. We cannot advance further in our investigation of probability without first having a close look at the meaning of events and event spaces. Some exposure to the rules governing event manipulation is also needed. The reader should realize that these concepts are fundamental to achieving an understanding of probability theory and that they should be understood early on. Visualization of event spaces and execution of the related computational analyses can be difficult. Proficiency comes with experience, and obviously the reader cannot be expected to become a virtuoso in the manipulation of event spaces through the reading of this volume.

This chapter is divided into three sections. The first section presents the important ideas from an intuitive viewpoint, complete with illustrative examples. The second section treats the reader to a more rigorous presentation of similar material. There is some redundancy in the content of these two sections, though the correspondence is not one-to-one. However, since the approaches are quite different, the reader will find that most of the material in the second section appears to be new. The final section is an introduction to the topic of event counting, and it deals principally with permutations and combinations.

2.1 Intuitive Approach

The essential ingredients to be considered are: i) a well-defined action or operation, ii) a closed system or space of all possible outcomes of this action, iii) the elements or components which comprise this space, and iv) the rules or laws which govern manipulation of these elements. The reader may well recollect that these are the essential ingredients of set theory. The similarities are not coincidental.

The action or operation must be so defined that its execution always yields a result whose description entitles it to belong to the space of possible outcomes under consideration. Usually, it is assumed that this action does not perceptibly alter the content of the space (with replacement assumption), but in some models it may (without replacement assumption). The statistical laws will depend strongly on which assumption is in effect if the space is discrete and has a limited content. More will be said about this point in Sec. 2.3. Actions or operations on well-defined spaces of outcomes are usually referred to as sampling.

The space containing all possible outcomes of a particular sampling procedure is called a sampling space or event space. In the parlance of set theory, it is a universal set whose elements are called events. Plural execution of a particular sampling procedure produces a sequence of specific outcomes. A relatively small collection of these outcomes is usually designated as a sample. Samples which are so large that they essentially resemble or even deplete the entire event space are sometimes referred to as populations. The terms population and event space are sometimes used interchangeably, but the reader should realize that there is a distinction.

Probability theory does not dictate how event spaces are to be constructed, but rather it establishes rules pertaining to operations upon a space once it is defined. There are generally many ways to structure and model a given physical problem. Often the diverse possibilities are equivalent, but judicious definition of an event space can lead to important practical simplifications. This aspect of probability theory practice is more of an art than it is a science, and the true worth of a statistician can be measured by his skills in this area. A well-developed talent for visualization of event spaces is essential.

As a general rule, event spaces should be developed, first and foremost, in terms of elementary (or simple) events. By definition, these are events which cannot be further decomposed into more basic entities.

Example 2.1

Consider the sampling procedure consisting of selecting and observing atoms from a spent reactor fuel rod. We could choose to define the possibilities (events) as: fissionable material, fertile material, fission products, cladding and other structural materials, and miscellaneous impurities. A sample would then be a sequence of outcomes corresponding to these five possibilities. However, this choice might not be very useful and could even be ambiguous (e.g., fertile materials such as ^{238}U are also fissionable for fast neutrons). Clearly, a better choice in this example would be to characterize a simple event by the pair (Z,N) , since this uniquely identifies which element and isotope is observed (the outcome) in any particular sampling action. In some applications it might also be important to keep track of whether an atom was in its ground state or resided as an excited isomer. Then, our model would have to include a third parameter E_x to indicate excitation energy of the nucleus. Consequently, a simple event would be identified uniquely by the triplet (Z,N,E_x) .

Example 2.1 offers the opportunity to introduce the concept of an augmented event space. We know that all atomic numbers Z from 1 to 108 are possible (confirmed elements). Also, neutron numbers N from 1 to 159 have been observed. Finally, nucleon binding energies up to $E_x \leq 10$ MeV are possible. However, only certain combinations (Z,N,E_x) for $1 \leq Z \leq 108$, $0 \leq N \leq 159$ and $0 \leq E_x < 10$ MeV actually correspond to atomic species which can exist and which might be found in the spent fuel rod. The true event space is comprised only of these realistic possible outcomes. When the nonphysical triplets are included, however, we have what is called an augmented space. Later in our treatment of probability it will be seen that it is sometimes useful to contemplate such augmented spaces in order to simplify certain mathematical operations.

A Venn diagram is a graphical representation of an ensemble of elementary events which form an event space. These pictorial diagrams can be very useful aids in visualizing event spaces and manipulations involving them.

Example 2.2

Figure 2.1 is a Venn diagram of (Z,N) pairs for an event space consisting of a few light nuclei. Shaded squares represent nuclei which have been observed, and the totality of these forms our event space. Blank squares represent nonphysical entities. The collection of all squares (shaded or blank) form an augmented event space.

The specific parameters which define an elementary event can be either discrete or continuous. It is evident in Example 2.1 that Z and N are discrete parameters, while E_x should be treated as continuous. In this chapter we generally focus on discrete quantities, but more is said about continuous ones in Chap. 3 and beyond.

Individual elementary events, or arbitrary collections of simple events within the space, can be interpreted as subspaces of the event space. Subspaces with well-defined attributes can also be considered as events. These are called compound events. In Example 2.1 we alluded to five possible compound events (e.g., fission products was one of these). Once a set of elementary events has been defined to form an event space, there is generally a variety of ways to partition it into subspaces. We will examine this in a quantitative way later in this report. First, we consider an example in which some compound events are explicitly illustrated.

Example 2.3

One day a nuclear researcher walks into his laboratory and discovers that a certain piece of electronic apparatus is malfunctioning. He traces the malfunction to a particular circuit board. Being a statistically inclined individual, he decides to analyze the possibilities for failure from this point of view before proceeding with the repairs. He therefore generates a failure-mode event space based on the status of the board's four integrated-circuit (IC) components. In this space an elementary event corresponds to a declaration of the condition of each of the four IC components as good (G) or bad (B). The sixteen elements which form this space are listed in Table 2.1. Based on this model, there are two obvious possibilities for forming compound events. Possibility A involves specification of the number of IC components involved, i.e., none, one, two, three or all four. Five compound events of this type span the space. Possibility B

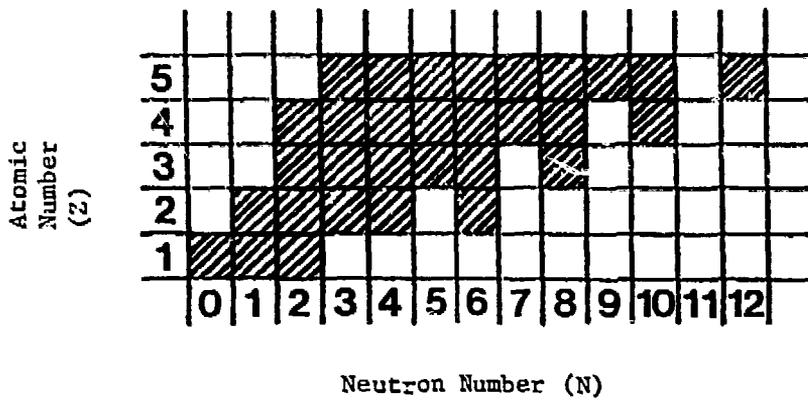


Figure 2.1: Venn diagram for an event space consisting of observable light nuclei (shaded). This space is augmented by the blank squares. See Ex. 2.2.

Table 2.1 Failure Modes for an Electronic Circuit Board with Four Integrated-Circuit (IC) Components. See Ex. 2.3.

Elementary Events ^a	Compound Events ^b	
	Possibility A	Possibility B
(G,G,G,G)	No bad IC components (1)	Non-IC failure (1)
(B,G,G,G)	One bad IC component (4)	IC failure (15)
(G,B,G,G)		
(G,G,B,G)		
(G,G,G,B)		
(B,B,G,G)		
(B,G,B,G)		
(B,G,G,B)		
(G,B,B,G)		
(G,B,G,B)		
(G,G,B,B)	Three bad IC components (4)	
(B,B,B,G)		
(B,B,G,B)		
(B,G,B,B)		
(G,B,B,B)	All bad IC components (1)	
(B,B,B,B)		

^aPosition in array (1,2,3,4) designated a particular IC component which is either good (G) or bad (B).

^bValue in parentheses (...) designates the number of elementary events in the subspace forming the compound event.

categorizes all failures into non-IC failures or IC failures, and thus involves two compound events. It is interesting to examine how many elementary events are required to form each of the compound events. For Possibility A, the reader may recognize that the number of elementary events corresponding to a failure of k IC components ($k = 0,4$) equals the well-known binomial coefficient C_{nk} for $n = 4$, namely

$$C_{4k} = 4!/[k!(4-k)!].$$

This point will be discussed further in Sec. 2.3. The total number of elementary events in the space representing the present problem is 2^4 . An important task in probability theory is that of determining the sizes of event spaces and of the subspaces which represent particular compound events. This issue will be addressed in more detail in other sections of this report.

Events are mutually exclusive if occurrence of one precludes occurrence of the other. In Example 2.3, the event involving two bad IC components is mutually exclusive from that involving one bad IC component. From Table 2.1, it is clear that the subsets representing these compound events would not overlap. Each simple event is mutually exclusive with respect to every other simple event by definition.

In some applications, order is important in the generation of an event space, while in others it is not. In Example 2.3, order is unimportant. If the failure of the circuit board is due to the first and third IC components being defective, the outcome is not affected by the order in which these two IC components are tested during the trouble-shooting process. The following example illustrates how order can be important in defining an event space that models physical reality.

Example 2.4

Consider an experiment designed to measure the energies of charged particles (CP) traveling in a particular direction. The CP energy is deduced from a determination of the times at which a particle passes through two separated, very-thin scintillation detectors D_1 and D_2 spaced by a distance L , as shown in Fig. 2.2.

We define a simple event by the recorded values of a pair of signals (one from each detector) at times t_1 and t_2 , respectively.

The choice of zero for the time scale is not important. However, causality is violated if $t_2 \leq t_1$, since it is assumed that the particles we are interested in pass through D_1 first. Actually $(t_2 - t_1) > L/c$ is required, since the particle velocity can never quite reach the speed of light, c , and the detectors D_1 and D_2

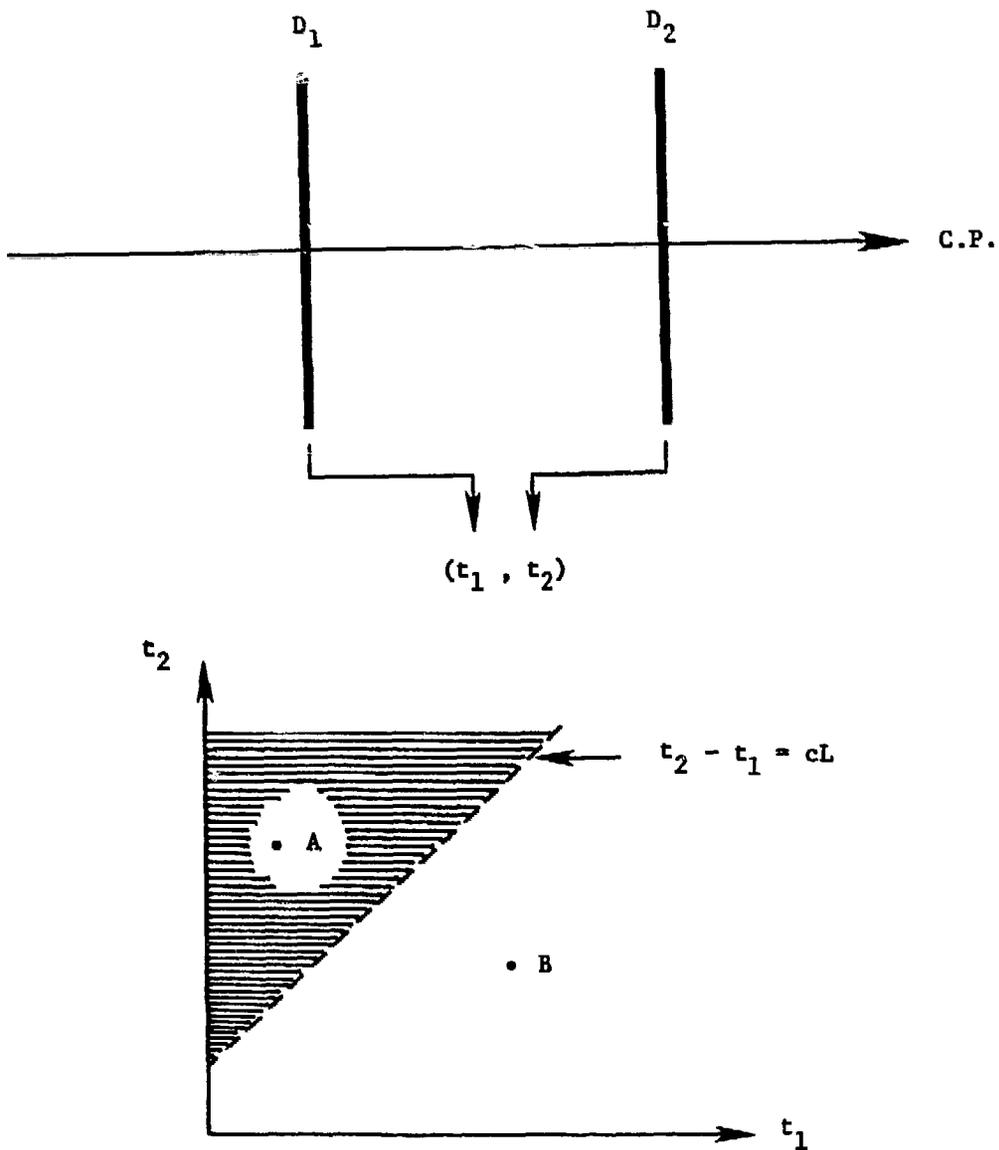


Figure 2.2: Venn diagram for an event space consisting of measured times for the passage of charged particles (CP) incident from the left through two detectors D_1 and D_2 separated by distance L . Points such as A (shaded region) are physically allowed, while those such as B are not. See Ex. 2.4.

have finite separation, L . Thus, the event space of physical possibilities can be represented by the Venn diagram shown in Fig. 2.2. A point, A , with $(t_2 - t_1) > L/c$ (shaded region) represents a valid physical event, while point B , with $(t_2 - t_1) \leq L/c$ is non-physical. The space of all pairs (t_1, t_2) is an augmented event space for which there is no condition on the relative magnitudes of t_1 and t_2 .

2.2 Theoretical Approach

This section deals with what is commonly referred to as set theory. The present treatment is abbreviated by necessity. The importance of understanding the definitions and concepts presented in this section will become evident to the reader as he proceeds through later sections of this report, and other reports in this series.

Basic to the discussion is the notion of a universal set, E , which we henceforth consider to be an event space, in accordance with the connotations in Sec. 2.1. The fundamental elements of E are simple events, designated by e , and interpreted as discussed in Sec. 2.1. From the point of view of dimensionality, E can be finite, consisting of a limited number of elements (e_1, e_2, \dots, e_n) , or denumerable, in which case the number of elements is infinite but countable, i.e., (e_1, e_2, \dots) , or nondenumerable, in which case we cannot explicitly list the individual elements. Most of the discussion here will pertain to finite or denumerable event spaces, since nondenumerable spaces require special treatment with regard to certain mathematical operations. We will consider how to deal with nondenumerable spaces of interest later in this report. An event space, E , must be complete in the sense that the ensemble of all elements e which form E must exhaust the possibilities for outcomes of a well-defined sampling activity, as discussed in Sec. 2.1.

Next, we discuss some fundamental terms of set theory.

Compound event or subspace:

A is a compound event or subspace of E if all simple events in A also belong to E . The notation is $A \in E$. Figure 2.3 is a Venn diagram which illustrates the notion of a subspace. By definition, a simple event is also a compound event. Henceforth, the terms event and subspace will be used interchangeably. Simple and compound events may not be distinguished, except possibly by context.

Containment:

If A_1 and A_2 are two events in E , and every simple event of A_2 is also in A_1 , then

$$A_1 \in E, A_2 \in E, \text{ and } A_2 \in A_1.$$

However, it need not follow that $A_1 \in A_2$. Figure 2.4 is a Venn Diagram which illustrates the case where $A_2 \in A_1$, but A_1 is not contained by A_2 . This we designate by $A_1 \notin A_2$.

Sure or certain event:

By definition, E is a subspace of E ($E \in E$), and E , since it includes all possible outcomes from sampling, is designated the sure event or certain event.

Impossible or null event:

In set theory, one is required to define the empty space, designated ϕ , for completeness. ϕ is a space with no simple elements e of E , yet $\phi \in E$ by definition. Also, $\phi \in A$, where A is any subspace of E .

Equality:

Events A_1 and A_2 are equal if they contain the same simple events. Then, $A_1 \in A_2$ and $A_2 \in A_1$.

Nested sequence of events:

A nested sequence of events is a finite or denumerable collection of events, e.g., (A_1, A_2, \dots, A_n) or (B_1, B_2, \dots) of E for which a distinct containment hierarchy exists. We say, e.g., that the sequence (A_1, A_2, \dots, A_n) is nonincreasing if and only if we have

$$A_n \in A_{n-1} \in \dots \in A_2 \in A_1 \quad .$$

On the other hand, the sequence (B_1, B_2, \dots) is nondecreasing if and only if we have

$$B_1 \in B_2 \in \dots \quad .$$

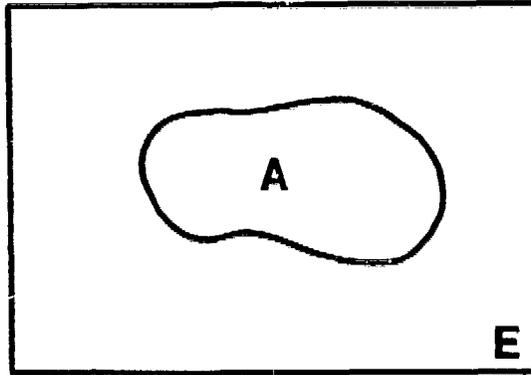


Figure 2.3: A is a subspace of E ($A \in E$).

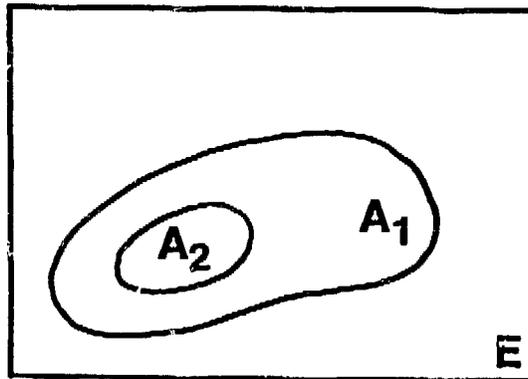


Figure 2.4: A_1 and A_2 are subspaces of E ($A_1 \in E, A_2 \in E$). A_2 is a subspace of A_1 ($A_2 \in A_1$), but the converse is not true ($A_1 \notin A_2$).

Figure 2.5 exhibits two nested sequences of events in E which are nonincreasing and nondecreasing, respectively.

The algebra of sets, or Boolean algebra as it is often called, governs the combination and manipulation of events. This algebra is based upon the operations discussed below.

Sum or union:

A is the sum or union of events A_1 and A_2 of E provided it contains those and only those elementary events of E which belong to at least one of the events A_1 and A_2 . The notation is

$$A = A_1 + A_2 \quad .$$

If there are several events ($A_1, A_2, \dots, A_i, \dots$) which form a union to yield A, the definition is similar but the notation

$$A = \sum_i A_i$$

is usually used. The union of events is illustrated via a Venn diagram in Fig. 2.6.

Difference or relative complement:

The difference A (occasionally called the relative complement) of events A_1 and A_2 of E, designated

$$A = A_1 - A_2 \quad ,$$

consists of all elementary events of A_1 , but excludes those which are also in A_2 . Note that $A_1 - A_2$ and $A_2 - A_1$ are conceptually different events and are never equal unless both happen to be the null event ϕ . Figure 2.7 is a Venn diagram which illustrates the notion of the difference of two events.

Product or intersection:

Consider two events A_1 and A_2 in E. The product or intersection of A_1 and A_2 , denoted $A_1 \times A_2$, is that event A of E which consists only of those simple events which simultaneously are contained in both A_1 and A_2 . Clearly,

$$A \in E, A \in A_1, \text{ and } A \in A_2.$$

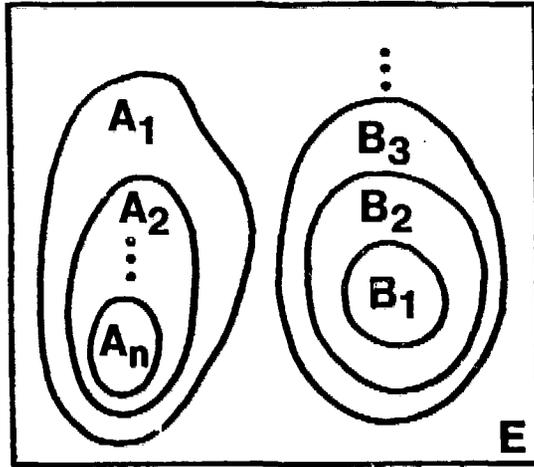


Figure 2.5: Nested sequences of events in E . (A_1, A_2, \dots, A_n) is nonincreasing, while (B_1, B_2, \dots) is nondecreasing.

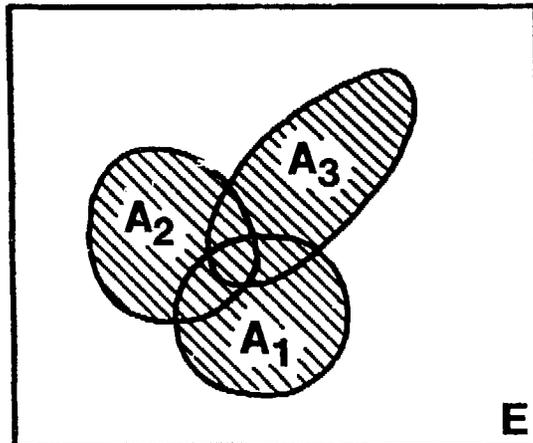


Figure 2.6: Union of events A_1 , A_2 , and A_3 of E is the shaded area A designated by $A = A_1 + A_2 + A_3$.

If several events A_1, A_2, \dots are involved in forming the product A , the notation

$$A = \prod_i A_i$$

is usually used. Figure 2.8 illustrates the notion of an event product.

Having defined the basic algebraic operations of event space, we are now in a position to offer a few more useful definitions.

Complement:

If A is an event in E , then the complement \bar{A} of A is the subspace which contains all the remaining elementary events of E that are not in A . It is thus true, e.g., that

$$A + \bar{A} = E, A \times \bar{A} = \phi, \bar{\bar{A}} = E - A, \text{ and } A = E - \bar{A},$$

based on the preceding definitions. Furthermore, it is possible to express the difference of two events using the notation of complementarity. If A_1 and A_2 belong to E , then

$$A_1 - A_2 = A_1 \times \bar{A}_2.$$

Figure 2.9 illustrates the notion of a complement.

Mutually exclusive or disjoint events:

Events A_1 and A_2 are said to be mutually exclusive or disjoint if both belong to E but have no simple events in common. Clearly, this condition is represented by the expression

$$A_1 \times A_2 = \phi.$$

as illustrated in Fig. 2.10. Whenever we have a collection of events (A_1, A_2, \dots) belonging to E which have the property that

$$A_i \times A_j = \phi$$

for any pair (i, j) , we say that the collection is pairwise disjoint. For any event A in E , it is obvious that A and its complement \bar{A} are a disjoint pair.

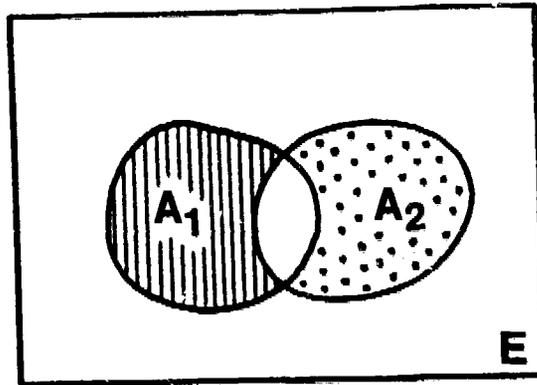


Figure 2.7: Differences of events A_1 and A_2 in E : $A_1 - A_2$ is the dashed region, while $A_2 - A_1$ is the dotted region.

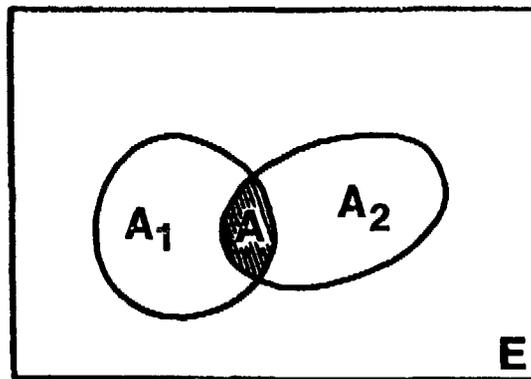


Figure 2.8: The shaded region A is the intersection $A_1 \times A_2$ of two events A_1 and A_2 in E .

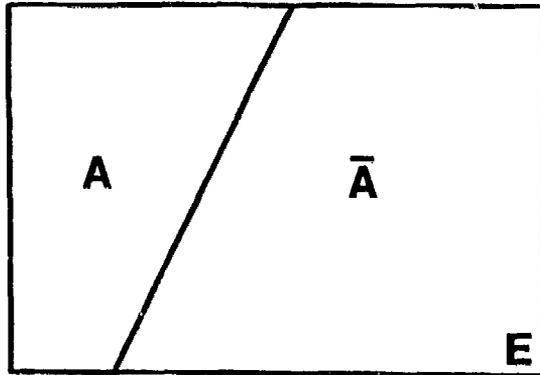


Figure 2.9: The complement \bar{A} consists of all simple events in E which do not belong to A .

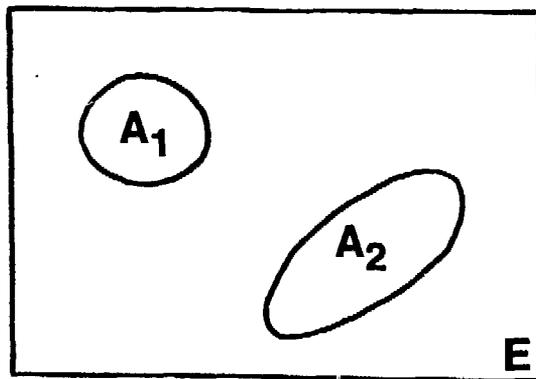


Figure 2.10: A_1 and A_2 are two mutually exclusive events belonging to E .

Partition:

Consider a pairwise disjoint collection of nonempty (not ϕ) events (A_1, A_2, \dots) belonging to E with the additional property that

$$\sum_i A_i = E \quad .$$

We then say that the collection forms a particular partition of E . Obviously, there are many different ways to partition E , as is evident from the examples in Fig. 2.11. It is also clear that, for any nonempty event A in E , the pair (A, \bar{A}) forms a partition of E , as is illustrated in Fig. 2.9. The concept of a partition is extremely important in the context of Bayes' theorem, which we will investigate in considerable detail in Chapter 8 of this report.

Limit of a sequence:

For a denumerable sequence of nested events (A_1, A_2, \dots), it is reasonable to define a limit. The manner of defining the limit depends upon whether the sequence is nonincreasing or nondecreasing. If (A_1, A_2, \dots) is nonincreasing, then the limit, A , is defined as

$$A = \pi_{i \geq 1} A_i \quad .$$

However, notice that for this type of sequence (with finite n), we have

$$\pi_{i=1}^n A_i = A_n \quad .$$

Thus, the limit, A , can also be written

$$A = \lim_{n \rightarrow \infty} A_n \quad .$$

If (A_1, A_2, \dots) is nondecreasing, then the limit, A , is defined by

$$A = \sum_{i \geq 1} A_i \quad .$$

However, notice that, for this type of sequence (with finite n), we have

$$\sum_{i=1}^n A_i = A_n \quad .$$

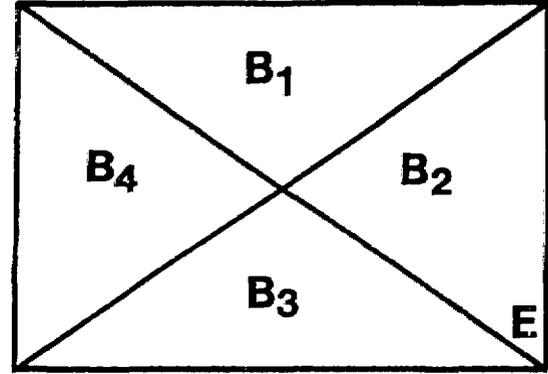
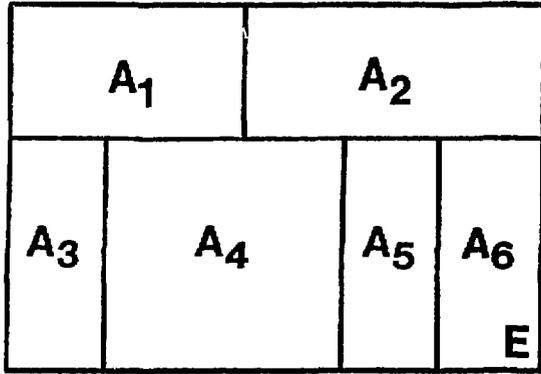


Figure 2.11: (A_1, \dots, A_6) and (B_1, \dots, B_4) represent two different partitions of E .

Thus, the limit A can be written as

$$A = \lim_{n \rightarrow \infty} A_n,$$

just as it was for a nonincreasing sequence of events.

Given the preceding definitions, we are now in a position to state the basic postulates of Boolean algebra, as applied to event sets.

Postulate No. 1: Laws of closure

For every pair of event sets A_1 and A_2 in E , there exist unique sets $A_1 + A_2$ and $A_1 \times A_2$ which also belong in E .

Postulate No. 2: Commutative laws

For every pair of event sets A_1 and A_2 , it holds that

$$A_1 + A_2 = A_2 + A_1$$

and

$$A_1 \times A_2 = A_2 \times A_1 .$$

Postulate No. 3: Associative laws

For any triplet of event sets A_1 , A_2 , and A_3 it holds that

$$(A_1 + A_2) + A_3 = A_1 + (A_2 + A_3)$$

and

$$(A_1 \times A_2) \times A_3 = A_1 \times (A_2 \times A_3) .$$

Postulate No. 4: Distributive laws

For any triplet of event sets A_1 , A_2 , and A_3 , it holds that

$$A_1 \times (A_2 + A_3) = (A_1 \times A_2) + (A_1 \times A_3)$$

and

$$A_1 + (A_2 \times A_3) = (A_1 + A_2) \times (A_1 + A_3) .$$

Postulate No. 5: Identity laws

There exist unique events, ϕ (the impossible or null event), and E (the sure or certain event), such that for any event A it holds that

$$A \times E = A$$

and

$$A + \phi = A$$

Postulate No. 6: Complementation law

Corresponding to each event A there exists a unique event \bar{A} called the complement, such that

$$A \times \bar{A} = \phi$$

and

$$A + \bar{A} = E$$

Many useful theorems and formulas that are employed in event-space analysis can be derived from these postulates. Before we examine some of these, it is worthwhile to digress to some extent in order to introduce the notion of a Borel field, or Sigma algebra as it is sometimes called. A Borel field Z based on the event space E is the set of all subsets of E. Since a subset of E is an event, it follows that Z is the exhaustive set of events based on E. By restricting our attention here to sets E that are finite or denumerable, we conveniently avoid difficulties with respect to defining the associated Borel field Z and subsequently considering its properties. Technically, the set Z based on E must possess the following five properties to be a true Borel field.

Property 1:

The event E (sure event) must be in Z.

Property 2:

The event ϕ (impossible event) must be in Z.

Property 3:

If a finite or denumerable collection of events A_1, A_2, \dots belong to Z, then their sum $\sum_i A_i$ also belongs to Z.

Property 4:

If events A_1 and A_2 belong to Z , then their difference $A_1 - A_2$ also belongs to Z .

Property 5:

If a finite or denumerable collection of events A_1, A_2, \dots belongs to Z , then their product $\prod_i A_i$ also belongs to Z .

It is shown in Section 2.3 that if E is finite and possesses n elements e , then the associated Borel field Z has 2^n elements z .

The notion of a Borel field has practical connotations. Given a space E of elementary events, then the corresponding Borel field represents the ensemble of all possible outcomes for an exhaustive collection of sampling procedures involving the elementary events of E .

Another feature of Borel fields is worthy of mention at this point. Let A be an event of E . Of course, it also belongs to the associated Borel field Z . Furthermore, we assume A is nontrivial (A is not ϕ). Thus, A consists of one or more simple events of E . One can now consider A as a distinct sample space of simple events, although we are aware that it is embedded in the larger space E . Then, one can define a new Borel field Z' based entirely on A . If B represents any arbitrary event of E , then the elements of Z' can be looked upon as being generated from products of the form $B \times A$. We shall see in Chapter 5 that this notion is the basis for dealing with conditional probability.

We now return to a consideration of the Boolean algebra of events. Given the preceding definitions, properties, and postulates, one can develop a rich area of mathematical theory which can only be touched on very briefly here. The careful reader will have already noticed that event manipulations differ in many respects from those with real numbers. Essentially this comes about because every real number is distinct, or elementary in our current parlance, while events can be compound and, thus, somehow "overlap" with other events. The fact that events need not be totally unique and distinct from other events leads to complications in set (or event) manipulations that force the uninitiated practitioner to develop a new arsenal of intuitive skills. Actually, the fundamental concepts of probability are generally rather simple. The difficulties one encounters in probability analysis, or in the consideration of uncertainties, are often traceable to the associated Boolean operations, not to the concepts themselves.

The following simple theorems can be readily demonstrated by means of Venn diagrams, though this will not be done here. If E is an event space, A and B are subspaces, and ϕ is the impossible or null event, then:

$$A + E = E, \quad (2.1)$$

$$A \times \bar{\phi} = \bar{\phi}, \quad (2.2)$$

$$A + A = A, \quad (2.3)$$

$$A \times A = A, \quad (2.4)$$

$$\bar{E} = \bar{\phi}, \quad (2.5)$$

$$\bar{\bar{\phi}} = E, \quad (2.6)$$

$$\bar{\bar{A}} = A, \quad (2.7)$$

$$A + (A \times B) = A, \quad (2.8)$$

$$A \times (A + B) = A, \quad (2.9)$$

where $\bar{\phi}$ denotes the complement of the impossible event, while the double bar over A denotes the complement of \bar{A} .

For interest, a rigorous proof of the theorem represented by Eq. (2.1) is provided, in terms of the previously mentioned postulates:

Step 1: $A + E = (A + E) \times E$, by Postulate 5.

Step 2: $(A + E) \times E = (A + E) \times (A + \bar{A})$, by Postulate 6 and substitution.

Step 3: $(A + E) \times (A + \bar{A}) = A + (E \times \bar{A})$, by Postulate 4 and substitution.

Step 4: $A + (E \times \bar{A}) = A + \bar{A}$, by Postulate 5 and substitution.

Step 5: $A + \bar{A} = E$, by Postulate 6.

Step 6: $A + E = E$, by substitution.

The following two equations are collectively known as the de Morgan laws:

$$\overline{(A + B)} = \bar{A} \times \bar{B} \quad , \quad (2.10)$$

$$\overline{(A \times B)} = \bar{A} + \bar{B} \quad . \quad (2.11)$$

Given a collection of n events A_1, A_2, \dots, A_n in E, one can express the de Morgan laws in the more general form

$$\overline{\left(\sum_{i=1}^n A_i \right)} = \prod_{i=1}^n \bar{A}_i \quad , \quad (2.12)$$

$$\overline{\left(\prod_{i=1}^n A_i \right)} = \sum_{i=1}^n \bar{A}_i \quad . \quad (2.13)$$

Given events A and B_1, B_2, \dots, B_m in E , one obtains the following two equations, known as the generalized distributive laws:

$$A \times \left(\sum_{i=1}^m B_i \right) = \sum_{i=1}^m (A \times B_i) \quad . \quad (2.14)$$

$$A + \left(\prod_{i=1}^m B_i \right) = \prod_{i=1}^m (A + B_i) \quad . \quad (2.15)$$

Two examples of rather complicated equations from Boolean algebra that can be derived from the definitions and postulates are:

$$\sum_{i=1}^n A_i = A_1 + \sum_{i=2}^n \left[A_i - \sum_{j=1}^{i-1} (A_j \times A_i) \right] \text{ for } n \geq 2, \quad (2.16)$$

$$\left(\sum_{i=1}^n A_i \right) \times \left(\sum_{j=1}^m B_j \right) = \sum_{i=1}^n \left[\sum_{j=1}^m (A_i \times B_j) \right] \quad . \quad (2.17)$$

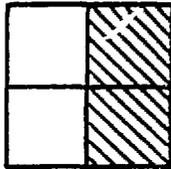
The interested reader can pursue this topic further by referring to texts which deal explicitly with set theory and Boolean algebra (e.g., Hau 57, Mah 68, Abb 69 and KM 76).

We will now close this section by considering two examples which illustrate some of the material presented up to this point.

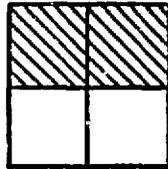
Example 2.5

Consider an arrangement whereby collimated neutrons, emitted from a reactor or accelerator source, impinge normally upon two separate, axially aligned, thin organic scintillation detectors A and B (see Fig. 2.12). The sampling process we consider is the emission of a single neutron from the source. There are four elementary events which form the sample space E which we will

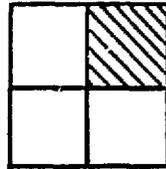
A: (1,0)
(1,1)



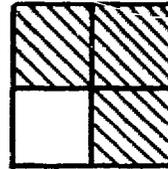
B: (0,1)
(1,1)



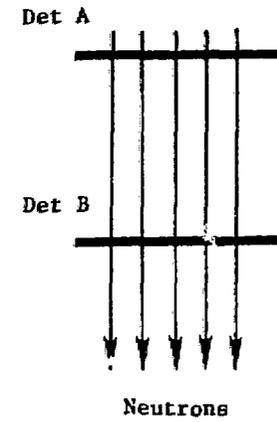
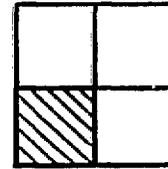
A X B: (1,1)



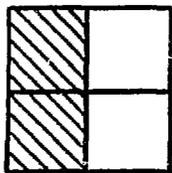
A + B: (1,0)
(0,1)
(1,1)



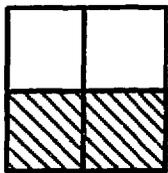
$\overline{(A + B)}$: (0,0)



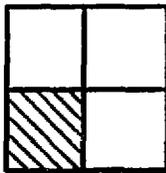
\bar{A} : (0,1)
(0,0)



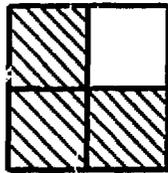
\bar{B} : (0,0)
(1,0)



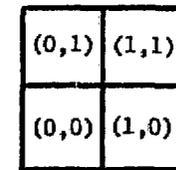
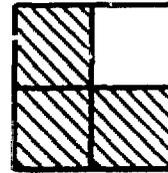
$\bar{A} \times \bar{B}$: (0,0)



$\bar{A} + \bar{B}$: (0,0)
(0,1)
(1,0)



$\overline{(\bar{A} \times \bar{B})}$: (0,0)
(0,1)
(1,0)



Sample
Space
E

Figure 2.12: Simple and compound events based on neutron detection by two organic scintillation detectors. See Example 2.5.

consider: i) (1,1): both A and B detect the neutron; ii) (1,0): A detects the neutron but B does not; iii) (0,1): B detects the neutron but A does not; and iv) (0,0): neither A nor B detects the neutron. Clearly, these labeled outcomes exhaust the possibilities. Our sample space E also contains the impossible event ϕ , as must all such spaces. Fig. 2.12 exhibits these possibilities, and it also indicates, by means of tables and Venn diagrams, several compound events that can be formed based on E. In Fig. 2.12, A represents the compound event in which detector A detects the neutron, while B represents the analogous compound event involving detector B. The validity of the de Morgan laws, given in Eqs. (2.10) and (2.11), is also demonstrated for this example in Fig. 2.12.

Example 2.6

In this example we consider a simple event space E and explicitly enumerate the subspaces which form its associated Borel field Z. Suppose that we possess a small quantity of ultrapure monoisotopic, elemental material which has been irradiated with neutrons in a reactor. This material consists of atoms in the ground state (G), or in one of two possible long-lived isomeric states (I1 or I2). The sampling procedure consists of selecting an individual atom and then determining which state it is in. The event space E has three simple events: e_1 (G), e_2 (I1) and e_3 (I2). It was previously indicated that the Borel field Z must have 2^3 (namely 8) elements. These are explicitly:

$$\begin{array}{ll}
 z_1 = \phi & z_5 = e_1 + e_2 \\
 z_2 = e_1 & z_6 = e_1 + e_3 \\
 z_3 = e_2 & z_7 = e_2 + e_3 \\
 z_4 = e_3 & z_8 = e_1 + e_2 + e_3
 \end{array}$$

Any event we can form by combining the elementary events of E must be equal to one of these z_i ($i = 1,8$). For example, suppose we consider $(e_1 + e_2) \times (e_1 + e_3)$. By Postulate 4 (distribution law), this event equals $e_1 + (e_2 \times e_3)$. However, $e_2 \times e_3$ is the impossible event ϕ because the e_i are elementary events and, as such, are nonintersecting. Thus, the event in question is $e_1 + \phi$, which equals e_1 and, therefore, z_2 , by Postulate 5.

A nontrivial subspace A of E is the space of all simple events corresponding to detection of an atom in an excited state. A consists of the two elements e_2 and e_3 . Furthermore, we can define a new Borel field Z' based on A alone. There are 2^2 (namely 4) elements in Z' .

These are explicitly:

$$\begin{aligned} z_1' &= \phi & z_3' &= e_3 \\ z_2' &= e_2 & z_4' &= e_2 + e_3 \end{aligned}$$

We can, e.g., demonstrate Property 3 (closure) for the Borel field Z' by considering $z_2' + z_3' + z_4'$. This is equivalent to $e_2 + e_3 + (e_2 + e_3)$ which, in turn, equals $(e_2 + e_3) + (e_2 + e_3)$ by Postulate 3 (associative law). Finally, this event equals $e_2 + e_3$, or z_4' by the theorem represented in Eq. (2.3).

2.3 Event Counting

It should be apparent to the reader, from the discussions in the two preceding sections, that counting the number of ways in which various types of compound events can be formed from simple events is an essential task for applications. This point was demonstrated in Example 2.3. So far in this report no general rules have been provided to indicate how such counting exercises are to be carried out. However, this matter will be addressed in the present section. There are a variety of counting algorithms that are used in practice, depending upon the problem at hand. Here, only a few of the more important ones will be reviewed. We have seen in the previous section that counting the number of subsets of a set is important. This by no means exhausts the applications for counting. In statistical physics, and in many other areas of basic and applied science, one is required to examine other counting procedures as well.

A few elementary definitions and concepts must be introduced first. A k-tuple is a collection of k quantities generated from elementary events belonging to an event space E. For present purposes we assume E to be finite, thus consisting of n elements. A k-tuple need not be a compound event in the sense discussed in Section 2.2, since a specific elementary event may be represented more than once in a particular k-tuple. However, every compound event corresponds to a certain k-tuple. Furthermore, k-tuples can be either ordered or not. For ordered k-tuples, or arrangements as they are sometime called, the position of every event in the array is important. Two k-tuples consisting of identical collections of one or more elementary events are distinguishable if ordering (or arrangement) is taken into consideration. Otherwise, they are indistinguishable. If each of the

available elementary events of E can only be represented once in forming k-tuples, then it is said that the k-tuples are formed without replacement. Such k-tuples have a one-to-one correspondence with k-fold subsets of E. Those k-tuples formed without regard to the multiplicity of the appearances of elementary events are said to be formed with replacement. Given an unordered k-tuple of events, there are a number of distinguishable ways that ordered k-tuples can be formed from it. Each of these arrangements is said to be a permutation of the unordered k-tuple. In counting exercises it is sometimes useful to visualize k-tuples in terms of occupancy. We then envision a collection of unique cells which may or may not be occupied. A k-tuple is equivalent to a k-fold specification of the occupancy status of these cells. The k entities which occupy these cells in some pattern may or may not be distinct, and multiple occupancy of cells may or may not be allowed, depending upon the statistical problem which is under consideration.

Example 2.7

Consider a space E consisting of the events e_1, e_2, e_3 , and e_4 . Then (e_1, e_2, e_3) , (e_2, e_1, e_3) , (e_1, e_1, e_4) and (e_1, e_4, e_1) are typical 3-tuples formed from E. Both (e_1, e_2, e_3) and (e_2, e_1, e_3) represent the same subspace of E, though they are distinct permutations of the three events e_1, e_2, e_3 , if ordering is considered. Neither (e_1, e_1, e_4) nor (e_1, e_4, e_1) can be considered a subspace of E. Each of these can be viewed as representing double occupancy of Cell No. 1 and single occupancy of Cell No. 4.

The basic principle of combinatorial analysis can be stated as follows: The number of ways W of forming ordered k-tuples is given by the formula

$$W = \prod_{i=1}^k N_i \quad , \quad (2.18)$$

where N_i is the number of ways available to form the i^{th} component of the k-tuple.

The simplest application of this principle is that of calculating the number of ordered k-tuples which can be formed from a set of n elementary events, with replacement. Since events can be "reused" according to the replacement assumption, there are n possibilities for selecting each of the components of the k-tuple. Therefore, each N_i equals n and

$$W = n^k \quad (2.19)$$

according to Eq. (2.18). It is interesting to note that even though it is hard to conceive of molecules in a classical gas as being "distinguishable", the Maxwell-Boltzmann Law of statistical physics is derived from this particular rule of counting, and it does seem to adequately describe the behavior of such an ensemble of particles under certain conditions of low particle density and high temperature (Mor 64).

The second application of Eq. (2.18) to be considered here is a determination of the number of ordered k-tuples which can be formed from a set of n events if no replacement is allowed. Under this restriction, $N_1 = n$, $N_2 = n-1, \dots, N_k = n-k+1$. So, an application of Eq. (2.18) leads to the expression

$$W = n(n-1)(n-2)\dots(n-k+1) \quad (2.20)$$

This formula does not form the basis for any known law of statistical physics, yet under the conditions for which the Maxwell-Boltzmann Law applies, this present counting rule leads to nearly the same result. Low particle density and high temperature imply that $k \ll n$, since, from an occupancy point of view, k represents the number of particles involved while n represents the number of states available for occupancy (which is large for high temperature).

Example 2.8

The predictions of Eqs. (2.19) and (2.20) for $k=5$ and $n=1000$ can be compared by examining the ratio $(1000^5)/[1000 \times 999 \times 998 \times 997 \times 996]$. The value of this ratio, to eight significant figures, is 1.0100654. The difference of 1% is negligible for most purposes.

Early in the present century it was discovered that certain microscopic physical phenomena could only be explained if the energy states available to atomic particles and their radiations were assumed to be quantized rather than continuous. Classical counting laws, which have in common the fact that individual particles are treated as distinguishable, also had to be abandoned in order to explain microscopic phenomena in the limit of low temperature and high particle density. It was learned that a particle of one certain class can never occupy the same quantum state as another similar particle (Pauli Exclusion Principle). Particles of this type are now known as fermions. Atomic particles from another class seem not to mind sharing quantum states with others of their kind. These particles are classified as bosons. For

nuclear scientists, the most important application of these fundamental principles of quantum statistics is in the consideration of nuclear structure, i.e., investigation of the manner in which nucleons (or clusters of them) occupy the various available states in their collective nuclear potential. We will now examine the two counting rules which form the basis of quantum statistics. Common to these two laws is the assumption that the individual particles which occupy the available quantum states are indistinguishable from each other. Thus, counting involves only unordered k-tuples from a space of n elementary events (or k-fold occupancy of n available cells or quantum states).

First, we consider the number of ways unordered k-tuples can be formed from n elementary events, without replacement. We recall that Eq. (2.20) provides an answer so long as ordering is preserved. If ordering becomes unimportant, the number of distinct possibilities must in general decrease. If we consider a particular ordered k-tuple, we should note that for $k \geq 2$ there are others which differ from the one we are considering only in the sense that they are permutations. Thus, when ordering is unimportant, we must divide the result of Eq. (2.20) by the number of ways that a collection of k quantities can be permuted. This number can be deduced by another application of Eq. (2.20), for $n=k$. The result is $\prod_{i=1}^k i$, conventionally designated $k!$ (k factorial), with $0! = 1$. Making use of the factorial notation, the counting law we seek is given by

$$W = C_{nk} = n!/[k!(n-k)!] \quad , \quad (2.21)$$

as first indicated in Example 2.3 of Sec. 2.1. The coefficient C_{nk} is called the binomial coefficient, for reasons to be discussed shortly. Example 2.3 demonstrates this counting rule. From the occupation viewpoint, we are considering the number of ways k objects can occupy n available cells, when the objects are indistinguishable and no two objects may occupy the same cell. Fermions behave this way, and the law of statistical physics that is based on this mode of counting is designated as Fermi-Dirac. The value W given by Eq. (2.21) is often referred to as the number of possible combinations of n quantities taken k at a time.

Before proceeding to the consideration of another distinct counting concept, we will address several topics that are convenient to consider at this point of the present development.

First, for large k the computation of $k!$ can be awkward. An approximation for $k!$, known as Stirling's approximation, is quite accurate as long as k is sufficiently large. The formula is

$$k! \approx (2\pi k)^{1/2} k^k e^{-k} \quad . \quad (2.22)$$

The worth of this formula becomes most evident when various factorial expressions appear in combination, e.g., as in Eq. (2.21). Since these combinations usually involve products and quotients, the logarithmic version of Eq. (2.22), i.e.,

$$\ell n(k!) \approx (1/2)\ell n(2\pi k) + k\ell n(k) - k \quad (2.23)$$

is often more useful in practice than Eq. (2.22). Table 2.2 provides a comparison of the directly calculated values $k!$ and $\ell n(k!)$, for $k = 1$ to 10, with corresponding results obtained from Stirling's approximation, according to Eqs. (2.22) and (2.23). For $k \geq 10$, the agreement for $\ell n(k!)$ is better than 0.1%, while that for $k!$ is better than 1%.

The counting rule corresponding to Eq. (2.21) is the one applicable to the determination of the number of subsets of size k which can be formed from the n -fold event space E . The size $W(Z)$ of the Borel field Z associated with E , including the impossible event ϕ , is thus

$$W(Z) = \sum_{k=0}^n C_{nk} \quad (2.24)$$

We shall digress for a moment and note that the binomial coefficient C_{nk} is thus named because of the role it plays in the well-known binomial theorem, attributed to Sir Isaac Newton. This theorem states that for two numbers a_1 and a_2 ,

$$(a_1 + a_2)^n = \sum_{k=0}^n C_{nk} a_1^k a_2^{n-k} \quad (2.25)$$

If we set $a_1 = 1$ and $a_2 = 1$, we deduce from Eqs. (2.24) and (2.25) that $W(Z)$ equals 2^n . This proves that the number of elements of the Borel field Z is 2^n , a fact which has been expressed previously.

The binomial coefficient is readily seen to be symmetric, from its definition, Eq. (2.21), so

$$C_{nk} = C_{n, n-k} \quad (2.26)$$

Likewise, for all n it is clear that

$$C_{n1} = C_{nn} = 1 \quad (2.27)$$

Furthermore, it can be proved rather easily that binomial coefficients satisfy the following useful recursion relation:

$$C_{n, k-1} + C_{nk} = C_{n+1, k} \quad (2.28)$$

This expression is the basis for the famous geometrical arrangement of integers known as Pascal's Triangle, which is shown in Fig. 2.13. Pascal's triangle provides an easily remembered algorithm for deriving binomial coefficients.

Table 2.2: Predictions of Stirling's Approximations for $\ln(k!)$ and $k!$

k	$\ln(k!)$		Actual	$k!$	Eq. (2.22)
	Actual	Eq. (2.23)			
1	0	- 0.081061	1		0.922137
2	0.693147	0.6518065	2		1.919004
3	1.791759	1.7640815	6		5.8362096
4	3.1780538	3.1572632	24		23.506175
5	4.787492	4.77084705	120		113.019168
6	6.579251	6.56537508	720		710.07818
7	8.5251614	8.51326465	5040		4980.3958
8	10.6046029	10.5941916	40320		39902.395
9	12.80182748	12.79257202	362880		359536.873
10	15.10441257	15.0960821	3628800		3598695.619

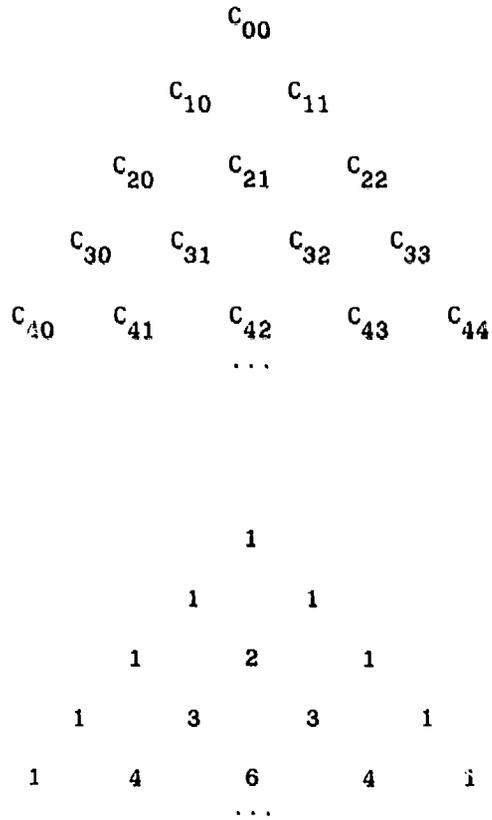


Figure 2.13: Pascal's triangle for the binomial coefficients. The pattern is based on Eqs. (2.26)-(2.28). The coefficients on any given row are obtained by summing adjacent coefficients from the preceding row.

A generalization of the problem of determining how many subsets of size k can be formed from a set of n events is that of determining how many ways one can partition a set of n events into r ordered subsets so that the first has size k_1 , the second has size k_2, \dots , and the r^{th} subset has size k_r . The constraint is

$$\sum_{j=1}^r k_j = n.$$

To approach this problem, one repeatedly applies the reasoning leading up to Eq. (2.21). The result is the following product of binomial coefficients:

$$W = C_{nk_1} C_{n-k_1, k_2} C_{n-k_1-k_2, k_3} \dots C_{n-k_1-\dots-k_{r-1}, k_r} \quad (2.29)$$

This coefficient can be written in the more compact form

$$W = C_{n; k_1 k_2 \dots k_r} = n! / (k_1! k_2! \dots k_r!) \quad (2.30)$$

It is called the multinomial coefficient owing to the role it plays in the multinomial theorem. This theorem states that for r numbers a_1, a_2, \dots, a_r

$$\left(\sum_{j=1}^r a_j \right)^n = \sum_{k_1} \sum_{k_2} \dots \sum_{k_r} C_{n; k_1 k_2 \dots k_r} a_1^{k_1} a_2^{k_2} \dots a_r^{k_r} \quad (2.31)$$

$(k_1 + k_2 + \dots + k_r = n)$

The special case of binary partitioning, which we have already discussed in detail, follows readily from Eqs. (2.30) and (2.31), for $r = 2$.

Finally, we turn our attention to yet another counting scheme. For this case, we consider forming unordered k -tuples with replacement allowed. From the occupancy point of view, multiple occupancy of the available cells is permitted and the occupants are indistinguishable. In statistical physics, this scheme is labeled Bose-Einstein. Ensembles of bosons at extremely low temperatures tend to accumulate in the available lowest-energy states (condensation), giving rise to nonclassical phenomena such as superfluidity. We state here, without proof, that the number of k -tuples that can be formed, subject to these conditions, is the particular binomial coefficient

$$W = C_{n+k-1, k} \quad (2.32)$$

This result can be proved by the induction method (e.g., Par 60), but, in general, the process is not as intuitively simple to visualize as the preceding counting rules. For small n and k it can be readily demonstrated that Eq. (2.32) gives the correct result, as the following example indicates.

Example 2.9

Consider forming pairs ($k = 2$) from a set of three ($n = 3$) elements, e_1 , e_2 , and e_3 . If the pairs are ordered and formed with replacement, Eq. (2.19) tells us that there are $3^2 = 9$ of them. They are explicitly:

(e_1, e_1)	(e_2, e_1)	(e_3, e_1)
(e_1, e_2)	(e_2, e_2)	(e_3, e_2)
(e_1, e_3)	(e_2, e_3)	(e_3, e_3)

If we arrange these differently and drop the ordering condition, we have:

(e_1, e_2)	(e_1, e_2) (e_2, e_1)	(e_1, e_3) (e_3, e_1)	(e_2, e_3) (e_3, e_2)
(e_2, e_2)			
(e_3, e_3)			

The pairs in the boxes are no longer distinct if ordering is dropped. Thus, for the Bose-Einstein condition we have only 6 distinct pairs, in agreement with Eq. (2.32).

For the convenience of the reader, the four basic counting laws we have considered in this section are summarized in Fig. 2.14. We close this section by considering an example which further illustrates some of the material discussed in this section.

Example 2.10

Consider four distinct energy levels as follows: $e_1 = 0$, $e_2 = e$, $e_3 = 2e$, and $e_4 = 3e$. Furthermore, assume that there are three particles to be distributed among these available energy levels. Let E be the total energy of the system corresponding to an arbitrary allowed-occupancy configuration for the particles. The objective is to determine the number of distinct states that can be formed for each possible total energy E of the system of particles, considering the four statistical counting rules summarized in Fig. 2.14. The results of this investigation are indicated in Figs. 2.15 and 2.16 and Table 2.3. Clearly, the choice of statistical counting law has a dramatic impact on the outcome for this simple example. We note that the distributions

	Distinguishable	Indistinguishable	Occupancy picture
With replacement	$W = n^k$ (Maxwell-Boltzmann)	$W = C_{n+k-1, k}$ (Bose-Einstein)	Without exclusion
Without replacement	$W = n(n-1)\dots(n-k+1)$	$W = C_{nk}$ (Fermi-Dirac)	With exclusion
Forming k-tuples	Ordered	Unordered	

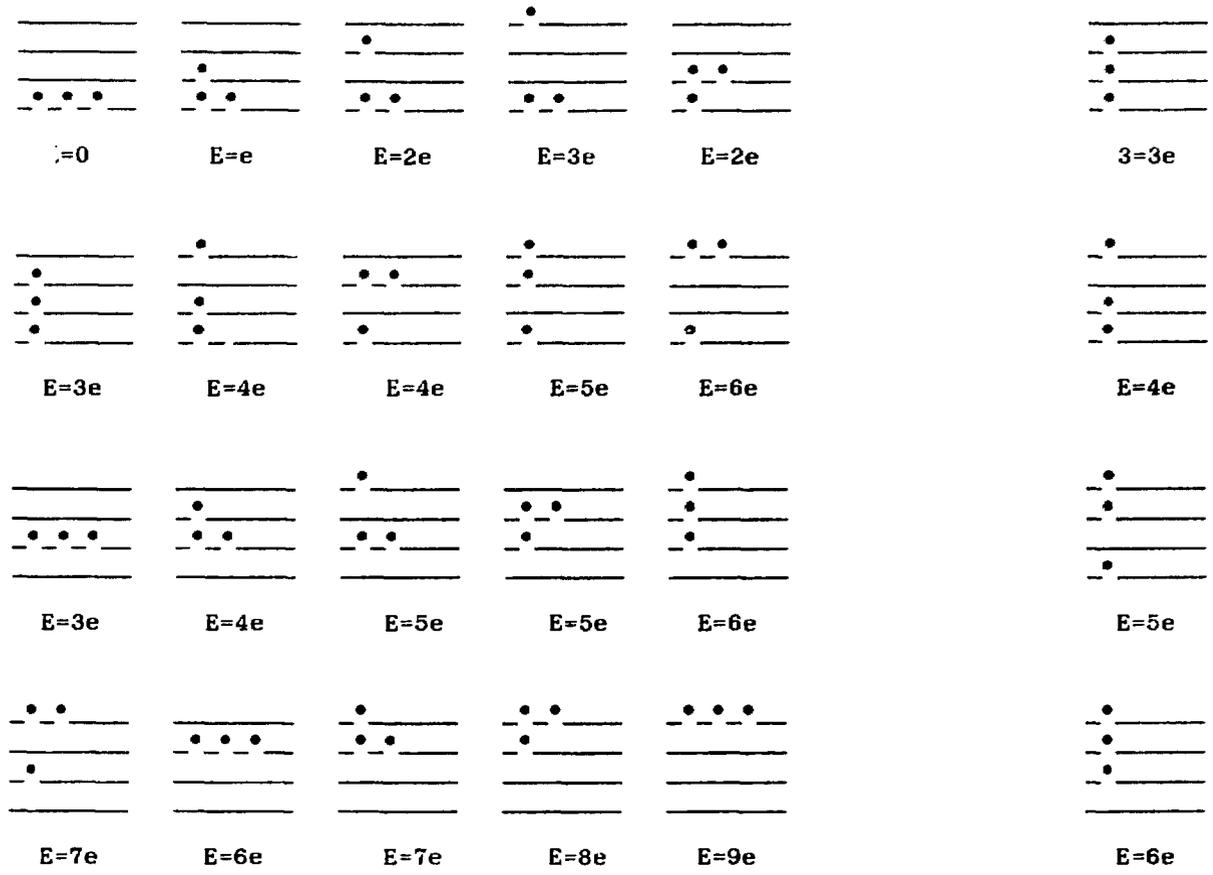
Figure 2.14: Summary of several important counting laws dealing with k items from a collection of n . The indicated C-coefficients are binomial coefficients as defined by Eq. (2.21). Both the k -tuple and occupancy viewpoints are indicated.

	$(e_1, a, b):$					$(e_2, a, b):$					$(e_1, a, b):$					$(e_2, a, b):$				
	e_4	3e	4e	5e	6e	e_4	4e	5e	6e	7e	e_4	*	4e	5e	*	e_4	4e	*	6e	*
	e_3	2e	3e	4e	5e	e_3	3e	4e	5e	6e	e_3	*	3e	*	5e	e_3	3e	*	*	6e
b	e_2	e	2e	3e	4e	e_2	2e	3e	4e	5e	e_2	*	*	3e	4e	e_2	*	*	*	*
	e_1	0	e	2e	3e	e_1	e	2e	3e	4e	e_1	*	*	*	*	e_1	*	*	3e	4e
		e_1	e_2	e_3	e_4		e_1	e_2	e_3	e_4		e_1	e_2	e_3	e_4		e_1	e_2	e_3	e_4
			a					a					a					a		
	$(e_3, a, b):$					$(e_4, a, b):$					$(e_3, a, b):$					$(e_4, a, b):$				
	e_4	5e	6e	7e	8e	e_4	6e	7e	8e	9e	e_4	5e	6e	*	*	e_4	*	*	*	*
	e_3	4e	5e	6e	7e	e_3	5e	6e	7e	8e	e_3	*	*	*	*	e_3	5e	6e	*	*
b	e_2	3e	4e	5e	6e	e_2	4e	5e	6e	7e	e_2	3e	*	*	6e	e_2	4e	*	6e	*
	e_1	2e	3e	4e	5e	e_1	3e	4e	5e	6e	e_1	*	3e	*	5e	e_1	*	4e	5e	*
		e_1	e_2	e_3	e_4		e_1	e_2	e_3	e_4		e_1	e_2	e_3	e_4		e_1	e_2	e_3	e_4
			a					a					a					a		

(A) No exclusion (Maxwell-Boltzmann), Eq. (2.19).

(B) With exclusion, Eq. (2.20).

Figure 2.15: Distinct ways of distributing three distinguishable particles among four energy states. Allowed total energies E appear in the square arrays. Asterisks mark the excluded configurations. See Ex. 2.10.



(A) No-exclusion (Bose-Einstein), Eq. (2.32)

(B) With exclusion (Fermi-Dirac) Eq. (2.21)

Figure 2.16: Distinct ways of distributing three indistinguishable particles among four energy states. Ensemble total energies are indicated under each diagram. See Ex. 2.10.

Table 2.3: Available Configurations with Indicated Ensemble Total Energies for Three Particles Distributed Among Four Available Energy States. See Example 2.10^a.

Ensemble Total Energy (E)	Maxwell-Boltzmann	Distinguishable with Exclusion	Bose-Einstein	Fermi-Dirac
0	1	-	1	-
e	3	-	1	-
2e	6	-	2	-
3e	10	6	3	1
4e	12	6	3	1
5e	12	6	3	1
6e	10	6	3	1
7e	6	-	2	-
8e	3	-	1	-
9e	1	-	1	-
Possible Configurations	64	24	20	4

^aRefer also to Figs. 2.14-2.16. Hyphens indicate excluded total energies.

of numbers of available states of a given total energy are symmetric, regardless of the statistics. Application of the exclusion principle for individual particles results in inaccessibility for configurations corresponding to many of the total energies which would otherwise be represented.

The link between nuclear data uncertainties and the material presented in this chapter has not yet been established, but it will be made clearly evident in due course in this series. In the next chapter we investigate some additional concepts which are needed in order to be able to make this connection. The reader should depart from the present chapter with a respect for the complexities associated with event definition and event counting. The reader should also reflect upon the number of interesting symmetries which have been observed. Skill in visualizing such symmetries is essential to the mastery of event-space analysis.

3. RANDOM VARIABLES AND THEIR FUNCTIONS

At this stage we are not yet ready to directly undertake a study of probability theory. First, it is necessary to explore three additional related aspects of the theory, namely, the notions of randomness, random variables, and functions of random variables. The relationship between random events and random variables must be established. The differences between discrete and continuous random variables also need to be understood. This chapter is divided into three sections. The first deals with the concept of randomness. The second discusses random variables, both discrete and continuous, and how they relate to observable events. Finally, the last section addresses three classes of random-variable functions which play important roles in the theory.

3.1 Concept of Randomness

We can easily conjure mental images, based on our personal experiences, that illustrate the concept of randomness. Terms such as "unpredictable," "unbiased," "indeterminate," etc., are often used to describe our interpretations of this notion. Generally, these preconceived concepts of randomness serve us well, since random phenomena are common in the experiences of life. Nevertheless, it is necessary at this point to refine this concept somewhat in order to be sure that we agree on the meaning of certain terms before proceeding to develop the topic further.

The following intuitive definition, quoted from Parzen (Par 60), serves well for present purposes: "A random (or chance or stochastic) phenomenon is an empirical phenomenon characterized by the property that its observation under a given set of circumstances does not always lead to the same observed outcome (so that there is no deterministic regularity), but rather to different outcomes in such a way that there is "statistical regularity."

This can be restated using terms from Chapter 2. When a process operates on an event space but there is no way to control individual outcomes of the procedure involved, then the process itself is random and the events which constitute the event space in question are random events. Although the exact outcome of any single random trial is unpredictable, a pattern is expected to emerge based on the accumulated experience of many trials (i.e., a sample of reasonable size). In fact, in the spirit of Parzen (Par 60), we will treat as random only those situations in which we anticipate a priori that such a pattern exists. It is essential to understand the distinction between randomness, with attendant statistical regularity, and chaos. Chaotic phenomena exhibit no discernible pattern, even for substantially large samples. However, one cannot learn whether a particular process is chaotic or statistically "well-behaved" by considering only a single trial. This must either be postulated on the basis of related experience or "common sense," or must be deduced from protracted sampling experiments. Suffice to say, statisticians earn their livings by dealing with these issues in practical applications, but experimenters are required to have a reasonable understanding for their own purposes.

Fortunately, natural processes are rarely found to be chaotic when they become well understood. On the other hand, results of individual trials generally do fluctuate in a random way, so that the outcomes are not deterministically predictable. The mathematical theory of probability has been developed to treat such random processes in a rigorous way. It is a useful theory because one can construct mathematical models within the framework of the theory which represent the features of many observable phenomena quite adequately. In later chapters of this report, and in subsequent reports of this series, we will, on numerous occasions, deal with mathematical analogues of observable phenomena, but we should not forget that the mathematical quantities which we will consider are useful only to the extent that the phenomena in question behave fundamentally in a way that is consistent with the postulates of the theory (i.e., controlled conditions).

The reader should not be left with the impression that all observable phenomena have to be distinctly categorized as either deterministic, random or chaotic. Intermediate situations can and do occur, and this fact adds to the complexity of interpreting and dealing with the real world. For example, much of the effort expended by experimental physicists involves identification and elimination of chaotic effects (e.g., malfunctioning equipment, electromagnetic fluctuations and interference, etc.) so that their experiments can be performed under the controlled conditions that are required in order to generate random, yet statistically well-behaved results, consistent with the theories which they seek to explore.

The following example illustrates the concepts discussed above.

Example 3.1

Consider an experimental setup which includes a 14-MeV neutron generator and a detector used to monitor neutron fluence. Neutrons are produced by bombarding a thin metal-target layer containing absorbed tritium with few-hundred-keV deuterons. The detector is assumed to be an organic scintillator. We seek to perform a particular experiment under conditions of controlled neutron production and stable neutron detection. Anyone having experience with such a facility can immediately identify a number of factors which influence the experimental configuration, e.g., accelerator-voltage stability, ion-source stability, target stability, geometry, detector-gain stability, etc. Can we identify any observables associated with this setup which behave in a deterministic manner? There are some trivial ones, such as discrete switch settings for the apparatus, which are assumed to be fixed from one measurement to the next (sequential samplings). These should be unambiguous if no one has disturbed the setup. Experimentalists should, of course, periodically check such things, because mysterious alterations are not uncommon, particularly if the apparatus has been left unattended. Some factors are nearly deterministic, e.g., in the present setup, the accelerator beam probably consists mainly of deuterons (atomic or

molecular) if a high-purity deuterium gas supply is attached to the ion source and the vacuum is good.

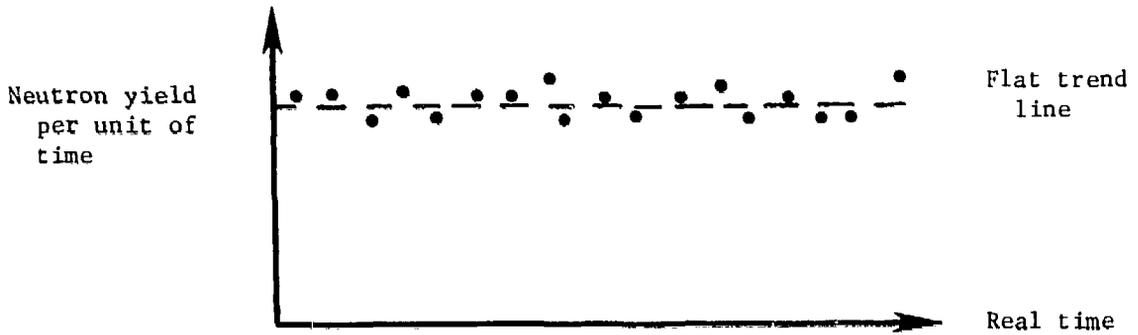
However, most of the parameters we might select to monitor during this experiment would behave randomly, even under ideal conditions. If, for example, we were to plot up the actual neutron yield deduced from the measured detector events per unit time for a relatively short time under acceptably stable conditions, we might obtain a plot as shown in Fig. 3.1(a). The trend is flat, but there are obviously fluctuations. The nature of these fluctuations over a large number of trials (and thus the uncertainty of the fluence determination) could be predicted in a statistical sense. Figure 3.1(b) exhibits a behavior that experimenters, who are experienced in this line of work, will recognize as more realistic than that of near-constant neutron output. In this case, the systematic decline in neutron yield over an extended time interval could probably be traced to a particular factor, e.g., steady depletion of available tritium in the target. One can usually accommodate such systematic behavior, once the origin is understood. Note, however, that unavoidable random fluctuations about the decreasing trend line persist. Figure 3.1(c) represents a chaotic situation. No experimenter would tolerate these sudden, unpredictable jumps in measured neutron output. Clearly, one or more components of the apparatus are malfunctioning. The experimenter would be forced to find a solution to this problem before proceeding with the experiment.

In summary, experimenters bear the responsibility of understanding their experiments to the extent that they can intelligently estimate the degree to which the observed experimental behavior at hand is deterministic, random, or chaotic. The manner in which they subsequently interpret their data must be consistent with these considerations.

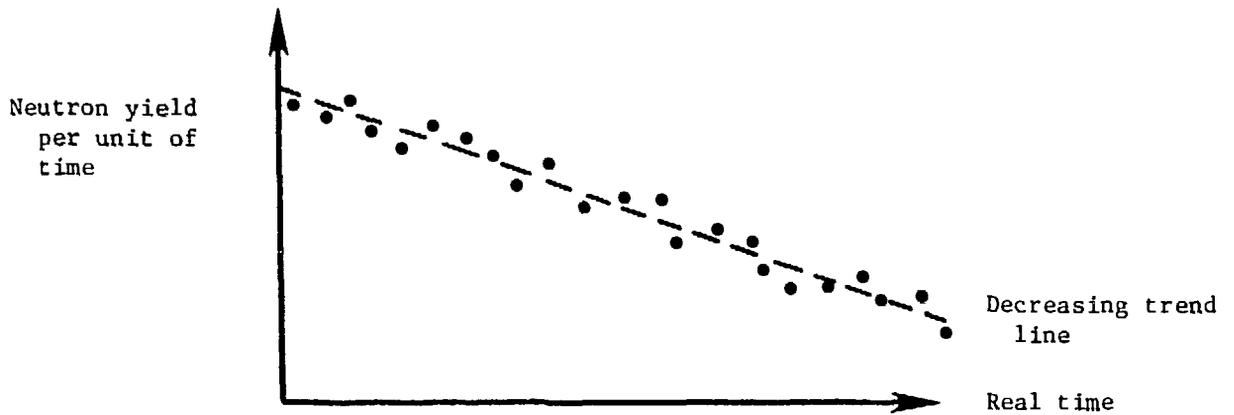
3.2 Random Variables

Considerable attention is devoted in Chap. 2 to developing the notion of events as the possible outcomes from random sampling procedures. No particular effort is made to consider how one might label or characterize these abstract random events. In practical situations, it is important to be able to label events, and the use of numerical parameterization is by far the most versatile approach. It is used, for instance, in Ex. 2.1. There, atomic number, neutron number, and nuclear-excitation energy are employed as useful parameters for unambiguous event specification. The use of real numbers to characterize events makes possible the marriage of statistical theory and numerical analysis, a very powerful combination which is essential for the present development of uncertainty theory.

(a) Normally "stable" neutron output



(b) Systematically declining neutron output



(c) Chaotic behavior of neutron output

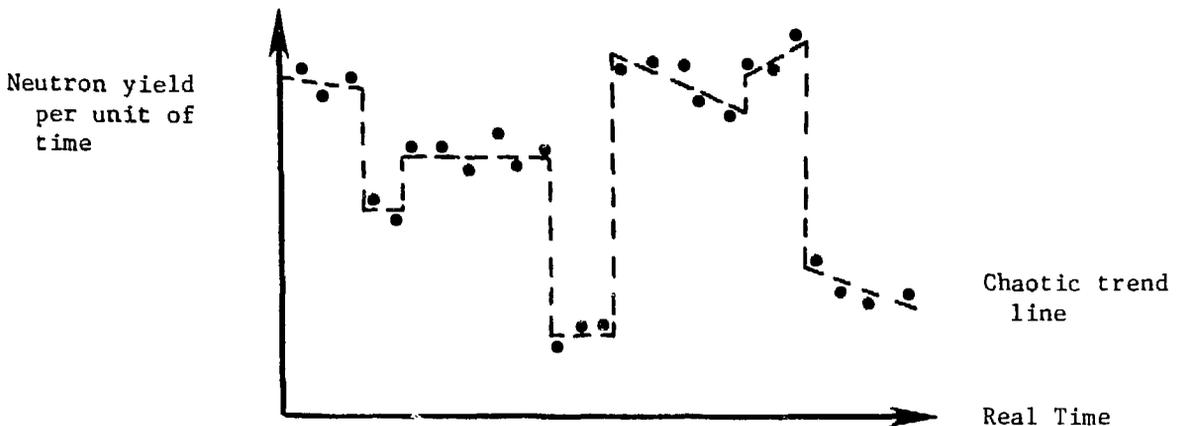


Figure 3.1: Examples of accumulated "data" from hypothetical neutron fluence measurements. See Ex. 3.1.

While numerical-valued random phenomena are encountered quite naturally in many realistic situations, there are also instances where the outcomes of a sampling procedure are more qualitative than quantitative in nature. What does one do then? The following example illustrates this point and indicates the approach which is normally used to deal with this issue.

Example 3.2

Consider an arbitrary experimental process (sampling procedure) where the outcome reduces to one of two possibilities: "on" or "off" (or something equivalent to this). The event space E consists of two elements, $e_1 = \text{"on"}$ and $e_2 = \text{"off"}$. One trivially notices that the integers 1 and 0 could be used without ambiguity to represent this apparently qualitative situation. Then, let $e_1 = 1$ and $e_2 = 0$ (of course the assignment is not unique). This apparently simple step of relating, for example, the status of a bipolar switching device to the integers 0 and 1 opens up the possibility of employing binary algebra for analysis. It is from this realization that the concept of a digital computer was born.

In more general terms, it turns out, for all practical purposes, that numbers can always be used to characterize the outcomes of statistical sampling procedures. Mathematicians point out that there are limitations associated with the requirement that the parameter sets so generated be amenable to the definition of probability. This is rarely a problem of concern for us, so we avoid the issue in this discussion.

To simplify matters, we first consider the entire class of events which can be characterized by the specification of single parameters. The characterization of events e in event space E in terms of a collection of representative numerical values is assumed to be accomplished by a function, say X , which operates on elements e in E in such a way that there is a particular correspondence between an event e and the real number $x = X(e)$. The function X , which has as its domain the event space E and yields real numbers such as x , is called a random variable. It is very important to grasp the concept that the random variable is the function X , not the individual values, $x = X(e)$, which this function produces when it operates on E . In routine practice, however, this distinction between a variable and the values it can assume is often blurred. For convenience, this will often be done in the present work. Nevertheless, the reader should always keep this distinction in mind. In a sense, the nomenclature "random variable" is somewhat misleading. A better choice would be "random function." However, we choose to adhere to the traditional convention. Events which can be specified by a single parameter in this way are designated single-variable events. Clearly, any number of other functions, e.g., F, Y, Z, \dots , could be

established on E , depending upon the problem at hand. This is illustrated in the following example.

Example 3.3

Consider a neutron activation experiment in which a number of very similar cylindrical disks with uniform material content are irradiated. For each sequential irradiation, the experimenter reaches into the sample container and selects a labeled sample "at random." Three obvious possibilities exist for single-variable categorization of the process of selecting a sample (the selected sample is an event): diameter (D), thickness (T), and mass (M). The functions D , T , and M are random variables by definition, each operate on the event space consisting of the complete sample collection. The real value provided by function D , when a particular sample is selected, is d . Likewise, values t and m are produced by functions T and M , respectively. The experimenter is free to employ the random variable which best satisfies his particular needs in order to represent the outcome of the sample-selection process. For cross-section calculation, the mass M is probably the one to favor. However, diameter D and thickness T impact upon various important corrections, and they may also have to be considered by the investigator.

Analysis of errors for the parameters that are chosen to describe the outcome of an experiment involves the implicit assumption that their observed frequency distributions (in samples taken from the ensemble of allowed values) are governed by known or assumed statistical laws, and, hence, that these parameters satisfy the conditions needed to designate them as random variables.

Next we consider the dimensionality of random variables. Their dimensionalities or cardinalities are distinct from those of the underlying event spaces E with which they are associated. A random variable X has dimensionality which is specified by the cardinality of the particular set of real numbers that corresponds to all the functional values obtainable for the random function X operating on E . E itself may be infinite, countable, or nondenumerable. This point is further clarified in the ensuing discussions and examples. Finite random variables are functions X which take on a finite set of values $x = X(e)$, for $e \in E$. Countable or denumerable random variables involve infinite sets of functional values which correspond one-to-one to the set of all integers. Finite and countable random variables are more generally designated as discrete random variables. The discrete values x_1, x_2, \dots are known as the jump points of X on E . Uncountable or nondenumerable random variables, more commonly called continuous random variables, are real-valued functions whose functional values form uncountable sets. The sets of numerical values produced by these functions can be mapped one-to-one into the set of all real numbers. The outcomes of many

measurement processes for realistic physical situations are usually treated this way (e.g., weight, length, time). Hybrid random variables must be used whenever the functional value spaces contain various components (subspaces) of different cardinality. All sorts of combinations are possible. The following hypothetical example represents a rather simple one.

Example 3.4

Consider the quantum mechanical problem of a particle interacting with a finite potential well. We assume existence of both bound and unbound states (solutions to Schrödinger's equation). The hypothetical sampling process is that of identifying which state the particle is occupying at a particular instant (although we realize that this is physically impossible according to the Heisenberg Uncertainty Principle). The allowed states form an event space, and particle energy E is a suitable representative random variable. It is a fact that the bound ($E < 0$) states will be quantized (i.e., discrete), and, furthermore, that there are a finite number of them in this situation. The unbound states are uncountable; thus, the allowed energy values form a continuum with ($0 \leq E < \infty$). In this example, particle energy is a hybrid random variable, since its allowed domain involves discrete values and a continuum region. Note that in this example the cardinalities of the event space (states) and the corresponding random-variable description (particle energy) are the same.

Not all events of practical interest can be characterized by the use of single random variables. An extension to multi-dimensional random variables, more commonly referred to as joint random variables or random vectors, is required. In practice, the number of required parameters is finite. Thus, we consider as an n-component random variable the collection of real-valued functions (F_1, \dots, F_n) having the property that when they operate on event e in E , they produce the array of numerical values (f_1, \dots, f_n), according to the rules $F_1(e) = f_1, \dots, F_n(e) = f_n$. For convenience, we use the vector notation $\bar{F} = (F_1, \dots, F_n)$, $\bar{f} = (f_1, \dots, f_n)$, and $\bar{F}(e) = \bar{f}$. Previously discussed concepts of dimensionality carry over to multi-component situations in an obvious way. Clearly, arbitrary mixtures of discrete and continuous random-variable components (hybrids) are also permitted in multi-dimensional problems. The essential feature is that well-defined relationships must exist.

The reader should surmise from the preceding discussion that the notion of random variables serves as a convenient aid in the numerical specification of events. By designating functional relationships between event space and corresponding spaces of numerical values, one avoids having to list tables of

numerical values explicitly and thereby gains the possibility of utilizing many useful features of real-valued function theory.

3.3 Random-Variable Functions

We have seen in the previous section that random variables are essentially "well-behaved" functions which operate on event spaces and yield, in a well-defined manner, real-number values that serve to unambiguously identify events. Thus, for instance, if E is an event space with elements e , and X is a random variable which serves to characterize E , X generates a new space S of real numbers $x = X(e)$ which acts as an image of E . An important point, first made in Section 3.2, should be reemphasized here: Transformation from an event space to an image space containing real numbers enables numerical analysis to be performed. In this section we focus on random variables, such as X , and their domains, such as S , but remain cognizant of the fact that they are really surrogates for the more fundamental event spaces that they represent. Generalization of this concept to joint random variables is straightforward.

Once the image space of real numbers is established, we are free to define functions on this space. Henceforth, we will usually bypass the previous practice of distinguishing function names from functional values by using lower case letters for values and capital letters for functions, though we may occasionally resort to this formal convention. Rather, the distinction will either be explicitly stated or assumed to be apparent from the context of the discussion. In this presentation, we will be interested in three special classes of functions of random variables. The distinctions are based mainly on their roles in the theory. We begin the discussion of these here and resume it in Chap. 6 and in later reports of this series. Since we wish to perform analysis using these functions, they must be "well-behaved" to the extent necessary to permit such mathematical operations to be undertaken as may be required in a particular application, e.g., differentiation or integration. Mathematicians labor over such matters, but we will generally be able to avoid detailed technical consideration of this issue because those functions of random variables which interest us for most practical applications tend to be well-behaved; otherwise, they would not have been found to be useful in the first place. The interested reader will find extensive discussions along this line in certain of the listed references (e.g., Fis 63 and Zeh 70). Generally, the functions to be considered will be bounded and devoid of singularities. Discontinuities, which play havoc with differentiation, are not uncommon, so care is required in this regard. There is also a frequent need to construct augmented spaces. For example, if a function f of a random variable x is physically defined only over the interval (x_{\min}, x_{\max}) , it may be useful to augment the space to $(-\infty, +\infty)$, namely, the set of all real numbers, and simply provide f with a dummy value, e.g. zero, outside the intended range. When this is done, discontinuities may be introduced, and they have to be handled with care.

We now state a fundamental fact without technical elaboration: Functions of random variables can also serve as random variables, provided that they are suitably well-behaved. However, not all functions which we

will choose to define on random variable spaces are intended to serve as random-variable functions. This point may seem trite, but its consequences are of great importance to the theory. From a physical point of view this represents a very reasonable state of affairs. One would expect that one could define quantities which would be distributed in statistically predictable ways by the operation of well-behaved functions on random-variable spaces. That this in fact can be done is, for example, essential to the notion of error propagation. It also enables statistical analyses to be performed using specially defined functions that are referred to as statistics. This topic will be addressed in later reports in this series.

Although we will not make any use of it in this volume, it is worthwhile for one to be aware of the notion of an indicator function for random variables x belonging to set A . The indicator function I_A is defined as

follows: $I_A(x) = 0$ if $x \in E$ but $x \notin A$, i.e., if $x \in \bar{A}$; however, $I_A(x) = 1$ if $x \in E$ and $x \in A$, i.e., $x \notin \bar{A}$. Thus, given E and $A \in E$, I_A maps every point in E into either 0 or 1, depending upon whether or not $x \in A$. Indicator functions are sometimes useful in proving theorems from probability theory.

The three classes of random-variable functions to be emphasized here cover most situations of applied interest. For lack of a better word, we refer to the first category as random-variable functions or transformation functions. These functions are intended to represent certain quantities, derived directly from other random variables, which correspond either to real physical parameters or to abstract ones which behave randomly and serve useful analytical purposes. The random-variable arguments for these functions can be discrete, continuous, or hybrid in nature. The simplest case might be represented by $y = f(x)$, in which the space X of real values x is mapped into an image space Y of real values y by the function f . In many instances, f will be sufficiently well-behaved so that there exists an inverse function, $g = f^{-1}$, with the property that all points of y in Y can be mapped one-to-one into equivalent points x in X . Such one-to-one mapping, with the existence of an inverse function, does not exist for all types of random-variable functions we will be interested in, so the reader must not assume its validity a priori. In the most general case, such a transformation can be represented by the mapping of vectors $\bar{x} = (x_1, \dots, x_n)$, belonging to an n -fold space, X , into vectors $\bar{y} = (y_1, \dots, y_m)$, belonging to an m -fold space Y , by means of a set \bar{f} of m functions (f_1, \dots, f_m) . Thus, $y_i = f_i(\bar{x})$ for $i = 1, m$, or equivalent $\bar{y} = \bar{f}(\bar{x})$.

Example 3.5

Consider again the sample collection described in Example 3.2. Let d represent diameter, t represent thickness, and m represent mass. There is a functional relationship between these parameters, namely, $m = m(d,t;\rho)$. Here, as is commonly done, we allow m to represent both the random variable (a function) and a typical value. The density ρ is assumed to be constant. Since it is not treated as a random variable like d and t , we choose to separate it from them by a semicolon. For simplicity, we suppress ρ and write the function m as

$$m(d,t) = \pi\rho d^2 t/4 \quad .$$

Thus, for $d_1 \leq d \leq d_2$, and $t_1 \leq t \leq t_2$, we have

$$\pi\rho d_1^2 t_1/4 \leq m \leq \pi\rho d_2^2 t_2/4 \quad .$$

The random variable m is treated as derived in this example, but it can also be directly measured (i.e., it can be deduced from the sampling process called weighing). The function m of d and t has no well-defined inverse in that knowledge of m does not yield a unique pair (d,t) . This is consistent with the common sense fact that a measurement of the mass cannot provide unambiguous information about the sample geometry. Knowledge about d , t , and ρ provides us with a far greater understanding of the nature of a particular sample than does the weight alone.

Data analysis or data reduction generally involves transformation from one collection of random variables to a lesser number of derived quantities. Inverses usually do not exist. Such transformations generally lead to loss of information. This is illustrated in the following example.

Example 3.6

Neutron cross sections are significant abstract parameters which are not directly measurable. For a simple hypothetical experiment, the cross section σ is a random variable derived, for example, from the following measurable random variables: Observed yield y , neutron fluence Ψ , and sample atoms n . Thus, $\sigma = \sigma(y,\Psi,n) = y/(\Psi n)$. It is evident that if only the cross section σ is given, we can say nothing about the experiment that produced it.

We turn our attention next to two classes of functions, involving random variables, that serve entirely different purposes. These functions cannot be interpreted as "random-variable functions" in the sense described above. The first of these classes is that of density functions. Density functions involve transformations from single or joint random-variable spaces to subsets of the set of all non-negative real numbers. Density functions are not intended to represent observables, or even certain abstract parameters somehow related to physical ones (e.g., cross sections). We shall see that, in fact, they are generally used in the quantification of probability. Since we have not yet introduced the notion of probability, this class of functions will not be discussed here in a probability context. The arguments of density functions can be discrete, continuous, or hybrid random variables. The term mass function is sometimes used to designate density functions with discrete arguments, since these are not true density functions. For convenience, we will use "density function" to denote both. Thus, for the single-variable case we have f operating on x to yield $f(x)$. Given f and a discrete collection of values (x_1, x_2, \dots) , one obtains the set of corresponding mass values (f_1, f_2, \dots) . For continuous-variable situations, if $x_{\min} \leq x < x_{\max}$, then for each x a corresponding density value $f(x)$ is generated.

Example 3.7

A function f , defined for random variable x such that $f(x) = c$ (a constant) for each x in the range $x_{\min} \leq x < x_{\max}$, is called a constant density function. We could easily augment the range of f to encompass the domain of all real numbers by declaring that $f(x) = 0$ for $-\infty \leq x < x_{\min}$ and $x_{\max} < x \leq +\infty$. Note the discontinuities which are introduced at $x = x_{\min}$ and $x = x_{\max}$ by this procedure.

Density functions involving several random variables are called multivariate or joint density functions. If \bar{x} represents the random variable collection (x_1, \dots, x_n) , we speak of $f_{12\dots n}$ as a joint density function with functional values $f_{12\dots n}(x_1, \dots, x_n)$ or $f_{12\dots n}(\bar{x})$. Unlike the situation for transformation functions, arrays of density functions are not defined. They are always scalar quantities.

Density functions are employed in conjunction with another class of functions of random variables called distribution functions. In fact, distribution functions are often defined in terms of corresponding density functions. In the following discussion, it will be assumed for convenience that the random-variable spaces are augmented as needed in order to be equivalent to the set of all the non-negative integers or to the set of all real numbers. If x is a single random variable and f is a density function, then the corresponding distribution function F is defined as

$$F(z) = \sum_{x < z} f(x), \quad (3.1)$$

if the domain of x is discrete, or

$$F(z) = \int_{-\infty}^z f(x) dx \quad (3.2)$$

if the domain of x is continuous. Referring to Fig. 3.2, for instance, the values of F corresponding to a particular z are equivalent to the shaded areas in Figs. 3.2(a) and 3.2(b), respectively, for a discrete and continuous case. Owing to the non-negative nature of density functions, and to the definition of distribution functions, it holds that $F(z) \geq 0$ for all z , and $F(z_b) \geq F(z_a)$ if $z_b \geq z_a$. The reader ought to realize that the term "distribution function" is frequently used interchangeably with "density function." In fact, we will sometimes do this in the volume, in conformance with standard practice. The distinctions indicated above must be kept in mind, however.

Extension to multivariate situations is straightforward. The distribution function $F_{12\dots n}$ associated with a continuous joint density function $f_{12\dots n}$ is defined as

$$F_{12\dots n}(z_1, \dots, z_n) = \int_{-\infty}^{z_1} dx_1 \dots \int_{-\infty}^{z_n} dx_n f_{12\dots n}(x_1, \dots, x_n). \quad (3.3)$$

A continuous density function is said to be uniformly well-behaved over a given space of values (z_1, \dots, z_n) if the following relationship exists between it and its corresponding distribution function in that region:

$$\partial^n F_{12\dots n}(z_1, \dots, z_n) / \partial z_1 \dots \partial z_n = f_{12\dots n}(z_1, \dots, z_n) \quad (3.4)$$

Density functions are said to be factorable in the variable x_i if there exist two functions f_i and $f_{12\dots i-1, i+1, \dots, n}$ such that

$$f_{12\dots n}(x_1, \dots, x_n) = f_i(x_i) f_{12\dots i-1, i+1, \dots, n}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \quad (3.5)$$

It is readily shown that the corresponding distribution function is also factorable; thus,

$$F_{12\dots n}(z_1, \dots, z_n) = F_i(z_i) F_{12\dots i-1, i+1, \dots, n}(z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_n) \quad (3.6)$$

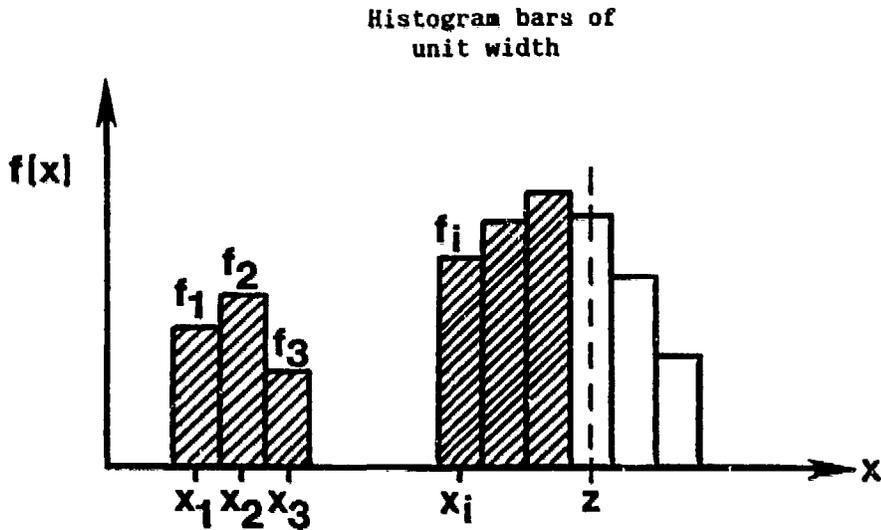


Figure 3.2(a): Histogram represents a discrete density (or mass) function $f_i = f(x_i)$. Shaded area represents the corresponding distribution function $F(z) = \sum_{x < z} f(x)$.

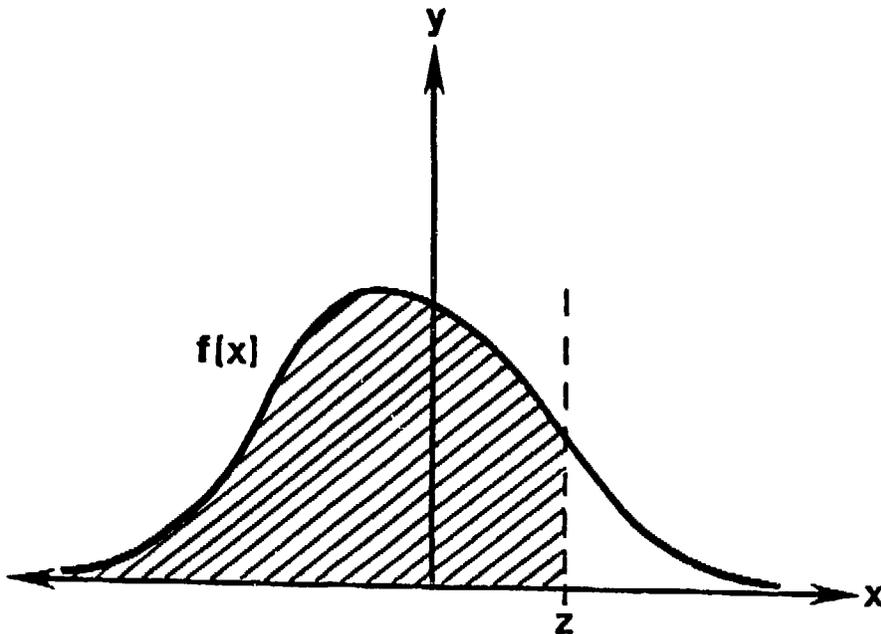


Figure 3.2(b): Curve represents a continuous density function $y = f(x)$. Shaded area represents the corresponding distribution function $F(z) = \int_{-\infty}^z f(x)dx$.

where

$$F_i(z_i) = \int_{-\infty}^{z_i} f_i(x_i) dx_i, \quad (3.7)$$

and

$$\begin{aligned} F_{12\dots i-1,i+1,\dots n}(z_1,\dots,z_{i-1},z_{i+1},\dots,z_n) = \\ \int_{-\infty}^{z_1} dx_1 \dots \int_{-\infty}^{z_{i-1}} dx_{i-1} \int_{-\infty}^{z_{i+1}} dx_{i+1} \dots \\ \int_{-\infty}^{z_n} dx_n f_{1\dots i-1,i+1,\dots n}(x_1,\dots,x_{i-1},x_{i+1},\dots,x_n). \end{aligned} \quad (3.8)$$

The factorability of distribution functions turns out to be an important issue when considering marginal probability and independence.

Finally, we note that probability theory deals with normalizable density functions having corresponding distribution functions which are bounded.

4. BASIC CONCEPTS OF PROBABILITY

The preceding chapters of this report serve to introduce the reader to certain basic concepts and mathematical tools which are essential for an understanding of probability theory. In the present chapter we come to grips with the concept of probability itself. Analogous to the approach used in Chap. 2, we begin in Sec. 4.1 with an intuitive discussion of probability. This is followed by a more technical exposition in Sec. 4.2. Finally, in Sec. 4.3, we consider some of the techniques used in probability calculations. This development emphasizes probability as it applies explicitly to event spaces. In Chap. 5, probability is considered in terms of density and distribution functions which operate on random-variable spaces. This exposition also strives to prepare the reader for the next report in this series, in which the relationship between probability theory and data uncertainties will be elucidated.

4.1. Intuitive Approach

Readers of this report will have come to realize by now that this exposition has, to this point, avoided discussion of the central concepts of probability theory, focusing instead on preliminary matters. This approach is intentional. In my opinion, many expositions on this subject tend to thrust the main ideas of the theory upon the reader before undertaking to provide the mathematical "infrastructure" needed for them to be properly assimilated. Automotive enthusiasts will agree that a sophisticated engine requires a carefully designed and adjusted chassis, an aerodynamic body, an efficient fuel and air delivery system, etc., in order to manifest its true capabilities. Of course, once the supporting components of a vehicle are in place, its heart, the engine, must then be installed. We proceed in this section to bring the theory of probability to life by adding to it the central working concept, namely, the notion of frequency of occurrence of certain well-defined attributes in random sampling from simple-event spaces.

The following example illustrates informally what is meant by frequency of occurrence, and it explores the link between this concept and some other previously-discussed notions.

Example 4.1

Consider an electromagnetic isotope separator. We place 100 mg of chemically pure elemental copper in an ion source and set the machine into operation. This separator is assumed to possess a well-established collection efficiency that is arbitrarily taken to be 1%. During operation, individual atoms of copper are randomly extracted from the ion source and injected into the separator. A particular copper atom, once released from the source material, is either effectively lost (99% of the time) or makes a successful journey to one of two collectors (1% of the time). Copper in elemental form has two isotopes, ^{63}Cu and

^{65}Cu . For successful journeys, it is assumed that ^{63}Cu atoms always go to Collector No. 1, while, analogously, ^{65}Cu atoms go to Collector No. 2. After 10 mg of the source material (10%) have been processed, the contents of Collectors No. 1 and No. 2 are weighed and it is found that Collector No. 1 contains ~ 70 micrograms of copper and Collector No. 2 contains ~ 30 micrograms of copper. Correcting for isotopic mass difference, it is noted that this result is consistent with the known natural isotopic abundances (Tul 85): ^{63}Cu (69.17 \pm 0.02%) and ^{65}Cu (30.83 \pm 0.02%).

Let us examine this process in terms of the ideas developed so far in this report. The random sampling process is clearly the extraction of individual atoms from the mass of material placed in the separator ion source. Each individual atom extracted represents a simple event belonging to a sampling space. Our particular sampling exercise produces a sample equivalent in size to ~ 10% of the entire sample space. In this instance the sample is large enough to serve as a population which is representative of the entire sample space. The observation itself consists of identifying whether a particular extracted atom (simple event) is ^{63}Cu or ^{65}Cu , and whether or not it makes a successful journey through the separator to a collector. If the journey through the separator is indeed successful, then it is essentially predetermined by the design of the instrumentation as to which collector will receive a particular isotopic species. In reality, we do not observe the outcome of individual atom extractions. That would be impossibly tedious. Instead, we weigh the collected material and indirectly deduce the total numbers of each isotope which have made a successful journey. We therefore are led to consider compound events, defined according to the two particular attributes which are of interest to us. Each compound event is a subspace of the entire space of simple events, and each such subspace contains a huge number of simple events.

The compound events can be identified by two random variables: mass number (A) and transit history (H). Thus, compound events (A,H) are: (63,0), (63,1), (65,0), (65,1), where H=0 (failure) or 1 (success). A and H are joint random variables, as described in Chap. 3. The transit success-to-failure frequency ratio for any particular sample (collection of extracted copper atoms) is determined by the efficiency of the separator. An efficiency of 1% implies that for a very large number of extracted atoms, about 1 out of every 100 will reach a collector. However, for a particular sample of small size, the observed success-to-failure ratio could be quite different from 1/99: In this example, however, the sample we have taken is a substantial portion of the entire event space, so it happens that the observed success-to-failure ratio is essentially indistinguishable from the 1/99 value characteristic of the entire space. Likewise, the

observed frequency ratio for isotope type of the extracted atoms will reflect the underlying isotopic abundance of elemental copper whenever the sample is large, as is the case here. That frequencies of occurrence of certain well-defined outcomes from sampling in fact tend to converge toward the underlying features of the entire sample space, as the size of the sample becomes large, is an observed fact of random phenomena in Nature. It would certainly have been pointless to develop a theory of probability if matters had been otherwise. This is merely an affirmation of our previously stated notion of statistical predictability. Figure 4.1 illustrates the current example. The simple-event space of individually extracted atoms (shown symbolically by dots) is partitioned according to attributes labeled by A and H. The allowed (A,H) pairs are mutually exclusive compound events. The final outcome (measured masses of copper in each particular collector) reflects the fact that each elementary event is equally likely in random sampling, and, therefore, the likelihoods of the various defined compound events are determined by relative frequencies with which elementary events are observed to possess the specific defined attributes that characterize the compound events in question.

The origin of probability theory as we know it today dates to the seventeenth century in France. Chevalier de Mere was known to be an ardent gambler. It seems that he was baffled by some questions concerning a game of chance which was then popular. He consulted the mathematician Blaise Pascal who, in turn, wrote to Pierre Fermat. This correspondence led to the earliest documented formulation of the theory. Great strides were made in its development in the eighteenth century, particularly as a result of contributions by Karl Gauss and Pierre Laplace. However, certain conceptual difficulties with the theory were not resolved until well into the twentieth century. Growth in the application of probabilistic methods has been very pronounced since the end of World War II, aided in no small measure by the advent of the digital computer.

Modern probability theory is based upon a series of postulates. However, since these are consistent with the notion of frequency of occurrence, introduced earlier in this chapter, we choose to pursue the frequency approach in this section and defer discussion of a more rigorous basis for the theory to Sec. 4.2.

First, consider a finite space E of simple events $e_1, e_2, \dots, e_i, \dots, e_n$. At the same time, define a sampling process which, for simplicity, consists of selecting a single simple event from E at random. Accordingly, one defines a scalar function P such that P assigns the value $P(e_i)$ to each event e_i . Furthermore, assume that $P(e_i) = 1/n$ for each e_i in E . This assumption is known as the Equal-Likelihood Postulate. The implication is that all the events of E are equally likely to occur when the sampling process is executed. Suppose A is a compound event consisting of one or more simple

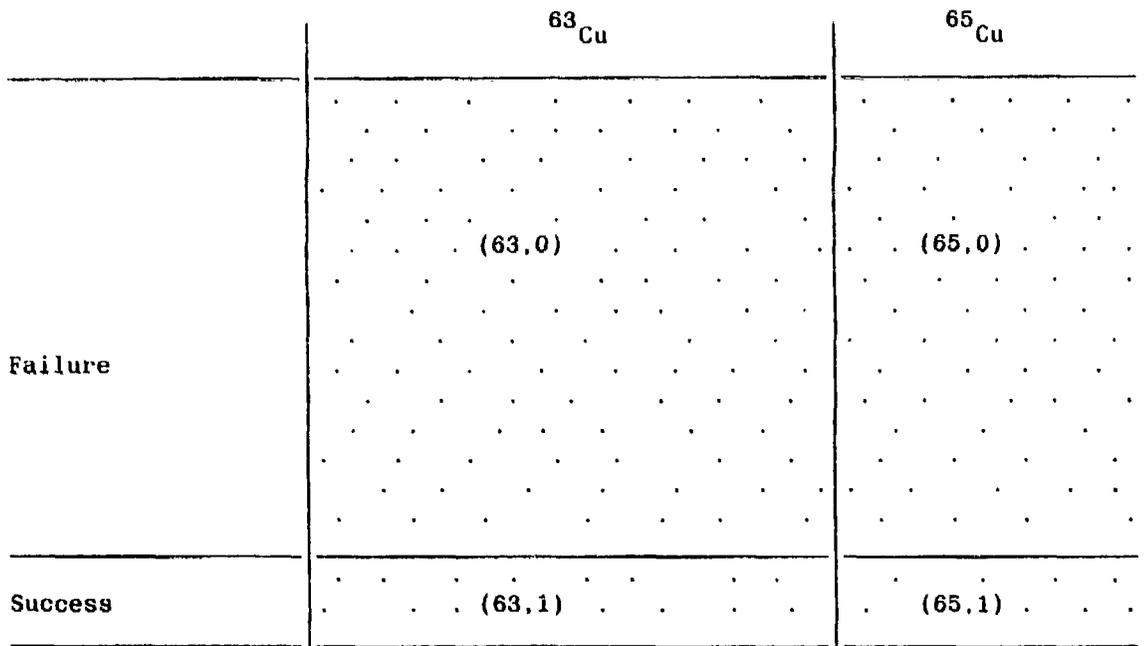


Figure 4.1: Event space for copper atoms processed by an isotope separator. Individual dots symbolize simple events, while the partition into various regions represent compound events. Each compound event contains many simple events. See example 4.1.

events of E which possess a particular attribute. To be definite, assume that A contains n_A distinct simple events E . The complement $\bar{A} = E - A$ contains $n - n_A$ events which by definition are distinct from those in A . The function P , when applied to A and \bar{A} , yields, respectively,

$$P(A) = n_A/n \text{ and } P(\bar{A}) = (n - n_A)/n.$$

Thus,

$$P(A) + P(\bar{A}) = 1,$$

or equivalently, $P(E) = 1$. Likewise, $P(\phi) = 0$. $P(A)$ represents the ratio of the number of simple events n_A with attribute characteristic of A to the total number of events n in E . The function P is the probability function, or simply probability, associated with a particular sampling procedure and particular event space E . It is very important to remember that the sampling procedure and the event space upon which it operates must both be considered in defining and calculating probability.

Example 4.2

Consider again the situation defined in Ex. 2.3. The simple events correspond to all of the distinct failure configurations for a circuit board with four IC components. In this example we concluded that there were 16 such configurations to consider (see Table 2.1). Under the category "Possibility A," we also defined five mutually exclusive compound events, corresponding to the particular attribute "number of defective IC components on the board". One of these compound events is "three-bad-component failure mode". According to Table 2.1, this compound event has 4 simple events associated with it. Thus, the probability of observing a "three-bad-component failure mode" when examining the board is $4/16$, or 0.25 , provided that all simple events are equally likely to occur.

For a finite sample space and a well-defined sampling rule, the definition of probability for various compound events belonging to the associated Borel field is unambiguous. The postulate of equal likelihood is almost universally assumed in applications, for this mathematical model appears to apply to a very large class of random phenomena in nature. For example, modern quantum theory relies very heavily on this postulate, since it is largely a statistically based theory. In those instances where it might appear to not be valid, it is generally possible to restore its applicability to the problem in question by seeking an alternative and more

fundamental definition of the simple-event space. It must not be forgotten that by definition each simple event must be unique, and therefore mutually exclusive from all other simple events of the sample space.

Difficulties can and do arise in probability analyses with finite sample spaces, owing to the fact that sampling from a finite space perturbs the nature of the space, unless the selected event e_i is "replaced" before undertaking subsequent sampling. The attendant problems do not reflect a fatal flaw in the theory, however, for they can be handled by being very careful with the counting procedures. The fundamental importance to probability theory of the counting rules discussed in Chap. 2 is now very evident.

Many applied problems require a mathematical model in which the sample space has such a large number N of simple events that it can, for all practical purposes, be considered infinite. We then run into a bit of a problem in defining probability in the manner discussed above. The probability of each simple event, $1/N$, is vanishingly small. In situations involving such large spaces, we are hardly ever interested in the probability of individual elementary events. Instead, we are concerned with the probability of occurrence of some attribute, i.e., with a compound event. Under these conditions probability can be conveniently envisioned in terms of observed frequency ratios. Assume that a sampling procedure is repeated a finite number of times, n , on an event space with a very large number of elementary events ($n \ll N$), and that for n_A of these times compound event A occurs. The frequency ratio $f(n_A, n) = n_A/n$ is clearly observable and, furthermore, tends to converge with increasing n . Therefore, we choose to interpret as the probability $P(A)$ the following limit:

$$P(A) = \lim_{n \rightarrow \infty} f(n_A, n) \quad (4.1)$$

Definition of probability according to Eq. (4.1) is considered to be unsatisfactory by most pure mathematicians, so they have elected to establish a theory of probability based on a minimal number of postulates, as discussed in Section 4.2. It is understandable that they should feel this way. In order for the limit indicated in Eq. (4.1) to exist, it must be proven that for any real number ϵ , no matter how small, there exists an integer m such that for all $n > m$, $|P(A) - f(n_A, n)| < \epsilon$. In fact, there is no way to prove that such a contention is always true for random phenomena, though there is a growing body of evidence to the effect that many natural processes do seem to behave in this way. In any case, the notion of frequency, or relative frequency, leads to what is known as a posteriori probability, namely a probability interpretation closely linked with experimentation. Even so, there remains the unresolvable dilemma that the assumption of equal likelihood for simple events is equivalent to acceptance of an a priori or postulated probability which can never be tested directly.

Example 4.3

The issue of the isotopic abundances of ^{63}Cu and ^{65}Cu in elemental copper, first discussed in Ex. 4.1, illustrates the manner in which the frequency notion of probability plays a role in the development of knowledge through experimentation. First, we take as premise that even if the universe is accepted to be finite, so that at any instant there exist well-defined values for the numbers of ^{63}Cu atoms (N_{63}) and ^{65}Cu atoms (N_{65}), it would be impossible to determine these numbers. We are actually quite certain that both N_{63} and N_{65} change with time (primarily via creation and destruction through nuclear reactions in stellar environments). What really concerns us is the relative abundance of these species on the planet Earth, an environment which we accept as reasonably stable (i.e., relatively "cold" from a nuclear point of view) and uniform (insofar as the isotopic constituency of the distributed elemental copper is concerned). Even then, explicit counting of all the copper on Earth is impossible, so we resort to examining finite samples of this material. Such samples are still of enormous size if we consider each individual atom to represent a simple event. Examination of a mass w of elemental copper with an isotope separator yields isotope value w_{63} and w_{65} for ^{63}Cu and ^{65}Cu , respectively. Let m_{63} , and m_{65} be the atomic masses of elemental copper, ^{63}Cu and ^{65}Cu , respectively, n , n_{63} , and n_{65} be the total numbers of elemental ^{63}Cu and ^{65}Cu atoms, respectively, given that w is the total sample mass. We are then led to assume, from an application of Eq. (4.1), that

$$P(^{63}\text{Cu}) \approx n_{63}/n \approx (mw_{63})/(m_{63}w)$$

and

$$P(^{65}\text{Cu}) \approx (mw_{65})/(m_{65}w).$$

Our knowledge of the ^{63}Cu and ^{65}Cu isotopic abundances is distilled from numerous such determinations, yet they are not known exactly. Regardless of future attempts at refinement, they will never be perfectly determined, not only because of the finiteness of the measured samples, but also because of measurement (sampling) imperfections.

The preceding discussions focused on consideration of a single compound event A. However, suppose that several compound events A_1, A_2, \dots, A_r are considered, and that they are pairwise mutually exclusive, i.e., that

$$A_i \times A_j = \phi \quad (i \neq j: 1, r).$$

Eq. (4.1) is applicable in that $P(A_i)$ is defined as the limit of the ratio n_{A_i}/n for very large samples n . If we define

$$A = \sum_{i=1}^r A_i,$$

i.e., as the union of all these events, then

$$P(A) = \sum_{i=1}^r P(A_i) \quad . \quad (4.2)$$

If the events A_1, A_2, \dots, A_r form a partition of E, then

$$\sum_{i=1}^r A_i = E,$$

and we have $P(A) = P(E) = 1$. Pairwise exclusivity is absolutely essential for Eq. (4.2) to be true. That is, if a simple event in E belongs to A_i , it cannot belong to another A_j ($j \neq i$). Section 4.3 discusses procedures for calculating probabilities of compound events which are not mutually exclusive. Knowledge of the rules of Boolean algebra, as discussed in Chap. 2, is essential for handling such cases.

The rules for probability calculation are meaningless unless one is dealing with stable, well-understood event spaces. We alluded to this condition in Ex. 4.3. In applications of probability theory (e.g., in experimental measurements), great care must be taken to perform investigations under controlled, reproducible conditions if one hopes to extract meaningful results. This is illustrated in the following example.

Example 4.4

There are good reasons to believe that the isotopic abundances of ^{63}Cu and ^{65}Cu in elemental copper are quite uniform throughout the planet Earth. Consequently, in measurements of these abundances, we can focus attention on good separation procedure and not worry about the material itself. What about

determination of the ${}^6\text{Li}$ and ${}^7\text{Li}$ isotopic abundances for elemental lithium? Here we have to be very careful to obtain documentation on the exact history of the batch of material on which an intended measurement of "natural" isotopic abundance is to be performed. During the last several decades, extensive separation work on lithium material has been undertaken in conjunction with various nuclear-energy programs. Some partially separated lithium material (generally partially depleted in ${}^6\text{Li}$) has worked its way back into commercial inventories. Much the same can be said about boron, where ${}^{10}\text{B}$ and ${}^{11}\text{B}$ isotopic ratios have been seen to vary in supposedly "elemental" material.

In summary, we see that probability theory offers a mathematical model which has been shown empirically to be applicable to the analysis of random phenomena. In Nature, frequency of occurrence of events in random sampling from event spaces is seen to be conceptually equivalent to what is defined as probability in the theory. However, in the theory itself, probability is postulated in terms of abstract functions operating on well-defined event spaces, thereby avoiding the ambiguities attendant upon the concept of frequency. The notion of experimental error or uncertainty is intimately related to the observation that repeated experimentation under supposedly well-controlled conditions inevitably leads to a sequence of unpredictable outcomes. Obviously, there exists a link between error and probability. This issue will be addressed formally in the second report of this series.

4.2 Axiomatic Approach

The challenge that faced mathematicians from the outset was that of developing a theory of probability which would not only be rigorous and consistent with other branches of mathematical theory, but would also apply to many observable situations. That is, the theory had to conform to the empirically deduced behavior of random phenomena. This was not an easy task, and its pursuit has generated vigorous disputes over a variety of conceptual and logical problems, some of which rage unsettled to this day. Nevertheless, for most applied purposes, the foundations of probability theory are now considered to be in acceptable form.

Consider an event space E . For convenience we first assume that it contains a finite number of simple events. Let A and B be two events belonging to E (either simple or compound events). Quite remarkably, it has been formally demonstrated that the complete theory of probability can be derived from the following three axioms attributed to A. Kolmogorov:

Axiom I: To each event $A \in E$ there is a non-negative number $P(A)$, its probability. Thus, $P(A) \geq 0$.

[Existence].

Axiom II: The certain event E (entire event space) has unit probability. Thus, $P(E) = 1$.

[Normalization].

Axiom III: If A and B are mutually exclusive events, that is if $A \times B = \phi$, then the probability of the union (or sum) of A and B is given by

$$P(A + B) = P(A) + P(B).$$

[Additivity].

These axioms are entirely consistent with the observable behavior of natural random phenomena, as is shown in Table 4.1. However, they in no way tell us how to go about actually calculating probability. For present purposes, the now-familiar Equal-Likelihood Postulate provides us with the calculational tool we require:

Postulate: [Equal-Likelihood Postulate]

If E is a finite space of simple events, and $A \in E$, then

$$P(A) = \frac{N(A)}{N(E)} \quad , \quad (4.3)$$

where $N(E)$ is the number of elementary events in E and $N(A)$ is the number of those events with attribute A. An equivalent way to express this postulate is to say that the probabilities of all simple events of E are equal and have the value $1/N(E)$.

In building a theory from the basic axioms, one must be very careful to avoid certain logical pitfalls. For example, if $A \in E$ and $P(A) = 1$, one cannot surmise that $A = E$ from an application of Axiom II. It might happen that $\bar{A} \neq \phi$, yet $P(\bar{A}) = 0$. This state of affairs is entirely consistent with both Axioms I and II. However, it is possible to logically deduce a large number of valid consequences from these axioms. A few of the important theorems will now be stated, mostly without proof.

Theorem: The impossible event ϕ has zero probability, however, the converse is not true.

Theorem: If $A \in E$, then

$$P(A) \leq 1 \quad . \quad (4.4)$$

Theorem: If $A \in E$, then

$$P(\bar{A}) = 1 - P(A) \quad . \quad (4.5)$$

Table 4.1 Equivalence Between Axioms of Probability Theory and Intuitive Observations

Intuitive Observation	Axiomatic Equivalent
I. The frequency of occurrence of random events with a particular characteristic oscillates around some fixed value when the number of trials becomes large.	For each $A \in E$ there is a value $P(A)$, called its probability, such that $P(A) \geq 0$.
II. If every possibility which we might sample is endowed with the same characteristic, we can be 100% certain that any individual trial will produce this characteristic.	The probability of the sure event E is unity. Thus, $P(E) = 1$.
III. There is an additive property in the frequency of occurrence of exclusive random events.	If $A \in E$ and $B \in E$, and $A \times B = \phi$, then $P(A + B) = P(A) + P(B)$.

Theorem: If $A \in E$ and $B \in E$, then

$$P(A \times B) \leq P(B).$$

Theorem: [Poincare Law of Probability Addition]

If $A \in E$ and $B \in E$, then

$$P(A + B) = P(A) + P(B) - P(A \times B) \quad (4.6)$$

The proof is straightforward. We first note that

$$A + B = A + (B - A \times B)$$

and

$$B = A \times B + (B - A \times B).$$

Furthermore, it happens that the particular events A and $B - (A \times B)$ are mutually exclusive, as well as are $A \times B$ and $B - (A \times B)$. The reader might find a Venn diagram such as Fig. 2.8 useful in visualizing these results. An application of Axiom III yields

$$P(A + B) = P(A) + P(B - (A \times B))$$

and

$$P(B) = P(A \times B) + P(B - (A \times B)).$$

The theorem is then proved by simple algebraic manipulations.

The following three theorems are closely related to the preceding one:

Theorem: [Boole's Inequality]

If $A \in E$ and $B \in E$, then, in general

$$P(A + B) \leq P(A) + P(B) \quad (4.7)$$

To prove this, we refer to the Poincare Law of Probability Addition, namely, that

$$P(A + B) = P(A) + P(B) - P(A \times B).$$

Now,

$$A \times B \in E;$$

thus,

$$P(A \times B) \geq 0$$

by Axiom I. A simple application of algebraic rules for inequalities proves the theorem.

The preceding two theorems can be generalized to a finite collection of events in E , namely, A_1, A_2, \dots, A_n :

Theorem: [Generalized Poincare Law of Probability Addition]

$$\begin{aligned}
 P\left(\sum_{k=1}^n A_k\right) &= \sum_{k=1}^n P(A_k) - \sum_{\substack{k_1, k_2=1 \\ (k_1 < k_2)}}^n P(A_{k_1} \times A_{k_2}) \\
 &+ \sum_{\substack{k_1, k_2, k_3=1 \\ (k_1 < k_2 < k_3)}}^n P(A_{k_1} \times A_{k_2} \times A_{k_3}) + \dots \\
 &+ (-1)^{n+1} P\left(\prod_{k=1}^n A_k\right). \tag{4.8}
 \end{aligned}$$

Theorem: [Generalized Boole's Inequality]

$$P\left(\sum_{k=1}^n A_k\right) \leq \sum_{k=1}^n P(A_k) \tag{4.9}$$

The following theorems are also useful:

Theorem: If $A \in E$ and $B \in E$, then

$$P(A - B) = P(A \times \bar{B}) = P(A) - P(A \times B).$$

Refer to Fig. 2.8. It is evident that

$$A - B = A \times \bar{B},$$

$$A = A \times B + A \times \bar{B},$$

and furthermore that

$$(A \times B) \times (A \times \bar{B}) = \phi.$$

Thus,

$$P(A) = P(A \times B) + P(A \times \bar{B}).$$

Note that if $B \in A$, then

$$A \times B = B$$

and

$$P(A - B) = P(A) - P(B).$$

Theorem: If A_1, A_2, \dots, A_n all belong to E and are pairwise mutually exclusive, then

$$P\left(\sum_{k=1}^n A_k\right) = \sum_{k=1}^n P(A_k) \quad . \quad (4.10)$$

This theorem is a special case of Eq. (4.9), Boole's Inequality.

Theorem: If A_1, A_2, \dots, A_n form a partition of E , then

$$P\left(\sum_{k=1}^n A_k\right) = 1.$$

This theorem is proved by noting that

$$\sum_{k=1}^n A_k = E \quad ,$$

from the definition of a partition. Then Axiom II can be applied.

Theorem: If A_1, A_2, \dots, A_n form a partition of E , and $B \in E$, then

$$P(B) = \sum_{k=1}^n P(B \times A_k) \quad .$$

Extension of the theorem to include denumerably infinite sequences of subsets A_1, A_2, \dots is relatively straightforward. This is generally accomplished by amending Axiom III as follows:

Axiom III': If E has a denumerably infinite number of elementary events, and A_1, A_2, \dots is a denumerably infinite sequence of pairwise mutually-exclusive events of E , then

$$\left(\sum_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k) \quad .$$

This augmented version of Axiom III allows us to prove theorems such as the following:

Theorem: If A_1, A_2, \dots is a nondecreasing sequence of events in E , and

$$A = \bigcup_{k=1}^{\infty} A_k,$$

then

$$P(A) = \lim_{k \rightarrow \infty} P(A_k) .$$

Theorem: If A_1, A_2, \dots is a nonincreasing sequence of events in E , and

$$A = \bigcap_{k=1}^{\infty} A_k,$$

then

$$P(A) = \lim_{k \rightarrow \infty} P(A_k) .$$

The reader who wishes to pursue the development of other theorems based on the axioms of Kolmogorov can refer to one or more of the textbooks given in the bibliography, e.g., Par 60, Fis 65, Tuc 67, and Zeh 70.

Another notion which proves to be very useful on occasion is that of marginal probability. If the occurrence of an event can be classified according to multiple criteria, then the term marginal probability is used whenever one or more criteria are ignored in the classification. Suppose $A_1, \dots, A_n; B_1, \dots, B_m$; and C_1, \dots, C_e are three partitions of E . They represent three different classifications. The marginal probability of A_i and C_k is

$$P(A_i \times C_k) = \sum_{j=1}^m P(A_i \times B_j \times C_k).$$

The marginal probability of C_k is

$$P(C_k) = \sum_{i=1}^n \sum_{j=1}^m P(A_i \times B_j \times C_k) = \sum_{i=1}^n P(A_i \times C_k) = \sum_{j=1}^m P(B_j \times C_k),$$

and so on.

If we move into the realm of uncountable event sets, matters become somewhat more complicated. For a great many cases, most of the rules we have established apply without difficulty. The essential point is that the event sets must have well-defined measure and thus be probabilizable. These difficulties can be resolved in the realm of measure theory, and the subject is treated in some of the texts listed in the bibliography, e.g., Fis 63. One obvious candidate for revision, when dealing with infinite event sets

(countable or uncountable), is the Equal-Likelihood Postulate. Eq. (4.3) must be revised to read

$$P(A) = \frac{M(A)}{M(E)} \quad (4.11)$$

where "M" designates a function which provides the appropriate measure of an event set of the type under consideration. In particular, event sets which are uncountably infinite, yet still probabilizable, are those where equally likely simple events can be represented by points in continuous random-variable spaces, while compound events are represented by well-defined regions in these spaces. Then, the measure of an event A, M(A), is simply the "volume" V_A of the corresponding finite region of the random-variable space. Even if M(E) is infinite, it is still reasonable to speak of the relative probability P(A)/P(B) of two events A and B, if M(A) and M(B) are both finite.

As previous indicated, the basic axioms of probability do not tell us how to assign probabilities to events; they merely set forth certain conditions which have to be met. A probability law is a specific functional law which enables one to calculate probability. At this level, the distinction between mathematicians and physicists becomes very apparent. Mathematicians tend to not concern themselves with the applicability of these laws. They merely accept the laws as interesting functions and explore the mathematical ramifications (deductive reasoning). Physicists, on the other hand, must discern which laws are applicable to various classes of observed phenomena. They often have to guess at the appropriateness of certain statistical laws, guided by experimentation and accumulated experience. Knowledge evolves from an initial position of ignorance (inductive reasoning). Chapters 7 and 8 examine in detail some of the probability laws which play a role in applied nuclear science.

The importance of the counting rules, first discussed in Chap. 2, stems directly from the Equal-Likelihood Postulate. Many of the probability laws applied in nuclear science can be derived from careful consideration of the way in which compound events are formed from equally likely simple events. This point was explored in Sec. 2.3. While the basic concepts of probability are not so difficult to grasp, we have already seen that the techniques of counting are difficult to master. Ash (Ash 70) describes the situation well: "Multiple counting is the nemesis of the combinatorial analyst." A fitting corollary to this principle is: "The physicist should likewise be on guard."

4.3 Probability Calculation

In most courses on probability the student is expected to develop skill in probability calculation through the analysis of a variety of problems. These exercises can be quite mind taxing, and they sometimes lead to rather remarkable results which may surprise the inexperienced student. Here, I would like to stress again that our objective is to understand uncertainties, not to become experts in probability analysis. We shall avoid becoming excessively distracted by this fascinating but decidedly peripheral topic.

Nevertheless, it is desirable to exercise our understanding of some of the concepts discussed so far in this chapter and in Chap. 2: We do this by reconsidering two of the earlier examples from a strictly probabilistic point of view.

Mendenhall (Men 67) describes a procedure for calculating probability, based on the sample-point approach, which is worthwhile repeating here:

Step 1: Define the experiment.

Step 2: List the simple events associated with the experiment and test each to make certain that they cannot be further decomposed. This defines the sample space E .

Step 3: Assign reasonable probabilities to the sample points e in E , making sure $\sum_{e \in E} P(e) = 1$.

Step 4: Define the event of interest, A , as a specific collection of sample points. (A sample point is in A if A occurs when the sample point occurs.) Test all sample points in E to locate those in A .

Step 5: Find $P(A)$ by summing the probabilities of the sample points in A .

Mendenhall (Men 67) stresses that combinational methods, such as those discussed in Sec. 2.3, are pertinent to this procedure because they assist in determining the total number of points in E , as well as those in A . When the sample points are, for instance, assumed to be equally likely, summation of the probability of the sample points in A , Step 5, can be accomplished by counting the points in A and multiplying by the probability per sample point.

Calculation of the probability of an event by using the five-step procedure described above is a systematic approach which will lead to the right solution if all the steps are correctly followed. A major possibility for error is introduced if one neglects to define the experiment clearly (Step 1) and thereby improperly specifies the simple events (Step 2). A second source of error is the failure to assign valid probabilities to the sample points. The procedure becomes tedious (and, for all practical purposes, unmanageable) when the number of sample points in E is large, except, as indicated above, in those cases where sets of sample points are equally likely (or equi-probable). When this occurs, summation can sometimes be accomplished by using the counting rules.

I have so far avoided discussing the link between probability and uncertainty in rigorous terms, and will continue to do so for the remainder of this report. However, in the next report of this series I will pick up on this topic in a formal way, as the necessary mathematical groundwork will by then have been well established. In the meantime, I choose to limit myself to piquing the interest of the reader by dabbling in this general area in a casual manner by way of the following two examples.

Example 4.5

Let us reconsider the circumstances described in Ex. 2.3. Now suppose that the experimenter chooses not to repair the malfunctioning apparatus by himself, but instead elicits the services of a professional repairman who will service the unit under a maintenance contract for a fixed fee, regardless of the problem. In establishing the cost for this service, one factor to be considered is how many IC components will typically have to be replaced for such a unit. We examine Table 2.1 and, assuming each elementary failure mode to be equally likely, deduce the following probabilities $P(k)$ for compound events consisting of the failure of k IC components (Possibility A): $P(0) = 1/16$, $P(1) = 1/4$, $P(2) = 3/8$, $P(3) = 1/4$, and $P(4) = 1/16$. Note that

$$\sum_{k=0}^4 P(k) = 1 \quad ,$$

as required. Clearly, the most common failure mode is that involving 2 IC components. Should the repair company figure its service pricing on the basis that a typical failure involves 2 IC components? We will eventually learn that the best choice, which is designated as the expected value or the most-probable or most-likely number of failures, turns out to be the weighted average

$$\langle k \rangle = \sum_{k=0}^4 k P(k).$$

In this problem, the answer is indeed $\langle k \rangle = 2$. The nomenclature $\langle \dots \rangle$ is a useful one for designating weighted averages of observables. Such weighted averages always involve the applicable probability function P .

We know that there is a spread of possibilities for k relative to the value 2, but how should this be quantified for purposes of uncertainty estimation? Again, we will formally learn, in the following report of this series, that the variance,

$$\langle (k-2)^2 \rangle = \sum_{k=0}^4 (k-2)^2 P(k).$$

should be used for this purpose. The square root of the variance proves to be a reasonable measure of the uncertainty in k , and it is known as the standard deviation of the probability distribution. In this example, the uncertainty provided by this measure is approximately 0.4. Practically speaking, the repair company does not really care about the uncertainty in the true cost of any individual repair job, but rather is concerned with

the uncertainty in the anticipated profits based on many such repair jobs. Since the probability of a repair job requiring more than 2 IC-component replacements is $P(k>2) = 5/16$, while the probability of requiring fewer than 2 IC components is $P(k<2) = 5/16$, the same, it is clear that losses to the company due to IC replacements exceeding the average 2 are generally cancelled by enhanced profits for repairs involving fewer than 2 IC components. It turns out that if the company repairs a great many such units in a given fiscal period, the profit uncertainty due to this effect will be very small.

Example 4.6

Example 2.10 provides the basis for an interesting physical application of probability which clearly illustrates the procedure of Mendenhall (Men 67). The elementary events to be considered are three-particle configurations allowed by the applicable statistical laws, as illustrated in Figs. 2.14 - 2.16. Corresponding to these are compound events with system total energy as the distinguishing attribute. Under conditions of thermal equilibrium, the elementary events do not have equal likelihood, but are weighted relative to each other by the Maxwell-Boltzmann Factor, $e^{-E/kT}$, where T is the temperature, E is the ensemble total kinetic energy, and k is Boltzmann's Constant (Mor 64). We assume that the particles do not interact with each other. The probability function P which satisfies all the necessary requirements must have the form

$$P(E) = [N(E)e^{-E/kt}] / Z \quad , \quad (4.12)$$

where

$$Z = \sum_E N(E)e^{-E/kt} \quad , \quad (4.13)$$

and $N(E)$ represents the number of available elementary states which have system total energy E (see Table 2.3). In other words, $P(E)$ is interpreted as the probability that the ensemble of three particles will have total energy E . Z is evaluated by summing over all allowed states (see Table 2.3) and, in statistical physics, is called a partition function (Mor 64). Specific numerical calculations will obviously depend upon system temperature, so we assume for demonstration purposes that $kT = 3e$. The resulting probabilities for each compound event are given in Table 4.2. The most-likely values for the ensemble total energy, derived from the formula

$$\langle E \rangle = \sum_E P(E)E \quad ,$$

Table 4.2 Calculated Probabilities for Three-Particle Total-Energy States. See Ex. 4.6^a.

Ensemble Total Energy (E)	Maxwell- Boltzmann	Distinguishable with Exclusion	Bose- Einstein	Fermi- Dirac
0	0.0570386	Forbidden	0.168815	Forbidden
e	0.1226098	Forbidden	0.120961	Forbidden
2e	0.175708	Forbidden	0.173345	Forbidden
3e	0.209833	0.384937	0.186311	0.384937
4e	0.180423	0.275819	0.133497	0.275819
5e	0.129278	0.197633	0.0956551	0.197633
6e	0.0771934	0.141610	0.0685399	0.141610
7e	0.0331869	Forbidden	0.0327407	Forbidden
8e	0.0118897	Forbidden	0.0117299	Forbidden
9e	0.00283978	Forbidden	0.00840481	Forbidden
<E>	3.2878 e	4.0959 e	2.8488 e	4.09591 e

^aRefer also to Figs. 2.14 - 2.16 and Table 2.3. Certain ensemble total-energy states are forbidden by the statistical laws, as indicated.

are also shown in Table 4.2. However, since we are dealing with a quantized physical system, we can never expect to actually observe the ensemble to have any of these calculated values $\langle E \rangle$.

Hopefully, the discussions and examples in this chapter will provide the reader with an indication of the direction to be taken in the present development of the subject of uncertainties. Uncertainty is evidently the consequence of properties of the underlying probability functions which govern physical processes, of the measurement procedures undertaken to learn about physical phenomena, or both.

5. CONDITIONAL PROBABILITY AND INDEPENDENCE

Probability, as defined in Chap. 4, is quite limited with regard to many important applications. In particular, it does not permit us to deal with questions having the general form: "What is the probability of B occurring if we know that A has actually occurred?" Most realistic situations involve such conditions (explicit or implicit), and outcomes based on probability depend a great deal on whether these conditions are met or not. The following example demonstrates this point.

Example 5.1

A Ge photon detector is employed, in a fixed geometry, to measure gamma-rays emitted from a source. Under normal conditions the detector is very stable, and the full-energy peak efficiency is ϵ for detection of a particular gamma-ray in the configuration of the experiment. Let us suppose that the decay mean life τ is very much longer than the counting time t . We observe N events in the peak following a particular count. We deduce that the activity A of the source is $A \approx N/(t\epsilon)$. An important hidden condition underlying this simple and very common situation is that we have assumed that the detection apparatus was operating stably during the entire counting time t . Just because it was operating properly when we started the count and when we stopped, it does not prove continuity over the whole time interval. A careful experimenter will instinctively review all the available evidence. For example, if he notices that the electric clock in the counting room had the right time when the count was stopped, this is convincing evidence that no detrimental electric power interruptions had occurred during the count.

In Chap. 2 it was indicated, in conjunction with a discussion of Borel fields, that one had to consider event counting with respect to subsets in order to deal with certain conditions that are imposed in probability analyses. This matter will be explored further in the present chapter. The principal concepts are discussed from an intuitive point of view in Sec. 5.1. Some formal aspects of the theory are then explored in Sec. 5.2. This aspect of probability theory is known as conditional probability.

5.1 Intuitive Approach

Let us consider an event space E with n_E elementary events. Let A and B be two arbitrary events of E , with the sole restriction that probability $P(A) > 0$. Now we return to the generic question posed early in this chapter: "What is the probability of event B , given the condition denoted as event A ?" We are immediately led to consider the event $B \times A$. However, the probability $P(B \times A)$ is not the answer we seek. The space E very likely includes events

for which A does not occur, and the function P is based on counting possibilities with respect to the entire space E. To be precise, we should denote P by P_E . Then $P_E(B \times A)$ is the probability that both A and B occur given that E has occurred! This is clearly not what we are looking for. We must restrict our consideration to the subspace A, and define a new probability function P_A which has the property that $P_A(A) = 1$. The answer to our original question can now be obtained: It is $P_A(B \times A)$. Here, P_A is a conditional probability function in which occurrence of event A is the condition which limits the range of possibilities we should consider among those available from E. Normally, the subscripts are omitted, and $P_A(B \times A)$ is expressed in the form $P(B/A)$. Note that "B/A" is not a quotient, but designates the concept: "B given A". We must keep in mind that "P" is being used to designate two conceptually distinct probabilities. What was labeled as P_E is called unconditional probability, or sometimes a priori probability, while P_A is called conditional probability, or a posteriori probability with respect to event A. The subscript "E" will usually not appear in representations of unconditional probability.

How do we calculate $P(B/A)$? Suppose n_A designates the number of elementary events in E which belong to A. We require $n_A > 0$. Furthermore, n_B and n_{AB} designate the simple-event measures for B and $B \times A$, respectively. The frequency definition of probability tells us that

$$P(A) = n_A/n_E,$$

$$P(B) = n_B/n_E,$$

and

$$P(B \times A) = n_{AB}/n_E.$$

$P(B/A)$ is therefore defined in the frequency sense by

$$P(B/A) = n_{AB}/n_A. \quad (5.1)$$

Consequently, one is led to the general relationship

$$P(B/A) = P(B \times A)/P(A), \text{ if } P(A) > 0, \quad (5.2)$$

between the conditional probability $P(B/A)$ and the unconditional probabilities $P(A)$ and $P(B \times A)$.

Furthermore, suppose that B is nontrivial and $P(B) > 0$. It is simple to show the validity of the interesting, symmetrical formula

$$P(B/A)P(A) = P(A/B)P(B). \quad (5.3)$$

The reader may find it helpful to refer to the Venn diagram in Fig. 5.1 in order to gain some insight concerning this matter. The following example will serve to further acquaint the reader with conditional probability.

Example 5.2

The situation first discussed Ex. 2.3 provides an excellent vehicle for illustrating the notion of conditional probability. The elements which denote the event space E are listed in Table 2.1. Let A_i be the event that IC No. i fails ($i = 1, 4$). Also, consider events $A_1 \times A_2$ and $A_1 \times A_2 \times A_3$. The simple elements of each of these events are listed explicitly in Table 5.1, using notation consistent with Table 2.1. Clearly, $n_1 = n_2 = n_3 = n_4 = 8$, $n_{12} = 4$, and $n_{123} = 2$ are the measures of these selected compound events. If we treat all elementary events as equally likely, then

$$P(A_i) = 8/16 = 1/2 \quad (i = 1, 4).$$

It is intuitively reasonable that the probability for failure of any particular IC component should be $1/2$ in this example.

The probability that both IC No. 1 and IC No. 2 fail is

$$P(A_1 \times A_2) = 4/16 = 1/4.$$

However, the conditional probability that IC No. 2 fails, given that IC No. 1 failed, is

$$P(A_2/A_1) = P(A_1 \times A_2)/P(A_1) = (1/4)/(1/2) = 1/2.$$

Also, note that

$$P(A_2/A_1)P(A_1) = P(A_1/A_2)P(A_2) = (1/2)(1/2) = 1/4.$$

The probability that IC Nos. 1, 2 and 3 fail is

$$P(A_1 \times A_2 \times A_3) = 2/16 = 1/8.$$

However the probability that IC No. 3 fails, given that IC Nos. 1 and 2 failed, is

$$P(A_3/A_1 \times A_2) = P(A_1 \times A_2 \times A_3)/P(A_1 \times A_2) = (1/8)/(1/4) = 1/2.$$

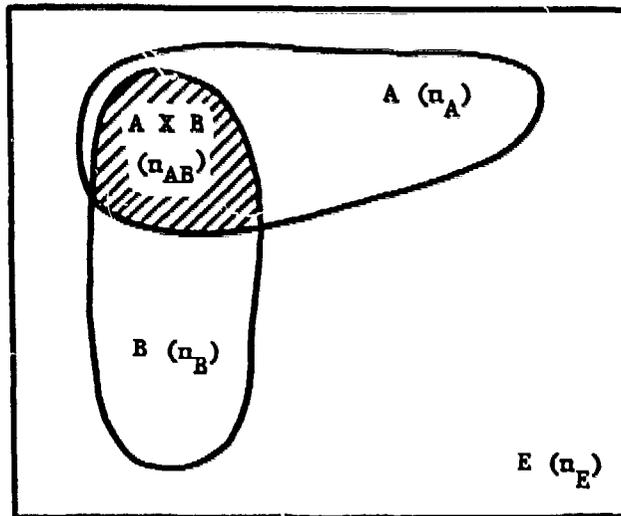


Figure 5.1: Venn diagram illustrating the relationship between events A, B, and $A \times B$ which are used to introduce the concept of conditional probability.

Table 5.1 Explicit Structure of Several Compound Events Formed from the Elementary Events Defined in Table 2.1. See Ex. 5.2.

A_1 :	(B,G,G,G) (B,B,G,G) (B,G,B,G) (B,G,G,B) (B,B,B,G) (B,B,G,B) (B,G,B,B) (B,B,B,B)	A_2 :	(G,B,G,G) (B,B,G,G) (G,B,B,G) (G,B,G,B) (B,B,B,G) (B,B,G,B) (G,B,B,B) (B,B,B,B)
A_3 :	(G,G,B,G) (B,G,B,G) (G,B,B,G) (G,G,B,B) (B,B,B,G) (B,G,B,B) (G,B,B,B) (B,B,B,B)	A_4 :	(G,G,G,B) (B,G,G,B) (G,B,G,B) (G,G,B,B) (B,B,G,B) (B,G,B,B) (G,B,B,B) (B,B,B,B)
$A_1 \times A_2$:	(B,B,G,G) (B,B,B,G) (B,B,G,B) (B,B,B,B)	$A_1 \times A_2 \times A_3$:	(B,B,B,G) (B,B,B,B)

One interesting feature that emerges in this example is that it looks as if

$$P(A_i/Q) = P(A_i),$$

regardless of the condition Q. This happens because we have postulated a scenario in which the probability of failure of any particular IC component in no way depends upon what happens to the other IC components. When this is the case, we say that the events Q and A_1 are stochastically independent. We could also demonstrate by examples that

$$P(A_i \times Q) = P(A_i)P(Q)$$

whenever A_i and Q are independent events of E.

Much has already been said about the concept of sampling. The term trial is usually used to designate one step among many in sequential sampling exercises. We have also introduced the concept of an ordered k-tuple (Sec. 2.3). Imagine for a moment that we have a space E of simple events e. We choose to perform k trials, with the proviso that no trial perturbs the essential nature of the space E. This requirement can be satisfied if the events are replaced after the trial or if E contains so many events e that it remains effectively unperturbed by the k sequential trials. Under these conditions, one must conclude that the outcome of any one of the k trials is unaffected by the fact that there have been prior trials. Trials of this nature are called independent trials. This notion is very important to our understanding of uncertainty. For example, in the evaluation of nuclear data it is quite often assumed that various data sets which are employed in an evaluation result from independent experiments (independent trials).

Associated with E is a probability function P such that P(e) is the probability of e, and

$$\sum_{e \in E} P(e) = 1.$$

The result of k trials is a collection of k outcomes which form the ordered k-tuple (an arrangement) (e_1, e_2, \dots, e_k) . This particular arrangement can be considered as an event h belonging to the event space H consisting of all possible k-tuples which could be formed from E with replacement. For independent trials, one defines a probability function for H which is related to that for E by the equation

$$P(h) = \prod_{i=1}^k P(e_i) \quad (5.4)$$

Once again we stress that two distinct probability functions, P_E and P_H , are really involved, but according to convention the subscripts are omitted and the space to which a particular probability function refers must be deduced from the context.

The following example illustrates the notion of probability for a sequence of independent trials:

Example 5.3

Consider an experimental procedure in which the outcome is either success (S) or failure (F). The event space E has two events; thus $E: \{S, F\}$. The probabilities are assumed to be $P(S) = s$ and $P(F) = f$, with $s + f = 1$. Two independent trials are to be performed. The event space H of possible outcomes has the form $H: \{(S, S), (S, F), (F, S), (F, F)\}$. Notice that (S, F) and (F, S) are distinct because H consists of ordered pairs. In accordance with Eq. (5.4), we assume that the probabilities for the outcomes $h \in H$ are: (S, S), s^2 ; (S, F), sf ; (F, S), $fs = sf$; (F, F), f^2 . Probability defined on H must satisfy the three axioms discussed in Sec. 4.2. Clearly, $P(h) \geq 0$ for each $h \in H$, so Axiom I is satisfied.

$$P(H) = s^2 + 2sf + f^2 = (s + f)^2 = 1^2 = 1,$$

so Axiom II is satisfied. Since all of the defined elementary events $h \in H$ are obviously mutually exclusive, any two mutually exclusive compound events generated from them will satisfy the additivity property, Axiom III. Thus, we see that probability as we have defined it for the space of all pairs of outcomes of two independent sequential trials, H, satisfies the required axioms and also conforms with our intuitive notion of what the probability should be.

We shall see in Chap. 7 that the trial procedure discussed in Ex. 5.3 belongs to a class of trials known as independent Bernoulli trials. Several probability functions that are important for applications evolve from Bernoulli trials.

5.2 Theoretical Approach

As indicated in Sec. 2.2, if E is an event space with a corresponding Borel field Z, and if $A \in E$ and is nontrivial, that is if $A \neq \emptyset$ and $P(A) > 0$, then one can define a new Borel field Z' which is the collection of all possible events of the form $B \times A$, for arbitrary events B of E. The conditional probability $P(B/A)$ is formally defined for all these events belonging to Z' by the expression

$$P(B/A) = P(B \times A)/P(A),$$

consistent with Eq. (5.2). In order for conditional probability defined in this way to represent a valid probability function, it must satisfy the three axioms discussed in Sec. 4.2. $P(A) > 0$ and $P(B \times A) \geq 0$, so Axiom I is thus satisfied. A represents the sure event with respect to the Borel field Z' . Furthermore,

$$P(A/A) = P(A \times A)/P(A) = P(A)/P(A) = 1.$$

Therefore, Axiom II is also satisfied. Let B and $C \in E$, with

$$B \times C = \phi.$$

Then

$$P(B + C/A) = P[(B + C) \times A]/P(A) = P(B \times A + C \times A)/P(A).$$

However,

$$(B \times A) \times (C \times A) = (B \times A) \times (A \times C) = B \times (A \times A) \times C = B \times A \times C =$$

$$(B \times C) \times A = \phi \times A = \phi$$

from the rules of Boolean algebra, as discussed in Sec. 2.2. Consequently, $(B \times A)$ and $(C \times A)$ are mutually exclusive. Therefore,

$$P(B \times A + C \times A)/P(A) = [P(B \times A) + P(C \times A)]/P(A) = P(B/A) + P(C/A).$$

Axiom III is therefore satisfied, so conditional probability, as defined above, is legitimate and applicable to every event in Z' .

There are many interesting features of conditional probability. A few of these will be expressed below in the form of theorems.

Theorem: If $A \in E$ with $P(A) > 0$ and $B \in E$ with $P(B) > 0$, then

$$P(B/A)P(A) = P(A/B)P(B),$$

consistent with Eq. (5.3.).

The proof involves a trivial application of the definition of conditional probability.

Clearly, $P(A/B)$ and $P(B/A)$ have different meanings, and they are equal in value only when $P(A) = P(B)$. This is also evident from the following theorem.

Theorem: If $A \in E$ with $P(A) > 0$ and $B \in E$ with $P(B) > 0$, and furthermore, $B \in A$, then

$$P(B/A) = P(B)/P(A)$$

while $P(A/B) = 1$.

Some other useful theorems related to conditional probability are:

Theorem: [Chain Rule]

If A, B and $C \in E$, then

$$P(A \times B \times C) = P(A) P(B/A) P(C/A \times B) \quad (5.5)$$

Let $Q = A \times B$; then

$$P(A \times B \times C) = P(Q \times C) = P(C/Q)P(Q);$$

but

$$P(Q) = P(B/A)P(A).$$

The theorem is proved by substitution.

Theorem: [Generalized Chain Rule]

If $A_i \in E \quad (i = 1, n)$, then

$$P\left(\prod_{i=1}^n A_i\right) = P(A_1)P(A_2/A_1)P(A_3/A_1 \times A_2) \dots P(A_n/\prod_{i=1}^{n-1} A_i) \quad (5.6)$$

Theorem: If $A \in E$ and $P(A) > 0$, then $P(E/A) = 1$.

Theorem: If A and $B \in E$ and $P(A) > 0$, then $P(B/A) \leq 1$.

Note that $E = B + \bar{B}$, so

$$P(E/A) = 1 = P(B/A) + P(\bar{B}/A),$$

since B and \bar{B} are mutually exclusive. However, $P(\bar{B}/A) \geq 0$. Therefore, $P(B/A) \leq 1$.

Theorem: If $A \in E$ with $P(A) > 0$ and $B \in E$ with $P(B) = 0$, then $P(B/A) = 0$.

$$B = B \times A + B \times \bar{A}.$$

Also, $(B \times A)$ and $(B \times \bar{A})$ are mutually exclusive. Therefore,

$$P(B) = P(B \times A) + P(B \times \bar{A}).$$

Since $P(B) = 0$, $P(B \times A) \geq 0$ and $P(B \times \bar{A}) \geq 0$, it must be the case that $P(B \times A) = 0$. Therefore,

$$P(B/A) = P(B \times A)/P(A) = 0.$$

Theorem: [Poincare Law of Conditional Probability Addition]

If A, B and $C \in E$ with $P(A) > 0$, then

$$P(B + C/A) = P(B/A) + P(C/A) - P(B \times C/A) \quad . \quad (5.7)$$

Theorem: [Rule of Total Probability]

Let A_1, \dots, A_n be a partition of E with $P(A_i) > 0$ ($i = 1, n$). Then if $B \in E$,

$$P(B) = \sum_{i=1}^n P(B/A_i)P(A_i) \quad . \quad (5.8)$$

Now,

$$B = \sum_{i=1}^n B \times A_i,$$

and, from the discussion in Sec. 4.2 on marginal probability, we know that

$$P(B) = \sum_{i=1}^n P(B \times A_i).$$

The theorem is thus proved from the definition of conditional probability, namely,

$$P(B/A_i) = P(B \times A_i)/P(A_i).$$

Figure 5.2 shows the relationship between event B and the events A_1, A_2, \dots, A_n which form a partition of E .

The rule of total probability is useful if direct computation of $P(B)$ is difficult, whereas calculation of the probabilities $P(A_i)$ and $P(B/A_i)$ is not so difficult. Although this is not the case for the following example, it nevertheless does demonstrate how this rule can be applied.

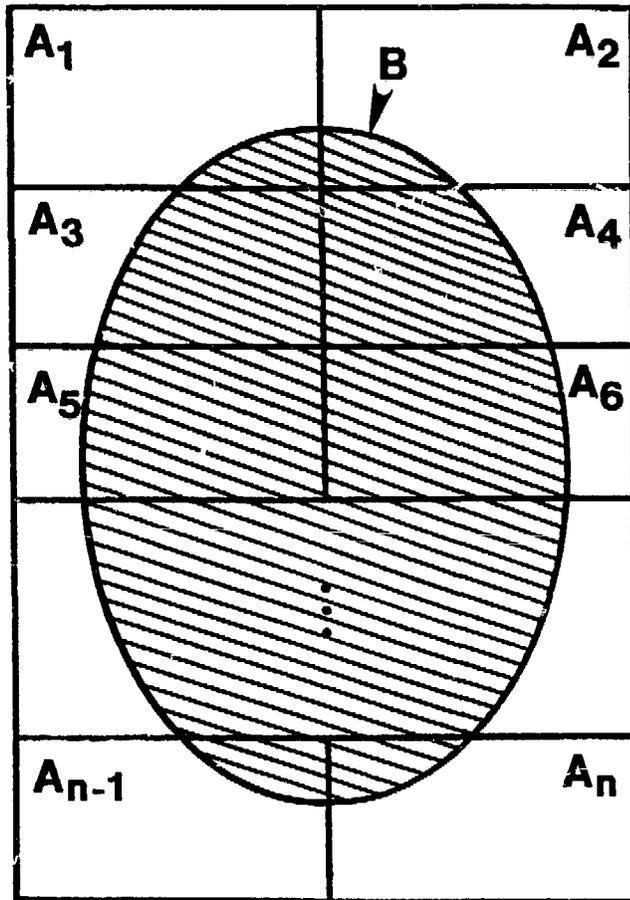


Figure 5.2: Venn diagram illustrating the relationship between event B belonging to E and the events A_1, A_2, \dots, A_n which form a partition of E .

Example 5.4

Let's reconsider Ex. 2.3. Referring to Table 2.1, we see that one possible partition of the event space is A_i ($i = 0,4$), where "i" designates the number of IC components which fail. Let B be the event that IC No. 1 fails. Table 5.2 summarizes the pertinent parameters of this example and explicitly demonstrates the application of Eq. (5.8) for evaluation of P(B).

The following theorem is of such fundamental importance to applications of probability theory that Chap. 8 of this report is devoted to providing examples and discussing related philosophical implications. The theorem is called Bayes' Theorem, after the eighteenth century clergyman, Rev. Thomas Bayes, who is usually credited with its discovery. First, we state the theorem and provide a proof.

Theorem: [Bayes' Theorem]

Let A_i ($i = 1,n$) be a partition of E with all $P(A_i) > 0$. Furthermore, let $B \in E$ with $P(B) > 0$. Then for every k, $1 \leq k \leq n$,

$$P(A_k/B) = P(B/A_k) P(A_k) / \left[\sum_{i=1}^n P(B/A_i) P(A_i) \right] \quad (5.9)$$

This theorem is a natural extension of the Rule of Total Probability. Thus,

$$P(B) = \sum_{i=1}^n P(B/A_i) P(A_i).$$

In a previous theorem, it was shown that

$$P(A_k/B) P(B) = P(B/A_k) P(A_k),$$

given that A_k and $B \in E$, and that $P(A_k) > 0$ and $P(B) > 0$. Equation (5.9) follows directly from substitution.

Although Bayes' theorem is discussed extensively in Chap. 8, it should be indicated at this point what the role of this theorem is in probability theory. So far, we have viewed the analysis of probability as proceeding from the definition of probabilities for simple events toward the calculation of probabilities for more complex events. In practice, however, what is often required is just the inverse; that is, given certain experimental observations, we wish to learn something about the parent population and generating mechanism which were responsible for our observations. This process is known as statistical inference. Suppose that event B has been observed. That event could be explained by several mutually exclusive causes

Table 5.2 Parameters Relevant to Example 5.4

Partition:

A_0 : No IC component fails.	$P(A_0) = 1/16$
A_1 : One IC component fails.	$P(A_1) = 4/16$
A_2 : Two IC components fail.	$P(A_2) = 6/16$
A_3 : Three IC components fail.	$P(A_3) = 4/16$
A_4 : Four IC components fail.	$P(A_4) = 1/16$

Marginal Probabilities:

$$P(B/A_0) = 0$$

$$P(B/A_1) = 1/4$$

$$P(B/A_2) = 3/6$$

$$P(B/A_3) = 3/4$$

$$P(B/A_4) = 1$$

$$P(B) = \sum_{i=0}^4 P(B/A_i)P(A_i)$$

$$= (0)(1/16) + (1/4)(4/16) + (3/6)(6/16) + (3/4)(4/16) + (1)(1/16)$$

$$= 8/16 = 1/2.$$

or hypotheses A_1, A_2, \dots, A_n (a partition of the space E of all possible explanations for B). Each hypothesis A_k has an a priori probability $P(A_k)$ of being true; that is, before the trial that produced B was executed, the $P(A_k)$ represented our "state-of-the-art" understanding of the situation. However, once the trial has been performed and B has been observed, we have acquired new knowledge. Consequently, our understanding of the probabilities for validity of the various hypotheses must be revised. We should now substitute a posteriori probabilities $P(A_k/B)$ for the original $P(A_k)$. Formally, Bayes' theorem provides the mechanism for calculating these revised probabilities. The quantity $P(B/A_i)$ is called a likelihood factor or likelihood. Calculation of the likelihood is usually straightforward, if A_i is a reasonable hypothesis. Our real problem lies in establishing the a priori probabilities $P(A_i)$. If we commence from a status of total ignorance, then there would seem to be no basis for favoring one hypothesis over another. The assumption that all the possible hypotheses should have equal a priori probabilities is known as Bayes' Postulate. The reader should note that this process can be repeated again and again. For example, for a second trial, the a priori probabilities are taken to be $P(A_i/B)$, since the first trial produced B , and one therefore has revised the $P(A_i)$. This approach toward statistical inference is appealing to many applied scientists. The reader who is at all familiar with nuclear data evaluation will surely recognize this as the approach often used in this field. That is, one eventually "learns" about the nature of fundamental nuclear properties by repeated and ever more refined experimentation. However, many statisticians are not comfortable with the concept of Bayesian statistical inference and therefore avoid it. That this remains a point of controversy is one indication that the field of statistics has not attained full maturity.

Two events A and B belonging to E are defined to be independent if

$$P(A \times B) = P(A) P(B).$$

From the definition of conditional probability, it then follows that

$$P(A/B) = P(A) \text{ if } P(B) > 0$$

and

$$P(B/A) = P(B) \text{ if } P(A) > 0.$$

If, for a collection of events A_i ($i = 1, n$),

$$P(A_i \times A_j) = P(A_i)P(A_j)$$

whenever $i \neq j$, then these events are said to be pairwise independent; they are not necessarily mutually independent. The terms independent en bloc or fully independent are also used to denote mutual independence. To be mutually independent, it is required that

$$P(A_{k_1} \times A_{k_2} \times \dots \times A_{k_s}) = P(A_{k_1}) P(A_{k_2}) \dots P(A_{k_s}) \quad (5.10)$$

for any $s = 1, n$ and $1 \leq k_1 < k_2 < \dots < k_s \leq n$.

A number of theorems can be proved relating to independent events. Among them are the following:

Theorem: If A and B are independent events then:

- (1) A and \bar{B} are independent,
- (2) \bar{A} and B are independent,
- (3) \bar{A} and \bar{B} are independent.

We prove Part (1) for demonstration purposes:

$$A = A \times E = A \times (B + \bar{B}) = A \times B + A \times \bar{B}.$$

Since $(A \times B)$ and $(A \times \bar{B})$ are mutually exclusive,

$$P(A) = P(A \times B) + P(A \times \bar{B}).$$

Thus,

$$P(A \times \bar{B}) = P(A) - P(A \times B) = P(A) - P(A)P(B) = P(A)[1 - P(B)].$$

Since B and \bar{B} are mutually independent,

$$1 - P(B) = P(\bar{B}).$$

Part (1) is therefore proved by substitution.

Theorem: If A_1, \dots, A_n and $B \in E$ and B is pairwise independent with respect to each A_i ($i = 1, n$), then

$$P\left(\left(\sum_{i=1}^n A_i\right) \times B\right) = P\left(\sum_{i=1}^n A_i\right) P(B).$$

Theorem: If A_1, A_2, \dots, A_n are mutually independent, then

$$P\left(\sum_{i=1}^n A_i\right) = 1 - \prod_{i=1}^n P(\bar{A}_i).$$

The notion of independent trials is discussed extensively in Sec. 5.1. Probability for the event space of k-tuples is formally defined by Eq. (5.4) when the trials are independent. One can show that probability defined this

way satisfies the three axioms of Sec. 4.2. We demonstrated this in the special case of Ex. 5.3, and pursue the issue no further here.

All trials which do not satisfy the condition of independence are denoted dependent trials. Let A denote the collective outcome of k sequential trials, the ith of which yields A_i . Clearly,

$$A = \prod_{i=1}^k A_i.$$

The probability $P(A)$, in the case of arbitrary dependencies, can be derived in principle using the Generalized Chain Rule for conditional probability. Usually this is impractical. For this reason, great pains are taken to set up experiments in such a way that they approximate independent trials.

Although we shall not pursue the matter to any great extent, for interest we now introduce a particular class of dependent trials known as Markov Trials. Again, let A_1, A_2, \dots, A_k be the events observed for k sequential trials. If these were independent trials, then for the jth trial,

$$P(A_j / \prod_{i=1}^{j-1} A_i) = P(A_j) \quad (j \leq k \leq 2).$$

In short, there would be no memory of any of the preceding trials at each stage of the chain of trials. However, Markov trials differ in the sense that

$$P(A_j / \prod_{i=1}^{j-1} A_i) = P(A_j / A_{j-1}) \quad (j \leq k \leq 2).$$

For Markov trials, the outcome of any particular trial depends upon the preceding trial, but not on any of the earlier ones. Markov-trial processes involve memory of the immediate past but nothing more.

6. PROBABILITY DISTRIBUTIONS

Random variables, random-variable transformations, and distribution and density functions were discussed in Chap. 3. Probability, as applied directly to events, was treated in Chaps. 4 and 5. The present chapter addresses the unification of these concepts, i.e., the merger of analysis and probability theory.

Without loss of generality, we restrict this discussion to event spaces which can be represented by one or more continuous random variables, each of which usually spans the entire range of real numbers $(-\infty, +\infty)$. First, consider events which can be represented by a single random variable x . Let α be any real number. We choose to ask the question: "What is the probability that a trial will yield a value of $x < \alpha$?" It is postulated that such a probability exists and that it satisfies the basic axioms of probability theory as enumerated in Chap. 4. It is designated as $P(x < \alpha)$, or $P(\alpha)$ for convenience. P is called a cumulative probability or probability distribution function, consistent with the convention of Sec. 3.3. The term univariate is often used to indicate that a single random variable is involved. Normally, one needs to be concerned only with probability functions which are either absolutely continuous, or are at most discontinuous in a "mild" way at a discrete number of points in the range $(-\infty, +\infty)$. We will not elaborate on what is meant by "mild discontinuities," but, for example, singularities are unacceptable while simple jump discontinuities can be accommodated. For simplicity, let us avoid further discussion of discontinuities. The interested reader is referred to the bibliography (e.g., Zeh 70) for consideration of the complications which they introduce. For an absolutely continuous probability distribution function P there exists a corresponding non-negative function p known as the probability density function. The relationship is

$$P(\alpha) = \int_{-\infty}^{\alpha} p(x)dx \text{ for } -\infty < \alpha < +\infty \quad . \quad (6.1)$$

The reader is reminded that in Sec. 3.3 we approached this topic from a different point of view. There the density function was treated as the more basic function, with the distribution function evolving from it. However, it is a tradition in probability theory to treat P as the more fundamental of the two, since P , in fact, is a true probability while p is not. For our purposes, we will assume that both P and p exist and will not concern ourselves further with which is the more fundamental. Some important properties of these functions that we will consider are:

- i) $P(\alpha) \geq 0$ for all real α . [Axiom I].
- ii) If $\beta \geq \alpha$, then $P(\beta) \geq P(\alpha)$. [Monotonic Property].
- iii) $\lim_{h \rightarrow 0} P(\alpha + h) = P(\alpha)$.
- iv) $\lim_{\alpha \rightarrow -\infty} P(\alpha) = 0$.

Note that this is equivalent to the statement that $P(\phi) = 0$, where ϕ is the null event.

$$v) \quad \lim_{\alpha \rightarrow +\infty} P(\alpha) = 1 \text{ [Axiom II].}$$

vi) $p(x) \geq 0$ for all real x as a consequence of the monotonic property of P .

We cannot express Axiom III (Additivity Property) until the notion of an interval is introduced. Let $I = I(\alpha, \beta)$ denote the collection of all values x such that $\alpha \leq x < \beta$. If A is the set of all $x < \alpha$ and B is the set of all $x \leq \beta$, then the interval I is the set $B \times A$. The probability of occurrence of I can be denoted as

$$P(I) = P[I(\alpha, \beta)] = P(\alpha; \beta).$$

It is related to $P(\alpha)$ and $P(\beta)$ by the relationship

$$P(\alpha; \beta) = P(\beta) - P(\alpha) \quad . \quad (6.2)$$

Note that here α could be $-\infty$ and β could be $+\infty$. Such intervals are designated as unbounded intervals. If both α and β are finite, the interval is bounded. If I_1 and I_2 are two nonintersecting intervals (mutually exclusive events in event language), then

$$P(I_1 + I_2) = P(I_1) + P(I_2).$$

This corresponds to a statement of Axiom III in the present context.

The following simple example demonstrates the Poincare Law of Probability Addition:

Example 6.1

Let A be the event that $x < \alpha$ while B is the event that $x < \beta$. Suppose $\beta > \alpha$. Clearly, $A \in B$. Now

$$P(A + B) = P(A) + P(B) - P(A \times B).$$

But $A \times B = A$, so $P(A + B) = P(B)$. Now $P(B) = P(\beta)$. This clearly is intuitively reasonable. Whenever $x < \alpha$, it also holds that $x < \beta$. The probability of x being less than α or β is just the probability of $x < \beta$.

A simpler relationship between P and p than indicated by Eq. (6.1) is apparent at the differential level. Consider the small interval $\alpha \leq x < \alpha + \Delta x$, where Δx represents a small increment of the continuous random variable x .

Then, referring to Eq. (6.2),

$$P(\alpha; \alpha + \Delta x) = P(\alpha + \Delta x) - P(\alpha) = \int_{-\infty}^{\alpha + \Delta x} p(x) dx - \int_{-\infty}^{\alpha} p(x) dx = \int_{\alpha}^{\alpha + \Delta x} p(x) dx .$$

In the limit as Δx becomes vanishingly small (approaches a differential dx), we have $P(\alpha + dx) - P(\alpha) = dP(\alpha) = p(\alpha)dx$, or $p(\alpha) = dP(\alpha)/dx = P'(\alpha)$. The density function p is just the derivative of the distribution function P , provided that P is continuous and differentiable over the region in question. This is also quite evident from Eq. (6.1), considering the definition of an integral. It is clear that there is no meaning to the question: "What is the probability of observing a particular value α ?" One must ask for the probability of finding x in some interval. That probability vanishes as the size of the interval approaches zero. Probability density is the local probability per unit of interval. It therefore possesses an intrinsic dimensionality, whereas probability itself is dimensionless.

A multivariate probability distribution or joint probability distribution $P_{1\dots n}$ is defined whenever a probability law applies to events which require more than one random variable to represent them. Consistent with Sec. 3.3, we can define a joint probability density function $p_{1\dots n}$ if the joint probability distribution $P_{1\dots n}$ is absolutely continuous. The relationship is

$$p_{1\dots n}(\alpha_1, \dots, \alpha_n) = \int_{-\infty}^{\alpha_1} dx_1 \dots \int_{-\infty}^{\alpha_n} dx_n p_{1\dots n}(x_1, \dots, x_n), \quad (6.3)$$

and $P_{1\dots n}$ is non-negative whenever it is defined.

Analogous to the notion of probability on an interval for an univariate distribution, one can define probability on a particular region or volume of n -dimensional space. If A designates that region, then

$$P_{1\dots n}(A) = \int_A dx_1 \dots dx_n p_{1\dots n}(x_1, \dots, x_n) . \quad (6.4)$$

Figure 6.1 shows a typical region for two continuous random variables.

Consider a probability law involving two random variables, (x_1, x_2) . The distribution function is designated P_{12} and the density function is p_{12} . One can define two new functions p_1 and p_2 as follows:

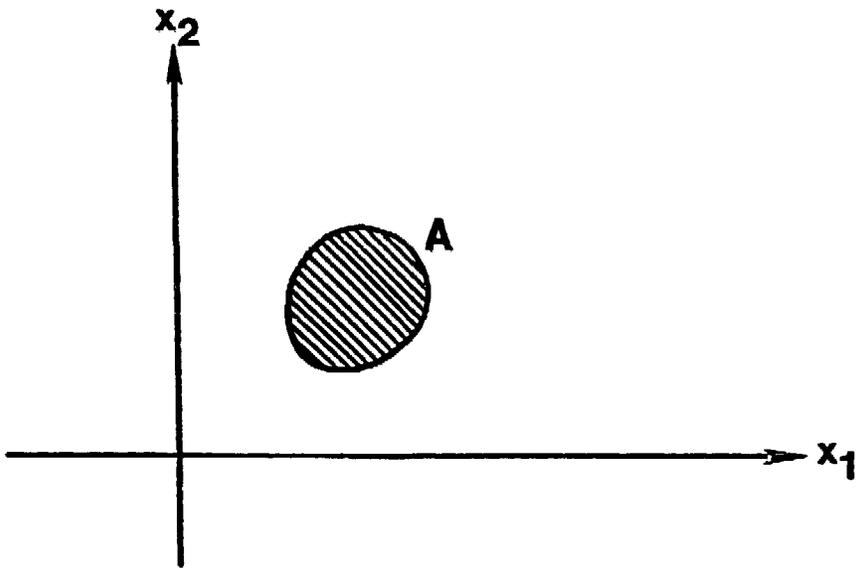


Figure 6.1: Typical region A for an event space described by two random variables

$$p_1(x_1) = \int_{-\infty}^{\infty} dx_2 p_{12}(x_1, x_2) \quad . \quad (6.5)$$

$$p_2(x_2) = \int_{-\infty}^{\infty} dx_1 p_{12}(x_1, x_2) \quad . \quad (6.6)$$

The function p_1 is called the marginal distribution of x_1 , and it is a density function with respect to the variable x_1 . Analogous statements can be made for x_2 . Two other functions, which are called conditional probability density functions, can be defined in terms of the unconditional probability density p_{12} and the marginal distributions p_1 and p_2 . They are:

$$p(x_2/x_1) = \frac{p_{12}(x_1, x_2)}{p_1(x_1)} \quad \text{if } p_1(x_1) > 0 \quad , \quad (6.7)$$

$$p(x_1/x_2) = \frac{p_{12}(x_1, x_2)}{p_2(x_2)} \quad \text{if } p_2(x_2) > 0 \quad . \quad (6.8)$$

The concept and structure of these definitions should be familiar from Sec. 5.2. The Rule of Total Probability can be expressed in the present context using this nomenclature. Thus,

$$p_1(x_1) = \int_{-\infty}^{\infty} dx_2 p_{12}(x_1, x_2) = \int_{-\infty}^{\infty} dx_2 p_2(x_2) p(x_1/x_2) \quad , \quad (6.9)$$

$$p_2(x_2) = \int_{-\infty}^{\infty} dx_1 p_{12}(x_1, x_2) = \int_{-\infty}^{\infty} dx_1 p_1(x_1) p(x_2/x_1) \quad . \quad (6.10)$$

One should keep in mind the following fundamental equation:

$$\{\text{conditional probability}\} = \frac{\{\text{unconditional probability}\}}{\{\text{marginal probability}\}} \quad , \quad (6.11)$$

provided that the marginal probability is nonvanishing. With this in mind, we consider the following question: "What is the probability of finding the second random variable in the interval $dI_2: (x_2, x_2 + dx_2)$ given that the first random variable is known to be in the interval $dI_1: (x_1, x_1 + dx_1)$." According to Eq. (6.11), the answer to this question is:

$$P(dI_2/dI_1) = p_{12}(x_1, x_2) dx_1 dx_2 / [p_1(x_1) dx_1] = p(x_2/x_1) dx_2 \quad .$$

Similarly,

$$p(dI_1/dI_2) = p(x_1/x_2) dx_1 \quad .$$

This discussion can be extended to random variable spaces of larger dimensionality; however, we will not pursue this matter further.

The concept of independence is easily expressed for probability density functions. The random variables x_1, \dots, x_n are deemed mutually independent provided that

$$p_1 \dots p_n(x_1, \dots, x_n) = \prod_{i=1}^n p_i(x_i) \quad (6.12)$$

In short, the joint probability density function is factorable into a product of marginal distributions for all of the random variables involved. For two independent random variables,

$$p_{12}(x_1, x_2) = p_1(x_1)p_2(x_2).$$

Therefore, from Eqs. (6.7) and (6.8) we have that $p(x_2/x_1) = p_2(x_2)$ and $p(x_1/x_2) = p_1(x_1)$. This is entirely consistent with the notion of independence which was developed in Chap. 5.

Let x be a continuous random variable which conforms to a space X . It represents events e belonging to the space E . Suppose that y is another random variable representing the events e of E such that $y = y(x)$ for $x \in X$. The resultant y conforms to a space Y . This is an example of a random-variable transformation, as discussed in Sec. 3.3. Suppose that we know that a probability density function p_x is defined for all $x \in X$. What are the required conditions for the corresponding function p_y to be defined for $y \in Y$, and what is its relation to p_x ? This is a very important question in practice so we will explore it here in some detail.

First, we approach this issue from a intuitive point of view. The incremental probability of an event for which x is in the interval $(x, x + dx)$ is simply $dP_x = p_x(x)dx$ ($dx > 0$ by convention). We assume for simplicity that the function y is continuous and differentiable over the whole space X , and furthermore that the derivative $y'(x) \neq 0$ anywhere in X . Thus $y'(x) < 0$ or $y'(x) > 0$ everywhere in X . Now we turn to a description of events in terms of the random variable y . For the stringent conditions of the present transformation, p_y does indeed exist over the entire space Y which is formed by pointwise mapping of the space X . Corresponding to differential dx in X is the differential dy in Y given by

$$dy = (dy/dx)dx \quad (6.14)$$

Since dy may be negative, we consider instead the positive quantity

$$|dy| = |dy/dx|dx \quad (6.15)$$

Here, the notation $|\dots|$ designates absolute value. The event for which x belongs to dx is also described by y belonging to $|dy|$ (a positive differential). The incremental probability expressed in terms of the variable y is thus

$$dP_y = p_y(y)|dy|.$$

The magnitude of the incremental probability must be independent of the choice of random variable used to describe the event, so

$$dP_x = p_x(x)dx = dP_y = p_y(y)|dy| = p_y(y)|dy/dx|dx.$$

From this we can deduce the important formula

$$p_y(y) = p_x(x)/|dy/dx| = p_x(x)/|y'(x)| = p_x(x)/|J| \quad (6.16)$$

Figure 6.2 illustrates the present situation graphically. This is the law of probability transformation for univariate random-variable events. The restrictive conditions insure that the inverse function y^{-1} exists. Thus,

$$x = y^{-1}(y) \text{ for } y \in Y$$

while

$$y = y(x) \text{ for } x \in X.$$

The factor $J = (dy/dx)$ is the simplest possible example of what is known more generally as a Jacobian for the transformation.

Let us explore the more general situation. Suppose events can be described by n -dimensional vectors $\bar{x} = (x_1, \dots, x_n)$ of random variables. They span an n -dimensional space X . Consider also the collection of n -dimensional vectors

$$\bar{y} = \bar{y}(\bar{x}) \text{ (or } y_i = y_i(\bar{x}) \text{ for } i = 1, n)$$

of random variables which form Y . X is mapped into Y , one-to-one. The transformation is assumed to be continuous and all partial derivatives $\partial y_i / \partial x_j$ exist on X . Furthermore, assume that the determinant J of the matrix of partial derivatives,

$$J = \det \begin{bmatrix} \partial y_1 / \partial x_1 & \dots & \partial y_1 / \partial x_n \\ \vdots & \dots & \vdots \\ \partial y_n / \partial x_1 & \dots & \partial y_n / \partial x_n \end{bmatrix},$$

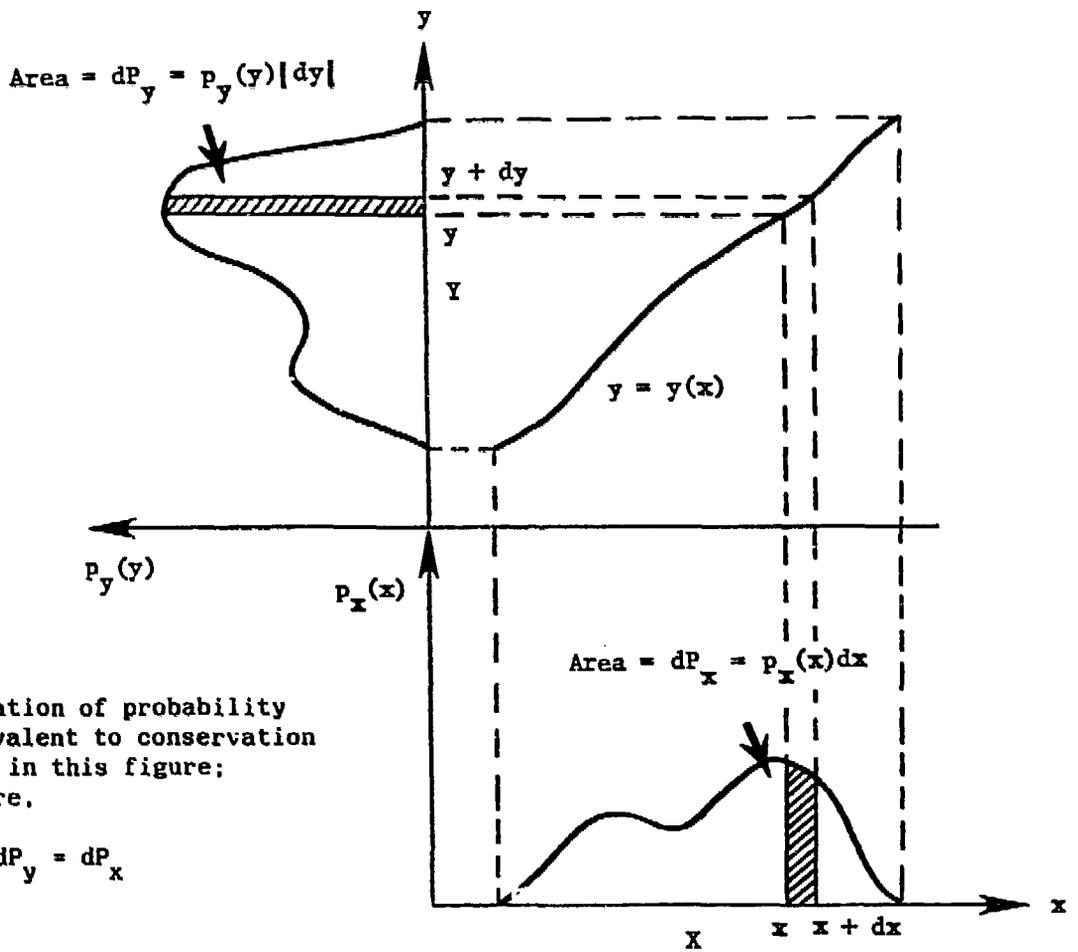


Figure 6.2: Graphical illustration of probability transformation.

namely, the Jacobian for this transformation, is nonvanishing. Then the probability transformation law is

$$p_y(\bar{y}) = p_x(\bar{x})/|J| \quad (6.17)$$

Eq. (6.16) is a special case of Eq. (6.17). Note that J can vary with \bar{x} ; it is not necessarily a constant. Since the inverse transformation exists under these conditions, we have $\bar{x} = \bar{y}^{-1}(\bar{y})$ for $\bar{y} \in Y$ while $\bar{y} = \bar{y}(\bar{x})$ for $\bar{x} \in X$.

Example 6.2

Consider the density function

$$p_x(x_1, x_2) = (1/\pi)\exp(-x_1^2 - x_2^2)$$

for $-\infty < x_1 < +\infty$ and $-\infty < x_2 < +\infty$.

Let $y_1 = x_1 + x_2$ and $y_2 = x_1 - x_2$. The partial derivatives of the transformation are:

$$\partial y_1 / \partial x_1 = 1,$$

$$\partial y_1 / \partial x_2 = 1,$$

$$\partial y_2 / \partial x_1 = 1,$$

$$\partial y_2 / \partial x_2 = -1.$$

Therefore, the Jacobian J is -2 . The inverse transformation exists and takes the form: $x_1 = (y_1 + y_2)/2$ and $x_2 = (y_1 - y_2)/2$.

According to Eq. (6.17), $p_y(\bar{y})$ can be derived from $p_x(\bar{x})$. The result of this analysis is:

$$\begin{aligned} p_y(y_1, y_2) &= (1/2)p_x[(y_1 + y_2)/2, (y_1 - y_2)/2] \\ &= (1/2\pi)\exp[-(y_1^2 + y_2^2)/2]. \end{aligned}$$

The concepts discussed in this chapter apply to all probability laws which satisfy certain fundamental requirements. Very little has been said about specific probability laws, e.g., their origins and applicability to the

description of real random phenomena. We will begin to explore this topic in earnest in Chap. 7, and will continue the discussion in later reports of this series. To close the present chapter, we mention for interest that Karl Pearson, one of the founders of modern statistics, discovered that solutions to the differential equation

$$\frac{1}{p(x)} \frac{d [p(x)]}{dx} = \frac{d - x}{a + bx + cx^2} \quad (6.18)$$

for various choices of the constants a, b, c, and d, remarkably yield most of the univariate probability density functions of importance in modern applications. We will demonstrate this fact later on appropriate occasions.

We now touch upon a point of particular interest to the nuclear data field because it is so often encountered in practice. Usually in statistics it is preferable to represent event spaces which appear to be effectively uncountable (e.g., a collection of atoms in a macroscopic sample) by continuous random variables. In nuclear data applications, however, one often proceeds in the opposite direction. For example, neutron energy is seemingly very well represented by a continuous random variable, and thus so are neutron fluence or a particular neutron cross section which is functionally related to the energy. However, it is common practice in reactor physics studies to generate an energy-group structure by dividing the neutron energy range of interest (e.g., 1×10^{-5} eV to 20 MeV) into a sequence of contiguous energy intervals called groups. These groups are defined by their upper and lower energies, called group limits. Sometimes a median energy is also identified for each group. Functions dependent upon neutron energy are also represented by a group structure. For example, consider the neutron fluence density function Ψ which is a function of neutron energy E. Associated with each energy interval ($E_{\ell i}, E_{hi}$) there is defined a group fluence Ψ_i . In order to conserve neutron number it is required that

$$\Psi_i = \int_{E_{\ell i}}^{E_{hi}} \Psi(E) dE \quad (6.19)$$

Consequently, if Ψ is the total neutron dose, then

$$\Psi = \sum_{i=1}^n \Psi_i = \int_{10^{-5} \text{ eV}}^{20 \text{ MeV}} \Psi(E) dE \quad (6.20)$$

The reason this is done is partly a matter of tradition in the field, but mostly it is for very practical reasons. In order to solve the complex differential and integral equations which often arise in reactor neutron transport studies, it is very useful to resort to finite-difference methods (e.g., Hil 52) which can be readily programmed on a digital computer. These methods require selection of a group structure and, subsequently, a group

representation of such pertinent energy-dependent parameters as neutron fluence and cross sections.

In neutron transport problems it is often necessary to know the probability of a neutron being found in a particular energy group. A neutron energy probability density function can be readily derived from the fluence function. It is

$$p(E) = \Psi(E)/\Psi.$$

However, for analysis involving a group-structure formulation, we require instead the mass distribution function (as defined in Sec. 3.3), where $P_i = \Psi_i/\Psi$. Clearly, the group-structure mass distribution function and the underlying density function are related via Eq. (6.19).

We will also see in a later report of this series that it is usually desirable to cast nuclear data evaluation endeavors into group formats even though the parameter being evaluated is naturally amenable to representation by a continuous random variable.

7. SOME IMPORTANT PROBABILITY LAWS GOVERNING THE FREQUENCY OF OCCURRENCE OF RANDOM EVENTS

In this chapter the emphasis shifts from consideration of general characteristics of probability laws to a discussion of some specific laws that are important to nuclear science. In order to be legitimate, all probability laws must satisfy the three basic axioms discussed in Sec. 4.2. From a mathematical point of view, discrete random variables present few technical difficulties. For continuous random variables, one generally deals with probability laws which are described by functions that are continuous and differentiable, except possibly at a few jump points. Not all functions that possess either necessary or desirable mathematical properties are useful in practice. In the final analysis the only laws that are worthy of consideration are those which have been found to provide a description of the actual behavior of random phenomena. Therefore, applied statistics is necessarily linked to the physical process of observation.

Our particular concern is with the description of random events in the field of nuclear science. In many ways, nuclear phenomena are unique. Small samples of material are composed of enormous numbers of individual atoms, yet the quantities of energy involved in nuclear transformations are usually extremely large, relatively speaking. Therefore, it is often possible to observe a single nuclear transformation at a macroscopic level, e.g., a track in a bubble chamber or the click of a Geiger counter. It should come as no surprise to the reader that the most important probability laws affecting nuclear phenomena have their origins in consideration of the frequency of occurrence of certain well-defined random quantum processes.

The discussion here will be limited to consideration of the following laws:

- i) the binomial law and its generalization the multinomial law (Sec. 7.1),
- ii) the Poisson law (Sec. 7.2).
- iii) the normal (Gaussian) law (Sec. 7.3).

These laws are related, and some of these relationships are explored at appropriate points in this chapter. Discussion of other probability laws that are useful in nuclear applications is deferred to future reports.

7.1 Binomial and Multinomial Probability Laws

The fundamental concepts leading to the binomial probability law were introduced earlier in this report (Secs. 2.3, 3.1, 3.2, 3.3, and 5.1, and Ex. 5.3). Here, the essential points are restated, and from those the binomial probability function is deduced. It is appropriate to designate it as a mass function because the random variable is discrete.

Consider an event space E consisting of elementary events e . Consider a partition of E into two spaces A_1 and A_2 . That is,

$$A_1 + A_2 = E,$$

$$A_2 = \bar{A}_1,$$

and

$$A_1 \times A_2 = \phi.$$

Now consider forming a new space E' in which the elementary events e' are ordered n -tuples of events taken from E . These n -tuples are formed by conducting n independent trials (experiments) in which events are selected (sampled) from E . Now, assume that all we are really interested in with regard to this process is whether any particular elementary event e of E possesses attribute A_1 (belongs to A_1) or possesses attribute A_2 (belongs to A_2). An essential assumption is that the experimental process of forming an n -tuple does not perturb (deplete) event space E . As indicated previously, one way to do this is to "replace" a particular event e after it is selected. The result is the same if E contains so many events that formation of an n -tuple does not alter the character of E . Let p be a probability function defined on E . Then $p(A_1)$ is the probability that a sampling experiment results in A_1 , while $p(A_2)$ is the probability that A_2 is obtained. They are designated p_1 and p_2 for convenience. We permit ourselves the indulgence of using lower case "p" to designate probability in contrast with our previous convention of reserving lower-case symbols for density functions. All of the n trials are identical in nature as well as independent, and the assumption that the character of E remains unperturbed by the sampling process leads us to the conclusion that for any particular trial, the probability is p_1 for observing A_1 and p_2 for observing A_2 . Naturally, $p_1 + p_2 = 1$ because A_1 and A_2 form a partition of E . Now we turn to consideration of the n -tuples e' and the space E' formed from from them. We ask: "What is the probability that n trials, as described above, will result in k_1 occurrences of A_1 ($k_1 \leq n$) and k_2 occurrences of A_2 , with $k_1 + k_2 = n$?" We denote this probability mass function by the expression $p(n; p_1, p_2; k_1, k_2)$, with k_1 the random variable which can assume $n + 1$ integer values between 0 and n (k_2 is derived from k_1). In fact, this is the binomial law we seek. Again, we have used "p" to designate probability functions for two different spaces, but in this field it is a common practice to which the reader should become accustomed.

Before answering the question stated in the preceding paragraph, we pose and answer a simpler question: "What is the probability that in n trials, the first k_1 in a row will yield A_1 , while the remaining $k_2 = n - k_1$ will

produce A_2 ?" The probability of such an event e' is simply $p_1^{k_1} p_2^{k_2}$, by virtue of independence and the identical nature of each trial. Since we really don't care about the order of the occurrences of A_1 and A_2 , $p(n; p_1, p_2; k_1, k_2)$ should, in general, be larger than this simple product of probabilities, reflecting existence of other ordered n -tuples with k_1 of A_1 and k_2 of A_2 which have the same probability $p_1^{k_1} p_2^{k_2}$. Thus,

$$p(n; p_1, p_2; k_1, k_2) = C_{nk_1} p_1^{k_1} p_2^{k_2} \quad (7.1)$$

where C_{nk_1} is the binomial coefficient from Eq. 2.21. Recall that C_{nk_1} reflects the number of ways k_1 indistinguishable events (occurrence of A_1) can be distributed among n positions in an ordered n -tuple. The binomial coefficient C_{nk_1} is identical to the multinomial coefficient $C_{n; k_1 k_2}$ defined in Eq. (2.30), so we can also write Eq. (7.1) in the form

$$p(n; p_1, p_2; k_1, k_2) = C_{n; k_1 k_2} p_1^{k_1} p_2^{k_2} \quad (7.2)$$

The detailed process described above is known as a Bernoulli experiment, after the eighteenth century mathematician Jacques Bernoulli.

The notation leading up to Eqs. (7.1) and (7.2) was selected for reasons that will be clear in the development below, but it is not conventionally used. More commonly, A_1 is designated as S (success) and A_2 as F (failure). The probability p_1 is simply designated as p , while $p_2 = 1 - p$. If $k_1 = k$ and $k_2 = n - k$, then one obtains the more familiar expression for the binomial probability law:

$$p(n, p; k) = C_{nk} p^k (1 - p)^{n - k} \quad (7.3)$$

While the concepts and notation surrounding the binomial law are fresh in mind, we proceed to a discussion of the multinomial probability law. Indeed, the concepts are identical. The only difference is that instead of two possibilities, A_1 and A_2 , one has r possibilities A_1, A_2, \dots, A_r ($r \leq n$) for the outcome of an individual trial. Thus,

$$\sum_{i=1}^r A_i = E,$$

and the A_i are all mutually exclusive (they form a partition of E). The nomenclature evolves from the case of two possibilities to r possibilities in an obvious way, with p_1, p_2, \dots, p_r representing probabilities for observing A_1, A_2, \dots, A_r at each trial, respectively. It should come as no surprise that the multinomial probability law assumes a form similar to Eq. (7.2), namely:

$$p(n; p_1, p_2, \dots, p_r; k_1, k_2, \dots, k_r) = C_{n; k_1 k_2 \dots k_r} p_1^{k_1} p_2^{k_2} \dots p_r^{k_r} \quad (7.4)$$

with

$$\sum_{i=1}^r k_i = n$$

Here, $p(n; p_1, p_2, \dots, p_r; k_1, k_2, \dots, k_r)$ is the probability that n trials, in an environment where r outcomes are possible, will generate k_1 of A_1 , k_2 of A_2, \dots, k_r of A_r . The multinomial coefficient $C_{n; k_1 k_2 \dots k_r}$, given by Eq.

2.30, represents the number of ordered n -tuples that satisfy our particular requirement, which has no concern for ordering.

Example 7.1

We now demonstrate some properties of the binomial probability law, as manifested in Eq. (7.3), by considering specific numerical values for p and n . In Fig. 7.1, values of $p(n, p; k)$, for $n = 20$, $k = 1$ to 20 and $p = 0.1, 0.3$ and 0.5 , are plotted (taken from Fig. 63, p. 133). Since the mass function is defined only for integer values k , the lines drawn between the mass points are intended only to serve as eyeguides. Three features of the binomial law are clearly evident in this figure: First, we note that for $p = 0.5$ the mass function is symmetric. Next, we note that the mass function appears to reach a maximum value for $k \approx pn$. Finally, it is evident that the "areas" under the plotted distributions are the same. This follows from Axiom II (Sec. 4.2) which requires that probability be normalized. In fact, from the binomial theorem, Eq. (2.25), it is clear that

$$\sum_{k=0}^n p(n, p; k) = 1 \text{ for every integer } n, \text{ and } p \leq 1.$$

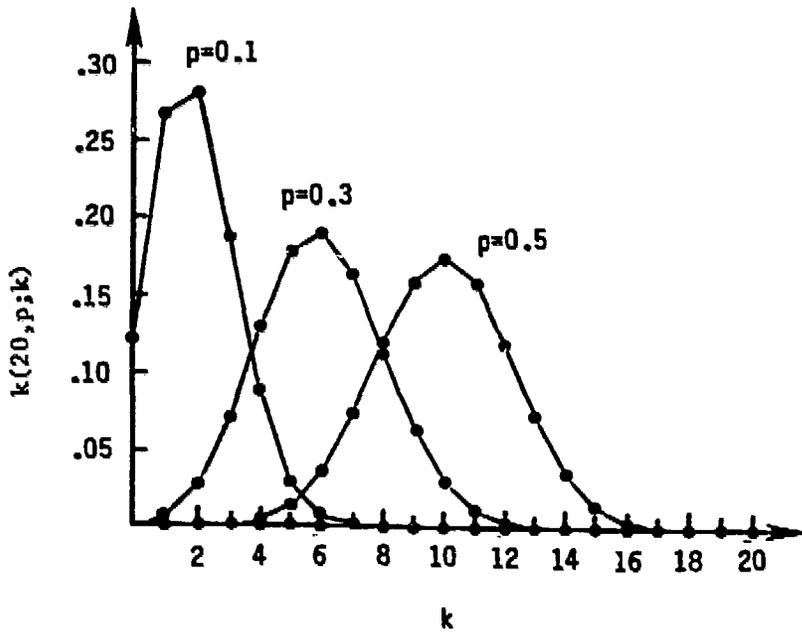


Figure 7.1: Plots of binomial probability mass functions $p(n,p;k)$ for $n = 20$, $k = 1$ to 20 and $p = 0.1, 0.3$ and 0.5 . Solid lines connecting mass points are there merely to serve as eyeguides. Figure is taken from Fis 63, p. 133. See Ex. 7.1.

Example 7.2

Now consider $p(n,p;k)$ for $p = 0.5$ and $n = 10$ and 20 . Each of these two mass functions is symmetric. For $n = 10$, symmetry is about $k = 5$, while for $n = 20$ symmetry is about $k = 10$. For each of these probability laws we ask the question: "What is the probability that k will equal or exceed by 20% the particular value of k corresponding to symmetry?" For $n = 10$, the desired probability is

$$\sum_{k=7}^{10} p(10,0.5;k) = 0.1719.$$

However, for $n = 20$, the desired probability is

$$\sum_{k=14}^{20} p(20,0.5;k) = 0.0577.$$

A general conclusion can be drawn from this comparison: As n increases, the probability of observing a number k of successes which exceeds the product pn by some fixed percentage decreases. Thus, the larger n becomes, the less likely one is to observe success fractions, for an n -fold collection of outcomes, which differ significantly from the probability p of observing a success when selecting a single event at random from E (a trial). This result, known as Bernoulli's Law of Large Numbers, is a feature of the binomial probability law. It is an intuitively reasonable result, for when n is large, the ensemble of results from n trials tends to mirror the larger event space. In essence, it approaches a population.

Laplace suggested a rather simple mathematical model, based on the binomial probability law, which can be used to describe how random error comes about in measurement processes. The notions behind this model, known as the Laplace Law of Error, are similar to those which lead to the Central Limit Theorem, a result which justifies widespread use of the normal probability law in statistics. Discussion of the Central Limit Theorem is deferred to Section 7.4. Here, we focus specifically on the Laplace Law of Error.

Let the true value of a parameter one seeks to measure be m_0 . In reality, one usually does not measure precisely m_0 but, rather, other values m , because the measurement process is perturbed by a number of unknown independent causes. Laplace assumed that n disturbances occur, with each disturbance contributing a deviation of the same magnitude ϵ . However, these "identical" disturbances could be of either a positive or negative nature

($\pm \epsilon$). Let $m(n, \epsilon; k)$ be an actual value of m measured when the n disturbances involve k of $+\epsilon$ and $(n - k)$ of $-\epsilon$. Thus,

$$m(n, \epsilon; k) = m_0 + k\epsilon - (n - k)\epsilon = m_0 + (2k - n)\epsilon \quad (7.5)$$

We now ask about the probability of observing the various values of m obtained from Eq. (7.5). As indicated above, this model of error assumes that $+\epsilon$ and $-\epsilon$ disturbances are equally likely. Clearly, this model establishes conditions identical to those upon which the binomial probability law is based. Consequently, the distribution for the $m(n, \epsilon; k)$ of Eq. (7.5) is

$$p(n, 1/2; k) = C_{nk} (1/2)^k (1/2)^{n-k} = C_{nk} (1/2)^n \quad (7.6)$$

The quantities $p(n, 1/2; k)$ are the basis for the interesting array of numbers shown in Fig. 7.2. The similarity of this array to the Pascal Triangle given in Fig. 2.13 is evident.

Example 7.3

Suppose $\epsilon = 0.01$ (1% disturbance) and $n = 20$. What is the probability of measuring a value m which deviates by at least 3ϵ from m_0 ? Since the binomial mass function is symmetric about

$$k = 10, \text{ the probability we desire is } 2 \left[\sum_{k=13}^{20} p(20, 1/2; k) \right] = 0.2632$$

($\sim 26\%$).

The following example raises an important physical problem that illustrates several concepts already discussed in this report. We examine a procedure for calculating the relative probabilities of populating the ground state and an isomeric level of a nucleus formed in a highly excited state by slow-neutron capture. De-excitation of the nucleus proceeds by a sequence of electromagnetic (EM) quantum transitions which are random in nature. This process has been investigated in detail by Poenitz (Poe 66). Here we discuss a highly simplified version of this problem for demonstrative purposes.

Example 7.4

Nuclear EM decay is governed by transition energy, spin-selection rules, and multipole-radiation probabilities. Models exist for calculating probabilities of electric- and magnetic-multipole transitions between two nuclear states, given the spins and parities of the states and their energy separation. We avoid these complexities by considering a hypothetical nucleus and a simple transition-probability law. The details of our model are as follows:

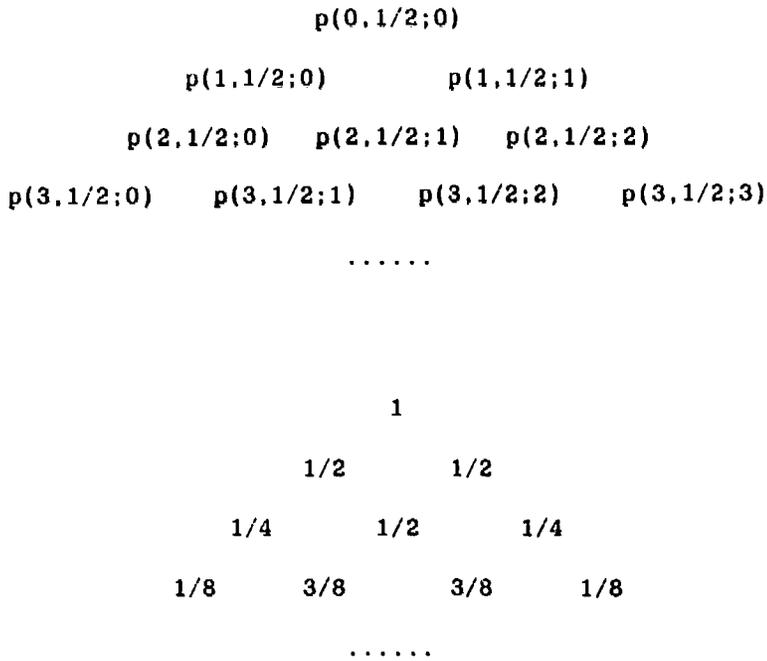


Figure 7.2: Array of values $p(n, 1/2; k)$ from Eq. (7.6) for $n = 0, 1, 2, \dots$ and $k = 0, 1, \dots, n$. Note similarity of this array to the Pascal Triangle from Fig. 2.13.

1) The nucleus:

It is assumed to have a ground state with spin $J = 0$ and four bands of excited states. The k th excited-state band has $k + 1$ levels with spins $J = 0, \dots, k$, respectively. We ignore the issue of parity. The spacing between levels within each excited-state band is assumed to be ΔE which is very small compared with the uniform inter-band spacing E . The excited-state bands are thus nearly equally spaced above the ground state. The level with spin $J = 0$ belonging to the first excited-state band is taken to be the isomer. See Fig. 7.3.

2) Neutron capture process:

Neutron capture is assumed to populate initially only the levels within the fourth excited-state band. The capture probabilities are considered to be proportional to the neutron-capture cross sections for each level within this band. For present purposes we employ capture cross sections $\sigma(J)$ appearing in Table 7.1. We suppose that resonance effects are largely responsible for the wide variation in these cross sections.

3) EM-transition selection rule:

The only transitions with significant probability are assumed to be inter-band, with spin change $\Delta J = 1$. Other possibilities are at the very least highly suppressed, if not forbidden. It is therefore clear why the $J = 0$ level of the first excited-state band is an isomer.

4) Transition probability rules:

An EM transition from state of spin J can proceed only to a state $J' = J + 1$ or $J' = J - 1$, by Item (3) above. When only one transition channel is open, we assume the transition probability is unity (an excited state must decay). When two decay channels are open (Bernoulli problem), then it is assumed that each has an equal probability of $1/2$.

The decay paths permitted by the preceding rules are shown in Fig. 7.4. From this figure it is evident that neutron capture for all even- J levels leads ultimately to population of the ground state, while capture for all the odd- J levels populates only the isomer. Starting at the capture level, EM decays produce a cascade "tree" in which each successive step of the cascade increases the range of accessible spin- J states. The probabilities for populating these states would indeed generate an array of binomial coefficients identical to Fig. 7.2 were it not for the fact that the number of available spin states in this particular problem becomes successively more limited at each stage of the cascade.

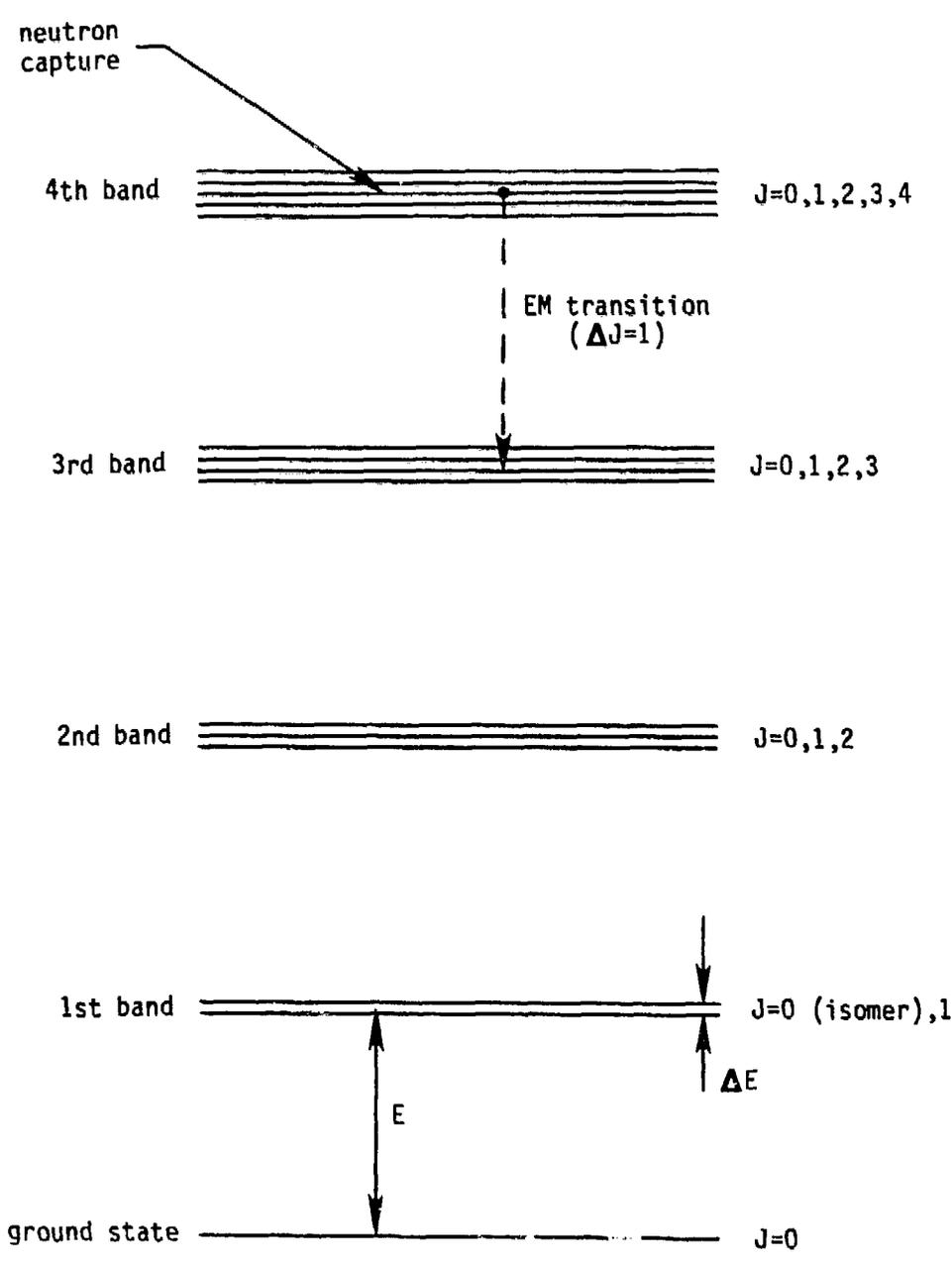


Figure 7.3: Level structure of the hypothetical nucleus which is considered in Example 7.4.

Table 7.1 Parameters Used in the Analysis of Example 7.4

J	$\sigma(J)^a$	$P(J)^b$	$P(\text{g.s./}J)^c$	$P(\text{isom./}J)^c$
0	8.5	0.0404	1	0
1	178.0	0.8468	0	1
2	22.0	0.1047	1	0
3	1.0	0.0048	0	1
4	0.7	0.0033	1	0

^a $\sigma(J)$ is the neutron capture cross section for the spin- J member of the fourth excited-state band of the hypothetical nucleus of Ex. 7.4. The cross section units are unimportant.

^b
$$P(J) = \sigma(J) / \sum_{J'=0}^4 \sigma(J').$$

^c Refer to Fig. 7.4.

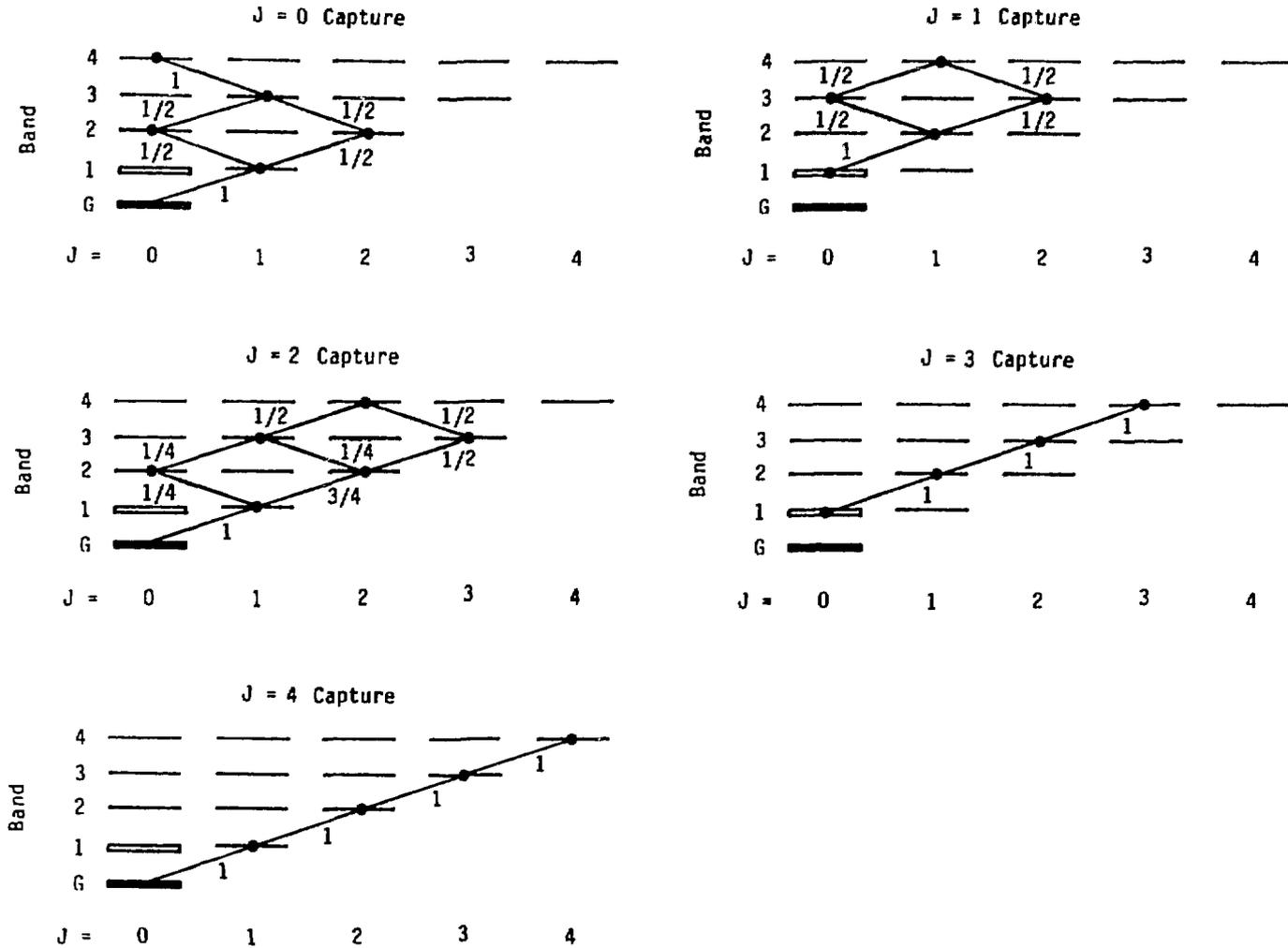


Figure 7.4: Decay paths permitted by the $\Delta J = 1$ selection rule, and their associated probabilities. Ground state is indicated by a darkened rectangle, while the isomer level is the open rectangle. See Ex. 7.4.

The rule of total probability, Eq. (5.8), is very useful in calculating the probability $P(\text{g.s.})$ of populating the ground state and $P(\text{isom.})$ of populating the isomer. Applied here, we have

$$P(\text{g.s.}) = \sum_{J=0}^4 P(\text{g.s.}/J)P(J)$$

and

$$P(\text{isom.}) = \sum_{J=0}^4 P(\text{isom.}/J)P(J).$$

$P(J)$ is the relative probability for neutron capture into the level with spin J in the fourth excited-state band. It is given by the formula

$$P(J) = \sigma(J) / \sum_{J'=0}^4 \sigma(J'),$$

where $\sigma(J)$ represents the previously mentioned neutron-capture cross section. $P(\text{g.s.}/J)$ and $P(\text{isom.}/J)$ are conditional probabilities for populating the ground state or isomeric state, respectively, given that neutron capture involved the spin- J level of the fourth excited state band. Table 7.1 summarizes all parameters needed for the analysis. The final result is, therefore, $P(\text{g.s.}) = 0.1484$ and $P(\text{isom.}) = 0.8516$.

In this particular example, capture into the $J = 1$ state of the fourth excited-state band is dominant. This qualitatively explains why this particular capture process results mainly in production of nuclei in the isomeric state.

We close this section by giving an example which illustrates use of the multinomial probability law.

Example 7.5

Consider an experiment with four independent detectors that are multiplexed into a single computer-based data acquisition system, as shown in Fig. 7.5. Data from the measurement apparatus are formulated into computer words consisting of two parts. The first part contains analog (A) information (e.g., pulse heights, timing information, etc.). The second part is a tag (T) which identifies the detector of origin of the particular event e which is characterized by the array (A,T). Note that $T = 1, 2, 3, \text{ or } 4$, since there are four detectors. We assume that the apparatus is adjusted (e.g., timing, gating, etc.) in such a way that every data word accepted by the computer represents a legitimate event

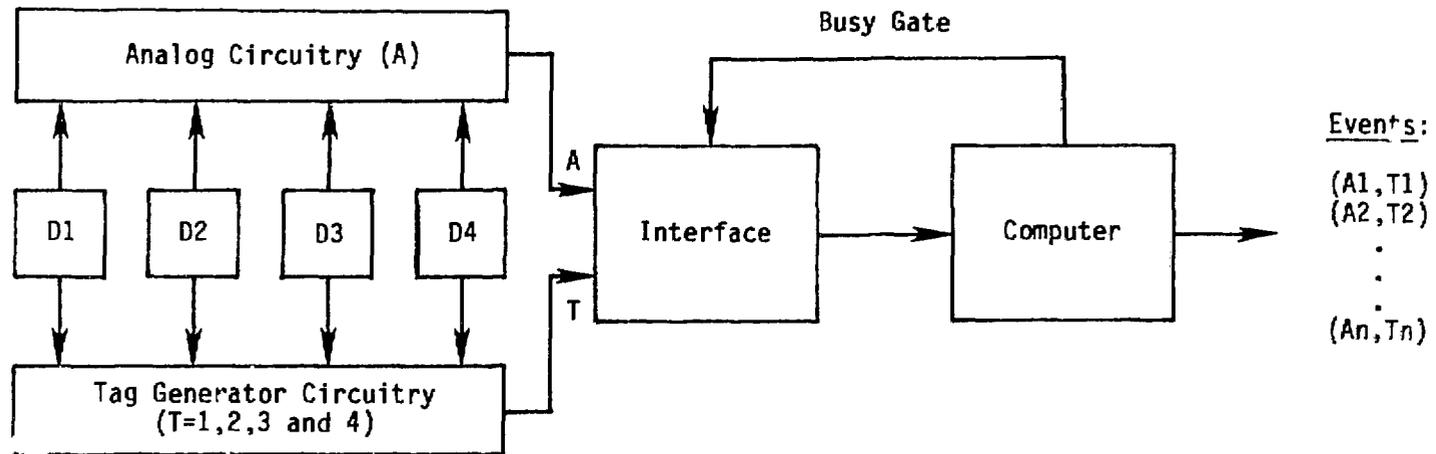


Figure 7.5: Schematic diagram of the experiment described in Example 7.5.

(i.e., no "defective" data are recorded). Furthermore, we assume that the natural system deadtime (due primarily to the time needed by the computer to process the events) introduces a bias against any particular detector (i.e., all detectors are equally afflicted). The data "words", (A,T), are recorded in sequence, e.g., on magnetic tape, word by word, in what is commonly referred to as the event mode. After the experiment, the event words are read back by a computer (playback) and combined (processed) according to the specific objectives of the experiment. The question we pose is the following: "What is the probability that after recording a sequence of n events, k_i of these will have tag

$T = i$ ($i = 1,4$), where $\sum_{i=1}^4 k_i = n$?" The conditions of the present

experiment satisfy the requirements for applicability of the multinomial law; therefore, the answer to the question we have posed is given by Eq. (7.4), with $r = 4$. The probabilities p_i ($i = 1,4$) are surely related to the radiation fluences Ψ_i and detector efficiencies ϵ_i applicable to each detector, as well as to the live-time fractions, λ_i . We therefore assume that

$$p_i = (\Psi_i \epsilon_i \lambda_i) / \sum_{j=1}^4 (\Psi_j \epsilon_j \lambda_j).$$

The λ -factors ultimately cancel because we have assumed that $\lambda_i = \lambda$ ($i = 1,4$). Suppose that $n = 12$ and $p_i = 1/4$ ($i = 1,4$). What is the relative probability that one will observe $k_1 = 1$, $k_2 = 2$, $k_3 = 4$, $k_4 = 5$, versus that of observing all $k_i = 3$ ($i = 1,4$)? From Eq. (7.4)

$$p(12;1/4,1/4,1/4,1/4;1,2,4,5) = (12!/1!2!4!5!) (1/4)^{12} \\ = 83160 (1/4)^{12}$$

and

$$p(12;1/4,1/4,1/4,1/4;3,3,3,3) = (12!/3!3!3!3!) (1/4)^{12} \\ = 369600 (1/4)^{12}$$

Thus, it is more probable by a factor of ~ 4.44 that an equal number of events will be observed from each detector than the other possibility indicated above.

In reality, the binomial probability law is rarely employed explicitly in statistical applications. Its significance stems from the fact that the more important Poisson and normal (Gaussian) probability laws are, in fact, limiting cases of the binomial law. This matter is explored in Sec. 7.4.

7.2 Poisson Probability Law

Suppose we wish to analyze a problem involving a large number n of independent random trials, with a small probability p that an event A will occur on a single trial. Furthermore, suppose that we are mainly interested in the probability of exactly k occurrences of A in n trials when $k \ll n$. In principle, the binomial probability function of Eq. (7.3) is applicable in this situation, but it is extremely awkward, if not impossible, to use for the given conditions. This will be apparent to the careful reader who worked through some of the examples given in the preceding section. The nineteenth century French mathematician S. Poisson suggested a probability law, known as the Poisson probability law in his honor, which turns out to be a good approximation to the binomial law under these conditions.

There are various ways to derive the Poisson probability law. Below, we discuss the most commonly used approach. This method explicitly demonstrates that the Poisson law is indeed a good approximation to the binomial law under the conditions indicated in the preceding paragraph. However, it is important to realize that the Poisson probability law is a distinct probability law which need not be considered as just a limiting case of the binomial probability law. In fact, this law has been demonstrated to be applicable to the analysis of certain random phenomena for which the conditions on the number of trials and probability per trial are not as severe as implied in the following derivation. We demonstrate below that this law satisfies the basic axioms of probability. Its importance in applied statistics stems from the fact that it seems to describe very well a large number of random phenomena of practical interest, especially in the area of nuclear processes.

Theorem: [Poisson Approximation]

Let n be a very large number of independent trials in which the probability p of occurrence of an event in a single trial is very small, i.e., $n \gg 1$ and $p \ll 1$. Let $\lambda = np$. Then, for $k \ll n$, the binomial mass function, Eq. (7.3), is approximated very well by the expression

$$p(n,p;k) \approx \lambda^k e^{-\lambda} / k! \quad (7.7)$$

Proof of the theorem involves appropriately grouping some factors and then making certain limiting approximations. Starting with binomial formula, Eq. (7.3), we carry out the following algebraic manipulations:

$$\begin{aligned}
p(n, p; k) &= C_{nk} p^k (1 - p)^{n - k} \\
&= \frac{n!}{k!(n - k)!} p^k (1 - p)^{n - k} \\
&= \frac{n(n - 1)(n - 2) \dots (n - k + 1)}{k!} p^k (1 - p)^{n - k} \\
&= \frac{n(n - 1)(n - 2) \dots (n - k + 1)}{k!} (\lambda/n)^k \frac{(1 - \lambda/n)^n}{(1 - \lambda/n)^k} \\
&= \frac{n(n - 1)(n - 2) \dots (n - k + 1)}{n^k} \frac{\lambda^k (1 - \lambda/n)^n}{k! (1 - \lambda/n)^k} \\
&= \frac{\lambda^k}{k!} (1 - \lambda/n)^n \left[\frac{1(1 - 1/n)(1 - 2/n) \dots [1 - (k - 1)/n]}{(1 - \lambda/n)^k} \right].
\end{aligned}$$

It is well known that

$$\lim_{n \rightarrow \infty} (1 - \lambda/n)^n = e^{-\lambda},$$

so for large n we therefore can approximate the factor $(1 - \lambda/n)^n$ by $e^{-\lambda}$. We now notice that most of the remaining factors are near unity since $k \ll n$ and $\lambda \ll n$. This proves the result stated in Eq. (7.7).

Since in the binomial probability law p can range from 0 to 1, another way to view the Poisson law is that it becomes a good approximation to the binomial law whenever the number of trials n is large and binomial law is very asymmetric, i.e., when p is $\ll 1$, and thus is far removed from $p = 0.5$, the point of symmetry.

An important difference between the Poisson probability law and binomial probability law is that in the former a single parameter λ explicitly replaces the two parameters n and p of the latter. However, λ is still interpreted as the product np in the limit of large n and small p . This transformation from n and p to λ via the product $\lambda = np$ is known as the Poisson condition or Poisson transformation. Clearly, it is essential to both the binomial and Poisson probability laws that there exists a fundamental probability p for the occurrence of an event A in any single random trial. This probability is an intrinsic feature of the space of events. Probability is not an observable, but the quantity k/n is, where k is the number of occurrences of event A , and n is the sample size. Our theory suggests that, for large n , k/n is likely to be close to p , regardless of n . The parameter λ , of course, depends on n and is therefore not as

fundamental as p . The reader will recall from the discussion of Bernoulli's law of large numbers (in Example 7.2) that, for a large number of independent trials n , the probability of finding that A has occurred k times is small if k is substantially different from np , i.e., from λ . This is just an equivalent way to state the particular concept under consideration. The Poisson law is therefore understood as that law which provides probabilities associated with various values of the integer random variable k , which represents the observed frequency of occurrence of A after a large number of trials, given that λ is known to be the most likely frequency of occurrence of A . In applications one is generally interested in the probabilities for various k in the vicinity of λ . However, the Poisson probability law is formally defined for all non-negative integers k , and for arbitrary positive real numbers λ , by the following mass function:

$$p(\lambda; k) \equiv \lambda^k e^{-\lambda} / k! \quad (\lambda > 0 \text{ and } k = 0, \infty) \quad (7.8)$$

Theorem:

The Poisson formula of Eq. (7.8) satisfies the basic axioms of probability, as described in Section 4.2.

To prove this, we demonstrate that the basic axioms are satisfied. Clearly, $p(\lambda; k) \geq 0$ for all allowed λ and k , so Axiom I is satisfied. The event space to which this probability applies is the set of all non-negative integers k . Each such event is mutually exclusive in the sense that for a particular set of n random trials one cannot observe both k occurrences of A and k' occurrences of A if $k' \neq k$. However, the probability of observing either k occurrences of A or k' occurrences of A is implicitly taken to be the sum $p(\lambda; k) + p(\lambda; k')$, in compliance with Axiom III. The theorem is proved if we can demonstrate normalization (Axiom II), i.e., if

$$\sum_{k=0}^{\infty} p(\lambda; k) = 1.$$

This is a straightforward task. From Eq. (7.8),

$$\sum_{k=0}^{\infty} p(\lambda; k) = \sum_{k=0}^{\infty} [\lambda^k e^{-\lambda} / k!] = e^{-\lambda} \sum_{k=0}^{\infty} (\lambda^k / k!) = e^{-\lambda} e^{\lambda} = 1,$$

from the well-known series expansion for the exponential factor e^{λ} .

One application of the Poisson law is in the calculation of failure frequencies for like components which are fabricated in large quantities and are known to fail relatively infrequently. This is demonstrated by the following example.

Example 7.6

Suppose that experience has shown that for a certain type of precision resistor obtained from a particular supplier, an intolerable defect (e.g. departure of the actual resistance from the nominal value by more than 20%) is encountered at an average frequency of 10 defects per 10^5 ; thus, $p = 10^{-4}$ (probability that any particular resistor will be defective) and $\lambda = np = 10$. The probabilities for observing defect numbers k ranging from zero to 20 per 10^5 , in any large batch obtained from this supplier, should be represented well by the Poisson law. These probabilities are listed in Table 7.2.

Although it might be known a priori that the Poisson probability law is applicable (due to the nature of the independent trials), the parameter λ must be deduced from experimentation in any practical application. For example, it would have been impossible to perform the analysis indicated in Example 7.6 without the knowledge that $\lambda \approx 10$, based on prior experience. This situation illustrates a basic feature of applied statistics, namely, that the fundamental parameters of any probability law must ultimately be deduced experimentally.

Example 7.6 also demonstrates that the probability mass function indeed peaks for $k = \lambda$. It also appears to be somewhat skewed, with larger probabilities for $k = \lambda - i$ than for $k = \lambda + i$, when i is a positive integer. This is a manifestation of the following theorem of Poisson probability which is presented without proof:

Theorem:

Let $\lambda = m$, an integer. Let i be another integer such that $1 \leq i \leq m$. Then, for the Poisson probability law, $p(m; m - i) > p(m; m + i)$.

We also note from Table 7.2 that for $\lambda = m$, an integer, $p(m; m - 1) = p(m; m)$. This is a general result which can also be expressed as a theorem:

Theorem:

Let $\lambda = m$, with m an integer greater than 1. Then, for the Poisson probability law, $p(m; m - 1) = p(m; m)$.

To prove this refer to Eq. (7.8). Thus,

$$\begin{aligned} p(m; m - 1) &= m^{m-1} e^{-m} / (m - 1)! \\ &= (m)m^{m-1} e^{-m} / [(m)(m - 1)!] \\ &= m^m e^{-m} / m! = p(m, m). \end{aligned}$$

When λ becomes very large ($\gg 1$), then the Poisson mass function approaches a symmetric shape for k in the vicinity of λ . This result follows from the next theorem which is offered here without proof:

Theorem:

Let λ be a large positive number ($\gg 1$). For all intents and purposes, we suppose that it is an integer. We also suppose that $k = \lambda + m$ and $k' = \lambda - m$, where m is a non-negative integer such that $m \ll \lambda$. Then

$$p(\lambda; k')/p(\lambda; k) = \lambda^{-2m} (\lambda + m)! / (\lambda - m)!.$$

This ratio is close to unity under the conditions stated here. The result is easy to demonstrate. Suppose, for example, that $m = 2$ and $\lambda \approx 1000$, then $p(\lambda; k')/p(\lambda; k) \approx 1.002$.

Since radioactive decay involves a very large number of atoms, and the probability is small that a particular atom will decay during a time interval which is short compared with the decay half life, one would expect that Poisson statistics should apply to such problems. Through a series of very thorough experiments conducted early in this century, E. Rutherford and coworkers demonstrated the validity of this contention. Indeed, the Poisson probability law is now considered to be the premier statistical law of nuclear science. The Law of Exponential Decay of radioactivity follows from similar assumptions. Consider an interval of time Δt which is much smaller than the decay half life for the atoms in a radioactive sample. Let p represent the probability per unit time that a particular atom will decay. Then $p\Delta t$ is the probability that it will decay in time Δt . Since it is assumed that the sample we are observing contains a very large number of atoms n which have not decayed, the most likely number of atoms actually expected to decay during a small time interval Δt is $\Delta n = np\Delta t$. This assumption is experimentally verifiable, and it is also consistent with both Bernoulli's law of large numbers and the Poisson probability law. For small Δt , $\Delta n \ll n$. If we consider the number of atoms in the sample which have not decayed to be a function of time, for instance, $n = n(t)$, then the change Δn in n during time Δt is given by the formula

$$\Delta n = -\Delta n = -np\Delta t.$$

Going to the differential limit yields the differential equation,

$$dn/dt = -pn, \tag{7.9}$$

which governs radioactive decay. The solution to this equation is the formula

$$n(t) = n(0)e^{-pt}, \tag{7.10}$$

which is designated the Law of Exponential Decay. It is generally known that the statistical uncertainty in the number of counts N recorded during a particular nuclear measurement is $N^{1/2}$. This result naturally must follow from the Poisson probability law, and we shall see in a future report

precisely how this comes about. Meanwhile, we demonstrate what this uncertainty signifies through the following example:

Example 7.7

Suppose that $\lambda = 100$. Values of the Poisson mass function $p(\lambda; k)$ for $k = 80$ to 120 are listed in Table 7.3. Note that the mass function is approximately symmetric for $k \approx \lambda$, in accordance with a previous theorem. It is also seen that $p(\lambda; k)$ drops to something on the order of half of the maximum value for k such that $|\lambda - k| \approx \lambda^{1/2}$. Thus, the statistical uncertainty $\lambda^{1/2}$ is a measure of the breadth of the Poisson probability mass function. In general, uncertainty is always a measure of the breadth of probability density (or mass) functions. In Example 7.2, it was demonstrated that a binomial distribution becomes "sharper" for large n , thereby quantitatively demonstrating Bernoulli's law of large numbers. The same is true for Poisson functions, where the fractional uncertainty varies as $\lambda^{-1/2}$. This is the origin of the well-known rule that long radioactivity counts produce better "statistics," i.e., better precision.

Statistical uncertainty associated with the Poisson distribution, as discussed in Example 7.7, is a fundamental reason why there is a limit to the resolution which is obtainable for pulse-height spectra recorded from a nuclear radiation detector. This point is demonstrated in the following example:

Example 7.8

Suppose that a 1-MeV photon gives up all its energy within a germanium diode detector. The signal generated by the detector has an amplitude which is proportional to the collected ionization charge, and that, in turn, is generally proportional to the energy deposited in the diode by the photon. We will assume that one quantized unit of ionization charge is collected per 1 eV of deposited energy, taking into consideration recombination and other losses. Thus, it is most likely (in the Poisson sense) that 10^6 quantized units of charge will be collected when a 1-MeV photon is completely absorbed by the detector. However, other charge numbers k are possible, and their probabilities are derived using the Poisson law. Consequently, the pulse-height full-energy peak that is formed in analysis of the signals produced by detecting many 1-MeV photons could be expected to resemble a Poisson distribution, all other considerations aside. The spectrum width, expressed in energy units (keV), would be of the order of $2(10^3/10^6)(1000) = 2$ keV. In fact, this represents an upper limit to the obtainable resolution since other random process, e.g., electronic noise, will serve to broaden the peak even further.

Table 7.2: Poisson Probabilities for $\lambda = 10$ and $k = 0$ to 20

k	p(10;k)	k	p(10;k)	k	p(10;k)
0	0.0000453	7	0.0901	14	0.0521
1	0.000453	8	0.113	15	0.0347
2	0.00227	9	0.125	16	0.0217
3	0.00757	10	0.125	17	0.0128
4	0.0189	11	0.114	18	0.00709
5	0.0378	12	0.0948	19	0.00373
6	0.0631	13	0.0729	20	0.00187

Table 7.3: Poisson Probabilities for $\lambda = 100$ and $k = 80$ to 120. The breadth of the distribution is characterized by the parameter $\sqrt{\lambda} = 10$ which is designated the statistical error in the most probable value λ .

k	p(100;k)	k	p(100;k)	k	p(100;k)
80	0.520350E-02	95	0.360446E-01	110	0.234404E-01
81	0.642390E-02	96	0.375469E-01	111	0.211159E-01
82	0.783383E-02	97	0.387033E-01	112	0.188533E-01
83	0.943729E-02	98	0.394955E-01	113	0.166847E-01
84	0.112351E-01	99	0.398929E-01	114	0.146364E-01
85	0.132172E-01	100	0.398953E-01	115	0.127272E-01
86	0.153688E-01	101	0.394980E-01	116	0.109702E-01
87	0.176667E-01	102	0.387222E-01	117	0.937643E-02
88	0.200753E-01	103	0.375927E-01	118	0.794603E-02
89	0.225548E-01	104	0.361482E-01	119	0.667736E-02
90	0.250620E-01	105	0.344276E-01	120	0.556486E-02
91	0.275404E-01	106	0.324782E-01		
92	0.299350E-01	107	0.303508E-01		
93	0.321881E-01	108	0.281041E-01		
94	0.342411E-01	109	0.257851E-01		

All nuclear radiation detectors require a certain amount of time to process an event once such an event has been detected. Because events occur randomly, there is a chance that some events will be lost because they happen to occur during the intervals of detection paralysis which follow detection of preceding events. The Poisson probability law is ideally suited to analyzing these losses. This is illustrated in the following example:

Example 7.9

Suppose that the average number of detector events per unit time is R , based on experience accumulated over a large number of random trials. Also, suppose that after it has accepted an event, the detector is unable to record another event for a time T which is the same for all such events. The average rate of events occurring during an interval of time T is $\lambda = RT$. One can approach this problem in either one of two equivalent ways:

Live time approach:

Let P_L be the probability that no pulses will be lost during the time interval T . According to the Poisson probability law, this is given by $P_L = p(RT;0) = (RT)^0 e^{-RT}/0! = e^{-RT}$. If $RT \ll 1$, which is generally the case for reasonable detector setups, then $P_L \approx 1 - RT$.

Deadtime approach:

Let P_D be the probability that one or more events will be lost during the time interval T . Applying the Poisson probability law again, we have

$$\begin{aligned}
 P_D &= \sum_{k=1}^{\infty} p(RT;k) = e^{-RT} \sum_{k=1}^{\infty} (RT)^k/k! = e^{-RT} \left[\sum_{k=0}^{\infty} (RT)^k/k! - 1 \right] \\
 &= e^{-RT} (e^{RT} - 1) = 1 - e^{-RT}.
 \end{aligned}$$

If $RT \ll 1$, then $P_D \approx RT$.

Note that regardless of the particular value for RT , $P_L + P_D = 1$. The detector either loses one or more events or it loses none during time interval T . There are no other possibilities.

Coincidence-measurement techniques are quite important in experimental nuclear physics. However, it is well known that in addition to true coincidence events, one can also obtain random (false) coincidences which can be misleading if suitable corrections are not applied. Interference from random coincidences become problematic if the timing resolution is poor, or if the count rates are high. In the following example, Poisson statistics is employed to derive a well-known formula for estimating random coincidences.

Example 7.10

Consider two separate detectors D_1 and D_2 which observe radiation from two distinct sources S_1 and S_2 , respectively, as shown in Fig. 7.6. The two detection channels are thus independent, and no true coincidences are to be expected. However, random coincidences are possible. Suppose that N_1 and N_2 are the measured average singles count rates for detectors D_1 , and D_2 respectively. Let T be the resolving time of the coincidence apparatus, i.e., if events are recorded in D_1 and D_2 within an elapsed time of T or less, there will be a coincidence recorded. Furthermore, suppose that $N_1 T \ll 1$ and $N_2 T \ll 1$. If D_1 registers a count, then it is very improbable that another count will occur in this channel within the following time interval T ; however, a count may occur in D_2 . In fact, the probability that one or more counts occur in D_2 in the particular time interval T initiated by an event in D_1 is

$$\sum_{k=1}^{\infty} (N_2 T)^k e^{-N_2 T} / k! = 1 - e^{-N_2 T} .$$

So, an approximation to the random coincidence rate in this limit is

$$N_c \approx N_1 (1 - e^{-N_2 T}) \approx N_1 N_2 T .$$

So long as the count rates in these detectors satisfy the requirements $N_1 T \ll 1$ and $N_2 T \ll 1$, we are led to the same result by this method of derivation regardless of which detector is first considered. However, if one or both of the detectors has a substantial count rate, the problem becomes more complex and must be approached in a more sophisticated manner. This problem is not considered in this report.

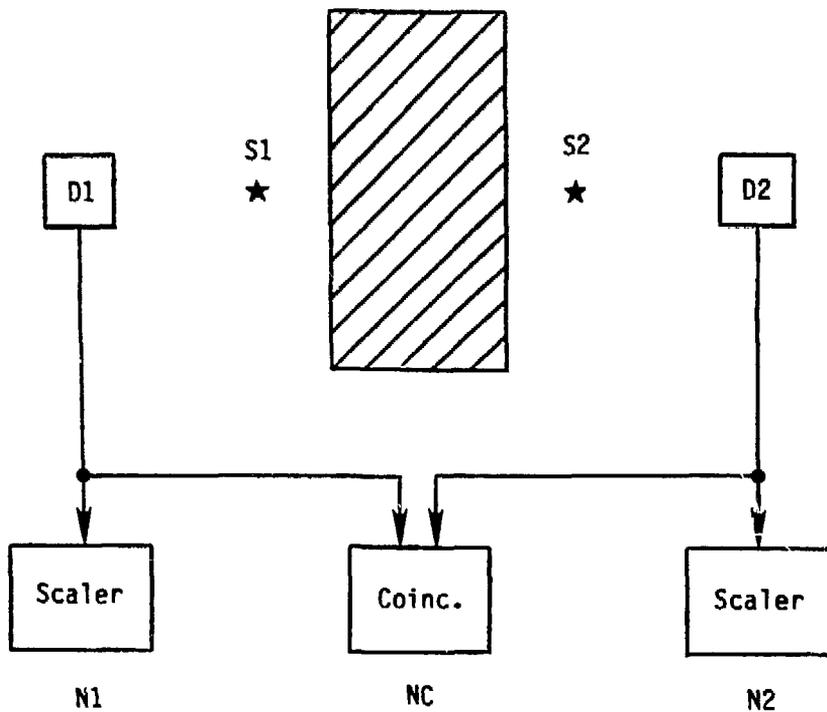


Figure 7.6: Schematic diagram of the experiment described in Example 7.10.

Examples 7.9 and 7.10 lead us to consideration of the so-called exponential or interval distribution. This distribution can be derived directly from the Poisson probability law, and it describes the distribution in the sizes of the time intervals between successive events in any random process in which the average rate is R events per unit time. Pursuing this notion further, we note that the probability that no events will be observed in the time interval from 0 to t is just

$$p(Rt;0) = (Rt)^0 e^{-Rt} / 0! = e^{-Rt}.$$

The probability that exactly one event will be observed in the sequential time interval from t to $t + dt$ is just

$$p(Rdt;1) = (Rdt)^1 e^{-Rdt} / 1! \approx Rdt,$$

for a differential time interval dt . The interval distribution describes the probability for a composite scenario, namely, no events for a time t followed by a single event in time dt . Owing to independence, this probability is the product of the two factors described above. In differential form, it is given by the equation

$$dP = R e^{-Rt} dt \quad . \quad (7.11)$$

It is easy to prove that the interval distribution is normalized, i.e., that

$$\int_{t=0}^{t=\infty} dP = R \int_0^{\infty} e^{-Rt} dt = 1 \quad .$$

We can thus consider

$$P(t) = R e^{-Rt}$$

as a continuous probability density function with a single random variable t ranging from 0 to ∞ , and a single fixed parameter R . We should again remind ourselves that only measurement, not theory, provides R .

Equation (7.11) is very useful for addressing problems involving anticipated count rates. The following example illustrates an application of the interval distribution.

Example 7.11

Consider an experiment with a mean rate of R events per unit time. We measure the actual elapsed time intervals between observed events until data have been recorded for a large number N of such events. How many of these intervals would we expect to have times $t < T$? From Eq. (7.11) we deduced that the probability of observing a time interval $t < T$, namely,

$$P(t < T) = R \int_0^T e^{-Rt} dt = 1 - e^{-RT}.$$

Consequently, we would anticipate $N(1 - e^{-RT})$ intervals shorter than T . Analogously, we would expect to observe Ne^{-RT} intervals exceeding time T . Clearly, the longer the time T is considered, the less probable it is that intervals of this duration will be observed.

Poisson probability satisfies an important property of additivity. We now state this in terms of two theorems. To prove these theorems requires mathematical techniques somewhat beyond the present treatment, so they will be stated without proof. However, the concept will be illustrated below with an example.

Theorem: [Addition Theorem]

Let k_1 and k_2 be two distinct random variables. k_1 is distributed according to the Poisson law $p(\lambda_1; k_1)$ while k_2 is distributed according to the Poisson law $p(\lambda_2; k_2)$, in accordance with Eq. (7.8). If $k = k_1 + k_2$ is treated as a random variable, then k is distributed according to the Poisson law $p(\lambda; k)$, where $\lambda = \lambda_1 + \lambda_2$.

This theorem has an inverse which is attributed to the Russian mathematician D. Raikov:

Theorem: [Raikov's Theorem]

If k_1 and k_2 are independent, and $k = k_1 + k_2$ follows a Poisson law, then each random variable, k_1 and k_2 , follows a Poisson law of its own.

Example 7.12

The addition theorem is very important in counting applications. Suppose two detectors D_1 and D_2 are independent and the counting processes obey Poisson statistics. Detector D_1 yields N_1 counts, while detector D_2 yields N_2 counts in a particular experiment. The statistical errors are $N_1^{1/2}$ and $N_2^{1/2}$, respectively. The sum $N = N_1 + N_2$ also obeys Poisson statistics, by the addition theorem, so its error is $N^{1/2}$. By Raikov's theorem, if the sum of two detector channels obeys Poisson statistics, then each individual channel must also. These theorems are commonly applied in the composition or decomposition of multi-channel spectra from a single detector, where the individual channel contents can be viewed as having been accumulated independently.

We close this section on Poisson probability with an example of Poisson statistics which offers us an interesting insight into the nature of modern scientific research in general.

Example 7.13

De Sola Price (Pri 63) conducted a study of many individual cases of scientific discovery. From this work he was able to reach the following interesting conclusion: The number of cases k of simultaneous scientific discovery, as evidenced by publications reporting them, seemed to conform proportionally to $(1/k!)$. We note that the Poisson probability formula

$$p(1;k) = (1)^k e^{-1}/k!$$

is also proportional to $(1/k!)$. Therefore, data of De Sola Price apparently support the contention that, while it is still most likely that any particular scientific discovery will be made just once, multiple discoveries do occur with Poisson frequencies owing to the apparent tendency for scientists to work independently of one another. In short, scientists tend to publish papers but rarely read the work of their peers! In view of the explosion of scientific knowledge in this century, and the acknowledged pressure placed upon scientists to be productive, this is not a surprising state of affairs.

7.3 Normal (Gaussian) Probability Law

The normal probability law is by far the best known and most widely used probability law of statistics. In this section we examine a few of its properties and offer some indication as to why it is hard to exaggerate its importance in realistic applications. The origins of this law stem from the early eighteenth century. Scientists began to observe an astonishing degree of regularity in the errors of measurement. The so-called "normal" curve (we have come to know of it informally as the bell-shaped curve) seemed very often to approximate the outcomes of multiple observations. The individuals credited with outlining the mathematical properties of this distribution and providing a theoretical basis for its application are: Abraham DeMoivre (1667-1745), Pierre Laplace (1749-1827), and Karl Gauss (1777-1855). In honor of the latter, the distribution is often referred to as a Gaussian distribution. Investigation of the properties of this distribution, and development of a better understanding of why it seems to be of such importance in Nature, persist as important research areas in the field of stochastic theory up to the present time.

Unlike the binomial and Poisson laws discussed in Sections 7.1 and 7.2, respectively, the normal law involves a continuous density function of a continuous random variable which we will designate as x . This variable can assume any real-number value from $-\infty$ to $+\infty$. Furthermore, there are two

constant parameters of the distribution, σ and λ , which are known as the standard deviation and expected value of the distribution, respectively. The density function for the normal probability law takes the form

$$p(\sigma, \lambda; x) = \exp[-(x - \lambda)^2 / 2\sigma^2] / (2\pi\sigma^2)^{1/2}, \quad (7.12)$$

for $-\infty < x < +\infty$. The parameter σ must be positive, since it is related to the width of the distribution. λ can be any finite real number. Referring to Pearson's differential equation (Eq. 6.18 in Chapter 6), it is rather easy to show that Eq. 7.12 is indeed a solution of this differential equation for the special case $a = \sigma^2$, $b = c = 0$, and $d = \lambda$. It can also be seen by referring to tables of definite integrals that this density function is normalized, i.e.,

$$\int_{-\infty}^{\infty} p(\sigma, \lambda; x) dx = 1$$

for arbitrary λ and σ ($\sigma > 0$).

A compilation of all the known properties of the normal distribution would fill many volumes. Here we only indicate a few of them, without proof:

Theorem: [Symmetry Theorem]

$$p(\sigma, \lambda; x = \lambda + z) = p(\sigma, \lambda; x = \lambda - z) \text{ for all real } z.$$

Theorem:

$$(a) \quad 0 < p(\sigma, \lambda; x) \leq p(\sigma, \lambda; \lambda) = (2\pi\sigma^2)^{-1/2} \text{ for all } x.$$

$$(b) \quad p(\sigma, \lambda; x = \lambda \pm \sigma) / p(\sigma, \lambda; \lambda) = e^{-1/2} \approx 0.6065,$$

$$(c) \quad p[\sigma, \lambda; x = \lambda \pm (2 \ln 2)^{1/2} \sigma] / p(\sigma, \lambda; \lambda) = 1/2,$$

$$(d) \quad p[\sigma, \lambda; x = \lambda \pm (2)^{1/2} \sigma] / p(\sigma, \lambda; \lambda) = 1/e \approx 0.3679.$$

Theorem:

The derivative of $p(\sigma, \lambda; x)$ with respect to x , exists everywhere and has the form

$$p'(\sigma, \lambda; x) = (\lambda - x)p(\sigma, \lambda; x) / \sigma^2.$$

Thus:

$$(a) \quad p'(\sigma, \lambda; \lambda) = 0,$$

$$(b) \quad p'(\sigma, \lambda; x) < 0 \text{ if } x > \lambda,$$

(c) $p'(\sigma, \lambda; x) > 0$ if $x < \lambda$,

(d) $p'(\sigma, \lambda; x = \lambda + z) = -p'(\sigma, \lambda; x = \lambda - z)$ for all real z ,

(e) $|p'(\sigma, \lambda; x)| \leq |p'(\sigma, \lambda; \lambda \pm \sigma)| = e^{-1/2}/(2\pi\sigma^4)^{1/2}$ for all x .

Theorem:

The tangents to the normal curve at $x = \lambda + \sigma$ and $x = \lambda - \sigma$ intercept the zero probability density axis at $x = \lambda + 2\sigma$ and $x = \lambda - 2\sigma$, respectively.

Theorem: [Addition Theorem]

Suppose a random variable x_1 is distributed according to $p(\sigma_1, \lambda_1; x_1)$, while x_2 is distributed according to $p(\sigma_2, \lambda_2; x_2)$. Next, suppose x is a random variable such that $x = x_1 + x_2$. Then x is distributed according to $p(\sigma, \lambda; x)$, with $\lambda = \lambda_1 + \lambda_2$ and $\sigma^2 = \sigma_1^2 + \sigma_2^2$.

Theorem: [Levy-Cramer Theorem]

Let x_1 and x_2 be independent random variables and let x be the normally distributed random variable that is the sum, $x = x_1 + x_2$. Then both x_1 and x_2 are normally distributed.

This theorem is essentially the inverse of the Addition Theorem, and it was first proven by H. Cramer.

It is evident from Eq. 7.12 that there is no unique normal distribution, but an infinitely large family of such distributions, characterized by the parameters σ and λ . This obvious fact makes for difficulty in tabulating the normal distribution. However, if we let $\sigma = 1$ and $\lambda = 0$, we obtain a special member of this family known as the standard normal distribution or standard Gaussian:

$$f(x) = p(1, 0; x) = \exp(-x^2/2)/(2\pi)^{1/2} \quad (7.13)$$

Furthermore, the integral

$$F(x) = \int_{-x}^{+x} f(z)dz \quad (0 \leq x \leq \infty) \quad (7.14)$$

is very useful since it represents an actual probability for observing the random variable z in the range $(-x, x)$, for $x \geq 0$, when this random variable is distributed according to the standard Gaussian. Tables of $f(x)$ and $F(x)$ are widely available, and an abbreviated version is provided in this report

as Table 7.4. The functions f and F , and tables thereof, are very useful in general because any normal distribution of the random variable x , which is characterized by λ and σ , can be transformed to the standard normal distribution in z by the linear random-variable transformation.

$$z = (x - \lambda)/\sigma. \quad (7.15)$$

The variable z in this form is known as a reduced variable or a standard variable for the distribution. The simplification afforded by this transformation is one of the many reasons why the normal distribution is so convenient for applications.

Statisticians often refer to one-sigma, two-sigma or three-sigma limits. Table 7.4 helps us understand what this means. From the values of $F(x)$ in this table, and Eq. 7.14, it is evident that the probability, for the standard Gaussian, that x will be observed to exceed 1 (one sigma) in repeated trials is $\sim 31.7\%$. From Eq. 7.15, it is seen that for arbitrary standard deviation σ and expected value λ , this is equivalent to $|x - \lambda| > \sigma$ (one sigma). Likewise, the probability, for the standard Gaussian, of $|x|$ exceeding 2 (two sigma) is seen to be $\sim 4.6\%$. Finally, the probability for the standard Gaussian that $|x|$ will exceed 3 (three sigma) is only $\sim 0.3\%$, which is nearly negligible for many practical purposes. It is the three-sigma rule which is generally used by statisticians as a test of confidence, and it is often applied to distributions other than the Gaussian. Then, σ is the standard deviation of the particular distribution in question. It is important to keep in mind that the "confidence" limit values indicated above (in percent) apply only to normal distributions. It should also be stressed that when a Gaussian distribution is used to approximate another type of probability distribution, no physical credence should be given to values of x beyond the range of the particular distribution in question. A similar caveat was established in Section 7.2, with regard to the Poisson distribution, which also has an infinite range defined for its random variable k .

We now examine in more detail several reasons why the Gaussian distribution holds a place of such eminence in statistics. Most of the arguments favoring its use fall into the following four categories: i) it is convenient to use, ii) it has been found (from basic or empirical considerations) to be the appropriate (or nearly appropriate) probability law for certain important physical processes, iii) it is a reasonably good approximation to several other probability distributions, and iv) the influence of the Central Limit Theorem. We have already presented evidence to the effect that Gaussians possess mathematical properties which make them convenient to use. This matter will be pursued further in later volumes of this report series. We should mention here that, by the use of random-variable transformations (see Section 3.3), it is sometimes possible to analyze certain phenomena in terms of Gaussian distributions even though they are not directly applicable. The following example illustrates this point:

Table 7.4 Numerical Values for the Standardized Normal Distribution^a

x	f(x) ^b	F(x) ^c	x	f(x) ^b	F(x) ^c
0	0.39894228	0	2.6	0.13583059(-1)	0.99067935
0.1	0.39695255	0.79695370(-1) ^d	2.7	0.10421011(-1)	0.99306748
0.2	0.39104269	0.15855852	2.8	0.79155151(-2)	0.99489094
0.3	0.38138782	0.23586098	2.9	0.59525848(-2)	0.99626938
0.4	0.36827015	0.31088031	3.0	0.44318910(-2)	0.99730177
0.5	0.35206535	0.38296014	3.1	0.32668533(-2)	0.99806556
0.6	0.33322462	0.45152710	3.2	0.23841153(-2)	0.99862641
0.7	0.31225393	0.51610393	3.3	0.17225901(-2)	0.99903377
0.8	0.28969154	0.57631818	3.4	0.12322355(-2)	0.99932672
0.9	0.26608522	0.63190636	3.5	0.87269508(-3)	0.99953529
1.0	0.24197068	0.68271369	3.6	0.61191124(-3)	0.99968230
1.1	0.21785217	0.72868966	3.7	0.42478781(-3)	0.99978490
1.2	0.19411861	0.76988008	3.8	0.29195199(-3)	0.99985580
1.3	0.17136866	0.80641618	3.9	0.19865914(-3)	0.99990429
1.4	0.14972757	0.83850168	4.0	0.13383285(-3)	0.99993713
1.5	0.12951772	0.86639860	4.1	0.89263514(-4)	0.99995916
1.6	0.11092098	0.89041259	4.2	0.58944366(-4)	0.99997378
1.7	0.94049237(-1)	0.91087859	4.3	0.38536094(-4)	0.99998339
1.8	0.78950324(-1)	0.92814740	4.4	0.24943084(-4)	0.99998964
1.9	0.65615982(-1)	0.94257362	4.5	0.15984155(-4)	0.99999367
2.0	0.53991129(-1)	0.95450535	4.6	0.10141128(-4)	0.99999624
2.1	0.43983751(-1)	0.96427581	4.7	0.63700076(-5)	0.99999787
2.2	0.35474737(-1)	0.97219693	4.8	0.39614182(-5)	0.99999889
2.3	0.28327169(-1)	0.97855492	4.9	0.24390376(-5)	0.99999950
2.4	0.22394648(-1)	0.98360750	5.0	0.14867686(-5)	0.99999989
2.5	0.17528404(-1)	0.98758277	5.1	0.89727451(-6)	1.00000000

^aCalculations performed using a microcomputer. Eight-significant figure reproduction from the output does not guarantee eight-significant-figure accuracy. Comparison with other tables indicates that f(x) is accurate to at least five significant figures, while F(x) is accurate to at least four significant figures, for $0 \leq x \leq 3$ (through three standard deviations).

^b $f(x) = \exp(-x^2/2)/(2\pi)^{1/2}$.

^c $F(x) = \int_{-x}^{+x} f(z)dz$.

^d0.79695370(-1) signifies $0.79695370 \times 10^{-1}$.

Example 7.14

Suppose that we consider a large number of similar shielding blocks with uniform density. As an exercise, we weigh each block and ask how the masses are distributed. It is known that the linear dimensions of objects follow distributions which appear to be well approximated by Gaussians, if the measurement sample sizes (number of blocks considered) are large. The mass M of a block is given in terms of uniform density ρ , length ℓ , width w , and height h by the formula

$$M = \rho \ell w h.$$

We shall assume for simplicity that the precision to which the blocks are cut is uniform for each of the three linear dimensions, and that there exist constant scale factors s_ℓ , s_w , and s_h for this particular problem, such that $\ell = s_\ell x$, $w = s_w x$ and $h = s_h x$. Under this assumption we express M in terms of a single random variable, namely, a generic linear dimension x , by means of the expression

$$M = s_\ell s_w s_h x^3.$$

Clearly, $M^{1/3}$ is proportional to x , so we expect $M^{1/3}$ to be distributed as a Gaussian from the preceding discussion. We may be satisfied to limit our statistical consideration to $M^{1/3}$. However, if ℓ , w , and h vary randomly, and independently of each other (which is more realistic), then it is more of a problem to examine the probability distribution for the mass M . We shall see in the next report of this series that it is generally not necessary to know the probability distribution in detail. Knowledge of a few of its moments will generally suffice. For this problem, it turns out that

$$\sigma_M^2 / \langle M \rangle^2 = \sigma_\ell^2 / \langle \ell \rangle^2 + \sigma_w^2 / \langle w \rangle^2 + \sigma_h^2 / \langle h \rangle^2,$$

where $\langle \dots \rangle$ denotes mean values and σ denotes standard deviations.

The normal probability law is not noteworthy as a truly fundamental distribution. In fact, few phenomena can be considered to behave statistically in naturally Gaussian fashion. Its importance stems mainly from the empirical observation that it often is a very good approximation to other more fundamental distributions of Nature, in the limit of large samples. Nevertheless, some physical phenomena involve such large samplings that Gaussian distributions are postulated as fundamental for all intents and purposes. The next two examples illustrate this point:

Example 7.15

The Maxwell-Boltzmann Law of Molecular Velocities evolves from the assumption of Gaussian probability distributions. Consider a uniform classical gas of molecules with mass M and temperature T (degrees absolute). Since there is no preferred direction in space, we can arbitrarily define a Cartesian coordinate system in real space and a corresponding one in velocity space. A point in velocity space is represented by (v_x, v_y, v_z) . Here, v_x represents the component of speed v for a particular molecule, as projected onto the x -axis, and we postulate that v_x is distributed according to the Gaussian

$$p_x(v_x) \sim \exp(-\beta^2 v_x^2)$$

with $-\infty < v_x < +\infty$. The same is true for v_y and v_z . These distributions are symmetric about zero because we postulate no preferred direction in space, i.e., zero flow for the gas. For convenience we omit normalization constants for these distributions. Since there is no net flow of the gas, v_x , v_y and v_z can be treated as independent random variables. Thus, the probability of finding a molecule within a small element $dv_x dv_y dv_z$ in velocity space (or perhaps we might consider this as essentially an element in phase space) is the product

$$p_x(v_x)p_y(v_y)p_z(v_z)dv_x dv_y dv_z \sim \exp[-\beta^2(v_x^2 + v_y^2 + v_z^2)]dv_x dv_y dv_z.$$

However, the speed v is related to its components v_x , v_y and v_z by

$$v^2 = v_x^2 + v_y^2 + v_z^2,$$

so the increment of probability associated with this small phase space element is $\sim \exp(-\beta^2 v^2)dv_x dv_y dv_z$. The Maxwell-Boltzmann distribution is concerned with molecular speed v , so we are interested in the probability of finding molecules having speeds between v and $v + dv$, namely, $p(v)dv$. To determine this probability we must sum over all cartesian phase space elements $dv_x dv_y dv_z$ for which the speed is between v and $v + dv$; namely,

$p(v)dv = \int \exp(-\beta^2 v^2)dv_x dv_y dv_z$. Because of the symmetric relationship between speed v and its components v_x , v_y and v_z , the phase space region is a thin spherical shell of radius v and thickness dv centered on the origin $(v_x, v_y, v_z) = (0,0,0)$. The shell volume is proportional to the volume element $v^2 dv$.

Therefore, the Maxwell-Boltzmann Law is expressed as the normalized probability density function

$$p(v) = 4\beta^2 v^2 \exp(-\beta^2 v^2) / \pi^{1/2}, \text{ for } 0 \leq v < \infty.$$

The constant β is related to the fundamental molecular parameters according to $\beta^2 = M/2kT$, where k is Boltzmann's Constant.

In Section 7.1 we discussed a simple model of error, based on binomial probability, which is attributed to Laplace. In the next example we examine another model of error, based on the normal distribution, which was suggested by John Herschel, and is thus known as Herschel's Model of Error.

Example 7.16

Herschel envisioned release of an essentially infinite number of identical grains of sand from a point in space onto a plane. That the continuous distribution of their individual landing positions is Gaussian can be derived from just two very plausible postulates. Of course, it is obviously assumed that there are random disturbances present which insure that the grains do not all fall directly below their point of release. The first postulate states that there is no preferred azimuthal bias. Thus, in cylindrical coordinates, the density function is

$$p(r, \theta) = h(r),$$

where r is the radial distance from the projection of the drop point onto the plane (i.e., the density function has only one random variable). We can define a Cartesian coordinate system with origin coinciding with the cylindrical coordinate system. The orientation is arbitrary, but we assume that $x = r \cos \theta$, $y = r \sin \theta$. Herschel's second postulate states that the distribution function, expressed in Cartesian coordinates, is independent with respect to the two coordinates, namely,

$$p(x, y) = f(x)g(y).$$

Let dA be an increment of area on the plane. Then,

$$h(r)dA = f(x)f(y)dA$$

so

$$h(r) = f(x)g(y)$$

and

$$\ell n h(r) = \ell n f(x) + \ell n g(y).$$

Then,

$$\ln h(r) = \ln f(r\cos\theta) + \ln g(r\sin\theta).$$

Take the derivative of both sides of this expression with respect to θ , then

$$d[\ln h(r)]/d\theta = [f(x)]^{-1}d[f(r\cos\theta)]/d\theta + [g(y)]^{-1}d[g(r\sin\theta)]/d\theta.$$

Therefore,

$$0 = -r\sin\theta [f(x)]^{-1}d[f(x)]/dx + r\cos\theta [g(y)]^{-1}d[g(y)]/dy.$$

by the chain rule of differentiation. So, in abbreviated terms,

$$f'(x)/[xf(x)] = g'(y)/[yg(y)].$$

Now, there should be no particular distinction between f and g or x and y , since our choice of a Cartesian coordinate system was arbitrary. Therefore, $f'(x)/[xf(x)]$ is independent of x for all x and must be a constant, which we choose to call $-C$. Since

$$f'(x)[xf(x)] = -C,$$

it must be the case that

$$f(x) \sim \exp(-Cx^2/2).$$

neglecting a normalization factor. It would be impossible to normalize f for $-\infty < x < \infty$ without requiring $C > 0$. Finally,

$$h(r) \sim e^{-Cr^2/2} \quad (r \geq 0).$$

Science museums often set up displays which demonstrate formation of a "bell-shaped" normal curve in the manner suggested by Herschel.

The normal probability law is often used as a mathematically convenient continuous approximation to the discrete binomial distribution for large n . As discussed in Section 7.2, the binomial distribution is difficult to calculate for large n . So long as $p \ll 1$, the Poisson distribution is useful. Otherwise, the normal law must be used. If n is large and $p \ll 1$, yet $\lambda = np \gg 1$, the normal law also approximates the Poisson law fairly well for $x = k \approx \lambda$. However, the Poisson distribution is decidedly asymmetric, while Gaussians are intrinsically symmetric. These relationships are clarified in Table 7.5. The reader can refer to this table for guidance on how and when to use the binomial, Poisson, or Gaussian distributions in practical applications.

Table 7.5 Use of Poisson and Normal Distributions to Approximate the Binomial Distribution.

Binomial Distribution:

$$p(n,p;k) = C_{nk} p^k (1-p)^{n-k} \quad [k = (0 \text{ to } n)].$$

Valid for all $n \geq 0$, and all $p > 0$, as is discussed in Sec. 7.1. For large n , this distribution is awkward to use and should be approximated by either the Poisson or normal distributions, as discussed below.

Poisson Distribution:

$$p(\lambda;k) = \lambda^k e^{-\lambda} / k! \quad [\lambda \equiv np > 0, k = 0 \text{ to } \infty].$$

Used to approximate the binomial distribution when $n \gg 1$ and $0 < p \ll 1$, as discussed in Sec. 7.2. Then, the binomial is very asymmetric. Often λ is not large in situations where this approximation is useful. This is a discrete, infinite distribution. Comparison with the binomial is meaningless for $k > n$. λ need not be an integer.

Normal Distributions:

$$p(\sigma,\lambda;x) = \exp[-(x - \lambda)^2 / 2\sigma^2] / (2\pi\sigma^2)^{1/2}$$

$$[-\infty < x < +\infty, \lambda = np, \sigma^2 = np(1-p)].$$

Used to approximate the binomial distribution when $n \gg 1$ and p is unrestricted. It is an especially good approximation when the binomial is nearly symmetric ($p \sim 0.5$). This is a continuous, infinite distribution. Direct comparison with the binomial is made only for $x = k$ ($k = 0$ to n). Comparison is meaningless for $x < 0$ or $x > n$. When $p \ll 1$ but n is so large that $\sigma^2 \approx \lambda$ and $\lambda \gg 1$, then the normal distribution (with $\sigma = \lambda$) is a reasonably good approximation to the corresponding Poisson distribution as well as to the binomial, particularly for x in the vicinity of λ .

It is not a straightforward matter to prove that a binomial distribution can be approximated by a Gaussian for large n , so we avoid the details and simply accept this result as fact. The essence of this result is embodied in the following famous theorem (De Moivre-Laplace) which is stated here without proof:

Theorem: [De Moivre-Laplace Theorem]

Let a and b be two non-negative integers with $a < b$, and let n and p be parameters of the binomial distribution $p(n,p;k)$. Then for large n ,

$$\sum_{k=a}^b p(n,p;k) \approx \int_{\alpha}^{\beta} p(1,0;x)dx,$$

where $p(1,0;x)$ is the standard normal distribution, while

$$\alpha = (a - np - 1/2)/[np(1 - p)]^{1/2}$$

and

$$\beta = (b - np + 1/2)/[np(1 - p)]^{1/2}.$$

Very crudely speaking, what this theorem states is that the sum of the areas of contiguous histogram segments, representing discrete binomial probabilities, approximately equals the area under the corresponding continuous Gaussian curve spanning the same region. While it has been stressed that n should be large for the binomial to be represented well by the corresponding Gaussian approximation from Table 7.5, it is remarkable how good this approximation becomes, even for relatively small n , when $p \approx 1/2$. This is demonstrated in the following example.

Example 7.17

Refer to Table 7.6 for values of the binomial distribution $p(n,p;k)$ for $n = 10$ and $p = 1/2$, and of the corresponding normal distribution $p(\sigma,\lambda;x)$ for $\sigma = (2.5)^{1/2}$ and $\lambda = 5$. The agreement is quite good. To demonstrate the validity of the De Moivre-Laplace theorem, we first consider

$$\sum_{k=2}^4 p(10,0.5;k) \approx 0.3662.$$

Then consider

$$\int_{1.5}^{4.5} dx \{ \exp[-(x - 5)^2/5] (5\pi)^{1/2} \} \approx 0.3622.$$

The difference is only $\sim 1\%$. This example is illustrated graphically in Fig. 7.7. The shaded area represents the integral, while the associated histogram rectangles represent components of the discrete sum.

Table 7.6 A Comparison of Binomial and Normal Distributions^a

k	Binomial ^b p(n,p;k)	Normal ^c p(σ,λ;x)	Difference (%)
0	9.77(-4)	1.70(-3)	+74.0
1	9.77(-3)	1.03(-2)	+ 5.4
2	4.39(-2)	4.17(-2)	- 5.0
3	0.117	0.113	- 3.4
4	0.205	0.207	+ 1.0
5	0.246	0.252	+ 2.4
6	0.205	0.207	+ 1.0
7	0.117	0.113	- 3.4
8	4.39(-2)	4.17(-2)	- 5.0
9	9.77(-3)	1.03(-2)	+ 5.4
10	9.77(-4)	1.70(-3)	+74.0

^aCalculations performed using n = 10, p = 0.5, λ = 5.0

$$\sigma^2 = np(1 - p) = 2.5.$$

$${}^b p(n,p;k) = C_{nk} p^k (1 - p)^{n-k}.$$

$${}^c p(\sigma,\lambda;x) = \exp[-(x - \lambda)^2 / 2\sigma^2] / (2\pi\sigma^2)^{1/2}, \text{ with } x = k.$$

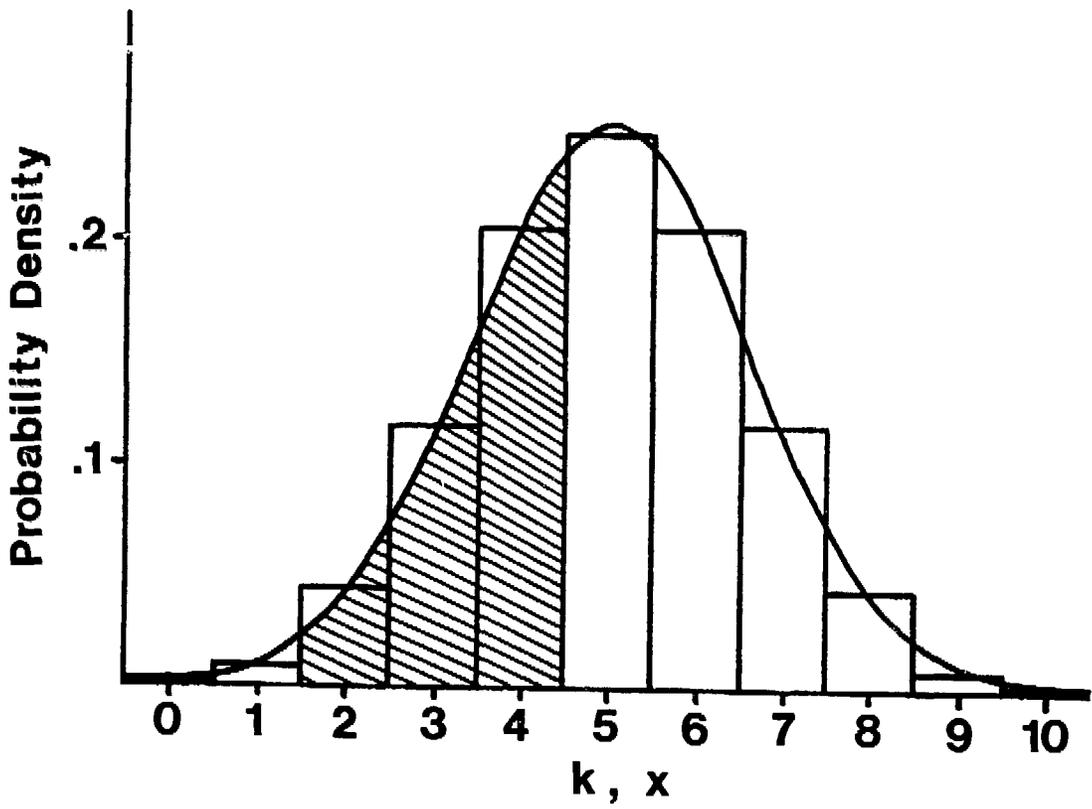


Figure 7.7: Graphical demonstration of the DeMoivre-Laplace Theorem for the special case discussed in Example 7.17.

The following example presents a special case in which the corresponding binomial, Poisson, and Gaussian distributions yield very nearly equal values over a narrow range.

Example 7.18

Let $n = 100000$ and $p = 0.01$. Therefore, $\lambda = np = 1000$ and $\sigma^2 = np(1 - p) = 990 \approx 1000$. Table 7.7 contains values for equivalent binomial, Poisson, and Gaussian distributions. The differences between these three distributions are smaller than 1% over the indicated limited range.

We alluded previously to the fact that normal distributions derive much of their importance from the fact that they seem to describe (at least approximately) a great many random phenomena observed in Nature and in every day life, even though there is no a priori reason to expect that this should be the case. One interesting clue as to why this actually happens may be found by considering an extremely important theorem of statistics known as the Central Limit Theorem. This theorem is stated in a variety of ways in textbooks, and it has been found to apply under even quite weak assumptions. Below we state a version of this theorem which, while not the strongest or most general statement of the concept one could make, is perhaps the easiest to understand and demonstrate:

Theorem: [Central Limit Theorem]

Suppose that a particular random property can be described by the random variable x . Suppose that we sample this random variable from a very large population in which x is distributed about an expected value λ with standard deviation σ . Now consider an experiment which is repeated m times. For each of the m repetitions, we sample x randomly from the population n times, independently, and then consider the random variable z , given by

$$z = \sum_{i=1}^n x_i.$$

This procedure yields a sequence of values z_k ($k = 1, m$). As m and n become large, it turns out that the random variable z becomes very nearly distributed as a Gaussian with expected value

$$\lambda' = n\lambda$$

and standard deviation

$$\sigma' = \sigma n^{1/2}.$$

Table 7.7 A Comparison of Equivalent Binomial, Poisson and Normal Distributions^a

k	Binomial ^b p(n,p;k)	Poisson ^c p(λ;k)	Normal ^d p(σ,λ;x)
990	0.01211	0.01206	0.01200
991	0.01222	0.01217	0.01211
992	0.01232	0.01227	0.01222
993	0.01241	0.01235	0.01231
994	0.01249	0.01243	0.01239
995	0.01255	0.01249	0.01246
996	0.01260	0.01254	0.01252
997	0.01264	0.01258	0.01256
998	0.01267	0.01260	0.01259
999	0.01268	0.01262	0.01261
1000	0.01268	0.01262	0.01262
1001	0.01267	0.01260	0.01261
1002	0.01264	0.01258	0.01259
1003	0.01260	0.01254	0.01256
1004	0.01255	0.01249	0.01252
1005	0.01249	0.01243	0.01246
1006	0.01241	0.01235	0.01239
1007	0.01233	0.01227	0.01231
1008	0.01223	0.01217	0.01222
1009	0.01212	0.01206	0.01211
1010	0.01200	0.01194	0.01200

^a Density functions are computed with the following parameters:
 $n = 1000.0$, $p = 0.01$, $\lambda = np = 1000.0$.

^b $p(n,p;k) = C_{nk} p^k (1-p)^{n-k}$

^c $p(\lambda;k) = \lambda^k e^{-\lambda} / k!$

^d $p(\sigma,\lambda;x) = \exp[-(x-\lambda)^2 / 2\sigma^2] / (2\pi\sigma^2)^{1/2}$, with $\sigma^2 = \lambda \approx np(1-p)$
 for $p \ll 1$ and $x = k$.

The larger n and m become, the better the representation by a Gaussian distribution. In practice, n need not be too large, but m needs to be rather sizable to obtain a reasonable Gaussian-like distribution. Also, we note that for fixed σ and λ

$$(\sigma'/\lambda') \sim n^{-1/2},$$

so this Gaussian becomes more sharply peaked as n increases.

Proof of the theorem is beyond the level of the present treatment, so we merely accept it. This is, indeed, a remarkable theorem with profound consequences. What it says, in essence, is that sums of random variables tend to be normally distributed regardless of the nature of the distribution of the individual components. In the preceding statement of the Central Limit Theorem it was assumed that each component of the sum was identically distributed. However, more general statements of this theorem relax this condition and allow the component random variables to be distributed arbitrarily. The only requirement is that the individual probability distributions be well-behaved. In particular they must be normalized and possess well-defined mean values and variances.

The Central Limit Theorem can be understood and appreciated through the following example:

Example 7.19

Envision a very large collection of cylindrical metal disks (foils). All the foils are made of the same material and have the same diameter. However, they differ in thickness. We assume that each foil belongs to one of three distinct categories based on thickness. One third of the foils are 0.01 cm thick, another third are 0.02 cm thick, while the rest are 0.03 cm thick. We suppose, for present purposes, that all foils belonging to a particular category are so identified.

Now consider the hypothetical exercise in which we make samples to irradiate in some experiment. We fabricate each sample by stacking and gluing together ten foils ($n = 10$) selected at random from an essentially inexhaustible supply. Each of the ten foils is selected independently and the samples themselves are independently fabricated. We fabricate a total of m of these samples and ask how the resulting sample thicknesses are distributed.

For simplicity, we suppose that the disk thickness x can assume only the values $x = 1, 2,$ and 3 . The probability distribution $p(x)$ is completely characterized by the values $p(1) = p(2) = p(3) = 1/3$. According to the definitions given in Ex. 4.5 of Sec. 4.3, the expected value of foil thickness, which we designate as $\langle x \rangle$, is given by

$$\langle x \rangle = \sum xp(x) = (1) (1/3) + (2) (1/3) + 3 (1/3) = 2.$$

This is the average value since, $p(x)$ is a uniform distribution. The variance in x is given by

$$\begin{aligned}\sigma^2 &= \sum (x - \langle x \rangle)^2 p(x) \\ &= (1 - 2)^2 (1/3) + (2 - 2)^2 (1/3) + (3 - 2)^2 (1/3) = 2/3.\end{aligned}$$

Thus, the standard deviation is $\sigma = (2/3)^{1/2} \approx 0.8165$. The thicknesses of samples formed by stacking ten randomly-selected foils are represented by the random variable

$$z = \sum_{i=1}^{10} x_i,$$

where each x_i is distributed as indicated above.

In this example we simulate the experiment described above on a microcomputer using a random number generator. This represents a particular application of the powerful Monte-Carlo Method which will be discussed in later volumes of this series. We simulate the random selection of individual foils from the population by generating a real random number R in range $0 < R < 1$ with a computer. If $0 < R \leq 1/3$, we assume $x = 1$. If $1/3 < R \leq 2/3$, we assume $x = 2$. Finally, if $2/3 < R \leq 1$, we assume $x = 3$. It is clear that z can assume any integer value in the range $10 \leq z \leq 30$, but no others. The simulation process is as follows: Using the computer we "fabricate" m samples. For each sample, we record the resulting thickness z , and by the Monte-Carlo procedure generate the array of numbers $N(z)$ for $z = 10, 11, \dots, 30$. $N(z)$ is the number of samples, out of a total of m , which are found to have thickness z . Since

$$\sum_{z=10}^{30} N(z) = m,$$

the normalized distribution we desire is just $p(m; z)$, where $p(m; z) = N(z)/m$ ($z = 10, 11, \dots, 30$). The results of three distinct simulation exercises of this nature, for $m = 10, 100$ and 1000 , respectively, are plotted in Fig. 7.8. The particular distribution for $m = 1000$ is compared with the Gaussian distribution $p(\sigma', \lambda'; z)$, with $\lambda' = 20$ and $\sigma' = (10)^{1/2} \sigma = (10)^{1/2} (0.8165) = 2.582$. It is clearly demonstrated that $p(m; z)$ becomes very well approximated by $p(\sigma, \lambda'; z)$ as m gets rather large. Remember that the distributions $p(m; z)$ are discrete and finite, while $p(\sigma, \lambda'; z)$ is continuous and infinite. It is understood, then, that comparison with a continuous Gaussian is meaningless where $p(m; z)$ is undefined.

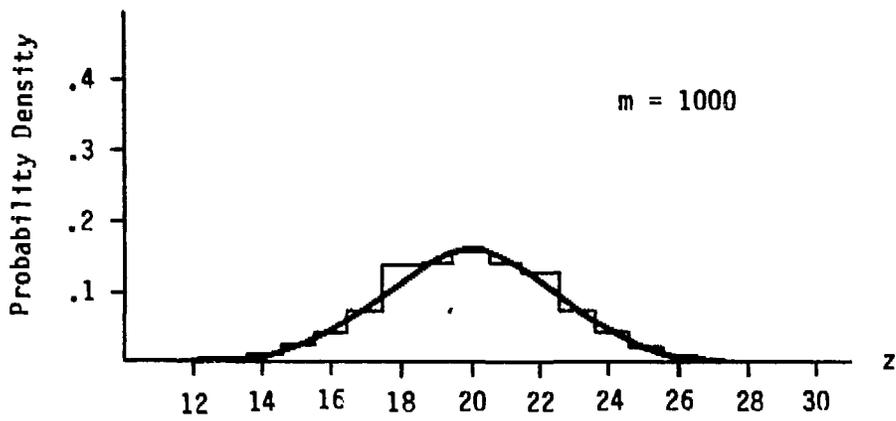
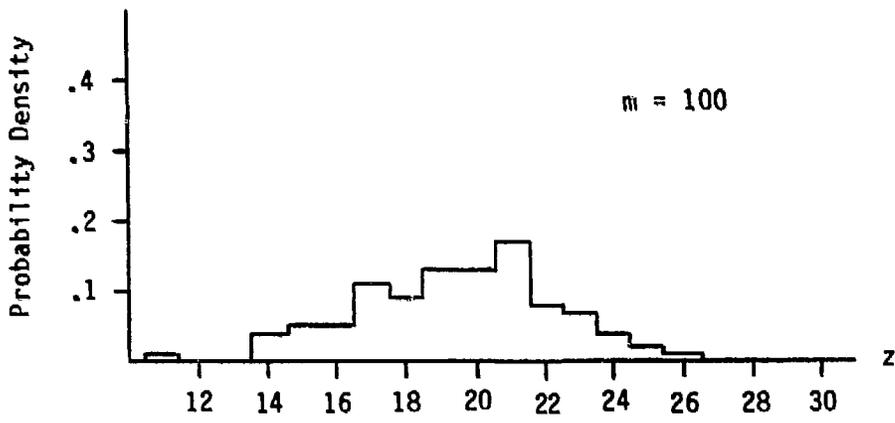
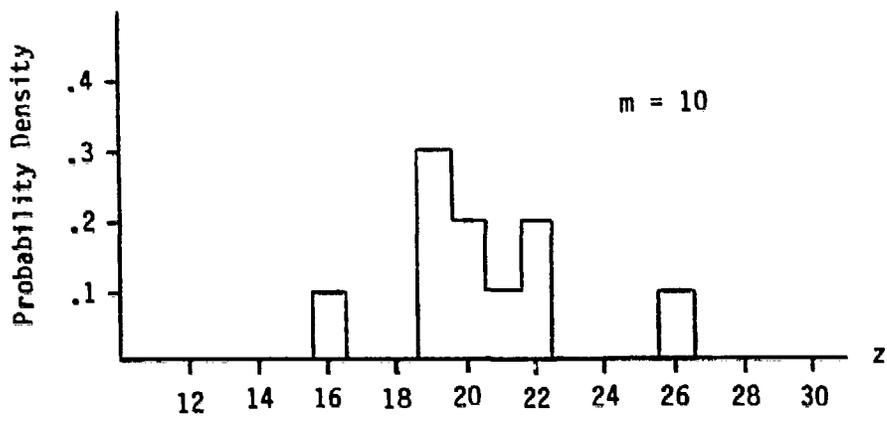


Figure 7.8: Graphical demonstration of the Central Limit Theorem for the special case discussed in Example 7.19.

The outcome of this computer simulation excited me greatly when I first did the calculations, and it has left me with a deep and lasting respect for the beauty and power of statistical analysis. I hope that the reader will be similarly impressed. It seems to me that the fundamental meaning of nuclear data uncertainty is clearly demonstrated in a very convincing fashion via this simple exercise.

In view of the preceding example, it is not hard to visualize why random variables that describe the linear dimensions of like macroscopic objects tend to be Gaussian-distributed for large populations. Macroscopic physical objects can be considered to be built up from many smaller components, each of which is distributed in its own way. When we consider the composite object, the Central Limit Theorem tells us that we need not be concerned with how these various components are distributed (so long as the distributions are physically reasonable), because this composite will be well described by a Gaussian. This very powerful result makes the job of statistical analysis much simpler. Once the mathematical form of the underlying distribution is known, the only remaining task is that of estimating its parameters. We shall see in a later report of this series that this is achieved using various sampling procedures and random variable statistics called estimators (as defined in Sec. 3.3).

We saw in Sec. 7.2 that the Poisson distribution is strictly univariate. However, in Sec. 7.1 it was indicated that the univariate binomial distribution has a multivariate equivalent known as the multinomial distribution. Gaussian distributions can also be multivariate. We now examine a few features of these more complicated multivariate normal distributions.

Using compact matrix notation, the general form for an n-dimensional Gaussian is

$$p(\bar{V}, \bar{\lambda}; \bar{x}) = \exp[-(1/2)(\bar{x} - \bar{\lambda})^+ \bar{V}^{-1}(\bar{x} - \bar{\lambda})] / \{(2\pi)^{n/2} [\det(\bar{V})]^{1/2}\}. \quad (7.16)$$

\bar{x} represents the random variable array (x_1, \dots, x_n) , a point in n-dimensional random variable space. $\bar{\lambda}$ represents the expected value of \bar{x} with respect to this distribution, and it is the array $(\lambda_1, \dots, \lambda_n)$. \bar{V} is an $n \times n$ symmetric, non-singular, positive-definite matrix called the variance-covariance matrix (or simply covariance matrix for short) of this distribution. The diagonal elements V_{ii} ($i = 1, n$) are variances, while the off-diagonal elements V_{ij} ($i \neq j = 1, n$) are called covariances. To be consistent with the univariate case, we express the variances V_{ii} in terms of standard deviation σ_i by $V_{ii} = \sigma_i^2$. There exists a related matrix \bar{C} called the correlation matrix. Its elements are given by the formula

$$C_{ij} = V_{ij}/(V_{ii}V_{jj})^{1/2} \quad (i, j = 1, n) \quad (7.17)$$

The quantity $\det(\bar{V})$ is the determinant of the matrix \bar{V} . If $\det(\bar{V}) \neq 0$, then the inverse \bar{V}^{-1} exists and \bar{V} is non-singular. The superscript "+" designates transposition. The matrix product

$$Q_n = (\bar{x} - \bar{\lambda})^+ \bar{V}^{-1} (\bar{x} - \bar{\lambda}) \quad (7.18)$$

is known as a generalized quadratic form of order n. We shall discuss these matters in far greater detail in future reports of this series.

Multivariate Gaussians share many properties in common with univariate Gaussians. We shall avoid consideration of these, since many of them are obvious. One feature of multivariate Gaussians which has no counterpart in the univariate case is embodied in the following theorem which is offered without proof:

Theorem:

The random variables x_i and x_j ($i \neq j = 1, n$) are mutually independent if, and only if $V_{ij} = 0$ ($i \neq j = 1, n$). Then

$$p(\bar{V}, \bar{\lambda}; \bar{x}) = \prod_{i=1}^n p[(V_{ii})^{1/2}, \lambda_i; x_i],$$

with $p[(V_{ii})^{1/2}, \lambda_i; x_i]$ given by Eq. (7.12).

Rather than dwelling further on general properties of multivariate Gaussians, we will examine the special case of bivariate Gaussians. Some of their features can be readily demonstrated graphically.

From Eq. (7.17), it is evident that the correlation matrix \bar{C} is characterized by a single parameter $\rho = C_{12} = C_{21}$ for the bivariate case. ρ in this special case is called the correlation coefficient. Thus, we can write the bivariate equivalent of Eq. (7.16) as

$$p(\sigma_1, \sigma_2, \lambda_1, \lambda_2; x_1, x_2) = \exp(-Q_2/2) / [2\pi\sigma_1\sigma_2(1 - \rho^2)^{1/2}] \quad (7.19)$$

with

$$Q_2 = [(x_1 - \lambda_1)^2/\sigma_1^2 - 2\rho(x_1 - \lambda_1)(x_2 - \lambda_2)/\sigma_1\sigma_2 + (x_2 - \lambda_2)^2/\sigma_2^2]/(1 - \rho^2). \quad (7.20)$$

The parameter ρ can span the range $-1 < \rho < +1$; however, we run into difficulty if $\rho = \pm 1$, according to Eqs. (7.19) and (7.20). We will see what this means shortly, but, for the present, assume that $\rho \neq \pm 1$. $p_2 > 0$ for all (x_1, x_2) , so

$$x_3 = p_2(\sigma_1, \sigma_2, \lambda_1, \lambda_2; x_1, x_2)$$

defines a surface which blankets the entire plane defined by $x_3 = 0$. The random variables x_1 and x_2 are defined with respect to a Cartesian coordinate system in this plane, and x_3 represents probability density. Now consider a plane $x_3 = c$ (constant), with

$$0 < c < 1/[2\pi\sigma_1\sigma_2(1 - \rho^2)^{1/2}].$$

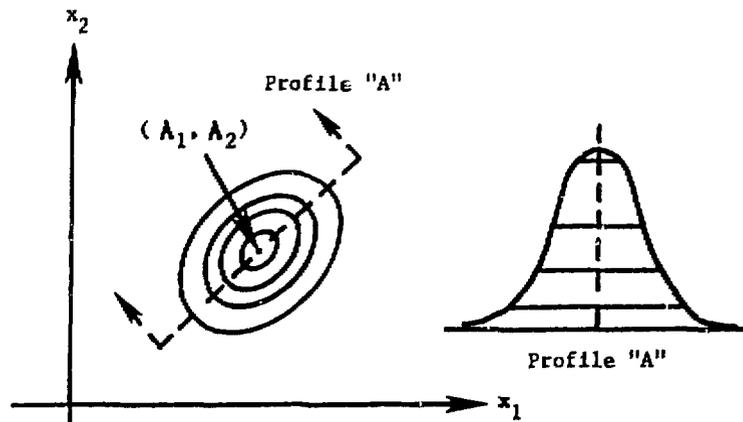
The set of all points (x_1, x_2) for which $x_3 = c$ is of interest. Since x_3 is constant, Q_2 must also be constant. In fact, if $x_3 = c$, then

$$Q_2 = c' = -2\ell n[2\pi c \sigma_1 \sigma_2 (1 - \rho^2)^{1/2}].$$

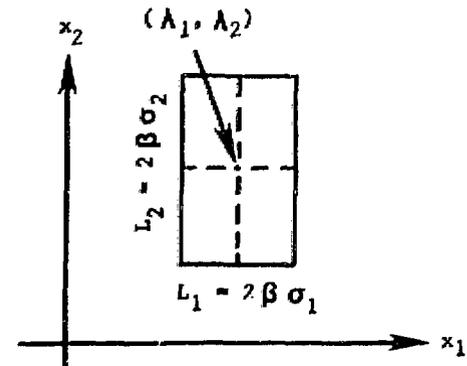
However, Eq. (7.20) (for Q_2 constant) is the general equation for an ellipse in terms of x_1 and x_2 . Thus, planes of constant probability intercept the bivariate Gaussian probability surface in a series of ellipses which are centered about $x_1 = \lambda_1$ and $x_2 = \lambda_2$, as shown in Fig. 7.9(a). We note in passing, without proof, that any plane perpendicular to the (x_1, x_2) plane, and passing through $x_1 = \lambda_1$ and $x_2 = \lambda_2$, slices the bivariate Gaussian surface into a univariate Gaussian profile, as shown in Fig. 7.9(a). Thus a bivariate probability distribution is "mound-like" with simple vertical univariate Gaussian profiles and elliptical horizontal profiles.

For constant probability $p = x_3 = c$, the corresponding (x_1, x_2) ellipse can be fitted into a rectangle centered about $x_1 = \lambda_1$ and $x_2 = \lambda_2$ with a side of length $\ell_1 = 2\beta\sigma_1$ in the x_1 direction and $\ell_2 = 2\beta\sigma_2$ in the x_2 direction (β is a nonnegative constant which depends upon the probability $p = x_3 = c$), as shown in Fig. 7.9(b).

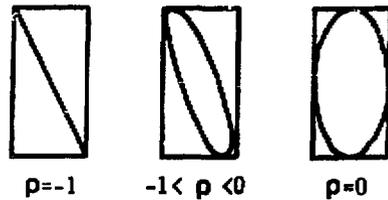
Suppose that we keep σ_1 and σ_2 fixed in magnitude, but consider the family of bivariate normal distributions generated by merely varying the correlation coefficient. The family of ellipses which is generated by varying ρ over the range $-1 < \rho < +1$, for the expression



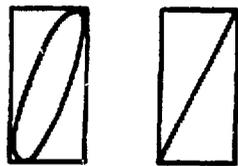
(a) Contour ellipses of constant probability and a typical Gaussian profile of the bivariate normal probability density function.



(b) Contour ellipses of constant probability are imbedded in rectangles of dimensions L_1 by L_2 , where β varies from 0 to infinity.

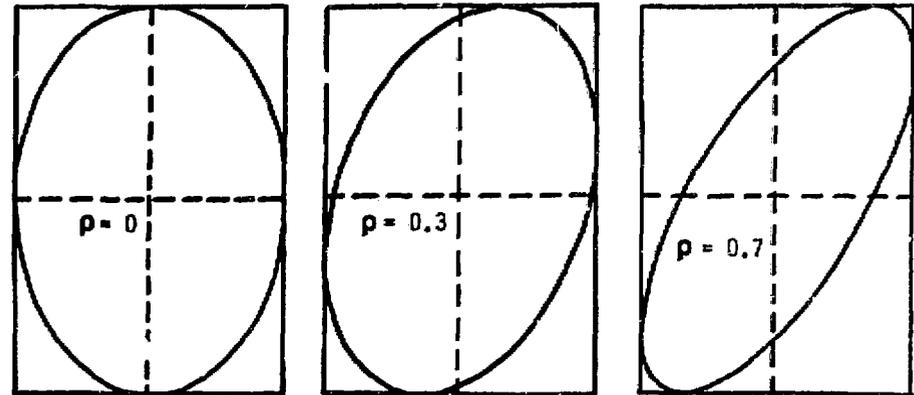


$\rho = -1$ $-1 < \rho < 0$ $\rho = 0$



$0 < \rho < +1$ $\rho = +1$

(c) Family of covariance ellipses for $-1 < \rho < +1$.



(d) Proportion of covariance ellipse vs. ρ , shown here for $\sigma_2 = \sqrt{2} \sigma_1$.

Figure 7.9: Illustration of various features of a bivariate normal probability density function.

$$[(x_1 - \lambda_1)^2/\sigma_1^2 - 2\rho(x_1 - \lambda_1)(x_2 - \lambda_1)/\sigma_1\sigma_2 + (x_2 - \lambda_2)^2/\sigma_2^2] = \text{constant},$$

is known as the covariance ellipse family or ellipses of covariance. The shapes and orientations of the ellipses depend upon ρ , σ_1 , and σ_2 , as shown in Fig. 7.9(c and d). When x_1 and x_2 are independently distributed, $\rho = 0$. Then the axes of the ellipses are parallel to the coordinate axes. Otherwise, the ellipses are tilted. However, by suitable transformation of the random variables (via formation of linear combinations of the original variables), it is possible to find an orthogonal, independent set of equivalent variables, thereby eliminating this tilt. For these bivariate distributions, this amounts to rotation of the (x_1, x_2) coordinates. When $\rho = \pm 1$, the ellipses degenerate into lines, as shown in Fig. 7.9(c and d). If $\rho = 1$, x_1 and x_2 are fully correlated and do not have distinct distributions. In fact, x_2 is always proportional to x_1 . If $\rho = -1$, x_1 and x_2 are fully anticorrelated and do not have distinct distributions. Then x_2 is always proportional to $-x_1$. In either case, the notion of a bivariate distribution for these two random variables becomes meaningless, and this is reflected by the fact that Eqs. (7.19) and (7.20) are then undefined.

8. BAYES' THEOREM AND APPLICATIONS

Bayes' theorem (or Bayes' Rule as it is commonly called) is stated and proved in Chap. 5. The notion of statistical inference is also introduced there, and the role of Bayes' theorem in this area of statistical theory is mentioned briefly. Furthermore, it is hinted in Chap. 5 that difficulties and controversies are associated with the issue of statistical inference in general, and with Bayes' Rule in particular. In this chapter we proceed to investigate this matter further. The intent here is for the reader to gain some insight into the function of Bayes' Rule in applications, and to acquire an appreciation for the subtleties of statistical inference in a wider sense. The reader is forewarned that this topic is so broad that it amounts to no less than an entire field of professional specialization. We will touch upon it again in later reports of this series, but will succeed in doing little more than just "scratching the surface." Debates over the interpretation of several important theorems of statistical inference, particularly in the field of applied statistics, have led to a dichotomy of statisticians into two distinct categories: classical statisticians and Bayesian statisticians. It is my intent to provide the reader with enough insight into this issue so that he will at least recognize and appreciate the two distinct philosophies involved. I do not care to promote one view point at the expense of others, but, as was mentioned in Chapter 5, I strongly suspect that the reader will be led to share my opinion that the process of nuclear-data development, as it is currently practiced, is essentially Bayesian in nature.

We begin by demonstrating Bayes' Rule, as embodied in Eq. 5.9, through two simple examples:

Example 8.1

Consider two indistinguishable containers, each holding five samples that are to be used in a neutron-activation experiment. The samples are either iron (Fe) or copper (Cu). However, they are physically similar in the sense that they are all disks which have nearly the same diameter, thickness, and weight. If an investigator were "blind," and thus were not able to distinguish the Fe from the Cu samples by their visual appearance, he might indeed not be able to distinguish them on the basis of his qualitative sense of "feel" alone. All that the blind investigator is presumed to know is that there are some of each type present. The information available to us, but not to our blind investigator, is that container 1 has 3 Fe samples and 2 Cu samples, while container 2 has 1 Fe sample and 4 Cu samples. This state of affairs is illustrated schematically in Fig. 8.1. Consider the following process: The blind investigator first selects a container at random, and then from it he selects a single sample at random. We choose to pose, and then to answer, two questions about this process:

First Question: What is the probability that the selected sample will be Fe?

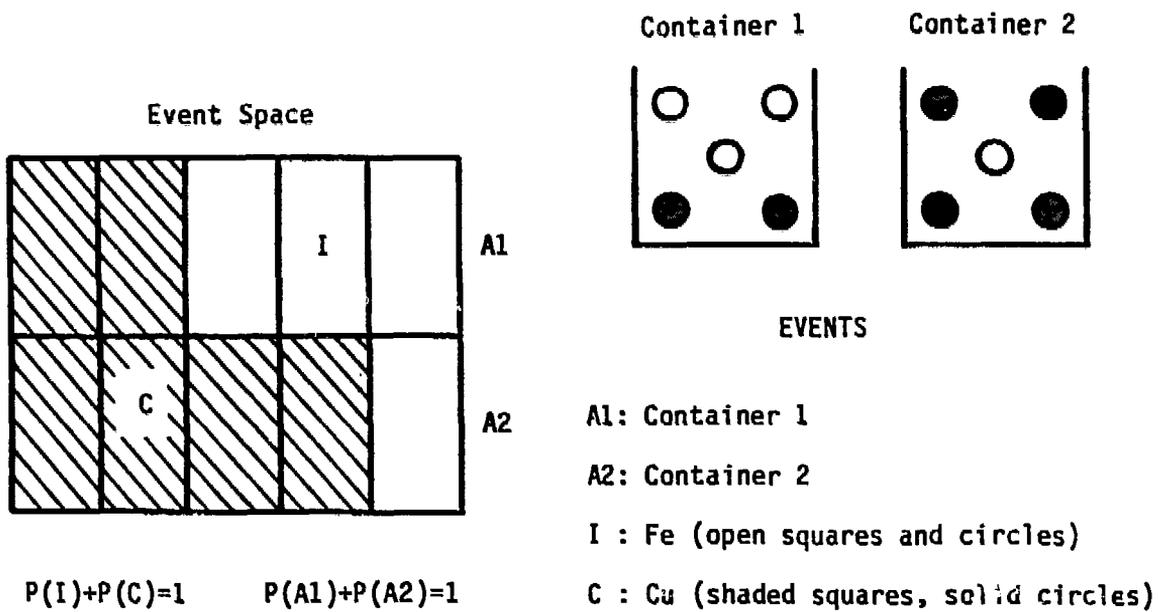


Figure 8.1: Illustration of details from Example 8.1.

Referring to Chap. 5, we recognize that this question can be answered by applying Eq. (5.8), namely, the rule of total probability. In accordance with the notion of Fig. 8.1,

$$I = (A_1 \times I) + (A_2 \times I).$$

Then,

$$P(I) = P(A_1 \times I) + P(A_2 \times I),$$

since

$$A_1 \times A_2 = \phi.$$

From Eq. 5.2, we have that

$$P(A_1 \times I) = P(I/A_1)P(A_1)$$

and

$$P(A_2 \times I) = P(I/A_2)P(A_2).$$

So,

$$P(I) = P(I/A_1)P(A_1) + P(I/A_2)P(A_2),$$

which is a statement of the rule of total probability, re-derived here in order to remind the reader of the concepts involved. The total probability for I is therefore a weighted average of the two conditional probabilities $P(I/A_1)$ and $P(I/A_2)$, with $P(A_1)$ and $P(A_2)$ serving as the weighting factors. It is very evident in this simple situation that

$$P(A_1) = P(A_2) = 1/2,$$

$$P(I/A_1) = 3/5$$

and

$$P(I/A_2) = 1/5.$$

Consequently,

$$P(I) = (1/2)(3/5) + (1/2)(1/5) = 2/5 \text{ (40\%)}$$

is the answer to our first question. We note in passing that the probability of selecting an Fe sample just equals the relative frequency of Fe samples in the entire collection of 10 samples, regardless of their distribution among the containers. The reason for this is that the selection of a container was conducted in an unbiased manner.

Second Question: Given that an Fe sample was selected, what is the probability that it came from Container 1?

The reader who has been carefully following the concepts and examples presented in this report should immediately sense that a

new element of logic is involved in this question. Previously, the reader has been asked to consider the prediction of probability for a certain event occurring, given a particular set of circumstances. Now we are asked to speculate about the circumstances themselves, given a particular outcome for a random procedure. Clearly, this problem forces us into pursuing a line of reasoning which proceeds from effect backward toward cause, rather than from cause forward toward effect, namely, logical inference rather than logical deduction. Bayes' Rule provides us with a formal algorithm for answering the question thus posed. What we seek is the conditional probability $P(A_1/I)$. Bayes' formula, namely, Eq. (5.9), states that

$$\begin{aligned}
 P(A_1/I) &= P(I/A_1)P(A_1)/[\sum_{k=1}^2 P(I/A_k)P(A_k)] \\
 &= P(I/A_1)P(A_1)/P(I).
 \end{aligned}$$

Since we indeed possess knowledge of all the parameters appearing on the right-hand side of this equation from the preceding deliberations, we can calculate $P(A_1/I)$. Thus

$$P(A_1/I) = (3/5)(1/2)/(2/5) = 3/4 \text{ (75\%).}$$

This answer makes a lot of sense, in view of our extensive knowledge of the facts in this situation. Actually, we have not learned much about the problem that we could not have already surmised from a casual glance at Fig. 8.1.

Consider, instead, what might be the outcome if our blind investigator had been called upon to answer this question. He executes the procedure of this example and is told that he has selected an Fe sample, but not which container was picked.

As this example is formulated, the blind investigator knows that there are two indistinguishable containers, each holding five physically similar samples of two types, but he does not know how many of the samples in all are Fe or how they are distributed among the containers. In order to apply Bayes' Rule, he is forced to speculate on the input parameters. For this particular situation, he has good reason to assume that

$$P(A_1) = P(A_2) = 1/2,$$

because even though he is blind he can arrange to select a container at random. However, he does not have enough information available to compute either $P(I/A_1)$ or $P(I/A_2)$. Forced to speculate on the matter, he proceeds to argue along the following

subjective lines: Because it makes good sense to store like samples together in the same container, he presumes that there exist 5 Fe and 5 Cu samples, and that they are segregated. Then he is faced with the following possibilities: either i)

$$P(I/A_1) = 1 \text{ and } P(I/A_2) = 0,$$

or ii)

$$P(I/A_1) = 0 \text{ and } P(I/A_2) = 1.$$

He applies Bayes' theorem as follows:

Possibility (i):

$$P(A_1/I) = (1)(1/2)/(1/2) = 1.$$

Possibility (ii):

$$P(A_1/I) = (0)(1/2)/(1/2) = 0.$$

Since the predictions are so different, he decides to hedge by averaging these two possibilities, thereby suggesting that

$$P(A_1/I) = 1/2 \text{ (50%).}$$

Actually his result, though at variance with the truth as we know it, is not too very far removed from it. In actual fact, he was merely fortunate, because it is obvious that he just did not have enough factual information available to him to cope with the problem in a meaningful way. For example, his speculation that there were equal numbers of Fe and Cu samples, and that they were segregated, bear little resemblance to reality.

The lesson to be learned from this example is that although Bayes' Rule is a completely valid mathematical result, its application is subject to question when the available (a priori) input information is speculative.

Next, we apply Bayes' Rule to an example involving radioactivity:

Example 8.2

Consider Fig. 8.2. We suppose that the indicated detector responds to indistinguishable radiation from two physically distinct, weak radioactive sources with emission strengths S_1 and S_2 , respectively. For example, these sources might each decay via positron emission, with the detector responding explicitly to the corresponding 0.511-MeV annihilation-radiation photons. We further assume that the efficiencies for detection of the photons

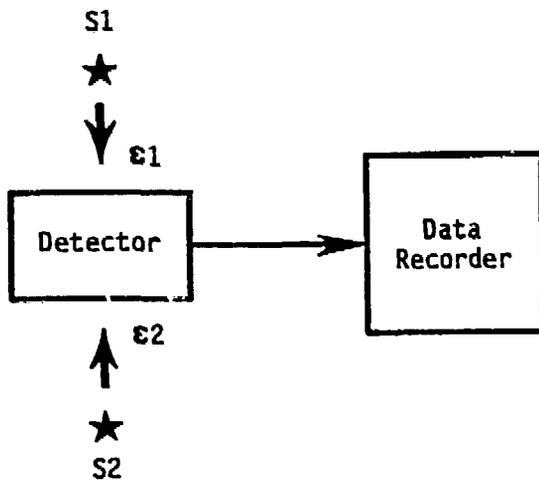


Figure 8.2: Schematic diagram of the experiment from Example 8.2.

from Source 1 and Source 2 are ϵ_1 and ϵ_2 , respectively. Since the emission rates from these sources are taken to be low, we shall make the assumption that the likelihood of two photons (one from each source) being detected within the required signal-processing time is negligible. Define the following events:

A_1 = emission of a photon from Source 1,

A_2 = emission of a photon from Source 2,

B = detector records a count.

The question that we wish to ask is the following: Given that the detector has recorded a count (event B has occurred), what is the probability that the photon that produced the count originated from Source 1? We can solve this problem using Bayes' Rule. Formally, the answer is

$$P(A_1/B) = P(B/A_1)P(A_1) / \left[\sum_{k=1}^2 P(B/A_k)P(A_k) \right].$$

The probabilities $P(A_1)$ and $P(A_2)$ are proportional to the respective source strengths. Thus,

$$P(A_1) \propto S_1$$

and

$$P(A_2) \propto S_2.$$

The conditional probabilities $P(B/A_1)$ and $P(B/A_2)$, i.e., the likelihood factors, are proportional to the respective efficiencies. Thus,

$$P(B/A_1) \propto \epsilon_1$$

and

$$P(B/A_2) \propto \epsilon_2.$$

Combining these results, we obtain

$$P(A_1/B) = \epsilon_1 S_1 / (\epsilon_1 S_1 + \epsilon_2 S_2)$$

from Bayes' Rule. To add life to this example, let us assume that $S_1 = 10S_2$ and $\epsilon_1 = 2\epsilon_2$. Then,

$$P(A_1/B) = (2)(10)\epsilon_2 S_2 / [(2)(10)\epsilon_2 S_2 + \epsilon_2 S_2] = 20/21 \text{ (95.2\%)}.$$

We can derive the probability that B was caused by a photon from Source 2 by a completely symmetrical development, namely,

$$P(A_2/B) = \epsilon_2 S_2 / [(2)(10)\epsilon_2 S_2 + \epsilon_2 S_2] = 1/21 \text{ (4.8\%)}.$$

This sort of analysis is quite useful in correcting the results of a radioactivity measurement for background effects. In such a situation, Source 1 is considered to be the primary source (foreground) while Source 2 represents the background source. Now if Source 1 is a point source, computation of ϵ_1 may be a manageable task. If Source 2 is background radiation, its distribution in space may not be known, and thus ϵ_2 may not be calculable. A key point is that Bayes' Rule in this instance requires only that the product (ϵS) be known for each source, since $(\epsilon S)_1 = \epsilon_1 S_1$ and $(\epsilon S)_2 = \epsilon_2 S_2$. Furthermore, since the detector count rates R_1 and R_2 for each source are equivalent to $(\epsilon S)_1$ and $(\epsilon S)_2$, respectively, the information required for the application of Bayes' theorem is directly measurable! In practice, we measure $(R_1 + R_2)$, namely, the detector count rate with both sources present (foreground plus background), and the background rate R_2 when Source 1 is removed. Then, in terms of measurable quantities, Bayes' Rule states

$$P(A_1/B) = R_1 / (R_1 + R_2) = [(R_1 + R_2) - R_2] / (R_1 + R_2).$$

So, we see that the well-known procedure for correcting radioactivity data for background effects originates from the notion of Bayesian statistical inference.

It is evident from these examples that use of Bayes' Rule demands that one become comfortable with the notion of backward or inductive reasoning, and possess a willingness to employ subjective information as a basis for statistical inference. In the field of nuclear data these are relatively familiar conditions, since there is rarely enough information at hand to be able to make predictions concerning nuclear parameters with a great degree of certainty. Complete faith in the process of learning through experience is thus implicit in this procedure. One must possess some confidence in the notion that, even though experience and subjective intuition may appear to form an uncertain foundation for the building of rigorous knowledge, uncertainty associated with this subjectivity will gradually be dispelled as factual information is accumulated through sound experimental procedure. An example which ought to instill a degree of confidence in the reader concerning this approach is presented later in this chapter.

Recall from Chap. 5 that Bayes' theorem is formally based upon a collection of events A_i ($i = 1, n$) which form a partition of the entire event

space E . These A_i are mutually exclusive and exhaustive. The manifestation of this condition in the domain of realistic applications is that all possible hypotheses or causes for an observed event must be taken into consideration, and they must indeed be distinct. In addition to considering all reasonable hypotheses, we must be in possession of tenable a priori estimates of the probabilities for validity of each of these hypotheses. It is then required to determine the conditional probabilities (likelihood factors) that the event, which in fact has actually been observed, might have been caused by each of the hypotheses under consideration. These stringent demands of Bayes' Rule often exceed what can be realistically met for many intended applications. The main problem that generally emerges is usually not one of calculating likelihood factors, but rather one of estimating the a priori probabilities of the various hypotheses.

Since a priori probabilities must often be subjectively estimated, there have indeed been instances of abuse and distorted reasoning associated with the application of Bayes' Rule, with the consequence that for many years it suffered from discredit in the eyes of numerous statisticians. While quite a few statisticians continue to view with considerable skepticism the methods which this theorem suggests, it has enjoyed a certain resurgence of popularity in recent years. In the imperfect world in which we live, prudent application of Bayes' Rule clearly provides us with a powerful tool for merging speculative information with factual knowledge, in order to advance our understanding of a particular situation. One approach which has come to be accepted rather widely as being a legitimate one is that in which a priori probabilities and likelihood factors required for the application of Bayes' Rule are derived from mathematical theories that appear to offer promising possibilities for explanation of the phenomena in question. An important criterion for a theory to be considered as reasonable and practical is that it offer clearly defined algorithms for calculation of the likelihoods associated with phenomena which have actually been observed. Experimentation, in concert with Bayesian analysis of the data, permits us to refine our understanding of the issue at hand--rejecting theories which appear to be at odds with observation while reinforcing those which are found to be consistent with reality. The a posteriori probabilities in this context represent an amalgamation of theory and experiment in the sense that the hypotheses for which theory and experiment are most consistent are favored by greater a posteriori probability than are those for which theory and experiment are at odds. Acceptance of this philosophical approach to inductive reasoning does not appear to be at odds with generally accepted methods of scientific investigation.

Bayesian analysis is certainly not the only method used in inductive reasoning. It is not our intent to dwell at length on this matter, but it is worthwhile to mention here another method that is based on what is known as the Principle of Maximum Likelihood. This principle states that when confronted with an exhaustive set of several hypotheses which might have caused a particular observed effect, we select as correct, unequivocally, the single hypothesis which is most likely to have produced that effect. This approach to inductive reasoning would appear to be less conservative than Bayesian reasoning because it seems to be inconsistent with the normal learning process, namely, that knowledge is acquired, and confidence in the

validity of certain hypotheses relative to other possible ones is developed in gradual steps. While the Principle of Maximum Likelihood obviously offers logical expediency, there would appear to be an attendant risk of prematurely forming rigid conclusions which might well be prone to bias. This risk could be significant when the observations are subject to considerable variance or possible systematic effects, as is often the case in nuclear-data research. We will encounter the Principle of Maximum Likelihood in other contexts in a later report in this series, and there we will learn that it does have a valid place in certain aspects of statistical theory. In fact, in some instances the application of this principle leads to results which are consistent with the Bayesian approach. For these reasons, the reader should not be led to regard the Principle of Maximum Likelihood as unworthy of further consideration.

An important problem of statistical inference is that of predicting the future outcomes from a sequence of random trials, given that a certain history of outcomes has been compiled for previous trials of identical nature. We limit consideration to sampling procedures where the outcomes can be categorized simply as either "success" or "failure," i.e., Bernoulli trials. This problem can be formally posed and answered in the form of a theorem:

Theorem: [Laplace's General Rule of Succession]

Given that n random trials have been performed with respect to an essentially infinite population, and that all of these have been successful (event A), the probability that the next n' trials will also be successful (event B) is given by the formula:

$$P(B/A) = (n + 1)/(n + n' + 1). \quad (8.1)$$

Proof of this theorem begins with the rule of total probability and ultimately involves approximation of certain discrete sums by definite integrals, an approximation that is well justified for large populations. The proof is given in Parzen (Par 60) and will not be repeated here.

A special case of this theorem is that for which $n' = 1$:

Theorem: [Laplace's Special Rule of Succession]

Given that n random trials have been performed with respect to an essentially infinite population, and that all of these have been successful (event A), the probability that the very next trial will also be successful (event B) is given by the formula

$$P(B/A) = (n + 1)/(n + 2). \quad (8.2)$$

The latter result generally is referred to simply as Laplace's Rule of Succession, a law which has garnered considerable notoriety in the annals of statistical history. Laplace himself was aware of the potential for abuse which it afforded. In jest, he once was quoted as using this theorem to "prove" that there was a finite probability (in fact, 1 part in 182614) that

the sun would not rise on the following day based on recorded observations that the sun had risen faithfully on each day during the preceding 5000 years!

As with all exercises in inductive reasoning, one can trust the predictions of Laplace's Rule (in the statistical sense) only to the extent that the process to which it is applied is random, with a legitimate range of possible outcomes. For example, in the case of Laplace's paradox, the rising of the sun is not governed by statistics (insofar as we are aware), and the event that the sun does not rise on a particular day is not just improbable, it is inconceivable! We know this to be the case based on knowledge which transcends the mere fact that the sun has been observed to rise each day during a period of 5000 years of recorded history.

We will now provide two further examples of the use of Laplace's Rule of Succession. From these it is hoped that the reader can "infer" where the hazards lie. The first example demonstrates proper use of the rule:

Example 8.3

A research laboratory stocks a sizable supply of a particular transistor which is commonly used in applications in that lab. This transistor is often replaced in the circuitry where it is found. The purchased transistors were randomly selected from a very large population, the manufacturer's inventory. The laboratory in question has had no previous experience with this vendor. It is anticipated that there will be some defective transistors in the lot, but no data to this effect are available a priori to the lab. Some time after the initial purchase, it is noted that the last 8 transistors used in the lab were good ones. What is the probability, based on the lab's experience, that the next transistor selected will also be good? Laplace's Law of Succession provides the answer 9/10 (90%). While there might be reason to expect the probability to be higher (otherwise the vendor would have a bad reputation), this result is not unreasonable, as a conservative choice. The validity is supported by the fact that the trials were performed at random from a large population, and that both success and failure were physically possible outcomes. Finally, the ninth trial is performed under the same circumstances as the previous eight.

The next example demonstrates improper use of Laplace's Rule of Succession:

Example 8.4

A scientist wishes on a particular day to perform a certain measurement outdoors using a temporary setup which is intolerant of rainy conditions. For example, he might wish to set up an

unprotected scintillation detector at a long distance from a neutron source in order to measure a certain spectrum by time-of-flight with exceptionally good resolution. On the previous eight days it did not rain. What is the probability that on the ninth day it will not rain? Laplace's Rule of Succession provides the answer 9/10 (90%). Common sense tells us that this is absurd! In order to have a reasonable chance of success in predicting the weather on a given day, one must scientifically analyze a great deal of information (e.g., the time of year, the climatic history of the region over an extended period of time, the current weather conditions in neighboring regions, etc.). Anyone who relies on Laplace's Rule of Succession to predict the weather is being naive.

The preceding considerations lead to a statement of what could aptly be called the Fundamental Principle of Applied Probability Theory: Before applying any theorem, particularly in the realm of statistical inference, be certain that the conditions under which the theorem is being applied are exactly those upon which the theorem is predicated, and, furthermore, that these conditions will continue to hold well into the future period in which one is attempting to make a prediction.

It has been indicated previously that Bayes' Theorem is an important concept of applied statistics because it offers an algorithm for combining old and new information in a manner that builds knowledge, i.e., it formalizes the learning process. We now outline in more detail how the Bayesian learning process proceeds, and ultimately demonstrate it with a very detailed example.

The Bayesian method offers its greatest potential when viewed as an iterative process. This process is outlined formally in Table 8.1. The reader should study this table carefully and, in particular, should keep in mind a very important caveat regarding the issue of "confidence." Implicit in the application of the process described in Table 8.1 are two distinct notions of confidence. First, there is the idea of confidence that hypothesis A_k is the correct one. After m trials, the confidence levels for the various A_k are properly reflected in the corresponding probabilities $P_m(A_k)$, as indicated in the table. However, there is another type of confidence with which one should be familiar. That is the "confidence" that occurrence of an event B will provide us with a strong indication as to which hypothesis is correct. This latter form of confidence is measured by $P(B/A_k)$, i.e., the likelihood factors. For example, if $P(B/A_j) \approx 1$ and $P(B/A_k) \approx 0$ for $k = 1, n$ (excluding j), then the observance of B strongly enhances our confidence that A_j is the correct hypothesis.

Those who work in the field of nuclear data will appreciate the following manifestation of this issue: One desires to know a particular cross section σ , which is assumed to have a precise and unchanging value,

Table 8.1 Bayesian "Learning" by Iterative Application of Bayes' Rule.

Step 1: Establish a mutually exclusive and exhaustive set of hypotheses A_k ($k = 1, n$).

For each A_k , develop an a priori probability $P_0(A_k)$ that the hypothesis is correct, based on the best available rational knowledge. If there is no a priori reason to favor one hypothesis over another, then assume $P_0(A_k) = 1/n$ (Bayes' Postulate) in order to avoid bias.

Step 2: Perform the first random trial. Event B_1 occurs. Calculate all the likelihoods $P(B_1/A_k)$ ($k = 1, n$). It is assumed that a requisite feature of each hypothesis A_k is that it provide an algorithm for calculating such likelihoods! Calculate all the a posteriori probabilities by means of Bayes' Rule.

$$P_1(A_k/B_1) = P(B_1/A_k)P_0(A_k) / \left[\sum_{i=1}^n P(B_1/A_i)P_0(A_i) \right] \quad (k = 1, n)$$

Step 3: Replace each a priori probability $P_0(A_k)$ by the a posteriori probability $P_1(A_k/B_1)$, i.e. let

$$P_1(A_k) = P_1(A_k/B_1) \quad (k = 1, n).$$

Step 4: Perform the second random trial. Event B_2 occurs. Calculate all the likelihoods $P(B_2/A_k)$ ($k = 1, n$). Calculate all the a posteriori probabilities by means of Bayes' Rule.

$$P_2(A_k/B_2) = P(B_2/A_k)P_1(A_k) / \left[\sum_{i=1}^n P(B_2/A_i)P_1(A_i) \right].$$

Table 8.1 (cont'd)

Step 5: Replace each a priori probability $P_1(A_k)$ by the a posteriori probability $P_2(A_k/B_2)$, i.e., let

$$P_2(A_k) = P_2(A_k/B_2) \quad (k = 1, n).$$

This iterative process is continued, with the results B_3, B_4, \dots, B_m of the successive random samplings incorporated into the learning process described above.

After m independent trials have been performed, each hypothesis A_k is found to have a probability $P_m(A_k)$ which represents the best estimate of the probability that A_k is the correct hypothesis. At no stage of this process is any legitimate hypothesis explicitly rejected. However, the experience gained from observation (learning) leads to continuous refinement of our confidence in the various hypotheses, with ultimate downgrading of those hypotheses that are at odds with the observations B_1, B_2, \dots, B_m , and enhancement of confidence in others that seem to be consistent with our observations.

If m is small, then the initial estimates $P_0(A_k)$ of the probabilities can be expected to have a noticeable impact on the final $P_m(A_k)$. For larger m , the effects of the initial estimates will be "washed out" by the bulk of new evidence. The number of iterations required to wash out the effects of the a priori estimates will depend largely on the nature of the likelihood factors. It is in these factors that one finds quantitative indication of the "precision" or "reliability" of the various observations B_1, B_2, \dots, B_m . In other words, the likelihood factors reflect the uncertainties associated with these observations. Since observations (measurements) are demanding of resources (time, manpower, and money), it is essential that this process embody reasonable hypotheses (theories) and observational procedures (experiments) which are reliable (accurate) and carefully designed to test the hypotheses under consideration.

although it is not known to us with any precision at the outset. Nuclear model calculations provide us with a range of possible values for σ , and an associated probability density function $p_0(\sigma)$. Repeated experimentation ought to lead us to the position where we can state with a certain level of confidence that σ must lie in the range $\langle\sigma\rangle - \Delta\sigma \leq \sigma \leq \langle\sigma\rangle + \Delta\sigma$. However, the number of experiments which have to be performed in order to achieve the desired level of accuracy depends strongly on the precision of the experimental process (sharpness of the likelihood factors), which ought to be reflected in the variance of the accumulated experimental results.

We proceed now to demonstrate, by means of a rather lengthy example, what has been said above regarding the Bayesian learning process, as summarized in Table 8.1.

Example 8.5

Consider a hypothetical situation in which a certain physical parameter has a precise integer value designated by $\langle k \rangle$. In this example we assume that a hypothetical investigator does not know what $\langle k \rangle$ is, even though we do know that in fact $\langle k \rangle = 5$. What our investigator does know, however, is that $\langle k \rangle$ might be any integer between 0 and 20. Furthermore, we suppose that when our hypothetical investigator performs random experiments in order to gather information relevant to the determination of $\langle k \rangle$, he will observe various values k between 0 and 20 according to the Poisson probability law

$$p(5;k) = 5^k e^{-5} / k!,$$

i.e., the value $k = 5$ is the expected value, but the values actually observed in the course of repeated sampling will scatter considerably around $k = 5$. Our investigator does not know, however, that the method he uses to conduct the sampling (the "measurement" procedure) is governed by this particular law. Nevertheless, he needs to make assumptions concerning the a priori probabilities, and he also requires rules for calculating likelihoods. The hypothesis A_k in this example corresponds to the statement that $\langle k \rangle = k$. Several possible choices for the a priori probabilities $P_0(A_k)$ which the investigator might make are considered in this example as distinct cases. The investigator chooses to employ Poisson distributions for the calculation of likelihoods. The sampling process yields a sequence of integers k_1, k_2, \dots, k_N , all in the range 0 to 20. Our investigator chooses to calculate likelihoods according to the formula

$$P(k_i/k) = k^i e^{-k} / k_i! \text{ for } i = 1, N \text{ and } k = 1, 20.$$

The Poisson formula does not apply when $k = 0$. Instead, it is assumed that $P(k_i/0) = 1$ if $k_i = 0$, and 0 if $k_i > 0$. This

assumption is consistent with the Poisson probability law in the limit of very small k .

A microcomputer with a random number generator is employed to simulate, by the Monte-Carlo method, the process of sampling integers from 0 to 20. Four different cases are treated, each corresponding to a different set of a priori assumptions. For each case, seven distinct simulated procedures are conducted. These procedures correspond to $N = 1, 5, 10, 20, 50, 100,$ and 1000 observations (experiments), respectively. Each experiment is conducted "independently" of all others, at least to the extent to which the various random numbers produced by the generator are truly independent.

First, we examine how well the Monte-Carlo method of sampling integers at random from 0 to 20 performs. The results of this analysis are plotted in Fig. 8.3. It is seen that after $N = 100$ observations, the accumulated results of sampling are distributed very nearly like the Poisson distribution $p(5;k)$, thereby assuring that the simulation process is satisfactory.

Case 1:

Our hypothetical investigator has no a priori information regarding $\langle k \rangle$ other than the range of possibilities, 0 to 20. He therefore employs Bayes' Postulate, assigning equal value $P_0(A_k) = 0.0476$ ($k = 0, 20$) to each probability, as shown in Fig. 8.4. The first procedure with $N = 1$ (a single experiment) produced $k_1 = 4$, and the conclusions indicated in Fig. 8.4. When Bayes' Postulate is applied, the a posteriori probabilities are dominated by the likelihoods when a single experiment is performed. Thus, for $N = 1$,

$$P(k) \propto P(k_1/k).$$

The second procedure involved five experiments. The first experiment produced $k_1 = 4$, coincidentally, the same value as did the first procedure. It is seen, after four more experiments, that the final outcome appears to have been influenced considerably by the result of the first experiment. Procedures involving larger numbers of experiments N produce distributions $P(k)$ which are more and more sharply peaked around $\langle k \rangle = 5$. This Bayesian procedure does seem to converge on the value $\langle k \rangle = 5$, with increasing confidence, as N becomes large. Our investigator measures his confidence that $\langle k \rangle = 5$ by calculating the probability that $\langle k \rangle \neq 5$, i.e., $P(\langle k \rangle \neq 5)$. The results are given in Fig. 8.4, and they are self-evident. The reader may be surprised that it requires so many trials to refine the knowledge of $\langle k \rangle$. This is true because the "measurement" process, as considered in this example, is not a very precise one. Furthermore, the likelihoods reflect a similar lack of

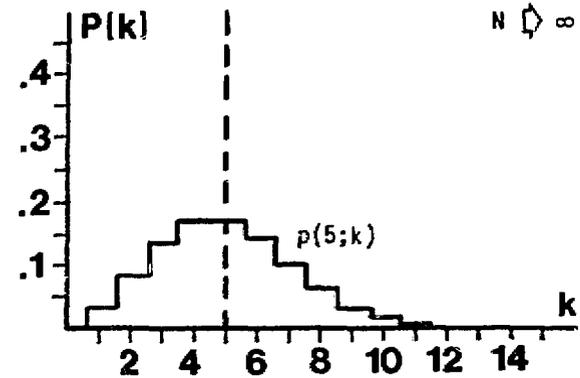
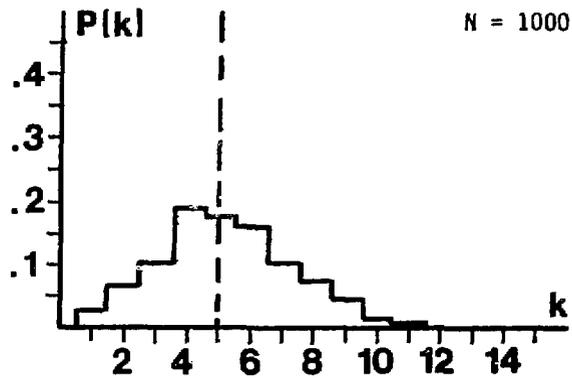
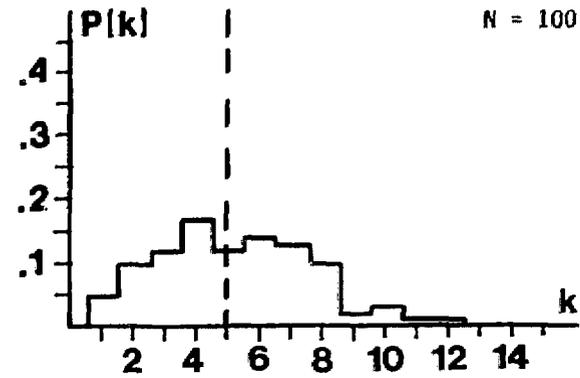
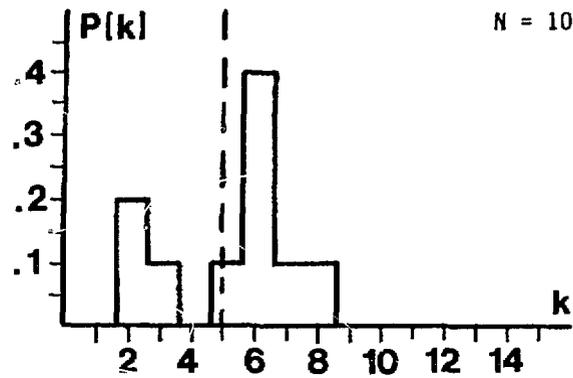


Figure 8.3: Normalized distributions $P(k)$ of outcomes from random sampling by Monte-Carlo of integers k between 0 and 20 from a population which is distributed according to the Poisson distribution $p(5;k) = 5^k \exp(-5)/k!$. As the number of sampled values N becomes large, the observed distribution of outcomes $P(k)$ approaches the fundamental Poisson distribution $p(5;k)$ of the population taken as a whole. Note that values of $k > 20$ are so unlikely that they can be neglected. See Example 8.5.

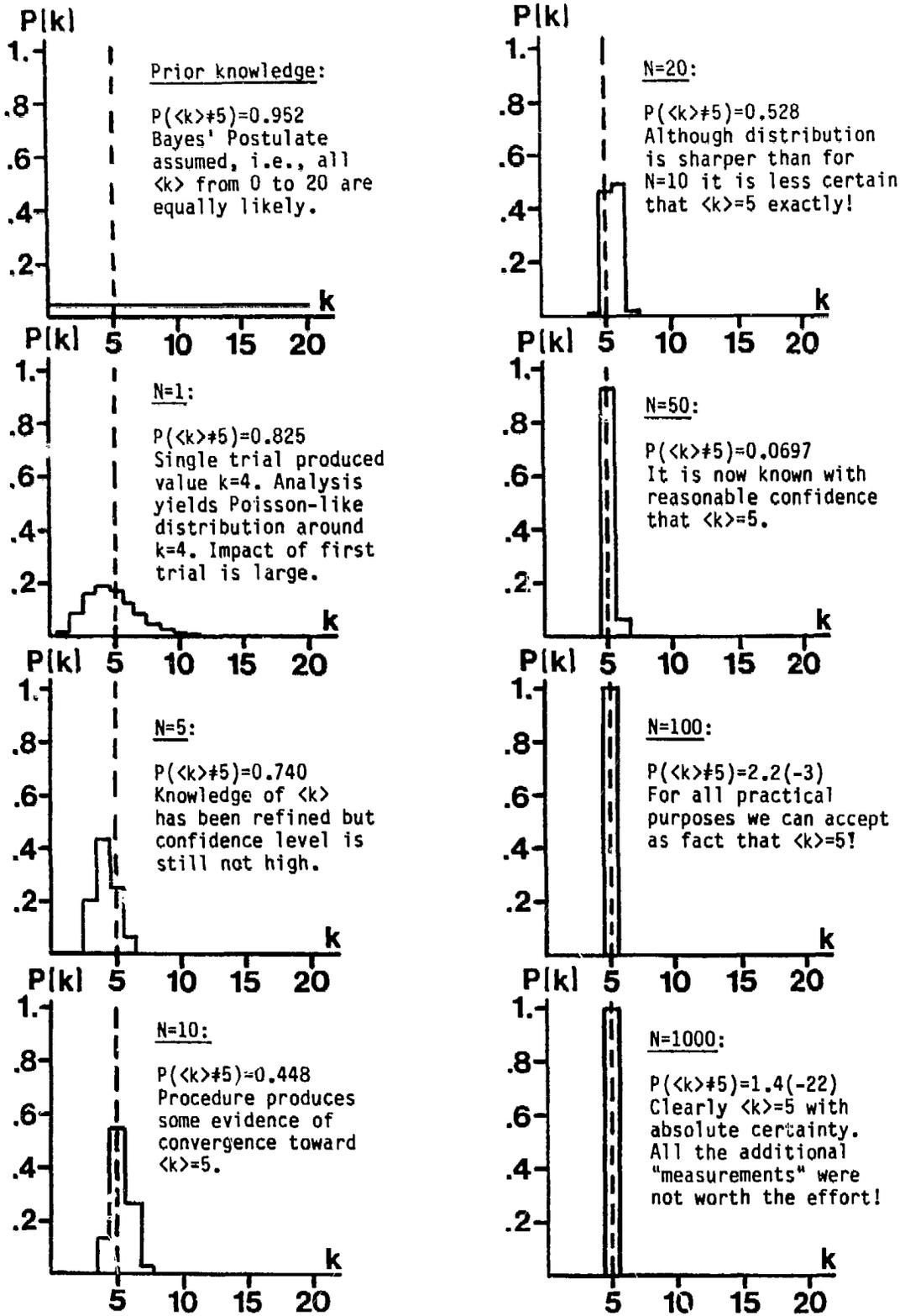


Figure 8.4: Results of Bayesian analyses of the simulated-measurement exercises described in Case 1 of Example 3.5.

selectivity. It is a well-known rule of experimental science that a few high quality measurements are more valuable than a host of experiments with low precision. Accordingly, any experimenter who is planning a measurement should be aware of the level of accuracy that will be necessary in order for the measurement to have an impact on the issue at hand.

Case 2:

Here, it is assumed by our investigator a priori that $\langle k \rangle$ is most likely to be 6, but that there is an associated Gaussian uncertainty distribution, with standard deviation $S = 1$. Since the allowed values of k are integers, the normal distribution must be interpreted accordingly. In reality, it is a distribution that is applicable only to continuous random variables. This is a more realistic situation, not unlike circumstances often encountered in the field of nuclear data. This a priori knowledge of the parameter in question is much more definitive than the situation represented in Case 1 (a nearly total absence of knowledge requiring application of Bayes' Postulate). We suppose that our investigator is led to this assumption either by examination of an existing prior data base or by consideration of a theoretical model. The results of various sets of experiments are summarized in Fig. 8.5. Two points ought to be made regarding this case: First, it still takes quite a few trials N to refine the knowledge of $\langle k \rangle$, even though the a priori assumption was not too distant from the truth. Again, this reflects the lack of "precision" of the "measurement" process. Another more puzzling result is that the final outcome for the experiment with $N = 100$ trials is less favorable than that for $N = 50$! The reader should keep in mind that these two sets of trials are distinct. Statistical artifacts of this nature also appear in the simulation exercises of Case 3 and Case 4, below. Whether such occurrences are encountered in realistic situations is a matter we are not prepared to explore at this point in the development of this subject.

Case 3:

In this case, the a priori assumption is that $\langle k \rangle$ is most likely to be 5, with a Gaussian-like uncertainty distribution of small width (standard deviation $\sigma = 0.5$). The results of this exercise are summarized in Fig. 8.6. Here, it is very evident that the knowledge of $\langle k \rangle$ is dominated by the a priori distribution through many experiments. The results of these experiments do not conflict with the a priori assumption, yet they do very little to refine the knowledge of $\langle k \rangle$ until N is quite large, since the "measurement" precision is rather poor. This case clearly illustrates the important point that once a physical quantity is known rather well, a few additional measurements are of little value unless they are very accurate ones.

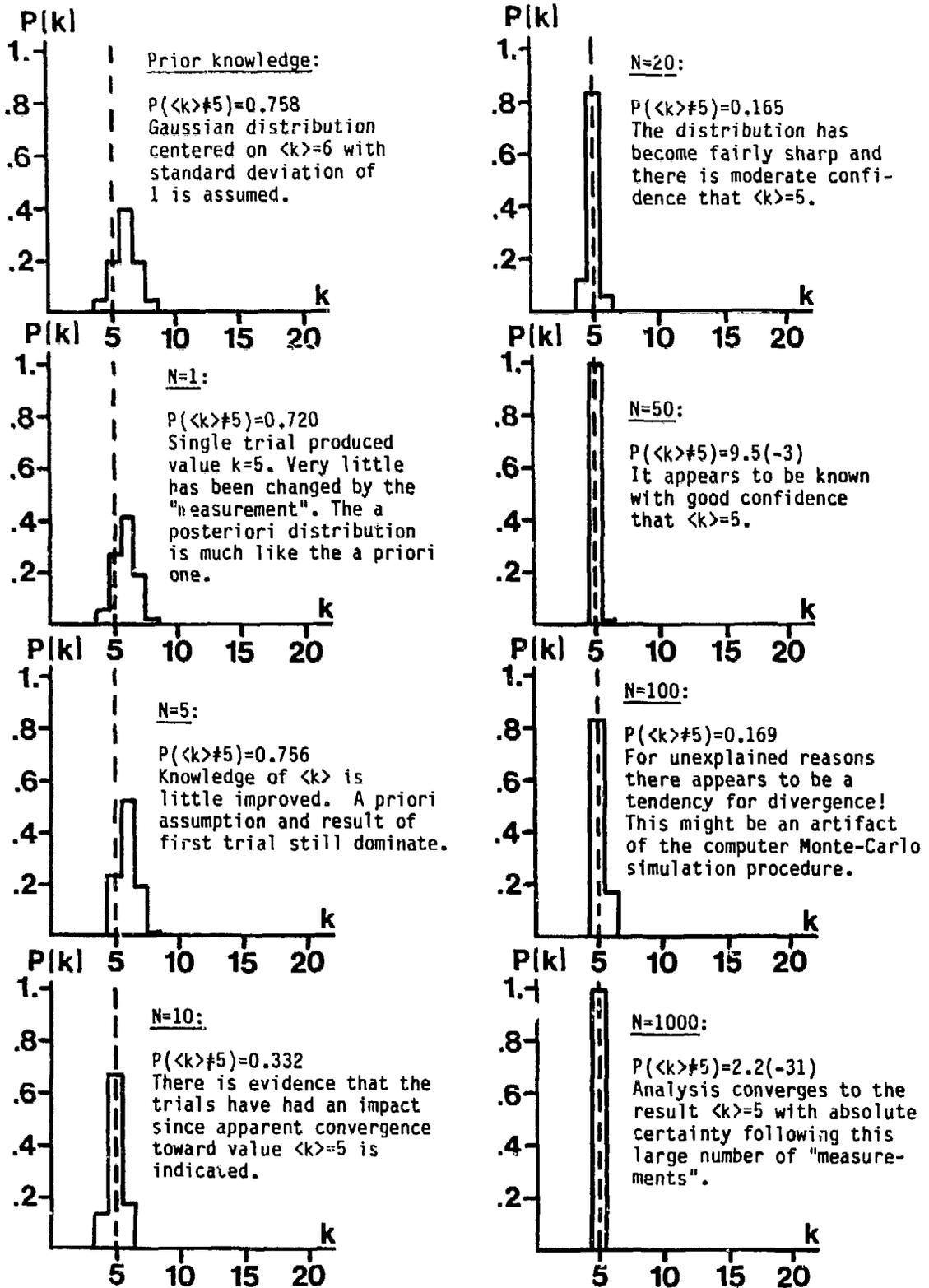


Figure 8.5: Results of Bayesian analyses of the simulated-measurement exercises described in Case 2 of Example 8.5.

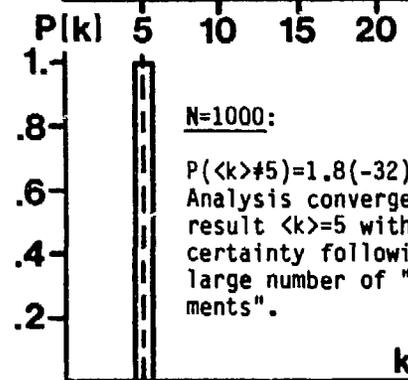
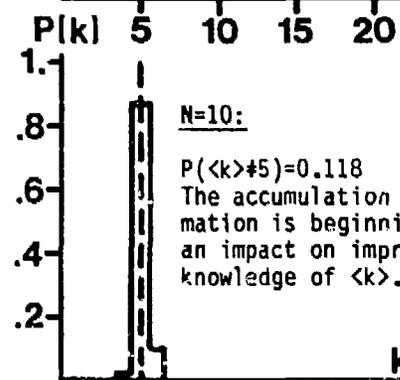
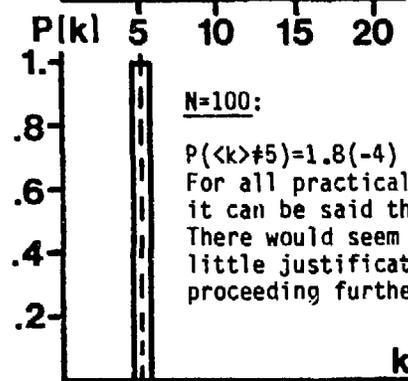
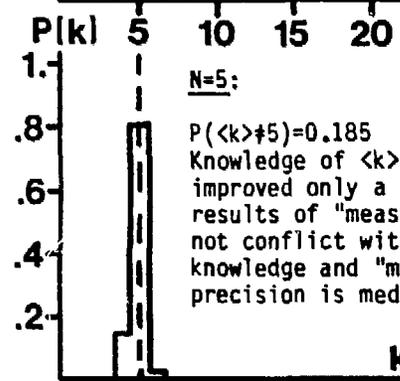
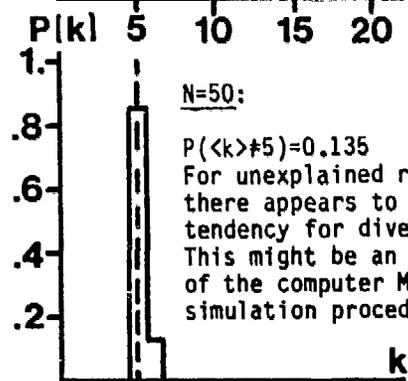
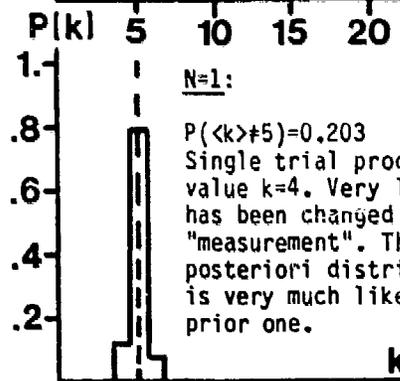
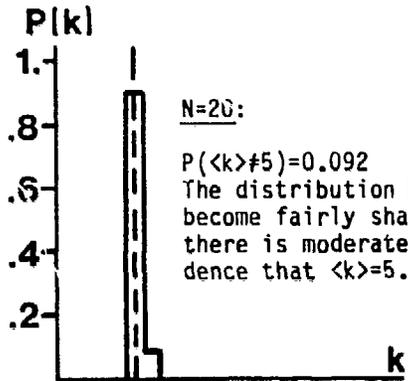
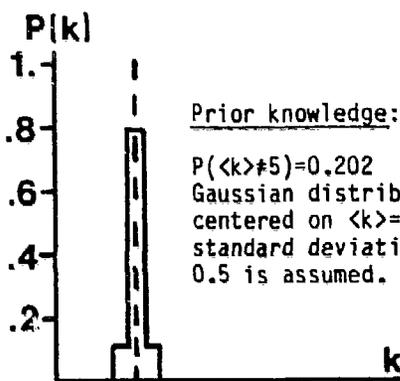


Figure 8.6: Results of Bayesian analyses of the simulated-measurement exercises described in Case 3 of Example 8.5.

Case 4:

Finally, the power of the Bayesian learning process to overcome the effect of an erroneous a priori assumption is tested. Referring to Fig. 8.7, it is assumed, a priori, that $\langle k \rangle = 8$, with a Gaussian-like uncertainty distribution of width characterized by the standard deviation $\sigma = 1$. From Fig. 8.7, it is evident that the "measured" results ultimately overwhelm the initial wrong assumption, but it takes many experiments. Thus, if knowledge of a physical quantity is in error at the outset, and this discrepancy is compounded by according too high a confidence level to the a priori assumed value, then a few more measurements of modest quality simply won't have enough impact to overcome initial prejudices, at least not in the Bayesian context.

Because of the expense and time-consuming nature of real physical experiments, it is quite unlikely that repetitions $N > 10$ would be practical in most instances. In the analyses for Cases 1 through 4, we have seen demonstrations that final results can be very sensitive to the initial assumptions and to the outcomes of the first few trials, with convergence via Bayesian learning assured only for large N . It would appear, then, that there is such a thing as luck (both good and bad) in applied statistics. To examine this notion, we consider Case 1 once again in the following context: Five distinct computer procedures are conducted, each involving $N = 10$ experiments. The outcomes are compared in Table 8.2, in terms of the confidence factors $P(\langle k \rangle \geq 5)$. The results are rather disturbing. Although there is greater confidence that $\langle k \rangle = 5$ after each set of experiments than existed before them, it is evident that a large variance exists in the final confidence levels. In one of the procedures (the fourth), the improvement in confidence is so miniscule that, for all practical purposes, the "measurements" were conducted in vain. The essence of the problem in this example is the fact that the "measurements" are not very precise, so the "measurement" process itself is not very selective, as reflected by the assumed Poisson-derived likelihoods. The most important lesson to be learned from this example, then, is that research effort should, as a general rule, be carefully planned to offer good accuracy and precision, and a high degree of selectivity with respect to the physical quantities under consideration! In the field of nuclear data it is all too often apparent that this rule has been ignored!

Modern nuclear-data evaluation procedures have generally evolved from either a Bayesian approach, or from the method of least squares which is more akin to the Principle of Maximum Likelihood. Under certain conditions, both of these approaches lead to the same formulas for combining new and old data, as is discussed by Peelle (Pee 82). The common result is most often referred to as the Generalized Least-Squares Method. We can pursue this issue no further until the concept of least squares has been developed, and this is a task for a future report in this series.

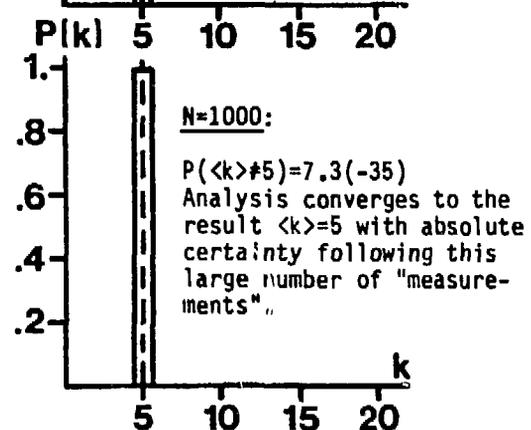
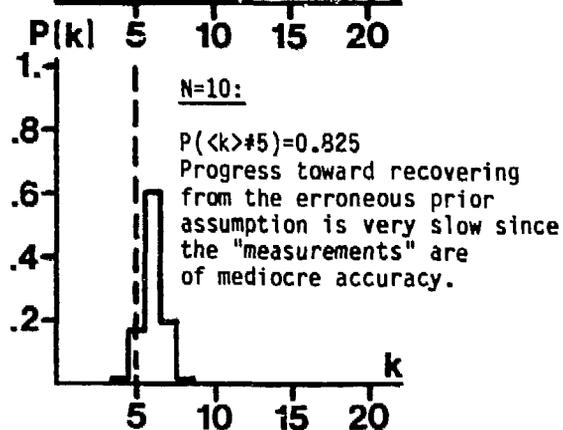
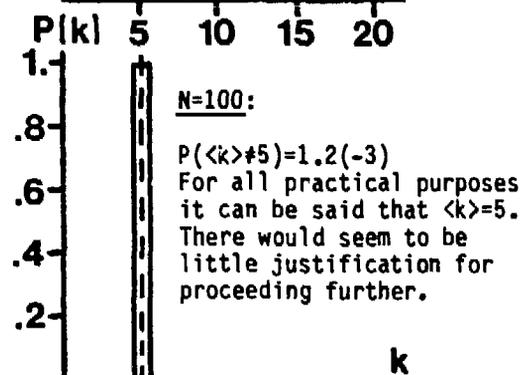
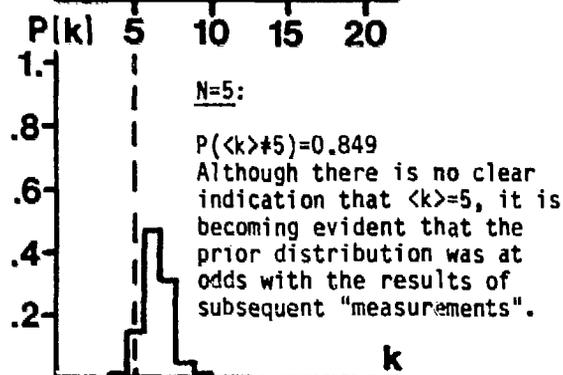
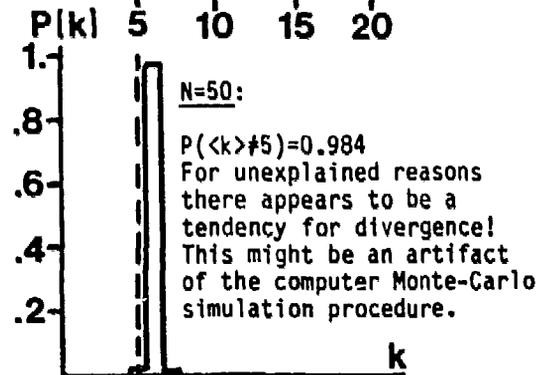
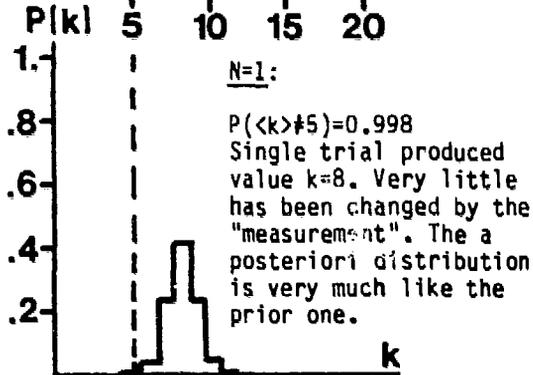
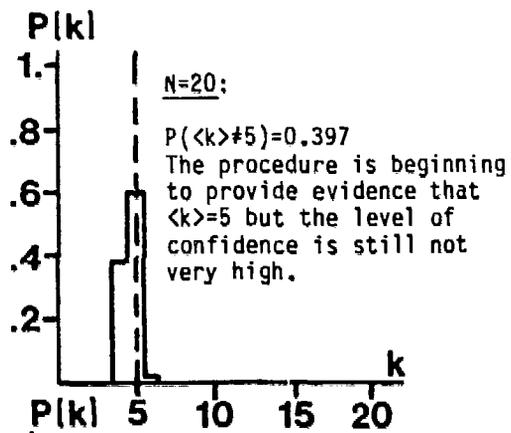
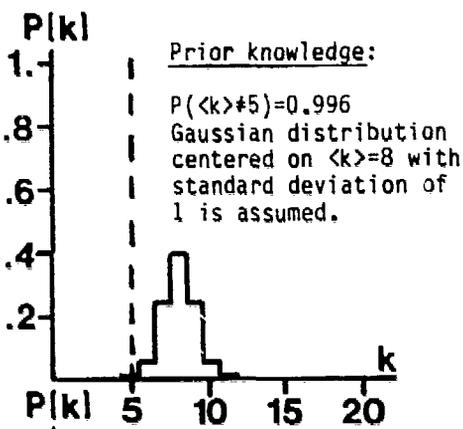


Figure 8.7: Results of Bayesian analyses of the simulated-measurement exercises described in Case 4 of Example 8.5.

Table 8.2 Comparison of five distinct computer procedures involving $N = 10$ experiments each^a

Procedure	$P_{10}(\langle k \rangle \neq 5)$ ^b
1 st	0.470
2 nd	0.437
3 rd	0.867
4 th	0.937
5 th	0.448

^a Refer to Ex. 8.5. A priori probabilities $P_0(k) = 0.0476$ ($k = 0, 20$), in accordance with Bayes' Postulate, as discussed for Case 1. Thus, $P_0(\langle k \rangle \neq 5) = 0.952$ measures the a priori confidence level in the assertion that $\langle k \rangle = 5$.

^b $P_{10}(\langle k \rangle \neq 5)$ measures the confidence that $\langle k \rangle = 5$ after ten Monte-Carlo experiments

Earlier in this chapter, it was noted that statisticians tend to group into two distinct categories, depending upon whether their philosophies regarding statistical inference are classical or Bayesian. It has been noted that classical statisticians avoid use of subjective probabilities. Another important difference in these two points of view has to do with the interpretation of probability and other parameters of statistical analysis. In the most general terms, classical statisticians interpret probability strictly in the sense of relative frequency. On the other hand, Bayesians view probability as a measure of rational degree of belief. For this reason, classical statisticians cannot evaluate probability until there is knowledge available on the structure of event space, while Bayesians are free to speculate on, or to postulate a priori probabilities in the absence of concrete factual evidence. Since these interpretations of probability are quite different, it is clear that there is room for controversy. Furthermore, the two different approaches to statistical inference may very well lead to distinct results in applications.

Often the same formulas are used by classical statisticians and Bayesians, since these result from rigorous mathematical theorems that are accepted by all statisticians. The differences lie in interpretation of the parameters used in these formulas. This state of affairs is clearly demonstrated in Table 8.3, where Bayes' Rule is considered from these two distinct perspectives.

The risks associated with use of subjective information in various processes of statistical inference are amply illustrated above, but they generally seem to be unavoidable in the field of nuclear data. The sentiment that nuclear-data evaluation is part art and part science has been expressed often in this community. I tend to agree with this contention. Hopefully, the discussions in the present chapter will help to provide the reader with some insight into which aspects of this discipline constitute "art" and which constitute "science". The reader will have attained a considerable level of sophistication if he understands these distinctions and can function comfortably in spite of the apparently unavoidable ambiguities of this discipline.

Table 8.3 Comparison of classical and Bayesian interpretations of the parameters of Bayes' Rule.

$$\text{Bayes' Rule: } P(A_k/B) = P(B/A_k)P(A_k) / \left[\sum_{i=1}^n P(B/A_i)P(A_i) \right]^a$$

Parameter	Classical Interpretation ^b	Bayesian Interpretation ^b
A_k	An event	A cause or hypothesis
$P(A_k)$	Relative frequency with which event A_k occurs (conventional probability)	A priori measure of the rational degree of belief that A_k is the correct cause or hypothesis)
$P(B/A_k)$	Relative frequency of occurrence of B given that A_k has occurred (conditional probability)	Likelihood that the cause or hypothesis A_k could have produced occurrence of B
$P(A_k/B)$	Relative frequency of occurrence of A_k given that B has occurred (conditional probability)	A posteriori measure of the rational degree of belief that A_k is the correct cause, or hypothesis, after knowledge that B has occurred is available

^a Eq. 5.9.

^b Both interpretations assume that random sampling is involved in an event space which exhausts the possibilities for the problem at hand.

ACKNOWLEDGEMENTS

I am indebted to the many people who have sent me copies of their papers, have made suggestions, and have brought to my attention important material which I might have otherwise overlooked. The insight provided by Dr. Francis Perey, through discussion and reference to his published papers over the past several years, has been of great value to me. The encouragement provided by Dr. Alan B. Smith and Dr. Andre Michaudon during the course of this endeavor has been greatly appreciated. Without the patient and persistent efforts of Marci Ambats and Phyllis Michaels, over a three-year period and under difficult circumstances, it would have been impossible to type, edit, and assemble the manuscript of this long report for reproduction. Their indispensable services were rendered in a manner truly above and beyond the normal call of duty.

This work has been supported by the U.S. Department of Energy, Energy Research Programs, under Contract W-31-109-Eng-38.

APPENDIX I: Notation

The notation used in this report is indicated below in order of appearance in the text. Chapters and page numbers are indicated for the convenience of the reader. Notational details which are of such common and standard usage that there is little possibility for confusion are not cited.

An attempt is made to maintain a reasonable degree of consistency in the use of notation throughout this report; however, some departures are inevitable and they should be clear from the context. In such instances, it is my judgement that the readers for whom this work is intended are not likely to be confused as a consequence of these particular inconsistencies.

Chapter 1: Introduction

No notation is introduced.

Chapter 2: Events and Event Spaces

C_{nk}	Pg. (10):	Binomial coefficient.
e, E	(12):	Events. In general, e denotes an elementary event while E denotes the entire event space. Other alphabetical letters (including Greek) are also used, generally to designate compound events.
(e_1, e_2, \dots)	(12):	A collection of elementary events.
$A \in E$	(13):	Event A is contained by event E (a subset).
$A_1 \notin A_2$	(13):	Event A_1 is not contained by event A_2 .
ϕ	(13):	Null event or empty event space.
$A_1 + A_2$	(15):	Union of events A_1 and A_2 .
$\sum_i A_i$	(15):	Union of collection of events typified by A_i .
$A_1 - A_2$	(15):	Difference or relative complement of events A_1 and A_2 .
$A_1 \times A_2$	(15):	Product or intersection of events A_1 and A_2 .
$\prod_i A_i$	(17):	Product of a collection of events typified by A_i .
\bar{A}	(17):	Complement of Event A .

Z	(23):	Borel field.
z	(24):	Element of a Borel field.
W	(30):	Often used to designate the number of compound events corresponding to a particular counting rule.
$C_{n;k_1,k_2,\dots,k_r}$	(36):	Multinomial coefficient.

Chapter 3: Random Variables and Their Functions

X, x	(47):	Random variable. Upper case alphabetical letters (including Greek) are usually used to designate the variable, while lower case symbols denote a specific numerical value which that variable assumes. Often, lower case symbols are used to designate either the variable or values which it assumes. In the framework of function theory, X designates a function, events e are operated upon by X to yield values x according to the relationship $x = X(e)$.
\bar{F}, \bar{f}	(49):	Vector arrays of random variables (functions) (F_1, F_2, \dots, F_n) and corresponding values (f_1, f_2, \dots, f_n) , where $f_i = F_i(e)$.
f^{-1}	(51):	Inverse function of f.
f(x)	(54):	Usually designates a density function.
F(z)	(54):	Usually designates a distribution function. F and f are related by $F(z) = \sum_{x < z} f(x)$ or $\int_{-\infty}^z dx f(x)$. Sometimes the term "distribution" function is used to designate a density function.
$f_{12\dots n}(x_1, x_2, \dots, x_n)$	(54):	Joint density function of n random variables.
$F_{12\dots n}(z_1, z_2, \dots, z_n)$	(54):	Joint distribution function of n random variables.

Chapter 4: Basic Concepts of Probability

P, P(e)	(59):	Probability, e.g., P(e) is the probability of event e. Sometimes, upper case "P" denotes probability density and lower case "p" signifies probability.
---------	-------	--

$N, N(A)$	(66):	Actual number of events, e.g., $N(A)$ denotes number of elementary events associated with compound event A . See "W", Chap. 2.
$M, M(A)$	(72):	Measure of an event set, e.g., $M(A)$ measures the "size" of A . Used instead of "N" when space is infinite.
$\langle k \rangle$	(74):	Expected value of k with respect to probabilities $P(k)$. $\langle k \rangle$ is the weighted average $\sum kP(k)$, inclusive of all possible k , such that $\sum P(k) = 1$.

Chapter 5: Conditional Probability and Independence

$P(B/A)$	(79):	Conditional probability of event B , given that event A has occurred.
----------	-------	---

Chapter 6: Probability Distributions

$P(x < \alpha), P(\alpha)$	(94):	Probability distribution.
$p, p(x)$	(94):	Probability density function.
$I(\alpha, \beta)$	(95):	An interval, namely, all x such that $\alpha \leq x \leq \beta$.
$P(\alpha; \beta)$	(95):	$P(\beta) - P(\alpha)$.
$p_{12\dots n}(x_1, \dots, x_n)$	(96):	Joint probability density function.
$P_{12\dots n}(\alpha_1, \dots, \alpha_n)$	(96):	Joint probability.
$P_{12\dots n}(A)$	(96):	Joint probability that $\bar{x} \in A$.
$p(x_2/x_1)$	(98):	Conditional probability density function. $p(x_2/x_1) = p_{12}(x_1, x_2)/p_1(x_1)$ if $p_1(x_1) > 0$.
$p_1(x_1)$	(98):	Marginal distribution of x_1 . For example, $p_1(x_1) = \int_{-\infty}^{\infty} dx_2 p_{12}(x_1, x_2)$.
J:		Jacobian or spin (see context).
det [...]	(100):	Determinant of the array in brackets.

Chapter 7: Some Important Probability Laws Governing the Frequency of Occurrence of Random Events.

- $p(n,p;k)$ (107): Binomial probability.
- $p(n;p_1, \dots, p_r; k_1, \dots, k_r)$ (108): Multinomial probability.
- σ : Cross section or standard deviation (see context).
- $p(\lambda;k)$ (122): Poisson probability.
- λ (121): Often used to designate mean value for Poisson or normal probability.
- $p(\sigma, \lambda; x)$ (133): Normal (Gaussian) probability.
- $f(x) = p(1,0;0)$ (134): Standard normal (Gaussian) probability function.
- \bar{V} (150): A matrix. Usually "V" designates a covariance matrix.
- $p(\bar{V}, \bar{\lambda}; \bar{x})$ (150): Multivariate Normal (Gaussian) probability.
- \bar{V}^+ (150): Transpose of matrix \bar{V} .
- \bar{V}^{-1} (151): Inverse of matrix \bar{V} .
- $\bar{A} \bar{B}, \bar{A} \cdot \bar{B}, \bar{A} \bar{x}, \bar{A} \cdot \bar{x}, \bar{y}^+ \bar{A} \bar{x}$: Examples of matrix products.
- V_{ij} (150): Element of matrix \bar{V} located in the i^{th} row and j^{th} column.
- \bar{C} (150): Usually designates a correlation matrix.
- Q_n (151): Quadratic form of order n .

Chapter 8: Bayes' Theorem and Applications

- $P_n(A_k/B_m)$ (167): Bayesian conditional (a posteriori) probability that A_k is the correct hypothesis given that B_m has been observed at the n^{th} iterative application of Bayes' Rule.

$P_{m-1}(A_k)$ (167): Bayesian a priori probability that A_k is the correct hypothesis just prior to m^{th} observation which produces B_m .

$P(\langle k \rangle \neq i)$ (170): Probability that $\langle k \rangle \neq i$, i.e., $1 - P(\langle k \rangle = i)$.

APPENDIX II: Important Formulas

Chapter 1: Introduction

No formulas

Chapter 2: Events and Event Spaces

[de Morgan Laws]:

$$\text{Eqn. (2.10) Pg. (25)} \quad \overline{(A + B)} = \bar{A} \times \bar{B}$$

$$(2.11) \quad (25) \quad \overline{(A \times B)} = \bar{A} + \bar{B}$$

[Binomial Coefficient]:

$$(2.21) \quad (32) \quad C_{nk} = n!/[k!(n-k)!]$$

[Stirling's Approximations]:

$$(2.22) \quad (32) \quad k! = (2\pi k)^{1/2} k^k e^{-k}$$

$$(2.23) \quad (33) \quad \ell n(k!) = (1/2)\ell n(2\pi k) + k\ell n k - k$$

[Binomial Theorem]:

$$(2.25) \quad (33) \quad (a_1 + a_2)^n = \sum_{k=0}^n C_{nk} a_1^k a_2^{n-k}$$

$$(2.30) \quad (36) \quad C_{n;k_1 k_2 \dots k_r} = n!/[k_1! k_2! \dots k_r!]$$

[Multinomial Theorem]:

$$(2.31) \quad (36)$$

$$\left(\sum_{j=1}^r a_j \right)^n = \sum_{k_1} \sum_{k_2} \dots \sum_{k_r} C_{n;k_1 k_2 \dots k_r} a_1^{k_1} a_2^{k_2} \dots a_r^{k_r}$$

for all k_j such that $k_1 + k_2 + \dots + k_r = n$.

Chapter 3: Random Variables and Their Functions

[Relationship between distribution function $F_{12\dots n}$ and joint density function $f_{12\dots n}$]:

$$(3.3) \quad (54)$$

$$F_{12\dots n}(z_1, \dots, z_n) = \int_{-\infty}^{z_1} dx_1 \int_{-\infty}^{z_2} dx_2 \dots \int_{-\infty}^{z_n} dx_n f_{12\dots n}(x_1, \dots, x_n)$$

Chapter 4: Basic Concepts of Probability

[Poincare Law of Probability Addition]:

$$(4.6) \quad (68) \quad P(A + B) = P(A) + P(B) - P(A \times B).$$

Chapter 5: Conditional Probability and Independence

$$(5.2) \quad (79) \quad P(B/A) = P(B \times A)/P(A) \text{ if } P(A) > 0.$$

$$(5.3) \quad (79) \quad P(B/A)P(A) = P(A/B)P(B) \text{ if } P(A), P(B) > 0.$$

[Chain Rule]:

$$(5.5) \quad (86) \quad P(A \times B \times C) = P(A)P(B/A)P(C/A \times B).$$

[Poincare Law of Addition Probability]:

$$(5.7) \quad (87) \quad P(B + C/A) = P(B/A) + P(C/A) - P(B \times C/A).$$

[Rule of Total Probability]:

$$(5.8) \quad (87) \quad P(B) = \sum_{i=1}^n P(B/A_i)P(A_i).$$

[Bayes' Theorem]:

$$(5.9) \quad (89) \quad P(A_k/B) = P(B/A_k)P(A_k) / \left[\sum_{i=1}^n P(B/A_i)P(A_i) \right].$$

Chapter 6: Probability Distributions

[Probability Density Transformation Law]:

$$(6.17) \quad (102) \quad p_y(\bar{y}) = p_x(\bar{x})/|J| \text{ where}$$

$$J = \det \begin{bmatrix} \partial y_1 / \partial x_1 & \dots & \partial y_1 / \partial x_n \\ \vdots & & \vdots \\ \partial y_n / \partial x_1 & \dots & \partial y_n / \partial x_n \end{bmatrix}$$

[Pearson's Formula]:

$$(6.18) \quad (103) \quad [1/p(x)] \{d[p(x)]/dx\} = (d - x)/(a + bx + cx^2).$$

Chapter 7: Some Important Probability Laws Governing the Frequency of Occurrence of Random Events

[Binomial Probability Law]:

$$(7.3) \quad (107) \quad p(n, p; k) = C_{nk} p^k (1 - p)^{n - k}$$

[Multinomial Probability Law]:

$$(7.4) \quad (108) \quad p(n; p_1, p_2, \dots, p_r; k_1, k_2, \dots, k_r) = C_{n; k_1 k_2 \dots k_r} p_1^{k_1} p_2^{k_2} \dots p_r^{k_r}$$

[Poisson Probability Law]:

$$(7.8) \quad (122) \quad p(\lambda; k) = \lambda^k e^{-\lambda} / k! \text{ for } \lambda > 0, k = 0, \infty.$$

[Interval Distribution]:

$$(7.11) \quad (130) \quad dP = R e^{-Rt} dt.$$

[Normal Probability Law]:

$$(7.12) \quad (133) \quad p(\sigma, \lambda; x) = \exp [-(x - \lambda)^2 / 2\sigma^2] / (2\pi\sigma^2)^{1/2}.$$

[Standard Normal Distribution Function]:

$$(7.13) \quad (134) \quad f(x) = p(1, 0; x) = \exp(-x^2/2) / (2\pi)^{1/2}.$$

[Multivariate Normal Probability Law]:

$$(7.16) \quad (150) \quad p(\bar{V}, \bar{\lambda}; \bar{x}) = \exp [-(1/2)(\bar{x} - \bar{\lambda})^T \bar{V}^{-1}(\bar{x} - \bar{\lambda})] / \{(2\pi)^{n/2} [\det(\bar{V})]^{1/2}\}.$$

Relationship between correlation matrix \bar{C} and covariance matrix \bar{V} :

$$(7.17) \quad (151) \quad C_{ij} = V_{ij} / (V_{ii} V_{jj})^{1/2} \text{ for } i, j = 1, n.$$

Chapter 8. Bayes' Theorem and Applications

[Laplace's General Rule of Succession]:

$$(8.1) \quad (164) \quad P(B/A) = (n + 1) / (n + n' + 1).$$

[Laplace's Special Rule of Succession]:

$$(8.2) \quad (164) \quad P(B/A) = (n + 1) / (n + 2).$$

APPENDIX III: A Guide to Additional Reading for Selected Topics

Chapter 1: Introduction

- **Systematic error**
You 61: pp. 56-62
Per 82: pp. 8-19

Chapter 2: Events and Event Spaces

- **Properties of event spaces**
Hau 57: pp. 11-27
Par 60: pp. 12-16, 22
Fre 62: pp. 10-13
Fis 63: pp. 5-11
Men 67: pp. 60, 62-65
Bur 68: pp. 77-82
Mah 68: pp. 15-46
Abb 69: pp. 21-58
Zeh 70: pp. 10-16, 37-41
KM 76: pp. 1-33
- **Algebra of event spaces (Boolean algebra)**
Par 60: pp. 12-16
Fre 62: pp. 5, 15
Mah 68: pp. 47-62
Abb 69: pp. 178-217
Ash 70: pp. 3-9
Zeh 70: pp. 16-17
KM 76: pp. 33-42
- **Borel fields**
Par 60: p. 150
Fis 63: p. 10
Zeh 70: pp. 39-40, 90
- **Sampling and combinatorial analysis**
Ney 50: pp. 34-39
Bro 60: pp. 16-19
Par 60: pp. 32-42
Fre 62: pp. 17-27
Men 67: pp. 54-59
Bur 68: pp. 68-76
Mah 68: pp. 69-78
BeV 69: pp. 28-30
Mar 71: p. 9
- **Maxwell-Boltzman, Fermi-Dirac and Bose-Einstein laws of statistical physics**
Fel 50: pp. 53-54
Par 60: pp. 70-71
Ash 70: pp. 20-21

- Stirling's formula
 - Ney 50: pp. 207-216
 - Par 60: p. 163
- Binomial and multinomial theorems
 - Par 60: pp. 37-41
 - Fre 62: pp. 24-26
 - Mah 68: pp. 77-78
 - Zeh 70: pp. 127-128, 252

Chapter 3: Random Variables and Their Functions

- Random phenomena
 - Ney 50: pp. 9
 - Par 60: pp. 1-8
- Random variables
 - Ney 50: pp. 166-173
 - Bro 60: pp. 11-16
 - Par 60: pp. 148-157, 268-275
 - Fre 62: pp. 61-65
 - Fis 63: pp. 29-34, 40-45
 - Men 69: pp. 70-71, 80-90
 - Zeh 70: pp. 48-88
- Random-variable functions
 - Par 60: pp. 151-157
 - Fre 62: pp. 61-65, 117-125
 - Fis 63: pp. 31-40
 - Ash 70: pp. 70-75

Chapter 4: Basic Concepts of Probability

- Notion of probability
 - Fel 50: pp. 4-7
 - Par 60: pp. 1-7
 - Fre 62: pp. 1-3, 32-34
 - Fis 63: pp. vii, 3-5
 - Bas 66: pp. 1, 5-6, 16-18
 - Bur 68: pp. 61-66
 - Zeh 70: pp. 20-22
 - Mar 71: p. 5
- Axioms and formal properties of probability
 - Ney 50: pp. 16-20, 44-55
 - Par 60: pp. 17-21, 25-28, 176-182
 - Fre 62: pp. 34-40
 - Fis 63: pp. 11-16
 - Bas 66: pp. 10-11
 - Tuc 67: pp. 5-8
 - Bra 70: pp. 6-7
 - Cra 70: pp. 1-17
 - Zeh 70: pp. 22-27

- Calculation of probability
 - Fel 50: pp. 17-40
 - Par 60: pp. 32-50, 76-85, 116-117
 - Bas 66: pp. 11-14
 - Men 69: p. 59
 - Ash 70: pp. 3-9, 15-23
 - Mar 71: pp. 6-7

Chapter 5: Conditional Probability and Independence

- Concept of conditional probability
 - Zeh 70: pp. 42-44
- Formal properties of conditional probability
 - Fel 50: pp. 78-81
 - Par 60: pp. 60-65
 - Fis 63: pp. 18-22
 - Mar 71: pp. 8
- Independence
 - Fel 50: pp. 85-91
 - Ney 50: pp. 55-65
 - Par 60: pp. 87-100, 113-114, 128-129
 - Fre 62: pp. 47-51
 - Fis 63: pp. 24-28
 - Zeh 70: pp. 48-52

Chapter 6: Probability Distributions

- Univariate probability density and distribution functions
 - Par 60: pp. 151-157
 - Fre 62: pp. 131
 - Bas 66: pp. 15-16
 - Tuc 67: pp. 15-17
 - Ash 70: p. 52
 - Bra 70: pp. 9-10
 - Mar 71: pp. 22-24
- Multivariate probability density and distribution functions
 - Par 60: pp. 285-292
 - Fre 62: pp. 81-85
 - Bra 70: pp. 19-21
- Transformations
 - Bra 70: pp. 27-31

Chapter 7: Some Important Probability Laws Governing the Frequency of Occurrence of Random Events

- Binomial and multinomial probability
 - Ney 50: pp. 179-182
 - Eva 55: pp. 747

- Fre 62: pp. 66-70, 85-87
 - Fis 63: pp. 130-134, 179-180
 - Men 69: pp. 101-102
 - Zeh 70: pp. 249-253
 - Mar 71: pp. 36-39
- Laplace's model of error and Pascal's triangle
 - Bra 70: pp. 53-54
 - Poisson probability
 - Eva 55: p. 750
 - Par 60: pp. 251-258
 - Fis 63: pp. 140-145
 - Bra 70: pp. 45-49
 - Relationship between binomial and Poisson probability
 - Ney 50: pp. 212-213
 - Bas 66: pp. 131-133
 - Applications of Poisson probability
 - Eva 55: pp. 753-754, 766-768, 785-817
 - Univariate normal probability
 - Eva 55: pp. 748-749
 - Bee 58: pp. 16-20
 - Bro 60: p. 33
 - Par 60: pp. 188-191, 237-238
 - Fre 62: pp. 128-131
 - Fis 63: pp. 147-151
 - Bas 66: pp. 87-90
 - Tuc 67: p. 17
 - Bur 68: pp. 153-159
 - Bev 69: pp. 43-45
 - Zeh 70: pp. 153-161
 - Mar 71: pp. 26-30
 - Multivariate normal probability
 - Bro 60: p. 150
 - Bas 66: pp. 102-109
 - Bur 68: pp. 289-293
 - Bra 70: pp. 61-67
 - Zeh 70: pp. 265-278
 - Mar 71: pp. 30-35
 - Central limit theorem
 - Par 60: p. 238
 - Men 67: pp. 124-129
 - Hershel's model of error
 - Bra 70: pp. 68-71

- The normal distribution as a limit to the binomial and Poisson distributions
 - Fel 50: pp. 133-143
 - Ney 50: pp. 215-234
 - Bro 60: pp. 105-112, 150
 - Par 60: pp. 239-249
 - Bas 66: pp. 131-133
 - Men 67: pp. 131-132
 - Bur 68: pp. 160-163
 - Mar 71: pp. 37-38, 41

Chapter 8: Bayes' Theorem and Application

- Philosophical implications of Bayes' theorem and Bayesian learning
 - Coo 25: pp. 87-100
 - Fel 50: p 85
 - Par 60: pp. 119-120
 - Fre 62: pp. 57-58
 - Fis 63: pp. 22-23
 - Bas 66: pp. 35-37
 - Bur 68: pp. 87-90, 357-358
 - Ash 70: pp. 35-38
 - Bra 70: p 95
 - Zeh 70: pp. 44-47, 416-419
 - Mar 71: pp. 5-6, 9-11
 - Pee 82: pp. 47-51
 - Pro 85: p. 236
- Laplace's Rule of Succession and the Maximum-Likelihood Principle
 - Par 60: pp. 122-124
 - Fis 63: pp. 440, 484-487
 - Zeh 70: pp. 353-370

BIBLIOGRAPHY

- Abb 69 James C. Abbott, Sets, Lattices, and Boolean Algebras, Allyn and Bacon, Inc., Boston (1969).
- Ash 70 Robert B. Ash, Basic Probability Theory, John Wiley and Sons, Inc., New York (1970).
- Bas 66 J. Bass, Elements of Probability Theory, Academic Press, New York (1966).
- Bee 58 Yardley Beers, Theory of Error, Addison-Wesley Publishing Co., Inc., Reading, MA (1958).
- Bev 69 Phillip R. Bevington, Data Reduction and Error Analysis for the Physical Sciences, McGraw-Hill Book Co., New York (1969).
- Bra 70 Siegmund Brandt, Statistical and Computational Methods in Data Analysis, North-Holland Publishing Co., Amsterdam (1970).
- Bro 60 K. A. Brownlee, Statistical Theory and Methodology in Science and Engineering, John Wiley and Sons, Inc., New York (1960).
- Bur 68 Robert L. Burford, Statistics: A Computer Approach, Charles E. Merrill Publishing Co., Columbus, OH (1968).
- Coo 25 Julian Lowell Coolidge, An Introduction to Mathematical Probability, Dover Publications, Inc., New York (1925).
- Cra 70 Harold Cramer, Random Variables and Probability Distributions, Cambridge University Press, Cambridge, England (1970).
- Eva 55 Robley D. Evans, The Atomic Nucleus, McGraw-Hill Book Co., New York (1955).
- Fel 50 William Feller, An Introduction to Probability Theory and its Applications, John Wiley and Sons, Inc., New York (1950).
- Fis 63 Marek Fisz, Probability Theory and Mathematical Statistics, John Wiley and Sons, Inc., New York (1963).
- Fre 62 John E. Freund, Mathematical Statistics, Prentice-Hall, Inc., Englewood Cliffs, NJ (1962).
- Hau 57 Felix Hausdorff, Set Theory, Chelsea Publishing Co., New York (1957).

- Hil 52 F. B. Hildebrand, Methods of Applied Mathematics, Prentice-Hall, Inc., Englewood Cliffs, NJ (1952).
- KM 76 K. Kuratowski and A. Mostowski, Set Theory, with an Introduction to Descriptive Set Theory, North-Holland Publishing Co., Amsterdam (1976).
- Mah 68 Lawrence P. Maher, Jr., Finite Sets: Theory, Counting, and Applications, Charles E. Merrill Publishing Co., Columbus, OH (1968).
- Man 81 W. Mannhart, "A Small Guide to Generating Covariances of Experimental Data," PTB-FMRB-84, Physikalisch-Technische Bundesanstalt-Braunschweig, FRG (1981).
- Mar 71 B. R. Martin, Statistics for Physicists, Academic Press, London (1971).
- Men 67 William Mendenhall, Introduction to Probability and Statistics, Wadsworth Publishing Co., Inc., Belmont, CA (1967).
- Mor 64 Phillip M. Morse, Thermal Physics, W. A. Benjamin, Inc., New York (1964).
- Ney 50 J. Neyman, First Course in Probability and Statistics, Henry Holt and Co., New York (1950).
- Par 60 Emanuel Parzen, Modern Probability Theory and its Applications, John Wiley and Sons, Inc., New York (1960).
- Pee 82 R. W. Peele, "Uncertainty in the Nuclear Data Used for Reactor Calculations," Advances in Nuclear Science, Vol. 14, ed. Lewins and Becker, Plenum Publishing Corp., New York (1982).
- Per 82 F. G. Perey, Application of Group Theory to Data Reduction, ORNL-5908, Oak Ridge National Laboratory (1982).
- Poe 66 W. P. Poenitz, Z. Physics 197, (1966) 262; KFK-523, Kernforschungszentrum Karlsruhe, Fed. Rep. of Germany (1966).
- Pri 63 D. J. DeSola Price, Little Science, Big Science, Columbia University Press, New York (1963).
- Pro 85 Harrison Bertrand Prosper, "A Bayesian Analysis of Experiments with Small Numbers of Events," Nuclear Instruments and Methods in Physics Research A241, 236 (1985).
- Smi 81 Donald L. Smith, "Covariance Matrices and Applications to the Field of Nuclear Data," ANL/NDM-62, Argonne National Laboratory (1981).

- Tuc 67 Howard G. Tucker, A Graduate Course in Probability, Academic Press, New York (1967).
- Tul 85 Jagdish K. Tuli, Nuclear Wallet Cards, National Nuclear Data Center, Brookhaven National Laboratory, (1985).
- You 61 W. J. Youdin, "Systematic Errors in Physical Constants," *Physics Today*, pp. 56-62 (September 1961).
- Zeh 70 Peter W. Zehna, Probability Distributions and Statistics, Allyn and Bacon, Inc., Boston (1970).

INDEX

A

- A priori probability
 - definition of (62, 79)
 - subjective estimation of in Bayesian inference (163)
- A posteriori probability
 - as an amalgamation of theory and experiment (163)
 - definition of (62, 79)
- Addition theorem
 - and normal probability (134)
 - and Poisson probability (131)
 - importance of in counting experiments (131)
- Algebraic manipulations
 - differences between events and numbers (24)
- Algebra of event spaces or sets (15)
- Algorithms
 - event counting (29, 30)
- Arrangements of events (29)
- Associative laws (22)
- Axioms of probability
 - equivalence to intuitive observations (67)
 - failure to provide procedures for calculation (66)
 - foundations of the theory (65,66)

B

- Background
 - educational level required of the reader (3)
- Backward reasoning
 - Bayesian basis for (162)
- Bayes, T., English clergyman (89)
- Bayesian inference
 - as a procedure for accepting or rejecting hypotheses (163)
 - relation to experimental measurements (163)
 - relation to mathematical theories (163)
- Bayesian learning
 - effect on of attaching too large a confidence level to an assumed hypothesis (176)
 - example to demonstrate process of (169-176)
 - is there an element of "luck" involved in it? (176)
 - overcoming effects of an erroneous observation (176)
- Bayesian statistics
 - philosophy of (179)
 - relation to inference (89,91)
- Bayes' postulate
 - definition of (91)
 - role of in scientific applications (91)
- Bayes' rule
 - application of in refinement of knowledge of a nuclear cross section (166, 169)
 - as a means for merging speculative and observed information (163)
 - as an algorithm for combining old and new data (166)

- as a substitute term for Bayes' theorem (155)
- examples to demonstrate (155-162)
- iterative use of (166)
- role of in correcting counting data for background (159-162)
- speculation on input parameters for the application of (158, 159)
- Bayes' theorem
 - philosophical concepts of (89, 91)
 - relation to rule of total probability (89)
 - statement of and proof (89)
- Bell-shaped curve (132)
- Bernoulli experiment (107)
- Bernoulli's law of large numbers
 - concept of (110)
 - relation to Poisson probability law (122)
- Bernoulli trials
 - as a basis for several important probability laws (84)
 - definition of independent (84)
- Binary model for certain random phenomena (47)
- Binomial coefficient (10, 32)
- Binomial probability law
 - and the binomial coefficient (107)
 - as descriptive of the outcome of Bernoulli trials (105, 106)
 - common form for (107)
 - conceptual derivation of (106, 107)
 - example to compare to normal and Poisson probability (146)
 - examples to demonstrate some features of (108-110)
 - mass function for (107)
 - multinomial format for mass function of (107)
 - rarity of use of in practice (119)
 - relation to frequency of occurrence of events (105)
 - when to use (141)
- Binomial theorem (33)
- Bivariate normal probability
 - density function for (151)
 - horizontal elliptical profiles for (152, 154)
 - vertical Gaussian profiles for (152)
- Boltzmann's constant (75, 139)
- Boolean algebra (15)
- Boole's inequality
 - concept of (68)
 - generalized form of (69)
- Borel field
 - concept of (23)
 - as applicable to subspaces (24)
 - example to demonstrate role of (28, 29)
 - properties of a (23, 24)
 - relation to probability (61)
 - size of a (24, 28, 29)
- Bose-Einstein statistics
 - an example to demonstrate use of (37)
 - concept of (36)
- Bosons
 - nature of (31, 32, 36)
- Bounded random-variable functions (50)

Bound states
quantized (49)

C

Cardinality of event spaces (49)
Causality and the order of occurrence of events (10-12)
Central limit theorem
 example to demonstrate use of (147-150)
 formal statement of (145, 147)
 general information on (145)
 relation to Laplace law of error (110)
 relation to normal probability (135)
 simple interpretation of (145, 147)
Chance phenomena (43)
Chaos
 concept of (43)
Chaotic phenomena (43, 44)
Certain event (13)
Classical statistics, philosophy of (179)
Coincidence measurements
 applicability of Poisson statistics to analysis of (128, 129)
Combinatorial analysis
 basic principle of (30)
Complement
 event (17, 19)
Complementation law (23)
Condensation
 relation to Bose-Einstein statistics (36)
Conditional probability
 an example to demonstrate use of (80, 83)
 application of fundamental axioms of probability to (85)
 calculation of (79)
 chain rule for (86)
 concept of (78, 79)
 density function for (98)
 formula relating it to unconditional probability (79, 98)
 functional form for (79)
 generalized chain rule for (86)
 relation to marginal probability (98)
 some basic theorems of (85-87)
Confidence
 aspects of in connection with application of Bayes' rule (166)
 measure of and normal probability (135)
Containment
 and the relationship of two events (13, 14)
Content of this report (xi, xii, 3, 4)
Continuous random variables (48, 94)
Continuum space of events (48, 49)
Convergence
 and the Bayesian learning process (170)
Correlated random variables
 negatively (154)
 positively (154)

- Correlation of two random variables
 - geometric interpretation of (152-154)
- Correlation matrix
 - relation to covariance matrix (150-151)
- Correlation parameter
 - role of in bivariate normal distribution (151-154)
- Countable random variables (48)
- Counting laws
 - example to demonstrate effect on quantum system total energy (37, 39-42)
 - summary of those commonly used in statistical physics (38)
- Counting of events
 - concepts (29)
 - importance in science (29)
- Count rates and Poisson probability (124)
- Covariance
 - definition of (150)
- Covariance ellipse
 - family for bivariate normal distribution (152-154)
- Covariance matrix
 - role of in multivariate normal distribution (150-151)
- Cramer, H., Swedish mathematician (134)
- Cumulative probability (94)

D

- Definitions
 - how they are identified in this report (3)
- de Mere, C., French mathematician (59)
- DeMoivre, A., French mathematician (132)
- DeMoivre-Laplace theorem
 - as a link between binomial and normal probability (140, 142)
 - example to demonstrate (142-145)
 - formal statement of (142)
 - simple interpretation of (142)
- de Morgan laws
 - concept of (25-26)
 - example to demonstrate (26-28)
 - generalized (25-26)
- Density functions
 - constant (53)
 - non-negative character of (53)
 - random variables as the arguments of (53)
 - relation to random variables (53)
- Denumerable random variables (48)
- Depletion of event-space elements (5)
- Detector deadtime
 - applicability of Poisson probability to the analysis of (127)
- Determinant of a covariance matrix (151)
- Deterministic regularity (43)
- Difference of event sets (15)
- Differential relationship between probability and probability density (95, 96)
- Discrete random variables (48)

- Disjoint events
 - concept of (17)
 - pairwise (17)
- Distinguishable events (29)
- Distribution function
 - continuous (54, 55)
 - definition of (53)
 - discrete (54, 55)
 - non-negative character of (54)
 - term often used as a substitute for density function (54)
- Distributive laws
 - concept of (22)
 - generalized (26)

E

- Electromagnetic multipole transitions (111)
- Ellipses of covariance for bivariate normal distribution (152-154)
- Energy-group structure (103)
- Equal-likelihood postulate
 - application in quantum theory (61)
 - as a foundation for probability calculation (66)
 - definition of (59)
 - in relation to the behavior of natural phenomena (61)
- Equality of events (13)
- Errors
 - equivalent term for uncertainties (x)
 - relation to probability and statistics (x)
- Estimators
 - role of in statistics (150)
- Events
 - compound (7, 12)
 - concept of (5, 6)
 - elementary (6)
 - mutually exclusive (10)
 - nested sequences of (13)
 - notion of subspaces of compound (12)
 - probabilizable nature of (71)
 - random (54)
 - simple (6)
 - subspace of (7)
- Event space
 - augmented (6)
 - closure of (22)
 - complete (12)
 - concept of (5-6)
 - continuous (7)
 - denumerable (12)
 - discrete (7)
 - finite (12)
 - nondenumerable (12)
 - nontrivial (24)
 - partitioning of (7)
 - size of an (10)

some basic theorems for (24, 25)
theory of and similarities to set theory (5, 6)
well-defined and stable (64)

Examples

purpose for inclusion in this report (3)

Expected value

notion of (74)

example to demonstrate concept (74, 75)

Exponential decay

derivation of (124)

law of (124)

role of in nuclear science (124)

F

Factorable

density functions (54)

distribution functions (54)

Factorability and marginal distributions (56)

Failure frequency

use of Poisson probability in the determination of (122)

Fermat, P., French mathematician (59)

Fermi-Dirac statistics (31, 32)

Fermions

nature of (31)

Finite difference methods

applicability in reactor physics (103, 104)

Finite random variables (48)

Flow

consideration of for gases in Maxwell-Boltzmann theory (138)

Formulas

important ones (Appendix II, 188-190)

where to find (3)

Frequency of occurrence

an example to demonstrate concept (57-59)

convergence of in large samples (59)

for infinite samples spaces (62)

inadequacy of as basis for probability in pure mathematics (62)

relation to probability (57)

relative (59)

role of in nuclear processes (105)

Function

as defined on event space (47)

Fundamental principle of applied probability theory (166)

G

Gauss, K., German mathematician (59, 132)

Gaussian distribution (132)

Gaussian probability

as an equivalent term for normal probability (105)

Generalized least-squares method

relation to Bayesian approach (176)

Group

- fluence (103)
- limits (103)
- representation of essentially continuous parameters (104)
- structure for a neutron-energy grid (103, 104)

H

- Herschell, J., English astronomer (139)
- Herschell's model of error
 - derivation of (139-140)
 - postulates of (139)
 - relation to normal distribution (139)
- Hybrid random variables (49)
- Hypotheses
 - need for complete range of in Bayesian statistics (162, 163)
 - role of in statistical inference (91)

I

- Identity laws (23)
- Image of an event space (50)
- Image space
 - role of in random-variable function mapping (51)
- Impossible event
 - definition of (13)
 - probability of (66)
- Independence
 - expressed in terms of probability density functions (99)
 - fully or en bloc (92)
 - mathematical condition for (92)
 - several basic theorems of (92)
 - stochastic (83)
- Independent events
 - mutually (91, 92)
 - pairwise (91, 92)
- Indicator functions (51)
- Indistinguishable events (29, 30)
- Inductive reasoning
 - building of knowledge from speculative origin through inclusion of new information gained by observation (162, 163)
 - relation to Bayes' Rule (162, 163)
- Inference
 - and Bayes' theorem (89, 91)
 - statistical (89, 91)
- Intersection of events (15-17)
- Interval distribution
 - derivation from Poisson probability law (130)
 - example to demonstrate (130, 131)
- Intervals
 - bounded (95)
 - non-intersecting (95)
 - notion of (95)
 - unbounded (95)

Inverse of a covariance matrix
condition for existence of (151)
role of in multivariate normal distribution (150, 151)

J

Jacobian of a transformation
concept of (100)
n-dimensional form (100, 102)

Joint
probability density function (54)
probability distribution function (96)
random variables (49)

Jump
discontinuities in probability density functions (94)
points (48)

K

k-tuples
applicability of probability axioms when formed by independent trials (92, 93)
formation without replacement (29, 30)
formation with replacement (29, 30)
probability of in relation to that for individual trials (83, 84)
role of events in (29, 30)

Kolmogorov, A., Russian mathematician (65)

L

Laplace, P., French mathematician (59, 132)

Laplace's general rule of succession
formal expression of (164)

Laplace's law of error
an example to demonstrate use of (111)
basis in binomial probability (110, 111)
derivation of (110, 111)
relation to Pascal's triangle (111)

Laplace's rule of succession
equivalence to Laplace's special rule of succession (164, 165)
example to demonstrate improper use of (165, 166)
example to demonstrate proper use of (165)
necessary conditions for validity of (164, 165)
potential for misuse (164, 165)

Laplace's special rule of succession
formal expression of (164)

Learning process
an iterative Bayesian algorithm for the (167)

Least-squares method
as a basis for nuclear-data evaluation (176)
relation to Bayes' theorem (91)

Levy, P., French mathematician (134)

Levy-Cramer theorem for normal probability (134)

Likelihood
 factor (91)
 relation to Bayes' theorem (91)
 theory must provide algorithm for calculation of (163)
Limitations of this report (4)
"Luck"
 does it exist in relation to Bayesian learning? (176)

M

Mapping
 one-to-one (51)
 role of for random-variable functions (51)
Marginal probability distribution (98)
Markov processes (93)
Mass function, concept of (53)
Maximum-likelihood principle
 as a method for inductive reasoning (163, 164)
 instances of consistency with Bayes' rule (164)
 risks associated with use of (163, 164)
 statement of (163)
Maxwell-Boltzmann
 conditions for applicability of M-B distribution (30, 31)
 factor (75)
 law of statistical physics (30 31)
 normal probability basis for M-B law of molecular velocity (138, 139)
Measure
 of an event set (71, 72)
 role of in definition of event probability (72)
 theory of (71, 72)
Measurement
 as an equivalent term for observation (168)
Median energy
 of a group (103)
Memory
 relation to independent and Markov trials (93)
Monotonic property of probability distributions (94)
Monte-Carlo method
 an example to demonstrate the (170)
 use of in applications (148)
Most-probable value
 notion of (74)
Multi-dimensional random variables (49)
Multinomial
 coefficient (36)
 theorem (36)
Multinomial probability law
 an example to demonstrate the use of (117-119)
 conceptual derivation of the (107, 108)
 mass function for (108)
 relation to frequency of occurrence (105)
Multivariate
 density function (53)
 probability distribution (96)

Multivariate normal probability
density function for (150, 151)
existence of (150, 151)
for independent random variables (151)
Mutually exclusive events
concept of (17, 19)
pairwise (64)

N

Neutron capture process
use of probability to describe (111, 113-117)
non-denumerable random variables (48-49)
Neutron transport
relation of probability to (103, 104)
Nonsingular
requirement for a covariance matrix to be (150, 151)
Normal curve (132-133)
Normalizability of density functions (56)
Normal probability law
as an approximation to binomial probability (140, 141)
as an approximation to Poisson probability (140, 141)
as a limiting approximation to more fundamental laws (135)
convenience of in use (135)
derivative for (133, 134)
example to compare to binomial and Poisson probability (146)
example to demonstrate use of (138, 139)
family of distributions for (134)
general features of (133, 134)
limits of applicability in realistic situations (135)
normalizability of (133)
parameters of (133)
origins of (132)
reasons for importance of in applied statistics (135)
relation to frequency of occurrence (105)
symmetry theorem for (133)
tangents to curve for (134)
univariate density function for (132, 133)
when to use (141)
why it often seems to be applicable (150)
Notation
major features of (Appendix I, 183-187)
use of in this report (3)
Nuclear-data research
accuracy assessment in (ix, x)
Bayesian nature of (155)
purpose for (ix)
quantitative nature of (ix)
social responsibility associated with (ix, x)
Nuclear phenomena
unique properties of (105)
Null event (13)
Numerical-valued random phenomena (47)

Q

- Objectives for the reader (2-3)
- Observations
 - averages of (74)
 - non-repeatability of and error (65)
- Occupancy
 - events (29, 30)
 - multiple (29, 30)
 - single (29, 30)
- One-sigma limit for normal distribution (135)
- Order
 - concept of for events (29, 30)
 - role of in defining event collections (29, 30)
- Ordered k-tuples
 - formed with replacement (30)
 - formed without replacement (30)
- Orthogonal transformation
 - use of in eliminating tilt of covariance ellipse (154)

P

- Partition
 - and Bayes' theorem (89)
 - binary (36)
 - event (20, 21)
 - multifold (36)
- Partition function
 - role of in statistical physics (75)
- Pascal, B., French mathematician (59)
- Pascal's triangle (33, 35)
- Pauli exclusion principle (31)
- Pearson, K., English mathematician (103)
- Pearson's differential equation for probability density
 - general form of (103)
 - parameterization of for normal probability (133)
- Permutation of events in k-tuple (30)
- Perturbation of finite sample spaces (62)
- Phase space
 - consideration of for Maxwell-Boltzmann theory (138)
- Poincare law of probability addition
 - concept of (68)
 - example to demonstrate (95)
 - for conditional probability (87)
 - generalized (69)
- Poisson, S., French mathematician (120)
- Poisson approximation (120)
- Poisson condition (121)
- Poisson probability law
 - a derivation for from binomial probability (120, 121)
 - an example to compare to binomial and normal probability (146)
 - an example to demonstrate use of (123)
 - as a limit to binomial law for rare events (120)
 - definition of (122)

- experimental determination of parameters for (123)
- importance of in analysis of nuclear processes (120, 124)
- interpretation of (122)
- relation to frequency of occurrence (105)
- satisfies the axioms of probability (122)
- skewed nature of mass function for (123)
- some theorems relating to the (123, 124)
- when to use it (141)
- Poisson transformation (121)
- Populations (5, 6)
- Positive definite matrix
 - requirement for in normal probability distribution (150)
- Postulated probability (62)
- Precision
 - of measurements and impact on Bayesian learning process (168)
- Presentation
 - manner of used in this report (xi)
- Probability
 - additivity axiom for denumerably infinite set of events (70)
 - additivity axiom in, for finite set of events (66)
 - calculation via sample-point method (73)
 - combinatorial methods in (73)
 - distribution function for (94)
 - effects of conditions on (78)
 - existence axiom (65)
 - function (61)
 - growth of importance of field of after World War II (59)
 - intuitive definition of (59, 61)
 - meaning of a law of (72)
 - nemesis of (72)
 - nondecreasing sequence of events (71)
 - nonincreasing sequence of events (71)
 - normalization axiom for (66)
 - of a sum of events forming a partition of event space (70)
 - practice of as an art (6, 179)
 - relation to errors and statistics (x, xi)
 - role of in mathematical science (5)
 - sum rule for pairwise-exclusive events (64)
- Probability density function
 - behavior under random-variable transformation (99-102)
 - cannot be interpreted as a true probability (94)
 - definition of (94)
 - factorability of joint in case of independence (99)
- Probability laws
 - comment on the usefulness of (105)
- Probability theory
 - avoidance of pitfalls in (66)
 - consistency with relative frequency concept (59)
 - developmental challenges in (65)
 - disputed areas in (65)
 - historical origins of (59)
 - need for minimum postulate basis of (62)
 - role of as a model for random phenomena (44, 59)
 - role of the postulates of (59)

- Samples
 - associated event space (5)
 - imperfections in process of selecting (63)
 - nature of (5)
 - replacement to avoid perturbation of finite event space (62)
- Scientific discoveries
 - frequency of multiple occurrences of (132)
- Selectivity
 - importance of in experimental design (176)
- Sequence of events
 - limit of a (20, 22)
 - nondecreasing (13, 15, 16)
 - nonincreasing (13, 15, 16)
- Sequential trials
 - probability of success in each trial (164)
- Set
 - equivalence of event space to (5)
 - universal (12)
- Sigma algebra (23)
- Simple events
 - uniqueness of (62)
- Single-variable events (47)
- Singularities
 - and random-variable functions (50)
 - unacceptability of in probability density functions (94)
- Speed versus velocity in molecular motion (138, 139)
- Spin selection rules (113, 114, 116)
- Spread
 - of observed values (74)
- Standard normal distribution
 - brief table of values for (136)
 - definition of (134)
 - some properties (134, 135)
- Statistical interpretations
 - difference between Bayesian and classical approach (179, 180)
- Statistical regularity (43)
- Statistical uncertainty
 - example to demonstrate meaning of (125)
- Statisticians
 - Bayesian (155)
 - classical (155)
- Statistics
 - relation to probability and errors (x, xi)
- Stirling's approximation (32, 33)
- Stochastic
 - phenomena (43)
 - role of normal probability in theory of behavior (132)
- Subjective information
 - risks of using as a basis for Bayesian inference (179)
- Success-to-failure ratio
 - relation to probability (58)
- Sum of events (15)

Procedures followed in this report (2, 3)
Product of events (15, 17, 18)

Q

Quadratic form
 role of in multivariate normal distribution (151)
Quantized energy states
 concept of (31)
Quantum statistics (32)

R

Radioactivity
 relation to Poisson probability (124)
Raikov, D., Russian mathematician (131)
Raikov's theorem for Poisson probability (131)
Random function (47)
Randomness
 relation to preconceived concept of (43)
 the notion of (43)
 an example to demonstrate difference between random and chaotic
 behavior (44-46)
Random-number generator
 use of in applications (148)
Random phenomena (43)
Random-variable functions
 on the nature of (50)
 properties of well-behaved (50, 51)
 relation to error propagation (51)
Random variables
 as a means to characterize random phenomena (45, 47)
 n-component definition of (49)
 relation to errors (48)
Random-variable transformations
 use of to widen scope of applicability of normal probability
 distribution (135)
Random vectors (49)
Readers
 for whom this report is intended (x, xi, 2)
Recursion formula for binomial coefficients (33)
References to other documents on probability (2)
Relative complement of events (15, 18)
Reliability
 of measurements and impact on Bayesian learning process (168)
Replacement of events (5)
Resolution
 example to demonstrate statistical nature of (125)
 relation to Poisson statistics (125)
Resource material for this report (2)
Rutherford, E., British physicist (124)

Superfluidity
 relation to Bose-Einstein statistics (36)
Sure event (13)
Symmetric matrix
 requirement for in multivariate normal distribution (150)
Symmetries
 importance of being able to visualize (42)
Systematic error
 limits of consideration in this report (4)

I

Theory
 equivalence to hypothesis (168)
Three-sigma rule
 for normal distribution (135)
 general significance of (135)
Total probability
 rule of (87)
 usefulness of rule of in applications (87)
Transformation
 example to demonstrate for probability (102)
 functions (51)
 information loss associated with (52)
 of random variables (51)
Trials
 dependent (93)
 independent (83)
 Markov (93)
 probabilities for dependent based on generalized chain rule (93)
 random (43)
Two-sigma limit
 from normal probability (135)

II

Uncertainty
 as a consequence of probability distribution (77)
 equivalent term for error (x , x_i)
 its relation to probability and statistics (x , x_i , 42)
Uncertainty analysis
 literature on (x)
Unconditional probability (79)
Uncountable random variables (48)
Uniformly well-behaved density function (54)
Union of events (15, 16)
Uniqueness of a complement (23)
Univariate probability distribution (94)
Unordered k -tuples
 formed with replacement (36)
 formed without replacement (32)

V

Variance

- example to demonstrate concept of (74)
- formal definition of (150)
- notion of (74)

Variance-covariance matrix

- role of in multivariate normal distribution (150)

Vector functions

- of a random variable (51)
- of a random vector (51)

Venn diagram (7)

Volume

- of n -dimensional event space (96)

W

Weighted average (74)