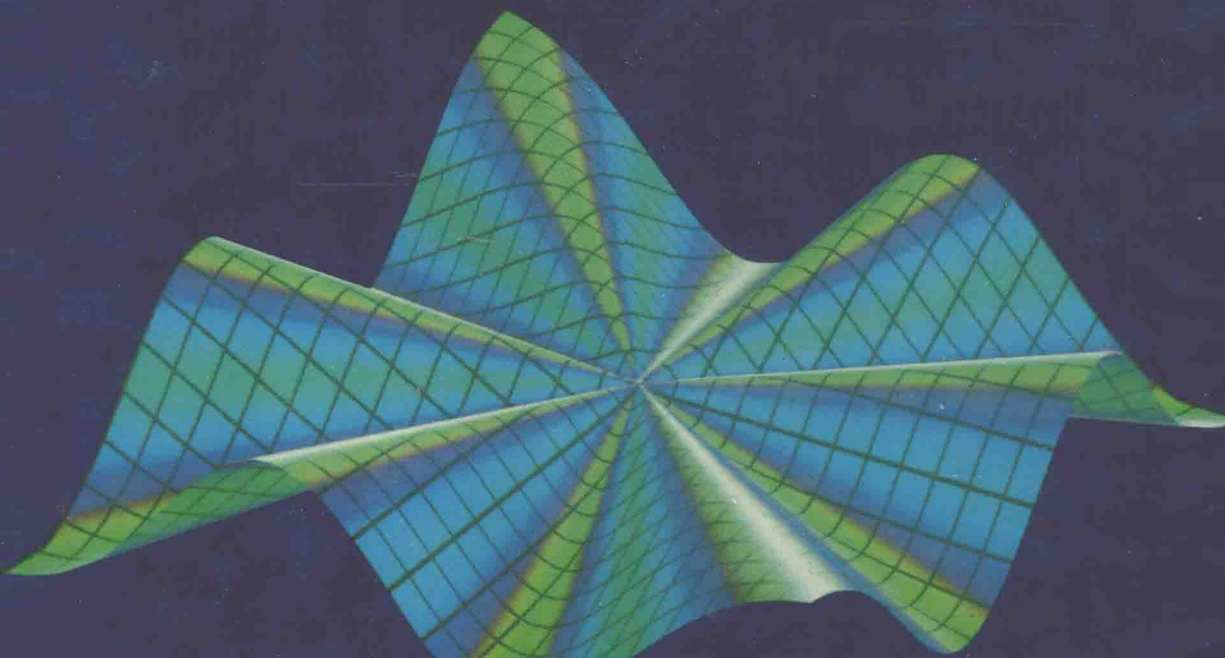
FOUNDATIONS OF

PROBABILITY & STATISTICS



RINAMAN

Foundations of Probability and Statistics

William C. Rinaman
Le Moyne College
Syracuse, New York

Saunders College Publishing
Harcourt Brace Jovanovich College Publishers
Forth Worth   Philadelphia   San Diego   New York   Orlando   Austin   San Antonio
Toronto   Montreal   London   Sydney   Tokyo

Text Typeface: Times Roman
Compositor: Technique Typesetting
Acquisitons Editor: Robert Stern
Managing Editor: Carol Field
Project Editor: Nancy Lubars
Copy Editor: Zanae Rodrigo
Manager of Art and Design: Carol Bleistine
Art Director: Doris Bruey
Art & Design Coordinator: Caroline McGowan
Text Designer: Alan Wendt
Cover Designer: Lawrence R. Didona
Text Artwork: Tech Graphics
Director of EDP: Tim Frelick
Production Manager: Robert Butler
Marketing Manager: Monica Wilson

Cover Credit: Illustration produced by James T. Hoffman of the Center for Geometry Analysis Numerics and Graphics, University of Massachusetts at Amherst. For more information, contact him via the INTERNet at jim@gang.umass.edu

Printed in the United States of America

Foundations of Probability and Statistics

0-03-071806-6

Library of Congress Catalog Card Number: 92-050763

3456 016 987654321

This book was printed on recycled, acid-free paper made from 10% post-consumer waste and 50% pre-consumer waste.
This cover was printed on recycled paper made from 10% post-consumer waste and 40% pre-consumer waste.

Preface

Introduction

A sound introduction to probability and statistics theory is required for many college majors. This text, written for junior and senior level undergraduates and beginning graduate students, provides that solid instruction. Designed for students with no previous experience in probability and statistics, this text uses many real-life examples to show how mathematical topics are related. Theories are introduced once solid relationships have been established.

This text encourages students to draw on knowledge gained in previous math classes. Many students entering their junior year have been exposed to a number of mathematical ideas, but they may have little experience applying these ideas to new subjects. Therefore, I have endeavored to show connections between material learned in earlier courses, such as calculus, and how they relate to the topics discussed here. Many junior level undergraduates need some help applying axioms and theorems to more complex mathematical ideas. For this reason, I have tried, in the earlier chapters, to provide some insight into why one would choose to prove a result in a particular manner. Later chapters offer less support so that students can develop their own skills.

Juniors and seniors in the mathematical and engineering sciences have the requisite background needed for this text. This means that for all chapters except Chapters 13 and 14 the only background assumed is three semesters of calculus. In Chapters 13 and 14 some rudimentary knowledge of matrix algebra is required. The first seven chapters provide more than enough material for a one-semester course in probability theory. Chapters 8 through 12 form the core of a one-semester course in statistical theory. Additional topics may be added, as desired, from Chapters 13 and 14.

Chapter Synopses

A brief synopsis of each chapter follows.

Chapter 1: The idea of random experiments is introduced. This is followed by the axioms of probability. From there some elementary consequences are discussed. Set theory and combinatorics are reviewed for later application. Probabilities for finite sample spaces of equally likely events are covered.

Chapter 2: The basic concepts of conditional probability and independent events are covered. In addition, Bayes' Rule is discussed.

Chapter 3: Random variables are defined along with the distribution function. The differences between discrete and continuous random variables are presented. Random vectors and the concept of independent random variables are introduced. Conditional distributions are also covered.

Chapter 4: Expected value is discussed. The expected values of functions of random variable, especially moments, are covered. Moment generating functions and their properties are considered. Conditional expectation is also briefly addressed.

Chapter 5: Coverage of the most commonly encountered discrete and continuous random variables is provided. The bivariate normal distribution is also considered.

Chapter 6: Standard methods for deriving the distributions of random variables, which are functions of other random variables, are introduced. The uses of distribution functions, moment generating functions, and Jacobians are discussed. These provide a basis for deriving the distribution of sample statistics.

Chapter 7: The general concepts of convergence in probability and distribution are discussed. Chebyshev's Inequality is introduced as a method for showing convergence in distribution. The Central Limit Theorem for independent and identically distributed random variables is given along with additional results for sum, products, and quotients of sequences of random variables.

Chapter 8: The role of descriptive statistics in data analysis is introduced. The quantile–quantile plot is emphasized as a method of determining the adequacy of a statistical model. Simple random sampling is covered.

Chapter 9: The chi-square, t and F distributions are defined. The independence of $\bar{X}$ and s^2 is proven. Order statistics are discussed in a very complete fashion for a text at this level.

Chapter 10: The point estimation problem is covered. Maximum likelihood and method of moments estimation are discussed as alternative approaches to estimation. The ideas of unbiasedness, efficiency, sufficiency and completeness are covered to show how it is possible to derive unbiased estimators having minimum variance. The large sample properties of estimators, such as consistency and being best asymptotically normal, are discussed. Robustness is introduced to address the possibility that the assumed model may differ from the actual one. Bayes estimators are introduced.

Chapter 11: The most standard methods for deriving confidence intervals are covered. One procedure for finding a nonparametric confidence interval for the median is discussed.

Chapter 12: Hypothesis testing is introduced. Types I and II errors are given as a method to discern between competing critical regions. Neyman-Pearson tests, uniformly most powerful tests, and likelihood ratio tests are discussed as methods to derive "good" tests. Chi-square testing is introduced. Nonparametric tests are covered to illustrate the case when the underlying distribution is not known.

Chapter 13: Least squares estimators are introduced and their properties analyzed for a simple regression model with normally distributed error terms. The calculus of matrices is covered to facilitate the discussion of multiple regression. Here a basic knowledge of matrix arithmetic is required. Multiple regression is presented in a matrix algebra setting. Correlation analysis is discussed. One emphasis here is to highlight the differences between regression and correlation. Nonparametric methods for simple regression and correlation are also considered.

Chapter 14: The basic concepts of experimental design are introduced. Single-factor, randomized complete block, and two-factor factorial designs are considered. The theoretical discussion here introduces the idea of idempotent matrices. A basic understanding of the rank of a matrix is assumed for this material. Nonparametric alternatives for the single-factor and randomized complete block cases are discussed.

Supplementary Material

An Instructors Manual is available with this text. It provides detailed solutions for all of the exercises in the text.

Acknowledgments

I wish to express my thanks to many people who have either directly and indirectly contributed to the completion of this text. The following reviewers are to be thanked for their helpful and insightful comments.

Dennis L. Young, Arizona State University

George R. Terrell, Virginia Polytechnic Institute

Basil P. Korin, The American University

William J. Studden, Purdue University

Ennis D. McCune, Stephen F. Austin State University

At Le Moyne College I wish to thank Lifang Hsu and Cathleen Zucco for their help and suggestions. I would also like to thank my students for their help in making the presentation more accessible. Another student, Jill Seamon, is to be thanked for her help during preparation of the manuscript. Finally, there is my wife Debra, whose patience and understanding made the whole thing possible.

Table of Contents

Chapter 1 **Random Experiments and Probability** 1

1.1 Introduction 1
1.2 Set Theory 3
1.3 Random Experiments 12
1.4 The Axioms of Probability 17
1.5 Counting Elementary Events 27
1.6 Probabilities for Finite Sample Spaces with Equally Likely Outcomes 41
Chapter Summary 53
Review Exercises 56

Chapter 2 **Conditional Probability and Independence** 59

2.1 Introduction 59
2.2 Conditional Probability 61
2.3 Bayes' Rule 71
2.4 Independence 77
Chapter Summary 92
Review Exercises 92

Chapter 3 **Random Variables** 95

3.1 Introduction 95
3.2 The Distribution Function 99

3.3 Discrete Random Variables 108
3.4 Continuous Random Variables 114
3.5 Random Vectors 121
3.6 Independent Random Variables 139
3.7 Conditional Distributions 145
Chapter Summary 149
Review Exercises 152

Chapter 4 **Expectation** 155

4.1 Introduction 155
4.2 Expected Value 156
4.3 Expectation of a Function of a Random Variable 162
4.4 Moments 173
4.5 Moment Generating Functions 189
4.6 Conditional Expectation 197
Chapter Summary 203
Review Exercises 204

Chapter 5 **Probability Distributions** 207

5.1 Introduction 207
5.2 Hypergeometric Distribution 208
5.3 Distributions Based on Independent Bernoulli Trials 211
5.4 Multinomial Distribution 218
5.5 Poisson Distribution 223
5.6 Uniform Distribution 228
5.7 Normal Distribution 232
5.8 Gamma Distribution 239
5.9 Other Continuous Distributions 244
5.9.1 Weibull Distribution 244
5.9.2 Beta Distribution 245
5.9.3 Cauchy Distribution 246
5.9.4 Double Exponential Distribution 247
5.9.5 Logistic Distribution 248
5.10 Bivariate Normal Distribution 250

Chapter Summary 253
Review Exercises 254

Chapter 6 **Distributions of Functions of Random Variables** 257

6.1 Introduction 257
6.2 Transformations of Discrete Random Variables 257
6.3 The Distribution Function Method 262
6.4 The Transformation of Variables Method 265
6.5 Moment Generating Function Method 270
6.6 Sums of Independent Random Variables 274
6.7 Distribution of Functions of Several Random Variables 277
Chapter Summary 286
Review Exercises 288

Chapter 7 **Limit Theorems** 291

7.1 Introduction 291
7.2 Convergence in Probability 292
7.3 Central Limit Theorem 308
7.4 More on Convergence in Distribution 315
Chapter Summary 319
Review Exercises 322

Chapter 8 **Statistical Models** 325

8.1 Introduction 325
8.2 Populations and Random Sampling 327
8.3 Numerical Summaries 332
- 8.3.1 Sample Mean 333
- 8.3.2 Sample Median 334
- 8.3.3 Sample Variance 334
- 8.3.4 Median Absolute Deviation 335
- 8.3.5 Percentile-Based Measures of Dispersion 335
- 8.3.6 Sample Skewness 337
- 8.3.7 Sample Kurtosis 338
- 8.3.8 An Example 338

8.4 Graphical Summaries 341
8.4.1 Quantile–Quantile Plots 342
8.4.2 Histograms 345
8.4.3 Stem-and-Leaf Plots 346
Chapter Summary 349
Review Exercises 350

Chapter 9 **Sampling Distributions** 351
9.1 Introduction 351
9.2 Distributions Based on Samples from Normal Populations 351
9.3 Independence of $\bar{X}$ and s^2 359
9.4 Order Statistics 362
Chapter Summary 377
Review Exercises 378

Chapter 10 **Point Estimation** 381
10.1 Introduction 381
10.2 Maximum Likelihood Estimation 382
10.3 Method of Moments 390
10.4 Properties of Estimators 394
10.4.1 Unbiasedness 394
10.4.2 Efficiency 398
10.4.3 Sufficiency 408
10.4.4 Minimal Sufficient Statistics 417
10.4.5 Completeness 419
10.5 Large Sample Properties of Estimators 430
10.5.1 Consistency 430
10.5.2 Asymptotic Efficiency 432
10.6 Bayes Estimators 434
10.7 Robust Estimation 441
Chapter Summary 444
Review Exercises 449

Chapter 11 **Confidence Intervals** 451

11.1 Introduction 451
11.2 Pivotal Quantity Method 453
11.3 Method Based on Sampling Distributions 463
11.4 Large Sample Confidence Intervals 468
11.5 A Nonparametric Confidence Interval 472
Chapter Summary 474
Review Exercises 476

Chapter 12 **Hypothesis Testing** 479

12.1 Introduction 479
12.2 Types of Errors 481
12.3 Testing Simple Hypotheses 489
12.4 Uniformly Most Powerful Tests 497
12.5 Likelihood Ratio Tests 509
12.6 Chi-Square Tests 528
12.7 Large Sample Tests 535
12.8 Sequential Probability Ratio Test 540
12.9 Nonparametric Tests 546
12.9.1 Sign Test 547
12.9.2 Wilcoxon Signed Rank Test 551
12.9.3 Wilcoxon Rank Sum Test 553
Chapter Summary 558
Review Exercises 564

Chapter 13 **Regression and Correlation** 569

13.1 Introduction 569
13.2 Simple Linear Regression 570
13.3 Matrix Calculus 589
13.4 Multiple Regression 595
13.5 Correlation 607
13.6 Nonparametric Methods 615
Chapter Summary 622
Review Exercises 630

Chapter 14 **Analysis of Variance** 633

14.1 Introduction 633
14.2 Experimental Design 634
14.3 One-Way Analysis of Variance 641
14.4 Two-Way Analysis of Variance 652
14.5 Estimation 667
14.6 Nonparametric Methods 673
Chapter Summary 681
Review Exercises 687

Appendix A **Statistical Tables** A.1

I—Binomial Distribution A.1
II—Poisson Distribution A.7
III—Normal Distribution A.13
IV—Chi-Square Distribution A.14
V—Student's t Distribution A.15
VI—F Distribution A.16
VII—Wilcoxon Signed Rank Test A.20
VIII—Wilcoxon Rank Sum Test A.27
IX—Spearman's Rho A.35
X—Kruskal-Wallis Test A.38
XI—Friedman Test A.53

Appendix B **Probability Distributions** A.59

Discrete Random Variables A.59
Continuous Random Variables A.61

Appendix C **Bibliography** A.63

Appendix D **Answers to Selected Odd-Numbered Exercises** A.65

Index I.1

1 Random Experiments and Probability

1.1 Introduction

Most people know intuitively what the word "probability" means. Many of us have encountered probabilities in our daily activities. For example, if we flip a coin it is our understanding that heads should result about half the time. When a local weather forecaster states that there is a 40 percent chance for rain tomorrow we have a general understanding that what is meant is that, given the current conditions, rain may occur 4 times out of 10. If we play bridge, we may draw on past experience to assess the likelihood that the cards held by our opponents have been distributed in a certain way in order to plan how we play a current hand. In fact, anyone who is consistently successful at games involving chance such as poker and backgammon is well aware of the relative frequency of the various possible outcomes in the play of the game.

The situations mentioned above have some things in common. In each we face a situation that is, at least in principle, well defined and repeatable. The possible outcomes are known, but the particular outcome that will take place this time cannot be predicted. What probability attempts to do is to determine a numerical value that tells what proportion of the time each possibility will occur.

One could correctly argue that the weather example is not like the others. It does not involve chance occurrences but rather represents a less than complete understanding of the physics of the atmosphere. This is the case in a number of areas where probability may be applied. Many physical situations are so complicated that it is not feasible to develop a mathematical model that accounts for all of the variables. In such cases it is common to develop models that use the main variables and then lump together the unused variables in a term that represents the unpredicted part of the model called the "noise." The noise in the model is what injects randomness into predictions.

Our goal is to investigate the mathematical structures that can be used to describe these types of situations, to define precisely what a probability is, and to study its properties. These concepts can be studied on a number of levels ranging from an intuitive approach, such as would be appropriate for a person mainly interested in applications, to a highly abstract approach, where probability is a corner of an area of real analysis known as measure theory. We aim for a middle-of-the-road approach—rigorous enough to give some appreciation for the mathematics involved but grounded enough in real-life applications to show how probability applies to real-life situations.

The origins of the study of probability theory can be traced to the 17th century where it was used to predict games of chance. A professional gambler named Chevalier de Mere posed the following problem to the French mathematician Pascal, now referred to as the *problem of the points*. Two players put up stakes and play a game of chance with the understanding that the winner takes all the money. How should the stakes be divided if the game is stopped before either player has won? In attacking this problem Pascal began and carried on a lengthy correspondence with another famous French mathematician, Fermat. At about the same time other famous mathematicians such as Christian Huygens and James Bernoulli worked on another problem that has become a classic, the *gambler's ruin problem*. In this problem two players, A and B, begin with some coins and play a game of chance. On each play of the game A wins a coin from B p percent of the time, and B wins a coin from A $100 - p$ percent of the time. The game runs until one player has won all of the coins. What proportion of the games will be won by A? We shall return to these problems in Chapter 2.

For a long time, probability theory concentrated in the area of games of chance, mainly because the areas of applications in the natural sciences such as errors of measurements, genetics, and statistics were not very well developed. Beginning in the 18th century the natural sciences began to use probabilistic ideas. The more advanced analytical methods were developed by mathematicians of the time such as DeMoivre, Laplace, Gauss, and Poisson. In the middle of the 19th century, the Russian mathematicians Chebyshev, Markov, and Liapunov introduced the random variable. This concept had an enormous impact on the study of probability.

In the 20th century, probability theory became a topic of study for its own sake. As we shall see in Section 1.2, the Soviet mathematician Andrei Kolmogorov developed an axiomatic basis for the subject. In addition, Emil Borel demonstrated connections between probability and measure theory. This connection has served as the basis for many recent developments in probability theory.

Today, probability theory is an important component in many areas of study. For example, the subjects of quantum mechanics and thermodynamics in physics are based heavily on the idea of random events. Statistics uses probability as its primary tool for assessing the quality of procedures. The topics of reliability, quality control, inventory theory, and queuing theory are applications of probability theory in operations research and industrial engineering. And in medicine, the subject of epidemiology models the spread of disease using probabilistic ideas. The application of probability theory to real-life problems is widespread, as will be demonstrated in this text.

1.2 Set Theory

Probability theory is based on the occurrence of events in a random experiment. Random experiments and events are most conveniently described mathematically by using sets and set operations. Random experiments will be formally defined and discussed in detail in Section 1.3. For now all we need to know is that an experiment is taken to be any activity that is conducted according to a well-defined set of rules where an observation of the outcome is made.

This section reviews set theories and the ideas that are useful in describing these sets. You may be familiar with the basic ideas in set theory from other mathematics courses. For more complete coverage, consult any discrete mathematics text. A more rigorous approach to the topic is given in *Naive Set Theory* by P. R. Halmos.

Definition 1.1

An *event* is a well-defined collection of outcomes of an experiment. The outcomes that belong to the set are called its *elements*. The collection of all possible outcomes of an experiment is called the *sample space*.

The term "well defined" simply means that if you are given a description of an event and an outcome it is possible to determine definitely if that outcome is an element of the event. Events are typically denoted by using capital letters such as A, B, and so on. The sample space is commonly denoted by S. Outcomes are denoted by using lowercase letters. If an object x is an outcome in the event A we would write this fact as

$$x \in A.$$

Similarly, if x is not an outcome in A we would write

$$x \notin A.$$

The exact composition of a given event is usually described in one of two ways. If the event has relatively few outcomes it is often convenient simply to describe the event by listing them in braces ($\{,\}$) and separating them by commas. For example, in an experiment where a number is selected from the digits the sample space S of digits would be described by

$$S = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}.$$

The event O of observing an odd digit would be described by

$$O = \{1, 3, 5, 7, 9\}.$$

The order in which elements appear in the list is meaningless. Thus, we could just as correctly define S by

$$S = \{7, 4, 2, 3, 5, 8, 1, 0, 9, 6\}.$$

In addition, if an outcome is in an event it appears only once in the list of elements. This means that $\{1, 2, 2\}$ is the same as $\{1, 2\}$ for the purpose of defining an event.

The other common way to define an event is to list those properties that characterize the outcomes in the event. Using this method we would describe S as follows.

$$S = \{x : x \text{ is an integer}, 0 \leq x \leq 9\}$$

This would read as "S is the set of outcomes x such that x is an integer and x is greater than or equal to 0 and less than or equal to 9." The colon is translated to read "such that," and the comma is read as "and." It is important to note that

$$E = \{x : x \text{ is an integer}\}$$

does not define the same sample space. The integer -1 is an element E but not of S. Thus, when defining a sample space or an event in this manner it is critical to give a list of properties that are satisfied by and only by each of the outcomes of the set.

We next look at the concept of 2 events being equal. The idea is quite intuitive. We would certainly be inclined to state that the event $A = \{1, 2, 3\}$ is the same as the event $\{3, 1, 2\}$. This notion is made formal by the following definition.

Definition 1.2

> Two events, A and B, are said to be *equal* if they contain exactly the same elements.

In formal set theory this is known as the *axiom of extension.* If the event A contains the same elements as the event B we should state that "A equals B" and write

$$A = B.$$

Similarly, if one event contains an outcome that is not in the other event then we would say "A is not equal to B" and write

$$A \neq B.$$

Example 1.1

Suppose we select a letter from the English alphabet. Let $V = \{a, e, i, o, u\}$. Here the sample space would be the letters of the English alphabet. We could define

$$A = \{x : x \text{ is a letter in the English alphabet}, x \text{ is a vowel}\}.$$

In this case $A = V$. ■

The opposite of the sample space, which contains all of the outcomes of an experiment, is the *empty set.* This event contains no outcomes and is denoted by $\varnothing$. An example of an event specification that results in the empty set is

$$\varnothing = \{x : x < 1,\ x > 1\}.$$

There are no real numbers that are simultaneously greater than and less than unity. Thus, the set is empty. It is logical to ask if there can be more than one empty set. The answer is no due to the axiom of extension. If A is empty and B is empty, then there are no elements in A and none in B. Hence there is no element that is in A that is not in B, which means that they are the same set. This kind of reasoning may seem a little strange, but arguments involving the empty set are typically carried out in this manner.

Example 1.2 Suppose we select a positive integer. The sample space S would be the positive integers. Then

$$A = \{x : x^2 \leq 5\} = \{1, 2\} \quad \text{while} \quad B = \{x : x^2 = 5\} = \emptyset.$$

■

If two events A and B are not equal, 3 things are possible. One event may be completely contained in the other such as $A = \{1, 2, 3, 4\}$ and $B = \{2, 3\}$. The 2 events have no outcomes in common such as $A = \{1, 2\}$ and $B = \{3, 4\}$. The last possibility is that some but not all outcomes are common to both sets such as $A = \{1, 3\}$ and $B = \{2, 3, 4\}$. When an event is completely contained in another, then the event that is contained is said to be a *subset* of the other. In the first example given above we would say that B is a subset of A and write it as

$$B \subset A.$$

It is common practice when A has outcomes that are not in B to refer to B as being a *proper subset* of A and to A as being a *superset* of B. A formal definition of subset is given below. If B is a subset of A, but not necessarily a proper subset, then we would write

$$B \subseteq A.$$

Definition 1.3 If, for any $x \in A$, we have $x \in B$ then $A \subset B$. If A is not a subset of B we would write

$$A \not\subset B.$$

All determinations of whether one event is a subset of another proceed in the same manner. Suppose we wish to determine if A is a subset of B. The method is to pick a general element from A and show that its belonging to A necessarily implies that it also belongs to B. Some basic properties of subsets are given by the following theorem.

Theorem 1.1 Let A, B, and C be any three events from the sample space S. Then the following are true.

1. $\emptyset \subseteq A \subseteq S$
2. $A \subseteq A$
3. If $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$.
4. $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$.