Special Continuous Distributions

Notation and Parameters	Continuous pdf $f(x)$	Mean	Variance	MGF $M_X(t)$
Student's t				
$X \sim t(\nu)$	$\dfrac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \dfrac{1}{\sqrt{\nu\pi}} \left(1 + \dfrac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$	0	$\dfrac{\nu}{\nu - 2}$	**
$\nu = 1, 2, \ldots$		$1 < \nu$	$2 < \nu$	
Snedecor's F				
$X \sim F(\nu_1, \nu_2)$	$\dfrac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \left(\dfrac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} x^{\frac{\nu_1}{2} - 1}$	$\dfrac{\nu_2}{\nu_2 - 2}$	$\dfrac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)}$	**
$\nu_1 = 1, 2, \ldots$ $\nu_2 = 1, 2, \ldots$	$\times \left(1 + \dfrac{\nu_1}{\nu_2}x\right)^{-\frac{\nu_1+\nu_2}{2}}$	$2 < \nu_2$	$4 < \nu_2$	
Beta				
$X \sim \text{BETA}(a, b)$	$\dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$	$\dfrac{a}{a+b}$	$\dfrac{ab}{(a+b+1)(a+b)^2}$	*
$0 < a$ $0 < b$	$0 < x < 1$			

*Not tractable.
**Does not exist.

Special Continuous Distributions

Notation and Parameters	Continuous pdf $f(x)$	Mean	Variance	MGF $M_X(t)$
Weibull				
$X \sim \text{WEI}(\theta, \beta)$	$\frac{\beta}{\theta^\beta} x^{\beta-1} e^{-(x/\theta)^\beta}$	$\theta\Gamma\left(1+\frac{1}{\beta}\right)$	$\theta^2\left[\Gamma\left(1+\frac{2}{\beta}\right)-\Gamma^2\left(1+\frac{1}{\beta}\right)\right]$	*
$0 < \theta$ $0 < \beta$	$0 < x$			
Extreme Value				
$X \sim \text{EV}(\theta, \eta)$	$\frac{1}{\theta}\exp\{[(x-\eta)/\theta]-\exp[(x-\eta)/\theta]\}$	$\eta - \gamma\theta$	$\frac{\pi^2\theta^2}{6}$	$e^{\eta t}\Gamma(1+\theta t)$
$0 < \theta$		$\gamma \doteq 0.5772$ (Euler's const.)		
Cauchy				
$X \sim \text{CAU}(\theta, \eta)$	$\frac{1}{\theta\pi\{1+[(x-\eta)/\theta]^2\}}$	**	**	**
$0 < \theta$				
Pareto				
$X \sim \text{PAR}(\theta, \kappa)$	$\frac{\kappa}{\theta(1+x/\theta)^{\kappa+1}}$	$\frac{\theta}{\kappa-1}$	$\frac{\theta^2\kappa}{(\kappa-2)(\kappa-1)^2}$	**
$0 < \theta$ $0 < \kappa$	$0 < x$	$1 < \kappa$	$2 < \kappa$	
Chi-Square				
$X \sim \chi^2(\nu)$	$\frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2}$	ν	2ν	$\left(\frac{1}{1-2t}\right)^{\nu/2}$
$\nu = 1, 2, \ldots$	$0 < x$			

INTRODUCTION TO PROBABILITY AND MATHEMATICAL STATISTICS

SECOND EDITION

Lee J. Bain
University of Missouri – Rolla

Max Engelhardt
University of Idaho

Duxbury
Thomson Learning™

Australia • Canada • Mexico • Singapore • Spain • United Kingdom • United States

The *Duxbury Classic Series* is a collection of authoritative works from respected authors. Reissued as paperbacks, these successful titles are now more affordable.

For more information about this or any other Duxbury product, contact:
DUXBURY
511 Forest Lodge Road
Pacific Grove, CA 93950 USA
www.duxbury.com
1-800-423-0563 (Thomson Learning Academic Resource Center)

Printed in the United States of America

10 9 8 7 6 5 4 3 2

Library of Congress Cataloging-in-Publication Data
Bain, Lee J.
Introduction to probability and mathematical statistics / Lee J. Bain, Max Engelhardt.—2nd ed.
p. cm.—(The Duxbury advanced series in statistics and decision sciences)
Includes bibliographical references and index.
ISBN 0-534-92930-3 (hard cover)
ISBN 0-534-38020-4 (paperback)
1. Probabilities. 2. Mathematical statistics. I. Engelhardt, Max. II. Title.
III. Series.
QA273.B2546 1991
519.2—dc20

91-25923
CIP

CONTENTS

CHAPTER 1

CHAPTER 2

CHAPTER 3

SPECIAL PROBABILITY DISTRIBUTIONS 90

CHAPTER 4

JOINT DISTRIBUTIONS 136

CHAPTER 5

PROPERTIES OF RANDOM VARIABLES 171

CHAPTER 6

FUNCTIONS OF RANDOM VARIABLES 193

CHAPTER 7

LIMITING DISTRIBUTIONS 231

CHAPTER 8

STATISTICS AND SAMPLING DISTRIBUTIONS 263

* Advanced (or optional) topics

CHAPTER 9

POINT ESTIMATION 288

CHAPTER 10

SUFFICIENCY AND COMPLETENESS 335

CHAPTER 11

INTERVAL ESTIMATION 358

CHAPTER 12

TESTS OF HYPOTHESES 389

CHAPTER 13

CONTINGENCY TABLES AND GOODNESS-OF-FIT 442

CHAPTER 14

NONPARAMETRIC METHODS 468

* Advanced (or optional) topics

PREFACE

This book provides an introduction to probability and mathematical statistics. Although the primary focus of the book is on a mathematical development of the subject, we also have included numerous examples and exercises that are oriented toward applications. We have attempted to achieve a level of presentation that is appropriate for senior-level undergraduates and beginning graduate students.

The second edition involves several major changes, many of which were suggested by reviewers and users of the first edition. Chapter 2 now is devoted to general properties of random variables and their distributions. The chapter now includes moments and moment generating functions, which occurred somewhat later in the first edition. Special distributions have been placed in Chapter 3. Chapter 8 is completely changed. It now considers sampling distributions and some basic properties of statistics. Chapter 15 is also new. It deals with regression and related aspects of linear models.

As with the first edition, the only prerequisite for covering the basic material is calculus, with the lone exception of the material on general linear models in Section 15.4; this assumes some familiarity with matrices. This material can be omitted if so desired.

Our intent was to produce a book that could be used as a textbook for a two-semester sequence in which the first semester is devoted to probability concepts and the second covers mathematical statistics. Chapters 1 through 7 include topics that usually are covered in a one-semester introductory course in probability, while Chapters 8 through 12 contain standard topics in mathematical statistics. Chapters 13 and 14 deal with goodness-of-fit and nonparametric statistics. These chapters tend to be more methods-oriented. Chapters 15 and 16 cover material in regression and reliability, and these would be considered as optional or special topics. In any event, judgment undoubtedly will be required in the choice of topics covered or the amount of time allotted to topics if the desired material is to be completed in a two-semester course.

It is our hope that those who use the book will find it both interesting and informative.

ACKNOWLEDGMENTS

We gratefully acknowledge the numerous suggestions provided by the following reviewers:

Dean H. Fearn
California State University–Hayward

Alan M. Johnson
University of Arkansas, Little Rock

Joseph Glaz
University of Connecticut

Benny P. Lo
Ohlone College

Victor Goodman
Rensselaer Polytechnic Institute

D. Ramachandran
California State University–Sacramento

Shu-ping C. Hodgson
Central Michigan University

Douglas A. Wolfe
Ohio State University

Robert A. Hultquist
Pennsylvania State University

Linda J. Young
Oklahoma State University

Thanks also are due to the following users of the first edition who were kind enough to relate their experiences to the authors: H. A. David, Iowa State University; Peter Griffin, California State University–Sacramento.

Finally, special thanks are due for the moral support of our wives, Harriet Bain and Linda Engelhardt.

Lee J. Bain
Max Engelhardt

CHAPTER 1

PROBABILITY

1.1 INTRODUCTION

In any scientific study of a physical phenomenon, it is desirable to have a mathematical model that makes it possible to describe or predict the observed value of some characteristic of interest. As an example, consider the velocity of a falling body after a certain length of time, t. The formula $v = gt$, where $g \doteq 32.17$ feet per second per second, provides a useful mathematical model for the velocity, in feet per second, of a body falling from rest in a vacuum. This is an example of a **deterministic model**. For such a model, carrying out repeated experiments under ideal conditions would result in essentially the same velocity each time, and this would be predicted by the model. On the other hand, such a model may not be adequate when the experiments are carried out under less than ideal conditions. There may be unknown or uncontrolled variables, such as air temperature or humidity, that might affect the outcome, as well as measurement error or other factors that might cause the results to vary on different performances of the experiment. Furthermore, we may not have sufficient knowledge to derive a more complicated model that could account for all causes of variation.

There are also other types of phenomena in which different results may naturally occur by chance, and for which a deterministic model would not be appropriate. For example, an experiment may consist of observing the number of particles emitted by a radioactive source, the time until failure of a manufactured component, or the outcome of a game of chance.

The motivation for the study of probability is to provide mathematical models for such nondeterministic situations; the corresponding mathematical models will be called **probability models** (or **probabilistic models**). The term **stochastic**, which is derived from the Greek word *stochos*, meaning "guess," is sometimes used instead of the term *probabilistic*.

A careful study of probability models requires some familiarity with the notation and terminology of set theory. We will assume that the reader has some knowledge of sets, but for convenience we have included a review of the basic ideas of set theory in Appendix A.

1.2

NOTATION AND TERMINOLOGY

The term **experiment** refers to the process of obtaining an observed result of some phenomenon. A performance of an experiment is called a **trial** of the experiment, and an observed result is called an **outcome**. This terminology is rather general, and it could pertain to such diverse activities as scientific experiments or games of chance. Our primary interest will be in situations where there is uncertainty about which outcome will occur when the experiment is performed. We will assume that an experiment is repeatable under essentially the same conditions, and that the set of all possible outcomes can be completely specified before experimentation.

Definition 1.2.1

The set of all possible outcomes of an experiment is called the **sample space**, denoted by S.

Note that one and only one of the possible outcomes will occur on any given trial of the experiments.

Example 1.2.1 An experiment consists of tossing two coins, and the observed face of each coin is of interest. The set of possible outcomes may be represented by the sample space

$$S = \{HH, HT, TH, TT\}$$

which simply lists all possible pairings of the symbols H (heads) and T (tails). An alternate way of representing such a sample space is to list all possible ordered pairs of the numbers 1 and 0, $S = \{(1, 1), (1, 0), (0, 1), (0, 0)\}$, where, for example, (1, 0) indicates that the first coin landed heads up and the second coin landed tails up.

Example 1.2.2 Suppose that in Example 1.2.1 we were not interested in the individual outcomes of the coins, but only in the total number of heads obtained from the two coins. An appropriate sample space could then be written as $S^* = \{0, 1, 2\}$. Thus, different sample spaces may be appropriate for the same experiment, depending on the characteristic of interest.

Example 1.2.3 If a coin is tossed repeatedly until a head occurs, then the natural sample space is $S = \{H, TH, TTH, \ldots\}$. If one is interested in the number of tosses required to obtain a head, then a possible sample space for this experiment would be the set of all positive integers, $S^* = \{1, 2, 3, \ldots\}$, and the outcomes would correspond directly to the number of tosses required to obtain the first head. We will show in the next chapter that an outcome corresponding to a sequence of tosses in which a head is never obtained need not be included in the sample space.

Example 1.2.4 A light bulb is placed in service and the time of operation until it burns out is measured. At least conceptually, the sample space for this experiment can be taken to be the set of nonnegative real numbers, $S = \{t \mid 0 \leqslant t < \infty\}$.

Note that if the actual failure time could be measured only to the nearest hour, then the sample space for the actual observed failure time would be the set of nonnegative integers, $S^* = \{0, 1, 2, 3, \ldots\}$. Even though S^* may be the observable sample space, one might prefer to describe the properties and behavior of light bulbs in terms of the conceptual sample space S. In cases of this type, the discreteness imposed by measurement limitations is sufficiently negligible that it can be ignored, and both the measured response and the conceptual response can be discussed relative to the conceptual sample space S.

A sample space S is said to be **finite** if it consists of a finite number of outcomes, say $S = \{e_1, e_2, \ldots, e_N\}$, and it is said to be **countably infinite** if its outcomes can be put into a one-to-one correspondence with the positive integers, say $S = \{e_1, e_2, \ldots\}$.

Definition 1.2.2

If a sample space S is either finite or countably infinite, then it is called a **discrete sample space**.

A set that is either finite or countably infinite also is said to be **countable**. This is the case in the first three examples. It is also true for the last example when failure times are recorded to the nearest hour, but not for the conceptual sample space. Because the conceptual space involves outcomes that may assume any value in some interval of real numbers (i.e., the set of nonnegative real numbers), it could be termed a **continuous sample space**, and it provides an example where a discrete sample space is not an appropriate model. Other, more complicated experiments exist, the sample spaces of which also could be characterized as continuous, such as experiments involving two or more continuous responses.

Example 1.2.5 Suppose a heat lamp is tested and X, the amount of light produced (in lumens), and Y, the amount of heat energy (in joules), are measured. An appropriate sample space would be the Cartesian product of the set of all nonnegative real numbers with itself,

$$S = [0, \infty) \times [0, \infty) = \{(x, y) \mid 0 \leqslant x < \infty \text{ and } 0 \leqslant y < \infty\}$$

Each variable would be capable of assuming any value in some subinterval of $[0, \infty)$.

Sometimes it is possible to determine bounds on such physical variables, but often it is more convenient to consider a conceptual model in which the variables are not bounded. If the likelihood of the variables in the conceptual model exceeding such bounds is negligible, then there is no practical difficulty in using the conceptual model.

Example 1.2.6 A thermograph is a machine that records temperature continuously by tracing a graph on a roll of paper as it moves through the machine. A thermographic recording is made during a 24-hour period. The observed result is the graph of a continuous real-valued function $f(t)$ defined on the time interval $[0, 24] = \{t \mid 0 \leqslant t \leqslant 24\}$, and an appropriate sample space would be a collection of such functions.

Definition 1.2.3

An **event** is a subset of the sample space S. If A is an event, then A has **occurred** if it contains the outcome that occurred.

To illustrate this concept, consider Example 1.2.1. The subset

$$A = \{\text{HH, HT, TH}\}$$

contains the outcomes that correspond to the event of obtaining "at least one head." As mentioned earlier, if one of the outcomes in A occurs, then we say that the event A has occurred. Similarly, if one of the outcomes in $B = \{HT, TH, TT\}$ occurs, then we say that the event "at least one tail" has occurred.

Set notation and terminology provide a useful framework for describing the possible outcomes and related physical events that may be of interest in an experiment. As suggested above, a subset of outcomes corresponds to a physical event, and the event or the subset is said to occur if any outcome in the subset occurs. The usual set operations of union, intersection, and complement provide a way of expressing new events in terms of events that already have been defined. For example, the event C of obtaining "at least one head and at least one tail" can be expressed as the intersection of A and B, $C = A \cap B = \{HT, TH\}$. Similarly, the event "at least one head or at least one tail" can be expressed as the union $A \cup B = \{HH, HT, TH, TT\}$, and the event "no heads" can be expressed as the complement of A relative to S, $A' = \{TT\}$.

A review of set notation and terminology is given in Appendix A.

In general, suppose S is the sample space for some experiments, and that A and B are events. The intersection $A \cap B$ represents the outcomes of the event "A and B," while the union $A \cup B$ represents the event "A or B." The complement A' corresponds to the event "not A." Other events also can be represented in terms of intersections, unions, and complements. For example, the event "A but not B" is said to occur if the outcome of the experiment belongs to $A \cap B'$, which sometimes is written as $A - B$. The event "exactly one of A or B" is said to occur if the outcome belongs to $(A \cap B') \cup (A' \cap B)$. The set $A' \cap B'$ corresponds to the event "neither A nor B." The set identity $A' \cap B' = (A \cup B)'$ is another way to represent this event. This is one of the set properties that usually are referred to as De Morgan's laws. The other such property is $A' \cup B' = (A \cap B)'$.

More generally, if $A_1, \ldots, A_k$ is a finite collection of events, occurrence of an outcome in the intersection $A_1 \cap \cdots \cap A_k$ (*or* $\bigcap_{i=1}^{k} A_i$) corresponds to the occurrence of the event "every A_i; $i = 1, \ldots, k$." The occurrence of an outcome in the union $A_1 \cup \cdots \cup A_k$ (*or* $\bigcup_{i=1}^{k} A_i$) corresponds to the occurrence of the event "at least one A_i; $i = 1, \ldots, k$." Similar remarks apply in the case of a countably infinite collection $A_1, A_2, \ldots$, with the notations $A_1 \cap A_2 \cap \cdots$ (*or* $\bigcap_{i=1}^{\infty} A_i$) for the intersection and $A_1 \cup A_2 \cup \cdots$ (*or* $\bigcup_{i=1}^{\infty} A_i$) for the union.

The intersection (or union) of a finite or countably infinite collection of events is called a **countable intersection** (or **union**).

We will consider the whole sample space S as a special type of event, called the **sure event**, and we also will include the empty set $\varnothing$ as an event, called the **null event**. Certainly, any set consisting of only a single outcome may be considered as an event.

Definition 1.2.4

An event is called an **elementary event** if it contains exactly one outcome of the experiment.

In a discrete sample space, any subset can be written as a countable union of elementary events, and we have no difficulty in associating every subset with an event in the discrete case.

In Example 1.2.1, the elementary events are {HH}, {HT}, {TH}, and {TT}, and any other event can be written as a finite union of these elementary events. Similarly, in Example 1.2.3, the elementary events are {H}, {TH}, {TTH}, ..., and any event can be represented as a countable union of these elementary events.

It is not as easy to represent events for the continuous examples. Rather than attempting to characterize these events rigorously, we will discuss some examples.

In Example 1.2.4, the light bulbs could fail during any time interval, and any interval of nonnegative real numbers would correspond to an interesting event for that experiment. Specifically, suppose the time until failure is measured in hours. The event that the light bulb "survives at most 10 hours" corresponds to the interval $A = [0, 10] = \{t \mid 0 \leqslant t \leqslant 10\}$. The event that the light bulb "survives more than 10 hours" is $A' = (10, \infty) = \{t \mid 10 < t < \infty\}$. If $B = [0, 15)$, then $C = B \cap A' = (10, 15)$ is the event of "failure between 10 and 15 hours."

In Example 1.2.5, any Cartesian product based on intervals of nonnegative real numbers would correspond to an event of interest. For example, the event

$$(10, 20) \times [5, \infty) = \{(x, y) \mid 10 < x < 20 \text{ and } 5 \leqslant y < \infty\}$$

corresponds to "the amount of light is between 10 and 20 lumens and the amount of energy is at least 5 joules." Such an event can be represented graphically as a rectangle in the xy plane with sides parallel to the coordinate axes.

In general, any physical event can be associated with a reasonable subset of S, and often a subset of S can be associated with some meaningful event. For mathematical reasons, though, when defining probability it is desirable to restrict the types of subsets that we will consider as events in some cases. Given a collection of events, we will want any countable union of these events to be an event. We also will want complements of events and countable intersections of events to be included in the collection of subsets that are defined to be events. We will assume that the collection of possible events includes all such subsets, but we will not attempt to describe all subsets that might be called events.

An important situation arises in the following developments when two events correspond to disjoint subsets.

Definition 1.2.5

Two events A and B are called **mutually exclusive** if $A \cap B = \varnothing$.

If events are mutually exclusive, then they have no outcomes in common. Thus, the occurrence of one event precludes the possibility of the other occurring. In Example 1.2.1, if A is the event "at least one head" and if we let B be the event "both tails," then A and B are mutually exclusive. Actually, in this example $B = A'$ (the complement of A). In general, complementary events are mutually exclusive, but the converse is not true. For example, if C is the event "both heads," then B and C are mutually exclusive, but not complementary.

The notion of mutually exclusive events can be extended easily to more than two events.

Definition 1.2.6

Events $A_1, A_2, A_3, \ldots,$ are said to be **mutually exclusive** if they are pairwise mutually exclusive. That is, if $A_i \cap A_j = \varnothing$ whenever $i \neq j$.

One possible approach to assigning probabilities to events involves the notion of relative frequency.

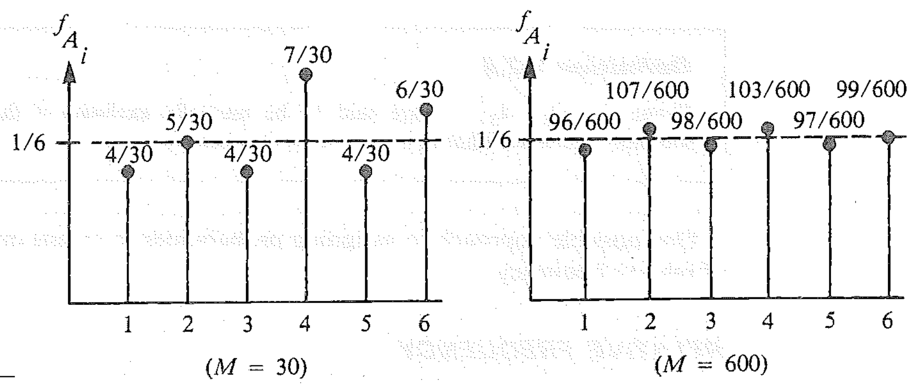
RELATIVE FREQUENCY

For the experiment of tossing a coin, we may declare that the probability of obtaining a head is 1/2. This could be interpreted in terms of the relative frequency with which a head is obtained on repeated tosses. Even though the coin may be tossed only once, conceivably it could be tossed many times, and experience leads us to expect a head on approximately one-half of the tosses. At least conceptually, as the number of tosses approaches infinity, the proportion of times a head occurs is expected to converge to some constant p. One then might define the probability of obtaining a head to be this conceptual limiting value. For a balanced coin, one would expect $p = 1/2$, but if the coin is unbalanced, or if the experiment is conducted under unusual conditions that tend to bias the outcomes in favor of either heads or tails, then this assignment would not be appropriate.

More generally, if $m(A)$ represents the number of times that the event A occurs among M trials of a given experiment, then $f_A = m(A)/M$ represents the **relative frequency** of occurrence of A on these trials of the experiment.

Example 1.2.7 An experiment consists of rolling an ordinary six-sided die. A natural sample space is the set of the first six positive integers, $S = \{1, 2, 3, 4, 5, 6\}$. A simulated die-rolling experiment is performed, using a "random number generator" on a computer. In Figure 1.1, the relative frequencies of the elementary events $A_1 = \{1\}$, $A_2 = \{2\}$, and so on are represented as the heights of vertical lines. The first graph shows the relative frequencies for the first $M = 30$ rolls, and the second graph gives the results for $M = 600$ rolls. By inspection of these graphs, obviously the relative frequencies tend to "stabilize" near some fixed value as M increases. Also included in the figure is a dotted line of height 1/6, which is the value that experience would suggest as the long-term relative frequency of the outcomes of rolling a die. Of course, in this example, the results are more relevant to the properties of the random number generator used to simulate the experiment than to those of actual dice.

FIGURE 1.1 Relative frequencies of elementary events for die-rolling experiment



If, for an event A, the limit of f_A as M approaches infinity exists, then one could assign probability to A by

$$P(A) = \lim_{M \to \infty} f_A \tag{1.2.1}$$

This expresses a property known as **statistical regularity.** Certain technical questions about this property require further discussion. For example, it is not clear under what conditions the limit in equation (1.2.1) will exist, or in what sense, or whether it will necessarily be the same for every sequence of trials. Our approach to this problem will be to define probability in terms of a set of axioms and eventually show that the desired limiting behavior follows.

To motivate the defining axioms of probability, consider the following properties of relative frequencies. If S is the sample space for an experiment and A is an event, then clearly $0 \leqslant m(A)$ and $m(S) = M$, because $m(A)$ counts the number of occurrences of A, and S occurs on each trial. Furthermore, if A and B are mutually exclusive events, then outcomes in A are distinct from outcomes in B, and consequently $m(A \cup B) = m(A) + m(B)$. More generally, if $A_1, A_2, \ldots$ are pairwise mutually exclusive, then $m(A_1 \cup A_2 \cup \cdots) = m(A_1) + m(A_2) + \cdots$. Thus, the following properties hold for relative frequencies:

$$0 \leqslant f_A \tag{1.2.2}$$

$$f_S = 1 \tag{1.2.3}$$

$$f_{A_1 \cup A_2 \cup \cdots} = f_{A_1} + f_{A_2} + \cdots \tag{1.2.4}$$

if $A_1, A_2, \ldots$ are pairwise mutually exclusive events.