Math Review Appendix: Sets, Functions, Permutations, Combinations, Notation, and Real Analysis

A1. Introduction

In this appendix we review basic results concerning set theory, relations and functions, combinations and permutations, summation and integration notation, and some fundamental concepts in real analysis. We also review the meaning of the terms definition, *axiom, theorem, corollary, and lemma*, which are labels that are affixed to a myriad of statements and results that constitute the theory of probability and mathematical statistics. The topics reviewed in this appendix constitute basic foundational material on which the study of mathematical statistics is based. Additional mathematical results, often of a more advanced nature, will be introduced throughout the text as the need arises.

A2. Definitions, Axioms, Theorems, Corollaries, and Lemmas

The development of the theory of probability and mathematical statistics involves a considerable number of statements consisting of definitions, axioms, theorems, corollaries, and lemmas. These terms will be used for organizing the various statements and results we will examine into these categories:

1. Descriptions of meaning;
2. Statements that are acceptable as true without proof;
3. Formulas or statements that require proof of validity;
4. Formulas or statements whose validity follows immediately from other true formulas or statements; and
5. Results, generally from other branches of mathematics, whose primary purpose is to facilitate the proof of validity of formulas or statements in mathematical statistics.

More formally, we present the following meaning of the terms.

Definition: A statement of the meaning of a word, word group, sign, or symbol.

Axiom (or postulate): A statement that has found general acceptance, or is thought to be worthy thereof, on the basis of an appeal to intrinsic merit or self-evidence, and thus requires no proof of validity.

Theorem (or proposition): A formula or statement that is deduced from other proved or accepted formulas or statements, and whose validity is thereby proved.

Corollary: A formula or statement that is immediately deducible from a proven theorem, and that requires little or no additional proof of validity.

Lemma: A proven auxiliary proposition stated for the expressed purpose of facilitating the proof of another proposition of more fundamental interest.

Thus, in the development of the theory of probability and mathematical statistics, axioms are the fundamental truths that are to be accepted at face value and not proven. Theorems and their corollaries are statements deducible from the fundamental truths and other proven statements and thus are *derived truths*. Lemmas represent proven results, often from fields outside of statistics per se, that are used in the proofs of other results of more primary interest.

We elaborate on the concept of a lemma, since our discussions will implicitly rely on lemmas more than any other type of statement, but we will generally choose not to exhaustively catalogue lemmas in the discussions. What constitutes a lemma and what does not depends on the problem context or one's point of view. *A* fundamental integration result from calculus could technically be referred to as a lemma when used in a proof of a statement in mathematical statistics, while in the study of calculus, it might be referred to as a theorem to be proved in and of itself. Since our study will require numerous auxiliary results from algebra, calculus, and matrix theory, exhaustively cataloging these results as lemmas would be cumbersome, and more importantly, not necessary given the prerequisites assumed for this course of study, namely, a familiarity with the basic concepts of algebra, univariate and multivariate calculus, and an introduction to matrix theory. We will have occasion to state a number of lemmas, but we will generally reserve this label for more exotic mathematical results that fall outside the realm of mathematics encompassed by the prerequisites.

A3. Elements of Set Theory

In the study of probability and mathematical statistics, sets are the fundamental objects to which probability will be assigned, and it is important that the concept of a set, and operations on sets, be well understood. In this section we review some basic properties of and operations on sets. This begs the following question: What is meant by the term *set*? In modern axiomatic developments of set theory, the concept of a set is taken to be primitive and incapable of being defined in terms of more basic ideas. For our purposes, a more intuitive notion of a set will suffice, and we avoid the complexity of an axiomatic development of the theory (see Marsden,[1] Appendix A, for a brief introduction to the axiomatic development). We base our definition of a set on the intuitive definition originally proposed by the founder of set theory, Georg Cantor (1845–1918).[2]

[1]J.E. Marsden, (1974), *Elementary Classical Analysis*, San Francisco: Freeman and Co.

[2]Cantor's original text reads: "Unter einer 'Menge' verstehen wir jede Zusammenfassung M von bestimmten wohlunterschiedenen Objekten m unserer Anschauung oder unseres Denkens (welche die 'Elemente' von M genannt werden) zu einem ganzen," (*Collected Papers*, p. 282). Our translation of Cantor's definition is, "By set we mean any collection, *M*, of clearly defined, distinguishable objects, *m*, (which will be called elements of *M*) which from our perspective or through our reasoning we understand to be a whole."

Definition A.1
Set

A **set** is a collection of objects with the following characteristics:

1. All objects in the collection are *clearly defined*, so that it is evident which objects are members of the collection and which are not;
2. All objects are *distinguishable*, so that objects in the collection do not appear more than once;
3. *Order is irrelevant* regarding the listing of objects in the collection, so two collections that contain the same objects but are listed in different order are nonetheless the *same set*; and
4. *Objects in the collection can be sets themselves,* so that a *set of sets* can be defined.

The objects in the collection of objects comprising a set are its **elements**. The term **members** is also used to refer to the objects in the collection. In order to signify that an object belongs to a given set, the symbol $\in$, will be used in an expression such as $x \in A$, which is to be read "x is an element (or member) of the set A." If an object is *not* a member of a given set, then a slash will be used as $x \notin A$ to denote that "x is not an element (or member) of the set A." Note that the slash symbol, /, is used to indicate negation of a relationship. The characteristics of sets presented in Definition A.1 will be clarified and elaborated upon in examples and discussions provided in subsequent subsections.

Set Defining Methods

Three basic methods are used in defining the objects in the collection constituting a given set: (1) **exhaustive listing;** (2) **verbal rule;** and (3) **mathematical rule**. An *exhaustive listing* requires that each and every object in a collection be individually identified either numerically, if the set is a collection of numbers, or by an explicit verbal description, if the collection is not of numbers. The object descriptions are conventionally separated by commas, and the entire group of descriptions is enclosed in brackets. The following are examples of sets defined by an exhaustive listing of the objects that are elements of the set:

Example A.1 $S_1 = \{\text{HEAD, TAIL}\}$ Here S_1 is the set of possible occurrences when tossing a coin into the air and observing its resting position. Note the set can be equivalently represented as $S_1 = \{\text{TAIL, HEAD}\}$. □

Example A.2 $S_2 = \{1,2,3,4,5,6\}$ Here S_2 is the set of positive integers from 1 to 6. Note that the set S_2 can be equivalently represented by the listing of the positive integers 1 to 6 in any order. □

A *verbal rule* is a verbal statement of characteristics that only the objects that are elements of a given set possess and that can be used as a test to determine set membership. The general form of the verbal rule is $\{x\text{: } \textit{verbal statement}\}$, which is to be read "the collection of all x for which *verbal statement* is true." The following are examples of sets described by verbal rules:

Example A.3 $S_3 = \{x: x \text{ is a college student}\}$ Here S_3 is the set of college students. An individual is an element of the set S_3 *iff* (if and only if) he or she is a college student. □

Example A.4 $S_4 = \{x: x \text{ is a positive integer}\}$ Here S_4 is the set of positive integers 1,2,3, ... A number is an element of the set S_4 *iff* it is a positive integer. □

A *mathematical rule* is of the same general form as a verbal rule, except the verbal statement is replaced by a mathematical expression of some type. The general form of the mathematical rule is {x: *mathematical expression*}, which is to be read "the collection of all x for which *mathematical expression* is true." The following are examples of sets described by mathematical rules:

Example A.5 $S_5 = \{x: x = 2k + 1, k = 0,1,2,3,\ldots\}$ Here S_5 is the set of odd positive integers. A number is an element of the set S_5 *iff* the number is equal to $2k + 1$ for some choice of $k = 0,1,2,3,\ldots$. □

Example A.6 $S_6 = \{x: 0 \leq x \leq 1\}$ Here S_6 is the set of numbers greater than or equal to 0 but less than or equal to 1. A number is an element of the set S_6 *iff* it is neither less than 0 nor greater than 1. □

The choice of method for describing the objects that constitute elements of a set depends on what is convenient and/or feasible for the case at hand. For example, exhaustive listing of the elements in set S_6 is impossible. On the other hand, there is some discretion that can be exercised, since, for example, a verbal rule could have adequately described the set S_5, say as $S_5 = \{x: x \text{ is an odd positive integer}\}$. A mixing of the basic methods might also be used, such as $S_5 = \{x: x = 2k + 1, k \text{ is zero or a positive integer}\}$. One can choose whatever method appears most useful in a given problem context.

Note that although our preceding examples of verbal and mathematical rules treat x as inherently one-dimensional, a vector interpretation of x is clearly permissible. For example, we can represent the set of points on the boundary or interior of a circle centered at (0,0) and having radius 1 as

$$S_7 = \left\{(x_1, x_2) : x_1^2 + x_2^2 \leq 1\right\}$$

or we can represent the set of input–output combinations associated with a two-input Cobb-Douglas production function as

$$S_8 = \left\{(y, x_1, x_2) : y = b_0\ x_1^{b_1}\ x_2^{b2}, x_1 \geq 0, x_2 \geq 0\right\}$$

for given numerical values of b_0, b_1, and b_2. Of course, the entries in the $\boldsymbol{x}$-vector need not be numbers, as in the set

$$S_9 = \{(x_1, x_2) : x_1 \text{ is an economist},\ x_2 \text{ is an accountant}\}.$$

Set Classifications

Sets are classified according to the number of elements they contain, and whether the elements are countable. We differentiate between sets that have a finite number of elements and sets whose elements are infinite in number, referring to a set of the former type as a **finite set** and a set of the latter type as an **infinite set**. In terms of countability, sets are classified as being either countable or uncountable. Note that when we count objects, we intuitively place the objects in a one-to-one correspondence with the positive integers, i.e., we identify objects one by one, and count "1,2,3,4,... ." Thus, a **countable set** is one whose elements can be placed in a one-to-one correspondence with some or all of the positive integers-any other set is referred to as an **uncountable set**.

A finite set is, of course, always countable, and thus it would be redundant to use the term "countable finite set." Sets are thus either **finite**, **countably infinite**, or **uncountably infinite**. Of the sets, S_1 through S_9 described earlier, S_1, S_2, S_3, and S_9 are finite, S_4 and S_5 are countably infinite, and S_6, S_7, and S_8 are uncountably infinite (why?).

Special Sets, Set Operations, and Set Relationships

We now proceed to a number of definitions and illustrations establishing relationships between sets, mathematical operations on sets, and the notions of the universal and empty sets.

Definition A.2
Subset

A is a **subset** of B, denoted as $A \subset B$ and read *A is contained in B*, *iff* every element of A is also an element of B.

Definition A.3
Equality of Sets

Set A is equal to set B, denoted as $A = B$, *iff* every element of A is also an element of B, and every element of B is also an element of A, i.e., *iff* $A \subset B$ and $B \subset$ A.

Definition A.4
Universal Set

The set containing all objects under consideration in a given problem setting, and from which all subsets are extracted, is the **universal set**.

Definition A.5
Empty or Null Set

The set containing no elements, denoted by $\emptyset$, is called the **empty**, or **null set**.

Definition A.6
Set Difference

Given any two subsets, A and B, of a universal set, the set of all elements in A that are *not* in B is called the **set difference** between A and B, and is denoted by $A{-}B$. If $A \subset B$, then $A{-}B = \emptyset$.

Definition A.7
Complement

Let A be a subset of a universal set, Ω. The **complement** of the set A is the set of all elements in Ω that are not in A, and is denoted by $\bar{A}$. Equivalently, $\bar{A} = \Omega - A$.

Definition A.8
Union

Let A and B be any two subsets of a universal set, Ω. Then, the **union** of the sets A and B is the set of all elements in Ω that are in *at least one* of the sets A or B, it is denoted by $A \cup B$.

Definition A.9
Intersection

Let A and B be any two subsets of a specified universal set, Ω. Then the **intersection** of the sets A and B is the set of all elements in Ω that are *in both* sets A and B, and is denoted by $A \cap B$.

Definition A.10
Mutually Exclusive (or Disjoint) Sets

Subsets A and B of a universal set, Ω, are said to be **mutually exclusive** or **disjoint** sets *iff* they have no elements in common, i.e., *iff* $A \cap B = \emptyset$.

We continue to use the slash, /, to indicate negation of a relationship (recall that / was previously used to indicate the negation of $\in$). Thus, $A \not\subset B$ denotes that A is not a subset of B, and $A \neq B$ denotes that A is not equal to B. We note here (and we shall state later as a theorem) that it is a logical requirement that $\emptyset$ is a subset of any set A, since if $\emptyset$ does not contain any elements, it cannot be the case that $\emptyset \not\subset A$, since the negation of $\subset$ would require the existence of an element in $\emptyset$ that was not in A.

Example A.7 Let the universal set be defined as $\Omega = \{x\text{: } 0 \leq x \leq 1\}$ and define three additional sets as

$A = \{x : 0 \leq x \leq .5\}, B = \{x : .25 \leq x \leq .75\}$ and $C = \{x : .75 < x \leq 1\}$.

Then we can establish the following set relationships:

$$\begin{aligned}
\bar{B} &= \{x : 0 \leq x < .25 \text{ or } .75 < x \leq 1\},\\
A \cup C &= \{x : 0 \leq x \leq .5 \text{ or } .75 < x \leq 1\},\\
\bar{B} &\subset A \cup C,\\
C \cap A &= C \cap B = \emptyset,\\
\bar{C} &= A \cup B = \{x : 0 \leq x \leq .75\},\\
A - B &= \{x : 0 \leq x < .25\},\\
A \cap B &= \{x : .25 \leq x \leq .5\}.
\end{aligned}$$

□

Note that although our definitions of subset, equality of sets, set difference, complement, union, and intersection explicitly involve only two sets A and B, it is implicit that the concepts can be applied to more complicated expressions involving an arbitrary number of sets. For example, since $A \cap B$ is itself a set, we can form its intersection with a set C as $(A \cap B) \cap C$, or form the set difference, $(A \cap B) - C$, or establish that $(A \cap B) =$ or $\neq C$, and so on. The point is that the concepts apply to sets, which themselves may have been constructed from other sets via various set operations.

Example A.8 Let Ω, A, B, and C be defined as in Example A.7. Then the following set relationships can be established:

$$\begin{aligned} A \cup B \cup C &= \Omega = \{x : 0 \leq x \leq 1\}, \\ (A \cup C) \cap B &= \{x : .25 \leq x \leq .5\}, \\ (A \cap B) \cap C &= \left(\overline{A \cup B}\right) \cap \bar{C} = \emptyset, \\ (B \cup C) - A &= \{x : .5 < x \leq 1\}, \\ ((A \cup B) - (A \cap B)) &\subset \bar{C} \cap \left(\bar{A} \cup \bar{B}\right). \end{aligned}$$

Can $\subset$ be replaced with $=$ in the last relationship? □

It is sometimes useful to conceptualize set relationships through illustrations called *Venn diagrams* (named after the 19th-century English logician, John Venn). In a Venn diagram, the universal set is generally denoted by a rectangle, with subsets of the universal set represented by various geometric shapes located within the bounds of the rectangle. Figure A.1 uses Venn diagrams to illustrate the set relationships defined previously.

Rules Governing Set Operations

Operations on sets must satisfy a number of basic rules. We state these basic rules as theorems, although we will not take the time to prove them here. The reader may wish to verify the plausibility of some of the theorems through the use of Venn diagrams. One of DeMorgan's laws will be proved to illustrate the formal proof method for the interested reader.

Theorem A.1 ***Idempotency Laws*** $A \cup A = A$ and $A \cap A = A$

Theorem A.2 ***Commutative Laws*** $A \cup B = B \cup A$ and $A \cap B = B \cap A$

Theorem A.3 ***Associative Laws*** $(A \cup B) \cup C = A \cup (B \cup C)$ and $(A \cap B) \cap C = A \cap (B \cap C)$

Theorem A.4 ***Distributive Laws*** $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ and $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

Theorem A.5 ***Identity Elements for*** $\cap$ ***and*** $\cup$ $A \cap \Omega = A$ (Ω is the identity element for $\cap$) $A \cup \emptyset = A$ ($\emptyset$ is the identity element for $\cup$)

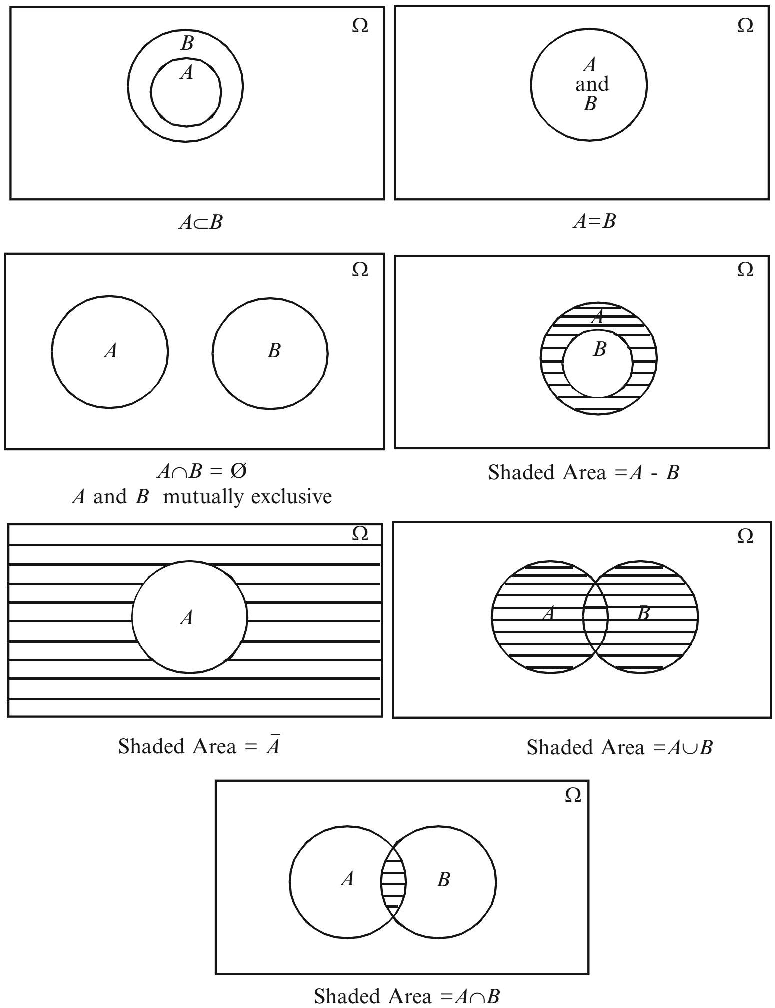


Figure A.1
Venn diagrams illustrating set relationships.

Theorem A.6 ***Intersection and Union of Complements*** $A \cup \bar{A} = \Omega$ and $A \cap \bar{A} = \emptyset$

Theorem A.7 ***Complements of Complements*** $\left(\bar{\bar{A}}\right) = A$

Theorem A.8 ***Intersection with the Null Set*** $A \cap \emptyset = \emptyset$

Theorem A.9 ***Null Set as a Subset*** If A is any set, then $\emptyset \subset A$

Theorem A.10 ***DeMorgan's Laws*** $(\overline{A \cup B}) = \bar{A} \cap \bar{B}$ and $(\overline{A \cap B}) = \bar{A} \cup \bar{B}$

Example A.9 Formal Proof of $(\overline{A \cap B}) = \bar{A} \cup \bar{B}$. By definition of the equality of sets, two sets are equal *iff* each is contained in the other. We first demonstrate that $\overline{A \cap B} \subset \bar{A} \cup \bar{B}$. By definition, $x \in \overline{A \cap B}$ implies that $x \notin A \cap B$. Suppose $x \notin \bar{A} \cup \bar{B}$. This implies

$x \notin \bar{A}$ and $x \notin \bar{B}$, which implies $x \in A$ and $x \in B$, i.e., $x \in A \cap B$, a contradiction. Therefore, if $x \in \overline{A \cap B}$, then $x \in \bar{A} \cup \bar{B}$, which implies $\overline{A \cap B} \subset \bar{A} \cup \bar{B}$. We next demonstrate that $\bar{A} \cup \bar{B} \subset \overline{A \cap B}$. Let $x \in \bar{A} \cup \bar{B}$. Then $x \notin A \cap B$, for if it were, then $x \in A$ and $x \in B$, contradicting that x belongs to at least one of $\bar{A}$ and $\bar{B}$. However, $x \notin A \cap B$ implies $x \in \overline{A \cap B}$, and thus $\bar{A} \cup \bar{B} \subset \overline{A \cap B}$. □

We remind the reader that since the sets used in Theorems A.1–A.10 could themselves be the result of set operations applied to other sets, the theorems are extendable in a myriad of ways to involve an arbitrary number of sets. For example, in the first of DeMorgan's laws listed as Theorem A.10, if $A = C \cup D$ and $B = E \cup F$, then by substitution,

$$\left(\overline{C \cup D \cup E \cup F}\right) = \left(\overline{C \cup D}\right) \cap \left(\overline{E \cup F}\right).$$

Then by applying Theorem A.10 to both $(C \cup D)$ and $(E \cup F)$, we obtain a generalization of DeMorgan's law as

$$\left(\overline{C \cup D \cup E \cup F}\right) = \bar{C} \cap \bar{D} \cap \bar{E} \cap \bar{F}.$$

Given the wide range of extensions that are possible, Theorems A.1–A.10 provide a surprisingly broad conceptual foundation for applying the rules governing set operations.

Some Useful Set Notation

Situations sometimes arise in which one is required to denote the union or intersection of a large number of sets. A convenient notation that represents such unions or intersections quite efficiently is available. Two types of notations are generally used, and they are differentiated on the basis of whether the union or intersection is of sets identified by a natural sequence of integer subscripts or whether the sets are identified by subscripts, say i's, that are elements of some set of subscripts, I, called an **index set**.

Definition A.11 ***Multiple Union Notation***

a. $\cup_{i=1}^{n} A_i = A_1 \cup A_2 \cup A_3 \ldots \cup A_n$.

b. $\cup_{i \in I} A_i =$ union of all sets A_i for which $i \in I$.

Definition A.12 ***Multiple Intersection Notation***

a. $\cap_{i=1}^{n} A_i = A_1 \cap A_2 \cap A_3 \ldots A_n$.

b. $\cap_{i \in I} A_i =$ intersection of all sets A_i for which $i \in I$.

Example A.10 Let the universal set be defined as $\Omega = \{x\colon 0 \le x \le 1\}$, and examine the following subsets of Ω:

$$A_1 = \{x:\ 0 \le x \le .25\}, A_2 = \{x:\ 0 \le x \le .5\},$$

$A_3 = \{x: \ 0 \leq x \leq .75\}, A_4 = \{x: \ .75 \leq x \leq 1\}$.

Define the index sets I_1 and I_2 as

$I_1 = \{1, \ 3\}$, and $I_2 = \{1, \ 3, \ 4\}$.

Then,

$$\begin{aligned}
&\cup_{i=1}^4 A_i = \cup_{i=2}^4 A_i = \cup_{i=3}^4 A_i = \{x : 0 \leq x \leq 1\} = \Omega, \\
&\cup_{i \in I_1} A_i = A_1 \cup A_3 = \{x : 0 \leq x \leq .75\}, \\
&\cup_{i \in I_2} A_i = A_1 \cup A_3 \cup A_4 = \{x : 0 \leq x \leq 1\} = \Omega, \\
&\cap_{i=1}^4 A_i = \cap_{i=2}^4 A_i = \emptyset, \\
&\cap_{i=3}^4 A_i = \{.75\}, \\
&\cap_{i \in I_1} A_i = A_1 \cap A_3 = \{x : 0 \leq x \leq .25\}, \\
&\cap_{i \in I_2} A_i = A_1 \cap A_3 \cap A_4 = \emptyset.
\end{aligned}$$

□

Whenever a set A is an interval subset of the *real line* (where the **real line** refers to all of the numbers between $-\infty$ and ∞), the set can be indicated in abbreviated form by the standard notation for intervals, stated in the following definition.

Definition A.13
Interval Set Notation

Let a and b be two numbers on the real line for which $a < b$. Then the following four sets, called intervals with endpoints a and b, can be defined as:

(a) Closed interval:

$[a, b] = \{x : a \leq x \leq b\}$,

(b) Half-open (or half-closed) intervals:

$(a, b] = \{x : a < x \leq b\}$, and

$[a, b) = \{x : a \leq x < b\}$,

(c) Open interval:

$(a, b) = \{x : a < x < b\}$.

Note that *weak* inequalities, $x \leq$ or $\leq x$, are signified by brackets] or [, respectively. *Strong* inequalities, $x <$ or $< x$, are signified by parentheses,) or (, respectively. Note further that whether the interval set contains its endpoints determines whether the set is closed.

As we have already done, (x,y) will also be used to denote coordinates in the two-dimensional plane. The context of the discussion will make clear whether we are referring to an open interval (a,b) or pair of coordinates (x,y).

A4. Relations, Point Functions, Set Functions

The concepts of point function and set function are central to a discussion of probability and statistics. We will see that probabilities can be represented by set functions and that in a large number of cases of practical interest, set functions can in turn be represented by a summation or integration operation applied to point functions. While readers may be somewhat familiar with point functions from introductory courses in algebra, the concept of a set function may not be familiar. We will review both function concepts within the broader context of the theory of relations. The relations context facilitates the presentation of a very general definition of "function" in which inputs into and outputs from the function may be objects of any kind including, but not limited to, numbers. The relations context also facilitates a demonstration of the significant similarities between the concept of a set function and the more familiar point function concept.

Cartesian Product

The concept of a relation can be made clear once we define what is meant by the Cartesian product of two sets A and B, named after the French mathematician Rene Descartes, (1596–1650).

Definition A.14 ***Cartesian Product of A and B***

Let A and B be two sets. Then the **Cartesian product** of A and B, denoted as $A \times B$, is the set of ordered pairs $A \times B = \{(x,y)\colon x \in A,\ y \in B\}$.

In words, $A \times B$ is the set of all possible pairs (x,y) such that x is an element of the set A and y is an element of the set B. Note carefully that the pairs are *ordered* in the sense that the first object in the pair must come from set A and the second object from set B.

Example A.11 Let $A = \{x\colon 1 \le x \le 2\}$ and $B = \{y\colon 2 \le y \le 4\}$. Then $A \times B = \{(x,y)\colon 1 \le x \le 2$ and $2 \le y \le 4\}$. (see Figure A.2). □

Example A.12 Let $A = \{x\colon x$ is a man$\}$ and $B = \{y\colon y$ is a woman$\}$. Then $A \times B = \{(x,y)\colon x$ is a man and y is a woman$\}$, which is the set of all possible man-woman pairings. □

In later chapters of the book we will have use for a more general notion of Cartesian product involving more than just two sets. The extension is given in the definition below.

Definition A.15 ***Cartesian Product (General)***

Let $A_1,\ldots,A_n$ be n sets. Then the Cartesian product of $A_1,\ldots,A_n$ is the set of ordered n-tuples

$$\times_{i=1}^{n} A_i = A_1 \times A_2 \times \ldots \times A_n = \{(x_1,\ldots,x_n) : x_i \in A_i, i = 1,\ldots,n\}.$$

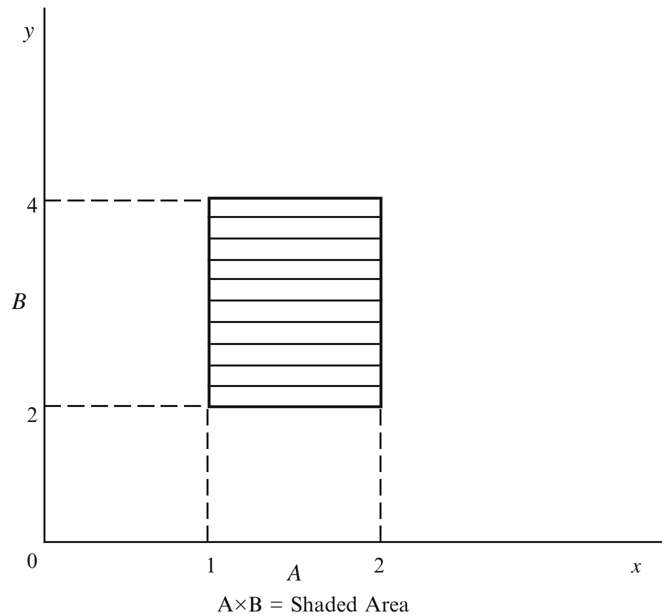


Figure A.2
$A \times B =$ shaded area.

In words, $\times_{i=1}^{n} A_i$ is the set of all possible n-tuples $(x_1,\ldots,x_n)$ such that x_1 is an element of set A_1, x_2 is an element of set A_2, and so on. Note that should the need arise, a general Cartesian product of sets could also be represented by the notation $\times_{i\in I} A_i$, where here the product is taken over all sets having subscript i in the index set I (recall Definitions A.11 and A.12, and the use of index set notation).

In certain cases, we may be interested in forming a Cartesian product of a set A with itself. While we might represent such a Cartesian product by the notation $\times_{i=1}^{n} A = \{(x_1,\ldots,x_n): x_i \in A,\ i = 1,\ldots,n\}$, such a Cartesian product is generally denoted by the notation A^n, and so for example, $A^2 = A \times A$.

Relation (Binary)

We now define what we mean by the term "binary relation."

Definition A.16 ***Binary Relation***

Any subset of the Cartesian product $A \times B$ is a **binary relation from A to B**.

Note that the adjective *binary* signifies that only two sets are involved in the relation. Relations that involve more than two sets can be defined, but for our purposes the concept of a binary relation will suffice.[3] Henceforth, we will use the word relation to mean binary relation. We should also mention that in the

[3] A higher order relation could be defined by taking a subset of the Cartesian product $\times_{i=1}^{n} A_i$, for example.

case where $B = A$, we will simply remain consistent with Definition A.16 and refer to a subset of $A \times A$ as a *relation from A to A*, although in this special case some authors prefer to call the subset of $A \times A$ a relation *on A*.

Now let $S \subset A \times B$. Thus by definition, S is a relation from A to B. We emphasize at this point that the choice of the letter S is quite arbitrary, and we could just as well have chosen any other letter to represent a subset of $A \times B$ defining a relation from A to B. If $(x,y) \in S$, we say that **x is in the relation** S **to** y or that **x is** S**-related to** y. An alternative notation for $(x,y) \in S$ is xSy. Also, we use S: $A \rightarrow B$ as an abbreviation for "the relation S from A to B."

As it now stands, the concept of a relation no doubt appears quite abstract. However, in practice, it is the context provided by the definition of the subset S and the definitions of the sets A and B that provide intuitive meaning to xSy. That is, x will be S-related to y because of some property satisfied by the (x,y) pair, the property being indicated in the set definition of S. The real-world objects being related will be clearly identified in the set definitions of A and B. A few examples will clarify the intuitive side of the relation concept.

Example A.13 Let $A = [0,\infty)$, and form the Cartesian product $A^2 = \{(x,y)$: $x \in A$ and $y \in A\}$. The set A^2 can thus be interpreted as the nonnegative (or first) quadrant of the Euclidean plane. Then $S = \{(x,y)$: $x \geq y$, $(x,y) \in A^2\}$ is a relation from A to A representing the set of points in the nonnegative quadrant for which the first coordinate has a value greater than or equal to the value of the second coordinate. The defining property of the relation S is "$\geq$." This is displayed in Figure A.3. □

Example A.14 Let $A = \{x$: x is an employed U.S. citizen$\}$ and $B = \{y$: y is a U.S. corporation$\}$. Then $A \times B = \{(x,y)$:x is an employed U.S. citizen and y is a U.S. corporation$\}$ is the set of all possible pairings of employed U.S. citizens with U.S. corporations. The relation $S = \{(x,y)$: *x is employed by y*, $(x,y) \in A \times B\}$ from A to B is the collection of U.S. citizens who are employed by U.S. corporations paired with their respective corporate affiliation. The defining property of the relation is the phrase "is employed by," and xSy *iff* x is a U.S. citizen employed by a U.S. corporation, and y is his or her corporate affiliation. □

Function

We are now in a position to define what is meant by the concept of a function. As indicated in the following definition, a function is simply a special type of relation. In the definition we introduce the symbol $\forall$, which stands **for every** or **for all**, the symbol $\exists$ which means **there exists.**

Definition A.17 ***Function***

> A function from A to B is a relation S: $A \rightarrow B$ such that $\forall a \in A$ $\exists$ one unique $b \in B$ such that $(a,b) \in S$.

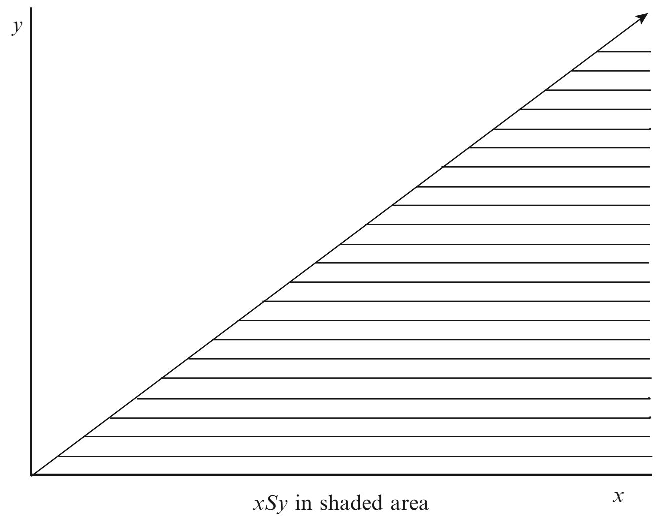


Figure A.3
xSy in shaded area.

A relation satisfying the above condition will often be given a special symbol to distinguish the relation as a function. A popular symbol used for this purpose is "f," where f: $A \rightarrow B$ is a common notation for designating the **function *f* from *A* to *B***. As we had remarked when choosing a symbol to represent a relation, the choice of the letter f is arbitrary, and when it is convenient or useful, any other letter or symbol could be used to depict a subset of $A \times B$ that represents a function from A to B. In the text we will often have occasion to use a variety of letters to designate various functions of interest.

The unique element $b \in B$ that the function f: $A \rightarrow B$ associates with a given element $x \in A$ is called **the image of *x* under** f and is represented symbolically by the notation **$f(x)$**. If $f(x)$ is a real number, the image of x under f is alternatively referred to as **the value of the function *f* at *x***. In the following example, we use $\mathbb{R}$ to denote the set of real numbers $(-\infty, \infty)$, i.e., $\mathbb{R}$ stands for the **real line**. Furthermore, the nonnegative subset of the real line is represented by $\mathbb{R}_{\geq 0}$, $= [0,\infty)$.

Example A.15 Let f: $\mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ be defined by $f = \{(x,y): y = x^2, x \in \mathbb{R}\}$. The *image of* -2 *under* f is $f(-2) = 4$. The *value of the function f at* 3 is $f(3) = 9$. □

Associated with a given function, f, are two important sets called the **domain** and **range** of the function.

Definition A.18
Domain and Range of a Function

The **domain** of a function f: $A \rightarrow B$ is defined as $D(f) = A$. The **range** of f is defined by $R(f) = \{y: y = f(x), x \in A\}$.

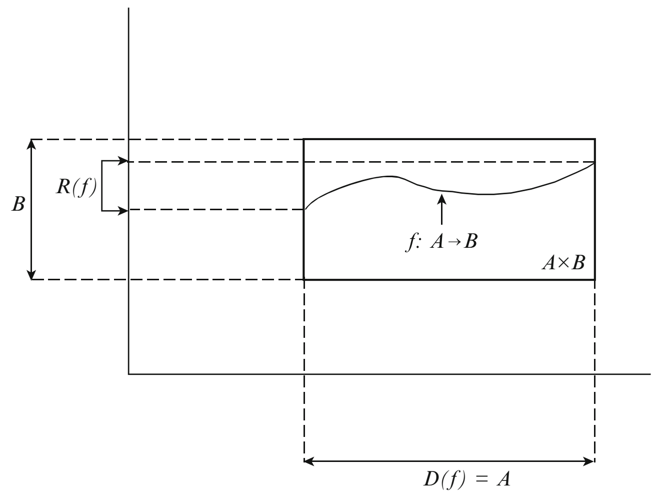


Figure A.4
Function with domain $D(f)$ and range $R(f)$.

Thus, the domain of a function f: $A \rightarrow B$ is simply the set A of the Cartesian product $A \times B$ associated with the function. The range of f is the collection of all elements in B that are images of the elements in A under the function f. It follows that $R(f) \subset B$. Figure A.4 provides a pictorial example of a function, including its *domain* and *range*.

Note that the concept of a function is completely general regarding the nature of the elements of the sets $D(f) = A$, $R(f)$, and B. The elements can be numbers, or other objects, or the elements can be sets themselves. For our work, it will suffice to deal only with *real-valued functions* meaning that $R(f)$ is a set of real numbers.

Definition A.19
Real-Valued Function

A function f: $A \rightarrow B$ such that $R(f) \subset \mathbb{R}$ is called a **real-valued function**.

The function defined in Example A.15 is a real-valued function. The following is another example.

Example A.16

Examine the Cobb-Douglas production function $f{:}\mathbb{R}^2_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ defined as $f = \left\{((x_1, x_2), y) : y = 10x_1^2 x_2, (x_1, x_2) \in \mathbb{R}^2_{\geq 0}\right\}$. Interpreting (x_1,x_2) as inputs into a production process, and y as the output of the process, we see that the domain of the production function is $d(f) = \mathbb{R}^2_{\geq 0}$, i.e., any nonnegative level of the input pair (x_1,x_2) is an admissible input level. The associated range of the production function is $\mathbb{R}(f) = \mathbb{R}_{\geq 0}$, i.e., any nonnegative level of output, y, is possible. Since $\mathbb{R}(f) = [0,\infty) \subset \mathbb{R}$, the production function is a real-valued function. □

In some cases the relation from A to B that defines the function f: $A \rightarrow B$ also defines a function from B to A. This relates to the concept of an **inverse function**. In particular, if for each element $y \in B$ there exists precisely one element $x \in A$ whose image under f is y, then such an *inverse function* exists.

Definition A.20
Inverse Function of f

Let f: $A \rightarrow B$ be a function from A to B. If $\mathbb{R}(f) = B$ and $\forall y \in B$ $\exists$ a unique $x \in A$ such that $y = f(x)$, then the relation $\{(y,x)$: $y = f(x)$, $y \in B\}$ is a function from B to A called the **inverse function of *f*** and denoted by f^{-1}: $B \rightarrow A$.

Note that neither of the functions in Example A.15 or Example A.16 are such that an inverse function exists. In Example A.15, the uniqueness condition of Definition A.20 is violated since $\forall y \neq 0$ there exist two values of x for which $y = x^2$, namely $x = \pm\sqrt{y}$. For example, when $y = 4$, $x = 2$ and -2 are each such that $y = x^2$. The reader can verify that Example A.16 also violates the uniqueness condition where an infinite number of (x_1, x_2) values satisfy $y = 10x_1^2x_2$ for a fixed value of y (defining level sets or isoquants of the production function). Also, note that an inverse function does not exist for the function illustrated in Figure A.4.

As an example of a function for which an inverse function does exist, consider the following.

Example A.17 Let f: $\mathbb{R} \rightarrow \mathbb{R}_+$ be defined as

$f = \{(x,y) : y = e^x, x \in \mathbb{R}\},$ where $\mathbb{R}_+ = (0,\infty)$.

Note that the *inverse function* can be represented as

$f^{-1} = \{(y,x) : x = \ln(y), y \in \mathbb{R}_+\}$

so that f^{-1}:$\mathbb{R}_+ \rightarrow \mathbb{R}$. It is clear that $\forall y \in \mathbb{R}_+$, $\exists$ one and only one $x \in \mathbb{R}$ such that $x = \ln(y)$. □

The final concept concerning functions that we will review here is the **inverse image** of $y \in R(f)$ or of $H \subset R(f)$. The inverse image of y is the set of domain elements $x \in D(f)$ such that $y = f(x)$, i.e., the collection of all x values in the domain of f whose image under the function f is y. The inverse image of y can be represented as the set $\{x$: $f(x) = y\}$, and when the inverse function exists, the inverse image of y can be represented as the value $f^{-1}(y)$. In Example A.17 above, the inverse image of 5 is $f^{-1}(5) = \ln(5) = 1.6094$, in Example A.15, the inverse image of 4 is $\{-2,2\}$, and in Example A.16, the inverse image of 3 is the isoquant $\left\{(x_1, x_2) : 3 = 10x_1^2x_2, (x_1, x_2) \in \mathbb{R}^2_{\geq 0}\right\}$. Similarly, the inverse image of $H \subset R(f)$ is the set of x values in the domain of f whose images under f equal some $y \in H$, i.e., the inverse image of H is $\{x$: $f(x) = y$, $y \in H\}$, and if the inverse function exists, $\{x$: $x = f^{-1}(y)$, $y \in H\}$.

Real-Valued Point Versus Set Functions

Two types of functions – point functions and set functions – are utilized extensively in modern discussions of probability and mathematical statistics. The reader should already have considerable experience with the application of real-valued point functions, since this type of function is the one that appears in elementary algebra and calculus courses and is central to discussions of utility, demand, production, and supply that the reader has encountered in his or her study of economic theory. Specifically, a **real-valued point function** is a

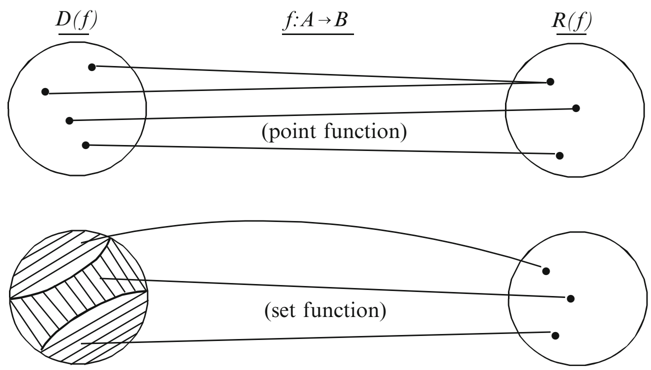


Figure A.5
Point versus set function.

real-valued function whose domain consists of a collection of points, where points are represented by coordinate vectors in $\mathbb{R}^n$. We have encountered examples of this type of function previously in Example A.15 through Example A.17. A typical ordered pair associated with a real-valued point function is of the form $(\mathbf{x},y)$, where $\mathbf{x}$ is a vector in $\mathbb{R}^n$ and y is a real number in $\mathbb{R}$.

A set function is more general than a point function in that its domain consists of a collection of *sets* rather than a collection of points.[4] A typical ordered pair belonging to a **real-valued set function** would have the form (A,y), where A is a set of some type of objects and y is a real number in $\mathbb{R}$. If the sets in the domain of the set function are contained in $\mathbb{R}^n$, i.e., they are sets of real numbers, then a real-valued set function is assigns a real number to each *set* of points in its domain in contrast to a real-valued point function which would assign a real number to each *point* in its domain. A pictorial illustration of a set function contrasted with a point function is given in Figure A.5.

Examples of set functions are presented below.

Example A.18 Let $\Omega = \{1,2,3\}$, and let A be the collection of all of the subsets of Ω, i.e., $A = \{A_1, A_2, \ldots, A_8\}$, where

$$A_1 = \{1\}, A_2 = \{2\}, A_3 = \{3\}, A_4 = \{1,2\}, A_5 = \{1,3\}, A_6 = \{2,3\},$$
$$A_7 = \{1,2,3\}, \text{ and } A_8 = \emptyset$$

The following is a real-valued set function $f\colon A \to \mathbb{R}$:

$$f = \left\{ (A_i, y) : y = \sum_{x \in A_i} x,\ A_i \subset A \right\},$$

where $\sum_{x \in A_i} x$ signifies the sum of the numerical values of all of the elements in the set A_i, and $\sum_{x \in \emptyset} x$ is defined to be zero. The range of the set function is

[4] Note that, in a sense, a point function can be viewed as a special case of a set function, since points can be interpreted as singleton (single element) sets. The set function concept is introduced to accommodate the case where one or more sets in its domain are *not* singleton.

$R(f) = \{0,1,2,3,4,5,6\}$ and the domain is the set of sets $D(f) = A$. The function can be represented in tabular form as

A_i	$f(A_i)$
A_1	1
A_2	2
A_3	3
A_4	3
A_5	4
A_6	5
A_7	6
A_8	0

□

Example A.19 Let $A = \{A_r\colon A_r = \{(x,y)\colon x^2 + y^2 \le r^2\},\ r \in [0,1]\}$, so that A is a set of sets, with typical element A_r represents the set of points in $\mathbb{R}^2$ that are on the boundary and in the interior of a circle centered at (0,0) with radius r. The following is a real valued set function $f\colon A \to \mathbb{R}$:

$$f = \{(A_r, y) : y = \pi r^2,\ A_r \subset A\}.$$

Note the set function assigns a real number representing the area to each set, A_r. The assignment is made for circles having a radius anywhere from 0 to 1. The range of the set function is given by $\mathbb{R}(f) = [0,\pi]$, and the domain is the *set of sets* $D(f) = A$. □

A special type of set function called the **size-of-set function** will prove to be quite useful.

Definition A.21 ***Size of Set Function***

Let A be any set of objects. The **size-of-set function**, N, is the set function that assigns to the set A the number of elements that are in set A, i.e., $N(A) = \sum_{x\in A} 1$.[5]

Applying the size-of-set function in Example A.18, note that $N(A) = 8$. In Example A.19 note that $N(A) = \infty$.

Another special (point) function that will be useful in our study is the **indicator function**, defined as follows:

Definition A.22 ***Indicator Function***

Let A be any subset of some universal set Ω. The indicator function, denoted by I_A, is a real-valued function with domain Ω and range $\{0, 1\}$ such that

$$I_A(x) = \left\{ \begin{matrix} 1 & \text{if} & x \in A \\ 0 & \text{if} & x \notin A \end{matrix} \right\}.$$

[5] Note that $\sum_{x\in A} 1$ signified that a collection of 1's are being summed together, the number in the collection being equal to the number of elements in the set A. If $A = \emptyset$, effectively no 1's are being added together, and thus $N(\emptyset) = 0$.

Note that the indicator function *indicates* the set A by assigning the number 1 to any x that is an element of A, while assigning zero to any x that is not an element of A. The main use of the indicator function is notational efficiency in defining functions, as the following example illustrates:

Example A.20 Let the function $f:\mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = \begin{bmatrix} 0 & \text{for} & x \in (-\infty, 0] \\ x & \text{for} & x \in (0, 2] \\ 3 - x & \text{for} & x \in (2, 3] \\ 0 & \text{for} & x \in (3, \infty) \end{bmatrix}$$

Utilizing the indicator function, we can alternatively represent $f(x)$ as

$$f(x) = xI_{(0,2]}(x) + (3 - x)I_{(2,3]}(x).$$ □

As a final note on the use of functions, we (as do the vast majority of other authors) will generally use a shorthand method for defining functions by simply specifying the relationship between elements in the domain of a function and their respective images in the range of the function. For example, we would define the function in Example A.19 by $f(A_r) = \pi r^2$ for $A_r \subset A$, or define the function in Example A.15 by $f(x) = x^2$ for $x \in \mathbb{R}$. In all cases, the reader should remember that a function is a set of ordered pairs $(x,f(x))$, or $(A,f(A))$. The reader will sometimes find in the literature phrases like **the function *f*(*x*)** or **the set function *f*(*A*)**. Literally speaking, such phrases are inconsistent, because $f(x)$ or $f(A)$ are not functions, but rather *images* of elements in the domain of the respective functions. In fact, the reader should *not* take such phrases literally, but rather interpret these phrases as shorthand for phrases such as *the function whose values are given by* $f(x)$ or *the set function whose values are given by* $f(A)$.

A5. Combinations and Permutations

In a number of situations involving probability assignments, it will be useful to have an efficient method for counting the number of *different* ways a group of r objects can be selected from a group of n distinct objects, $n \geq r$. Obviously, if two groups of r objects do not contain the same r objects, they must be considered *different*. But what if two groups of r objects do contain the same objects, except the objects in the group are arranged in different orders? Are the two groups to be considered *different*? If difference in order constitutes difference in groups, then we are dealing with the notion of **permutations**. On the other hand, if the order of listing the objects in a group is not used as a basis for distinguishing between groups, then we are dealing with the notion of **combinations**.

In order to establish a formula for determining the number of permutations of n distinct objects taken r at a time, the following example is suggestive.

Example A.21 Examine the number of different ways a group of three letters can be selected from the letters, *a,b,c,d*, where difference in order of listing is taken to mean difference in groups. Note that the first letter can be chosen in four different ways. After we have chosen one of the letters for the first selection, the second selection can be any of the remaining three letters. Finally, after we have chosen two letters in the first two selections, there are then two letters left to be potentially chosen for the third selection. Thus, there are $4 \cdot 3 \cdot 2 = 24$ different ways of selecting a group of three letters from the letters *a,b,c,d* if difference in the order of listing constitutes difference in groups (The reader should attempt to list the 24 different groups).

The logic of the preceding example can be applied to establish a general formula for determining the number of permutations of n distinct objects taken r at a time:

$$(n)_r = \frac{n!}{(n-r)!} = n(n-1)(n-2)\cdots(n-r+1),$$

where ! denotes the **factorial operation**, i.e.,[6]

$$n! = n(n-1)(n-2)(n-3)\ldots 1.$$

So, for example, $4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$. In Example A.21, $n = 4$ and $r = 3$, so that $(4)_3 = 4!/1! = 24$.

In order to establish a formula for determining the number of combinations of n distinct objects taken r at a time, examine the number of different ways a group of three letters can be selected from the letters *a,b,c,d*, where difference in order of listing does *not* imply the groups are different. Recall that we discovered that there were 24 permutations of the four letters *a,b,c,d* selected three at a time. Now note that any three letters, say *a,b,c*, can be arranged in $(3)_3 = 6$ different orders, which represents "overcounting" from the combinations point of view. Reducing the number of permutations by the degree of "overcounting" results in the number of combinations, i.e., there are $24/6 = 4$ combinations of the four letters taken three at a time, namely (a,b,c), (a,b,d), (a,c,d), and (b,c,d). □

In the preceding example, the number of permutations of $n(=4)$ objects taken $r(=3)$ at a time was reduced by a factor of $r!(=3!)$, where the latter value represents the number of possible permutations of r objects. This suggests the general formula for the number of combinations of n objects taken r at a time:

$$\binom{n}{r} = \frac{(n)_r}{r!} = \frac{n!}{(n-r)!\,r!}$$

[6]By definition, we take $0! = 1$.

In Example A.21, we have

$$\binom{4}{3} = \frac{4!}{1!\,3!} = 4$$

as the appropriate number of combinations.

The concept of combinations is useful in determining the number of subsets that can be constructed from a finite set A. Note that in counting the number of subsets, changes in the order of listing set elements does not produce a different set, e.g., the sets $\{a,b,c\}$ and $\{c,a,b\}$ are the same set of letters (recall the definition of a set). Then the total number of subsets of a set A containing n elements is given by the number of different subsets defined by taking no elements (i.e., the null set) plus the number of different subsets defined by taking one element, plus the number of different subsets defined by taking two elements, ..., and finally, the number of different subsets defined by taking all n elements (i.e., the set A itself). Thus, the total number of different subsets of A can be written as

$$\sum_{r=0}^{n} \binom{n}{r} = \sum_{r=0}^{n} \frac{n!}{(n-r)!\,r!}$$

This sum can be greatly simplified by recalling that

$$(x+y)^n = \sum_{r=0}^{n} \binom{n}{r} x^r y^{n-r},\, n = 1, 2, \ldots,$$

which is the **binomial theorem**. Then letting $x = y = 1$, we have that

$$2^n = \sum_{r=0}^{n} \binom{n}{r}$$

so that 2^n is the number of different subsets contained in a set A that has n elements. The set containing all 2^n subsets of a set, A, of n elements is called the **power set** of A.

Example A.22 ***Power Set*** In Example A.18, recall that we identified a total of eight subsets of the set $\Omega = \{1,2,3\}$. This is the number of subsets we would expect from our discussion above, i.e., since $n = 3$, there are $2^3 = 8$ subsets of Ω. □

It should be noted that $\binom{n}{r}$ is *defined* to be 0 whenever $n < r$, or whenever n and/or r are < 0 or are not integer valued. The rationale for $\binom{n}{r} = 0$ in each of these cases is that there is no way to define subsets of size r from a collection of n objects for the designated values of n and r.

When n is large, the calculation of $n!$ needed in the previous formulas pertaining to numbers of permutations or combinations can be quite formidable. A result known as Stirling's formula can provide a useful approximation to $n!$ for large n.

Table A.1 Summation and integration notation

Notation	Definition
$\sum_{i=\ell}^{n} x_i$	Sum the values of $x_\ell, x_{\ell+1},\ldots,x_n$, i.e., $x_\ell + x_{\ell+1} + \ldots + x_n$
$\sum_{i\in I} x_i$	Sum the values of the x_i's, for $i \in I$
$\sum_{x\in A} x$	Sum the values of $x \in A$
$\sum_{x=a}^{b} x$	Sum the values of x in the sequence of integers from a to b, i.e., $a + (a + 1) + (a + 2) + \ldots + b$
$\sum_{i=\ell}^{n} \sum_{j=k}^{m} x_{ij}$	Sum the values of the x_{ij}'s for $i = \ell, \ell + 1,\ldots,n$ and $j = k, k + 1,\ldots,m$
$\sum_{i\in I} \sum_{j\in J} x_{ij}$ or $\sum_{(i,j)\in A} x_{ij}$	Sum the values of the x_{ij}'s for $i \in I$ and $j \in J$, or for $(i,j) \in A$
$\sum_{x_1\in A_1} \cdots \sum_{x_n\in A_n} f(x_1,\ldots,x_n)$	Sum the values of $f(x_1,\ldots,x_n)$ for $x_i \in A_i$, $i = 1,\ldots,n$
$\sum_{(x_1,\ldots,x_n)\in A} f(x_1,\ldots,x_n)$	Sum the values of $f(x_1,\ldots,x_n)$ for $(x_1,\ldots,x_n) \in A$
$\sum_{x_1=a_1}^{b_1} \cdots \sum_{x_n=a_n}^{b_n} f(x_1,\ldots,x_n)$	Sum the values of $f(x_1,\ldots,x_n)$ for x_i in the sequence of integers a_i to b_i, $i = 1,\ldots,n$
$\int_a^b f(x)dx$	Integral of the function $f(x)$ from a to b (a and/or b can be $-\infty$ and ∞)
$\int_{x\in A} f(x)dx$	Integral of the function $f(x)$ over the set of points A
$\int_{x_1\in A_1} \cdots \int_{x_n\in A_n} f(x_1,\ldots,x_n)dx_n \ldots dx_1$	Iterated integral of the function $f(x_1,\ldots,x_n)$ over the points $x_i \in A_i$, $i = 1,\ldots,n$
$\int_{(x_1,\ldots x_n)\in A} f(x_1,\ldots,x_n)dx_1 \ldots dx_n$	Multiple integral of the function $f(x_1,\ldots,x_n)$ over the points $(x_1,\ldots,x_n) \in A$
$\int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f(x_1,\ldots,x_n)dx_n \ldots dx_1$	Iterated integral of the function $f(x_1,\ldots,x_n)$ for x_i in the (open, half open-half closed, or closed) interval a_i to b_i, for $i = 1,\ldots,n$

Definition A.23
Stirling's Formula[7]

$n! \approx (2\pi)^{1/2}\, n^{n+.5}\, e^{-n}$ for large n.[8]

A logical question to ask regarding the use of Stirling's formula is how large is "large n"? Stirling's formula invariably underestimates $n!$ but the percentage error is ≤ 1 percent for $n \geq 10$, and monotonically decreases as $n \to \infty$.

A6. Summation and Integration Notation

We will use a number of variations on summation and integration notation in this text. The meaning of the various types of notation are presented in Table A.1.

We illustrate the use of some of the notation in the following examples.

Example A.23
Summation Notation

Let $A_1 = \{1,2,3\}$, $A_2 = \{2,4,6\}$, $A = A_1 \times A_2$, $B = \{(x_1,x_2)\colon\ x_1 \in A_1;\ \ x_2 \in \{x_1, x_1 + 1,\ldots,3x_1\}\}$, $y = (y_1,y_2,\ldots,y_n)$, and $f(x_1,x_2) = x_1 + 2x_2^2$. Then

[7]See Feller (1968), *An Introduction to the Theory of Probability and Its Applications*, 3rd ed., pp. 52–54.

[8]Note that ≈ means "approximately equal to."

$$\sum_{x\in A_1} x = 1+2+3 = 6, \sum_{x\in A_2} x^2 = 2^2+4^2+6^2 = 56,$$

$$\sum_{i\in A_2} y_i = y_2+y_4+y_6, \sum_{i\in A_1} y_i = \sum_{i=1}^{3} y_i = y_1+y_2+y_3$$

$$\sum_{x_1\in A_1}\sum_{x_2\in A_2} f(x_1,x_2) = \sum_{x_1\in A_1}\sum_{x_2\in A_2}\left(x_1+2x_2^2\right) = 354,$$

$$\sum_{(x_1,x_2)\in A} f(x_1,x_2) = \sum_{x_1\in A_1}\sum_{x_2\in A_2} f(x_1,x_2) = 354,$$

$$\sum_{(x_1,x_2)\in B} f(x_1,x_2) = \sum_{x_1=1}^{3}\sum_{x_2=x_1}^{3x_1}\left(x_1+2x_2^2\right) = 802.$$

□

Example A.24
Integration Notation

Let $A_1 = [0,3]$, $A_2 = [2,4]$, $A = A_1 \times A_2$, $B = \{(x_1,x_2): x_1 \in A_1, 0 < x_2 < x_1^2\}$, and $f(x_1, x_2) = x_1x_2^2$. Then

$$\int_{x\in A_1} 2x dx = \int_0^3 2x dx = \frac{2x^2}{2}\Big|_0^3 = 9,$$

$$\int_{x\in A_1\cap A_2} x^2 dx = \int_2^3 x^2 dx = \frac{x^3}{3}\Big|_2^3 = \frac{19}{3},$$

$$\int_{x_1\in A_1}\int_{x_2\in A_2} f(x_1,x_2)dx_2dx_1 = \int_0^3\int_2^4 f(x_1,x_2)dx_2dx_1$$

$$= \int_0^3 \frac{x_1x_2^3}{3}\Big|_2^4 dx_1 = \int_0^3 \frac{56}{3}x_1dx_1$$

$$= \frac{56x_1^2}{6}\Big|_0^3 = 84,$$

$$\int_{(x_1,x_2)\in A} f(x_1,x_2)dx_1dx_2 = \int_{x_1\in A_1}\int_{x_2\in A_2} f(x_1,x_2)dx_2dx_1 = 84,$$

$$\int_{(x_1,x_2)\in B} f(x_1,x_2)dx_1dx_2 = \int_0^3\int_0^{x_1^2} f(x_1,x_2)dx_2dx_1$$

$$= \int_0^3 \frac{x_1x_2^3}{3}\Big|_0^{x_1^2} dx_1 = \int_0^3 \frac{x_1^7}{3}dx_1$$

$$= \frac{x_1^8}{24}\Big|_0^3 = 273.375.$$

□

Regarding matrix differentiation notation, we utilize the following conventions. Let $g(\mathbf{x})$ and $\mathbf{y}(\mathbf{x})$ be a scalar and $(n \times 1)$ vector function of the $(k \times 1)$ vector $\mathbf{x}$, respectively. Then

Derivative	Matrix dimension	***(i,j)*th entry**
$\frac{\partial g(\mathbf{x})}{\partial \mathbf{x}}$	$(k \times 1)$	$\frac{\partial g}{\partial x_i}$
$\frac{\partial g(\mathbf{x})}{\partial \mathbf{x}'} = \frac{\partial g(\mathbf{x})'}{\partial \mathbf{x}}$	$(1 \times k)$	$\frac{\partial g}{\partial x_j}$
$\frac{\partial g(\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}'}$	$(k \times k)$	$\frac{\partial^2 g}{\partial x_i \partial x_j}$
$\frac{\partial \mathbf{y}(\mathbf{x})}{\partial \mathbf{x}}$	$(k \times n)$	$\frac{\partial y_j}{\partial x_i}$
$\frac{\partial \mathbf{y}(\mathbf{x})}{\partial \mathbf{x}'} = \frac{\partial \mathbf{y}(\mathbf{x})'}{\partial \mathbf{x}}$	$(n \times k)$	$\frac{\partial y_i}{\partial x_j}$

A7. Elements of Real Analysis

In this section, we present a number of prerequisite results from real analysis that facilitate the development and understanding of various types of asymptotic probability behavior. In particular, the concepts of sequences, limits, continuity of a function, and orders of magnitude of a sequence will be examined.

Sequences of Numbers and Random Variables

We begin with the notion of a **sequence**. In the definition, we refer to the set of **natural numbers**, which is simply the set of positive integers in their natural order, 1, 2, 3, ...

Definition A.24 ***Sequence***

Let A be any set. A sequence in A is a function having the natural numbers, N, for its domain, and its range contained in A, i.e., $f\colon N \rightarrow A$, is a sequence in A.

When utilizing the concept of a sequence, we (and others) will often suppress the function aspect of its definition and concentrate on the ordered collection of image elements of the function. Thus, given a sequence defined by $\{(n,y)\colon y = f(n),\ n \in N\}$, we will equivalently refer to the collection of image elements $\{y_1, y_2, y_3, \ldots\}$ as the sequence, where $y_n = f(n)$. The subscripts on the elements of the set $\{y_1, y_2, y_3, \ldots\}$ serve to define the order of the elements in the sequence. Furthermore, we will utilize the notation $\{y_n\}$ as an abbreviation for the sequence $\{y_1, y_2, y_3, \ldots\}$[9] In the following examples of sequences, we continue to use N to denote the set of natural numbers.

[9] Another common abbreviated notation that is sometimes used to denote a sequence is given by (y_n). The notation we have adopted is more prevalent in the statistics literature. While there will be no confusion in this text, in general, the reader will have to rely on the context of a discussion to determine whether $\{y_n\}$ refers to a sequence or to a set containing the single element y_n.

Example A.25
Sequences in $\mathbb{R}$

(a) $\{2,4,8,\ldots\}$, which is defined by the function $y = 2^n$, $n \in N$.

(b) $\{1,1/3,1/9,1/27,\ldots\}$, which is defined by the function $y = (1/3)^{n-1}$, $n \in N$.
(c) $\{-3,-1,1,3,\ldots\}$, which is defined by the function $y = 2n - 5$, $n \in N$. □

Example A.26
A Sequence of Matrices

Let $\mathbf{x}_n$ be an $(n \times 2)$ matrix whose ith row is defined by the (1×2) vector $[1 \ i]$, so that

$$\mathbf{x}_n = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \\ 1 & n \end{bmatrix}.$$

Then

$$\left\{ \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & \frac{3}{2} \\ \frac{3}{2} & \frac{5}{2} \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 2 & \frac{14}{3} \end{bmatrix}, \ldots \right\}$$

is a sequence of *matrices* $\{\mathbf{y_1},\mathbf{y_2},\mathbf{y_3},\ldots\}$ defined by the function $\mathbf{y}_n = \frac{1}{n}\mathbf{x}'_n\mathbf{x}_n, n \in N$, where the nth element of the sequence is defined as

$$\mathbf{y}_n = \begin{bmatrix} 1 & \frac{\left(\sum_{i=1}^n i\right)}{n} \\ \frac{\left(\sum_{i=1}^n i\right)}{n} & \frac{\left(\sum_{i=1}^n i^2\right)}{n} \end{bmatrix} = \begin{bmatrix} 1 & \frac{(n+1)}{2} \\ \frac{(n+1)}{2} & \frac{(n+1)(2n+1)}{6} \end{bmatrix}.$$ □

In statistical practice one frequently encounters **sequences of random variables**. In this case, the set A in Definition A.24 is a collection of random variables, and the function f: $N \rightarrow A$ defining the sequence places the random variables in A in a specific order. That is, the sequence of random variables $\{Y_1, Y_2, Y_3,\ldots\}$ is simply an ordered collection of random variables. In our study of asymptotics, the elements in the sequence of random variables will often be defined as functions of other random variables, such as $Y_n = g_n(X_1,\ldots,X_n)$, and we will be interested in studying the characteristics of the sequence of probability distributions associated with the sequence of Y_n's as $n \rightarrow \infty$.

The following are examples of sequences of random variables. We introduce the notation $Y \sim f(y)$ to indicate that Y **has probability density** $f(y)$, or that Y **is distributed as** $f(y)$. This notation can also be used as $Y \sim F(y)$ to denote that Y has the CDF $F(y)$. The acronym ***iid*** used ahead stands for **independent and identically distributed**, meaning that the random variables in a collection are independent and each of the random variables has the same PDF or probability distribution.

Example A.27

Let $X_1,\ldots, X_n$ be *iid* random variables, each with PDF $N(\mu, \sigma^2)$, where X_i represents the miles per gallon obtained from the ith automobile of a certain type tested for fuel efficiency. Examine the sequence of random variables

$\{Y_1, Y_2, Y_3, \ldots\}$, where $Y_n = n^{-1} \sum_{i=1}^{n} X_i, n \in N$.

Note that the nth element of the sequence represents the *average* miles per gallon obtained from n of the automobiles tested, and

$$Y_n \sim N\left(\mu, \frac{\sigma^2}{n}\right),$$

so that we can define a **sequence of probability density functions** associated with the sequence of random variables as

$$\left\{N(\mu, \sigma^2), N\left(\mu, \tfrac{\sigma^2}{2}\right), N\left(\mu, \tfrac{\sigma^2}{3}\right), \ldots\right\}.$$ □

Example A.28 Let $X_1, \ldots, X_n$ be *iid* Bernoulli-type random variables each with density function $p^z(1-p)^{1-z} I_{\{0,1\}}(z)$, where X_i indicates whether the ith customer entering a store makes a purchase $(x_i = 1)$ or not $(x_i = 0)$. Examine the sequence of random variables $\{Y_1, Y_2, Y_3, \ldots\}$, where $Y_n = \sum_{i=1}^{n} X_i$, $n \in N$. Note that the nth element of the sequence represents how many of the first n customers make a purchase, and Y_n has a binomial distribution with parameters n and p, as BIN(n,p) or

$$Y_n \sim \binom{n}{y_n} p^{y_n} (1-p)^{n-y_n} I_{\{0,1,2\ldots n\}}(y_n).$$

The sequence of probability density functions associated with the sequence of random variables is given by {BIN(1,p), BIN(2,p), BIN(3,p),. . .}. □

Limit of a Real Number Sequence

We now examine the concept of the **limit of a real number sequence**. We begin with a sequence whose elements are scalars and then extend the result to a sequence whose elements are vectors of real numbers.

Definition A.25 ***Limit of a Real Number Sequence***

> Let $\{y_n\}$ be a sequence whose elements are scalar real numbers. Suppose there exists a real number, y, such that for every real $\varepsilon > 0$ there exists an integer $N(\varepsilon)$ for which $n \geq N(\varepsilon) \Rightarrow |y_n - y| < \varepsilon$. Then y is the **limit of the sequence $\{y_n\}$**, and the sequence $\{y_n\}$ is said to **converge to** y as $n \to \infty$. The existence of the limit is denoted by $y_n \to y$ or $\lim_{n \to \infty} y_n = y$. If the limit does not exist, the sequence is said to be **divergent**.

The definition of the limit implies that for a sufficiently large choice of n, y_n (and $y_{n+1}, y_{n+2}, y_{n+3}, \ldots$) becomes arbitrarily close to the number y. This is so since, by definition, we can choose $\varepsilon > 0$ to be arbitrarily small and yet there exists an n large enough (namely $n \geq N(\varepsilon)$) such that $y - \varepsilon < y_n < y + \varepsilon$. Figure A.5 provides a graphical illustration of the limit concept.

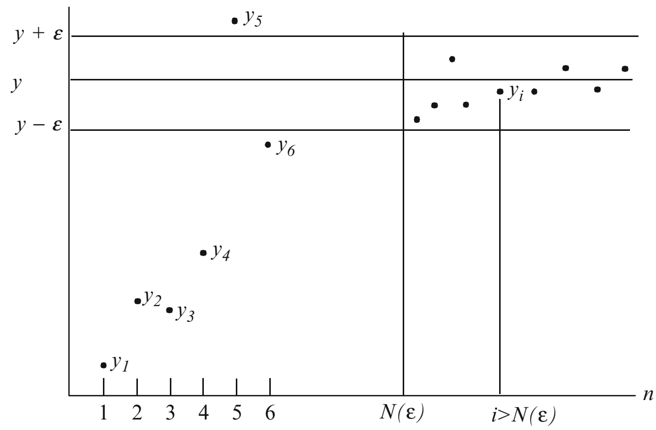


Figure A.6
Illustration of the sequence $\{y_n\}$ for which $\lim_{n\to\infty} y_n = y$.

In the figure, it is seen that for all elements of $y_i \in \{y_n\}$ for which i is large enough, i.e., for $i > N(\varepsilon)$, the value of y_i is contained in the interval $(y - \varepsilon, y + \varepsilon)$. In other words, y_i is within ε − distance of y for $i > N(\varepsilon)$. Furthermore, for every choice of $\varepsilon > 0$, no matter how small, there exists an $N(\varepsilon)$ for which a figure such as Figure A.6 applies.

It can be shown that for the limit of a sequence of real numbers to exist, it is necessary (but not sufficient) that the sequence is **bounded** (Bartle, *The Elements of Real Analysis*, 2nd Edition, John Wiley and Sons, New York, p. 93).

Definition A.26
Bounded Sequence of Real Numbers

The sequence of real numbers $\{y_n\}$ is **bounded** *iff* there exists a finite number $m > 0$ such that $|y_n| \leq m \ \forall \ n \in N$; otherwise the sequence is said to be **unbounded**.

Thus, for a sequence of real numbers to be bounded, there must exist a positive number that is larger than the absolute value of each and every number in the sequence. For a sequence that has no limit and is *also* unbounded, we write $y_n \to \infty$ (or $y_n \to -\infty$), denoting that the sequence **diverges to infinity** (or **diverges to negative infinity**), if $\forall m > 0$ there exists a positive integer $N(m)$ such that $y_n > m$ (or $y_n < -m$) $\forall n > N(m)$.

Example A.29
Boundedness and existence of a Limit for Real Number Sequences

(a) $y_n = 3 + n^{-2}$, $n \in N$. This sequence is bounded, since $|y_n| \leq 4 \ \forall \ n \in N$. Also, the sequence has a limit, where $y_n \to 3$. This follows since, $\forall \ \varepsilon > 0$, $|y_n - 3| < \varepsilon \ \forall \ n > \varepsilon^{-1/2}$, and there always exists an integer $N(\varepsilon) \geq \varepsilon^{-1/2}$ (e.g., trunc $(\varepsilon^{-1/2}) + 1$).

(b) $y_n = \sin(n)$, $n \in N$ (let n be measured in degrees). The sequence is bounded, since $|\sin(x)| \leq 1 \ \forall \ x$. The sequence does *not* have a limit, since $\sin(x)$ cycles between the values of -1 and 1.

(c) $y_n = n^2 - 3n + 1$, $n \in N$. The sequence is not bounded, since there does not exist a finite number $m > 0$ for which $n^2 - 3n + 1 \leq m \ \forall \ n \in N$. Since the sequence is unbounded, the sequence does not have a limit.

Also note that $y_n \to \infty$, that is, the sequence diverges to infinity, because $\forall\, m > 0,\ n^2 - 3n + 1 > m$ when $n > N(m) = \text{trunc}\,([3 + \sqrt{5+4m}\,]/2) + 1$ (use the quadratic formula).□

The preceding examples illustrate that boundedness of a sequence is not sufficient for the existence of a limit for the sequence. We add that it is proper to speak of *the* limit of a sequence since the limit will be *unique* if it exists at all (Bartle, p. 93).

The limit concept can be extended to a sequence whose elements are real-valued vectors or matrices. We introduce the notation $y_n[i,j]$ to indicate the (i,j)th element of the $(q \times k)$ matrix $\mathbf{y}_n$ in a sequence of matrices, and similarly $y_n[i]$ denotes the ith element of the $(q \times 1)$ vector $\mathbf{y}_n$ in a sequence of vectors. The extension of the limit concept amounts to viewing the matrix sequence as encompassing mk sequences, $\{y_n[i,j]\}$, one for each matrix element, with each element examined for convergence (by Definition A.25). Limits of vector sequences follow by letting $k = 1$.

Definition A.27
Limit of a Real-Valued Matrix Sequence

Let $\{\mathbf{y}_n\}$ be a sequence whose elements are $(q \times k)$ real-valued matrices. Suppose there exists a $(q \times k)$ matrix of real numbers $\mathbf{y}$ such that $y_n[i,j] \to y[i,j]$ for $i = 1,\ldots,q$ and $j = 1,\ldots,k$. Then the matrix $\mathbf{y}$ is the **limit of the matrix sequence** $\{\mathbf{y}_n\}$, and the sequence of matrices $\{\mathbf{y}_n\}$ is said to **converge to the matrix y** as $n \to \infty$. The existence of the limit is denoted by $\mathbf{y}_n \to \mathbf{y}$, or by $\lim_{n\to\infty}\mathbf{y}_n = \mathbf{y}$. If the limit does not exist, the sequence is said to be **divergent**.

The definition of the limit implies that for a sufficiently large choice of n, the matrix $\mathbf{y}_n$ (and $\mathbf{y}_{n+1}, \mathbf{y}_{n+2}, \mathbf{y}_{n+3},\ldots$) becomes arbitrarily close to the matrix $\mathbf{y}$, element by element. The definition also implies that for the real-valued matrix to have a limit, the sequence of matrices must be bounded elementwise, i.e., $|y_n[i,j]| \le m\ \forall\, n \in N$ and $\forall i,\, j$ or else $y_n[i,j] \not\to y[i,j]$ for some i and j, and then $\mathbf{y}_n \not\to \mathbf{y}$. Regarding divergence to infinity (or negative infinity), since there are essentially qk convergence conditions involved when examining sequences of $(q \times k)$ matrices, patterns of divergence and convergence of the various elements of the matrices can be quite diverse.

Example A.30
Boundedness and Limits of Matrices

(a) Recall the sequence of matrices in Example A.26. In this case, only the sequence $\{y_n[1,1]\}$ is bounded. All other sequences of matrix elements are unbounded and, in fact, diverge to infinity, i.e., $y_n[i,j] \to \infty$ for $(i,j) \ne (1,1)$. Since all of the sequences of matrix elements must be bounded for the matrix sequence to converge, the matrix sequence does not have a limit.

(b) Let $\{\mathbf{y}_n\}$ be a sequence of matrices such that $\mathbf{y}_n = \begin{bmatrix} 3n^{-1} & n^{-1} \\ 3 & 1+n^{-1} \end{bmatrix}$, $n \in N$.

All four sequences of matrix elements are bounded, since $|3n^{-1}| \le 3$, $|\,n^{-1}| \le 1$, $|3| \le 3$, and $|1 + n^{-1}| \le 2\ \forall\, n \in N$. Furthermore, limits exist for all four sequences of matrix elements, since $3n^{-1} \to 0$, $n^{-1} \to 0$, $3 \to 3$, and $1 + n^{-1} \to 1$. Thus, $\mathbf{y}_n \to \mathbf{y} = \begin{bmatrix} 0 & 0 \\ 3 & 1 \end{bmatrix}$. □

One might be interested in a sequence $\{y_n\}$ that is defined via a function of the elements of other sequences. For example, we may be interested in the sequence $\{y_n\}$ that is defined by adding corresponding elements in the sequences $\{x_n\}$ and $\{z_n\}$, as $y_n = x_n + z_n$. Of course, we can analyze the properties of the sequence $\{y_n\}$ directly to establish whether the sequence converges, but if the convergence properties of $\{x_n\}$ and $\{z_n\}$ are known, the following lemma can expedite the analysis if the functions are defined via addition, subtraction, or multiplication. We precede the statement of the lemma with definitions for adding, subtracting, and multiplying sequences. Note that x[.,i] refers to the ith column of the matrix **x**.

Definition A.28
Adding, Subtracting, Multiplying Sequences

Let $\{\mathbf{x}_n\}$ and $\{\mathbf{z}_n\}$ be sequences of conformable real-valued matrices.

(a) *Summation*: The summation of $\{\mathbf{x}_n\}$ and $\{\mathbf{z}_n\}$, $\{\mathbf{x}_n\} + \{\mathbf{z}_n\}$, is a sequence $\{\mathbf{y}_n\}$ defined by $\mathbf{y}_n = \mathbf{x}_n + \mathbf{z}_n$, $\forall n$.

(b) *Difference*: The difference between $\{\mathbf{x}_n\}$ and $\{\mathbf{z}_n\}$, $\{\mathbf{x}_n\} - \{\mathbf{z}_n\}$, is a sequence $\{\mathbf{y}_n\}$ defined by $\mathbf{y}_n = \mathbf{x}_n - \mathbf{z}_n$, $\forall n$.

(c) *Product*: The product of $\{\mathbf{x}_n\}$ and $\{\mathbf{z}_n\}$, $\{\mathbf{x}_n\}\ \{\mathbf{z}_n\}$, is a sequence $\{\mathbf{y}_n\}$ defined by $\mathbf{y}_n = \mathbf{x}_n\mathbf{z}_n$, $\forall n$.

Lemma A.1
Combinations of Sequences

Let $\{\mathbf{x}_n\}$ and $\{\mathbf{z}_n\}$ be convergent sequences of conformable real-valued matrices such that $\mathbf{x}_n \rightarrow \mathbf{x}$ and $\mathbf{z}_n \rightarrow \mathbf{z}$. Then

(a) $\mathbf{x}_n + \mathbf{z}_n \rightarrow \mathbf{x} + \mathbf{z}$,

(b) $\mathbf{x}_n - \mathbf{z}_n \rightarrow \mathbf{x} - \mathbf{z}$,

(c) $\mathbf{x}_n\mathbf{z}_n \rightarrow \mathbf{xz}$,

(d) if $\{a_n\} \rightarrow a$, then $a_n\mathbf{x}_n \rightarrow a\mathbf{x}$,

(e) if $\{b_n\} \rightarrow b \neq 0$, then $b_n^{-1}\mathbf{x}_n \rightarrow b^{-1}\mathbf{x}$,

(f) $\sum_{i=1}^{k} \mathbf{x}_n[.,i] \rightarrow \sum_{i=1}^{k} \mathbf{x}[.,i]$

(g) if $\{\mathbf{z}_n\}$ is a sequence of nonsingular matrices that converges to the nonsingular matrix $\mathbf{z}$, then $\mathbf{z}_n^{-1} \rightarrow \mathbf{z}^{-1}$ and $\mathbf{z}_n^{-1}\mathbf{x}_n \rightarrow \mathbf{z}^{-1}\mathbf{x}$.

Proof Bartle, pp. 100–101, and Definitions 5.4 and 5.5.

Note that since the sequences $\{\mathbf{x}_n\}$ and $\{\mathbf{z}_n\}$ can themselves be defined in terms of combinations of other sequences, the lemma actually implies convergence results involving more than just two sequences. For example, letting $\mathbf{z}_n = \mathbf{a}_n + \mathbf{b}_n \rightarrow \mathbf{a} + \mathbf{b}$, then $\mathbf{a}_n + \mathbf{b}_n + \mathbf{z}_n \rightarrow \mathbf{a} + \mathbf{b} + \mathbf{z}$ and $(\mathbf{a}_n + \mathbf{b}_n)\mathbf{z}_n \rightarrow (\mathbf{a} + \mathbf{b})\mathbf{z}$.

Example A.31
Convergence Properties of Sequences

(a) Let $\{x_n\}$ and $\{z_n\}$ be defined as $x_n = 3 + n^{-1/2}$ and $z_n = 2\exp(-2/n)$ for $n \in N$, respectively. Note that $x_n \to 3$ and $z_n \to 2$. Then using Lemma A.1, $x_n + z_n \to 5$, $x_n - z_n \to 1$, and $x_n z_n \to 6$. Let $\{a_n\}$ be defined by $a_n = 5(n+1)/n$ for $n \in N$, and note that $a_n \to 5$. Also define the vector sequence $\{\mathbf{y}_n\}$ by $\underset{(2\times 1)}{\mathbf{y}_n} = \begin{bmatrix} x_n \\ z_n \end{bmatrix}$ so that $\mathbf{y}_n \to \begin{bmatrix} 3 \\ 2 \end{bmatrix}$. Then from Lemma A.1, $a_n \mathbf{y}_n \to \begin{bmatrix} 15 \\ 10 \end{bmatrix}$ and $a_n^{-1}\mathbf{y}_n \to \begin{bmatrix} 3/5 \\ 2/5 \end{bmatrix}$.

(b) Let $\{\mathbf{w}_n\}$ be a matrix sequence defined by

$$\mathbf{w}_n = \begin{bmatrix} 2+n^{-1} & \dfrac{2}{n} \\ 3 & \dfrac{(n+1)}{n} \end{bmatrix} \text{ for } n \in N,$$

and let $\{\mathbf{x}_n\}$ be a vector sequence defined by

$$\mathbf{x}_n = \begin{pmatrix} 1+n^{-1} \\ 2\exp(n^{-1}) \end{pmatrix} \text{ for } n \in N.$$

Note that $\mathbf{w}_n \to \begin{bmatrix} 2 & 0 \\ 3 & 1 \end{bmatrix}$ and $x_n \to \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. Using Lemma A.1, it follows that $\mathbf{w}_n^{-1} \to \begin{bmatrix} 2 & 0 \\ 3 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} .5 & 0 \\ -1.5 & 1 \end{bmatrix}$, $\mathbf{w}_n\,\mathbf{x}_n \to \begin{bmatrix} 2 \\ 5 \end{bmatrix}$, and $\mathbf{w}_n^{-1}\mathbf{x}_n \to \begin{bmatrix} .5 \\ .5 \end{bmatrix}$.

Note further that $\mathbf{w}_n\,[.,1] + \mathbf{w}_n\,[.,2] + \mathbf{x}_n \to \begin{bmatrix} 3 \\ 6 \end{bmatrix}$. □

Continuous Functions

Continuous functions play a prominent role in a number of important asymptotic results. We define the concept of a continuous function below using two alternative but completely equivalent characterizations. We remind the reader that $d(\mathbf{x},\mathbf{w}) = [(\mathbf{x}-\mathbf{w})'(\mathbf{x}-\mathbf{w})]^{1/2}$ is the distance between points $\mathbf{x}$ and $\mathbf{w}$.

Definition A.29
Continuous Functions

The function[10] $g\colon A \to \mathbb{R}$, for $A \subset \mathbb{R}^m$, is continuous at the point $\mathbf{x} \in A$ iff either:

(a) $\forall \varepsilon > 0$, $\exists \delta(\varepsilon) > 0$ such that $\mathbf{w} \in A$ and $d(\mathbf{x},\mathbf{w}) < \delta(\varepsilon)$ implies $|g(\mathbf{w}) - g(\mathbf{x})| < \varepsilon$

(b) $\forall$ sequence $\{\mathbf{x}_n\}$ in A for which $\mathbf{x}_n \to \mathbf{x}$, it is also true that $g(\mathbf{x}_n) \to g(\mathbf{x})$.

[10]This definition can be altered to provide definitions for *continuity from the right* and *continuity from the left*. For continuity from the right, the condition $\mathbf{w} \geq \mathbf{x}$ is added in part (a). The condition $\mathbf{x}_n \geq \mathbf{x}\ \forall n$ is added to part (b). For continuity from the left, the conditions become $\mathbf{w} \leq \mathbf{x}$ and $\mathbf{x}_n \leq \mathbf{x}\ \forall n$.

The $k \times 1$ *vector* function $\mathbf{g}$: $A \rightarrow \mathbb{R}^k$ is continuous at the point $\mathbf{x} \in A$ *iff* each coordinate function $g_j(\mathbf{x})$ is continuous at the point $\mathbf{x}$, $j = 1,\ldots,k$. The function $\mathbf{g}$ is said to be continuous on the set $B \subset A$ if the function is continuous at every point in B.

Intuitively, a function is continuous at a point $\mathbf{x}$ if, when the function is evaluated at domain elements that are closer and closer to $\mathbf{x}$, the value of the function is closer and closer to the value of the function at $\mathbf{x}$ (in an elementwise vector comparison sense if $k \geq 2$). In the simple case where $m = k = 1$, the graph of a function that is continuous on an interval set $B = (a,b)$ can be drawn $\forall\, x \in B$ "without lifting the pencil from the paper," i.e., the graph is an unbroken curve.

Example A.32
Continuity Properties of Functions

(a) Let f:$(0,\infty) \rightarrow \mathbb{R}$ be defined by $y = x^{-1}$. Intuitively we expect f to be continuous on $(0,\infty)$, since its graph is an unbroken curve. To demonstrate formally that f is continuous on $(0,\infty)$, let $x_0 \in (0,\infty)$, and note that

$$|f(x) - f(x_0)| = \left|x^{-1} - x_0^{-1}\right| = \left|\frac{x_0 - x}{xx_0}\right| = \frac{|x_0 - x|}{xx_0}$$

where we have eliminated the denominator from the absolute value operator because in the domain of f, $x > 0$. It follows that if $|x - x_0| < \delta$, then $|f(x) - f(x_0)| < \delta/xx_0$.

Now choose any $\varepsilon > 0$. For f to be continuous at x_0, it must be the case that δ can be chosen such that $|x - x_0| < \delta \Rightarrow |f(x) - f(x_0)| < \varepsilon$. Such choices of δ exist, one being equal to $\delta = x_0^2\, \varepsilon\, /(1 + x_0\, \varepsilon)$. Since the argument can be applied $\forall x_0 \in (0,\infty)$ and $\forall\, \varepsilon > 0$, f is continuous on $(0,\infty)$.

(b) Let f: $\mathbb{R}^2 \rightarrow \mathbb{R}^2$be defined by the coordinate functions

$$\begin{bmatrix} y[1] \\ y[2] \end{bmatrix} = \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} x[1]^2 + 2x[2] \\ x[1] \end{bmatrix}.$$

Let $\mathbf{x}_* \in \mathbb{R}^2$, and let $\{\mathbf{x}_n\}$ be any sequence in $\mathbb{R}^2$ for which $\mathbf{x}_n \rightarrow \mathbf{x}_*$.To demonstrate continuity of f at $\mathbf{x}_*$, we will show that $\mathbf{x}_n \rightarrow \mathbf{x}_* \Rightarrow f(\mathbf{x}_n) \rightarrow f(\mathbf{x}_*)$. Examine $f_1(x)$ first. Note that $f_1(\mathbf{x}_n) = x_n[1]^2 + 2x_n[2] = (x_n[1]x_n[1]) + 2x_n[2]$ can be interpreted as the summation of the sequence $\{x_n[1]\}\{x_n[1]\}$ and the sequence $2\{x_n[2]\}$. Since $x_n[1] \rightarrow x_*[1]$ and $x_n[2] \rightarrow x_*[2]$ by assumption, it follows from Lemma A.1 that $\{x_n[1]\}\,\{x_n[1]\} \rightarrow x_*[1]^2$, $2\{x_n[2]\} \rightarrow 2\,x_*[2]$, and thus $f_1(\mathbf{x}_n) \rightarrow x_*[1]^2 + 2\,x_*[2] = f_1(\mathbf{x}_*)$. Verifying convergence of the second component function is straightforward, because $f_2(\mathbf{x}_n) = x_n[1] \rightarrow x_*[1] = f_2(\mathbf{x}_*)$. Thus, f is continuous at $\mathbf{x}_*$, and since the above argument holds for any $\mathbf{x}_* \in \mathbb{R}^2$, f is continuous on $\mathbb{R}^2$.

(c) Let f: $\mathbb{R} \rightarrow \mathbb{R}$ be defined by $f(x) = I_{(0,\infty)}(x)$. Intuitively, since the graph of the function has a break at $x = 0$, we expect that the function is not continuous on $\mathbb{R}$. To formally demonstrate that f is discontinuous at $x = 0$, it will be shown that there does *not* exist a $\delta(\varepsilon) > 0$ for every $\varepsilon > 0$ such that $|x - 0| < \delta(\varepsilon) \Rightarrow |f(x) - f(0)| < \varepsilon$. Choose any $\varepsilon \in (0,1)$, and note that $f(0) = 0$. Given *any* choice of $\delta(\varepsilon) > 0$, let $x = \delta(\varepsilon)/2$, so that $|x - 0| < \delta(\varepsilon)$.

Then $|f(x) - f(0)| = 1 \nless \varepsilon$. Therefore, f is not continuous on $\mathbb{R}$. (Note: f *is* continuous on $(0,\infty)$ and on $(-\infty,0)$, as the reader might wish to verify). □

Convergence of a Function Sequence

Convergence of a sequence of functions refers to a case where the function definitions themselves can change as n changes, so that we can conceptualize a sequence of image values $\{f_n(x)\} = \{f_1(x), f_2(x), f_3(x), \ldots\}$ for *each* value of x (which could be a vector) in the *common* domain of the sequence of functions. Interest centers on whether there exists a "limit" function definition, $f(x)$, such that $f_n(x) \to f(x)$ for all x in some subset of the common domain of the sequence of functions. The subset of x-values on which $\{f_n(x)\}$ converges to $f(x)$ could be the entire domain, or some smaller subset of points, or the null set (i.e., $\{f_n(x)\}$ does not converge for any x).

We formalize the concept of convergence of a sequence of functions in the next definition. In the definition, the notation $\{f_n\}$ refers to the sequence of function definitions, as opposed to $\{f_n(x)\}$ which denotes the sequence of image values generated by the sequence of function definitions when evaluated at the point x.

Definition A.30
Convergence of a Function Sequence

Let $\{f_n\}$ be a sequence of functions $f_n : D \to \mathbb{R}^\ell$ for $n = 1, 2, \ldots,$ having common domain $D \subset \mathbb{R}^m$. Let $f : D_0 \to \mathbb{R}^\ell$ be a function with domain $D_0 \subset D$. The function sequence $\{f_n\}$ is said to converge on D_0 to f if $f_n(\mathbf{x}) \to f(\mathbf{x})$, $\forall \mathbf{x} \in D_0$. If $\{f_n\}$ converges to f on D_0, f is called the **limiting function** of $\{f_n\}$ on D_0, and $\{f_n\}$ is said to be **convergent** on D_0.

Intuitively, if f is the limit function of $\{f_n\}$ on D_0, then for large enough n, $f_n(\mathbf{x}) \approx f(\mathbf{x})$ for $x \in D_0$ since $f_n(\mathbf{x})$ converges to $f(\mathbf{x})$ on D_0. Then $f(\mathbf{x})$ can be viewed as an approximation to $f_n(\mathbf{x})$ on D_0 when n is large.

Example A.33
Convergence of Function Sequences

(a) Let the function sequence $\{f_n\}$ be defined by $f_n(x) = n^{-1} + 2x^2$ for $x \in \mathbb{R}$. Define the function f: $\mathbb{R} \to \mathbb{R}$ by $f(x) = 2x^2$. Then f is the limiting function of $\{f_n\}$ on $\mathbb{R}$. To see this, note that $f_n(x) \to 2x^2 = f(x)\ \forall\, x \in \mathbb{R}$. For large n, $f_n(x) \approx f(x)\ \forall\, x \in \mathbb{R}$.

(b) Let the vector function sequence $\{f_n\}$ be defined by

$$f_n(x) = \begin{bmatrix} f_{1n}(\mathbf{x}) \\ f_{2n}(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} \dfrac{(2x_1^2 + 3nx_1x_2)}{n} \\ x_1^2 + x_2^2 \end{bmatrix} \quad \text{for } (x_1, x_2) \in \mathbb{R}^2.$$

Define the function f:$\mathbb{R}^2 \to \mathbb{R}^2$ by

$$f(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} 3x_1\ x_2 \\ x_1^2 + x_2^2 \end{bmatrix} \text{ for } (x_1, x_2) \in \mathbb{R}^2.$$

Then f is the limiting function of $\{f_n\}$ on $\mathbb{R}^2$. To see this, note that $f_{1n}(\mathbf{x}) = 2x_1^2/n + 3x_1\ x_2 \to 0 + 3x_1\ x_2 = f_1(\mathbf{x})$, and $f_{2n}(\mathbf{x}) = x_1^2 + x_2^2 = f_2(\mathbf{x})$ $\forall \mathbf{x} \in \mathbb{R}^2$, so that $f_n(\mathbf{x}) \to f(\mathbf{x})\ \forall\ \mathbf{x} \in \mathbb{R}^2$.

(c) Let the function sequence $\{f_n\}$ be defined by $f_n(x) = x - 2\exp(-nx)$ for $x \in \mathbb{R}$. Let f: $\mathbb{R} \rightarrow \mathbb{R}$ be defined by $f(x) = x - 2I_{\{0\}}(x)$ for $x \in D_0 = [0,\infty]$. Then f is the limiting function of $\{f_n\}$ on the set $[0,\infty)$. The sequence $\{f_n\}$ does not converge for $x < 0$. To justify these conclusions, note that, by Lemma A.1, $\lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} x - \lim_{n\to\infty} 2\exp(-nx) \quad = x - 2I_{(0)}(x) \forall x \geq 0$. When $x < 0$, $2\exp(-nx) \rightarrow \infty$, and thus $f_n(x)$ does not converge. □

Order of Magnitude of a Sequence

In analyses involving sequences, it is sometimes useful to be able to characterize or compare sequences and/or terms in a sequence relative to their order of magnitude. In particular, when the definition of a sequence contains a number of terms, the orders of magnitude of the terms will distinguish which terms make dominant contributions to the magnitude of the sequence as n increases. The concept is defined below.

Definition A.31
Order of Magnitude of a Sequence

Let $\{x_n\}$ be a real number sequence, and let $\{\mathbf{w}_n\}$ be a real-valued matrix sequence.

(a) The sequence $\{x_n\}$ is said to be **at most of order** n^k, denoted by $O(n^k)$, if there exists a finite real number c such that $|n^{-k} x_n| \leq c \ \forall\, n \in N$.

(b) The sequence $\{x_n\}$ is said to be **of order smaller than** n^k, denoted by $o(n^k)$, if $n^{-k}x_n \rightarrow 0$.

(c) If $\{w_n[i,j]\}$ is $O(n^k)$ (or $o(n^k)$) $\forall\, i$ and j, then the matrix sequence $\{\mathbf{w}_n\}$ is said to be $O(n^k)$ (or $o(n^k)$).

Intuitively, a sequence $\{x_n\}$ is $O(n^k)$ if the corresponding sequence $\{n^{-k}x_n\}$ is such that all elements $n^{-k}x_n$ are bounded in absolute value by some positive number c. A sequence $\{x_n\}$ is $o(n^k)$ if the product sequence $\{n^{-k}x_n\}$ converges to zero. Note that if $\{x_n\}$ is $O(n^k)$, then $\{x_n\}$ is $o(n^{k+\in}) \ \forall \in > 0$, and if $\{x_n\}$ is $o(n^k)$, then it is also $O(n^k)$. Notationally, the case $O(n^o)$ or $o(n^o)$ is most often represented as $O(1)$ or $o(1)$.

Example A.34
Order of Magnitude of a Sequence

(a) Let $\{x_n\}$ be defined by $x_n = 3n^3 - n^2 + 2$, for $n \in N$. Then $\{x_n\}$ is $O(n^3)$, since $n^{-3}x_n = 3 - n^{-1} + 2n^{-3}$ is bounded. Also, $\{x_n\}$ is $o(n^{3+\varepsilon})$ for any $\varepsilon > 0$ since $n^{-3-\varepsilon}x_n = 3n^{-\varepsilon} - n^{-1-\varepsilon} + 2n^{-3-\varepsilon} \rightarrow 0$.

(b) Let $\{x_n\}$ be defined by $x_n = 3 + n^{-1}$, for $n \in N$. Then $\{x_n\}$ is $O(1)$, since $x_n = 3 + n^{-1}$ is bounded, and $\{x_n\}$ is $o(n^\varepsilon)\ \forall \varepsilon > 0$, since $n^{-\varepsilon}x_n = 3n^{-\varepsilon} + n^{-1-\varepsilon} \rightarrow 0$.

(c) Let the vector sequence $\{\mathbf{x}_n\}$ be defined by $\begin{bmatrix} x_n[1] \\ x_n[2] \end{bmatrix} = \begin{bmatrix} 3n^{-1} \\ n^{-1} \end{bmatrix}$. Then the vector sequence $\{\mathbf{x}_n\}$ is $o(1)$ and $O(1)$, since $x_n \rightarrow \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. □

Some useful results regarding the order of magnitude of sums and products of sequences are given in the following lemma.

Lemma A.2

Let $\{x_n\}$ and $\{z_n\}$ be real number sequences. The following relationships between orders of magnitude hold:

IF		THEN	
$\{x_n\}$	$\{z_n\}$	$\{x_n + z_n\}$	$\{x_n z_n\}$
$O(n^k)$	$O(n^m)$	$O(n^{\max(k,m)})$	$O(n^{k+m})$
$o(n^k)$	$o(n^m)$	$o(n^{\max(k,m)})$	$o(n^{k+m})$
$O(n^k)$	$o(n^m)$	$O(n^{\max(k,m)})$	$o(n^{k+m})$

Proof H. White (1984), *Asymptotic Theory for Econometricians.* Orlando, Academic Press, p. 15.

Given that the sequences referred to in Lemma A.2 can themselves be functions of other sequences, the results can be extended in a myriad of ways to an arbitrary finite number of sequences. In the following lemma we state some useful extensions.

Lemma A.3

Let $\{\mathbf{x}_n\}$ and $\{\mathbf{w}_n\}$ be $(m \times \ell)$ and $(r \times m)$ matrix sequences, respectively.

(a) If $\{\mathbf{x}_n\}$ is such that $\{\mathbf{x}_n[.,i]\}$ is $O(n^{k_i})$ (or $o(n^{k_i})$) for $i = 1,\ldots,\ell$, then $\{ \sum_{i=1}^{\ell} \mathbf{x}_n[.,i]\}$ is $O(n^{k\max})$ (or $o(n^{k\max})$) where $k\max = \max(k_1,\ldots,k_\ell)$.

(b) (Special case of (a)): If $\{\mathbf{x}_n\}$ is $O(n^k)$ (or $o(n^k)$), then $\{ \sum_{i=1}^{\ell} \mathbf{x}_n[.,i]\}$ is $O(n^k)$ (or $o(n^k)$).

(c) If $\{\mathbf{x}_n\}$ is $O(n^k)$ and $\{\mathbf{w}_n\}$ is $\left\{ \begin{matrix} O(n^d) \\ o(n^d) \end{matrix} \right\}$, then $\{\mathbf{w}_n\mathbf{x}_n\}$ is $\left\{ \begin{matrix} O(n^{k+d}) \\ o(n^{k+d}) \end{matrix} \right\}$ and $\{n^{-v}\mathbf{w}_n\mathbf{x}_n\}$ is $\left\{ \begin{matrix} O(n^{k+d-v}) \\ o(n^{k+d-v}) \end{matrix} \right\}$.

(d) (Special case of (c)): If $\{\mathbf{x}_n\}$ is $O(n^k)$ (or $o(n^k)$), then $\{\mathbf{x}_n'\mathbf{x}_n\}$ is $O(n^{2k}\}$ (or $o(n^{2k})$) and $\{n^{-v}\mathbf{x}_n'\mathbf{x}_n\}$ is $O(n^{2k-v})$ (or $o(n^{2k-v})$).

Proof This follows from Lemma A.2 and mathematical induction.

We illustrate the application of some of the preceding results in the following example.

Example A.35
Orders of Magnitude for Combinations of Sequences

(a) Let $\{x_n\}$ be defined by $x_n = 2n^{-1} + 5n$ and let $\{z_n\}$ be defined by $z_n = n^2 + 2n$. Note that $\{x_n\}$ is $O(n^1)$ and $\{z_n\}$ is $O(n^2)$. It follows immediately, from Lemma A.2 that $\{x_n + z_n\} = \{n^2 + 7n + 2n^{-1}\}$ is $O(n^2)$ (and $o(n^{2+\varepsilon})$ for $\varepsilon > 0$) and $\{x_n z_n\} = \{5n^3 + 10n^2 + 2n + 4\}$ is $O(n^3)$ (and $o(n^{3+\varepsilon})$ for $\varepsilon > 0$).

(b) Let $\{\mathbf{x}_n\}$ and $\{\mathbf{w}_n\}$ be defined by

$$\mathbf{x}_n = \begin{bmatrix} 7+n^{-1} \\ n^{-1} \end{bmatrix} \text{ and } \mathbf{w}_n = \begin{bmatrix} n^2+2n+1 & 3\,n^2+7 \\ n^2 & n^2+n \end{bmatrix}$$

so that $\mathbf{x}_n$ is $O(1)$ and $\mathbf{w}_n$ is $O(n^2)$. It follows immediately from Lemma A.3 that

$$\{\mathbf{w}_n\,\mathbf{x}_n\} = \left\{ \begin{bmatrix} 7\,n^2+18n+9+8\,n^{-1} \\ 7\,n^2+2n+1 \end{bmatrix} \right\}$$

is $O(n^2)$ and $o(n^{2+\varepsilon})$ for $\varepsilon > 0$, and $\{\mathbf{x}_n'\mathbf{x}_n\} = \{49+14n^{-1}+2n^{-2}\}^{\frac{1}{2}}$ is $O(1)$ and $o(n^{\varepsilon})$ for $\varepsilon > 0$. It also follows, for example, that $\{n\mathbf{x}_n'\mathbf{x}_n\}$ and $\{n^{-1}\mathbf{w}_n\mathbf{x}_n\}$ are both $O(n^1)$ and $o(n^{1+\varepsilon})$ for $\varepsilon > 0$. □

Keywords, Phrases, and Symbols

$\in$
$\mathbb{R}_{\geq 0}$, $\mathbb{R}_+$
$A^2 = A \times A$
Associative laws
Axiom (or postulate)
Binary relation from A to B, S: $A \rightarrow B$
Binomial theorem
Cartesian product
Closed, open, half-open intervals
Combinations, $\binom{n}{r}$
Commutative laws
Complement, $\bar{A}$
Complements of complements
Contained in, $\subset$
Corollary
Countable set
Definition
DeMorgan's laws
Distributive laws
Element
Empty or null set, $\emptyset$
Equality of sets, $=$
Exhaustive listing
Finite set
For every, $\forall$
Function from A to B, f:$A \rightarrow B$
Idempotency laws
Identity elements
iff (if and only if)
Image of x under f
Index set
Indicator function, $I_A(x)$
Infinite set
Integration notation
Intersection and union of complements
Intersection with null set
Intersection, $\cap$
Interval set notation
Inverse function, f^{-1}:$B \rightarrow A$
Inverse image
Lemma
Mathematical rule
Multiple intersection notation
Multiple union notation
Mutually exclusive (disjoint)
Negation, /
Null set as a subset
Permutations, n_r
Point function
QED
Real line
Real-valued function
Set
Set difference, $-$
Set function
Size of set function
Stirling's formula
Subset
Such that (such that)
Summation notation
The function $f(x)$
The set function $f(A)$
Theorem (or proposition)
There exists, $\exists$
Uncountable set
Union, $\cup$
Universal set
Venn diagram
Verbal rule
xSy

Problems

1. Using either an exhaustive listing, verbal rule, or mathematical rule, define the following sets:

(a) The set of all senior citizens receiving social security payments in the United States.

(b) The set of all positive numbers that are positive integer powers of the number 10 (i.e., 10^1, 10^2, etc.).

(c) The set of all possible outcomes resulting from rolling a red and a green die and calculating the values of $y - x$, where y = number of dots on the red die, x = number of dots on the green die.

(d) the set of all two-tuples (x_1, x_2) where x_1 is any real number and x_2 is related to x_1 by raising the number e to the power x_1.

2. Label the sets you have identified in Problem (1) as being either finite, countably infinite, or uncountably infinite, and explain your choice.

3. For each set below, state whether the set is finite, countably infinite, or uncountably infinite.

(a) $S = \{x$: x is a U.S. citizen who has purchased a Japanese car during the past year$\}$.

(b) $S = \{(x,y): y \leq x^2, x \text{ is a positive integer}, y \in \mathbb{R}_{\geq 0}\}$.

(c) $S = \{p$: p is the price of a quart of milk sold at a retail store in the U.S. on Friday, September 13, 1991$\}$.

(d) $S = \{x: x = 2y, \text{ y is a positive integer}\}$.

4. Let the universal set be $\Omega = [0,10]$, and define the following subsets of Ω.

$A = [0, 2), B = [2, 7], C = [5, 6], D = \{2\},$

$E = \{x : x = y^{-1}, \text{ y is an even positive integer} \geq 4\}.$

(a) Define the following sets:

$A \cup B, A \cap B, \overline{A \cup C}, (A \cup D) \cap B, B - C, A \cap E, \bar{D} \cap B$

(b) For each of the sets in (a), indicate whether the set is finite, countably infinite, or uncountably infinite.

5. Let the universal set be defined by $\Omega = [-5,5]$, and define the following subsets of Ω:

$A_1 = [-2, 1)$
$A_2 = (1, 2)$
$A_3 = [2, 5]$
$A_4 = [-5, -2]$

Also, define an index set $I = \{1,3,4\}$.

(a) Define $\cup_{i \in I} A_i$.

(b) Define $\cup_{i=1}^{4} A_i$

(c) Define $A_1 \cap A_2$

(d) Define $A_4 - A_1$

(e) Define $\bar{A}_4$

6. Define the universal set, Ω, as $\Omega = \{x : 0 \leq x \leq 5 \text{ or } 10 \leq x \leq 20\}$, and define the following subsets of Ω as

$A_1 = \{x : 0 \leq x < 2.5\},$
$A_2 = \{x : 15 < x \leq 20\},$
$A_3 = \{x : 2.5 \leq x \leq 5 \text{ or } 10 \leq x \leq 20\},$
$A_4 = \{x : 0 \leq x \leq 5 \text{ or } 10 \leq x \leq 15\}.$

In addition, define the following two index sets as:

$I_1 = \{1, 3\}, I_2 = \{1, 4\}.$

Define the following sets:

(a) $\cup_{i \in I_1} A_i$

(b) $\cap_{i=1}^{4} A_i$

(c) $\cap_{i \in I_2} A_i$

(d) $\cap_{i=1}^{2} A_i$

(e) $A_1 - A_2$

(f) $A_4 - A_3$

(g) $\bar{A}_3$

(h) $A_2 - \cup_{i \in I_1} A_i$

7. In each situation below indicate where the relation is a function. If so, determine the domain and range of the function.

(a) $A = [0,10)$, $B = [0, 1\text{n}(11)]$, $S = \{(x,y): y = 1\text{n}(1 + x), (x,y) \in A \times B\}$

(b) Consider S^{-1}, the inverse of S in (a).

(c) $A = \mathbb{R}^2_{\geq 0}$, $B = [0,\infty)$, $S = \{((x_1, x_2), y): y = 5\, x_1\, x_2^2, ((x_1, x_2), y) \in A \times B\}$

(d) Consider S^{-1}, the inverse of S in (c).

8. For each relation below, state whether the relation is a function, and state whether an inverse function exists. Explicitly define the inverse function if it exists.

(a) Let $P = \{\$.01, \$.02, \ldots, \$1.00, \$1.01, \ldots\}$ represent a set of possible prices for a given commodity, and let $Q = [0,\infty)$ represent possible levels of quantity demanded. Define $S: P \rightarrow Q$ as $S = \{(p,q): q = 20p^{-1.5}, (p,q) \in P \times Q\}$

(b) Let $A = \{D: D = \{(x_1, x_2): x_1 \in [a_1, b_1], x_2 \in [a_2, b_2]\}, a_1 < b_1, a_2 < b_2\}$ be a set of rectangular sets. Define $S: A \rightarrow \mathbb{R}$ as

$S = \{(D, y) : y = \text{area of } D, \ (D, y) \in A \times \mathbb{R}_+\}$

9. Let $A = (0,\infty)$, and examine the following relation on A:

$$S = \{(x, y) : y = 2 + 3x,\ (x, y) \in A_2)\}.$$

(a) Is $2S8$?

(b) Is S a function?

(c) Does an inverse function exist?

(d) If you can, define $f(2)$ and $f^{-1}(5)$.

(e) What is D(S)? What is R(S)?

10. Define a universal set as $\Omega = \{x{:}0 \leq x \leq 5\}$, and consider the set function

$$P(A) = .5 \int_{x \in A} x dx + 12.5$$

where the domain of the set function is all subsets $A \subset \Omega$ of the form $A = [a, b]$, for $0 \leq a \leq b \leq 5$.

(a) What is the image of the set $A = [0, 2]$ under P?

(b) What is the image of Ω under P?

(c) What is the image of set $A = [3,3]$ under P?

(d) What is the range of the set function?

(e) What is the inverse image of 2?

11. Define a set function that will assign the appropriate area to all rectangles of the form $[x_1, x_2] \times [y_1, y_2]$, $x_2 \geq x_1$ and $y_2 \geq y_1$, contained in $\mathbb{R}^2$. Be sure to identify the domain and range of the set function.

12. A statistics class has 20 students in attendance, and in the room where the class meets, there are 25 desks available for the students. How many different ways can the students leave five desks unoccupied?

13. There are 15 students in an econometrics class that you are attending.

(a) How many different ways can a three-person committee be formed to give a class report?

(b) Of the number of possible three-person committees indicated in (a), how many involve you?

14. Competing for the title of Miss America are 50 contestants from each of the 50 states plus one contestant from the District of Columbia. How many different ways can the contestants be assigned the titles of Miss America, first runner up,..., fourth runner up?

15. Let $A_1 = \{x{:}\ x \text{ is a positive integer}\}$, $A_2 = \{1, 2, 3, 4, 5\}$, $B = \{(x_1, x_2){:}\ (x_1, x_2) \in A_1 \times A_2, x_1 \leq x_2\}$, and $y_i = i^2$. Calculate the values of the following sums.

(a) $\sum_{x \in A_2} x$

(b) $\sum_{i \in A_2} y_i$

(c) $\sum_{x_1 \in A_1} \sum_{x_2 \in A_2} (1/2)^{x_1} x_2^2$

(d) $\sum_{x \in (A_1 - A_2)} (1/3)^x$

(e) $\sum_{(x_1, x_2) \in B} (x_1 + x_2)$

16. Let $A_1 = [0, \infty)$, $A_2 = [1, 10]$, and $B = \{(x_1, x_2){:}\ (x_1, x_2) \in A_1 \times A_2,\ x_2 > x_1\}$. Calculate the values of the following integrals.

(a) $\int_{x \in A_1} (1/2) e^{-x/2} dx$

(b) $\int_{x_1 \in A_1} \int_{x_2 \in A_2} x_2 e^{-x_1} dx_2 dx_1$

(c) $\int_{(x_1, x_2) \in B} (x_1 + x_2) dx_1 dx_2$

(d) $\int_0^2 \int_{x_2 \in A_1 \cap A_2} x_1 x_2^2 dx_2 dx_1$

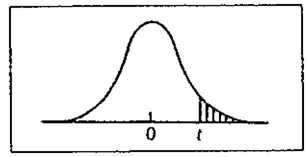
Useful Tables

Table B.1 Cumulative normal distribution $F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$

x	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9830	0.9834	0.9838	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9978	0.9979	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Source: Reprinted, by permission of the publisher, from A. M. Mood, F. A. Graybill, and D. C. Boes, Introduction to the Theory of Statistics*, 3d ed., New York: McGraw-Hill, 1974, p. 552*

Table B.2 Student's *t* distribution. The first column lists the number of degrees of freedom (v). The headings of the other columns give probabilities (P) for t to exceed the entry value. Use symmetry for negative t values



P / v	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
120	1.289	1.658	1.980	2.358	2.617
∞	1.282	1.645	1.960	2.326	2.576

Source: Reprinted, by permission of the publisher, from P. G. Hoel, Introduction to Mathematical Statistics, *4th ed., New York: John Wiley and Sons, Inc., 1971, p. 393*

Table B.3 Chi-square distribution. The first column lists the number of degrees of freedom (v). The headings of the other columns give probabilities (P) for the χ_v^2 random variable to exceed the entry value

P / v	0.995	0.990	0.975	0.950	0.900	0.750
1	392704×10^{-10}	157088×10^{-9}	982069×10^{-9}	393214×10^{-8}	0.0157908	0.1015308
2	0.0100251	0.0201007	0.0506356	0.102587	0.210720	0.575364
3	0.0717212	0.114832	0.215795	0.351846	0.584375	1.212534
4	0.206990	0.297110	0.484419	0.710721	1.063623	1.92255
5	0.411740	0.554300	0.831211	1.145476	1.61031	2.67460
6	0.675727	0.872085	1.237347	1.63539	2.20413	3.45460
7	0.989265	1.239043	1.68987	2.16735	2.83311	4.25485
8	1.344419	1.646482	2.17973	2.73264	3.48954	5.07064
9	1.734926	2.087912	2.70039	3.32511	1.16816	5.89883
10	2.15585	2.55821	3.24697	3.94030	4.86518	6.73720
11	2.60321	3.15347	3.81575	4.57481	5.57779	7.58412
12	3.07382	3.57056	4.40379	5.22603	6.30380	8.43842
13	3.56503	4.10691	5.00874	5.89186	7.04150	9.29906
14	4.07468	4.66043	5.62872	6.57063	7.78953	10.1653
15	4.60094	5.22935	6.26214	7.26094	8.54675	11.0365
16	5.14224	5.81221	6.90766	7.96164	9.31223	11.9122
17	5.69724	6.40776	7.56418	8.67176	10.0852	12.7919
18	6.26481	7.01491	8.23075	9.39046	10.8649	13.6753
19	6.84398	7.63273	8.90655	10.1170	11.6509	14.5620
20	7.43386	8.26040	9.59083	10.8508	12.4426	15.4518
21	8.03366	8.89720	10.28293	11.5613	13.2396	16.3444
22	8.64272	9.54279	10.3923	12.3380	14.0415	17.2396
23	9.26042	10.19567	11.6885	13.0905	14.8479	18.1373
24	9.88623	10.8564	12.4011	13.8484	15.6587	19.0372
25	10.5197	11.5240	13.1197	14.6114	16.4734	19.9393
26	11.1603	12.1981	13.8439	15.3791	17.2919	20.8434
27	11.8076	12.8786	14.5733	16.1513	18.1138	21.7494
28	12.4613	13.5648	15.3079	16.9279	18.9392	22.6572
29	13.1211	14.2565	16.0471	17.7083	19.7677	23.5666
30	13.7867	14.9535	16.7908	18.4926	20.5992	24.4776
40	20.7065	22.1643	24.4331	26.5093	29.0505	33.6603
50	27.9907	29.7067	32.3574	34.7642	37.6886	42.9421
60	35.5346	37.4848	40.4817	43.1879	46.4589	52.2938
70	43.2752	45.4418	48.7576	51.7393	55.3290	61.6983
80	51.1720	53.5400	57.1532	60.3915	64.2778	71.1445
90	59.1963	61.7541	65.6466	69.1260	73.2912	80.6247
100	67.3276	70.0648	74.2219	77.9295	82.3581	90.1332

Table B.3 (*continued*)

P / v	0.500	0.250	0.100	0.050	0.025	0.010	0.005
1	0.454937	1.32330	2.70554	3.84146	5.02389	6.63490	7.87944
2	1.38629	2.77259	4.60517	5.99147	7.37776	9.21034	10.5966
3	2.36597	1.10835	6.25139	7.81473	9.34840	11.3449	12.8381
4	3.35670	5.38527	7.77944	9.48773	11.1433	13.2767	14.8602
5	4.35146	6.62568	9.23635	11.0705	12.8325	15.0863	16.7496
6	5.34812	7.84080	10.6446	12.5916	14.4494	16.8119	18.5476
7	6.34581	9.03715	12.0170	14.0671	16.0128	18.4753	20.2777
8	7.34412	10.2188	13.3616	15.5073	17.5346	20.0902	21.9550
9	8.34283	11.3887	14.6837	16.9190	19.0228	21.6660	23.5893
10	9.34182	12.5489	15.9871	18.3070	20.4831	23.2093	25.1882
11	10.3410	13.7007	17.2750	19.6751	21.9200	24.7250	26.7569
12	11.3403	14.8454	18.5494	21.0261	23.3367	26.2170	28.2995
13	12.3398	15.9839	19.8119	22.3621	24.7356	27.6883	29.8194
14	13.3393	17.1170	21.0642	23.6848	26.1190	29.1413	31.3193
15	14.3389	18.2451	22.3072	24.9958	27.4884	30.5779	32.8013
16	15.3385	19.3688	23.5418	26.2962	28.8454	31.9999	34.2672
17	16.3381	20.4887	24.4690	27.5871	30.1910	33.4087	35.7185
18	17.3379	21.6049	25.9894	28.8693	31.5264	34.8053	37.1564
19	18.3376	22.7178	27.2036	30.1435	32.8523	36.1908	38.5822
20	19.3374	23.8277	28.4120	31.4104	34.1696	37.5662	39.9968
21	20.3372	24.9348	29.6151	32.6705	35.4789	38.9321	41.4010
22	21.3370	26.0393	30.8133	33.9244	36.7807	40.2894	42.7956
23	22.3369	27.1413	32.0069	35.1725	38.0757	41.6384	44.1813
24	23.3367	28.2412	33.1963	36.4151	39.3641	42.9798	45.5585
25	24.3366	29.3389	34.3816	37.6525	40.6465	44.3141	46.9278
26	25.3364	30.4345	35.5631	38.8852	41.9232	45.6417	48.2899
27	26.3363	31.5284	36.7412	40.1133	43.1944	46.9630	49.6449
28	27.3363	32.6205	37.9159	41.3372	44.4607	48.2782	50.9933
29	28.3362	33.7109	39.0875	42.5569	45.7222	49.5879	52.3356
30	29.3360	34.7998	40.2560	43.7729	46.9792	50.8922	53.6720
40	39.3354	45.6160	51.8050	55.7585	59.3417	63.6907	66.7659
50	49.3349	56.3336	63.1671	67.5048	71.4202	76.1539	79.4900
60	59.3347	66.9814	74.3970	79.0819	83.2976	88.3794	91.9517
70	69.3344	77.5766	85.5271	90.5312	95.0231	100.425	104.215
80	79.3343	88.1303	96.5782	101.879	106.629	112.329	116.321
90	89.3342	98.6499	107.565	113.145	118.136	124.116	128.299
100	99.3341	109.141	118.498	124.342	129.561	135.807	140.169

Source: Reprinted, by permission of the Biometrika Trustees from C. M. Thompson, "Tables of Percentage Points of the χ^2 Distribution," Biometrika *32 (1941): 188–189*

Table B.4 *F*-distribution: 5 % points. The first column lists the number of denominator degrees of freedom (v_2). The headings of the other columns list the numerator degrees of freedom (v_1). The table entry is the value of c for which $P(F_{v_1,v_2} \geq c) = .05$

v_1 / v_2	1	2	3	4	5	6	7	8	9
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385
3	10.128	9.5521	9.2766	9.1172	9.0135	8.9406	8.8868	8.8452	8.8123
4	7.7086	6.9443	6.5914	6.3883	6.3560	6.1631	6.0942	6.0410	5.9988
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725
6	5.9874	5.1433	4.7571	4.5337	1.3874	4.2839	4.2066	4.1468	4.0990
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767
8	5.3177	4.4590	4.0662	3.8378	3.6875	3.5806	3.5005	3.4381	3.3881
9	5.1174	4.2565	3.8626	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789
10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962
12	4.7472	3.8856	3.4903	3.2592	3.1059	2.9961	3.9134	2.8486	2.7964
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144
14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876
16	4.4940	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377
17	4.4513	3.5915	3.1968	2.9647	2.8100	2.6987	2.6143	2.5480	2.4943
18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563
19	4.3808	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227
20	4.3513	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928
21	4.3248	3.4668	3.0725	2.8401	2.6848	2.5727	2.4876	2.4205	2.3661
22	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419
23	4.2793	3.4221	3.0280	2.7955	2.6400	2.5277	2.4422	2.3748	2.3201
24	4.2597	3.4028	3.0088	3.7763	2.6207	2.5082	2.4226	2.3551	2.3002
25	4.2417	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	2.3371	2.2821
26	4.2252	3.3690	2.9751	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655
27	4.2100	3.3541	2.9604	2.7278	2.5719	2.4591	2.3732	2.3053	2.2501
28	4.1960	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.2360
29	4.1830	3.3277	2.9340	2.7014	2.5454	2.4324	2.3463	2.2782	2.2229
30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107
40	4.0848	3.2317	2.8387	2.6060	2.4495	2.3359	2.2490	2.1802	2.1240
60	4.0012	3.1504	2.7581	2.5252	2.3683	2.2540	2.1665	2.0970	2.0401
120	3.9201	3.0718	2.6802	2.4472	2.2900	2.1750	2.0867	2.0164	1.9588
∞	3.8415	2.9957	2.6049	2.3719	2.2141	2.0986	2.0096	1.9354	1.8799

Table B.4 (*continued*)

v_1 / v_2	10	12	15	20	24	30	40	60	120	∞
1	241.88	243.91	245.95	248.01	249.05	250.09	251.14	252.20	253.25	254.32
2	19.396	19.413	19.429	19.446	19.454	19.462	19.471	19.479	19.487	19.496
3	8.7855	8.7446	8.7029	8.6602	8.6385	8.6166	8.5944	8.5720	8.5494	8.5265
4	5.9644	5.9117	5.8578	5.8025	5.7744	5.7459	5.7170	5.6878	5.6581	5.6281
5	4.7351	4.6777	4.6188	4.5581	4.5272	4.4957	4.4638	4.4314	4.3984	4.3650
6	4.0600	3.9999	3.9381	3.8742	3.8415	3.8082	3.7743	3.7398	3.7047	3.6688
7	3.6365	3.5747	3.5108	3.4445	3.4105	3.3758	3.3404	3.3043	3.2674	3.2298
8	3.3472	3.2840	3.2184	3.1503	3.1152	3.0794	3.0428	3.0053	2.9669	2.9276
9	3.1373	3.0729	3.0061	2.9365	2.9005	2.8637	2.8259	2.7872	2.7475	2.7067
10	2.9782	2.9130	2.8450	2.7740	2.7372	2.6996	2.6609	2.6211	2.5801	2.5379
11	2.8536	2.7876	2.7186	2.6464	2.6090	2.5705	2.5309	2.4901	2.4480	2.4045
12	2.7534	2.6866	2.6169	2.5436	2.5055	2.4663	2.4259	2.3842	2.3410	2.2962
13	2.6710	2.6037	2.5331	2.4589	2.4202	2.3803	2.3392	2.2966	2.2524	2.2064
14	2.6021	2.5342	2.4630	2.3879	2.3487	2.3082	2.2664	2.2230	2.1778	2.1307
15	2.5437	2.4753	2.4035	2.3275	2.2878	2.2468	2.2043	2.1601	2.1141	2.0658
16	2.4935	2.4247	2.3522	2.2756	2.2354	2.1938	2.1507	2.1058	2.0589	2.0096
17	2.4499	2.3807	2.3077	2.2304	2.1898	2.1477	2.1040	2.0584	2.0107	1.9604
18	2.4117	2.3421	2.2686	2.1906	2.1497	2.1071	2.0629	2.0166	1.9681	1.9168
19	2.3779	2.3080	2.2341	2.1555	2.1141	2.0712	2.0264	1.9796	1.9302	1.8780
20	2.3479	2.2776	2.2033	2.1242	2.0825	2.0391	1.9938	1.9464	1.8963	1.8432
21	2.3210	2.2504	2.1757	2.0960	2.0540	2.0102	1.9645	1.9165	1.8657	1.8117
22	2.2967	2.2258	2.1508	2.0707	2.0283	1.9842	1.9380	1.8895	1.8380	1.7831
23	2.2747	2.2036	2.1282	1.0476	2.0050	1.9605	1.9139	1.8649	1.8128	1.7570
24	2.2547	2.1834	2.1077	2.0267	1.9838	1.9390	1.8920	1.8424	1.7897	1.7331
25	2.2365	2.1649	2.0889	2.0075	1.9643	1.9192	1.8718	1.8217	1.7684	1.7110
26	2.2197	2.1479	2.0716	1.9898	1.9464	1.9010	1.8533	1.8027	1.7488	1.6906
27	2.2043	2.1323	2.0558	1.9736	1.9299	1.8842	1.8361	1.7851	1.7307	1.6717
28	2.1900	2.1179	2.0411	1.9586	1.9147	1.8687	1.8203	1.7689	1.7138	1.6541
29	2.1768	2.1045	2.0245	1.9446	1.9005	1.8543	1.8055	1.7537	1.6981	1.6377
30	2.1646	2.0921	2.0148	1.9317	1.8874	1.8409	1.7918	1.7396	1.6835	1.6223
40	2.0772	2.0035	1.9245	1.8389	1.7929	1.7444	1.6928	1.6373	1.5766	1.5089
60	1.9926	1.9174	1.8364	1.7480	1.7001	1.6491	1.5943	1.5343	1.4673	1.3893
120	1.9105	1.8337	1.7505	1.6587	1.6084	1.5543	1.4952	1.4290	1.3519	1.2539
∞	1.8307	1.7522	1.6664	1.5705	1.5173	1.4591	1.3940	1.3180	1.2214	1.0000

Source: Reprinted, by permission of the Biometrika Trustees from M. Merrington and C. M. Thompson, "Tables of Percentage Points of the Inverted Beta(F) Distribution," Biometrika *33 (1943): 80–81*

Table B.5 *F*-distribution: 1 % points. The first column lists the number of denominator degrees of freedom (v_2). The headings of the other columns list the numerator degrees of freedom (v_1). The table entry is the value of *c* for which $P(F_{v_1,v_2} \geq c) = .01$

v_1 / v_2	1	2	3	4	5	6	7	8	9
1	4,052.2	4,999.5	5,403.3	5,624.6	5,763.7	5,859.0	5,928.3	5,981.6	6,022.5
2	98.503	99.000	99.166	99.249	99.299	99.332	99.356	99.374	99.388
3	34.116	30.817	29.457	28.710	28.237	27.911	27.672	27.489	27.345
4	21.198	18.000	16.694	15.977	15.522	15.207	14.976	14.799	14.659
5	16.258	13.274	12.060	11.392	10.967	10.672	10.456	10.289	10.158
6	13.745	10.925	9.7795	9.1483	8.7459	8.4661	8.2600	8.1016	7.9761
7	12.246	9.5466	8.4513	7.8467	7.4604	7.1914	6.9928	6.8401	6.7188
8	11.259	8.6491	7.5910	7.0060	6.6318	6.3707	6.1776	6.0289	5.9106
9	10.561	8.0215	6.9919	6.4221	6.0569	5.8018	5.6129	5.4671	5.3511
10	10.044	7.5594	6.5523	5.9943	5.6363	5.3858	5.2001	5.0567	4.9424
11	9.6460	7.2057	6.2167	5.6683	5.3160	5.0692	4.8861	4.7445	4.6315
12	9.3302	6.9266	5.9526	5.4119	5.0643	4.8206	4.6395	4.4994	4.3875
13	9.0738	6.7010	5.7394	5.2053	4.8616	4.6204	4.4410	4.3021	4.1911
14	8.8616	6.5149	5.5639	5.0354	4.6950	4.4558	4.2779	4.1399	4.0297
15	8.6831	6.3589	5.4170	4.8932	4.5556	4.3183	4.1415	4.0045	3.8948
16	8.5310	6.2262	5.2922	4.7726	4.4374	4.2016	4.0259	3.8896	3.7804
17	8.3997	6.1121	5.1850	4.6690	4.3359	4.1015	3.9267	3.7910	3.6822
18	8.2854	6.0129	5.0919	4.5790	4.2479	4.0146	3.8406	3.7054	3.5971
19	8.1850	5.9259	5.0103	4.5003	4.1708	3.9386	3.7653	3.6305	3.5225
20	8.0906	5.8489	4.9382	4.4307	4.1027	3.8714	3.6987	3.5644	3.4567
21	8.0166	5.7804	4.8740	4.3688	4.0421	3.8117	3.6396	3.5056	3.3981
22	7.9454	5.7190	4.8166	4.3134	3.9880	3.7583	3.5867	3.4530	3.3458
23	7.8811	5.6637	4.7649	4.2635	3.9392	3.7102	3.5390	3.4057	3.2986
24	7.8229	5.6131	4.7181	4.2184	3.8951	3.6667	3.4959	3.3629	3.2560
25	7.7689	5.5680	4.6755	4.1774	3.8550	3.6272	3.4568	3.3239	3.2172
26	7.7213	5.5263	4.6366	4.1400	3.8183	3.5911	3.4210	3.2884	3.1818
27	7.6767	5.4881	4.6009	4.1056	3.7848	3.5580	3.3882	3.2558	3.1494
28	7.6356	5.4529	4.5681	4.0740	3.7539	3.5276	3.3581	3.2259	3.1195
29	7.5976	5.4205	4.5378	4.0449	3.7254	3.4995	3.3302	3.1982	3.0920
30	7.5625	5.3904	4.5097	4.0179	3.6990	3.4735	3.3045	3.1726	3.0665
40	7.3141	5.1785	4.3126	3.8283	3.5138	3.2910	3.1238	2.9930	2.8876
60	7.0771	4.9774	4.1259	3.6491	3.3389	3.1187	2.9530	2.8233	2.7185
120	6.8510	4.7865	3.9493	3.4796	3.1735	2.9559	2.7918	2.6629	2.5586
∞	6.6349	4.6052	3.7816	3.3192	3.0173	2.8020	2.6393	2.5113	2.4073

Table B.5 (*continued*)

ν_1 / ν_2	10	12	15	20	24	30	40	60	120	∞
1	6,055.8	6,106.3	6,157.3	6,208.7	6,234.6	6,260.7	6,286.8	6,313.0	6,339.4	6,366.0
2	99.399	99.416	99.449	99.458	99.458	99.466	99.474	99.483	99.491	99.501
3	27.229	27.052	26.872	26.690	26.598	26.505	26.411	26.316	26.221	26.125
4	14.546	14.374	14.198	14.020	13.929	13.838	13.745	13.652	13.558	13.463
5	10.051	9.8883	9.7222	9.5527	9.4665	9.3793	9.2912	9.2020	9.1118	9.0204
6	7.8741	7.7183	7.5590	7.3958	7.3127	7.2285	7.1432	7.0568	6.9690	6.8801
7	6.6201	6.4691	6.3143	6.1554	6.0743	5.9921	5.9084	5.8236	5.7372	5.6495
8	5.8143	5.6668	5.5151	5.3591	5.2793	5.1981	5.1156	5.0316	4.9460	4.8588
9	5.2565	5.1114	4.9621	4.8080	4.7290	4.6486	4.5667	4.4831	4.3978	4.3105
10	4.8492	4.7059	4.5582	4.4054	4.3269	4.2469	4.1653	4.0819	3.9965	3.9090
11	4.5393	4.3974	4.2509	4.0990	4.0209	3.9411	3.8596	3.7761	3.6904	3.6025
12	4.2961	4.1553	4.0096	3.8584	3.7805	3.7008	3.6192	3.5355	3.4494	3.3608
13	4.1003	3.9603	3.8154	3.6646	3.5868	3.5070	3.4253	3.3413	3.2548	3.1654
14	3.9394	3.8001	3.6557	3.5052	3.4274	3.3476	3.2656	3.1813	3.0942	3.0040
15	3.8049	3.6662	3.5255	3.3719	3.2940	3.2141	3.1319	3.0471	2.9595	2.8684
16	3.6909	3.5527	3.4089	3.2588	3.1808	3.1007	3.0182	2.9330	2.8447	2.7528
17	3.5931	3.4552	3.3117	3.1615	3.0835	3.0032	2.9205	2.8348	2.7459	2.6530
18	3.5082	3.3706	3.2273	3.0771	2.9990	2.9185	3.8354	2.7493	2.6597	2.5660
19	3.4338	3.2965	3.1533	3.0031	2.9249	2.8442	2.7608	2.6742	2.5839	2.4893
20	3.3682	3.2311	3.0880	2.9377	2.8594	2.7785	2.6947	2.6077	2.5168	2.4212
21	3.3098	3.1729	3.0299	2.8796	2.8011	2.7200	2.6359	2.5484	2.4568	2.3603
22	3.2576	3.1209	2.9780	2.8274	2.7488	2.6675	2.5831	2.4951	2.4029	2.3055
23	3.2106	3.0740	2.9311	2.7805	2.7017	2.6202	2.5355	2.4471	2.3542	2.2559
24	3.1681	3.0316	2.8887	2.7380	2.6591	2.5773	2.4923	2.4035	2.3099	2.2107
25	3.1294	2.9331	2.8502	2.6993	2.6203	2.5383	2.4530	2.3667	2.2695	2.1694
26	3.0941	2.9576	2.8150	2.6640	2.5848	2.5026	2.4170	2.3273	2.2325	2.1315
27	3.0618	2.2956	2.7827	2.6316	2.5522	2.4699	2.3840	2.2938	2.1984	2.0965
28	3.0320	2.8959	2.7530	2.6017	2.5223	2.4397	2.3535	2.2629	2.1670	2.0642
29	3.0045	2.8685	2.7256	2.5742	2.4946	2.4118	2.3253	2.2344	2.1378	2.0342
30	2.9791	2.8431	2.7002	2.5487	2.4689	2.3680	2.2992	2.2079	2.1107	2.0062
40	2.8005	2.6648	2.5216	2.3689	2.2880	2.2034	2.1142	2.0194	1.9172	1.8047
60	2.6318	2.4961	2.3523	2.1978	2.1154	2.0285	1.9360	1.8363	1.7263	1.6006
120	2.4721	2.3363	2.1915	2.0346	1.9500	1.8600	1.7628	1.6557	1.5330	1.3805
∞	2.3209	2.1848	2.0385	1.8783	1.7908	1.6964	1.5923	1.4730	1.3246	1.0000

Source: Reprinted, by permission of the Biometrika Trustees from M. Merrington and C. M. Thompson, "Tables of Percentage Points of the Inverted Beta(F) Distribution," Biometrika *33 (1943): 84–85*

Index

A

Almost-sure convergence
  Cauchy's condition, 253–254
  characterization, 251
  continuous functions, 253
  definition, 250–251
  independent random variables, 252
  not imply convergence, 252
  and plim, 251–252
  with probability 1, 251
  to random variable, 253
  sums and products, 253
Asymptotically equivalent tests, 626
Asymptotically unbiased tests, 624
Asymptotic distribution
  aggregate demand, beverage product, 290
  binomial density, 291
  CLTs (*see* Central limit theorems (CLTs))
  color screens for notebook computers, 287–288
  continuity correction, 291
  customers, grocery store, 288–289
  exponential distribution, 290
  gamma density, 289
  laws of large numbers (*see* Laws of large numbers)
  non-random and random sequence, 232
  normally distributed random variables (*see* Delta method)
  n-variate random variables, 231–232
  probability distribution, 231
  proofs and proof references, theorems, 276–287
  random variables, convergence (*see* Convergence of random variables)
Asymptotic pivotal quantity, 664
Asymptotic properties, estimators
  consistent
    population
      mean, 387
    sampling density, 387
    sufficient condition, 387–389
  CUAN, 391
  definition, 385
  nonuniqueness problem, 386
  relative efficiency and admissibility
    CAN (*see* Consistent asymptotically normal (CAN) estimators)
    exponential probability distribution, 389
    function, normal random variable, 390
  Slutsky's theorem, 386
Asymptotics, GLR tests
  consistency of
    for binomial distribution, 623–624
    description, 622–623
    for exponential distribution, 623
  description, 622
  distribution
    under local alternatives, 625–627, 685–686
    under null hypothesis, 624–625, 683–685
Axiomatic approach
  certain event probability, 11–12
  countable additivity, 13–14

Axiomatic approach (*cont.*)
event space, 10
mathematical induction principle, 12–13
nonnegativity, 11
probability set function, 14–15
probability space, 16
properties, 9–10

B

Bayes' rule
blood test accuracy, 31–33
description, 30
total probability, 30–31
two event case, 31
Behrens-Fisher problems, 690
Bernoulli, discrete density functions
density definition, 177
density, p = .3., 178
moments and MGF, 177
outcomes, 178
parameterization, 177
probability, dichotomous outcome, 178
probability space, 177–178
Bernoulli distribution
admissible estimators, mean, 378–379
MVUE, mean, 383–384
population
Lehman-Scheffé minimal sufficiency approach, 398–399
sufficient statistic, 392–394
Best linear unbiased estimator (BLUE)
definition, 384
Gauss-Markov theorem, 441–442
linear unbiased estimator, ξ'β, 442
mean, population distribution, 384–385
Beta, continuous density functions
and binomial density, 201–202
density definition, 200
density shapes, 200, 201
order statistics, 201
parameterization, 200
PDFs, 200
proportion outcomes, 200, 202
Bias matrix, 377
Binomial, discrete density functions
density definition, 178
density, n = 5, p = 0.3., 179
n Bernoulli trials, 178
parameterization, 178
PDF for n = 5, p = .3, 179, 180
sequences, 179
Bivariate normal population distribution, 694
BLUE. *See* Best linear unbiased estimator (BLUE)
Bonferroni joint tests, 644–645

C

CAN estimators. *See* Consistent asymptotically normal (CAN) estimators
Cauchy-Schwarz inequality, 151
CDF. *See* Cumulative distribution function (CDF)
Central limit theorems (CLTs)
asymptotic distributions, 261–262
control charts, 288
definition, 261
dependent scalar random variables, 270–271
independent scalar random variables (*see* Independent scalar random variables, CLTs)
multivariate central limit theorems, 272–273
sequences, random variables, 261
triangular arrays (*see* Triangular arrays of random variables, CLTs)
Change of variables technique
continuous random variables, 333
log-normal distribution, 335–336
multivariate and invertible, 338–339
in noninvertible function case, 337–338
piecewise invertible and univariate, 336–337
univariate and invertible, 334–335
Characteristic function, MGFs, 146
Chebyshev's inequality, 135–136
Chi-square, continuous density functions
additivity, 198–199
χ^2 densities, 198
density definition, 197
inverse additivity, 199
nonintegral, 198
parameterization, 197
table, 199–200
Chi-square distribution, 736–737
CIs. *See* Confidence intervals (CIs)
Classical general linear model (GLM)
β and σ^2 parameters, 637
assumption violation
consistency sufficient condition 1, 457
definite symmetric matrix, 459
degree, multicollinearity, 461
generalized least squares estimator, 458
heteroskedasticity/autocorrelation, 456
Markov's inequality, 459
matrix theory lemma, 457
perfect multicollinearity, 460
properties, LS estimator, 455
simultaneous determination, Y and X outcomes, 462
symmetric matrix, 462
asymptotic normality, $\hat{\beta}$, 447–448, 509
average squared prediction error, 446
bivariate case, 437
BLUE, 441–442
classical assumptions, 463–464
coefficient of determination, 440–441
definition, 436
design matrix, 437
disturbance vector, 450, 509–510
homoskedasticity, 438
linear model representation, 443
matrix calculus, 439–440
matrix multiplication, 445
ε, multivariate normal
Bonferroni joint tests, 644–645
F-tests, 637–640
χ^2-tests, 645–648
T-tests, 640–643
nonnegative-valued random variable, 446
partitioned inversion, 445
point estimation problem, 452, 453
polynomial function, explanation variables, 438
population distribution, 452
probability model, 439
production process, 443
regression hyperplane, 438
sample mean and variance, 452
semiparametric models
asymptotic χ^2 tests, 648–649
asymptotic normal tests, 649–652
description, 648
tests, σ^2, 652–653
Slutsky's theorem, 449–450
$\hat{S}^2$–non-iid residuals, 451–452, 510–511
statistical characteristics, 436
transformed model, 444
zero autocovariance/autocorrelation, 438

Classical hypothesis testing theory
applications, 597
exponential class, densities, 569–582
monotone likelihood ratio approach, 563–569
multiparameter exponential class, UMPU
hypothesis testing, 585–586, 601–602
nuisance parameters, 582–583
parameterization, exponential class density, 586
population distribution, 586
power functions, 584–585
random variable's distribution, 587
testing one and two-sided hypothesis, 589
Neyman-Pearson approach, 553–563
statistical theory and procedures, 596
Classical probability
application, 6
and fair dice, 6–7
CLTs. *See* Central limit theorems (CLTs)
Composite hypotheses, GLR tests
in Bernoulli population, 620–622
for exponential distribution, 619–620
vs. simple null, 618
UMP and unbiased, 620
UMP level α H_o *vs.* H_a, 618–619
Conditional density functions
bivariate case, 85–87
conditional CDFs, 85–87
description, 84–85
elementary event, 87
mean value theorem, integrals, 88
Conditional expectation, random variables
analysis, regression, 126
bivariate case, 126–127
expectation-bivariate, 126
generalized iterated theorem, 127–129
regression, multivariate and bivariate case, 129–132
substitution theorem, 128
and unconditional (*see* Unconditional expectation, random variables)
Conditional probability
description, 19–20
and frequency definition, 21
multiplication rule (*see* Multiplication rule)
sample selection, 23
sample space, 20
theorems, 22
Confidence intervals (CIs)
definitions, one-sided/bound, 655–656
description, 654
duality with rejection regions
description, 656–657
ellipsoid, GLM, 660–661
θ, exponential distribution, 659
μ, normal distribution, 657
p, Bernoulli distribution, 659–660
Rβ, GLM, 657–658
hypothesis testing, 654
level γ and coefficient 1–α, 655
lower and/or upper limits, 655
pivotal quantities, 654
properties of, confidence regions (*see* Confidence regions)
random set A(**X**), 655–656
scalar **R**(**Θ**), 655
unknown **R**(**Θ**), 656
Confidence regions
consistent level γ, 663
and hypothesis tests, relationships, 661
from pivotal quantities (*see* Pivotal quantities, confidence regions)
statistical tests, 667
UMA level γ, 662
UMAU, 662
unbiased, 661–662
Consistent asymptotically normal (CAN) estimators
asymptotic relative efficiency and admissibility, 389
definition, 388
variance/covariance matrix, 390, 391
Consistent Uniformly Asymptotically Normal (CUAN) class, 391
Continuous density functions
Beta, 200–202
gamma, exponential and chi-square subfamilies, 193–195
gamma subfamily name
chi-square, 197–200
exponential, 195–197
logistic, 202–203
uniform, 192–193
Continuous random variables
characteristics, 54–55
class of, 57
fundamental theorem, calculus, 54
lower and upper bounded events, 56–57
real-valued sample space, 53–54
Convergence of random variables
almost-sure convergence, 250–254
in distribution (*see* Distribution, convergence)
in mean square/quadratic mean, 247–250
modes relationships, 254
in probability (*see* Probability, convergence)
Convergence with probability 1. *See* Almost-sure convergence
Correlation, sample
definition, 321, 322
linear association, outcomes, 322–323
matrix, 325–326
properties, 323–324
Covariance, sample
definition, 319
matrix, 325
outcomes, moments, 319
properties of, 319–321
Cramer-Rao Lower Bound (CRLB)
asymptotic efficiency, 422–423
attainment theorem
and exponential class, 421–422
exponential population, 421
regularity conditions, 420
definition, 411
estimation, parameter vector, 418
exponential population, 419–420
Hessian form
evaluation, applicability, 417–418
exponential class, 416–417
regularity conditions, 416
joint density function, random sample, 414–415
regularity
conditions, 411–412
exponential class, distributions, 413
exponential population, 412–413
geometric distribution, exponential class, 413–414
scalar function case, 418
Cramer-Wold device
Bernoulli vectors, 272–273
description, 272
independent bounded random vectors, 273
multivariate Lindberg-Levy CLT, 272
normal limiting distributions, 272
CRLB. *See* Cramer-Rao Lower Bound (CRLB)
CUAN class. *See* Consistent Uniformly Asymptotically Normal (CUAN) class

Cumulant-generating function
definition, 147
and multivariate MGFs
definition, 148
and marginal, 148–149
Cumulative distribution function (CDF)
discontinuity, 65–66
discrete, 63–64
formal definition, 63
mixed discrete continuous RV, 64
vs. PDFs
component, 68
description, 65–66
theorems, 67
properties, 64–65
random variable, 63
sampling
gamma distribution, 331
Leibnitz's rules, 332
set definition, 62
truncation function, 64
Cumulative normal distribution, 734

D

Degenerate random variable and density function, 155
Delta method
asymptotically normally distributed random variables, 273–274
Bernoulli RVs, 274–275
definition, 274
first-order partial derivatives, 274
products and ratios, 275–276
vector function, 275
Young's form, Taylor's theorem, 274
DeMorgan's laws, 704
Density functions. *See also* Univariate random variables
algebraic representation, 176
continuous (*see* Continuous density functions)
discrete (*see* Discrete density functions)
domain elements, 176
exponential class of, 220–222
functional forms of, 176
normal family of
conditional normal densities, 215–219
description, 203
marginal normal densities, 212–215
multivariate normal density, 208–212
univariate normal, 203–208
parameter space, 175
parametric families, 175
PDFs, 176
statistical applications, 175
Dependent scalar random variables, CLTs
definition, 270
least squares estimator, 271
m-dependence, 271
Discrete density functions
Bernoulli, 177–178
binomial, 178–180
hypergeometric, 188–190
multinomial, 180–181
multivariate hypergeometric, 190–191
negative binomial and geometric subfamily, 181–183
Poisson (*see* Poisson, discrete density functions)
uniform, 176–177
Discrete random variables
class of, 57
concept, 52
point function, 51–52
probability mass function, 53
set function, 50
Discrete uniform population distribution, 694
Distribution, convergence
asymptotic distribution
definition, 237–238
vs. limiting distribution, 238
vs. no limiting distribution, 238
to constant, 233–234
continuous functions
description, 238
limiting distribution, 238–239
scalar function, 239
vector function, 239
cumulative distribution function, 234–235
definition, 233
limiting density and distribution, 234
moment generating functions, 236
vs. in probability, 241–242
random sequence, 236
sequence, density functions, 236

E

EDF. *See* Empirical distribution function (EDF)
Empirical distribution function (EDF)
description, 306
multivariate case
definition, 311–312
properties, 312
sample moments
empirical substitution principle, 312–313
origin and mean, 313–315
scalar case
and binomial, 308–309
definition, 307
empirical probabilities, 311
Glivenko-Cantelli theorem, 311
properties, 309–310
Empirical substitution principle, 312–313, 318
Equality of proportions, 694
Equal variances, 690
Equivalent-events approach, 332–333
Estimand, 373
Estimators and estimator properties, point estimation theory
asymptotic (*see* Asymptotic properties, estimators)
and estimand, 373
and estimate, 373
evaluation, performance, 374–375
finite sample (*see* Finite samples, estimators)
procedure, 374
Events, probability theory
definition, 4
disjoint, 5–6
elementary, 4
occurrence, 5
Excess kurtosis, 138, 147
Expectations, random variables
approximation of area, under $f(x)$, 115
conditional (*see* Conditional expectation, random variables)
continuous mass, weightless rod, 114
continuous RV, 116
density function "weights", weightless rod, 111, 112
discrete case, 113
E(X) existence, bounded R(X), 117
function
concave/convex, 121
cumulative distribution function, 118
density function, 119
Jensen's inequalities, 121
probabilities, 120–121
profit function, 119–120
gain, insurance, 113–114

multivariate random variables, 123–125
scalar random variables, 122
Experiment, probability theory, 2–3
Exponential class, densities
definition, 220
differentiability and continuity, power functions, 570–571
integral/summation sign, 570
linear transformations
noncentral *t*-distribution, 594
polar coordination, 592
power function, test, 594
probability mass, 591
reparameterization, 590
student *t*-distribution, 593
t-statistic, 593
normal family C, 221–222
UMPU level
Bernoulli population distribution, 572–573
binomial and normal distributions, 582
nonlinear equation solver, 580
normal population distribution, 574–576
power function, 577–578
α-size rejection region, 578–579
test statistic, 581
two sided tests, 582, 600–601
Exponential, continuous density functions
density definition, 196
"good as new" electronics, 197
memoryless property, 196–197
nonnegative subset, 196
parameterization, 195

F

F-density
definition and derivation, 343–344
F-family, 344–346
F-distribution, 738–739
Finite samples
estimators
bias matrix, 377
BLUE (*see* Best linear unbiased estimator (BLUE))
definition, 375
estimator admissibility, 378
estimator bias, 376
mean, Bernoulli distribution, 378–379
MSE (*see* Mean square error (MSE))
MVUEs (*see* Minimum variance unbiased estimators (MVUEs))
relative efficiency, 376, 377, 380–381
SMSE (*see* Strong mean square error (SMSE))
unbiased estimator, 381–382
GLR tests
composite hypotheses (*see* Composite hypotheses, GLR tests)
simple hypotheses (*see* Simple hypotheses, GLR tests)
UMP/even unbiased, 616
F-tests
characteristic roots, 638–639
joint explanatory variable significance, 639–640
noncentral F-distribution, 638
small sample distribution, W, 638
Wald-type test, null hypothesis, 637
Function sequences
convergence, 728–729
order of magnitude, 729

G

Gamma, exponential and chi-square subfamilies
$\alpha = 3.$, 193, 194
additivity, 194–195
$\beta = 1$, 193
definition and functions, density, 193
inverse additivity, 195
parameterization, 193
scaling, random variables, 195
time probability, machine breakdowns, 194
Generalized likelihood ratio (GLR) tests
asymptotics (*see* Asymptotics, GLR tests)
constrained maximum likelihood, 616
finite sample (*see* Finite samples, GLR tests)
functions of $\mathbf{\Theta}$, 615
function values, 615
of size α, 615
test statistic $\lambda(\mathbf{x})$, 616
Generalized method of moments (GMM)
asymptotic covariance matrix, 507
estimator properties, 504
first-order Taylor series representation, 506
instrumental variables estimator., 505–506
nonrandom positive definite symmetric matrix, 504–505, 513–514
sample moment conditions, 502
two-stage least squares estimator, 508
vector function, 503
weighted squared Euclidean distance, 496
General random sampling
commonalities, 304–305
definition, 303–304
Glivenko-Cantelli theorem, 311
GLM. *See* Classical General Linear Model (GLM)
GLR. *See* Generalized likelihood ratio (GLR) tests
GMM. *See* Generalized method of moments (GMM)
χ^2 Goodness of fit test
likelihood function, multinomial distribution, 669
ML method, 670
multivariate, LLCLT, 669
null hypothesis, 668
population distribution, exponential PDF, 670–671
test rule, 670

H

Heuristic approach, statistical test construction
classical UMP and UMPU, 610
H_0 *vs.* H_a
Poisson based on asymptotic normality, 613–614
Poisson distribution, 610–611
principle, 610
significance, regression parameter, 611–613
test consistency, 614–615
Hypergeometric, discrete density functions
density definition, 188
drawing objects with replacement, 190
drawing randomly without replacement, 189
moments and MGF, 189
parameterization, 188
probability space, 189
Hypothesis testing theory

Hypothesis testing theory (*cont.*)
acceptable region, 529–530
classical (*see* Classical hypothesis testing theory)
control charting, 606
definition, 529
general procedure, 530, 531
noncentral *t*-distribution, 597–598
null and alternative hypotheses, 541–542
parametric and properties (*see* Parametric hypothesis testing)
Pareto distribution, 605–606
population distribution, 604
power function graph, 603, 606
randomized test, 604–605
rejection region, 529–530
simple/composite hypothesis, 603–604
statistical (*see* Statistical hypothesis)
test statistics, 540–541
type I, type II error
applications, 537
Bernoulli random variables, 535
binomial distribution, 536
error probabilities, 533
hypothesis testing applications, 533
ideal statistical test, 532
probabilities, type II error, 539
protection against type I error, 537
size, statistical test, 538
test decision errors, 540
true probability distribution, 531
UMPU tests, 604

I

Independence
probability theory
complement event pairs, 26–27
description, 25–26
vs. disjointness, 27–28
event case, 28–29
two event case, 26
Venn diagram, 29–30
random variables
bivariate, 91–94
concept, 90
n-variate, 94–95
random vectors and functions, 96–98
Independent and identically distributed (iid) random variables
scalar, 262–265
SLLNs, 258–259
WLLNs, 255–256
Independent scalar random variables, CLTs
Bernoulli-type, 264
bounded, 266
chi-square, 264–265
continuity correction, 264
least squares estimator, 266–267
Liapounov CLT, 267
LLCLT, 262–263, 265–266
non-IID uniform RVs, 267–268
Iterated expectation theorem, 127–129, 132

J

Jacobian matrix, 338, 340, 343–344
Jarque-Bera (JB) test
asymptotic distribution, 677
null hypothesis, 676
test statistics, 676
JB. *See* Jarque-Bera (JB) test
Jensen's inequalities, 121, 164
Joint density, random samples, 300, 304–305, 330–331
Joint moments, covariance and correlation
bivariate function dependence, 153
bounds, correlation, 152
calculation, 150–151
Cauchy-Schwarz inequality, 151
covariance bound, 151–152
definition, 150, 152
expectation, product, 151
and independence, 152–153
linear association and degeneracy
bivariate random variable, continuous case, 155, 156
degenerate random variable and density function, 155
equivalent probability statement, 154
line integrals, 156
Markov's inequality, 153–154
prediction, Y outcomes, 156–157
mean, 150
moments, 150
origin, 149

K

Kolmogorov's inequality, 258
Kolmogorov-Smirnov (K-S) test, 671–672
K-S. *See* Kolmogorov-Smirnov (K-S) test
Kurtosis, 138

L

Lagrangian multiplier (LM) tests
asymptotic chi-square-distributed test statistic, 632
estimators, $\boldsymbol{\Theta}$ and λ, 632
finite/asymptotic probability distribution, 631
functional constraints, 632
functions of $\boldsymbol{\Theta}$, 627
in MLE (*see* Maximum likelihood estimator (MLE))
objective function, 631–632
power considerations, 633
restricted and unrestricted estimation, probability model, 627
restriction, parameters, 633
vector, 633
Laws of large numbers
convergence behavior, 255
definition, 254–255
SLLNs, 257–260
WLLNs, 255–257
Least squares (LS) estimator
GLM (*see* Classical general linear model (GLM))
NLS, 467–471
normality
finite sample distributions, 464–465
MVUE property, 465–466
normal approximation, 467
regression models, 434–436
Lehmann–Scheffé completeness theorem, MVUEs, 423–424
Lehmann–Scheffé minimal sufficiency theorem
Bernoulli, 398–399
determination, statistics, 425–426
g-function, 397
population distribution
Bernoulli, 398–399
gamma, 399
uniform, 399–400
vector function, 398
Lilliefors tests, 673–674
Lindberg-Levy CLT (LLCLT)
description, 262
distributions and asymptotic distributions, 263
non-IID case, 265

LLCLT. *See* Lindberg-Levy CLT (LLCLT)
LM. *See* Lagrangian multiplier (LM) tests
Local alternatives
  consistent tests, 625
  GLR statistics, 625
  H_0: R(**Θ**) = r *vs.* H_a: R(**Θ**) ≠ r, 626, 631
  and noncentral chi-square distribution, 633
  random variable sequence, 626
  $R(\beta)-r = n^{-1/2}\,\Phi$, 650
  unrestricted and/or restricted MLEs, 637
  Wald test, 649
Log-normal distribution, 335–336

M

Marginal cumulative distribution functions (MCDFs), 84
Marginal normal densities, density functions
  block diagonal Cov (X) ⇒ independent subvectors, 219
  conditional, $N(\mu,\Sigma)$, 215–216
  density via reordering, 214–215
  description, 212
  diagonal Cov(X) ⇒ independence, 218–219
  for $N(\mu,\Sigma)$, 213–214
  partitioned inversion and determinants, 216–217
  PDF, linear combinations, 212
  trivariate normal, 214, 217–218
Marginal probability density functions (MPDF)
  bivariate case
    continuous, 82
    discrete, 80–82
    joint density function, 79
  definition, 79
  MCDFs, 84
  n-variate case
    continuous, 83–84
    description, 82–83
    discrete, 83
  probability space, 78–79
Markov's inequality, 134
Mathematical induction principle, 12–13
Maximum likelihood (ML)
  invariance principle, 472
  likelihood function, 472
  MLE (*see* Maximum likelihood estimator (MLE))
Maximum likelihood estimator (MLE)
  approach, 492
  asymptotic properties
    direct evaluation, 480
    exponential density function, 484
    exponential distribution, 488–489
    gamma distribution, 489–491
    gamma population distribution, 480–482
    hessian matrix, 486
    Jensen's inequality, 483
    maximization problem, 485–486
    normality and efficiency, 487
    normality-sufficient conditions, 487–488, 512
    parameter vectors, 484–485, 511
  attainment, CRLB, 477
  Bernoulli population distribution, 493
  exponential distribution, 474
  GLM, 494–495
  hypergeometric distribution, 476–477
  implementation, 472–474
  invariance principle, 492–494, 513
  LM tests
    asymptotic distribution, 631, 688
    first order conditions, 628
    H_0: $\alpha = 1$ *vs.* H_o: $\alpha \neq 1$, Gamma population distribution, 629–630
    H_0: R($\mathbf{\Theta}_0$)=r *vs.* H_a: R($\mathbf{\Theta}_0$)≠r, 628–629, 686–688
  Neyman Factorization theorem, 478
  normal distribution, 474–475
  parameterization, density functions., 491
  point estimation, 495–496
  reparameterization, density family, 492
  uniform distribution, 476
  unique unbiased/ bias-adjusted, 478–479
MCDFs. *See* Marginal cumulative distribution functions (MCDFs)
Means and variances, random variables
  and correlations calculation, 161–162
  covariance matrix, 158–161
  linear combination
    matrix, 158–160
    vector, 158, 159
Mean square, convergence
  to constant vector, 248
  correlation convergence, 248–249
  definition, 247
  description, 248
  probability and distribution, 249–250
Mean square error (MSE)
  matrix, estimator, 376–377
  and relative efficiency, 376
  scalar case, 376
Mean value theorem, integrals, 88
Memoryless property, 182
Method of moments (MOM)
  consistency and asymptotic normality, 501–502
  estimated sample moments, 496
  gamma and normal distribution parameters, 500
  gamma population distribution, 498–499
  normal distribution parameters, 499
  population distribution, 497
  sample moment conditions, 498
  statistical model, 497
MGFs. *See* Moment-generating functions (MGFs)
Minimal sufficient statistics
  definition, 395
  domain elements, 396
  identification, 426
  Lehmann–Scheffé theorem (*see* Lehmann–Scheffé minimal sufficiency theorem)
  nonredundant coordinate functions, 397
Minimum variance linear unbiased estimators (MVLUEs). *See* Best linear unbiased estimator (BLUE)
Minimum variance unbiased estimators (MVUEs)
  and complete sufficient statistics, 423–424
  CRLB (*see* Cramer-Rao Lower Bound (CRLB))
  mean, Bernoulli distribution, 383–384
  MVLUE/BLUE and efficiency, 381–382
  scalar, 408–409
  uniqueness, scalar and vector, 409–412
Mixed discrete-continuous random variables
  category, 60
  density function components, 61
Mixture distribution, 681
ML. *See* Maximum likelihood (ML)
MLE. *See* Maximum likelihood estimator (MLE)
MOM. *See* Method of moments (MOM)

Moment-generating functions (MGFs)
bivariate, 173
central moments, 144
continuous RV, 143–144
definition, 142
density function, 172
determination, mean and variance, 169
function *g(t)*, 143
gamma function, 169
identification, PDF, 145–146
inversion, 146–147
mean and variance, 173
multivariate and cumulant-generating function, 148
properties, 144
sampling distributions, 330–331
uniqueness, 145
Moments, random variables
binomial theorem, 138
central moments, 139
Chebyshev's inequality, 135–136
definition, 132–133
existence, 139–140
kurtosis and excess kurtosis, 138
Markov's inequality, 134
mean, random variable, 133
MGFs (*see* Moment-generating functions (MGFs))
nonmoment measures, probability density characteristics, 141–142
nonsymmetric density function, 139
origin, 169
origin, *r*th central moment, 133
skewed density functions, 138–139
standard deviation, 134
symmetric and skewed densities, 136
μ_3, symmetry and skewness, 137
variance, 134, 139
Monotone likelihood ratio approach
Bernoulli distribution, 564
density functions, 563
exponential class, densities, 564
Gamma distribution, 564–565
hypergeometric distribution, 565–566
UMP level
exponential distribution, 567–568
hypergeometric distribution, 568
nonexistence, 568
α tests, 566–567, 600
MPDF. *See* Marginal probability density functions (MPDF)
MSE. *See* Mean square error (MSE)
Multinomial, discrete density functions
density definitions, 180
moments and MGF, 180
parameterization, 180
probability, customer payment modes, 181
probability space, 180
sequences, *n* repetitions, 180
Multiplication rule
one-stage inspections, 24
three card draw, 25
two card draw, 24
two events, 24
two-stage inspections, 25
Multivariate CDFs and duality
concept, 74
continuous bivariate, 76–77
continuous multivariate PDFs, 75
discrete bivariate, 75–76
Multivariate central limit theorems. *See* Cramer-Wold device
Multivariate concepts, continuous case, 98–100
Multivariate hypergeometric, discrete density functions
binomial density, 191
density definition, 190
moments and MGF, 191
parameterization, 190
Multivariate mixed discrete-continuous and composite random variables, 78
Multivariate normal density, density functions
central limit theorems, 212
density definition, 208
iso-density ellipses of, 211
n-variate, 209
parameterization, 208
three-dimensions and an iso-density ellipse, 209, 210
Multivariate random variables
continuous, 70
discrete, 70
expectation of matrix, 125
independent random variables, 124
joint density function, 123
n-variate, 69
properties and classes
discrete and continuous, 70–71
support and range equivalence, 71
real-valued vector function, 68–69
MVUEs. *See* Minimum variance unbiased estimators (MVUEs)

N

Negative binomial and geometric subfamily
Bernoulli trials, 181–182
defect after x trials, 182–183
density definition, 181
meeting quota, x trials, 182
memoryless property, 182
moments and MGF, 181
parameterization, 181
Pascal distribution, 181
Neyman factorization theorem
density function, random sample, 394–395
joint density, 395
Neyman-Pearson approach
probability calculation, 554, 598–600
simple hypothesis, 553–554
UMP
Bernoulli, 554–556
exponential mean *vs.* composite alternative, 560–561
nonexistence, 561–562
rejection region, 559–560
two exponential means, 556–558
UMPU
Bernoulli random variable, 558–559
one-sided and two-sided alternative hypotheses, 563
Nonaxiomatic probability
classical, 6–7
relative frequency, 7–8
subjective, 8–9
Noncentral χ^2-distributions
definition, 680
mixture distribution, 681
parameterization, 680
power function, statistical tests, 681
Noncentral F-distribution
definition, 682
degrees of freedom, 682
parameterization, 682
power function, statistical tests, 682
Noncentral *t*-distribution
application, 597–598
moments, 597
parameterization, 597
Non-independent and identically distributed (non-iid) random variables
independent scalar random variables, 265–268
SLLNs, 259–260
WLLNs, 256–257

Nonlinear least squares (NLS) estimator
computational approaches, 468
functional flexibility, 467
Gauss-Newton method, 468
noise component moment, 471
random residual vector outcomes, 469–470
sampling properties, 468–469
statistical properties, 471
Nonparametric models
definition, 366
regression, 366–368
Nonparametric tests, distributional assumptions
description, 667
of functional forms
χ^2 goodness of fit test, 668–671
JB test, 676–677
K-S and Lilliefors Tests, 671–674
SW test, 674–676
iid assumption
runs, 677–678
WW runs test, 678–680
maintained hypothesis, 668
probability, density functions, 668

O

Order statistics
definition, 350
extreme values, sampling distribution, 353
*k*th order statistic, 350–351
observations, largest and smallest, 349–350
pairs, sampling distribution, 352
waiting time, 351–352
Outcome of random samples, 299, 301, 303, 304

P

Paired comparisons of mean, 691–692
Parameter estimation problems
model, 368
objectives, 367
semiparametric and nonparametric models, 368
Parametric families
continuous (*see* Continuous density functions)
discrete (*see* Discrete density functions)
Parametric hypothesis testing
Bernoulli distribution, 542
definition, 542
maintained hypothesis, 543–544
power function, 544–545
properties, statistical tests
asymptotic, 552
consistency, 549
power function, 545–546
P-values, 549–550
rejection region, 550
test size and level, 546–547
UMP, 547–549
unbiasedness, 547
Parametric models
definition, 365
regression, 366–368
Pascal distribution, 181
PDFs. *See* Probability density functions (PDFs)
Permutations and combinations
binomial theorem, 717
description, 715
factorial operation, 716
power set, 718
PIT. *See* Probability integral transformation (PIT)
Pittman drift, local alternatives, 625
Pivotal quantities, confidence regions
continuous population, PDFs, 665–666
definition, 663
location-scale parameter families, PDFs, 665
mean of power distribution, 666–667
method of, 663–665
Point and set function
binary relation, 708–709
Cartesian product, 707–708
concepts, 707
domain and range, function, 710
function, 709–710
indicator function, 714
inverse function, 712
real-valued function, 711
size, 714
Point estimation methods
BLUE, 518
CRLB, 434
daily cash transactions, 516–517
GMM (*see* Generalized Method of Moments (GMM))
least-squares estimator, 519
linear and nonlinear models, 520–521
LS estimator (*see* Least squares (LS) estimator)
method of ML (*see* Maximum Likelihood (ML))
minimal sufficient statistics, 520
MOM (*see* Method of moments (MOM))
MVUE, 518
population distribution, 517
statistical model, 516
t-distribution, 515–516
Point estimation theory
estimators
and estimator properties (*see* Estimators and estimator properties, point estimation theory)
probability, 430
unbiased/consistent, 430
Poisson distribution, 426
math coprocessor installation, personal computer, 425
mean and variance, estimator, 430
MVUEs (*see* Minimum variance unbiased estimators (MVUEs))
nonparametric models, 366–368
objectives, 364
parametric models, 365–368
population distribution mean, freshmen's I.Q., 427
population mean, 429–430
probability models (*see* Probability models)
proportion, voters, 426–427
random sampling, exponential population distribution, 428
resets, circuit breakers, 430–431
semiparametric models, 365–368
sufficient statistics (*see* Sufficient statistics, point estimation theory)
test plots, wheat, 428–429
toy manufacturer, MVUE estimation, 429
trigger mechanisms, air bags, 427–428
Poisson, discrete density functions
background and application, 183–184
density definitions, 183
moments and MGF, 183
number, machine breakdowns, 188
parameterization, 183
$\Rightarrow$ poisson density, 186–188
process and PDF, 185–186
x defects in *n* trials, 184–185

Population distribution
defective/nondefective keyboards, 300–301
definition, 299
EDF, 306
gamma family of densities, 301
marginal, 319
method-of-moments, 312
order statistics, 350
random sampling with replacement, 299–300
sample correlation, 322
Probability, convergence
to constant, 240
definition, 240
and in distribution
constant, 244–245
difference and convergence, 243
limiting distribution, 244
RVs and parameters, 245
Slutsky's theorem, 245, 246
vs. in distribution, 241–242
order of magnitude
bounded in probability, 246
definition, 246
random sequence, 246–247
real numbers/matrices, 247
outcomes, random variable, 239
plim operator
addition, multiplication and division, 243
of continuous function, 242
definition, 242
matrix functions, 243
scalar additive and multiplicative functions, 243
probability limits of matrices, 241
Probability density functions (PDFs)
continuous (*see* Continuous random variables)
discrete (*see* Discrete random variables)
support and range equivalence, 58–59
verification, 59–60
Probability integral transformation (PIT)
definition, 346
exponential distribution, 346
exponential RV, 347
inverse PIT, 347–348
uniform distribution, 346–347
Probability models
abbreviated notation, random variable, 365
definition, 364, 369
parametric functional specification, sampling distribution, 369–371
point estimation problem
parameter space, 371–372
phraseology, 373
Probability samples
commonalities in structure, 304–305
definition, 298–299
random sample, 299
sample outcomes, 296
Probability space
bivariate continuous RV, 73–74
bivariate discrete RV, 71–72
Probability theory
axiomatic definition (*see* Axiomatic approach)
Bayes' Rule, 30–33
belief functions, 2
concept, 1
conditional (*see* Conditional probability)
description, 1
digression, events, 33–34
events, 4–6
experiment (*see* Experiment, probability theory)
fiducial probability, 2
independence (*see* Independence, probability theory)
nonaxiomatic (*see* Nonaxiomatic probability)
outcome propensity, 1–2
sample space, 3–4
structural probability, 2
theorems, 16–19

Q

Quantile of X, 142

R

Random samples
definition, 299
general random sampling, 303–305
with replacement, 300–301
sample mean and variance
filling process, 329–330
linear and quadratic forms, 326–327
under normality, distribution, 329
normal population distribution, 328
t-distribution, 326
simple random sampling, 299–300
simulation (*see* Simulations, random samples)
size of, 299
without replacement, 302–303
Random sampling without replacement
characteristics, 302
defectives, 303
definition, 302
joint density, 302–303
Random sampling with replacement
defectives, 300–301
definition, 300
joint density, 300
Random variables
calculation, expected value, 169
CDF (*see* Cumulative distribution function (CDF))
commodity exchange, bivariate, 167
cumulant-generating function, 147–149
densities, sample spaces, 45
expectation (*see* Expectations, random variables)
game of chance, 168–169
joint moments, covariance and correlation (*see* Joint moments, covariance and correlation)
mean
bivariate, 174
median, mode and quantiles, 172
trivariate, 174
and variances, 157–162
moments (*see* Moments, random variables)
net dollar return, project, 171–172
PDFs, 173
positive semidefiniteness, 162–165
regression curve, tablet computers, 172–173
Stieltjes integration and implicit support convention, 117–118
variance, dollar return, 174
Rao-Blackwell theorem
scalar case, 401–402
vector case, 402
Real analysis elements
adding, subtracting and multiplying sequences, 725
boundedness and limits, matrices, 724–725
bounded sequence, real numbers, 723
combinations, sequences, 725
continuity properties, functions, 726–728

convergence properties, sequences, 726
probability density functions, 722
random variables sequences, 721
real number sequence, 722
real-valued matrix sequence, 724
sequences, 720
Real numbers
bounded sequence, 723
matrix, 724
set, 710
vectors, 722
Regression function, bivariate case
conditional expectation, 130
definition, 129
investment return, 130–131
Regression models
linear
explanatory variables, 435
levels, generality, 434
Weierstrass's approximation theorem, 435
nonlinear, 436
Relative frequency probability
coin tossing and, 7–8
event, 7

S

Sample moments
correlation (*see* Correlation, sample)
covariance (*see* Covariance, sample)
definition, 312
empirical substitution principle, 312–313
mean, 315–316
multivariate asymptotic normality, 316
in multivariate case, 318–319
origin and mean
from EDF, 314
on EDF, 313–314
properties of origin, 314–315
variance (*see* Sample variance)
Sample space, probability theory
continuous, 4
discrete, 4
element/member, 3–4
Sample variance
calculation, 356–357, 360
covariance matrix, 325
definition, 316, 317
properties, 317–318
and sample covariance, 323
and sample mean, 318
Sampling
characteristics, 297–298
distributions
CDF approach, 331–332
change of variables approach, 333–339
description, 330
equivalent-events approach, 332–333
MGF approach, 330–331
variables approach, 333–339
EDF (*see* Empirical distribution function (EDF))
error, 298
F-density, 343–346
moments (*see* Sample moments)
normal population distributions, 354
order statistics (*see* Order statistics)
population, 297
probability samples (*see* Probability samples)
random sample (*see* Random samples)
statistical inference, 295–296, 306
stochastic process, 297
t-density, 339–343
t-distribution, 326
Semiparametric models
definition, 365–366
regression, 366–368
Set theory
associative laws, 703
characteristics, 699
classifications, 701
commutative laws, 703
complements of complements, 702, 704
DeMorgan's Laws, 704
description, 698–699
distributive laws, 703
empty/null set, 701
equality of sets, 701
finite set, 701
idempotency laws, 703
intersection and union, complements, 702, 704
intersection, null set, 704
interval set notation, 706
multiple intersection notation, 705–706
multiple union notation, 705
mutually exclusive/disjoint sets, 702–703
set difference, 701
subset, 701
union, 702
universal set, 701
verbal rule, 699–700
Shapiro-Wilks (SW) test
description, 674
residuals of GLM, normality, 675–676
statistic calculation, 675
Simple hypotheses, GLR tests
of exponential distribution, 617–618
Neyman-Pearson lemma, 616–617
Simulations, random samples
Bernoulli distribution, 349
exponential distribution, 348–349
PIT (*see* Probability integral transformation (PIT))
SLLNs. *See* Strong law of large numbers (SLLNs)
Slutsky's theorem, 245, 246, 386
SMSE. *See* Strong mean square error (SMSE)
Statistical hypothesis
Bernoulli process, 525–526
composite
exponential population, 525
non-iid normally-distributed sample, 526
semiparametric model, 526–527
definition, 524–525
set-defining conditions, 528–529
Statistics
definition, 306
vs. nonobservables, 306
observable random variable, 306
statistical inference, 295–296
Stieltjes integration and support convention
expected value, 118
operator, 117–118
Strong law of large numbers (SLLNs)
description, 257–258
iid random variables, 258–259
Kolmogorov's inequality, 257–258
non-iid random variables, 259–260
Strong mean square error (SMSE)
relative efficiency and admissibility, 380–381
superiority, 379
Student's *t* distribution, 735
Substitution theorem, 128, 132
Sufficient statistics, point estimation theory

Sufficient statistics (*cont.*)
Bernoulli population distribution, 392–394
definition, 392
exponential class
completeness, 405–407
density functions, 400–401
invertible functions, 407–408
Lehmann–Scheffé minimal sufficiency theorem, 397–400
minimal sufficient statistics, 395–400
Neyman factorization theorem (*see* Neyman factorization theorem)
objective, 391
Rao-Blackwell theorem (*see* Rao-Blackwell theorem)
Summation and integration notation, 718–720
SW. *See* Shapiro-Wilks (SW) test

T

t-Density
convergence, 342
definition and derivation, 339–341
standard normal random variable, 341
symmetry, 342, 343
t-family, 341–342
χ^2-Tests, GLM
equal tails test, 646
GLR approach, 645–646
rejection regions, 647
variances, GLM under normality, 647–648
Test statistics
asymptotic tests, 541
convergence properties, 552
definition, 540
exponential class of densities, UMPU, 581
Neyman-Pearson Lemma, 558
Triangular arrays of random variables, CLTs
column, 269
definition, 268–269
least squares estimator, 270
Liapounov CLTs, 269–270
row averages, 269
T-tests
constant elasticity GLM, 642–643
F-statistic, 640
individual explanatory variable significance, 641–642
sequence, 641
unbiasedness, 640

U

UMA. *See* Uniformly Most Accurate (UMA) Level γ
UMAU. *See* Uniformly Most Accurate Unbiased (UMAU)
Unbiased confidence regions, 661–662
Unconditional expectation, random variables
bivariate RV linear function, iteration, 128, 129
nonlinear function, iteration, 128, 129
Unequal variances, 690
Uniform
continuous density functions
density definition, 192
parameterization, 192
probability spaces, 192
Teller services, 192–193
discrete density functions
density definition, 176
moments, 176
parameterization, 176
probabilities, correctly at random, 177
probability $f(x;N) = 1/N$, 176
Uniformly Most Accurate (UMA) Level γ, 662
Uniformly Most Accurate Unbiased (UMAU), 662
Uniformly most powerful (UMP) test
monotone likelihood ratio approach
exponential distribution, 567–568
hypergeometric distribution, 568
nonexistence, 568
α tests, 566–567
Neyman-Pearson approach
Bernoulli means, 554–556
exponential mean, composite alternative hypothesis, 560–561
nonexistence, 561–562
rejection region, 559–560
two exponential means, 556–558
Uniformly most powerful unbiased (UMPU) tests
exponential class of densities
Bernoulli population distribution, 572–573
binomial and normal distributions, 582
nonlinear equation solver, 580
normal population distribution, 574–576
power function, 577–578
α-size rejection region, 578–579
test statistic, 581
two-sided test, 571
multiparameter exponential class
hypothesis testing, 585–586
nuisance parameters, 582–583
parameterization, density, 586
population distribution, 586
power functions, 584–585, 588
random variable's distribution, 587
testing one and two-sided hypotheses, 589
Neyman-Pearson approach
Bernoulli random variable, 558–559
one-sided and two-sided alternative hypotheses, 563
Univariate normal, density functions
assign probabilities, 205–206
behavior of, 204
and chi-square, 206–208
density definition, 203
parameterization, 203
random variable, 205
standard normal distribution, 204–205
table, 208
Univariate random variables
concept, 46
equivalent events, 48–49
induced probability space, 50
outcome, 46–47
probability space, 47–48

V

Variances, 689–690

W

Wald tests
asymptotic distribution, under H_a: $\mathbf{R}(\boldsymbol{\Theta}) \neq$ r, 636
of asymptotic size α, 634–635
description, 633
finite sample properties, 634
GMM estimator, 636–637
outcome, unrestricted estimator, 634

R(Θ) values, 633–634
variance, exponential population, 635
Wald-Wolfowitz (WW) runs test, 678–680
Weak law of large numbers (WLLNs)
description, 255–256
random variables
iid, 255–256
non-iid, 256–257
Weierstrass's approximation theorem, 435
WLLNs. *See* Weak law of large numbers (WLLNs)
WW. *See* Wald-Wolfowitz (WW) runs test

Y

Young's form, first-order Taylor series expansion, 274

Z

ZAP Electric Co. manufactures electric circuit breakers, 38
Z-statistics
asymptotic level α tests
$R(\boldsymbol{\beta})$, 649
σ^2, 653
outcome, 651, 652
and Wald, 650