D.1 ALGEBRA OF SETS

The totality of objects under consideration is called the universal set and is denoted by \( S \). Each object in \( S \) is called an element of \( S \). If a set \( A \) is a collection of elements that are also in \( S \), then \( A \) is said to be a subset of \( S \). In applications of probability, \( S \) usually denotes the sample space. An event \( A \) is a collection of possible outcomes of an experiment and is a subset of \( S \). We say that event \( A \) has occurred if the outcome of the experiment is an element of \( A \). The set or event \( A \) may be described by listing all of its elements or by defining the properties that its elements must satisfy.

**EXAMPLE D.1-1**
A four-sided die, called a tetrahedron, has four faces that are equilateral triangles. These faces are numbered 1, 2, 3, 4. When the tetrahedron is rolled, the outcome of the experiment is the number of the face that is down. If the tetrahedron is rolled twice and we keep track of the first roll and the second roll, then the sample space is that displayed in Figure D.1-1 on the next page.

Let \( A \) be the event that the second roll is a 1 or a 2. That is,
\[
A = \{(x, y): \ y = 1 \text{ or } \ y = 2\}.
\]

Let
\[
B = \{(x, y): \ x + y = 6\} = \{(2, 4), (3, 3), (4, 2)\},
\]
and let
\[
C = \{(x, y): \ x + y \geq 7\} = \{(4, 3), (3, 4), (4, 4)\}.
\]

Events \( A \), \( B \), and \( C \) are shown in Figure D.1-1.
When \( a \) is an element of \( A \), we write \( a \in A \). When \( a \) is not an element of \( A \), we write \( a \notin A \). So, in Example D.1-1, we have \((3,1) \in A \) and \((1,3) \notin A \). If every element of a set \( A \) is also an element of a set \( B \), then \( A \) is a \textit{subset} of \( B \), and we write \( A \subseteq B \). In probability, if event \( B \) occurs whenever event \( A \) occurs, then \( A \subseteq B \). The two sets \( A \) and \( B \) are equal (i.e., \( A = B \)) if \( A \subseteq B \) and \( B \subseteq A \). Note that it is always true that \( A \subseteq A \) and \( A \subseteq S \), where \( S \) is the universal set. We denote the subset that contains no elements by \( \emptyset \). This set is called the \textit{null} or \textit{empty} set. For all sets \( A \), \( \emptyset \subseteq A \).

The set of elements of either \( A \) or \( B \) or possibly both \( A \) and \( B \) is called the \textit{union} of \( A \) and \( B \) and is denoted \( A \cup B \). The set of elements of both \( A \) and \( B \) is called the \textit{intersection} of \( A \) and \( B \) and is denoted \( A \cap B \). The \textit{complement} of a set \( A \) is the set of elements of the universal set \( S \) that are not in the set \( A \) and is denoted \( A' \).

The operations of union and intersection may be extended to more than two sets. Let \( A_1, A_2, \ldots, A_n \) be a finite collection of sets. Then the \textit{union}

\[
A_1 \cup A_2 \cup \cdots \cup A_n = \bigcup_{k=1}^{n} A_k
\]

is the set of elements that belong to at least one \( A_k \), \( k = 1, 2, \ldots, n \). The \textit{intersection}

\[
A_1 \cap A_2 \cap \cdots \cap A_n = \bigcap_{k=1}^{n} A_k
\]

is the set of all elements that belong to every \( A_k \), \( k = 1, 2, \ldots, n \). Similarly, let \( A_1, A_2, \ldots, A_n \) be a countable collection of sets. Then \( x \) belongs to the \textit{union}

\[
A_1 \cup A_2 \cup \cdots = \bigcup_{k=1}^{\infty} A_k
\]
if $x$ belongs to at least one $A_k$, $k = 1, 2, 3, \ldots$. Also, $x$ belongs to the intersection

$$A_1 \cap A_2 \cap A_3 \cap \cdots = \bigcap_{k=1}^{\infty} A_k$$

if $x$ belongs to every $A_k$, $k = 1, 2, 3, \ldots$.

**EXAMPLE D.1-2** Let

$$A_k = \left\{ x : \frac{10}{k+1} \leq x \leq 10 \right\}, \quad k = 1, 2, 3, \ldots$$

Then

$$\bigcup_{k=1}^{n} A_k = \left\{ x : \frac{10}{n} \leq x \leq 10 \right\};$$

$$\bigcup_{k=1}^{\infty} A_k = \{ x : 0 < x \leq 10 \}.$$

Note that the number zero is not in this latter union, since it is not in one of the sets $A_1, A_2, A_3, \ldots$. Also,

$$\bigcap_{k=1}^{n} A_k = \{ x : 5 \leq x \leq 10 \} = A_1$$

and

$$\bigcap_{k=1}^{\infty} A_k = \{ x : 5 \leq x \leq 10 \} = A_1,$$

since $A_1 \subseteq A_k$, $k = 1, 2, 3, \ldots$.

A convenient way to illustrate operations on sets is with a Venn diagram. In Figure D.1-2 on the next page, the universal set $S$ is represented by the rectangle and its interior, and the subsets of $S$ are represented by the points enclosed by the circles, as well as by the complement of the union of those subsets. The sets under consideration are the shaded regions.

Set operations satisfy several properties. For example, if $A$, $B$, and $C$ are subsets of $S$, we have the following laws:

**Commutative Laws:**

$$A \cup B = B \cup A$$

$$A \cap B = B \cap A$$

**Associative Laws:**

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

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

**Distributive Laws**

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

**De Morgan’s Laws**

$$(A \cup B)' = A' \cap B'$$

$$(A \cap B)' = A' \cup B'$$

A Venn diagram will be used to justify the first of De Morgan’s laws. In Figure D.1-3(a), $A \cup B$ is represented by horizontal lines, and thus $(A \cup B)'$ is the
region represented by vertical lines. In Figure D.1-3(b), $A'$ is indicated with horizontal lines and $B'$ is indicated with vertical lines. An element belongs to $A' \cap B'$ if it belongs to both $A'$ and $B'$. Thus the crosshatched region represents $A' \cap B'$. Clearly, this crosshatched region is the same as that shaded with vertical lines in Figure D.1-3(a).

### D.2 MATHEMATICAL TOOLS FOR THE HYPERGEOMETRIC DISTRIBUTION

Let $X$ have a hypergeometric distribution. That is, the p.m.f. of $X$ is

$$f(x) = \frac{\binom{N_1}{x} \cdot \binom{N_2}{n-x}}{\binom{N}{n}}$$

$$= \frac{\binom{N_1}{x} \cdot \binom{N_2}{n-x}}{\binom{N}{n}}, \quad x \leq n, \ x \leq N_1, \ n - x \leq N_2.$$
To show that \( \sum_{x=0}^{n} f(x) = 1 \) and to find the mean and variance of \( X \), we use the following theorem.

**Theorem D.2-1**

\[
\binom{N}{n} = \sum_{x=0}^{N} \binom{N_1}{x} \binom{N_2}{n-x}
\]

where \( N = N_1 + N_2 \) and it is understood that \( \binom{k}{j} = 0 \) if \( j > k \).

**Proof.** Because \( N = N_1 + N_2 \), we have the identity

\[
(1 + y)^N = (1 + y)^{N_1} (1 + y)^{N_2}.
\]  \hspace{1cm} (D.2-1)

We will expand each of these binomials, and since the polynomials on each side are identically equal, the coefficients of \( y^n \) on each side of Equation D.2-1 must be equal. Using the binomial expansion, we find that the expansion of the left side of Equation D.2-1 is

\[
(1 + y)^N = \sum_{k=0}^{N} \binom{N}{k} y^k
\]

\[
= \binom{N}{0} y^0 + \binom{N}{1} y + \cdots + \binom{N}{n} y^n + \cdots + \binom{N}{N} y^N.
\]

The right side of Equation D.2-1 becomes

\[
(1 + y)^{N_1} (1 + y)^{N_2} = \left[ \binom{N_1}{0} y^0 + \binom{N_1}{1} y + \cdots + \binom{N_1}{n} y^n + \cdots + \binom{N_1}{N_1} y^{N_1} \right] \times \left[ \binom{N_2}{0} y^0 + \binom{N_2}{1} y + \cdots + \binom{N_2}{n} y^n + \cdots + \binom{N_2}{N_2} y^{N_2} \right].
\]

The coefficient of \( y^n \) in this product is

\[
\binom{N_1}{0} \binom{N_2}{0} + \binom{N_1}{1} \binom{N_2}{1} y + \cdots + \binom{N_1}{n} \binom{N_2}{n} y^n + \cdots + \binom{N_1}{N_1} \binom{N_2}{N_2} y^{N_1 + N_2}
\]

and this sum must be equal to \( \binom{N}{n} \), the coefficient of \( y^n \) on the left side of Equation D.2-1.

Using Theorem D.2-1, we find that if \( X \) has a hypergeometric distribution with p.m.f. \( f(x) \), then

\[
\sum_{x=0}^{n} f(x) = \sum_{x=0}^{n} \binom{N_1}{x} \binom{N_2}{n-x} \binom{N}{n} = 1.
\]

To find the mean and variance of a hypergeometric random variable, it is useful to note that, with \( n > 0 \),

\[
\binom{N}{n} = \frac{N!}{n!(N-n)!} = \frac{N}{n} \frac{(N-1)!}{(n-1)!(N-n)!} = \frac{N}{n} \binom{N-1}{n-1}
\]
Appendix D  Review of Selected Mathematical Techniques

The mean of a hypergeometric random variable $X$ is

$$
\mu = \sum_{\xi=0}^{x(n)} \xi f(\xi) \\
= \sum_{\xi=0}^{x(n)} \frac{N_1!}{\xi! (N_1 - \xi)!} \frac{N_2!}{(n - \xi)! (N_2 - n + \xi)!} \binom{N}{n}
$$

$$
= \frac{N_1}{N} \sum_{\xi=0}^{x(n)} \frac{(N_1 - 1)!}{\xi! (N_1 - \xi)!} \frac{N_2!}{(n - \xi)! (N_2 - n + \xi)!} \binom{N}{n}
$$

If we now make the change of variables $k = x - 1$ in the summation and replace

$$
\binom{N}{n}
$$

in the denominator, the previous equation becomes

$$
\mu = \frac{N_1}{N} \sum_{k=0}^{N(N_1)} \frac{(N_1 - 1)!}{k! (N_1 - 1 - k)!} \frac{N_2!}{(n - k - 1)! (N_2 - n + k + 1)!} \binom{N}{n - 1}
$$

$$
= n \left( \frac{N}{N} \right) \sum_{k=0}^{N(N_1)} \frac{(N_1 - 1) (N_2)}{(n - 1 - k) \binom{N}{n - 1}} = n \left( \frac{N_1}{N} \right)
$$

because, from Theorem D.2-1, the summation in the expression for $\mu$ is equal to

$$
\binom{N}{n - 1}.
$$

Note that

$$
\text{Var}(X) = \sigma^2 = E[(X - \mu)^2] = E[X^2] - \mu^2 = E[X(X - 1)] + E(X) - \mu^2.
$$

So, to find the variance of $X$, we first find $E[X(X - 1)]$:

$$
E[X(X - 1)] = \sum_{x=0}^{n} x(x - 1)f(x) = \sum_{x=2}^{n} \frac{x(x - 1)}{x! (N_1 - x)!} \frac{N_1!}{(n - x)! (N_2 - n + x)!} \binom{N}{n}
$$
Section D.3 Limits

\[ E[X(X - 1)] = N_1(N_1 - 1) \sum_{k=0}^{n} \binom{N_1 - 2}{k} \binom{N_2}{n - 2 - k} \frac{N_2!}{n!(N - n)!} \left( \frac{N - 2}{n - 2} \right) \]

In the summation, let \( k = x - 2 \), and in the denominator, note that
\[ \binom{N}{n} = \frac{N!}{n!(N - n)!} = \frac{N(N - 1)}{n(n - 1)} \frac{N - 2}{n - 2} \]

Thus, from Theorem D.2-1,
\[ \sigma^2 = \frac{N_1(N_1 - 1)(n)(n - 1)}{N(N - 1)} + \frac{nN_1}{N} - \left( \frac{nN_1}{N} \right)^2 \]

Hence, the variance of a hypergeometric random variable is, after some algebraic manipulations,
\[ \sigma^2 = \frac{N_1(N_1 - 1)(n)(n - 1)}{N(N - 1)} + \frac{nN_1}{N} - \left( \frac{nN_1}{N} \right)^2 \]

D.3 LIMITS

We refer the reader to the many fine books on calculus for the definition of a limit and the other concepts used in that subject. Here we simply remind you of some of the techniques we find most useful in probability and statistics.

Early in a calculus course, the existence of the following limit, denoted by the letter \( e \), is discussed:
\[ e = \lim_{r \to 1^+} (1 + r)^{1/r} = \lim_{n \to \infty} \left( 1 + \frac{1}{n} \right)^n. \]

Of course, \( e \) is an irrational number, which, to six significant figures, equals 2.71828.

Often, it is rather easy to see the value of certain limits. For example, with \( -1 < r < 1 \), the sum of the geometric progression allows us to write
\[ \lim_{n \to \infty} (1 + r + r^2 + \cdots + r^{n-1}) = \lim_{n \to \infty} \left( \frac{1 - r^n}{1 - r} \right) = \frac{1}{1 - r}. \]

That is, the limit of the ratio \( (1 - r^n)/(1 - r) \) is not difficult to find because \( \lim_{n \to \infty} r^n = 0 \) when \( -1 < r < 1 \).

However, it is not that easy to determine the limit of every ratio; for example, consider
\[ \lim_{b \to \infty} (be^{-b}) = \lim_{b \to \infty} \left( \frac{k}{e^b} \right) \]

Copyright 2010 Pearson Education, Inc.
Appendix D  Review of Selected Mathematical Techniques

Since both the numerator and the denominator of the latter ratio are unbounded, we can use L'Hôpital's rule, taking the limit of the ratio of the derivative of the numerator to the derivative of the denominator. We then have

\[ \lim_{b \to \infty} \left( \frac{b}{n} \right) = \lim_{b \to \infty} \left( \frac{1}{n} \right) = 0. \]

This result can be used in the evaluation of the integral

\[ \int_0^\infty xe^{-x} \, dx = \lim_{b \to \infty} \int_0^b xe^{-x} \, dx \]
\[ = \lim_{b \to \infty} \left[ -xe^{-x} - e^{-x} \right]_0^b \]
\[ = \lim_{b \to \infty} \left[ 1 - be^{-b} - e^{-b} \right] = 1. \]

Note that

\[ D_e[ -xe^{-x} - e^{-x} ] = xe^{-x} - e^{-x} + e^{-x} = xe^{-x}; \]
that is, \(-xe^{-x} - e^{-x}\) is the antiderivative of \(xe^{-x}\).

Another limit of importance is

\[ \lim_{n \to \infty} \left( 1 + \frac{b}{n} \right)^n = \lim_{n \to \infty} e^{n \ln(1 + b/n)}, \]
where \(b\) is a constant.

Since the exponential function is continuous, the limit can be taken to the exponent. That is,

\[ \lim_{n \to \infty} e^{n \ln(1 + b/n)} = \exp \left( \lim_{n \to \infty} n \ln(1 + b/n) \right). \]

By L'Hôpital's rule, the limit in the exponent is equal to

\[ \lim_{n \to \infty} \frac{-b/n^2}{\ln(1 + b/n)} = \lim_{n \to \infty} \frac{1 + b/n}{-1/n^2} = \lim_{n \to \infty} \frac{b}{1 + b/n} = b. \]

Since this limit is equal to \(b\), the original limit is

\[ \lim_{n \to \infty} \left( 1 + \frac{b}{n} \right)^n = e^b. \]

Applications of this limit in probability occur with \(b = -1\), yielding

\[ \lim_{n \to \infty} \left( 1 - \frac{1}{n} \right)^n = e^{-1}. \]

D.4  infinite series

A function \(f(x)\) possessing derivatives of all orders at \(x = b\) can be expanded in the following Taylor series:

\[ f(x) = f(b) + \frac{f'(b)}{1!} (x - b) + \frac{f''(b)}{2!} (x - b)^2 + \frac{f'''(b)}{3!} (x - b)^3 + \cdots. \]
Section D.4 Infinite Series

If \( b = 0 \), we obtain the special case that is often called the **Maclaurin series**:

\[
f(x) = f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \cdots.
\]

For example, if \( f(x) = e^x \), so that all derivatives of \( f(x) = e^x \) are \( f^{(r)}(x) = e^x \), then \( f^{(r)}(0) = 1 \), for \( r = 1, 2, 3, \ldots \). Thus, the Maclaurin series expansion of \( f(x) = e^x \) is

\[
e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots.
\]

The **ratio test**,

\[
\lim_{n \to \infty} \left| \frac{x^{n+1}/(n+1)!}{x^n/(n-1)!} \right| = \lim_{n \to \infty} \left| \frac{x}{n} \right| = 0,
\]

shows that the Maclaurin series expansion of \( e^x \) converges for all real values of \( x \).

Note, for examples, that

\[
e = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots
\]

and

\[
e^{-1} = 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + \frac{(-1)^n}{n!} + \cdots.
\]

As another example, consider

\[
h(w) = (1 - w)^{-r},
\]

where \( r \) is a positive integer. Here

\[
h'(w) = r(1 - w)^{-(r+1)},
\]

\[
h''(w) = (r)(r + 1)(1 - w)^{-(r+2)},
\]

\[
h'''(w) = (r)(r + 1)(r + 2)(1 - w)^{-(r+3)},
\]

\[
\vdots
\]

In general, \( h^{(k)}(0) = (r)(r + 1) \cdots (r + k - 1) = (r + k - 1)!/(r - 1)! \). Thus,

\[
(1 - w)^{-r} = 1 + \frac{(r + 1 - 1)!}{r - 1}w + \frac{(r + 2 - 1)!}{(r - 1)!}w^2 + \cdots + \frac{(r + k - 1)!}{(r - 1)!}w^k + \cdots
\]

\[
= \sum_{k=0}^{\infty} \binom{r + k - 1}{r - 1}w^k.
\]

This is often called the negative binomial series. Using the ratio test, we obtain

\[
\lim_{n \to \infty} \left| \frac{w^n(r + n - 1)!/[r - 1]!n!}{w^{n-1}(r + n - 2)!/[r - 1]!(n - 1)!} \right| = |w|.
\]

Thus, the series converges when \(|w| < 1\), or \(-1 < w < 1\).
Appendix D Review of Selected Mathematical Techniques

A negative binomial random variable receives its name from this negative binomial series. Before showing that relationship, we note that, for $-1 < w < 1$,

\[
h(w) = \sum_{k=0}^{\infty} \binom{r + k - 1}{r - 1} w^k = (1 - w)^{-r},
\]

\[
h'(w) = \sum_{k=1}^{\infty} \binom{r + k - 1}{r - 1} k w^{k-1} = r(1 - w)^{-r-1},
\]

\[
h''(w) = \sum_{k=2}^{\infty} \binom{r + k - 1}{r - 1} k(k-1) w^{k-2} = r(r + 1)(1 - w)^{-r-2}.
\]

The p.m.f. of a negative binomial random variable $X$ is

\[
g(x) = \binom{x - 1}{r - 1} p^r q^{x - r}, \quad x = r, r + 1, r + 2, \ldots.
\]

In the series expansion for $h(w) = (1 - w)^{-r}$, let $x = k + r$. Then

\[
\sum_{x=r}^{\infty} \binom{x - 1}{r - 1} w^{x-r} = (1 - w)^{-r}.
\]

Letting $w = q$ in this equation, we see that

\[
\sum_{x=r}^{\infty} g(x) = \sum_{x=r}^{\infty} \binom{x - 1}{r - 1} p^r q^{x-r} = p'(1 - q)^{-r} = 1.
\]

That is, $g(x)$ does satisfy the properties of a p.m.f.

To find the mean of $X$, we first find

\[
E(X - r) = \sum_{x=r}^{\infty} (x - r) \binom{x - 1}{r - 1} p^r q^{x-r} = \sum_{x=r+1}^{\infty} (x - r) \binom{x - 1}{r - 1} p^r q^{x-r}.
\]

Letting $k = x - r$ in this latter summation and using the expansion of $h'(w)$ gives us

\[
E(X - r) = \sum_{k=1}^{\infty} \binom{r + k - 1}{r - 1} p^r q^k
\]

\[
= p' q \sum_{k=1}^{\infty} \binom{r + k - 1}{r - 1} k q^{k-1}
\]

\[
= p' q r (1 - q)^{-r} = r \left( \frac{q}{p} \right).
\]

Thus,

\[
E(X) = r + r \left( \frac{q}{p} \right) = r \left( 1 + \frac{q}{p} \right) = r \left( \frac{1}{p} \right).
\]
Similarly, using \( h''(w) \), we can show that
\[
E[(X - r)(X - r - 1)] = \left( \frac{q^2}{p^2} \right) (r)(r + 1).
\]

Hence,
\[
\text{Var}(X) = \text{Var}(X - r) = \left( \frac{q^2}{p^2} \right) (r)(r + 1) + r \left( \frac{q^2}{p^2} \right) - r \left( \frac{q^2}{p^2} \right) = r \left( \frac{q^2}{p^2} \right)
\]

A special case of the negative binomial series occurs when \( r = 1 \), whereupon we obtain the well-known geometric series
\[
(1 - w)^{-1} = 1 + w + w^2 + w^3 + \cdots,
\]
provided that \(-1 < w < 1\).

The geometric series gives its name to the geometric probability distribution. Perhaps you recall the geometric series
\[
g(r) = \sum_{k=0}^\infty ar^k = \frac{a}{1 - r}, \quad (D.4-1)
\]
for \(-1 < r < 1\). To find the mean and variance of a geometric random variable \( X \), simply let \( r = 1 \) in the respective formulas for the mean and variance of a negative binomial random variable. However, if you want to find the mean and variance directly, you can use
\[
g'(r) = \sum_{k=1}^\infty akr^{k-1} = \frac{a}{(1 - r)^2}, \quad (D.4-2)
\]
and
\[
g''(r) = \sum_{k=2}^\infty ak(k-1)r^{k-2} = \frac{2a}{(1 - r)^3}, \quad (D.4-3)
\]
to find \( E(X) \) and \( E[X(X - 1)] \), respectively.

In applications associated with the geometric random variable, it is also useful to recall that the \( n \)th partial sum of a geometric series is
\[
x_n = \sum_{k=0}^{n-1} ar^k = \frac{a(1 - r^n)}{1 - r}.
\]

A bonus in this section is a logarithmic series that produces a useful tool in daily life. Consider
\[
f(x) = \ln(1 + x),
\]
\[
f'(x) = (1 + x)^{-1},
\]
\[
f''(x) = (-1)(1 + x)^{-2},
\]
\[
f'''(x) = (-1)(-2)(1 + x)^{-3},
\]
\[
\vdots.
\]
Appendix D  Review of Selected Mathematical Techniques

Thus, \( f^{(r)}(0) = (-1)^{r-1}(r-1)! \) and
\[
\ln(1 + x) = \frac{0!}{1!} x - \frac{1!}{2!} x^2 + \frac{2!}{3!} x^3 - \frac{3!}{4!} x^4 + \cdots \\
= x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots ,
\]
which converges for \(-1 < x < 1\).

Now consider the following question: “How long does it take money to double in value if the interest rate is \( i \)?” Assuming that compounding is on an annual basis and that you begin with $1, after one year you have \( 1 \times (1 + i) \), and after two years the number of dollars you have is
\[
(1 + i) + i(1 + i) = (1 + i)^2 .
\]
Continuing this process, we find that the equation that we have to solve is
\[
(1 + i)^n = 2 ,
\]
the solution of which is
\[
\ln(1 + x) = \frac{\ln 2}{\ln(1 + i)} .
\]
To approximate the value of \( n \), recall that \( \ln 2 \approx 0.693 \) and use the series expansion of \( f(x) = \ln(1 + x) \) to obtain
\[
\ln \left( \frac{1 + i}{0.693} \right) = \frac{x^2}{2} - \frac{x^3}{3} + \cdots 
\]
Due to the alternating series in the denominator, the denominator is a little less than \( i \). Frequently, brokers increase the numerator a little (say, to 0.72) and simply divide by \( i \), obtaining the “well-known Rule of 72,” namely,
\[
\ln \left( \frac{1 + i}{0.693} \right) = \frac{0.72}{i} - \frac{x^2}{2} + \frac{x^3}{3} - \cdots 
\]
For example, if \( i = 0.08 \), then \( n \approx 72/8 = 9 \) provides an excellent approximation. (The answer is about 9.006.) Many people find that the Rule of 72 is extremely useful in dealing with money matters.

D.5  INTEGRATION

Say \( F(t) = f(t), a \leq t \leq b \). Then
\[
\int_a^b f(t) \, dt = F(b) - F(a) .
\]
Thus, if \( u(x) \) is such that \( u'(x) \) exists and \( a \leq u(x) \), then
\[
\int_a^{u(x)} f(t) \, dt = F[u(x)] - F(a) .
\]
Taking derivatives of this latter equation, we obtain
\[ D_s \left[ \int_a^b f(t) \, dt \right] F'(u(x)) u'(x) = f(u(x)) u'(x). \]

For example, with \( 0 < \nu \),
\[ D_s \left[ 2 \int_0^{\sqrt{\nu}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \, dt \right] = \left( \frac{2}{\sqrt{2\pi}} e^{-\nu/2} \right) \frac{1}{2\sqrt{\nu}} \sqrt{\nu}^{(1/2)-1} e^{-\nu/2}. \]

This formula is needed in proving that if \( Z \) is \( N(0, 1) \), then \( Z^2 \) is \( \chi^2(1) \).

The preceding example could be worked by first changing variables in the integral—that is, first using the fact that
\[ \int_a^b f(x) \, dx = \int_{u(a)}^{u(b)} f[u(y)] w'(y) \, dy, \]
where the monotonically increasing (decreasing) function \( x = w(y) \) has derivative \( w'(y) \) and inverse function \( y = u(x) \). In that example, \( a = 0, b = \sqrt{\nu}, z = \sqrt{\nu}, \)
\( z' = 1/2 \sqrt{\nu} = t \), so that
\[ 2 \int_0^{\sqrt{\nu}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \, dt = 2 \int_0^{\sqrt{\nu}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \left( \frac{1}{\sqrt{\nu}} \right) \, dt. \]

The derivative of the latter, by one form of the fundamental theorem of calculus, is
\[ 2 \frac{1}{\sqrt{2\pi}} e^{-\nu/2} \left( \frac{1}{2\sqrt{\nu}} \right) = \frac{\nu^{(1/2)-1} e^{-\nu/2}}{\sqrt{2\pi} 2^{1/2}}. \]

Integration by parts is frequently needed. It is based upon the derivative of the product of two functions of \( x \)—say \( u(x) \) and \( v(x) \). The derivative is
\[ D_s [u(x) v(x)] = u(x) v'(x) + v(x) u'(x). \]

Thus,
\[ [u(x) v(x)] = \int_a^b u(x) v'(x) \, dx + \int_a^b v(x) u'(x) \, dx \]
or, equivalently,
\[ \int_a^b u(x) v'(x) \, dx = [u(x) v(x)] - \int_a^b v(x) u'(x) \, dx. \]

For example, letting \( u(x) = x \) and \( v'(x) = e^{-x} \), we obtain
\[ \int_0^b xe^{-x} \, dx = [-xe^{-x}]_0^b - \int_0^b (-e^{-x}) \, dx \]
\[ = -be^{-b} - [e^{-x}]_0^b = -be^{-b} - e^{-b} + 1, \]
because \( u'(x) = 1 \) and \( v(x) = e^{-x} \).
We really only make some suggestions about functions of two variables, say, 
\[ z = f(x, y). \]

But these remarks can be extended to more than two variables. The two first partial derivatives with respect to \( x \) and \( y \), denoted, respectively, by \( \frac{\partial z}{\partial x} \) and \( \frac{\partial z}{\partial y} \), can be found...
in the usual manner of differentiating by treating the “other” variable as a constant. For instance,
\[
\frac{\partial (x^2y + \sin x)}{\partial x} = 2xy + \cos x
\]
and
\[
\frac{\partial (e^{xy})}{\partial y} = (e^{xy})(2xy).
\]

The second partial derivatives are simply first partial derivatives of the first partial derivatives. If \( z = e^{xy} \), then
\[
\frac{\partial^2 z}{\partial x \partial y} = \frac{\partial}{\partial x}(2xye^{xy}) = 2xye^{xy} + 2ye^{xy}.
\]

For notation, we use
\[
\frac{\partial^2 z}{\partial x^2} = \frac{\partial}{\partial x}\left( \frac{\partial z}{\partial x} \right), \quad \frac{\partial^2 z}{\partial y^2} = \frac{\partial}{\partial y}\left( \frac{\partial z}{\partial y} \right),
\]
\[
\frac{\partial^2 z}{\partial x \partial y} = \frac{\partial}{\partial x}\left( \frac{\partial z}{\partial y} \right), \quad \frac{\partial^2 z}{\partial y \partial x} = \frac{\partial}{\partial y}\left( \frac{\partial z}{\partial x} \right)
\]

In general,
\[
\frac{\partial^2 z}{\partial x \partial y} = \frac{\partial^2 z}{\partial y \partial x}
\]
provided that the partial derivatives involved are continuous functions.

As you might guess, at a relative maximum or minimum of \( z = f(x, y) \), we have
\[
\frac{\partial z}{\partial x} = 0 \quad \text{and} \quad \frac{\partial z}{\partial y} = 0,
\]
provided that the derivatives exist. To assure us that we have a maximum or minimum, we need
\[
\left( \frac{\partial^2 z}{\partial x \partial y} \right)^2 - \left( \frac{\partial^2 z}{\partial x^2} \right)\left( \frac{\partial^2 z}{\partial y^2} \right) < 0.
\]

Moreover, we have a relative minimum if \( \frac{\partial^2 z}{\partial x^2} > 0 \) and a relative maximum if \( \frac{\partial^2 z}{\partial x^2} < 0 \).

A major problem in statistics, called least squares, is to find \( a \) and \( b \) such that the point \((a, b)\) minimizes
\[
K(a, b) = \sum_{i=1}^{n} (y_i - a - bx_i)^2.
\]

Thus, the solution of the two equations
\[
\frac{\partial K}{\partial a} = \sum_{i=1}^{n} 2(y_i - a - bx_i)(-1) = 0
\]
and
\[
\frac{\partial K}{\partial b} = \sum_{i=1}^{n} 2(y_i - a - bx_i)(-x_i) = 0
\]
could give us a point \((a, b)\) that minimizes \(K(a, b)\). Taking second partial derivatives, we obtain

\[
\frac{\partial^2 K}{\partial a^2} = \sum_{i=1}^{n} 2(-1)(-1) = 2n > 0,
\]

\[
\frac{\partial^2 K}{\partial b^2} = \sum_{i=1}^{n} 2(-x_i)(-x_i) = 2 \sum_{i=1}^{n} x_i^2 > 0,
\]

and

\[
\frac{\partial^2 K}{\partial a \partial b} = \sum_{i=1}^{n} 2(-1)(-x_i) = 2 \sum_{i=1}^{n} x_i,
\]

and note that

\[
\left( \sum_{i=1}^{n} x_i \right)^2 - (2n) \left( \sum_{i=1}^{n} x_i^2 \right) < 0
\]

because \((\sum_{i=1}^{n} x_i)^2 < n \sum_{i=1}^{n} x_i^2\) provided that not all the \(x_i\) are equal. Noting that \(\frac{\partial^2 K}{\partial a^2} > 0\), we see that the solution of the two equations \(\frac{\partial K}{\partial a} = 0\) and \(\frac{\partial K}{\partial b} = 0\), provides the only minimizing solution.

The double integral

\[
\iint_{A} f(x, y) \, dx \, dy
\]

can usually be evaluated by an iteration—that is, by evaluating two successive single integrals. For example, say \(A = \{(x, y) : 0 \leq x \leq 1, 0 \leq y \leq x\}\), as given in Figure D.6-1.
Then
\[
\int_0^1 \left( \int_0^x (x + x^3 y^2) \, dy \right) \, dx = \int_0^1 \left[ \frac{x^2}{2} + \frac{x^4 y^3}{4} \right]_0^x \, dx \\
= \int_0^1 \left( \frac{x^2}{2} + \frac{x^6}{4} \right) \, dx = \left[ \frac{x^3}{3} + \frac{x^7}{7} \right]_0^1 \\
= \frac{1}{3} + \frac{1}{21} = \frac{8}{21}
\]

When placing the limits on the iterated integral, note that, for each fixed \(x\) between 0 and 1, \(y\) is restricted to the interval from 0 to \(x\). Also, in the inner integral on \(y\), \(x\) is treated as a constant.

In evaluating this double integral, we could have restricted \(y\) to the interval from 0 to 1. Then \(x\) would be between \(y\) and 1. That is, we would have evaluated the iterated integral
\[
\int_0^1 \left( \int_y^1 (x + x^3 y^2) \, dx \right) \, dy = \int_0^1 \left[ \frac{y}{2} + \frac{y^4 x^3}{4} \right]_y^1 \, dy \\
= \int_0^1 \frac{1}{2} + \frac{y^3}{4} - \frac{y^2}{2} - \frac{y^6}{4} \, dy \\
= \left[ \frac{y}{2} - \frac{y^3}{3} - \frac{y^7}{4} \right]_0^1 \\
= \frac{1}{2} - \frac{1}{12} - \frac{1}{28} = \frac{8}{21}.
\]

Finally, we can change variables in a double integral
\[
\iint_A f(x, y) \, dx \, dy.
\]

If \(f(x, y)\) is a joint p.d.f. of random variables \(X\) and \(Y\) of the continuous type, then the double integral represents \(P[\{X, Y\} \in A]\). Consider only one-to-one transformations—say, \(z = u_1(x, y)\) and \(w = u_2(x, y)\)—with inverse transformation given by \(x = v_1(z, w)\) and \(y = v_2(z, w)\). The determinant of order two,
\[
J = \begin{vmatrix}
\frac{\partial x}{\partial z} & \frac{\partial x}{\partial w} \\
\frac{\partial y}{\partial z} & \frac{\partial y}{\partial w}
\end{vmatrix},
\]
is called the Jacobian of the inverse transformation. Moreover, say the region \(A\) maps onto the region \(B\) in \((z, w)\) space. Since we are usually dealing with probabilities in
this book, we fixed the sign of the integral so that it is positive (by using the absolute value of the Jacobian). Then it follows that
\[ \int f(x, y) \, dx \, dy = \int f[v_1(z, w), v_2(z, w)] \, |J| \, dz \, dw. \]
To illustrate, let
\[ f(x, y) = \frac{1}{2\pi} e^{-\frac{(x^2+y^2)}{2}}, \quad -\infty < x < \infty, \quad -\infty < y < \infty, \]
which is the joint p.d.f. of two independent normal variables, each with mean 0 and variance 1. Say \( A = \{(x, y): 0 \leq x^2 + y^2 \leq 1\} \), and consider
\[ P(A) = \int f(x, y) \, dx \, dy. \]
This integration is impossible to deal with directly in the \( x, y \) variables. However, consider the inverse transformation to polar coordinates, namely,
\[ x = r \cos \theta, \quad y = r \sin \theta \]
with Jacobian
\[ J = \begin{vmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{vmatrix} = r(\cos^2 \theta + \sin^2 \theta) = r. \]
Since \( A \) maps onto \( B = \{(r, \theta): 0 \leq r \leq 1, 0 \leq \theta < 2\pi \} \), we have
\[
P(A) = \int_0^{2\pi} \left( \int_0^1 \frac{1}{2\pi} e^{-\frac{r^2}{2}} \, r \, dr \right) \, d\theta \\
= \int_0^{2\pi} \left[ -\frac{1}{2\pi} e^{-\frac{r^2}{2}} \right]_0^1 \, d\theta \\
= \int_0^{2\pi} \frac{1}{2\pi} \left( 1 - e^{-\frac{1}{2}} \right) \, d\theta \\
= \frac{1}{2\pi} \left( 1 - e^{-\frac{1}{2}} \right) 2\pi = 1 - e^{-\frac{1}{2}}.
\]