1

Mathematical Foundations 1

Point-Set Concepts, Set and Measure Functions, Normed Linear Spaces, and Integration

1.1 Set Notation and Operations

1.1.1 Sets and Set Inclusion

We may generally think of a **set** as a collection or grouping of items without regard to structure or order. (Sets will be represented by capital letters, e.g., A, B, C, ….) An **element** is an item within or a member of a set. (Elements are denoted by small case letters, e.g., a, b, c, ….) A set of sets will be termed a **class** (script capital letters will denote a class of sets, e.g., $\mathcal{A}, \mathcal{B}, \mathcal{C}, \ldots$); and a set of classes will be called a **family**.

Let us define a **space** (denoted Ω) as a type of *master* or *universal* set—it is the context in which discussions of sets occur. In this regard, an element of Ω is a point ω. To define a set X, let us write $X = \{x | \text{the } x\text{'s possess some defining property}\}$, that is, this reads "$X$ is the set of all elements x such that the x's have some unique characteristic," where "such that" is written "|."

The set containing no elements is called the **empty set** (denoted ϕ)—it is a member of every set. What about the size of a set? A set may be *finite* (it is either empty or consists of n elements, n a positive integer), *infinite* (e.g., the set of positive integers), or *countably infinite* (its elements can be put into one-to-one correspondence with the counting numbers).

We next look to inclusion symbols. Specifically, we first consider **element inclusion**. Element x being a member of set X is symbolized as $x \in X$. If x is not a member of, say, set Y, we write $x \notin Y$. Next comes **set inclusion** (a subset notation). A set A is termed a **subset** of set B (denoted $A \subseteq B$) if B contains the same elements that A does and possibly additional elements that are not found in A. If A is not a subset of B, we write $A \nsubseteq B$. Actually, two cases are subsumed in $A \subseteq B$: (1) either $A \subset B$ (A is then called a **proper subset** of B, meaning that B is a set that is larger than A; or (2) $A = B$ (A and B contain exactly the same

Stochastic Differential Equations: An Introduction with Applications in Population Dynamics Modeling, First Edition. Michael J. Panik.

elements and thus are **equal**). More formally, $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$. If equality between sets A and B does not hold, we write $A \neq B$.

1.1.2 Set Algebra

Given sets A and B within Ω, their **union** (denoted $A \cup B$) is the set of elements that are in A, or in B, or in both A and B. Here, we are employing the *inclusive or*. Symbolically, $A \cup B = \{x | x \in A \text{ or } x \in B\}$ (Figure 1.1a). The **intersection** of sets A and B (denoted $A \cap B$) is the set of elements common to both A and B, that is, $A \cap B = \{x | x \in A \text{ and } x \in B\}$ (Figure 1.1b). The **complement** of a set A is the set of elements within Ω that lie outside of A (denoted A'). Here, $A' = \{x | x \notin A\}$ (Figure 1.1c).

If sets A and B do not intersect and thus have no elements in common, then A and B are said to be **disjoint** or **mutually exclusive** and we write $A \cap B = \emptyset$. The **difference** between sets A and B (denoted $A - B$) is the set of elements in A but not in B or $A - B = A \cap B'$. Thus, $A - B = \{x | x \in A \text{ and } x \notin B\}$ (Figure 1.1d). The **symmetric difference** between sets A and B (denoted $A \Delta B$) is the

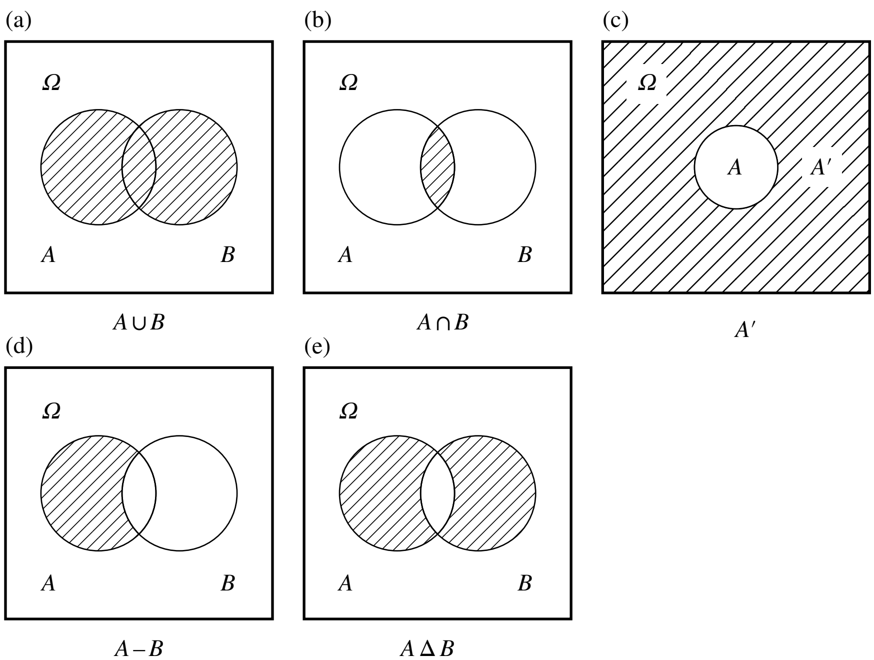

Figure 1.1 (a) Union of *A* and *B*, (b) intersection of *A* and *B*, (c) complement of *A*, (d) difference of *A* and *B*, and (e) symmetric difference of *A* and *B*.

union of their differences in reverse order or $A \Delta B = (A - B) \cup (B - A) = (A \cap B') \cup (B \cap A')$ (Figure 1.1e).

A few essential properties of these set operations now follow. Specifically for sets A, B, and C within Ω:

UNION

$A \cup A = A$, $A \cup \Omega = \Omega$, $A \cup \varnothing = A$
$A \cup B = B \cup A$ (commutative property)
$A \cup (B \cup C) = (A \cup B) \cup C$ (associative property)
$A \subseteq B$ if and only if $A \cup B = B$

INTERSECTION

$A \cap A = A$, $A \cap \Omega = A$, $A \cap \varnothing = \varnothing$
$A \cap B = B \cap A$ (commutative property)
$A \cap (B \cap C) = (A \cap B) \cap C$ (associative property)
$A \subseteq B$ if and only if $A \cap B = A$

COMPLEMENT

$(A')' = A$, $\Omega' = \varnothing$, $\varnothing' = \Omega$
$A \cup A' = \Omega$, $A \cap A' = \varnothing$

$$\left.\begin{aligned}(A \cup B)' &= A' \cap B' \\ (A \cap B)' &= A' \cup B'\end{aligned}\right\}\text{De Morgan's laws}$$

DIFFERENCE

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

SYMMETRIC DIFFERENCE

$A \Delta A = \varnothing$, $A \Delta \varnothing = A$
$A \Delta B = B \Delta A$ (commutative property)
$A \Delta (B \Delta C) = (A \Delta B) \Delta C$ (associative property)
$A \cap (B \Delta C) = (A \cap B) \Delta (A \cap C)$

DISTRIBUTIVE LAWS (connect the operations of union and intersection)

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

If $\{A_i, i = 1, \ldots, n\}$ is any arbitrary finite class of sets, then the extension of the union and intersection operations to this class can be written, respectively, as

$$\cup_{i=1}^{n} A_i \text{ and } \cap_{i=1}^{n} A_i.$$

Hence, the union of a class of sets is the collection of elements belonging to at least one of them; the intersection of a class of sets is the set of elements common to all of them. In fact, given these notions, De Morgan's laws may be extended to

$$\left(\cup_{i=1}^{n} A_i\right)' = \cap_{i=1}^{n} A_i' \text{ and } \left(\cap_{i=1}^{n} A_i\right)' = \cup_{i=1}^{n} A_i'.$$

Furthermore, if $\{A_i, i = 1, \ldots, n\}$ and $\{B_j, j = 1, \ldots, m\}$ are two finite classes of sets with $\{A_i\} \subseteq \{B_j\}$, then

$$\cup_{i=1}^{n} A_i \subseteq \cup_{j=1}^{m} B_j \text{ and } \cap_{j=1}^{m} B_j \subseteq \cap_{i=1}^{n} A_i.$$

In addition, if $\{A_i, i = 1, 2, \ldots\}$ represents a **sequence of sets**, then their union and intersection appears as

$$\cup_{i=1}^{\infty} A_i \text{ and } \cap_{i=1}^{\infty} A_i,$$

respectively.

1.2 Single-Valued Functions

Given two nonempty sets X and Y (which may or may not be equal), a **single-valued function** or **point-to-point mapping** $f: X \rightarrow Y$ is a rule or law of correspondence that associates with point $x \in X$ a unique point $y \in Y$. Here, $y = f(x)$ is the **image of x** under rule f. While set X is called the **domain of f** (denoted D_f), the collection of those y's that are the image of at least one $x \in X$ is called the **range of f** and denoted R_f. Clearly the range of f is a subset of Y (Figure 1.2a). If $R_f \subset Y$, then f is an **into mapping**. In addition, if $R_f = Y$ (i.e., *every* $y \in Y$ is the image of at least one $x \in X$ or all the y's are accounted for in the mapping process), then f is termed an **onto** or **surjective mapping**. Moreover, f is said to be **one-to-one** or **injective** if no $y \in Y$ is the image of more than one $x \in X$ (i.e., $x_1 \neq x_2$ implies $f(x_1) \neq f(x_2)$). Finally, f is called **bijective** if it is both one-to-one and onto or both surjective and injective. If the range of f consists of but a single element, then f is termed a **constant function**.

Given a nonempty set X, if Y consists entirely of real numbers or $Y = R$, then $f: X \rightarrow Y$ is termed a **real-valued function** or **mapping** of a point $x \in X$ into a unique real number $y \in R$.[1] Hence, the image of each point $x \in X$ is a real scalar $y = f(x) \in R$.

1 A discussion of real numbers is offered in Section 1.3.

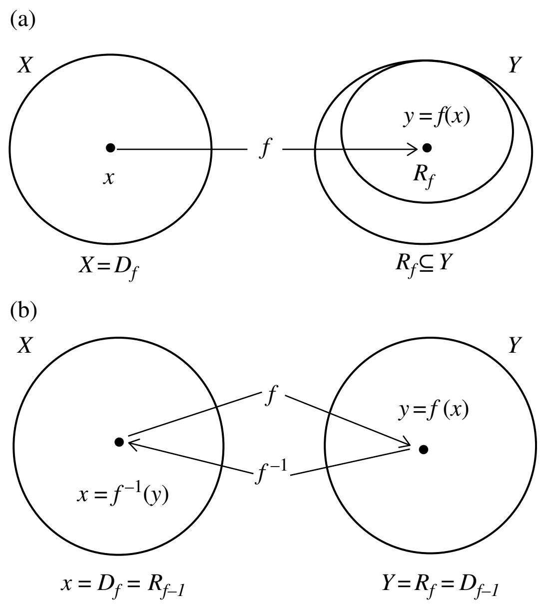

Figure 1.2 (a) f is an into mapping and (b) f is one-to-one and onto.

For sets X and Y with set $A \subset X$, let f_1: $A \rightarrow Y$ be a point-to-point mapping of A into Y and f_2: $X \rightarrow Y$ be a point-to-point mapping of X into Y. Then f_1 is said to be a **restriction** of f_2 and f_2 is termed an **extension** of f_1 if and only if for each $x \in A$, $f_1(x) = f_2(x)$.

Let $X_1, X_2, \ldots, X_n$ represent a class of nonempty sets. The **product set of** X_1, $X_2, \ldots, X_n$ (denoted $X_1 \times X_2 \times \cdots \times X_n$) is the set of all ordered n-tuples $(x_1, x_2, \ldots, x_n)$, where $x_i \in X_i$ for each $i = 1, \ldots, n$. Familiar particularizations of this definition are $R^1 = R$ (the real line); $R^2 = R \times R$ is the two-dimensional coordinate plane (made up of all ordered pairs (x_1, x_2), where both $x_1 \in R$ and $x_2 \in R$); and $R^n = R \times R \times \cdots \times R$ (the product is taken n times) depicts the collection of ordered n-tuples of real numbers. In this regard, for f a point-to-point mapping of X into Y, the subset $G_f = \{(x, y) | x \in X, y = f(x) \in Y\}$ of $X \times Y$ is called the **graph of** $\boldsymbol{f}$.

If the point-to-point mapping f is bijective (f is one-to-one and onto), then its single-valued **inverse mapping** $f^{-1}: Y \rightarrow X$ exists. Thus to each point $y \in Y$, there corresponds a unique **inverse image** point $x \in X$ such that $x = f^{-1}(y) = f^{-1}(f(x))$ so that x is termed the **inverse function** of y. Here, the domain $D_{f^{-1}}$ of f^{-1} is Y, and its range $R_{f^{-1}}$ is X. Clearly, f^{-1} must also be bijective (Figure 1.2b).

1.3 Real and Extended Real Numbers

We noted in Section 1.2 that a function f is *real valued* if its range is the set of real numbers. Let us now explore some of the salient features of real numbers—properties that will be utilized later on.

The **real number system** may be characterized as a complete, ordered field, where a **field** is a set F of elements together with the operations of addition and multiplication. Moreover, both addition and multiplication are associative and commutative, additive and multiplicative inverse and identity elements exist, and multiplication distributes over addition. Set F is **ordered** if there is a binary order relation "<" in F that satisfies the following conditions:

1. For any elements x, y in F, either $x < y$, $y < x$, or $x = y$.
2. For any elements x, y, and z in F, if $x < y$ and $y < z$, then $x < z$.

Now, if F is an **ordered field**, then the order relation must be connected to the field operations according to the following conditions:

1. If $x < y$, then $x + z < y + z$.
2. If x, y, and z is positive, then $zx < zy$.

Looking to the completeness property of the real number system, let us note first that a set A ($\neq \emptyset$) of real numbers is **bounded above** if there is a real number b (the **upper bound** for A) such that $a \leq b$ for every $a \in A$. The **least upper bound** or **supremum** of A (denoted sup A) is a real number b such that (1) $a \leq b$ for every $a \in A$; and (2) if $a \leq c$ for every $a \in A$, then $b \leq c$. So if b is an upper bound for A such that no smaller element of A is also an upper bound for A, then b is the least upper bound for A. In a similar vein, we can state that a set A ($\neq \emptyset$) of real numbers is **bounded below** if there is a real number b (the **lower bound** for A) such that $b \leq a$ for every $a \in A$. The **greatest lower bound** or **infimum** of A (written *inf* A) is a real number b such that (1) $b \leq a$ for every $a \in A$; and (2) if $c \leq a$ for every $a \in A$, then $c \leq b$. Hence, if b is a lower bound for A such that no larger element of A is also a lower bound for A, then b is the greatest lower bound for A. Clearly the supremum and infimum for A must be unique.

Armed with these considerations, we can state the **completeness property** as every nonempty subset A of the ordered field F of real numbers which has an upper bound in F has a least upper bound in F.

If we admit the elements $\{-\infty\}$ and $\{+\infty\}$ to our discussion of real numbers R, then the **extended real number system** (denoted R^*) consists of the set of real numbers R together with $\pm\infty$, that is, $R^* = R \cup \{-\infty\} \cup \{+\infty\}$.

1.4 Metric Spaces

Given a space Ω, a **metric** defined on Ω is an everywhere finite real-valued function μ of ordered pairs (x, y) of points of Ω or $\mu(x,y) : \Omega \times \Omega \to [0, +\infty)$ satisfying the following conditions:

1. For $x \in \Omega$, $\mu(x, x) = 0$ (reflexitivity).
2. For $x, y \in \Omega$, $\mu(x, y) \geq 0$ and $\mu(x, y) = 0$ if and only if $x = y$.
3. For $x, y \in \Omega$, $\mu(x, y) = \mu(y, x)$ (symmetry).
4. For $x, y, z \in \Omega$, $\mu(x, y) \leq \mu(x, z) + (z, y)$ (triangle inequality).

Here, μ serves to define the **distance** between x and y. A **metric space** consists of the space Ω and a metric μ defined on Ω. Hence, a metric space will be denoted (Ω, μ). For instance, if $\Omega = R$, then R is a metric space if $\mu(x,y) = |x - y|$ (the distance between points x and y on the real line). In addition, if $\Omega = R^n$, then R^n can be considered a metric space if

$$\mu(x,y) = \left[\sum_{i=1}^{n} |x_i - y_i|^2\right]^{\frac{1}{2}}, \tag{1.1}$$

where again $\mu(x, y)$ is interpreted as the distance between $x, y \in R^n$.[2]

Suppose Ω is a metric space with metric μ and X ($\neq \emptyset$) is an arbitrary subset of Ω. *If* μ is defined only for points in X, then (X, μ) is also a metric space. Then under this restriction on μ, X is termed a **subspace of** Ω.

The importance of a metric space is that it incorporates a concept of distance (μ) that is applicable to the points within Ω. In addition, this distance function will enable us to tackle issues concerning the convergence of sequences in Ω and continuous functions defined on Ω.

---

2 Equation (1.1) is actually a generalization of the absolute value function $|x - y|$. To see this, let us define on R^n a **norm** (denoted $\|\cdot\|$)—a function $\|\cdot\| : R^n \to [0, +\infty)$ which assigns to each $x \in R^n$ some number $\|x\|$ such that

a. $\|x\| \geq 0$ and $\|x\| = 0$ if and only if $x = 0$;
b. $\|x + y\| \leq \|x\| + \|y\|$ (triangle inequality);
c. for a scalar c, $\|cx\| = |c|\|x\|$ (homogeneity); and
d. $\left|\sum_{i=1}^{n} x_i y_i\right| \leq \|x\|\|y\|$ (Cauchy–Schwarz inequality).

Then the distance between points $x, y \in R^n$ induced by the norm "$\|\cdot\|$" on R^n is

$$\|x - y\| = \left[\sum_{i=1}^{n} |x_i - y_i|^2\right]^{\frac{1}{2}} \tag{1.2}$$

or Equation (1.1). So if $\Omega = R^n$ and μ is given by (1.1), then R^n is a metric space with metric (1.1).

1.5 Limits of Sequences

Let X be a subset of R^n. A **sequence of points** in X is a function whose domain is the set of all positive integers I and whose range appears in X. If the value of the function at $n \in I$ is $x_n \in X$, then the range of the sequence will be denoted by $\{x_n\} = \{x_1, x_2, \ldots\}$ and interpreted as "the sequence of points $x_1, x_2, \ldots$ in X." (Note that the sequence of points $\{x_n\}$ mapped into X *is not* a subset of X.) By deleting certain elements of the sequence $\{x_n\}$, we obtain the **subsequence** $\{x_n\}_{n \in J}$, where J is a subset of the positive integers.

A sequence $\{x_n\}$ in R^n **converges to a limit** $\bar{x}$ if and only if $lim_{n \to \infty} \mu(x_n, \bar{x}) = lim_{n \to \infty} \|x_n - \bar{x}\| = 0$. (This is alternatively expressed as $lim_{n \to \infty} x_n = \bar{x}\, or\, x_n \to \bar{x}$ as $n \to \infty$.) That is, $\bar{x}$ is the limit of $\{x_n\}$ if for each $\varepsilon > 0$ there exists an index value $\bar{n}_\varepsilon$ such that $n > \bar{n}_\varepsilon$ implies $\|x_n - \bar{x}\| < \varepsilon$. If we think of the condition $\|x_n - \bar{x}\| < \varepsilon$ as defining an open sphere of radius ε about $\bar{x}$, then we can say that $\{x_n\}$ converges to $\bar{x}$ if for each open sphere of radius $\varepsilon > 0$ centered on $\bar{x}$, there exists an $\bar{n}_\varepsilon$ such that x_n is within this open sphere for all $n > \bar{n}_\varepsilon$. Hence, the said sphere contains all points of $\{x_n\}$ from $x_{\bar{n}_\varepsilon}$ on, that is, $\bar{x}$ is the limit of the sequence $\{x_n\}$ in R^n if, given $\varepsilon > 0$, all but a finite number of terms of the sequence are within ε of $\bar{x}$.

A point $\hat{x} \in R^n$ is a **limit (cluster) point** of an infinite sequence $\{x_k\}$ if and only if there exists an infinite subsequence $\{x_k\}_{k \in K}$ of $\{x_k\}$ that converges to $\hat{x}$, that is, there exists an infinite subsequence $\{x_k\}$ such that $lim_{j \to \infty} \|x_{k_j} - \hat{x}\| = 0$ or $x_{k_j} \to \hat{x}$ as $j \to \infty$. Stated alternatively, $\hat{x}$ is a limit point of $\{x_k\}$ if, given a $\delta > 0$ and an index value $\bar{k}$, there exists some $k > \bar{k}$ such that $\|x_k - \hat{x}\| < \delta$ for infinitely many terms of $\{x_k\}$.

What is the distinction between the limit of a sequence and a limit point of a sequence? To answer this question, we state the following:

a. $\bar{x}$ is a limit of a sequence $\{x_k\}$ in R^n if, given a small and positive $\varepsilon \in R$, all but a finite number of terms of the sequence are within ε of $\bar{x}$.
b. $\hat{x}$ is a limit point of $\{x_k\}$ in R^n if, given a real scalar $\varepsilon > 0$ and given $\bar{k}$, infinitely many terms of the sequence are within ε of $\hat{x}$.

Thus, a sequence $\{x_k\}$ in R^n may have a limit but no limit point. However, if a convergent sequence $\{x_k\}$ in R^n has infinitely many distinct points, then its limit is a limit point of $\{x_k\}$. Likewise, $\{x_k\}$ may possess a limit point but no limit. In fact, if the sequence $\{x_k\}$ in R^n has a limit point $\hat{x}$, then there is a subsequence $\{x_k\}_{k \in K}$ of $\{x_k\}$ that has $\hat{x}$ as a limit; but this does not necessarily mean that the entire sequence $\{x_k\}$ converges to $\hat{x}$.[3]

3 If $x_k = n =$ constant for all k, then $\{x_k\}$ converges to the limit n. But since the range of this sequence contains only a single point, it is evident that the sequence has no limit point. If $x_k = 1/k$, then the sequence $\{x_k\}$ converges to a limit of zero, which is also a limit point. In addition, if $x_k = (-1)^k$, then the sequence $\{x_k\}$ has limit points at ±1, but has no limit.

A sufficient condition that at least one limit point of an infinite sequence $\{x_k\}$ in R^n exists is that $\{x_k\}$ is **bounded**, that is, there exists a scalar $M \in R$ such that $\|x_k\| \leq M$ for all k. In this regard, if an infinite sequence of points $\{x_k\}$ in R^n is bounded and it has only one limit point, then the sequence converges and has as its limit that single limit point.

The preceding definition of the limit of the sequence $\{x_n\}$ explicitly incorporated the actual limit $\bar{x}$. If one does not know the actual value of $\bar{x}$, then the following theorem enables us to prove that a sequence converges even if its actual limit is unknown. To this end, we state first that a sequence is a **Cauchy sequence** if for each $\varepsilon > 0$ there exists an index value $N_{\varepsilon/2}$ such that $m, n > N_{\varepsilon/2}$ implies $d(x_m, x_n) = \|x_m - x_n\| < \varepsilon$.[4] Second, R^n is said to be **complete** in that to every Cauchy sequence $\{x_n\}$ defined on R^n there corresponds a point $\bar{x}$ such that $lim_{n\to\infty} x_n = \bar{x}$. Given these concepts, we may now state the

> **Cauchy Convergence Criterion**: Given that R^n is complete, a sequence $\{x_n\}$ in R^n converges to a limit $\bar{x}$ if and only if it is a Cauchy sequence, that is, a necessary and sufficient condition for $\{x_n\}$ to be convergent in R^n is that $d(x_m, x_n) \to 0$ as $m, n \to \infty$.

Hence, every convergent sequence on R^n is a Cauchy sequence. The implication of this statement is that if the terms of a sequence approach a limit, then, beyond some point, the distance between pairs of terms diminishes.

It should be evident from the preceding discussion that a **complete metric space** is a metric space in which every Cauchy sequence converges, that is, the space contains a point $\bar{x}$ to which the sequence converges or $lim_{n\to\infty} x_n = \bar{x}$. In this regard, it should also be evident that the real line R is a complete metric space as is R^n.

We next define the **limit superior** and **limit inferior** of a sequence $\{x_n\}$ of real numbers as, respectively,

$$\begin{aligned} &\text{a. } \lim_{n\to\infty} sup\, x_n = \lim_{n\to\infty}\left(\sup_{m\geq n} x_m\right) \text{ and} \\ &\text{b. } \lim_{n\to\infty} inf\, x_n = \lim_{n\to\infty}\left(\inf_{m\geq n} x_m\right). \end{aligned} \tag{1.3}$$

Hence, the limit superior of the sequence $\{x_n\}$ is the largest number $\bar{x}$ such that there is a subsequence of $\{x_n\}$ that converges to $\bar{x}$—and no subsequence converges to a higher value. Similarly, the limit inferior is the smallest limit attainable for some convergent subsequence of $\{x_n\}$—and no subsequence converges to a lower value. Looked at in another fashion, for, say, Equation (1.3a), a

4 That is, for $\varepsilon > 0$ there exists a positive integer $N_{\varepsilon/2}$ such that $m \geq N_{\varepsilon/2}$ implies $d(x_m, \bar{x}) < \varepsilon/2$; and $n \geq N_{\varepsilon/2}$ implies $d(x_n, \bar{x}) < \varepsilon/2$. Hence, both $m, n > N_{\varepsilon/2}$ imply, via the triangle inequality, that

$$d(x_m, x_n) \leq d(x_m, \bar{x}) + d(x_n, \bar{x}) < \tfrac{\varepsilon}{2} + \tfrac{\varepsilon}{2} = \varepsilon.$$

number $\bar{x}$ is the limit superior of a sequence $\{x_n\}$ if (1) for every $x < \bar{x}$, we have $x < x_n$ for infinitely many n's; and (2) for every $x > \bar{x}$, we have $x < x_n$ for only finitely many n's. Generally speaking, when there are multiple points around which the terms of a sequence tend to "pile up," the limit superior and limit inferior select the largest and smallest of these points, respectively.

We noted earlier that a sequence defined on a subset X of R^n is a function whose range is $\{x_n\}$. If this function is **bounded**, then its range $\{x_n\}$ is bounded from both above and below. In fact, if $\{x_n\}$ is a bounded sequence of real numbers, then the limit superior and limit inferior both exist. It is also important to note that $lim_{n\to\infty} x_n$ exists if and only if the limit superior and limit inferior are equal. We end this discussion of limits by mentioning that since any set of extended real numbers has both a supremum and an infimum, it follows that every sequence of extended real numbers has both a limit superior and a limit inferior.

1.6 Point-Set Theory

Let δ be any positive scalar. A **δ-neighborhood of a point $x_0 \in R^n$** or **sphere of radius δ about x_0** is the set $\delta(x_0) = \{x | \|x - x_0\| < \delta, \delta > 0\}$. A point $\bar{x}$ is an **interior point** of a set X in R^n if there exists a δ-neighborhood about $\bar{x}$ that contains only points of X.

A set X in R^n is said to be **open** if, given any point $x_0 \in X$, there exists a positive scalar δ such that $\delta(x_0) \subseteq X$. Hence, X is open if it contains only interior points. Moreover,

a. $\emptyset$, $\delta(x_0)$, and R^n are all open sets.
b. Any union of open sets in R^n is open; and any finite intersection of open sets in R^n is open.

Let X be a set in R^n. The **complementary set of X**, denoted X', is the collection of all points of R^n lying outside of X. A point $\bar{x} \in X'$ is an exterior point of X in R^n if there exists a δ-neighborhood of $\bar{x}$ that contains only points of X'. A point $\bar{x}$ is a **boundary point** of a set X in R^n if every δ-neighborhood of $\bar{x}$ encompasses points in X and in X'.

A set X in R^n is **bounded** if there exists a scalar $M \in R$ such that $\|x\| \le M$ for all $x \in X$. Stated alternatively, X is bounded if it has a finite **diameter** $d(X) = sup\{\|x - y\| | x, y \in X\}$.

A set X in R^n has an **open cover** if there exist a collection $\{G_i\}$ of open subsets from R^n such that $X \subseteq \cup_i G$. The open cover $\{G_i\}$ of X in R^n is said to contain a **finite subcover** if there are finitely many indices $i_1, \ldots, i_m$ for which $X \subseteq \cup_{j=1}^{m} G_{i_j}$.

A point $\bar{x}$ is termed a **point of closure** of a set X in R^n if every δ-neighborhood of $\bar{x}$ contains at least one point of X, that is, $\delta(\bar{x}) \cap X \neq \phi$. It is important to note that a point of closure of X need not be a member of X; however, every element within X is also a point of closure of X. A subset X of R^n is **closed** if every point of closure of X is contained in X. The **closure** of a set X in R^n, denoted $\bar{X}$, is the set of points of closure of X. Clearly, a set X in R^n is closed if and only if $X = \bar{X}$. A set X in R^n has a **closed cover** if there exists a collection $\{G_i\}$ of closed subsets from R^n such that $X \subseteq \cup_i G$.

Closely related to the concept of a point of closure of X is the notion of a **limit (cluster) point** of a set X in R^n. Specifically, $\bar{x}$ is a limit point of X if each δ-neighborhood about $\bar{x}$ contains at least one point of X different from $\bar{x}$, that is, points of X different from $\bar{x}$ tend to "pile up" at $\bar{x}$. So if $\bar{x}$ is a limit point of a set X in R^n, then $X \cap \delta(\bar{x})$ is an infinite set—every δ-neighborhood of $\bar{x}$ contains infinitely many points of X. Moreover,

a. If X is a finite set in R^n, then it has no limit point.
b. The limit point of X need not be an element of X.
c. The collection of all limit points of X in R^n is called the **derived set** and will be denoted X^d.

Based on the preceding discussion, we can alternatively characterize a set X in R^n as **closed** if it contains each of its limit points or if $X^d \subseteq X$. In addition, we can equivalently state that the **closure** of a set X in R^n is X together with its collection of limit points or $\bar{X} = X \cup X^d$. Furthermore,

a. Ø, a single point, and R^n are all closed sets.
b. Any finite union of closed sets in R^n is closed; any intersection of closed sets in R^n is closed.
c. The closure of any set X in R^n is the smallest closed set containing X.
d. A subset X in R^n is closed if and only if its complementary set X' is open.
e. A subset X in R^n is closed if and only if X contains its boundary.

Let's now briefly relate the concepts of a limit and a limit point of a sequence in R^n to some of the preceding point-set notions that we just developed. In particular, we shall take another look at the *point of closure* concept. To this end, a limit point (as well as a limit) of a sequence $\{x_k\}$ in R^n is a **point of closure** of a set X in R^n if X contains $\{x_k\}$. Conversely, if $\hat{x}$ is a point of closure of a set X in R^n, then there exists a sequence $\{x_k\}$ in X (and hence also a subsequence $\{x_k\}_{k \in K}$ in X) such that $\hat{x}$ is a limit point of $\{x_k\}$ (and thus a limit of $\{x_k\}_{k \in K}$). Hence, the **closure** of $X, \bar{X}$, consists of all limit points of convergent sequences $\{x_k\}$ from X.

Similarly, we note that a subset X in R^n is **closed** if and only if every convergent sequence of points $\{x_k\}$ from X has a limit in X, that is, X is closed if for $\{x_k\}$ in X, $lim_{k \to \infty} x_k = \hat{x} \in X$. Also, a set X in R^n is **bounded** if every sequence of points $\{x_k\}$ formed from X is bounded. In addition, if a set X in R^n is both closed and bounded, then it is termed **compact**. (Equivalently, a set X in R^n is compact if it has the **finite intersection property**: every finite subclass has a nonempty intersection.) We mention briefly the following:

a. A closed subset of a compact set X in R^n is compact.
b. The union of a finite number of compact sets in R^n is compact; the intersection of any number of compact sets in R^n is compact.
c. A set X in R^n is compact if and only if it is closed and bounded.
d. Any finite set of points in R^n is compact.
e. If X in R^n is a set consisting of a convergent sequence $\{x_k\}$ and its limit $\bar{x} = lim_{k\to\infty} x_k$, then X is compact. Conversely, if X in R^n is compact, every sequence $\{x_k\}$ has a convergent subsequence $\{x_k\}_{k\in K}$ whose limit belongs to X.

A set X in R^n is **locally compact** if each of its points has a δ-neighborhood with compact closure, that is, for each $x \in X$, $\overline{\delta(x)}$ is compact. In this regard, any compact space is locally compact but not conversely, for example, R^n is locally compact but not compact.

1.7 Continuous Functions

For metric spaces X and Y with metrics d_1 and d_2, respectively, let $f: X \to Y$ be a point-to-point mapping of X into Y. f is said to be **continuous at a point** $x_0 \in X$ if either

a. for any $\varepsilon > 0$ there exists a $\delta_\varepsilon > 0$ such that $d_1(x,x_0) < \delta_\varepsilon$ implies $d_2(f(x),f(x_0)) < \varepsilon$. (Note that the subscript on δ means that "δ depends upon the ε chosen."); or
b. for each ε-neighborhood of $f(x_0)$, $\varepsilon(f(x_0))$, there exists a δ_ε-neighborhood about x_0, $\delta_\varepsilon(x_0)$, such that $f(\delta_\varepsilon(x_0)) \subseteq \varepsilon(f(x_0))$, that is, points "near" x_0 are mapped by f into points "near" $f(x_0)$.

In general, the point-to-point mapping f: $X \to Y$ is **continuous on** $\boldsymbol{X}$ if it is continuous at each point of X.

Theorems 1.7.1 and 1.7.2 provide us with a set of necessary and sufficient conditions for the continuity of a point-to-point mapping at a specific point $x_0 \in X$ and at any arbitrary $x \in X$, respectively. Specifically, we start with Theorem 1.7.1.

Theorem 1.7.1 (continuity in terms of convergent sequences). For metric spaces X and Y, the point-to-point mapping f of X into Y is continuous at $x_0 \in X$ if and only if $x_k \to x_0$ implies $f(x_k) \to f(x_0)$ for every subsequence $\{x_k\}$ in X.

Hence, f is a continuous mapping of X into Y if it "sends convergent sequences in X into convergent sequences in Y." Next comes Theorem 1.7.2.

Theorem 1.7.2 (continuity in terms of open (resp. closed) sets). For metric spaces X and Y, let f be a point-to-point mapping of X into Y. Then, (a) f is continuous if and only if $f^{-1}(A)$ is open in X whenever set A is open in Y; and (b) f is continuous if and only if $f^{-1}(A)$ is closed in X whenever A is closed in Y.

Thus, f is continuous if it "pulls open (resp. closed) sets back to open (resp. closed) sets," that is, the inverse images of open (resp. closed) sets are open (resp. closed).

We next consider Theorem 1.7.3 which states that continuous mappings preserve compactness. That is,

Theorem 1.7.3 For metric spaces X and Y, let f be a continuous point-to-point mapping from X into Y. If A is a compact subset of X, then so is its range $f(A)$.

Next, let X be a subset of R^n. A continuous point-to-point mapping $g: R^n \rightarrow X$ is termed a **retraction mapping** on R^n if $g(x) = x$ for all $x \in X$. Here, X is called a **retraction** of R^n. If X is contained within an arbitrary subset A of R^n, then $g: A \rightarrow X$ is a retraction of A onto X if $g(x) = x$ for all $x \in X$.

1.8 Operations on Sequences of Sets

Let $\{A_i\}, i = 1, 2, \ldots$, represent a sequence of sets in a metric space X. If $\{A_i\}$ is such that $A_i \subseteq A_{i+1}, i = 1, 2, \ldots$, then $\{A_i\}$ is said to be a **nondecreasing sequence** (if $A_i \subset A_{i+1}$, then $\{A_i\}$ is said to be an **expanding sequence**). In addition, if $\{A_i\}$ is such that $A_i \supseteq A_{i+1}, i = 1, 2, \ldots$, then $\{A_i\}$ is called a **nonincreasing sequence** (if $A_i \supset A_{i+1}, i = 1, 2, \ldots$, then $\{A_i\}$ is termed a **contracting sequence**). A **monotone sequence** of sets is one which is either an expanding or contracting sequence.

If the sequence $\{A_i\}, i = 1, 2, \ldots$, in X is nondecreasing or nonincreasing, then its limit exists and we have the following:

$$\lim_{i \rightarrow \infty} A_i = \cup_{i=1}^{\infty} A_i \quad (\text{if } \{A_i\} \text{ is nondecreasing});$$

$$\lim_{n \rightarrow \infty} A_i = \cap_{i=1}^{\infty} A_i \quad (\text{if } \{A_i\} \text{ is nonincreasing}).$$

In addition, for any sequence of sets $\{A_i\}, i = 1, 2, \ldots$, in X,

$$sup\{A_i\} = \cup_{i=1}^{\infty} A_i,\ inf\{A_i\} = \cap_{i=1}^{\infty} A_i;$$

with

$$sup\left(\cup_{i=1}^{\infty} A_i\right) = sup\{sup A_i, i = 1, 2, \ldots\},$$

$$inf\left(\cup_{i=1}^{\infty} A_i\right) = inf\{inf A_i, i = 1, 2, \ldots\}; \text{and}$$

$$sup\left(\cap_{i=1}^{\infty}A_i\right) \leq inf\{supA_i, i=1,2,\ldots\},$$
$$inf\left(\cap_{i=1}^{\infty}A_i\right) \geq sup\{inf A_i, i=1,2,\ldots\}.$$

Let $\{A_i\}, i=1,2,\ldots$, again depict a sequence of sets in X.

Then there are subsets $E_i \subset A_i$ of disjoint sets, with $E_j \cap E_k = \phi$ for $j \neq k$, such that

$$\cup_{i=1}^{\infty}E_i = \cup_{i=1}^{\infty}A_i.$$

We next consider the concepts of the limit superior and limit inferior of a sequence of sets $\{A_i\}, i=1,2,\ldots$, in a metric space X. To this end, the **limit superior of a sequence** $\{A_i\}$ is defined as

$$\begin{aligned}\lim_{i\to\infty} sup A_i &= \cap_{i=1}^{\infty}\left(\cup_{k\geq i}A_k\right)\\ &= (A_1 \cup A_2 \cup \cdots)\cap(A_2 \cup A_3 \cup \cdots)\cap\cdots\\ &= \{x \in X | x \in A_i \text{ for infinitely many } i\}.\end{aligned}$$

Hence, *lim sup* A_i is the set S of points such that, for every positive integer i, there exists a positive integer $k \geq i$ such that $S \in A_i$; thus S consists of those points that belong to A_i for an infinite number of i values. Looked at in another fashion, if $x \in S$, then x is in all of $\cup_{k\geq i}A_k$. Hence, no matter how large of an i value is chosen, you can find a $k \geq i$ for which x is a member of A_i.

Similarly, the **limit inferior of a sequence** $\{A_i\}$ is

$$\begin{aligned}\lim_{i\to\infty} inf A_i &= \cup_{i=1}^{\infty}\left(\cap_{k\geq i}A_k\right)\\ &= (A_1 \cap A_2 \cap \cdots)\cup(A_2 \cap A_3 \cap \cdots)\cup\cdots\\ &= \{x \in X | x \in A_i \text{ for all but finitely many } i\}.\end{aligned}$$

Thus, *lim inf* A_i is the set I of points such that, for some positive integer i, $I \in A_i$ for all positive integers $k \geq i$; hence, I consists of those points that belong to A_i for all except a finite number of i values. Stated alternatively, if $x \in I$, then x is an element of $\cap_{k\geq i}A_k$ so that $x \in A_i$ for $k \geq i$—x must be in I with only finitely many exceptions, that is, for $x \in I$, there is an index value such that x is in every A_i in the remaining portion of the limit.[5]

5 Alternative definitions of the limit superior and limit inferior of a sequence of sets are the following. Again, let $\{A_i\}$ be a sequence of sets in a metric space X. Then

$$\lim_{i\to\infty} sup A_i = \left\{x \in X | \lim_{i\to\infty} inf\, d(x, A_i) = 0\right\};$$
$$\lim_{i\to\infty} inf A_i = \{x \in X | d(x, A_i) = 0\},$$

where $d(x, A_i)$ is the distance from x to A_i.

We note briefly that if $\{A_i\}, i = 1, 2, \ldots,$ is any sequence of sets in a metric space X, then *lim inf* $A_i \subset$ *lim sup* A_i. A sequence of sets $\{A_i\}$ is **convergent** (or a subset A of X is the **limit** of $\{A_i\}$) if

$$\lim_{i\to\infty} sup A_i = \lim_{i\to\infty} inf A_i = \lim_{i\to\infty} A_i = A.$$

Here, A is termed the **limit set**. In this vein, any monotone sequence of sets $\{A_i\}, i = 1, 2, \ldots,$ is convergent.

1.9 Classes of Subsets of Ω

1.9.1 Topological Space

We previously defined a metric space (Ω, μ) as consisting of the space Ω and a metric μ defined on Ω. Let $\mathcal{A}$ denote the class of open sets in the metric space. Then $\mathcal{A}$ satisfies the following conditions:

i. $\emptyset, \Omega \in \mathcal{A}$.
ii. If $A_1, A_2, \ldots, A_n \in \mathcal{A}$, then $\cap_{i=1}^{n} A_i \in \mathcal{A}$ (the intersection of every *finite* class of sets in $\mathcal{A}$ is itself a set in $\mathcal{A}$).
iii. If $A_\alpha \in \mathcal{A}$ for $\alpha \in I$, then $\cup_{\alpha\in I} A_\alpha \in \mathcal{A}$ (the union of every arbitrary class of sets in $\mathcal{A}$ is itself a set in $\mathcal{A}$).

Armed with properties (i)–(iii), let us generalize a metric space to that of a topological space. That is, given a nonempty space Ω and a given class $\mathcal{A}$ of subsets of Ω consisting of the "open sets" in Ω, a class $\mathcal{J}$ of subsets of Ω is called a **topology** on Ω if (i)–(iii) hold. (Thus, the class of open sets $\mathcal{A}$ determines the topology in Ω.) Hence, a **topological space** consists of Ω and a topology $\mathcal{J}$ on Ω and is denoted $(\Omega, \cdot)$.

A subset A of a topological space $(\Omega, \mathcal{J})$ is said to be (everywhere) **dense** if its closure $\bar{A}$ equals $(\Omega, \cdot)$. Hence, A is dense if and only if (a) A intersects every nonempty set; or (b) the only open set disjoint from A is $\emptyset$.

1.9.2 σ-Algebra of Sets and the Borel σ-Algebra

A **ring** $\mathcal{R}$ is a nonempty class of subsets that contains $\emptyset$ and is closed under the operations of union, intersection, and difference. A **σ-ring** is a ring $\mathcal{R}$ that is closed under countable unions and intersections, that is, if $A_i \in \mathcal{R}$, $i = 1, 2, \ldots,$ then $A = \cup_{i=1}^{\infty} A_i \in \mathcal{R}$ and $\cap_{i=1}^{\infty} A_i = \left(A - \cup_{i=1}^{\infty} (A - A_i)\right) \in \mathcal{R}$.

Next, we can define a **σ-algebra** as a class of sets $\mathcal{F}$ that contains Ω and is a σ-ring. More formally, for Ω a given space, a σ-algebra on Ω is a family $\mathcal{F}$ of subsets of Ω that satisfies the following conditions:

i. $\Omega \in \mathcal{F}$.
ii. If a set $A \in \mathcal{F}$, then its complement $A' \in \mathcal{F}$, where $A' = \Omega - A$.

iii. If $\{A_i\}_{i \geq 1} \in \mathcal{F}$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$, that is, countable unions of sets in $\mathcal{F}$ are also in $\mathcal{F}$.

Note that since $\Omega \in \mathcal{F}$, we must have $\Omega' = \emptyset \in \mathcal{F}$; with $\left(\cup_{i=1}^{\infty} A_i\right)' = \cap_{i=1}^{\infty} A_i'$, it follows that $\mathcal{F}$ is closed under countable intersections as well. Note also that if $\{A_i\}_{i \geq 1} \in \mathcal{F}$, then $lim_{i \to \infty} A_i \in \mathcal{F}$, $lim_{i \to \infty} sup A_i \in \mathcal{F}$, and $lim_{i \to \infty} inf A_i \in \mathcal{F}$. The pair $(\Omega, \mathcal{F})$ is called a **measurable space** and the sets in $\mathcal{F}$ are termed ($\mathcal{F}$-) **measurable sets**.

Given a family $\mathcal{C}$ of subsets of Ω, there exists a smallest σ-algebra $\sigma(\mathcal{C})$ on Ω that contains $\mathcal{C}$, is contained in every σ-algebra that contains $\mathcal{C}$, and is unique. Here, $\sigma(\mathcal{C})$ is termed the **σ-algebra generated by** $\mathcal{C}$ and is specified as

$$\sigma(\mathcal{C}) = \cap \left\{ \mathcal{H}_j | \mathcal{H}_j \ \text{ a } \sigma\text{-algebra on } \Omega, \mathcal{C} \subset \mathcal{H}_j \right\}.$$

(For instance, if $\mathcal{C} = \{E\}$, $E \subset \Omega$, then $\sigma(\mathcal{C}) = \{\emptyset, E, E', \Omega\}$.) Now, if $\Omega = R^n$ and $\mathcal{C}$ is a family of open sets in R^n, then $\mathcal{B}^n = \sigma(\mathcal{C})$ is called the **Borel σ-algebra on Ω** and an element $B \in \mathcal{B}^n$ is called a **Borel set**. Hence, the Borel σ-algebra on Ω is the smallest σ-algebra generated by all the open subsets of R^n; and the **class of Borel sets** $\mathcal{B}^n$ in R^n is the σ-algebra generated by the open sets in R^n. In fact, the class of half-open intervals in R^n generates the σ-algebra $\mathcal{B}^n$ of Borel sets in R^n. Borel sets also include all open and closed sets, all countable unions of closed sets, among others.

Since σ-algebras will be of paramount importance in our subsequent analysis (especially in our review of the essentials of probability theory), let us consider Example 1.1.

Example 1.1 To keep the analysis manageable, suppose $\Omega = \{1,2,3,4\}$. Then possibly, $\mathcal{F} = \{\emptyset, \{1,2\}, \{3,4\}, \Omega\}$. Does $\mathcal{F}$ satisfy (i)–(iii) given earlier? If so, then $\mathcal{F}$ is a legitimate σ-algebra. Specifically,

1. As constructed, $\emptyset \in \mathcal{F}$ and $\Omega \in \mathcal{F}$. Hence, (i) holds.
2. $\emptyset' = \Omega, \Omega' = \emptyset \; \{1,2\}' = \{3,4\}, \{3,4\}' = \{1,2\}$. Clearly each of these subsets is a member of $\mathcal{F}$ and thus (ii) is valid.
3. Since $\mathcal{F}$ contains four disjoint subsets, partition the index set $I = \{1,2,3,4\}$ into four disjoint subsets according to

$$I_1 = \{i | A_i = \emptyset\} \neq \emptyset \; I_2 = \{i | A_i = \{1,2\}\} \neq \emptyset$$
$$I_3 = \{i | A_i = \{3,4\}\} \neq \emptyset, \text{ and } I_4 = \{i | A_i = \Omega\} \neq \emptyset.$$

Then

$$\begin{aligned}
\cup_{i=1}^{\infty} A_i &= \bigcup_{i \in I} A_i \\
&= \left(\bigcup_{i \in I_1} A_i\right) \cup \left(\bigcup_{i \in I_2} A_i\right) \cup \left(\bigcup_{i \in I_3} A_i\right) \cup \left(\bigcup_{i \in I_4} A_i\right) \\
&= \emptyset \cup \{1,2\} \cup \{3,4\} \cup \Omega \in \mathcal{F}
\end{aligned}$$

and thus $\mathcal{F}$ is a σ-algebra on Ω. ■

1.10 Set and Measure Functions

1.10.1 Set Functions

We previously defined the concept of a point-to-point function or mapping as a rule f that associates with a point x from a nonempty set X a unique point $y = f(x)$ in a nonempty set Y, where X, *a set of points*, was called the domain of the function. Now, let's consider a real-valued function whose domain is a *class of sets*, that is, we have a function of sets rather than a function of points. In this regard, consider a function $\mu: \mathcal{C} \to R^*$, where $\mathcal{C}$ is a nonempty class of sets and R^* denotes the set of extended real numbers. Thus μ is a rule that associates with each set $E \in \mathcal{C}$ a unique element $\mu(E)$, which is either a real number or $\pm\infty$. Some important types of set functions follow.

First, a set function $\mu: \mathcal{C} \to R^*$ is said to be (finitely) **additive** if

i. $\mu(\varnothing) = 0$; and
ii. for every finite collection $E_1, E_2, \dots, E_n$ of disjoint sets ($E_j \cap E_k = \varnothing, j \neq k$) in $\mathcal{C}$ such that $\cup_{i=1}^n E_i \in \mathcal{C}$, we have $\mu\left(\cup_{i=1}^n E_i\right) = \sum_{i=1}^n \mu(E_i) \in R^*$.

Remember that the domain $\mathcal{C}$ of μ is a finitely additive class of sets $\{E_i, i = 1, \dots, n\}$ and $\sum_{i=1}^n \mu(E_i)$ is defined in R^*. (If $\Omega = R$, $\mathcal{C}$ is the class of all *finite* intervals of R, and if E is taken to be (a, b) or $(a, b]$ or $[a, b)$ or $[a, b]$, then $\mu(E) = b - a$.)

It should be evident that a suitable domain of definition of an additive set function μ is a ring R since, if $E_i \in R, i = 1, \dots, n$, then $\cup_{i=1}^n E_i \in R$. So if $\mathcal{C}$ is a ring, then the set function $\mu: \to R^*$ is additive if and only if $\mu(\varnothing) = 0$ and, if E_j and E_k are disjoint sets in $\mathcal{C}$, then $\mu\left(E_j \cup E_k\right) = \mu\left(E_j\right) + \mu(E_k)$. In this regard, suppose $\mu: \mathcal{C} \to R^*$ is an additive set function defined as a ring $\mathcal{C}$ with sets $E_j, E_k \in \mathcal{C}$. Then

i. if $E_j \subset E_k$ and $\mu(E_j)$ is finite, then $\mu(E_k - E_j) = \mu(E_k) - \mu(E_j) \geq 0$;
ii. if $E_j \subset E_k$ and $\mu(E_j)$ is infinite, then $\mu(E_j) = \mu(E_k)$;
iii. if $E_j \subset E_k$ and $\mu(E_k)$ is finite, then $\mu(E_j)$ is finite; and
iv. if $\mu(E_k) = +\infty$, then $\mu(E_j) \neq -\infty$.

Next, a set function $\mu: \mathcal{C} \to R^*$ is termed **σ-additive** (or **countably** or **completely additive**) provided

i. the domain of μ is a σ-ring of sets $\mathcal{C}$;
ii. $\mu(\varnothing) = 0$; and
iii. for any disjoint sequence $E_1, E_2, \dots$ of sets in $\mathcal{C}$ such that $\cup_{i=1}^\infty E_i \in \mathcal{C}$, we have
$$\mu\left(\cup_{i=1}^\infty E_i\right) = \sum_{i=1}^\infty \mu(E_i) \in R^*.$$

Here, the domain $\mathcal{C}$ of μ is a *countably additive* class of sets $\{E_i, i = 1, 2, \dots\}$ and $\sum_{i=1}^\infty \mu(E_i) \in R^*$ is defined in the extended real numbers. Clearly, a σ-additive set function is also (finitely) additive, though the converse is not generally true. However, if $\mathcal{C}$ is a finite class of sets, then the additivity of μ: $\mathcal{C} \rightarrow R^*$ implies σ-additivity.

A set function μ: $\rightarrow R^*$ is said to be ***σ*-finite** if, for each set $E \in \mathcal{C}$, there is a sequence of sets $E_i \in \mathcal{C}$, $i = 1, 2, \ldots$, such that $E = \cup_{i=1}^{\infty} E_i$ and $\mu(E_i) < +\infty$ for all i. As this definition reveals, additivity is not a property of σ-finite set functions. For instance, consider the Borel σ-algebra in R^n that is generated by the collection of all "cubes" of the form $C = (a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_n, b_n)$, with $b_i > a_i, i = 1, \ldots, n$. Then, $\mu(C) = \prod_{i=1}^{n}(b_i - a_i)$. Here, μ is σ-finite since $R^n = \cup_{i=1}^{\infty}(-i, i)^n$.

A set function μ defined on $\mathcal{C}$ is **nondecreasing** if $\mu(E_k) \geq \mu(E_j)$ whenever $E_j \subset E_k$; it is **nonincreasing** if $\mu(E_k) \leq \mu(E_j)$ when $E_j \subset E_k$; and it is said to be **monotone** if it is either nondecreasing or nonincreasing. Now, if μ is additive and nondecreasing (resp. nonincreasing), then it is everywhere non-negative (resp. nonpositive). In fact, the reverse implication holds, that is, if μ is additive and everywhere non-negative (resp. nonpositive), then it is also nondecreasing (resp. nonincreasing).

1.10.2 Measure Functions

Let R^+ denote the set of non-negative real numbers together with $+\infty$, that is, $R^+ = \{x \in R^* | x \geq 0\}$. A **measure function** on a σ-ring $\mathcal{C}$ is any non-negative σ-additive set function μ: $\mathcal{C} \rightarrow R^+$. (For any subset $A \in \mathcal{C}$ we assume that $-\infty < \mu(A) < +\infty$.) Note that since a measure function μ on $\mathcal{C}$ is non-negative, it must also be nondecreasing. A **Borel measure** is a measure function μ on the σ-algebra $\mathcal{B}$ of Borel subsets of a given topological space (Ω, $\mathcal{J}$), that is,

$$\mu : \mathcal{B} \rightarrow [0, +\infty).$$

A couple of important characteristics of measure functions are as follows:

i. If μ is a measure function on $\mathcal{C}$ and if $\{E_i, i = 1, 2, \ldots\}$ is any sequence of sets from $\mathcal{C}$, then μ is **countably subadditive** or $\mu(\cup_{i=1}^{\infty} E_i) \leq \sum_{i=1}^{\infty} \mu(E_i)$.
ii. If μ is a measure function on $\mathcal{C}$ and if $\{E_i, i = 1, 2, \ldots\}$ is any sequence of sets from $\mathcal{C}$ with $\mu(\cup_{i=1}^{\infty} E_i) < +\infty$, then $\mu(lim_{i\rightarrow\infty} E_i) = lim_{i\rightarrow\infty} \mu(E_i)$.

We next examine the continuity of set functions. To this end, suppose $\mathcal{R}$ is a ring and the set function μ: $\mathcal{R} \rightarrow R^*$ is additive with $\mu(A) > -\infty$ for all sets $A \in \mathcal{R}$:

i. μ is **continuous from below** at A if $lim_{i\rightarrow\infty} \mu(E_i) = \mu(A)$ for every monotone increasing sequence $\{E_i\}$ in $\mathcal{R}$ that converges to A.
ii. μ is **continuous from above** at A if $lim_{i\rightarrow\infty} \mu(E_i) = \mu(A)$ for every monotone decreasing sequence $\{E_i\}$ in $\mathcal{R}$ for which $\mu(E_i) < +\infty$ for some i.
iii. μ is continuous at $A \in \mathcal{R}$ if it is continuous at A from both above and below. Moreover, under the aforementioned assumptions, if

iv. μ is σ-additive (and thus additive) on $\mathcal{R}$, then μ is continuous at A for all sets $A \in \mathcal{R}$.

Given two classes $\mathcal{C}$ and $\mathfrak{D}$ of subsets of Ω, with $\mathcal{C} \subset \mathfrak{D}$, and set functions μ: $\mathcal{C} \to R^*$ and τ: $\mathfrak{D} \to R^*$ respectively, τ is termed an **extension** of μ if, for all $A \in \mathcal{C}$, $\tau(A) = \mu(A)$; and μ is called a **restriction** of τ to $\mathcal{C}$.

In later chapters, we shall be concerned with issues pertaining to the convergence of sequences of random variables. To adequately address these issues, we need to be able to define measures on countable unions and intersections of "measurable sets." To accomplish this task, we need to assume that the collection of measurable sets is a σ-algebra $\mathcal{F}$, that is, $\mathcal{F}$ contains Ω and is a σ-ring. This requirement enables us to confine our analysis, for the most part, to "measure spaces," where a **measure space** is a triple $(\Omega, \mathcal{F}, \mu)$ consisting of a space Ω, a σ-algebra $\mathcal{F}$ on Ω (a collection subsets of Ω), and μ: $\mathcal{F} \to R^+$ is a measure on $\mathcal{F}$.

1.10.3 Outer Measure Functions

Suppose $\mathcal{C}$ is the class of all subsets of a space Ω. Then μ^*: $\mathcal{C} \to R^+$ is an **outer measure function** on Ω if

i. $\mu^*(\varnothing) = 0$;
ii. μ^* is nondecreasing (i.e., for subsets $E_j \subset E_k$, $\mu(E_j) \le \mu(E_k)$); and
iii. μ^* is countably subadditive—i.e., for any sequence $\{E_i,\ i = 1, 2, \ldots\}$ of subsets of Ω,

$$\mu^*\left(\cup_{i=1}^{\infty} E_i\right) \le \sum_{i=1}^{\infty} \mu^*(E_i). \tag{1.4}$$

In sum, μ^* is said to be non-negative, monotone, and countably subadditive. We note that every measure on the class of all subsets of Ω is an outer measure on Ω; and, in defining an outer measure on Ω, no "additivity" requirement was in effect.

Given that μ^* is an outer measure on Ω, a subset E is said to be **measurable** with respect to μ^*, or simply μ^*-measurable, if for every set $A \subset \Omega$,

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E') \tag{1.5}$$

(given that $A = (A \cap E) \cup (A \cap E')$). Thus, a subset E of Ω is μ^*-measurable if it partitions a set $A \subset \Omega$ into two subsets, $A \cap E$ and $A \cap E'$, on which μ^* is additive. As this definition reveals, a set E is not innately measurable—its measurability depends upon the outer measure employed. That is, to define the measurability of a set E, we start with an arbitrary set A and we examine the effect of E on the outer measure of A, $\mu^*(A)$. If E is measurable, then it is sufficiently "well-behaved" in that it does not partition A in a way that compromises the additivity of μ^*, that is, if we partition A into $A \cap E$ and $A \cap E'$, then the outer measures of $A \cap E$ and $A \cap E'$ add up correctly to μ^* (A).

Since μ^* is countably subadditive, we have, from (1.4), $\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E')$ for all sets A, $E \in \Omega$. Hence, E is μ^*-measurable if and only if

$$\mu^*(A) \geq \mu^*(A \cap E) + \mu^*(A \cap E') \tag{1.6}$$

for every set $A \in \Omega$. Since this inequality holds for any set A for which $\mu^*(A) = +\infty$, it follows that a necessary and sufficient condition for E to be μ^*-measurable is that $\mu^*(A) < +\infty$ for every $A \in \Omega$.

Some important properties of outer measures are the following:

i. If $E \in \Omega$ is μ^*-measurable, then E' is also μ^*-measurable.
ii. If $\mu^*(E) = 0$, then $E \in \Omega$ is measurable.
iii. Any finite union of μ^*-measurable sets in Ω is μ^*-measurable.
iv. If $\{E_i, i = 1, 2, \ldots\}$ is a sequence of disjoint μ^*-measurable sets in Ω and if $\mathcal{G} = \cup_{j=1}^{\infty} E_j$, then for any set $A \in \Omega$, $\mu^*(A \cap \mathcal{G}) = \sum_{j=1}^{\infty} \mu^*(A \cap E_j)$.
v. Any countable union of μ^*-measurable sets in Ω is μ^*-measurable.
vi. Any countable union of disjoint μ^*-measurable sets in Ω is μ^*-measurable.
vii. If $\{E_i, i = 1, 2, \ldots\}$ is a sequence of disjoint μ^*-measurable sets in Ω and if, for each n, $\mathcal{G}_n = \cup_{j=1}^{n} E_j$, then, for each set $A \in \Omega$, $\mu^*(A \cap \mathcal{G}_n) = \sum_{j=1}^{n} \mu^*(A \cap E_j)$.

Why are outer measures important? Simply because they are useful for constructing measure functions. That is, given that the outer measure μ^* has as its domain the class of all subsets of the space Ω, a *restriction* of μ^* to a "smaller" domain always generates a measure function. In this regard, suppose μ^* is an outer measure function on Ω and let $\mathcal{E}$ be the class of μ^*-measurable sets. Then $\mathcal{E}$ is a completely additive class (a σ-algebra) and the restriction of μ^* to $\mathcal{E}$ is a measure function μ.

An outer measure μ^* is said to be **regular** if, for every subset $A \in \Omega$, there is a μ^*-measurable set $E \supset A$ such that $\mu^*(E) = \mu^*(A)$. (Here, E is said to be a **measurable cover** for A.) Thus, an outer measure is regular if it effectuates measurable sets in a manner that guarantees that every set $A \in \Omega$ has a measurable cover E.

Key properties of regular outer measures are the following:

i. If μ^* is a regular outer measure on Ω and $\{E_i, i = 1, 2, \ldots\}$ is an increasing sequence of sets, then $\mu^*(lim_{i\to\infty} E_i) = lim_{i\to\infty} \mu^*(E_i)$.
ii. If μ^* is a regular outer measure on Ω for which $\mu^*(\Omega) < +\infty$, then a subset $E \in \Omega$ is measurable if and only if $\mu^*(\Omega) = \mu^*(\Omega \cap E) + \mu^*(\Omega \cap E') = \mu^*(E) + \mu^*(E')$. (This result follows from Equation (1.5) with $A = \Omega$, since (1.5) must hold for *any* set A.)

Next, let Ω be a metric space. An outer measure μ^* on Ω is a **metric outer measure** if $\mu^*(\emptyset) = 0$; μ^* is nondecreasing and countably subadditive; and μ^*

is additive on separated sets (i.e., for subsets E and F in Ω with $d(E, F) > 0$, $\mu^*(E \cup F) = \mu^* E + \mu^*(F)$).[6] We note briefly the following:

i. If μ^* is a metric outer measure, then any closed set is measurable.
ii. If μ^* is a metric outer measure, then every Borel set is measurable (since the class $\mathcal{E}$ of μ^*-measurable sets contains the open sets, and thus contains $\mathcal{B}$, the class of Borel sets).

1.10.4 Complete Measure Functions

Given a measure function $\mu: \mathcal{C} \rightarrow R^+$, the class $\mathcal{C}$ of subsets of Ω is **complete with respect to μ** if $E \subset F$, $F \in \mathcal{C}$, and $\mu(F) = 0$ implies $E \in \mathcal{C}$. Now, if $\mu: \mathcal{C} \rightarrow R^+$ is such that $\mathcal{C}$ is complete with respect to μ, then μ is said to be **complete**. Hence, μ is complete if its domain contains all subsets of sets of measure zero, that is, every subset of a set of measure zero is measurable.

For a measure space $(\Omega, \mathcal{F}, \mu)$, the **completion of $\mathcal{F}$**, denoted $\mathcal{F}_c$, with respect to a measure μ on $\mathcal{F}$ involves all subsets $A \in \Omega$ such that there exist sets $E, F \in \mathcal{F}$, with $E \subset A \subset F$, and $\mu(F - E) = 0$. The **completion of μ**, μ_c, is defined on $\mathcal{F}_c$ as $\mu_c(A) = \mu_c \ (E) = \mu_c(F)$; it is the unique extension of μ to $\mathcal{F}_c$. For $A \in \mathcal{F}_c, \mu_c(A) = inf\{\mu(F) | F \in \mathcal{F}, A \subset F\} = sup\{\mu(E) | E \in \mathcal{F}, E \subset A\}$.

The **complete measure space** $(\Omega, \mathcal{F}_c, \mu_c)$ is thus the completion of $(\Omega, \mathcal{F}, \mu)$. In fact, $(\Omega, \mathcal{F}_c, \mu_c)$ is the smallest complete measure space that contains $(\Omega, \mathcal{F}, \mu)$.

If a measure μ is obtained by restricting an outer measure μ^* to $\mathcal{E}$, the class of sets of Ω that are μ^*-measurable, then μ is a complete measure. In fact, any measure generated by an outer measure is complete.

1.10.5 Lebesgue Measure

In what follows, our discussion will focus in large part on a class $\mathcal{M}$ of open sets (containing $\emptyset$) in $\Omega = R$. This will then facilitate our development of the Lebesgue integral.

Let us express the **length of a bounded interval** I (which may be open, closed, or half-open) with endpoints a and b, $a < b$, as $l(I) = b - a$. Our objective herein is to extend this "length" concept to arbitrary subsets of R, for example, for a subset $E \subset R$, the notion of the "length of E" is simply its measure $\mu(E)$. In particular, we need to explore the concept of *Lebesgue measure* of a set E, $\mu(E)$, and specify the family of *Lebesgue measurable sets*. Our starting point is the concept of Lebesgue outer measure.

6 For sets $E, F \in R^n$, the **distance between sets** E, F is $d(E,F) = inf\{\|x - y\| | x \in E, y \in F\}$. If $E \cap F \neq \emptyset, d(E,F) = 0$.

For each subset $E \subset R$, the **Lebesgue outer measure** $\mu^*(E)$ is defined as

$$\mu^*(E) = inf\left\{\sum_{i=1}^{n} l(I_i) | \{I_i\} \text{ is a sequence of open intervals with } E \subset \cup_{i=1}^{n} I_i\right\}.$$

What is the significance of this expression? Suppose E can be covered by multiple sets of open intervals, where the union of each particular set of open intervals contains E. Since the total length of any set of intervals can overestimate the measure of E (it may contain points not in E), we need to take the greatest lower bound of the lengths of the interval sets in order to isolate the covering set whose length fits *E as closely as possible* and whose constituent intervals *do not overlap*.

Given the discussion on outer measures in Section 1.10.3, it follows that the key properties of Lebesgue outer measures are the following:

i. For every set $E \subset R, 0 \leq \mu^*(E) \leq +\infty$.
ii. μ^* is nondecreasing.
iii. μ^* is countably subadditive, that is, for any sequence $\{E_i, i = 1, 2, \ldots\}$ of subsets of R,

$$\mu^*\left(\cup_{i=1}^{\infty} E_i\right) \leq \sum_{i=1}^{n} \mu^*(E_i).$$

iv. μ^* generalizes or extends the concept of "length" in that $\mu^*(I) = l(I)$.

How does the concept of *Lebesgue outer measure* translate to the notion of *Lebesgue measure* itself? In order to transition from $\mu^*(E)$ to $\mu(E)$, we need an additional condition on E. Specifically, a set $E \subset R$ is **Lebesgue measurable** if for every set $A \subset R$,

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E'). \tag{1.5}$$

This requirement is not new (see the discussion underlying Equation (1.5) of Section 1.10.3). As explained therein, if for every A the partition of A induced by E (the sets $A \cap E$ and $A \cap E'$) has outer measures that correctly add up to the outer measure of A itself, then set E is "well-behaved" in that E does not adversely impact or distort the outer measure of A when E is used to partition A. The upshot of all this is that, under (1.5), $\mu^*(E)$ yields $\mu(E)$. That is, if E is Lebesgue measurable, then the **Lebesgue measure** of E is defined to be its outer measure $\mu^*(E)$ and simply written as $\mu(E)$.

As far as the properties of Lebesgue measure $\mu(E)$ are concerned, they mirror those of $\mu^*(E)$ (see properties (i)–(iv)), but with one key exception—property (iii) involving *countable subadditivity* is replaced by *countable additivity*: if $\{E_i, i = 1, 2, \ldots\}$ is a sequence of disjoint subsets of R, then

$$(\text{iii})' \mu\left(\cup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mu(E_i).$$

How should the family of Lebesgue measurable sets (denoted $\mathcal{M}$) be defined? Clearly, we need to specify the *largest* family $\mathcal{M}$ of subsets of R for which μ: $\mathcal{M} \to R^+$ and properties (i), (ii), (iii)′, and (iv) hold. Hence, the family **of Lebesgue measurable sets** $\mathcal{M}$ encompasses the collection of all open intervals as well as all finite unions of intervals on the real line. Then for $E \in \mathcal{M}$, $\mu(E)$, the Lebesgue measure of E, is the total length of E when E is decomposed into the union of a finite number of disjoint intervals.

We note in passing that $\emptyset$ and R are Lebesgue measurable with $\mu(\emptyset) = 0$ and $\mu(R) = +\infty$, respectively; open and closed intervals of real numbers are Lebesgue measurable; every open set and every closed set is Lebesgue measurable; every Borel set (which includes countable sets, open and closed intervals, all open sets and all closed sets) is Lebesgue measurable; any countable set of real numbers has Lebesgue measure equal to zero; if E is Lebesgue measurable, then so is E'; and if $\{E_i, i = 1, 2, \ldots\}$ is a sequence of Lebesgue measurable sets, then $\cup_{i=1}^{\infty} E_i$ and $\cap_{i=1}^{\infty} E_i$ are Lebesgue measurable sets.

Example 1.2 Let $E = [a,b] \subset \mathcal{M}$ with $\{x_1, x_2, x_3\} \subset [a,b], a < x_1 < x_2 < x_3 < b$. Consider the set $A = E\ \{x_1, x_2, x_3\}$. For the measure function μ: $\mathcal{M} \to R^+$,

$$\begin{aligned}\mu(A) &= \mu\{[a,x_1) \cup (x_1,x_2) \cup (x_2,x_3) \cup (x_3,b]\}\\ &= \mu[a,x_1) + \mu(x_1,x_2) + \mu(x_2,x_3) + \mu(x_3,b]\\ &= (x_1 - a) + (x_2 - x_1) + (x_3 - x_2) + (b - x_3) = b - a = \mu(E). \blacksquare\end{aligned}$$

1.10.6 Measurable Functions

Let $(X, \mathfrak{D})$ and $(Y, \mathcal{G})$ be measurable spaces, where $\mathfrak{D}$ is a σ-algebra on X and is a σ-algebra on Y, respectively. A **measurable function** is a mapping f: $X \to Y$ such that $f^{-1}(G) \in \mathfrak{D}$ for every set $G \in \mathcal{G}$. Clearly, the measurability of f depends upon $\mathfrak{D}$ and $\mathcal{G}$ and not on the particular measures defined on these σ-algebras. As this definition indicates, measurable functions are defined in terms of inverse images of sets. (Thus, measurable functions are mappings that occur between measurable spaces in much the same way that continuous functions are mappings between topological spaces.) To elaborate on this notion, if f: $\Omega \to \mathcal{G}$ and $A \in \cdot$, let $f^{-1}(A) = \{x \in \Omega | f(x) \in A\}$ and call $f^{-1}(A)$ the **inverse image** of set A under rule f. (Note: $f^{-1}(A)$ contains all of the points in the domain Ω of f mapped by f into A; it does *not* denote the inverse function of f.) Key properties of the inverse image of A are the following:

i. $f^{-1}(A') = (f^{-1}(A))'$ for all $A \in \mathcal{G}$.
ii. If A, $B \in \mathcal{G}$, then $f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B)$.
iii. If $\{A_k\} \subset \mathcal{G}$, then $f^{-1}\left(\cup_{k=1}^{\infty} A_k\right) = \cup_{k=1}^{\infty} f^{-1}(A_k)$.

In addition, if $\mathcal{C}$ is a collection of subsets of $\mathcal{G}$, let $f^{-1}(\mathcal{C}) = \{f^{-1}(A)|A \in \mathcal{C}\}$. In this regard, if f: $\Omega \to \mathcal{G}$ and $\mathcal{C}$ is a collection of subsets of $\mathcal{G}$, then (a) for $\mathcal{C}$ a σ-algebra on $\mathcal{G}$, $f^{-1}(\mathcal{C})$ is a σ-algebra on Ω; and (b) $f^{-1}(\mathcal{F}(\mathcal{C})) = \mathcal{F}(f^{-1}(\mathcal{C}))$, $(\Omega, \mathcal{F})$ a measurable space.

A measurable mapping $g: X \to Y$ on a measure space $(\Omega, \mathcal{F}, \mu)$ is **measure preserving** if $\mu(g^{-1}(A)) = \mu(A)$ for all measurable sets A.

We previously termed the σ-algebra generated by intervals (open, closed, half-open) in R the **Borel σ-algebra** $\mathcal{B}$. In this regard, if $(\Omega, \mathcal{F})$ is a measurable space, the mapping f: $\Omega \to \mathcal{R}$ is **$\mathcal{F}$-measurable** if $f^{-1}(B) \in \mathcal{F}$ for every Borel set $B \in R$. In fact, the collection of sets $f^{-1}(B)$, where B is contained within the Borel subsets of R, is a σ-algebra on Ω. In addition, if the collection $\mathcal{C}$ of Borel subsets of R generates the Borel σ-algebra, then f: $\Omega \to R$ is $\mathcal{F}$-measurable if and only if $f^{-1}(\mathcal{C}) \subset \mathcal{F}$. Equivalently, the mapping f: $\Omega \to R$ is $\mathcal{F}$-measurable if and only if the set $\{x \in \Omega | f(x) \leq a\}$ is measurable (i.e., it is a member of $\mathcal{F}$) for all $a \in R$. (Note: "≤" can be replaced by "<, ≥, >.")

The **indicator** or **characteristic function** of a set $A \in \Omega$ is defined as

$$\chi_A(x) = \begin{cases} 1, x \in A; \\ 0, x \notin A. \end{cases} \tag{1.7}$$

If A, B are two subsets of Ω, then

$$\chi_{A \cap B} = min\{\chi_A, \chi_B\} = \chi_A \cdot \chi_B;$$
$$\chi_{A \cup B} = max\{\chi_A, \chi_B\} = \chi_A + \chi_B - \chi_A \cdot \chi_B; \text{ and}$$
$$\chi_{A'} = 1 - \chi_A.$$

Moreover, if $A_i, i = 1, \ldots, n,$ and $B_j, j = 1, \ldots, m,$ are subsets of Ω and $X = \sum_{i=1}^{n} x_i \chi_{A_i}$ and $Y = \sum_{j=1}^{m} y_j \chi_{B_j}$, then

$$X \cdot Y = \sum_{i=1}^{n} \sum_{j=1}^{m} x_i y_j \chi_{A_i \cap B_j}.$$

In addition, if $\{A_i\}_{i=1}^{n}$ and $\{B_j\}_{j=1}^{m}$ are partitions of Ω, then $\{A_i \cap B_j\}_{\text{all } i,j}$ is also a partition of Ω, and thus

$$X + Y = \sum_{i=1}^{n} \sum_{j=1}^{m} (x_i + y_j) \chi_{A_i \cap B_j}.$$

If $\mathcal{F}$ is a σ-algebra on Ω, then $\emptyset$ and Ω are members of $\mathcal{F}$. In addition, with set $A \in \mathcal{F}$, it follows that $A' \in \mathcal{F}$. Hence, $\mathcal{F}_{\chi_A} = \{\emptyset, A, A', \Omega\}$, and thus χ_A **is $\mathcal{F}$-measurable** if and only if $A \in \mathcal{F}$. Note also that if X and Y are $\mathcal{F}$-measurable functions on Ω, then $X + Y, X - Y, X \cdot Y$, and cX(c a real scalar) are all $\mathcal{F}$-measurable. Suppose $Y(x) \neq 0$ for all $x \in \Omega$. Then X/Y is also $\mathcal{F}$-measurable.

Suppose $f:\Omega \rightarrow R^*$ is a measurable function with $A=\{x|f(x)\geq 0\}$ and $B=\{x|f(x)\leq 0\}$. If $f^+=f\cdot\chi_A$ and $f^-=-f\cdot\chi_B$, then the **positive part of *f*** is defined as

$$f^+ = max\{f(x),0\} = \begin{cases} f(x), f(x)\geq 0; \\ 0, f(x)<0; \end{cases}$$

and the **negative part of *f*** is defined as

$$f^- = max\{-f(x),0\} = \begin{cases} -f(x), f(x)\leq 0; \\ 0, f(x)\geq 0, \end{cases}$$

where f^+ and f^- are both positive functions on Ω. With $\mathcal{F}$ a σ-algebra on Ω and f is measurable, sets $G=f^{-1}(\{x|x\geq 0\})$ and $H=f^{-1}(\{x|x\leq 0\})$ are in the σ-algebra generated by $\mathcal{F}$ (denoted $\mathcal{F}_f$). Hence, f^+, f^- and $|f|$ are all $\mathcal{F}_f$-measurable. The upshot of this discussion is that an arbitrary measurable function f can be written in a canonical way as the difference between two positive measurable functions as $f=f^+-f^-$. In addition, $|f|=f^++f^-$.

A function Φ: $\Omega \rightarrow R$ defined on a measurable space (Ω, $\mathcal{F}$) is a **simple function** if there are disjoint measurable sets $A_1,\ldots,A_n$ and real scalars $c_1,\ldots,c_n$ such that

$$\Phi = \sum_{i=1}^{n} c_i \chi_{A_i}. \quad (1.8)$$

Clearly, Φ takes on finitely many, finite values $c_i, i=1,\ldots,n$. Since simple functions are measurable, any measurable function may be approximated by simple functions. In fact, for f: $\Omega \rightarrow R^+$ a non-negative measurable function, there is a monotone increasing sequence $\{\Phi_i\}$ of simple functions that converges pointwise to f.

Given a measure space (Ω, $\mathcal{F}$, μ), if $\Omega = \cup_{i=1}^{n} A_i$ and the sets A_i are disjoint, then these sets are said to form a (finite) **dissection** of Ω. They are said to form an **$\mathcal{F}$-dissection** if $A_i \in \mathcal{F}, i=1,\ldots,n$. A function f: $\Omega \rightarrow R$ is termed **$\mathcal{F}$-simple** if it can be expressed as $f(x)=\sum_{i=1}^{n} c_i\chi_{A_i}$, where the A_i's, $i=1,\ldots,n$, form an $\mathcal{F}$-dissection of Ω. Thus, $f(x)$ takes on a constant value c_i on the set A_i, given that the A_i's are disjoint subsets of $\mathcal{F}$.

A sequence of measurable functions $\{f_n\}$ from a measure space (Ω, $\mathcal{F}$, μ) to $R*$ **converges pointwise** to a function f: $\Omega \rightarrow R^*$ if $lim_{n\rightarrow\infty} f_n(x) = f(x)$ for every $x \in \Omega$. Moreover, f itself is measurable. A sequence $\{f_n\}$ **converges pointwise a.e.**[7] to f if it converges pointwise to f except on a set M of measure zero.

7 A **set of measure zero** is a measurable set M such that $\mu(M)=0$. A property or condition that holds for all $x \in \Omega - M$, where M is a set of measure zero, is said to hold **almost everywhere** (abbreviated "a.e.") or "except on a set of measure zero." Note: a subset of a set of measure zero need not be measurable; but if it is measurable, then it must have measure zero.

If $(\Omega, \mathcal{F}, \mu)$ is a complete measure space and $\{f_n\}$ converges pointwise a.e. to f, then f is measurable.

We note briefly that if the measurable space is $(R^n, \mathcal{B}^n)$, a $\mathcal{B}^n$-measurable function is termed a **Borel-measurable function**.

1.10.7 Lebesgue Measurable Functions

Suppose $(X, \mathfrak{D})$ and $(Y, \mathcal{G})$ are measurable spaces, with X and Y equipped with the σ-algebras $\mathfrak{D}$ and $\mathcal{G}$, respectively. Then, as indicated in Section 1.10.6, the function $f: X \to Y$ is **measurable** if the anti-image of E under f is in $\mathfrak{D}$ for every $E \subset \mathcal{G}$, that is, $f^{-1}(E) = \{x \in X | f(x) \in E\} \in \mathfrak{D}$ for all $E \subset \mathcal{G}$.

Let us now get a bit more specific. Suppose $(R, \mathcal{L})$ and $(R, \mathcal{B})$ are measurable spaces, with $\mathcal{L}$ the σ-algebra of Lebesgue measurable sets and $\mathcal{B}$ the Borel σ-algebra on R. (Remember that $\mathcal{B}$ is the smallest σ-algebra containing all the open sets.) The function $f: R \to R$ is **Lebesgue measurable** if the anti-image of B under f is a Lebesgue measurable subset of R for every Borel subset B of R, that is, $f^{-1}(B) = \{x \in R | f(x) \in B\} \in \mathcal{L}$ for all $B \in \mathcal{B}$. (Clearly, the domain and range of f involve different σ-algebras defined on the same set R.) In very basic terms, for a bounded interval I, a function $f: I \to R$ is Lebesgue measurable if, for every open set $B \subseteq R$, the anti-image $f^{-1}(B)$ is measurable in I.

An important alternative way of specifying a function that is Lebesgue measurable is the following. If $(X, \mathcal{L})$ is a measurable space, then $f: X \to R$ is **Lebesgue measurable** if and only if $f^{-1}([a, +\infty]) = \{x \in X | f(x) > a\} \in \mathcal{L}$ for all $a \in R$ (Figure 1.3). (Note: equivalent statements involve ">" being replaced by " ≥ " or "<" or " ≤ .")

To summarize: A function f between measurable spaces is measurable if the anti-image of each measurable set is *measurable*. A function f is *Lebesgue*

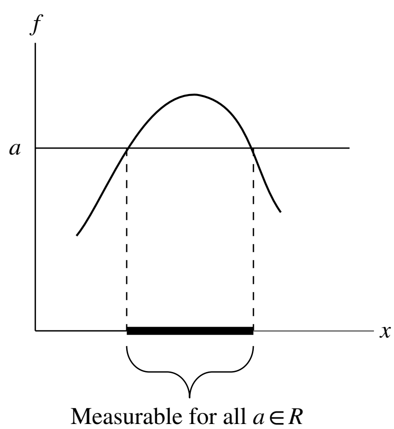

Figure 1.3 A measurable subset of $\mathcal{L}$.

measurable if and only if the anti-image of each of the sets $[a, +\infty]$ is a Lebesgue measurable set.

We end this discussion by commenting that continuous functions, monotone functions, step functions, and Riemann integrable functions are all Lebesgue measurable. Moreover, if $f, g: R \to R$ are Lebesgue measurable functions and $c \in R$, then $cf, f+g, fg, |f|$ and $|g|$, and $max\{f, g\}$ are all Lebesgue measurable.

1.11 Normed Linear Spaces

1.11.1 Space of Bounded Real-Valued Functions

Given a nonempty set A, suppose that each pair of elements $x, y \in A$ can be operated on, by a process called **addition** ("+"), to yield a new element $x + y = z \in A$, where the operator "+" satisfies the following:

a. $x + y = y + x$ (commutative law).
b. $x + (y + z) = (x + y) + z$ (associative law).
c. $0 \in A$ such that $x + 0 = x$ for every $x \in A$ (zero is the **additive identity**).
d. $-x \in A$ such that $x + (-x) = 0$ ($-x$ is the **additive inverse**).

Suppose also that, for each real scalar α and for each element $x \in A$, α and x can be operated on, by a process called **scalar multiplication** ("·"), to yield a new element $\alpha x = y \in A$, where the operator "·" satisfies the following:

$$\left.\begin{array}{l}\text{a. } \alpha(x+y) = \alpha x + \alpha y \\ \text{b. } (\alpha+\beta)x = \alpha x + \beta x\end{array}\right\} \text{distributive laws.}$$

c. $(\alpha\beta)x = \alpha(\beta x)$ (associative law).

d. $1 \cdot x = x$ (1 is the **multiplicative identity**).

Now, if A is closed under the operations of addition and scalar multiplication, then A (which can be viewed as an algebraic system) is termed a **linear space** (or **vector space**). A nonempty subset C of a linear space A constitutes a **linear subspace** of A if $x + y$ is in C when $x, y \in C$; and $\alpha x \in C$ (α a real scalar) when $x \in C$. As was the case with A itself, 0 and $-x$ are elements of C whenever $x \in C$.

In Section 1.4, we introduced the concept of a norm—a function that assigns to each x within a space a real number $\|x\|$ such that the properties of nonnegativity, homogeneity, and the triangle inequality hold. If the norm "$\|\cdot\|$" is defined on a linear space A, then A becomes a **normed linear space**. We also noted in Section 1.4 that a normed linear space is a metric space with respect to the metric $d(x, y) = \|x - y\|$ induced by the norm (Equation (1.2)).

We next consider the notion of a **function space**—a linear space whose elements are functions defined on some nonempty set X, with pointwise addition and scalar multiplication satisfying the following:

$$\begin{aligned}
&\text{a. } (f + g)(x) = f(x) + g(x).\\
&\text{b. } (\alpha f)(x) = \alpha f(x).\\
&\text{c. the zero element } 0 \in X \text{ is the constant function "0"}\\
&\quad \text{whose only value is the scalar } 0.\\
&\text{d. } (-f)(x) = -f(x).
\end{aligned} \tag{1.9}$$

Consider now the set of all real-valued functions defined on X. Clearly, X is a *real linear space* whose elements satisfy (1.9). If we separate out from X the subset B of all bounded real-valued functions ($f \varepsilon B$ is **bounded** if there exists a real scalar M such that $|f(x)| \leq M$), then B is itself a linear space. In addition, if we define on the elements of B the norm

$$\|f\| = sup|f(x)|, \tag{1.10}$$

then B is a metric space.

Suppose $\{f_n\}$ is a sequence of real-valued functions defined on B and that, for each $x \in B$, $\{f_n(x)\}$ is a Cauchy sequence (to review, $\{f_n(x)\}$ is a Cauchy sequence if for all $\varepsilon > 0$ there exists an $N_{\varepsilon/2} > 0$ such that for all $m, n > N_{\varepsilon/2}, \|f_m(x) - f_n(x)\| < \varepsilon$). Thus, B is complete in that, for each $x \in B$ and every Cauchy sequence $\{f_n\}$ defined on B, there exists a well-defined continuous **limit function** $f(x) = lim_{n\to\infty} f_n(x)$ so that $\{f_n(x)\}$ **converges pointwise** to $f(x)$. In fact, any normed space with the property that every Cauchy sequence defined on it is convergent is complete.

The preceding discussion enables us to conclude that B constitutes an important type of function space, namely, a **Banach space**—a complete normed linear (metric) space. We next turn to another type of function space that is also a Banach space.

1.11.2 Space of Bounded Continuous Real-Valued Functions

A key property that the elements of a function space can possess is *continuity*. To explore this characteristic, let's assume at the outset that the set of all real-valued functions is defined on a metric space X. Furthermore, given B above (as defined earlier, B is the subset of X containing all bounded real-valued functions), let $C(X) \subset B$ denote the set of all bounded continuous functions defined on X.

It should thus be evident that

a. if f, g are continuous real-valued functions defined on X, then, pointwise, $f + g$ and αf (α a real scalar) are also continuous;

b. $C(X)$ is a linear subspace of the linear (metric) space B; and
c. $C(X)$ is a closed[8] subset of the linear (metric) space B.

But remember that B is a Banach space and, since a closed linear subspace of a Banach space is also a Banach space, it follows that $C(X)$, the set of all bounded continuous real-valued functions defined on a metric space X with norm (1.10) is a Banach space.

1.11.3 Some Classical Banach Spaces

1. Let R^n denote the set of all **vectors** or ordered n-tuples $x = (x_1, x_2, \ldots, x_n)$ of real numbers. For elements $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ in R^n, let us define, coordinatewise, addition and scalar multiplication as

$$x + y = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n),$$
$$\alpha x = (\alpha x_1, \alpha x_2, \ldots, \alpha x_n), \alpha \text{ a real scalar,}$$

respectively. In addition,, the zero (or null) element $0 = (0, 0, \ldots, 0)$ (also an n-tuple) and $-x = (-x_1, -x_2, \ldots, -x_n)$ are elements in R^n. For any element $x \in R^n$, let us define the norm of x, $\|x\|$, by

$$\|x\| = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{\frac{1}{2}}. \tag{1.11}$$

(See Section 1.4 for a discussion of the properties of this norm.) Given (1.11), it is evident that R^n can be characterized as a normed linear (metric) space (also called n-dimensional **Euclidean space**—since (1.11) is the **Euclidean norm)**. Moreover, it is complete with respect to the metric (1.2) and thus amounts to a complete normed linear space or Banach space.

2. Let $L^p(\mu)$ denote the set of all measurable functions f defined on a measure space $(\Omega, \mathcal{F}, \mu)$ and having the property that $|f(x)|^p$ is integrable, with **p-norm**

$$\|f\|_p = \left(\int |f(x)|^p d\mu(x) \right)^{\frac{1}{p}}, 1 \le p < +\infty. \tag{1.12}$$

For $p \in [1, +\infty), L^p(\mu)$ is complete with respect to (1.12) and thus constitutes a Banach space.

8 Suppose $f \in B$ with $f \in \overline{C(X)}$ (the closure of $C(X)$). Let d be the metric on X, with $\varepsilon > 0$ given. Since f is in $\overline{C(X)}$, there exists a function f_0 in $C(X)$ such that $\|f - f_0\| < \varepsilon'$ implies $|f(x) - f(x_0)| < \varepsilon'$ for each $x \in X$, where ε' is proportional to ε. With f_0 continuous at x_0, there exists a $\delta_{\varepsilon'} > 0$ such that $d(x, x_0) = \|x - x_0\| < \delta_{\varepsilon'}$ implies $|f(x_0) - f_0(x_0)| < \varepsilon'$. Since $\|x - x_0\| < \delta_\varepsilon$ implies that $|f(x) - f(x_0)| < \varepsilon$, we see that f is continuous at x_0 (x_0 arbitrary). Hence, $f \in C(X) = \overline{C(X)}$ so that C(X) must be closed (Simmons, 1963, p. 83).

3. Let H be an n-dimensional linear (vector) space with the **inner product norm** defined by

$$\|x\| = (x,x)^{\frac{1}{2}}, \tag{1.13}$$

where (x, x) is the **inner product**[9] defined by

$$(x,x) = \sum_{i=1}^{n} |x_i|^2, x \in H.$$

(A linear space equipped with an inner product is called an **inner product space**.) Clearly, H is a normed linear space and is complete with respect to the norm given by (1.13); it will be called a **Hilbert space**—a complete normed inner product space. Although H is always a Banach space whose norm is determined by an inner product $(\cdot,\cdot)$ (i.e., $\|f\| = (f,f)^{1/2}$ for all f in the space), the converse does not generally hold. So what is the essential difference between a Banach space and a Hilbert space? The difference is in the source of the norm, that is, for a Banach space, the norm is defined directly as $\|\cdot\| : B \to [0, +\infty)$ for all points x, y (and scalar c) satisfying the properties outlined earlier in footnote 2; and for a Hilbert space, the norm is defined by an inner product (Equation (1.13)). The inner product is not defined on a Banach space.

The ordered n-tuples (vectors) $x, y \in R^n$ are said to be **orthogonal** if $(x,y) = \sum_{i=1}^{n} x_i y_i = 0$. Elements x and y within a Hilbert space (H) are orthogonal if $(x, y) = 0$ and **orthonormal** if, in addition, $\|x\| = \|y\| = 1$. An **orthonormal set** in H is a nonempty subset of H that consists of mutually orthogonal unit vectors $e_i, i = 1, 2, \ldots$. (A **unit vector** e_i has a "1" as its ith component and "0's" elsewhere.) That is, an orthonormal set is a non-empty subset $\{e_i, i = 1, 2, \ldots\}$ of H with the properties (1) $(e_i, e_j) = 0, i \neq j$; and (2) $\|e_i\| = 1$ for all i.

An orthonormal sequence $\{e_i, i = 1, 2, \ldots\}$ in H is **complete** if the only member of H that is orthogonal to every e_i is the **null vector** 0 (which contains all zero components). (Stated alternatively, we cannot find a vector e such that $\{\{e_i\}, e\}$ is an orthonormal set that properly contains $\{e_i, i = 1, 2, \ldots\}$.)

Suppose $\{e_1, e_2, \ldots, e_n\}$ is a finite orthonormal set in H. If x is any vector in H, then

1. $\sum_{i=1}^{n} |(x, e_i)|^2 \leq \|x\|^2$;

9 The inner product (x, x) satisfies the following conditions:

a. For $x \in H$, $(x, x) \geq 0$ and $(x, x) = 0$ if and only if $x = 0$ (positive semidefiniteness).
b. For $x, y \in H$, $(x,y) = (y,x)$ (symmetry).
c. For $x, y \in H$, with a, b real scalars, $((ax_1 + bx_2), y) = (ax_1, y) + (bx_2, y)$ (linear in its first argument).
d. For $x, y \in H$, $|(x,y)| \leq \|x\|\|y\|$ (**Cauchy–Schwarz inequality**).

2. $x = \sum_{i=1}^{n} (x, e_i) e_i$; and
3. $\left(x - \sum_{i=1}^{n} (x, e_i) e_i, e_j \right) = 0$ for each j.

An **orthonormal basis** for a Hilbert space (H) is a basis[10] consisting of non-zero orthonormal vectors. Such vectors are linearly independent and span H in the sense that every element in H can be written as a linear combination of the basis vectors. In fact, every Hilbert space contains a maximal orthonormal set that serves as a basis.

Suppose in (1.12), we set $p = 2$. Then the class of real-valued square-integrable functions

$$L^2(\mu) = \|f\|_2 = \left(\int |f|^2 d\mu(x) \right)^{\frac{1}{2}} < +\infty \tag{1.12.1}$$

is a Hilbert space. In addition, if (Ω, ϑ, μ) is a measure space and the functions $f, g \in L^2(\mu)$, then the inner product of f and g is

$$(f, g) = \int fg d\mu, \tag{1.14}$$

where $|(f \cdot g)| \leq \|f\| \|g\|$.

The space $L^2(a, b)$, the collection of Borel measurable real-valued square integrable functions f on (a, b) (i.e., $\int_a^b |f(t)|^2 dt < +\infty$), is a Hilbert space. For this space, the inner product is $(f, g) = \int_a^b f(t) g(t) dt$, and the associated norm and metric are, respectively, $\|f\|_2 = \left(\int_a^b |f(t)|^2 dt \right)^{1/2}$ and $d(f, g) = \|f - g\| = \left(\int_a^b |f(t) - g(t)|^2 dt \right)^{1/2}$. (Here, the functions f, g are considered equal if they differ on (a, b) only on a set of measure zero.)

1.12 Integration

Our approach in this section is to first define the *integral of a non-negative simple function*. We then define the *integral of a non-negative measurable*

---

10 To review, a vector $x \in R^n$ is a **linear combination** of the vectors $x_j \in R^n, j = 1, \ldots, m$, if there exists scalars $\lambda_j, j = 1, \ldots, m$, such that $x = \sum_{j=1}^{m} \lambda_j x_j$. A set of vectors is **linearly independent** if the trivial combination $0x_1 + \cdots + 0x_n$ is the only linear combination of the x_j which equals the null vector. (The set of vectors $\{x_j, j = 1, \ldots, m\}$ is said to be **linearly dependent** if there exists scalars $\lambda_j, j = 1, \ldots, m$, not all zero such that $\sum_{j=1}^{m} \lambda_j x_j = 0$.) The vectors $x_j, j = 1, \ldots, m$, **span** R^n if every element of R^n can be written as a linear combination of the x_j's. Hence, the x_j's constitute a **spanning set** for R^n. A **basis** for R^n is a linearly independent set of vectors from R^n which spans R^n. Thus, every vector in R^n can be expressed as a linear combination of the basis vectors.

function via an approximation by simple functions. Next comes the definition of the *integral of a measurable function*, followed by the specification of the *integral of a measurable function on a measurable set*. In what follows, (Ω, $\mathcal{F}$, μ) is taken to be a measure space. However, if $\Omega = R$ admits the Borel σ-algebra, $\mathcal{F}$ is the σ-algebra of Lebesgue measurable sets in R, and the measure $\mu : \mathcal{F} \to [0, +\infty]$ is given by $\mu(E) = \mu^*(E), E \in \mathcal{F}$, then the integrals defined below are also *Lebesgue integrals*. (Readers not familiar with the *Lebesgue integral* are encouraged to read Appendix B to this chapter along with Taylor (1973) before tackling this section.)

1.12.1 Integral of a Non-negative Simple Function

Recall (Section 1.10.5) that a non-negative simple function has the form

$$\emptyset(x) = \sum_{i=1}^{n} c_i \chi_{E_i}(x), c_i \geq 0, i = 1, \ldots, n, \tag{1.15}$$

where the indicator function χ_{E_i} is defined as

$$\chi_{E_i}(x) = \begin{cases} 1, x \in E_i; \\ 0, x \in E_i'. \end{cases}$$

The **integral of a non-negative simple function** with respect to μ is defined in terms of the integral operator "$\int$" as

$$\int_{\Omega} \emptyset \, d\mu = \sum_{i=1}^{n} c_i \mu(E_i), \tag{1.16}$$

where $E_i = \{x | \emptyset(x) = c_i\}, \int_{\Omega} \chi_{E_i} d\mu = \int_{E_i} d\mu = \mu(E_i) < +\infty$ and the sum on the right-hand side of (1.16) is well defined since each of its terms is non-negative. (It is important to note that since the specification of a simple function in terms of indicator functions is not unique, this definition of the integral is independent of the actual specification used.) For the simple function given in Equation (1.15), suppose set $A \in \mathcal{F}$ is measurable. Then the **integral of a non-negative simple function over a set** A is defined as

$$\int_{A} \emptyset \, d\mu = \sum_{i=1}^{n} c_i \mu(E_i \cap A). \tag{1.16.1}$$

As far as the essential properties of the integral operator "$\int$" are concerned, it is *linear* as well as *order preserving* on the class of non-negative simple functions. That is, given two non-negative simple functions $\emptyset = \sum_{i=1}^{n} c_i \chi_{E_i}$ and $\psi = \sum_{j=1}^{m} d_j \chi_{F_j}$, the simple function

$\emptyset + \psi = \sum_{i=1}^{n} \sum_{j=1}^{m} (c_i + d_j)\chi_{E_i \cap F_j}$, and thus

$$\int_{\Omega} (\emptyset + \psi) d\mu = \int_{\Omega} \emptyset \, d\mu + \int_{\Omega} \psi d\mu; \left(\text{“} \int \text{” is } \textbf{linear} \right) \tag{1.17}$$

while, for $\emptyset \geq \psi$,

$$\int_{\Omega} \emptyset \, d\mu \geq \int_{\Omega} \psi d\mu. \left(\text{“} \int \text{” is } \textbf{order preserving or monotonic} \right) \tag{1.18}$$

1.12.2 Integral of a Non-negative Measurable Function Using Simple Functions

Suppose the non-negative function f: $\Omega \to R^+$ is measurable. It was noted in Section 1.10.5 that there exists a monotone increasing sequence $\{f_n\}$ of simple functions that converge pointwise to f. Given that $\int_{\Omega} f_n d\mu$ is defined for all n, and the said sequence is monotonic, it follows that the limit of $\int_{\Omega} f_n d\mu$ is an element of R^+. Hence, we may define the operation of **integration for non-negative measurable functions** as

$$\int_{\Omega} f d\mu = \lim_{n \to \infty} \int_{\Omega} f_n d\mu. \tag{1.19}$$

Since $\int_{\Omega} f d\mu$ may be finite or infinite in R^+, we may conclude that a non-negative measurable function f is integrable with respect to a measure μ if the limit in (1.19) is finite. In addition, if $f \geq 0$ is measurable, then **integration for non-negative measurable functions over a set** A is defined as

$$\int_{A} f d\mu = \sup_{\emptyset} \int_{A} \emptyset \, d\mu < +\infty. \tag{1.19.1}$$

where the supremum is taken over all simple functions $\emptyset$ with $0 \leq \emptyset \leq f$.

1.12.3 Integral of a Measurable Function

Suppose the function f: $\Omega \to R^+$ is measurable. Then, as indicated earlier in Section 1.10.5, so are f_+ and f_-. If f_+ and f_- are integrable with respect to μ, then $f = f_+ - f_-$ itself is integrable with respect to μ, and thus

$$\int_{\Omega} f d\mu = \int_{\Omega} f_+ d\mu - \int_{\Omega} f_- d\mu \tag{1.20}$$

so that this expression defines **integration for the class of integrable measurable functions**. Also, for set $A \in \mathcal{F}$, f is integrable over a set A if

$$\int_A |f|\, d\mu = \int_A f_+\, d\mu + \int_A f_-\, d\mu < +\infty. \tag{1.20.1}$$

1.12.4 Integral of a Measurable Function on a Measurable Set

Let set $A \in \mathcal{F}$. Suppose $\int f\chi_A d\mu$ is defined (e.g., either $f\chi_A$ is non-negative and measurable, or $f\chi_A$ is measurable and integrable). Then

$$\int_A f d\mu = \int_A f\chi_A d\mu. \tag{1.21}$$

Thus, ***f* is integrable over a set *A* if $f\chi_A$ is integrable**. (Note that if $A \in \mathcal{F}$ and $\mu(A) = 0$, then $f\colon \Omega \to R^*$ is integrable over A with $\int_A f d\mu = 0$.)

For a measure space $(\Omega, \mathcal{F}, \mu)$ and $f\colon \Omega \to R^*$ an integrable function with respect to μ over Ω, some additional properties of the integral operator "$\int$" are the following:

i. For A and B disjoint sets in $\mathcal{F}$,

$$\int_{A\cup B} f d\mu = \int_A f d\mu + \int_B f d\mu.$$

ii. $|f|$ is integrable and $\left|\int_\Omega f d\mu\right| = \int_\Omega |f| d\mu$.

iii. For a constant $c \in R$, cf is integrable and $\int_\Omega cf d\mu = c\int_\Omega f d\mu$.

iv. If $f \geq 0$, then $\int_\Omega f d\mu \geq 0$; but if $f \geq 0$ and $\int f d\mu = 0$, then $f = 0$ a.e.

v. If $g\colon \Omega \to R^*$ is integrable with respect to μ over Ω, then if $f = g$ a.e., it follows that

$$\int_\Omega f d\mu = \int_\Omega g d\mu.$$

vi. If sets $A, B \in \mathcal{F}$ with $A \subset B$ and $f \geq 0$, then $\int_A f d\mu \leq \int_B f d\mu$.

vii. Let μ be the counting measure[11] on $\Omega = \{1, 2, 3, \ldots\}$ and define the measurable function $f\colon \Omega \to R$ as $f(j) = a_j, j \in \Omega$. Then

---

11 Let $(\Omega, \mathcal{F}, \nu)$ be a measure space. The **counting measure** ν on Ω is defined as $\nu(A)$ = number of elements in $A \in \mathcal{F}$. This measure is finite if Ω is a finite set; it is σ-finite if Ω is countable.

$$\int_{\Omega} f(j)\, d\mu(j) = \sum_{j=1}^{\infty} a_j.$$

This integral is well-defined if $f \geq 0$, or if the sum on the right-hand converges absolutely. (If $\sum_{j=1}^{\infty} |a_j|$ is convergent, then $\sum_{j=1}^{\infty} |a_j|$ is termed **absolutely convergent**.) In either instance we say that f is integrable with respect to μ.

viii. A measurable function f is integrable on $A \in \mathcal{F}$ if and only if $|f|$ is integrable on A.

ix. If f is integrable on set $A \in \mathcal{F}$, if g is measurable, and if $|g| \leq f$ a.e. on A, then g is integrable on A and $\int_A g d\mu \leq \int_A f d\mu$.

x. If f is *any* function and for set $A \in \mathcal{F}$, if $\mu(A) = 0$, then $\int_A f d\mu = 0$.

1.12.5 Convergence of Sequences of Functions

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. A sequence of functions $\{f_n\}$, where $f_n : \Omega \to R^+$, **converges pointwise** to a function $f : \Omega \to R^+$ if $lim_{n\to\infty} f_n(x) = f(x)$ for every $x \in \Omega$. Here f is termed a **limiting function**. The sequence $\{f_n\}$ **converges pointwise a.e.** to f if it converges pointwise to f on $\Omega - A$, where $A \in \mathcal{F}$ is a set of measure zero.

In this regard, let $\{f_n\}$ be a sequence of functions that converges pointwise to a limiting function f. When can we legitimately conclude that $\int_{\Omega} f_n d\mu$ converges to $\int_{\Omega} f d\mu$? Two conditions that guarantee the convergence of the integrals $\int_{\Omega} f_n d\mu$ are (1) the monotone convergence of the sequence $\{f_n\}$; and (2) a uniform bound on $\{f_n\}$ by an integrable function.

To set the stage for a discussion of the first condition, let us define a sequence of functions $\{f_n\}$, where $f_n : \Omega \to R^+$, as **monotone increasing** if $f_1(x) \leq \cdots \leq f_n(x) \leq \ldots$ for every $x \in \Omega$. We then have Theorem 1.12.1.

Theorem 1.12.1 (Lebesgue) Monotone Convergence Theorem (MCT)
Let $\{f_n\}$ be a monotone increasing sequence of non-negative measurable functions $f_n : \Omega \to [0, +\infty]$ on a measure space $(\Omega, \mathcal{F}, \mu)$ and let $f : \Omega \to [0, +\infty]$ be the pointwise limit of $\{f_n\}$ or $f(x) = lim_{n\to\infty} f_n(x)$. Then

$$\lim_{n\to\infty} \int_{\Omega} f_n d\mu = \int_{\Omega} f d\mu.$$

(Note that if f is integrable on Ω $\left(lim_{n\to\infty} \int_{\Omega} f_n d\mu < +\infty\right)$, this theorem posits the convergence of the integrals $\int_{\Omega} f_n d\mu$ to $\int_{\Omega} f d\mu$. If f is not integrable on Ω, then possibly f_n is integrable for all n and $\int_\Omega f_n d\mu \to +\infty$ as $n \to +\infty$.) A consequence of the MCT is Corollary 1.12.1.

Corollary 1.12.1 (corollary to the MCT). Let $\{f_n\}$, $f: \Omega \to [0, +\infty]$, be a sequence of non-negative measurable functions and set $f = \sum_{n=1}^{\infty} f_n$. Then

$$\int_\Omega f d\mu = \sum_{n=1}^{\infty} \int_\Omega f_n d\mu.$$

A generalization of the MCT is provided by Lemma 1.12.1.

Lemma 1.12.1 Fatou's lemma
Let $\{f_n\}$ be a sequence of non-negative measurable functions $f_n : \Omega \to [0, \infty]$ on a measure space $(\Omega, \mathcal{F}, \mu)$. Then

$$\int_\Omega \left(\lim_{n\to\infty} inf f_n \right) d\mu \le \lim_{n\to\infty} inf \int_\Omega f_n d\mu.$$

As this lemma indicates, the limit of the integrals on the right-hand side of this inequality is always at least as large as the integral of the limit function on the left-hand side.

The second aforementioned condition is incorporated in Theorem 1.12.2.

Theorem 1.12.2 (Lebesgue) Dominated Convergence Theorem (DCT)
Let $\{f_n\}$ be a sequence of integrable functions, where $f_n: \Omega \to R^*$, on a measure space $(\Omega, \mathcal{F}, \mu)$ that converge pointwise to a limit function $f: \Omega \to R^*$. If there is an integrable function $g : \Omega \to [0, \infty]$ such that $|f_n(x)| \le g(x)$ for all $x \in \Omega$ (and independent of n), then f is integrable and

$$\lim_{n\to\infty} \int_\Omega f_n d\mu = \int_\Omega f d\mu.$$

We close this section with an additional useful Theorem 1.12.3.

Theorem 1.12.3 Let $\{A_n\}$ be a sequence of disjoint measurable sets with $A = \cup_{n=1}^{\infty} A_n$ and let f be a non-negative measurable function that is integrable on each set A_n. Then f is integrable on A if and only if $\sum_{n=1}^{\infty} \int_{A_n} f d\mu < +\infty$, so that

$$\int_A f d\mu = \sum_{n=1}^{\infty} \int_{A_n} f d\mu.$$