

CHAPTER 4

Differentiation

Differentiation originated in connection with the problems of drawing tangents to curves and of finding maxima and minima of functions. Pierre de Fermat (1601–1665), the founder of the modern theory of numbers, is credited with having put forth the main ideas on which differential calculus is based.

In this chapter, we shall introduce the notion of differentiation and study its applications in the determination of maxima and minima of functions. We shall restrict our attention to real-valued functions defined on R , the set of real numbers. The study of differentiation in connection with multivariable functions, that is, functions defined on R^n ($n \geq 1$), will be considered in Chapter 7.

4.1. THE DERIVATIVE OF A FUNCTION

The notion of differentiation was motivated by the need to find the tangent to a curve at a given point. Fermat's approach to this problem was inspired by a geometric reasoning. His method uses the idea of a tangent as the limiting position of a secant when two of its points of intersection with the curve tend to coincide. This has led to the modern notation associated with the derivative of a function, which we now introduce.

Definition 4.1.1. Let $f(x)$ be a function defined in a neighborhood $N_r(x_0)$ of a point x_0 . Consider the ratio

$$\phi(h) = \frac{f(x_0 + h) - f(x_0)}{h}, \quad (4.1)$$

where h is a nonzero increment of x_0 such that $-r < h < r$. If $\phi(h)$ has a limit as $h \rightarrow 0$, then this limit is called the derivative of $f(x)$ at x_0 and is

denoted by $f'(x_0)$. It is also common to use the notation

$$\left. \frac{df(x)}{dx} \right|_{x=x_0} = f'(x_0).$$

We thus have

$$f'(x_0) = \lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h}. \quad (4.2)$$

By putting $x = x_0 + h$, formula (4.2) can be written as

$$f'(x_0) = \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0}.$$

If $f'(x_0)$ exists, then $f(x)$ is said to be differentiable at $x = x_0$. Geometrically, $f'(x_0)$ is the slope of the tangent to the graph of the function $y = f(x)$ at the point (x_0, y_0) , where $y_0 = f(x_0)$. If $f(x)$ has a derivative at every point of a set D , then $f(x)$ is said to be differentiable on D .

It is important to note that in order for $f'(x_0)$ to exist, the left-sided and right-sided limits of $\phi(h)$ in formula (4.1) must exist and be equal as $h \rightarrow 0$, or as x approaches x_0 from either side. It is possible to consider only one-sided derivatives at $x = x_0$. These occur when $\phi(h)$ has just a one-sided limit as $h \rightarrow 0$. We shall not, however, concern ourselves with such derivatives in this chapter. $\square$

Functions that are differentiable at a point must necessarily be continuous there. This will be shown in the next theorem.

Theorem 4.1.1. Let $f(x)$ be defined in a neighborhood of a point x_0 . If $f(x)$ has a derivative at x_0 , then it must be continuous at x_0 .

Proof. From Definition 4.1.1 we can write

$$f(x_0 + h) - f(x_0) = h\phi(h).$$

If the derivative of $f(x)$ exists at x_0 , then $\phi(h) \rightarrow f'(x_0)$ as $h \rightarrow 0$. It follows from Theorem 3.2.1(2) that

$$f(x_0 + h) - f(x_0) \rightarrow 0$$

as $h \rightarrow 0$. Thus for a given $\epsilon > 0$ there exists a $\delta > 0$ such that

$$|f(x_0 + h) - f(x_0)| < \epsilon$$

if $|h| < \delta$. This indicates that $f(x)$ is continuous at x_0 . $\square$

It should be noted that even though continuity is a necessary condition for differentiability, it is not a sufficient condition, as can be seen from the following example: Let $f(x)$ be defined as

$$f(x) = \begin{cases} x \sin \frac{1}{x}, & x \neq 0, \\ 0, & x = 0. \end{cases}$$

This function is continuous at $x = 0$, since $f(0) = \lim_{x \rightarrow 0} f(x) = 0$ by the fact that

$$\left| x \sin \frac{1}{x} \right| \leq |x|$$

for all x . However, $f(x)$ is not differentiable at $x = 0$. This is because when $x = 0$,

$$\begin{aligned} \phi(h) &= \frac{f(h) - f(0)}{h} \\ &= \frac{h \sin \frac{1}{h} - 0}{h}, \quad \text{since } h \neq 0, \\ &= \sin \frac{1}{h}, \end{aligned}$$

which does not have a limit as $h \rightarrow 0$. Hence, $f'(0)$ does not exist.

If $f(x)$ is differentiable on a set D , then $f'(x)$ is a function defined on D . In the event $f'(x)$ itself is differentiable on D , then its derivative is called the second derivative of $f(x)$ and is denoted by $f''(x)$. It is also common to use the notation

$$\frac{d^2 f(x)}{dx^2} = f''(x).$$

By the same token, we can define the n th ($n \geq 2$) derivative of $f(x)$ as the derivative of the $(n - 1)$ st derivative of $f(x)$. We denote this derivative by

$$\frac{d^n f(x)}{dx^n} = f^{(n)}(x), \quad n = 2, 3, \dots$$

We shall now discuss some rules that pertain to differentiation. The reader is expected to know how to differentiate certain elementary functions such as polynomial, exponential, and trigonometric functions.

Theorem 4.1.2. Let $f(x)$ and $g(x)$ be defined and differentiable on a set D . Then

1. $[\alpha f(x) + \beta g(x)]' = \alpha f'(x) + \beta g'(x)$, where α and β are constants.
2. $[f(x)g(x)]' = f'(x)g(x) + f(x)g'(x)$.
3. $[f(x)/g(x)]' = [f'(x)g(x) - f(x)g'(x)]/g^2(x)$ if $g(x) \neq 0$.

Proof. The proof of (1) is straightforward. To prove (2) we write

$$\begin{aligned}
 & \lim_{h \rightarrow 0} \frac{f(x+h)g(x+h) - f(x)g(x)}{h} \\
 &= \lim_{h \rightarrow 0} \frac{[f(x+h) - f(x)]g(x+h) + f(x)[g(x+h) - g(x)]}{h} \\
 &= \lim_{h \rightarrow 0} g(x+h) \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \\
 &\quad + f(x) \lim_{h \rightarrow 0} \frac{g(x+h) - g(x)}{h}.
 \end{aligned}$$

However, $\lim_{h \rightarrow 0} g(x+h) = g(x)$, since $g(x)$ is continuous (because it is differentiable). Hence,

$$\lim_{h \rightarrow 0} \frac{f(x+h)g(x+h) - f(x)g(x)}{h} = g(x)f'(x) + f(x)g'(x).$$

Now, to prove (3) we write

$$\begin{aligned}
 & \lim_{h \rightarrow 0} \frac{f(x+h)/g(x+h) - f(x)/g(x)}{h} \\
 &= \lim_{h \rightarrow 0} \frac{g(x)f(x+h) - f(x)g(x+h)}{hg(x)g(x+h)} \\
 &= \lim_{h \rightarrow 0} \frac{g(x)[f(x+h) - f(x)] - f(x)[g(x+h) - g(x)]}{hg(x)g(x+h)} \\
 &= \frac{\lim_{h \rightarrow 0} \{g(x)[f(x+h) - f(x)]/h - f(x)[g(x+h) - g(x)]/h\}}{g(x)\lim_{h \rightarrow 0} g(x+h)} \\
 &= \frac{g(x)f'(x) - f(x)g'(x)}{g^2(x)}. \quad \square
 \end{aligned}$$

Theorem 4.1.3 (The Chain Rule). Let $f: D_1 \rightarrow R$ and $g: D_2 \rightarrow R$ be two functions. Suppose that $f(D_1) \subset D_2$. If $f(x)$ is differentiable on D_1 and $g(x)$ is differentiable on D_2 , then the composite function $h(x) = g[f(x)]$ is differentiable on D_1 , and

$$\frac{dg[f(x)]}{dx} = \frac{dg[f(x)]}{df(x)} \frac{df(x)}{dx}.$$

Proof. Let $z = f(x)$ and $t = f(x + h)$. By the fact that $g(z)$ is differentiable we can write

$$\begin{aligned} g[f(x + h)] - g[f(x)] &= g(t) - g(z) \\ &= (t - z)g'(z) + o(t - z), \end{aligned} \quad (4.3)$$

where, if we recall, the o -notation was introduced in Section 3.3. We then have

$$\frac{g[f(x + h)] - g[f(x)]}{h} = \frac{t - z}{h} g'(z) + \frac{o(t - z)}{t - z} \cdot \frac{t - z}{h}. \quad (4.4)$$

As $h \rightarrow 0$, $t \rightarrow z$, and hence

$$\lim_{h \rightarrow 0} \frac{t - z}{h} = \lim_{h \rightarrow 0} \frac{f(x + h) - f(x)}{h} = \frac{df(x)}{dx}.$$

Now, by taking the limits of both sides of (4.4) as $h \rightarrow 0$ and noting that

$$\lim_{h \rightarrow 0} \frac{o(t - z)}{t - z} = \lim_{t \rightarrow z} \frac{o(t - z)}{t - z} = 0,$$

we conclude that

$$\frac{dg[f(x)]}{dx} = \frac{df(x)}{dx} \frac{dg[f(x)]}{df(x)}. \quad \square$$

NOTE 4.1.1. We recall that $f(x)$ must be continuous in order for $f'(x)$ to exist. However, if $f'(x)$ exists, it does not have to be continuous. Care should be exercised when showing that $f'(x)$ is continuous. For example, let us consider the function

$$f(x) = \begin{cases} x^2 \sin \frac{1}{x}, & x \neq 0, \\ 0, & x = 0. \end{cases}$$

Suppose that it is desired to show that $f'(x)$ exists, and if so, to determine if it is continuous. To do so, let us first find out if $f'(x)$ exists at $x = 0$:

$$\begin{aligned} f'(0) &= \lim_{h \rightarrow 0} \frac{f(h) - f(0)}{h} \\ &= \lim_{h \rightarrow 0} \frac{h^2 \sin \frac{1}{h}}{h} \\ &= \lim_{h \rightarrow 0} h \sin \frac{1}{h} = 0. \end{aligned}$$

Thus the derivative of $f(x)$ exists at $x = 0$ and is equal to zero. For $x \neq 0$, it is clear that the derivative of $f(x)$ exists. By applying Theorem 4.1.2 and using our knowledge of the derivatives of elementary functions, $f'(x)$ can be written as

$$f'(x) = \begin{cases} 2x \sin \frac{1}{x} - \cos \frac{1}{x}, & x \neq 0, \\ 0, & x = 0. \end{cases}$$

We note that $f'(x)$ exists for all x , but is not continuous at $x = 0$, since

$$\lim_{x \rightarrow 0} f'(x) = \lim_{x \rightarrow 0} \left(2x \sin \frac{1}{x} - \cos \frac{1}{x} \right)$$

does not exist, because $\cos(1/x)$ has no limit as $x \rightarrow 0$. However, for any nonzero value of x , $f'(x)$ is continuous.

If $f(x)$ is a convex function, then we have the following interesting result, whose proof can be found in Roberts and Varberg (1973, Theorem C, page 7):

Theorem 4.1.4. If $f: (a, b) \rightarrow R$ is convex on the open interval (a, b) , then the set S where $f'(x)$ fails to exist is either finite or countable. Moreover, $f'(x)$ is continuous on $(a, b) - S$, the complement of S with respect to (a, b) .

For example, the function $f(x) = |x|$ is convex on R . Its derivatives does not exist at $x = 0$ (why?), but is continuous everywhere else.

The sign of $f'(x)$ provides information about the behavior of $f(x)$ in a neighborhood of x . More specifically, we have the following theorem:

Theorem 4.1.5. Let $f: D \rightarrow R$, where D is an open set. Suppose that $f'(x)$ is positive at a point $x_0 \in D$. Then there is a neighborhood $N_\delta(x_0) \subset D$ such that for each x in this neighborhood, $f(x) > f(x_0)$ if $x > x_0$, and $f(x) < f(x_0)$ if $x < x_0$.

Proof. Let $\epsilon = f'(x_0)/2$. Then, there exists a $\delta > 0$ such that

$$f'(x_0) - \epsilon < \frac{f(x) - f(x_0)}{x - x_0} < f'(x_0) + \epsilon$$

if $|x - x_0| < \delta$. Hence, if $x > x_0$,

$$f(x) - f(x_0) > \frac{(x - x_0)f'(x_0)}{2},$$

which shows that $f(x) > f(x_0)$ since $f'(x_0) > 0$. Furthermore, since

$$\frac{f(x) - f(x_0)}{x - x_0} > 0,$$

then $f(x) < f(x_0)$ if $x < x_0$. $\square$

If $f'(x_0) < 0$, it can be similarly shown that $f(x) < f(x_0)$ if $x > x_0$, and $f(x) > f(x_0)$ if $x < x_0$.

4.2. THE MEAN VALUE THEOREM

This is one of the most important theorems in differential calculus. It is also known as the theorem of the mean. Before proving the mean value theorem, let us prove a special case of it known as *Rolle's theorem*.

Theorem 4.2.1 (Rolle's Theorem). Let $f(x)$ be continuous on the closed interval $[a, b]$ and differentiable on the open interval (a, b) . If $f(a) = f(b)$, then there exists a point c , $a < c < b$, such that $f'(c) = 0$.

Proof. Let d denote the common value of $f(a)$ and $f(b)$. Define $h(x) = f(x) - d$. Then $h(a) = h(b) = 0$. If $h(x)$ is also zero on (a, b) , then $h'(x) = 0$ for $a < x < b$ and the theorem is proved. Let us therefore assume that $h(x) \neq 0$ for some $x \in (a, b)$. Since $h(x)$ is continuous on $[a, b]$ [because $f(x)$ is], then by Corollary 3.4.1 it must achieve its supremum M at a point ξ in $[a, b]$, and its infimum m at a point η in $[a, b]$. If $h(x) > 0$ for some $x \in (a, b)$, then we must obviously have $a < \xi < b$, because $h(x)$ vanishes at both end points. We now claim that $h'(\xi) = 0$. If $h'(\xi) > 0$ or < 0 , then by Theorem 4.1.5, there exists a point x_1 in a neighborhood $N_\delta(\xi) \subset (a, b)$ at which $h(x_1) > h(\xi)$, a contradiction, since $h(\xi) = M$. Thus $h'(\xi) = 0$, which implies that $f'(\xi) = 0$, since $h'(x) = f'(x)$ for all $x \in (a, b)$. We can similarly arrive at the conclusion that $f'(\eta) = 0$ if $h(x) < 0$ for some $x \in (a, b)$. In this case, if $h'(\eta) \neq 0$, then by Theorem 4.1.5 there exists a point x_2 in a neigh-

neighborhood $N_{\delta_2}(\eta) \subset (a, b)$ at which $h(x_2) < h(\eta) = m$, a contradiction, since m is the infimum of $h(x)$ over $[a, b]$.

Thus in both cases, whether $h(x) > 0$ or < 0 for some $x \in (a, b)$, we must have a point c , $a < c < b$, such that $f'(c) = 0$. $\square$

Rolle's theorem has the following geometric interpretation: If $f(x)$ satisfies the conditions of Theorem 4.2.1, then the graph of $y = f(x)$ must have a tangent line that is parallel to the x -axis at some point c between a and b . Note that there can be several points like c inside (a, b) . For example, the function $y = x^3 - 5x^2 + 3x - 1$ satisfies the conditions of Rolle's theorem on the interval $[a, b]$, where $a = 0$ and $b = (5 + \sqrt{13})/2$. In this case, $f(a) = f(b) = -1$, and $f'(x) = 3x^2 - 10x + 3$ vanishes at $x = \frac{1}{3}$ and $x = 3$.

Theorem 4.2.2 (The Mean Value Theorem). If $f(x)$ is continuous on the closed interval $[a, b]$ and differentiable on the open interval (a, b) , then there exists a point c , $a < c < b$, such that

$$f(b) = f(a) + (b - a)f'(c).$$

Proof. Consider the function

$$\Phi(x) = f(x) - f(a) - A(x - a),$$

where

$$A = \frac{f(b) - f(a)}{b - a}.$$

The function $\Phi(x)$ is continuous on $[a, b]$ and is differentiable on (a, b) , since $\Phi'(x) = f'(x) - A$. Furthermore, $\Phi(a) = \Phi(b) = 0$. It follows from Rolle's theorem that there exists a point c , $a < c < b$, such that $\Phi'(c) = 0$. Thus,

$$f'(c) = \frac{f(b) - f(a)}{b - a},$$

which proves the theorem. $\square$

The mean value theorem has also a nice geometric interpretation. If the graph of the function $y = f(x)$ has a tangent line at each point of its length between two points P_1 and P_2 (see Figure 4.1), then there must be a point Q between P_1 and P_2 at which the tangent line is parallel to the secant line through P_1 and P_2 . Note that there can be several points on the curve between P_1 and P_2 that have the same property as Q , as can be seen from Figure 4.1.

The mean value theorem is useful in the derivation of several interesting results, as will be seen in the remainder of this chapter.

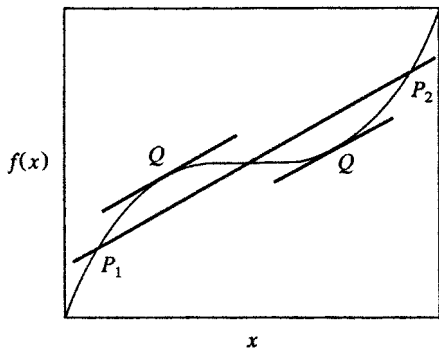


Figure 4.1. Tangent lines parallel to the secant line.

Corollary 4.2.1. If $f(x)$ has a derivative $f'(x)$ that is nonnegative (non-positive) on an interval (a, b) , then $f(x)$ is monotone increasing (decreasing) on (a, b) . If $f'(x)$ is positive (negative) on (a, b) , then $f(x)$ is strictly monotone increasing (decreasing) there.

Proof. Let x_1 and x_2 be two points in (a, b) such that $x_1 < x_2$. By the mean value theorem, there exists a point x_0 , $x_1 < x_0 < x_2$, such that

$$f(x_2) = f(x_1) + (x_2 - x_1)f'(x_0).$$

If $f'(x_0) \geq 0$, then $f(x_2) \geq f(x_1)$ and $f(x)$ is monotone increasing. Similarly, if $f'(x_0) \leq 0$, then $f(x_2) \leq f(x_1)$ and $f(x)$ is monotone decreasing. If, however, $f'(x) > 0$, or $f'(x) < 0$ on (a, b) , then strict monotonicity follows over (a, b) . $\square$

Theorem 4.2.3. If $f(x)$ is monotone increasing [decreasing] on an interval (a, b) , and if $f(x)$ is differentiable there, then $f'(x) \geq 0$ [$f'(x) \leq 0$] on (a, b) .

Proof. Let $x_0 \in (a, b)$. There exists a neighborhood $N_r(x_0) \subset (a, b)$. Then, for any $x \in N_r(x_0)$ such that $x \neq x_0$, the ratio

$$\frac{f(x) - f(x_0)}{x - x_0}$$

is nonnegative. This is true because $f(x) \geq f(x_0)$ if $x > x_0$ and $f(x) \leq f(x_0)$ if $x < x_0$. By taking the limit of this ratio as $x \rightarrow x_0$, we claim that $f'(x_0) \geq 0$. To prove this claim, suppose that $f'(x_0) < 0$. Then there exists a $\delta > 0$ such that

$$\left| \frac{f(x) - f(x_0)}{x - x_0} - f'(x_0) \right| < -\frac{1}{2}f'(x_0)$$

if $|x - x_0| < \delta$. It follows that

$$\frac{f(x) - f(x_0)}{x - x_0} < \frac{1}{2}f'(x_0) < 0.$$

Thus $f(x) < f(x_0)$ if $x > x_0$, which is a contradiction. Hence, $f'(x_0) \geq 0$. A similar argument can be used to show that $f'(x_0) \leq 0$ when $f(x)$ is monotone decreasing. $\square$

Note that strict monotonicity on a set D does not necessarily imply that $f'(x) > 0$, or $f'(x) < 0$, for all x in D . For example, the function $f(x) = x^3$ is strictly monotone increasing for all x , but $f'(0) = 0$.

We recall from Theorem 3.5.1 that strict monotonicity of $f(x)$ is a sufficient condition for the existence of the inverse function f^{-1} . The next theorem shows that under certain conditions, f^{-1} is a differentiable function.

Theorem 4.2.4. Suppose that $f(x)$ is strictly monotone increasing (or decreasing) and continuous on an interval $[a, b]$. If $f'(x)$ exists and is different from zero at $x_0 \in (a, b)$, then the inverse function f^{-1} is differentiable at $y_0 = f(x_0)$ and its derivative is given by

$$\left. \frac{df^{-1}(y)}{dy} \right|_{y=y_0} = \frac{1}{f'(x_0)}.$$

Proof. By Theorem 3.5.2, $f^{-1}(y)$ exists and is continuous. Let $x_0 \in (a, b)$, and let $N_r(x_0) \subset (a, b)$ for some $r > 0$. Then, for any $x \in N_r(x_0)$,

$$\begin{aligned} \frac{f^{-1}(y) - f^{-1}(y_0)}{y - y_0} &= \frac{x - x_0}{f(x) - f(x_0)} \\ &= \frac{1}{[f(x) - f(x_0)]/(x - x_0)}, \end{aligned} \quad (4.5)$$

where $y = f(x)$. Now, since both f and f^{-1} are continuous, then $x \rightarrow x_0$ if and only if $y \rightarrow y_0$. By taking the limits of all sides in formula (4.5), we conclude that the derivative of f^{-1} at y_0 exists and is equal to

$$\left. \frac{df^{-1}(y)}{dy} \right|_{y=y_0} = \frac{1}{f'(x_0)}. \quad \square$$

The following theorem gives a more general version of the mean value theorem. It is due to Augustin-Louis Cauchy and has an important application in calculating certain limits, as will be seen later in this chapter.

Theorem 4.2.5 (Cauchy's Mean Value Theorem). If $f(x)$ and $g(x)$ are continuous on the closed interval $[a, b]$ and differentiable on the open interval (a, b) , then there exists a point c , $a < c < b$, such that

$$[f(b) - f(a)]g'(c) = [g(b) - g(a)]f'(c),$$

Proof. The proof is based on using Rolle's theorem in a manner similar to that of Theorem 4.2.2. Define the function $\psi(x)$ as

$$\psi(x) = [f(b) - f(x)][g(b) - g(a)] - [f(b) - f(a)][g(b) - g(x)].$$

This function is continuous on $[a, b]$ and is differentiable on (a, b) , since

$$\psi'(x) = -f'(x)[g(b) - g(a)] + g'(x)[f(b) - f(a)].$$

Furthermore, $\psi(a) = \psi(b) = 0$. Thus by Rolle's theorem, there exists a point c , $a < c < b$, such that $\psi'(c) = 0$, that is,

$$-f'(c)[g(b) - g(a)] + g'(c)[f(b) - f(a)] = 0. \quad (4.6)$$

In particular, if $g(b) - g(a) \neq 0$ and $f'(x)$ and $g'(x)$ do not vanish at the same point in (a, b) , then formula (4.6) can be written as

$$\frac{f'(c)}{g'(c)} = \frac{f(b) - f(a)}{g(b) - g(a)}. \quad \square$$

An immediate application of this theorem is a very popular method in calculating the limits of certain ratios of functions. This method is attributed to Guillaume Francois Marquis de l'Hospital (1661–1704) and is known as *l'Hospital's rule*. It deals with the limit of the ratio $f(x)/g(x)$ as $x \rightarrow a$ when both the numerator and the denominator tend simultaneously to zero or to infinity as $x \rightarrow a$. In either case, we have what is called an indeterminate ratio caused by having $0/0$ or ∞/∞ as $x \rightarrow a$.

Theorem 4.2.6 (l'Hospital's Rule). Let $f(x)$ and $g(x)$ be continuous on the closed interval $[a, b]$ and differentiable on the open interval (a, b) . Suppose that we have the following:

1. $g(x)$ and $g'(x)$ are not zero at any point inside (a, b) .
2. $\lim_{x \rightarrow a^+} f'(x)/g'(x)$ exists.
3. $f(x) \rightarrow 0$ and $g(x) \rightarrow 0$ as $x \rightarrow a^+$, or $f(x) \rightarrow \infty$ and $g(x) \rightarrow \infty$ as $x \rightarrow a^+$.

Then,

$$\lim_{x \rightarrow a^+} \frac{f(x)}{g(x)} = \lim_{x \rightarrow a^+} \frac{f'(x)}{g'(x)}.$$

Proof. For the sake of simplicity, we shall drop the $+$ sign from a^+ and simply write $x \rightarrow a$ when x approaches a from the right. Let us consider the following cases:

CASE 1. $f(x) \rightarrow 0$ and $g(x) \rightarrow 0$ as $x \rightarrow a$, where a is finite. Let $x \in (a, b)$. By applying Cauchy's mean value theorem on the interval $[a, x]$ we get

$$\frac{f(x)}{g(x)} = \frac{f(x) - f(a)}{g(x) - g(a)} = \frac{f'(c)}{g'(c)}.$$

where $a < c < x$. Note that $f(a) = g(a) = 0$, since $f(x)$ and $g(x)$ are continuous and their limits are equal to zero when $x \rightarrow a$. Now, as $x \rightarrow a, c \rightarrow a$; hence

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \lim_{c \rightarrow a} \frac{f'(c)}{g'(c)} = \lim_{x \rightarrow a} \frac{f'(x)}{g'(x)}.$$

CASE 2. $f(x) \rightarrow 0$ and $g(x) \rightarrow 0$ as $x \rightarrow \infty$. Let $z = 1/x$. As $x \rightarrow \infty, z \rightarrow 0$. Then

$$\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = \lim_{z \rightarrow 0} \frac{f_1(z)}{g_1(z)}, \quad (4.7)$$

where $f_1(z) = f(1/z)$ and $g_1(z) = g(1/z)$. These functions are continuous since $f(x)$ and $g(x)$ are, and $z \neq 0$ as $z \rightarrow 0$ (see Theorem 3.4.2). Here, we find it necessary to set $f_1(0) = g_1(0) = 0$ so that $f_1(z)$ and $g_1(z)$ will be continuous at $z = 0$, since their limits are equal to zero. This is equivalent to defining $f(x)$ and $g(x)$ to be zero at infinity in the extended real number system. Furthermore, by the chain rule of Theorem 4.1.3 we have

$$\begin{aligned} f'_1(z) &= f'(x) \left(-\frac{1}{z^2} \right), \\ g'_1(z) &= g'(x) \left(-\frac{1}{z^2} \right). \end{aligned}$$

If we now apply Case 1, we get

$$\begin{aligned} \lim_{z \rightarrow 0} \frac{f_1(z)}{g_1(z)} &= \lim_{z \rightarrow 0} \frac{f'_1(z)}{g'_1(z)} \\ &= \lim_{x \rightarrow \infty} \frac{f'(x)}{g'(x)}. \end{aligned} \quad (4.8)$$

From (4.7) and (4.8) we then conclude that

$$\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = \lim_{x \rightarrow \infty} \frac{f'(x)}{g'(x)}.$$

CASE 3. $f(x) \rightarrow \infty$ and $g(x) \rightarrow \infty$ as $x \rightarrow a$, where a is finite. Let $\lim_{x \rightarrow a} [f'(x)/g'(x)] = L$. Then for a given $\epsilon > 0$ there exists a $\delta > 0$ such that $a + \delta < b$ and

$$\left| \frac{f'(x)}{g'(x)} - L \right| < \epsilon, \quad (4.9)$$

if $a < x < a + \delta$. By applying Cauchy's mean value theorem on the interval $[x, a + \delta]$ we get

$$\frac{f(x) - f(a + \delta)}{g(x) - g(a + \delta)} = \frac{f'(d)}{g'(d)},$$

where $x < d < a + \delta$. From inequality (4.9) we then have

$$\left| \frac{f(x) - f(a + \delta)}{g(x) - g(a + \delta)} - L \right| < \epsilon$$

for all x such that $a < x < a + \delta$. It follows that

$$\begin{aligned} L &= \lim_{x \rightarrow a} \frac{f(x) - f(a + \delta)}{g(x) - g(a + \delta)} \\ &= \lim_{x \rightarrow a} \frac{f(x)}{g(x)} \lim_{x \rightarrow a} \frac{1 - f(a + \delta)/f(x)}{1 - g(a + \delta)/g(x)} \\ &= \lim_{x \rightarrow a} \frac{f(x)}{g(x)} \end{aligned}$$

since both $f(x)$ and $g(x)$ tend to ∞ as $x \rightarrow a$.

CASE 4. $f(x) \rightarrow \infty$ and $g(x) \rightarrow \infty$ as $x \rightarrow \infty$. This can be easily shown by using the techniques applied in Cases 2 and 3.

CASE 5. $\lim_{x \rightarrow a} f'(x)/g'(x) = \infty$, where a is finite or infinite. Let us consider the ratio $g(x)/f(x)$. We have that

$$\lim_{x \rightarrow a} \frac{g'(x)}{f'(x)} = 0.$$

Hence,

$$\lim_{x \rightarrow a} \frac{g(x)}{f(x)} = \lim_{x \rightarrow a} \frac{g'(x)}{f'(x)} = 0.$$

If A is any positive number, then there exists a $\delta > 0$ such that

$$\frac{g(x)}{f(x)} < \frac{1}{A},$$

if $a < x < a + \delta$. Thus for such values of x ,

$$\frac{f(x)}{g(x)} > A,$$

which implies that

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \infty.$$

When applying l'Hospital's rule, the ratio $f'(x)/g'(x)$ may assume the indeterminate form $0/0$ or ∞/∞ as $x \rightarrow a$. In this case, higher-order derivatives of $f(x)$ and $g(x)$, assuming such derivatives exist, will be needed. In general, if the first $n - 1$ ($n \geq 1$) derivatives of $f(x)$ and $g(x)$ tend simultaneously to zero or to ∞ as $x \rightarrow a$, and if the n th derivatives, $f^{(n)}(x)$ and $g^{(n)}(x)$, exist and satisfy the same conditions as those imposed on $f'(x)$ and $g'(x)$ in Theorem 4.2.6, then

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \lim_{x \rightarrow a} \frac{f^{(n)}(x)}{g^{(n)}(x)}. \quad \square$$

A Historical Note

According to Eves (1976, page 342), in 1696 the Marquis de l'Hospital assembled the lecture notes of his teacher Johann Bernoulli (1667–1748) into the world's first textbook on calculus. In this book, the so-called l'Hospital's rule is found. It is perhaps more accurate to refer to this rule as the Bernoulli-l'Hospital rule. Note that the name l'Hospital follows the old French spelling and the letter s is not to be pronounced. In modern French this name is spelled as l'Hôpital.

$$\text{EXAMPLE 4.2.1. } \lim_{x \rightarrow 0} \frac{\sin x}{x} = \lim_{x \rightarrow 0} \frac{\cos x}{1} = 1.$$

This is a well-known limit. It implies that $\sin x$ and x are asymptotically equal, that is, $\sin x \sim x$ as $x \rightarrow 0$ (see Section 3.3).

$$\text{EXAMPLE 4.2.2. } \lim_{x \rightarrow 0} \frac{1 - \cos x}{x^2} = \lim_{x \rightarrow 0} \frac{\sin x}{2x} = \lim_{x \rightarrow 0} \frac{\cos x}{2} = \frac{1}{2}.$$

We note here that l'Hospital's rule was applied twice before reaching the limit $\frac{1}{2}$.

EXAMPLE 4.2.3. $\lim_{x \rightarrow \infty} \frac{a^x}{x}$, where $a > 1$.

This is of the form ∞/∞ as $x \rightarrow \infty$. Since $a^x = e^{x \log a}$, then

$$\lim_{x \rightarrow \infty} \frac{a^x}{x} = \lim_{x \rightarrow \infty} \frac{e^{x \log a} (\log a)}{1} = \infty.$$

This is also a well-known limit. On the basis of this result it can be shown that (see Exercise 4.12) the following hold:

1. $\lim_{x \rightarrow \infty} \frac{a^x}{x^m} = \infty$, where $a > 1, m > 0$.
2. $\lim_{x \rightarrow \infty} \frac{\log x}{x^m} = 0$, where $m > 0$.

EXAMPLE 4.2.4. $\lim_{x \rightarrow 0^+} x^x$.

This is of the form 0^0 as $x \rightarrow 0^+$, which is indeterminate. It can be reduced to the form $0/0$ or ∞/∞ so that l'Hospital's rule can apply. To do so we write x^x as

$$x^x = e^{x \log x}.$$

However,

$$x \log x = \frac{\log x}{1/x},$$

which is of the form $-\infty/\infty$ as $x \rightarrow 0^+$. By l'Hospital's rule we then have

$$\begin{aligned} \lim_{x \rightarrow 0^+} (x \log x) &= \lim_{x \rightarrow 0^+} \frac{1/x}{-1/x^2} \\ &= \lim_{x \rightarrow 0^+} (-x) \\ &= 0. \end{aligned}$$

It follows that

$$\begin{aligned} \lim_{x \rightarrow 0^+} x^x &= \lim_{x \rightarrow 0^+} e^{x \log x} \\ &= \exp \left[\lim_{x \rightarrow 0^+} (x \log x) \right] \\ &= 1. \end{aligned}$$

EXAMPLE 4.2.5. $\lim_{x \rightarrow \infty} x \log \left(\frac{x+1}{x-1} \right)$.

This is of the form $\infty \times 0$ as $x \rightarrow \infty$, which is indeterminate. But

$$x \log \left(\frac{x+1}{x-1} \right) = \frac{\log \left(\frac{x+1}{x-1} \right)}{1/x}$$

is of the form $0/0$ as $x \rightarrow \infty$. Hence,

$$\begin{aligned}\lim_{x \rightarrow \infty} x \log \left(\frac{x+1}{x-1} \right) &= \lim_{x \rightarrow \infty} \frac{\frac{-2}{(x-1)(x+1)}}{-1/x^2} \\ &= \lim_{x \rightarrow \infty} \frac{2}{(1-1/x)(1+1/x)} \\ &= 2.\end{aligned}$$

We can see from the foregoing examples that the use of l'Hospital's rule can facilitate the process of finding the limit of the ratio $f(x)/g(x)$ as $x \rightarrow a$. In many cases, it is easier to work with $f'(x)/g'(x)$ than with the original ratio. Many other indeterminate forms can also be resolved by l'Hospital's rule by first reducing them to the form $0/0$ or ∞/∞ as was shown in Examples 4.2.4 and 4.2.5.

It is important here to remember that the application of l'Hospital's rule requires that the limit of $f'(x)/g'(x)$ exist as a finite number or be equal to infinity in the extended real number system as $x \rightarrow a$. If this is not the case, then it does not necessarily follow that the limit of $f(x)/g(x)$ does not exist. For example, consider $f(x) = x^2 \sin(1/x)$ and $g(x) = x$. Here, $f(x)/g(x)$ tends to zero as $x \rightarrow 0$, as was seen earlier in this chapter. However, the ratio

$$\frac{f'(x)}{g'(x)} = 2x \sin \frac{1}{x} - \cos \frac{1}{x}$$

has no limit as $x \rightarrow 0$, since it oscillates inside a small neighborhood of the origin.

4.3. TAYLOR'S THEOREM

This theorem is also known as the general mean value theorem, since it is considered as an extension of the mean value theorem. It was formulated by the English mathematician Brook Taylor (1685–1731) in 1712 and has since become a very important theorem in calculus. Taylor used his theorem to expand functions into infinite series. However, full recognition of the importance of Taylor's expansion was not realized until 1755, when Leonhard Euler (1707–1783) applied it in his differential calculus, and still later, when Joseph Louis Lagrange (1736–1813) used it as the foundation of his theory of functions.

Theorem 4.3.1 (Taylor's Theorem). If the $(n-1)$ st ($n \geq 1$) derivative of $f(x)$, namely $f^{(n-1)}(x)$, is continuous on the closed interval $[a, b]$ and the n th derivative $f^{(n)}(x)$ exists on the open interval (a, b) , then for each $x \in [a, b]$

we have

$$\begin{aligned} f(x) = & f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2!}f''(a) \\ & + \cdots + \frac{(x-a)^{n-1}}{(n-1)!}f^{(n-1)}(a) + \frac{(x-a)^n}{n!}f^{(n)}(\xi), \end{aligned}$$

where $a < \xi < x$.

Proof. The method to prove this theorem is very similar to the one used for Theorem 4.2.2. For a fixed x in $[a, b]$ let the function $\psi_n(t)$ be defined as

$$\psi_n(t) = g_n(t) - \left(\frac{x-t}{x-a} \right)^n g_n(a),$$

where $a \leq t \leq b$ and

$$\begin{aligned} g_n(t) = & f(x) - f(t) - (x-t)f'(t) \\ & - \frac{(x-t)^2}{2!}f''(t) - \cdots - \frac{(x-t)^{n-1}}{(n-1)!}f^{(n-1)}(t). \end{aligned} \quad (4.10)$$

The function $\psi_n(t)$ has the following properties:

1. $\psi_n(a) = 0$ and $\psi_n(x) = 0$.
2. $\psi_n(t)$ is a continuous function of t on $[a, b]$.
3. The derivative of $\psi_n(t)$ with respect to t exists on (a, b) . This derivative is equal to

$$\begin{aligned} \psi'_n(t) = & g'_n(t) + \frac{n(x-t)^{n-1}}{(x-a)^n}g_n(a) \\ = & -f'(t) + f'(t) - (x-t)f''(t) + (x-t)f''(t) \\ & - \cdots + \frac{(x-t)^{n-2}}{(n-2)!}f^{(n-1)}(t) \\ & - \frac{(x-t)^{n-1}}{(n-1)!}f^{(n)}(t) + \frac{n(x-t)^{n-1}}{(x-a)^n}g_n(a) \\ = & - \frac{(x-t)^{n-1}}{(n-1)!}f^{(n)}(t) + \frac{n(x-t)^{n-1}}{(x-a)^n}g_n(a). \end{aligned}$$

By applying Rolle's theorem to $\psi_n(t)$ on the interval $[a, x]$ we can assert that there exists a value ξ , $a < \xi < x$, such that $\psi'_n(\xi) = 0$, that is,

$$-\frac{(x-\xi)^{n-1}}{(n-1)!}f^{(n)}(\xi) + \frac{n(x-\xi)^{n-1}}{(x-a)^n}g_n(a) = 0,$$

or

$$g_n(a) = \frac{(x-a)^n}{n!}f^{(n)}(\xi). \quad (4.11)$$

Using formula (4.10) in (4.11), we finally get

$$\begin{aligned} f(x) = & f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2!}f''(a) \\ & + \cdots + \frac{(x-a)^{n-1}}{(n-1)!}f^{(n-1)}(a) + \frac{(x-a)^n}{n!}f^{(n)}(\xi). \end{aligned} \quad (4.12)$$

This is known as *Taylor's formula*. It can also be expressed as

$$\begin{aligned} f(a+h) = & f(a) + hf'(a) + \frac{h^2}{2!}f''(a) \\ & + \cdots + \frac{h^{n-1}}{(n-1)!}f^{(n-1)}(a) + \frac{h^n}{n!}f^{(n)}(a + \theta_n h), \end{aligned} \quad (4.13)$$

where $h = x - a$ and $0 < \theta_n < 1$. □

In particular, if $f(x)$ has derivatives of all orders in some neighborhood $N_r(a)$ of the point a , formula (4.12) can provide a series expansion of $f(x)$ for $x \in N_r(a)$ as $n \rightarrow \infty$. The last term in formula (4.12), or formula (4.13), is called the remainder of Taylor's series and is denoted by R_n . Thus,

$$\begin{aligned} R_n = & \frac{(x-a)^n}{n!}f^{(n)}(\xi) \\ = & \frac{h^n}{n!}f^{(n)}(a + \theta_n h). \end{aligned}$$

If $R_n \rightarrow 0$ as $n \rightarrow \infty$, then

$$f(x) = f(a) + \sum_{n=1}^{\infty} \frac{(x-a)^n}{n!} f^{(n)}(a), \quad (4.14)$$

or

$$f(a+h) = f(a) + \sum_{n=1}^{\infty} \frac{h^n}{n!} f^{(n)}(a). \quad (4.15)$$

This results in what is known as *Taylor's series*. Thus the validity of Taylor's series is contingent on having $R_n \rightarrow 0$ as $n \rightarrow \infty$, and on having derivatives of all orders for $f(x)$ in $N_r(a)$. The existence of these derivatives alone is not sufficient to guarantee a valid expansion.

A special form of Taylor's series is *Maclaurin's series*, which results when $a = 0$. Formula (4.14) then reduces to

$$f(x) = f(0) + \sum_{n=1}^{\infty} \frac{x^n}{n!} f^{(n)}(0). \quad (4.16)$$

In this case, the remainder takes the form

$$R_n = \frac{x^n}{n!} f^{(n)}(\theta_n x). \quad (4.17)$$

The sum of the first n terms in Maclaurin's series provides an approximation for the value of $f(x)$. The size of the remainder determines how close the sum is to $f(x)$. Since the remainder depends on θ_n , which lies in the interval $(0, 1)$, an upper bound on R_n that is free of θ_n will therefore be needed to assess the accuracy of the approximation. For example, let us consider the function $f(x) = \cos x$. In this case,

$$f^{(n)}(x) = \cos\left(x + \frac{n\pi}{2}\right), \quad n = 1, 2, \dots,$$

and

$$\begin{aligned} f^{(n)}(0) &= \cos\left(\frac{n\pi}{2}\right) \\ &= \begin{cases} 0, & n \text{ odd,} \\ (-1)^{n/2}, & n \text{ even.} \end{cases} \end{aligned}$$

Formula (4.16) becomes

$$\begin{aligned}\cos x &= 1 + \sum_{n=1}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!} \\ &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots + (-1)^n \frac{x^{2n}}{(2n)!} + R_{2n+1},\end{aligned}$$

where from formula (4.17) R_{2n+1} is

$$R_{2n+1} = \frac{x^{2n+1}}{(2n+1)!} \cos \left[\theta_{2n+1} x + \frac{(2n+1)\pi}{2} \right].$$

An upper bound on $|R_{2n+1}|$ is then given by

$$|R_{2n+1}| \leq \frac{|x|^{2n+1}}{(2n+1)!}.$$

Therefore, the error of approximating $\cos x$ with the sum

$$s_{2n} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots + (-1)^n \frac{x^{2n}}{(2n)!}$$

does not exceed $|x|^{2n+1}/(2n+1)!$, where x is measured in radians. For example, if $x = \pi/3$ and $n = 3$, the sum

$$s_6 = 1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} = 0.49996$$

approximates $\cos(\pi/3)$ with an error not exceeding

$$\frac{|x|^7}{7!} = 0.00027.$$

The true value of $\cos(\pi/3)$ is 0.5.

4.4. MAXIMA AND MINIMA OF A FUNCTION

In this section we consider the problem of finding the extreme values of a function $y = f(x)$ whose derivative $f'(x)$ exists in any open set inside its domain of definition.

Definition 4.4.1. A function $f: D \rightarrow R$ has a local (or relative) maximum at a point $x_0 \in D$ if there exists a $\delta > 0$ such that $f(x) \leq f(x_0)$ for all $x \in N_\delta(x_0) \cap D$. The function f has a local (or relative) minimum at x_0 if $f(x) \geq f(x_0)$ for all $x \in N_\delta(x_0) \cap D$. Local maxima and minima are referred to as local optima (or local extrema). $\square$

Definition 4.4.2. A function $f: D \rightarrow R$ has an absolute maximum (minimum) over D if there exists a point $x^* \in D$ such that $f(x) \leq f(x^*)$ [$f(x) \geq f(x^*)$] for all $x \in D$. Absolute maxima and minima are called absolute optima (or extrema). $\square$

The determination of local optima of $f(x)$ can be greatly facilitated if $f(x)$ is differentiable.

Theorem 4.4.1. Let $f(x)$ be differentiable on the open interval (a, b) . If $f(x)$ has a local maximum, or a local minimum, at a point x_0 in (a, b) , then $f'(x_0) = 0$.

Proof. Suppose that $f(x)$ has a local maximum at x_0 . Then, $f(x) \leq f(x_0)$ for all x in some neighborhood $N_\delta(x_0) \subset (a, b)$. It follows that

$$\frac{f(x) - f(x_0)}{x - x_0} \begin{cases} \leq 0 & \text{if } x > x_0, \\ \geq 0 & \text{if } x < x_0, \end{cases} \quad (4.18)$$

for all x in $N_\delta(x_0)$. As $x \rightarrow x_0^+$, the ratio in (4.18) will have a nonpositive limit, and if $x \rightarrow x_0^-$, the ratio will have a nonnegative limit. Since $f'(x_0)$ exists, these two limits must be equal and equal to $f'(x_0)$ as $x \rightarrow x_0$. We therefore conclude that $f'(x_0) = 0$. The proof when $f(x)$ has a local minimum is similar. $\square$

It is important here to note that $f'(x_0) = 0$ is a necessary condition for a differentiable function to have a local optimum at x_0 . It is not, however, a sufficient condition. That is, if $f'(x_0) = 0$, then it is not necessarily true that x_0 is a point of local optimum. For example, the function $f(x) = x^3$ has a zero derivative at the origin, but $f(x)$ does not have a local optimum there (why not?). In general, a value x_0 for which $f'(x_0) = 0$ is called a stationary value for the function. Thus a stationary value does not necessarily correspond to a local optimum.

We should also note that Theorem 4.4.1 assumes that $f(x)$ is differentiable in a neighborhood of x_0 . If this condition is not fulfilled, the theorem ceases to be true. The existence of $f'(x_0)$ is not prerequisite for $f(x)$ to have a local optimum at x_0 . In fact, $f(x)$ can have a local optimum at x_0 even if $f'(x_0)$ does not exist. For example, $f(x) = |x|$ has a local minimum at $x = 0$, but $f'(0)$ does not exist.

We recall from Corollary 3.4.1 that if $f(x)$ is continuous on $[a, b]$, then it must achieve its absolute optima at some points inside $[a, b]$. These points can be interior points, that is, points that belong to the open interval (a, b) , or they can be end (boundary) points. In particular, if $f'(x)$ exists on (a, b) , to determine the locations of the absolute optima we must solve the equation $f'(x) = 0$ and then compare the values of $f(x)$ at the roots of this equation with $f(a)$ and $f(b)$. The largest (smallest) of these values is the absolute maximum (minimum). In the event $f'(x) \neq 0$ on (a, b) , then $f(x)$ must achieve its absolute optimum at an end point.

4.4.1. A Sufficient Condition for a Local Optimum

We shall make use of Taylor's expansion to come up with a sufficient condition for $f(x)$ to have a local optimum at $x = x_0$.

Suppose that $f(x)$ has n derivatives in a neighborhood $N_\delta(x_0)$ such that $f'(x_0) = f''(x_0) = \cdots = f^{(n-1)}(x_0) = 0$, but $f^{(n)}(x_0) \neq 0$. Then by Taylor's theorem we have

$$f(x) = f(x_0) + \frac{h^n}{n!} f^{(n)}(x_0 + \theta_n h)$$

for any x in $N_\delta(x_0)$, where $h = x - x_0$ and $0 < \theta_n < 1$. Furthermore, if we assume that $f^{(n)}(x)$ is continuous at x_0 , then

$$f^{(n)}(x_0 + \theta_n h) = f^{(n)}(x_0) + o(1),$$

where, as we recall from Section 3.3, $o(1) \rightarrow 0$ as $h \rightarrow 0$. We can therefore write

$$f(x) - f(x_0) = \frac{h^n}{n!} f^{(n)}(x_0) + o(h^n). \quad (4.19)$$

In order for $f(x)$ to have a local optimum at x_0 , $f(x) - f(x_0)$ must have the same sign (positive or negative) for small values of h inside a neighborhood of 0. But, from (4.19), the sign of $f(x) - f(x_0)$ is determined by the sign of $h^n f^{(n)}(x_0)$. We can then conclude that if n is even, then a local optimum is achieved at x_0 . In this case, a local maximum occurs at x_0 if $f^{(n)}(x_0) < 0$, whereas $f^{(n)}(x_0) > 0$ indicates that x_0 is a point of local minimum. If, however, n is odd, then x_0 is not a point of local optimum, since $f(x) - f(x_0)$ changes sign around x_0 . In this case, the point on the graph of $y = f(x)$ whose abscissa is x_0 is called a *saddle point*.

In particular, if $f'(x_0) = 0$ and $f''(x_0) \neq 0$, then x_0 is a point of local optimum. When $f''(x_0) < 0$, $f(x)$ has a local maximum at x_0 , and when $f''(x_0) > 0$, $f(x)$ has a local minimum at x_0 .

EXAMPLE 4.4.1. Let $f(x) = 2x^3 - 3x^2 - 12x + 6$. Then $f'(x) = 6x^2 - 6x - 12 = 0$ at $x = -1, 2$, and

$$f''(x) = \begin{cases} 12x - 6 = -18 & \text{at } x = -1, \\ 18 & \text{at } x = 2. \end{cases}$$

We have then a local maximum at $x = -1$ and a local minimum at $x = 2$.

EXAMPLE 4.4.2. $f(x) = x^4 - 1$. In this case,

$$f'(x) = 4x^3 = 0 \quad \text{at } x = 0,$$

$$f''(x) = 12x^2 = 0 \quad \text{at } x = 0,$$

$$f'''(x) = 24x = 0 \quad \text{at } x = 0,$$

$$f^{(4)}(x) = 24.$$

Then, $x = 0$ is a point of local minimum.

EXAMPLE 4.4.3. Consider $f(x) = (x + 5)^2(x^3 - 10)$. We have

$$f'(x) = 5(x + 5)(x - 1)(x + 2)^2,$$

$$f''(x) = 10(x + 2)(2x^2 + 8x - 1),$$

$$f'''(x) = 10(6x^2 + 24x + 15).$$

Here, $f'(x) = 0$ at $x = -5, -2$, and 1 . At $x = -5$ there is a local maximum, since $f''(-5) = -270 < 0$. At $x = 1$ we have a local minimum, since $f''(1) = 270 > 0$. However, at $x = -2$ a saddle point occurs, since $f''(-2) = 0$ and $f'''(-2) = -90 \neq 0$.

EXAMPLE 4.4.4. $f(x) = (2x + 1)/(x + 4)$, $0 \leq x \leq 5$. Then

$$f'(x) = 7/(x + 4)^2.$$

In this case, $f'(x)$ does not vanish anywhere in $(0, 5)$. Thus $f(x)$ has no local maxima or local minima in that open interval. Being continuous on $[0, 5]$, $f(x)$ must achieve its absolute optima at the end points. Since $f'(x) > 0$, $f(x)$ is strictly monotone increasing on $[0, 5]$ by Corollary 4.2.1. Its absolute minimum and absolute maximum are therefore attained at $x = 0$ and $x = 5$, respectively.

4.5. APPLICATIONS IN STATISTICS

Differential calculus has many applications in statistics. Let us consider some of these applications.

4.5.1. Functions of Random Variables

Let Y be a continuous random variable whose cumulative distribution function is $F(y) = P(Y \leq y)$. If $F(y)$ is differentiable for all y , then its derivative $F'(y)$ is called the density function of Y and is denoted by $f(y)$. Continuous random variables for which $f(y)$ exists are said to be absolutely continuous.

Let Y be an absolutely continuous random variable, and let W be another random variable which can be expressed as a function of Y of the form $W = \psi(Y)$. Suppose that this function is strictly monotone and differentiable over its domain. By Theorem 3.5.1, ψ has a unique inverse ψ^{-1} , which is also differentiable by Theorem 4.2.4. Let $G(w)$ denote the cumulative distribution function of W .

If ψ is strictly monotone increasing, then

$$G(w) = P(W \leq w) = P[Y \leq \psi^{-1}(w)] = F[\psi^{-1}(w)].$$

If it is strictly monotone decreasing, then

$$G(w) = P(W \leq w) = P[Y \geq \psi^{-1}(w)] = 1 - F[\psi^{-1}(w)].$$

By differentiating $G(w)$ using the chain rule we obtain the density function $g(w)$ for W , namely,

$$\begin{aligned} g(w) &= \frac{dF[\psi^{-1}(w)]}{d\psi^{-1}(w)} \frac{d\psi^{-1}(w)}{dw} \\ &= f[\psi^{-1}(w)] \frac{d\psi^{-1}(w)}{dw} \end{aligned} \quad (4.20)$$

if ψ is strictly monotone increasing, and

$$\begin{aligned} g(w) &= - \frac{dF[\psi^{-1}(w)]}{d\psi^{-1}(w)} \frac{d\psi^{-1}(w)}{dw} \\ &= -f[\psi^{-1}(w)] \frac{d\psi^{-1}(w)}{dw} \end{aligned} \quad (4.21)$$

if ψ is strictly monotone decreasing. By combining (4.20) and (4.21) we obtain

$$g(w) = f[\psi^{-1}(w)] \left| \frac{d\psi^{-1}(w)}{dw} \right|. \quad (4.22)$$

For example, suppose that Y has the uniform distribution $U(0, 1)$ whose density function is

$$f(y) = \begin{cases} 1, & 0 < y < 1, \\ 0 & \text{elsewhere.} \end{cases}$$

Let $W = -\log Y$. Using formula (4.22), the density function of W is given by

$$g(w) = \begin{cases} e^{-w}, & 0 < w < \infty, \\ 0 & \text{elsewhere.} \end{cases}$$

The Mean and Variance of $W = \psi(Y)$

The mean and variance of the random variable W can be obtained by using its density function:

$$\begin{aligned} E(W) &= \int_{-\infty}^{\infty} wg(w) dw, \\ \text{Var}(W) &= E[W - E(W)]^2 \\ &= \int_{-\infty}^{\infty} [w - E(W)]^2 g(w) dw. \end{aligned}$$

In some cases, however, the exact distribution of Y may not be known, or $g(w)$ may be a complicated function to integrate. In such cases, approximate expressions for the mean and variance of W can be obtained by applying Taylor's expansion around the mean of Y , μ . If we assume that $\psi''(y)$ exists, then

$$\psi(y) = \psi(\mu) + (y - \mu)\psi'(\mu) + o(y - \mu).$$

If $o(y - \mu)$ is small enough, first-order approximations of $E(W)$ and $\text{Var}(W)$ can be obtained, namely,

$$E(W) \approx \psi(\mu), \quad \text{since } E(Y - \mu) = 0; \quad \text{Var}(W) \approx \sigma^2 [\psi'(\mu)]^2, \quad (4.23)$$

where $\sigma^2 = \text{Var}(Y)$, and the symbol $\approx$ denotes approximate equality. If $o(y - \mu)$ is not small enough, then higher-order approximations can be utilized provided that certain derivatives of $\psi(y)$ exist. For example, if $\psi'''(y)$ exists, then

$$\psi(y) = \psi(\mu) + (y - \mu)\psi'(\mu) + \frac{1}{2}(y - \mu)^2\psi''(\mu) + o[(y - \mu)^2].$$

In this case, if $o[(y - \mu)^2]$ is small enough, then second-order approximations can be obtained for $E(W)$ and $\text{Var}(W)$ of the form

$$\begin{aligned} E(W) &\approx \psi(\mu) + \frac{1}{2}\sigma^2\psi''(\mu), \quad \text{since } E(Y - \mu)^2 = \sigma^2, \\ \text{Var}(W) &\approx E\{Q(Y) - E[Q(Y)]\}^2, \end{aligned}$$

where

$$Q(Y) = \psi(\mu) + (Y - \mu)\psi'(\mu) + \frac{1}{2}(Y - \mu)^2\psi''(\mu).$$

Thus,

$$\begin{aligned}\text{Var}(W) \approx \sigma^2[\psi'(\mu)]^2 + \frac{1}{4}[\psi''(\mu)]^2 \text{Var}[(Y - \mu)^2] \\ + \psi'(\mu)\psi''(\mu)E[(Y - \mu)^3].\end{aligned}$$

Variance Stabilizing Transformations

One of the basic assumptions of regression and analysis of variance is the constancy of the variance σ^2 of a response variable Y on which experimental data are obtained. This assumption is often referred to as the assumption of homoscedasticity. There are situations, however, in which σ^2 is not constant for all the data. When this happens, Y is said to be heteroscedastic. Heteroscedasticity can cause problems and difficulties in connection with the statistical analysis of the data (for a survey of the problems of heteroscedasticity, see Judge et al., 1980).

Some situations that lead to heteroscedasticity are (see Wetherill et al., 1986, page 200):

- i. *The Use of Averaged Data.* In many experimental situations, the data used in a regression program consist of averages of samples that are different in size. This happens sometimes in survey analysis.
- ii. *Variances Depending on the Explanatory Variables.* The variance of an observation can sometimes depend on the explanatory (or input) variables in the hypothesized model, as is the case with some econometric models. For example, if the response variable is household expenditure and one explanatory variable is household income, then the variance of the observations may be a function of household income.
- iii. *Variances Depending on the Mean Response.* The response variable Y may have a distribution whose variance is a function of its mean, that is, $\sigma^2 = h(\mu)$, where μ is the mean of Y . The Poisson distribution, for example, has the property that $\sigma^2 = \mu$. Thus as μ changes (as a function of some explanatory variables), then so will σ^2 . The following example illustrates this situation (see Chatterjee and Price, 1977, page 39): Let Y be the number of accidents, and x be the speed of operating a lathe in a machine shop. Suppose that a linear relationship is assumed between Y and x of the form

$$Y = \beta_0 + \beta_1 x + \epsilon,$$

where ϵ is a random error with a zero mean. Here, Y has the Poisson distribution with mean $\mu = \beta_0 + \beta_1 x$. The variance of Y , being equal to μ , will not be constant, since it depends on x .

Heteroscedasticity due to dependence on the mean response can be removed, or at least reduced, by a suitable transformation of the response variable Y . So let us suppose that $\sigma^2 = h(\mu)$. Let $W = \psi(Y)$. We need to find a proper transformation ψ that causes W to have almost the constant variance property. If this can be accomplished, then ψ is referred to as a variance stabilizing transformation.

If the first-order approximation of $\text{Var}(W)$ by Taylor's expansion is adequate, then by formula (4.23) we can select ψ so that

$$h(\mu) [\psi'(\mu)]^2 = c, \quad (4.24)$$

where c is a constant. Without loss of generality, let $c = 1$. A solution of (4.24) is given by

$$\psi(\mu) = \int \frac{d\mu}{\sqrt{h(\mu)}}.$$

Thus if $W = \psi(Y)$, then $\text{Var}(W)$ will have a variance approximately equal to one. For example, if $h(\mu) = \mu$, as is the case with the Poisson distribution, then

$$\psi(\mu) = \int \frac{d\mu}{\sqrt{\mu}} = 2\sqrt{\mu}.$$

Hence, $W = 2\sqrt{Y}$ will have a variance approximately equal to one. (In this case, it is more common to use the transformation $W = \sqrt{Y}$ which has a variance approximately equal to 0.25). Thus in the earlier example regarding the relationship between the number of accidents and the speed of operating a lathe, we need to regress $\sqrt{Y}$ against x in order to ensure approximate homoscedasticity.

The relationship (if any) between σ^2 and μ may be determined by theoretical considerations based on a knowledge of the type of data used—for example, Poisson data. In practice, however, such knowledge may not be known a priori. In this case, the appropriate transformation is selected empirically on the basis of residual analysis of the data. See, for example, Box and Draper (1987, Chapter 8), Montgomery and Peck (1982, Chapter 3). If possible, a transformation is selected to correct nonnormality (if the original data are not believed to be normally distributed) as well as heteroscedasticity. In this respect, a useful family of transformations introduced by Box and Cox (1964) can be used. These authors considered the power family of transformations defined by

$$\psi(Y) = \begin{cases} (Y^\lambda - 1)/\lambda, & \lambda \neq 0, \\ \log Y, & \lambda = 0. \end{cases}$$

This family may only be applied when Y has positive values. Furthermore, since by l'Hospital's rule

$$\lim_{\lambda \rightarrow 0} \frac{Y^\lambda - 1}{\lambda} = \lim_{\lambda \rightarrow 0} Y^\lambda \log Y = \log Y,$$

the Box-Cox transformation is a continuous function of λ . An estimate of λ can be obtained from the data using the method of maximum likelihood (see Montgomery and Peck, 1982, Section 3.7.1; Box and Draper, 1987, Section 8.4).

Asymptotic Distributions

The asymptotic distributions of functions of random variables are of special interest in statistical limit theory. By definition, a sequence of random variables $\{Y_n\}_{n=1}^\infty$ converges in distribution to the random variable Y if

$$\lim_{n \rightarrow \infty} F_n(y) = F(y)$$

at each point y where $F(y)$ is continuous, where $F_n(y)$ is the cumulative distribution function of Y_n ($n = 1, 2, \dots$) and $F(y)$ is the cumulative distribution function of Y (see Section 5.3 concerning sequences of functions). This form of convergence is denoted by writing

$$Y_n \xrightarrow{d} Y.$$

An illustration of convergence in distribution is provided by the well-known *central limit theorem*. It states that if $\{Y_n\}_{n=1}^\infty$ is a sequence of independent and identically distributed random variables with common mean and variance, μ and σ^2 , respectively, that are both finite, and if $\bar{Y}_n = \sum_{i=1}^n Y_i/n$ is the sample mean of a sample size n , then as $n \rightarrow \infty$,

$$\frac{\bar{Y}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} Z,$$

where Z has the standard normal distribution $N(0, 1)$.

An extension of the central limit theorem that includes functions of random variables is given by the following theorem:

Theorem 4.5.1. Let $\{Y_n\}_{n=1}^\infty$ be a sequence of independent and identically distributed random variables with mean μ and variance σ^2 (both finite), and let $\bar{Y}_n$ be the sample mean of a sample of size n . If $\psi(y)$ is a function whose derivative $\psi'(y)$ exists and is continuous in a neighborhood of μ such that

$\psi'(\mu) \neq 0$, then as $n \rightarrow \infty$,

$$\frac{\psi(\bar{Y}_n) - \psi(\mu)}{\sigma|\psi'(\mu)|/\sqrt{n}} \xrightarrow{d} Z.$$

Proof. See Wilks (1962, page 259). $\square$

On the basis of Theorem 4.5.1 we can assert that when n is large enough, $\psi(\bar{Y}_n)$ is approximately distributed as a normal variate with a mean $\psi(\mu)$ and a standard deviation $(\sigma/\sqrt{n})|\psi'(\mu)|$. For example, if $\psi(y) = y^2$, then as $n \rightarrow \infty$,

$$\frac{\bar{Y}_n^2 - \mu^2}{2|\mu|\sigma/\sqrt{n}} \xrightarrow{d} Z.$$

4.5.2. Approximating Response Functions

Perhaps the most prevalent use of Taylor's expansion in statistics is in the area of linear models. Let Y denote a response variable, such as the yield of a product, whose mean $\mu(x)$ is believed to depend on an explanatory (or input) variable x such as temperature or pressure. The true relationship between μ and x is usually unknown. However, if $\mu(x)$ is considered to have derivatives of all orders, then it is possible to approximate its values by using low-order terms of a Taylor's series over a limited range of interest. In this case, $\mu(x)$ can be represented approximately by a polynomial of degree d (≥ 1) of the form

$$\mu(x) = \beta_0 + \sum_{j=1}^d \beta_j x^j,$$

where $\beta_0, \beta_1, \dots, \beta_d$ are unknown parameters. Estimates of these parameters are obtained by running n ($\geq d+1$) experiments in which n observations, $y_1, y_2, \dots, y_n$, on Y are obtained for specified values of x . This leads us to the linear model

$$y_i = \beta_0 + \sum_{j=1}^d \beta_j x_i^j + \epsilon_i, \quad i = 1, 2, \dots, n, \quad (4.25)$$

where ϵ_i is a random error. The method of least squares can then be used to estimate the unknown parameters in (4.25). The adequacy of model (4.25) to represent the true mean response $\mu(x)$ can be checked using the given data provided that replicated observations are available at some points inside the region of interest. For more details concerning the adequacy of fit of linear models and the method of least squares, see, for example, Box and Draper (1987, Chapters 2 and 3) and Khuri and Cornell (1996, Chapter 2).

4.5.3. The Poisson Process

A random phenomenon that arises through a process which continues in time (or space) in a manner controlled by the laws of probability is called a stochastic process. A particular example of such a process is the Poisson process, which is associated with the number of events that take place over a period of time—for example, the arrival of customers at a service counter, or the arrival of α -rays, emitted from a radioactive source, at a Geiger counter.

Define $p_n(t)$ as the probability of n arrivals during a time interval of length t . For a Poisson process, the following postulates are assumed to hold:

1. The probability of exactly one arrival during a small time interval of length h is approximately proportional to h , that is,

$$p_1(h) = \lambda h + o(h)$$

as $h \rightarrow 0$, where λ is a constant.

2. The probability of more than one arrival during a small time interval of length h is negligible, that is,

$$\sum_{n>1} p_n(h) = o(h)$$

as $h \rightarrow 0$.

3. The probability of an arrival occurring during a small time interval $(t, t+h)$ does not depend on what happened prior to t . This means that the events defined according to the number of arrivals occurring during nonoverlapping time intervals are independent.

On the basis of the above postulates, an expression for $p_n(t)$ can be found as follows: For $n \geq 1$ and for small h we have approximately

$$\begin{aligned} p_n(t+h) &= p_n(t)p_0(h) + p_{n-1}(t)p_1(h) \\ &= p_n(t)[1 - \lambda h + o(h)] + p_{n-1}(t)[\lambda h + o(h)], \end{aligned} \quad (4.26)$$

since the probability of no arrivals during the time interval $(t, t+h)$ is approximately equal to $1 - p_1(h)$. For $n = 0$ we have

$$\begin{aligned} p_0(t+h) &= p_0(t)p_0(h) \\ &= p_0(t)[1 - \lambda h + o(h)]. \end{aligned} \quad (4.27)$$

From (4.26) and (4.27) we then get for $n \geq 1$,

$$\frac{p_n(t+h) - p_n(t)}{h} = p_n(t) \left[-\lambda + \frac{o(h)}{h} \right] + p_{n-1}(t) \left[\lambda + \frac{o(h)}{h} \right],$$

and for $n = 0$,

$$\frac{p_0(t+h) - p_0(t)}{h} = p_0(t) \left[-\lambda + \frac{o(h)}{h} \right].$$

By taking the limit as $h \rightarrow 0$ we obtain the derivatives

$$p'_n(t) = -\lambda p_n(t) + \lambda p_{n-1}(t), \quad n \geq 1 \quad (4.28)$$

$$p'_0(t) = -\lambda p_0(t). \quad (4.29)$$

From (4.29) the solution for $p_0(t)$ is given by

$$p_0(t) = e^{-\lambda t}, \quad (4.30)$$

since $p_0(t) = 1$ when $t = 0$ (that is, initially there were no arrivals). By substituting (4.30) in (4.28) when $n = 1$ we get

$$p'_1(t) = -\lambda p_1(t) + \lambda e^{-\lambda t}. \quad (4.31)$$

If we now multiply the two sides of (4.31) by $e^{\lambda t}$ we obtain

$$e^{\lambda t} p'_1(t) + \lambda p_1(t) e^{\lambda t} = \lambda,$$

or

$$\left[e^{\lambda t} p_1(t) \right]' = \lambda.$$

Hence,

$$e^{\lambda t} p_1(t) = \lambda t + c,$$

where c is a constant. This constant must be equal to zero, since $p_1(0) = 0$. We then have

$$p_1(t) = \lambda t e^{-\lambda t}.$$

By continuing in this process and using equation (4.28) we can find $p_2(t)$, then $p_3(t), \dots$, etc. In general, it can be shown that

$$p_n(t) = \frac{e^{-\lambda t} (\lambda t)^n}{n!}, \quad n = 0, 1, 2, \dots \quad (4.32)$$

In particular, if $t = 1$, then formula (4.32) gives the probability of n arrivals during one unit of time, namely,

$$p_n(1) = \frac{e^{-\lambda} \lambda^n}{n!}, \quad n = 0, 1, \dots$$

This gives the probability mass function of a Poisson random variable with mean λ .

4.5.4. Minimizing the Sum of Absolute Deviations

Consider a data set consisting of n observations $y_1, y_2, \dots, y_n$. For an arbitrary real number a , let $D(a)$ denote the sum of absolute deviations of the data from a , that is,

$$D(a) = \sum_{i=1}^n |y_i - a|.$$

For a given a , $D(a)$ represents a measure of spread, or variation, for the data set. Since the value of $D(a)$ varies with a , it may be of interest to determine its minimum. We now show that $D(a)$ is minimized when $a = \mu^*$, where μ^* denotes the median of the data set. By definition, μ^* is a value that falls in the middle when the observations are arranged in order of magnitude. It is a measure of location like the mean. If we write $y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(n)}$ for the ordered y_i 's, then when n is odd, μ^* is the unique value $y_{(n/2+1/2)}$; whereas when n is even, μ^* is any value such that $y_{(n/2)} \leq \mu^* \leq y_{(n/2+1)}$. In the latter case, μ^* is sometimes chosen as the middle of the interval.

There are several ways to show that μ^* minimizes $D(a)$. The following simple proof is due to Blyth (1990):

On the interval $y_{(k)} < a < y_{(k+1)}$, $k = 1, 2, \dots, n-1$, we have

$$\begin{aligned} D(a) &= \sum_{i=1}^n |y_{(i)} - a| \\ &= \sum_{i=1}^k (a - y_{(i)}) + \sum_{i=k+1}^n (y_{(i)} - a) \\ &= ka - \sum_{i=1}^k y_{(i)} + \sum_{i=k+1}^n y_{(i)} - (n-k)a. \end{aligned}$$

The function $D(a)$ is continuous for all a and is differentiable everywhere except at $y_1, y_2, \dots, y_n$. For $a \neq y_i$ ($i = 1, 2, \dots, n$), the derivative $D'(a)$ is given by

$$D'(a) = 2(k - n/2).$$

If $k \neq n/2$, then $D'(a) \neq 0$ on $(y_{(k)}, y_{(k+1)})$, and by Corollary 4.2.1, $D(a)$ must be strictly monotone on $[y_{(k)}, y_{(k+1)}]$, $k = 1, 2, \dots, n-1$.

Now, when n is odd, $D(a)$ is strictly monotone decreasing for $a \leq y_{(n/2+1/2)}$, because $D'(a) < 0$ over $(y_{(k)}, y_{(k+1)})$ for $k < n/2$. It is strictly monotone increasing for $a \geq y_{(n/2+1/2)}$, because $D'(a) > 0$ over $(y_{(k)}, y_{(k+1)})$ for $k > n/2$. Hence, $\mu^* = y_{(n/2+1/2)}$ is a point of absolute minimum for $D(a)$. Furthermore, when n is even, $D(a)$ is strictly monotone decreasing for $a \leq y_{(n/2)}$, because $D'(a) < 0$ over $(y_{(k)}, y_{(k+1)})$ for $k < n/2$. Also, $D(a)$ is constant over $(y_{(n/2)}, y_{(n/2+1)})$, because $D'(a) = 0$ for $k = n/2$, and is strictly monotone increasing for $a \geq y_{(n/2+1)}$, because $D'(a) > 0$ over $(y_{(k)}, y_{(k+1)})$ for $k > n/2$. This indicates that $D(a)$ achieves its absolute minimum at any point μ^* such that $y_{(n/2)} \leq \mu^* \leq y_{(n/2+1)}$, which completes the proof.

FURTHER READING AND ANNOTATED BIBLIOGRAPHY

- Apostol, T. M. (1964). *Mathematical Analysis*. Addison-Wesley, Reading, Massachusetts. (Chap. 5 discusses differentiation of functions of one variable.)
- Blyth, C. R. (1990). "Minimizing the sum of absolute deviations." *Amer. Statist.*, **44**, 329.
- Box, G. E. P., and D. R. Cox (1964). "An analysis of transformations." *J. Roy. Statist. Soc. Ser. B*, **26**, 211–243.
- Box, G. E. P., and N. R. Draper (1987). *Empirical Model-Building and Response Surfaces*. Wiley, New York. (Chap. 2 introduces the idea of approximating the mean of a response variable using low-order polynomials; Chap. 3 discusses the method of least squares for fitting empirical models; the use of transformations, including those for stabilizing variances, is described in Chap. 8.)
- Box, G. E. P., W. G. Hunter, and J. S. Hunter (1978). *Statistics for Experimenters*. Wiley, New York. (Various transformations are listed in Chap. 7, which include the Box–Cox and variance stabilizing transformations.)
- Buck, R. C. (1956). *Advanced Calculus*, McGraw-Hill, New York. (Chap. 2 discusses the mean value theorem and l'Hospital's rule.)
- Chatterjee, S., and B. Price (1977). *Regression Analysis by Example*. Wiley, New York. (Chap. 2 includes a discussion concerning variance stabilizing transformations, in addition to detection and removal of the effects of heteroscedasticity in regression analysis.)
- Cooke, W. P. (1988). "L'Hôpital's rule in a Poisson derivation." *Amer. Math. Monthly*, **95**, 253–254.
- Eggermont, P. P. B. (1988). "Noncentral difference quotients and the derivative." *Amer. Math. Monthly*, **95**, 551–553.
- Eves, H. (1976). *An Introduction to the History of Mathematics*, 4th ed. Holt, Rinehart and Winston, New York.
- Fulks, W. (1978). *Advanced Calculus*, 3rd ed. Wiley, New York. (Differentiation is the subject of Chap. 4.)

- Georgiev, A. A. (1984). "Kernel estimates of functions and their derivatives with applications," *Statist. Probab. Lett.*, **2**, 45–50.
- Hardy, G. H. (1955). *A Course of Pure Mathematics*, 10th ed. The University Press, Cambridge, England. (Chap. 6 covers differentiation and provides some interesting examples.)
- Hogg, R. V., and A. T. Craig (1965). *Introduction to Mathematical Statistics*, 2nd ed. Macmillan, New York. (Chap. 4 discusses distributions of functions of random variables.)
- James, A. T., and R. A. J. Conyers (1985). "Estimation of a derivative by a difference quotient: Its application to hepatocyte lactate metabolism." *Biometrics*, **41**, 467–476.
- Judge, G. G., W. E. Griffiths, R. C. Hill, and T. C. Lee (1980). *The Theory and Practice of Econometrics*. Wiley, New York.
- Khuri, A. I., and J. A. Cornell (1996). *Response Surfaces*, 2nd ed. Dekker, New York. (Chaps. 1 and 2 discuss the polynomial representation of a response surface and the method of least squares.)
- Lindgren, B. W. (1976). *Statistical Theory*, 3rd ed. Macmillan, New York. (Section 3.2 discusses the development of the Poisson process.)
- Menon, V. V., B. Prasad, and R. S. Singh (1984). "Non-parametric recursive estimates of a probability density function and its derivatives." *J. Statist. Plann. Inference*, **9**, 73–82.
- Montgomery, D. C., and E. A. Peck (1982). *Introduction to Linear Regression Analysis*. Wiley, New York. (Chap. 3 presents several methods useful for checking the validity of the basic regression assumptions. Several variance stabilizing transformations are also listed.)
- Parzen, E. (1962). *Stochastic Processes*. Holden-Day, San Francisco. (Chap. 1 introduces the definition of stochastic processes including the Poisson process.)
- Roberts, A. W., and D. E. Varberg (1973). *Convex Functions*. Academic Press, New York. (Chap. 1 discusses a characterization of convex functions using derivatives; Chap. 5 discusses maxima and minima of differentiable functions.)
- Roussas, G. G. (1973). *A First Course in Mathematical Statistics*. Addison-Wesley, Reading, Massachusetts. (Chap. 3 discusses absolutely continuous random variables.)
- Rudin, W. (1964). *Principles of Mathematical Analysis*. 2nd ed. McGraw-Hill, New York. (Differentiation is discussed in Chap. 5.)
- Sagan, H. (1974). *Advanced Calculus*. Houghton Mifflin, Boston. (Chap. 3 discusses differentiation.)
- Wetherill, G. B., P. Duncombe, M. Kenward, J. Köllerström, S. R. Paul, and B. J. Vowden (1986). *Regression Analysis with Applications*. Chapman and Hall, London, England. (Section 9.2 discusses the sources of heteroscedasticity in regression analysis.)
- Wilks, S. S. (1962). *Mathematical Statistics*. Wiley, New York. (Chap. 9 considers limit theorems including asymptotic distributions of functions of the sample mean.)

EXERCISES

In Mathematics

- 4.1. Let $f(x)$ be defined in a neighborhood of the origin. Show that if $f'(0)$ exists, then

$$\lim_{h \rightarrow 0} \frac{f(h) - f(-h)}{2h} = f'(0).$$

Give a counterexample to show that the converse is not true in general, that is, if the above limit exists, then it is not necessary that $f'(0)$ exists.

- 4.2. Let $f(x)$ and $g(x)$ have derivatives up to order n on $[a, b]$. Let $h(x) = f(x)g(x)$. Show that

$$h^{(n)}(x) = \sum_{k=0}^n \binom{n}{k} f^{(k)}(x) g^{(n-k)}(x).$$

(This is known as Leibniz's formula.)

- 4.3. Suppose that $f(x)$ has a derivative at a point x_0 , $a < x_0 < b$. Show that there exists a neighborhood $N_\delta(x_0)$ and a positive number A such that

$$|f(x) - f(x_0)| < A|x - x_0|$$

for all $x \in N_\delta(x_0)$, $x \neq x_0$.

- 4.4. Suppose that $f(x)$ is differentiable on $(0, \infty)$ and $f'(x) \rightarrow 0$ as $x \rightarrow \infty$. Let $g(x) = f(x+1) - f(x)$. Prove that $g(x) \rightarrow 0$ as $x \rightarrow \infty$.

- 4.5. Let the function $f(x)$ be defined as

$$f(x) = \begin{cases} x^3 - 2x, & x \geq 1, \\ ax^2 - bx + 1, & x < 1. \end{cases}$$

For what values of a and b does $f(x)$ have a continuous derivative?

- 4.6. Suppose that $f(x)$ is twice differentiable on $(0, \infty)$. Let m_0, m_1, m_2 be the least upper bounds of $|f(x)|$, $|f'(x)|$, and $|f''(x)|$, respectively, on $(0, \infty)$.

(a) Show that

$$|f'(x)| \leq \frac{m_0}{h} + hm_2$$

for all x in $(0, \infty)$ and for every $h > 0$.

(b) Deduce from (a) that

$$m_1^2 \leq 4m_0m_2.$$

4.7. Suppose that $\lim_{x \rightarrow x_0} f'(x)$ exists. Does it follow that $f(x)$ is differentiable at x_0 ? Give a proof to show that the statement is correct or produce a counterexample to show that it is false.

4.8. Show that $D(a) = \sum_{i=1}^n |y_i - a|$ has no derivatives with respect to a at $y_1, y_2, \dots, y_n$.

4.9. Suppose that the function $f(x)$ is such that $f'(x)$ and $f''(x)$ are continuous in a neighborhood of the origin and satisfies $f(0) = 0$. Show that

$$\lim_{x \rightarrow 0} \frac{d}{dx} \left[\frac{f(x)}{x} \right] = \frac{1}{2} f''(0).$$

4.10. Show that if $f'(x)$ exists and is bounded for all x , then $f(x)$ is uniformly continuous on R , the set of real numbers.

4.11. Suppose that $g: R \rightarrow R$ and that $|g'(x)| < M$ for all $x \in R$, where M is a positive constant. Define $f(x) = x + cg(x)$, where c is a positive constant. Show that it is possible to choose c small enough so that f is a one-to-one function.

4.12. Suppose that $f(x)$ is continuous on $[0, \infty)$, $f'(x)$ exists on $(0, \infty)$, $f(0) = 0$, and $f'(x)$ is monotone increasing on $(0, \infty)$. Show that $g(x)$ is monotone increasing on $(0, \infty)$ where $g(x) = f(x)/x$.

4.13. Show that if $a > 1$ and $m > 0$, then

(a) $\lim_{x \rightarrow \infty} (a^x/x^m) = \infty,$

(b) $\lim_{x \rightarrow \infty} [(\log x)/x^m] = 0.$

4.14. Apply l'Hospital's rule to find the limit

$$\lim_{x \rightarrow \infty} \left(1 + \frac{1}{x} \right)^x.$$

- 4.15.** (a) Find $\lim_{x \rightarrow 0^+} (\sin x)^x$.
(b) Find $\lim_{x \rightarrow 0^+} (e^{-1/x}/x)$.

- 4.16.** Show that

$$\lim_{x \rightarrow 0} [1 + ax + o(x)]^{1/x} = e^a,$$

where a is a constant and $o(x)$ is any function whose order of magnitude is less than that of x as $x \rightarrow 0$.

- 4.17.** Consider the functions $f(x) = 4x^3 + 6x^2 - 10x + 2$ and $g(x) = 3x^4 + 4x^3 - 5x^2 + 1$. Show that

$$\frac{f'(x)}{g'(x)} \neq \frac{f(1) - f(0)}{g(1) - g(0)}$$

for any $x \in (0, 1)$. Does this contradict Cauchy's mean value theorem?

- 4.18.** Suppose that $f(x)$ is differentiable for $a \leq x \leq b$. If $f'(a) < f'(b)$ and γ is a number such that $f'(a) < \gamma < f'(b)$, show that there exists a ξ , $a < \xi < b$, for which $f'(\xi) = \gamma$ [a similar result holds if $f'(a) > f'(b)$]. [Hint: Consider the function $g(x) = f(x) - \gamma(x - a)$. Show that $g(x)$ has a minimum at ξ .]

- 4.19.** Suppose that $f(x)$ is differentiable on (a, b) . Let $x_1, x_2, \dots, x_n$ be in (a, b) , and let $\lambda_1, \lambda_2, \dots, \lambda_n$ be positive numbers such that $\sum_{i=1}^n \lambda_i = 1$. Show that there exists a point c in (a, b) such that

$$\sum_{i=1}^n \lambda_i f'(x_i) = f'(c).$$

[Note: This is a generalization of the result in Exercise 4.18.]

- 4.20.** Let $x_1, x_2, \dots, x_n$ and $y_1, y_2, \dots, y_n$ be in (a, b) such that $x_i < y_i$ ($i = 1, 2, \dots, n$). Show that if $f(x)$ is differentiable on (a, b) , then there exists a point c in (a, b) such that

$$\sum_{i=1}^n [f(y_i) - f(x_i)] = f'(c) \sum_{i=1}^n (y_i - x_i).$$

- 4.21.** Give a Maclaurin's series expansion of the function $f(x) = \log(1 + x)$.

- 4.22.** Discuss the maxima and minima of the function $f(x) = (x^4 + 3)/(x^2 + 2)$.

- 4.23. Determine if $f(x) = e^{3x}/x$ has an absolute minimum on $(0, \infty)$.
- 4.24. For what values of a and b is the function

$$f(x) = \frac{1}{x^2 + ax + b}$$

bounded on the interval $[-1, 1]$? Find the absolute maximum on that interval.

In Statistics

- 4.25. Let Y be a continuous random variable whose cumulative distribution function, $F(y)$, is strictly monotone. Let $G(y)$ be another strictly monotone, continuous cumulative distribution function. Show that the cumulative distribution function of the random variable $G^{-1}[F(Y)]$ is $G(y)$.

- 4.26. Let Y have the cumulative distribution function

$$F(y) = \begin{cases} 1 - e^{-y}, & y \geq 0, \\ 0, & y < 0. \end{cases}$$

Find the density function of $W = \sqrt{Y}$.

- 4.27. Let Y be normally distributed with mean 1 and variance 0.04. Let $W = Y^3$.

- (a) Find the density function of W .
- (b) Find the exact mean and variance of W .
- (c) Find approximate values for the mean and variance of W using Taylor's expansion, and compare the results with those of (b).

- 4.28. Let Z be normally distributed with mean 0 and variance 1. Let $Y = Z^2$. Find the density function of Y .

[Note: The function $\psi(z) = z^2$ is not strictly monotone for all z .]

- 4.29. Let X be a random variable that denotes the age at failure of a component. The failure rate is defined as the probability of failure in a finite interval of time, say of length h , given the age of the component, say x . This failure rate is therefore equal to

$$P(x \leq X \leq x + h | X \geq x).$$

Consider the following limit:

$$\lim_{h \rightarrow 0} \frac{1}{h} P(x \leq X \leq x + h | X \geq x).$$

If this limit exists, then it is called the hazard rate, or instantaneous failure rate.

- (a) Give an expression for the failure rate in terms of $F(x)$, the cumulative distribution function of X .
- (b) Suppose that X has the exponential distribution with the cumulative distribution function

$$F(x) = 1 - e^{-x/\sigma}, \quad x \geq 0,$$

where σ is a positive constant. Show that X has a constant hazard rate.

- (c) Show that any random variable with a constant hazard rate must have the exponential distribution.
- 4.30.** Consider a Poisson process with parameter λ over the interval $(0, t)$. Divide this interval into n equal subintervals of length $h = t/n$. We consider that we have a “success” in a given subinterval if one arrival occurs in that subinterval. If there are no arrivals, then we consider that we have a “failure.” Let Y_n denote the number of “successes” in the n subintervals of length h . Then we have approximately

$$P(Y_n = r) = \binom{n}{r} p_n^r (1 - p_n)^{n-r}, \quad r = 0, 1, \dots, n,$$

where p_n is approximately equal to $\lambda h = \lambda t/n$. Show that

$$\lim_{n \rightarrow \infty} P(Y_n = r) = \frac{e^{-\lambda t} (\lambda t)^r}{r!}.$$