

CHAPTER 12

Approximation of Integrals

Integration plays an important role in many fields of science and engineering. For applications, numerical values of integrals are often required. However, in many cases, the evaluation of integrals, or *quadrature*, by elementary functions may not be feasible. Hence, approximating the value of an integral in a reliable fashion is a problem of utmost importance. Numerical quadrature is in fact one of the oldest branches of mathematics: the determination, approximately or exactly, of the areas of regions bounded by lines or curves, a subject which was studied by the ancient Babylonians (see Haber, 1970). The word “quadrature” indicates the process of measuring an area inside a curve by finding a square having the same area. Probably no other problem has exercised a greater or a longer attraction than that of constructing a square equal in area to a given circle. Thousands of people have worked on this problem, including the ancient Egyptians as far back as 1800 B.C.

In this chapter, we provide an exposition of methods for approximating integrals, including those that are multidimensional.

12.1. THE TRAPEZOIDAL METHOD

This is the simplest method of approximating an integral of the form $\int_a^b f(x) dx$, which represents the area bounded by the curve of the function $y = f(x)$ and the two lines $x = a$, $x = b$. The method is based on approximating the curve by a series of straight line segments. As a result, the area is approximated with a series of trapezoids. For this purpose, the interval from a to b is divided into n equal parts by the partition points $a = x_0, x_1, x_2, \dots, x_n = b$. For the i th trapezoid, which lies between x_{i-1} and x_i , its width is $h = (1/n)(b - a)$ and its area is given by

$$A_i = \frac{h}{2} [f(x_{i-1}) + f(x_i)], \quad i = 1, 2, \dots, n. \quad (12.1)$$

The sum, S_n , of $A_1, A_2, \dots, A_n$ provides an approximation to the integral $\int_a^b f(x) dx$.

$$\begin{aligned} S_n &= \sum_{i=1}^n A_i \\ &= \frac{h}{2} \{ [f(x_0) + f(x_1)] + [f(x_1) + f(x_2)] + \cdots + [f(x_{n-1}) + f(x_n)] \} \\ &= \frac{h}{2} \left[f(x_0) + f(x_n) + 2 \sum_{i=1}^{n-1} f(x_i) \right]. \end{aligned} \quad (12.2)$$

12.1.1. Accuracy of the Approximation

The accuracy in the trapezoidal method depends on the number n of trapezoids we take. The next theorem provides information concerning the error or approximation.

Theorem 12.1.1. Suppose that $f(x)$ has a continuous second derivative on $[a, b]$, and $|f''(x)| \leq M_2$ for all x in $[a, b]$. Then

$$\left| \int_a^b f(x) dx - S_n \right| \leq \frac{(b-a)^3 M_2}{12n^2},$$

where S_n is given by formula (12.2).

Proof. Consider the partition points $a = x_0, x_1, x_2, \dots, x_n = b$ such that $h = x_i - x_{i-1} = (1/n)(b-a)$, $i = 1, 2, \dots, n$. The integral of $f(x)$ from x_{i-1} to x_i is

$$I_i = \int_{x_{i-1}}^{x_i} f(x) dx. \quad (12.3)$$

Now, in the trapezoidal method, $f(x)$ is approximated in the interval $[x_{i-1}, x_i]$ by the right-hand side of the straight-line equation,

$$\begin{aligned} p_i(x) &= f(x_{i-1}) + \frac{1}{h} [f(x_i) - f(x_{i-1})](x - x_{i-1}) \\ &= \frac{x_i - x}{h} f(x_{i-1}) + \frac{x - x_{i-1}}{h} f(x_i) \\ &= \frac{x - x_i}{x_{i-1} - x_i} f(x_{i-1}) + \frac{x - x_{i-1}}{x_i - x_{i-1}} f(x_i), \quad i = 1, 2, \dots, n. \end{aligned}$$

Note that $p_i(x)$ is a linear Lagrange interpolating polynomial (of degree $n = 1$) with x_{i-1} and x_i as its points of interpolation [see formula (9.14)].

Using Theorem 9.2.2, the error of interpolation resulting from approximating $f(x)$ with $p_i(x)$ over $[x_{i-1}, x_i]$ is given by

$$f(x) - p_i(x) = \frac{1}{2!} f''(\xi_i)(x - x_{i-1})(x - x_i), \quad i = 1, 2, \dots, n, \quad (12.4)$$

where $x_{i-1} < \xi_i < x_i$. Formula (12.4) results from applying formula (9.15). Hence, the error of approximating I_i with A_i in (12.1) is

$$\begin{aligned} I_i - A_i &= \int_{x_{i-1}}^{x_i} [f(x) - p_i(x)] dx \\ &= \frac{1}{2!} f''(\xi_i) \int_{x_{i-1}}^{x_i} (x - x_{i-1})(x - x_i) dx \\ &= \frac{1}{2!} f''(\xi_i) \left[\frac{1}{3} (x_i^3 - x_{i-1}^3) - \frac{1}{2} (x_{i-1} + x_i)(x_i^2 - x_{i-1}^2) + x_{i-1} x_i (x_i - x_{i-1}) \right] \\ &= \frac{h}{2!} f''(\xi_i) \left[\frac{1}{3} (x_i^2 + x_{i-1} x_i + x_{i-1}^2) - \frac{1}{2} (x_{i-1} + x_i)^2 + x_{i-1} x_i \right] \\ &= \frac{h}{2!} f''(\xi_i) \left[\frac{1}{6} (2x_{i-1} x_i - x_{i-1}^2 - x_i^2) \right] \\ &= -\frac{h^3}{12} f''(\xi_i), \quad i = 1, 2, \dots, n. \end{aligned} \quad (12.5)$$

The total error of approximating $\int_a^b f(x) dx$ with S_n is then given by

$$\int_a^b f(x) dx - S_n = -\frac{h^3}{12} \sum_{i=1}^n f''(\xi_i).$$

It follows that

$$\begin{aligned} \left| \int_a^b f(x) dx - S_n \right| &\leq \frac{nh^3 M_2}{12} \\ &= \frac{(b-a)^3 M_2}{12n^2}. \quad \square \end{aligned} \quad (12.6)$$

An alternative procedure to approximating the integral $\int_a^b f(x) dx$ by a sum of trapezoids is to approximate $\int_{x_{i-1}}^{x_i} f(x) dx$ by a trapezoid bounded from above by the tangent to the curve of $y = f(x)$ at the point $x_{i-1} + h/2$, which is the midpoint of the interval $[x_{i-1}, x_i]$. In this case, the area of the i th trapezoid is

$$A_i^* = hf \left(x_{i-1} + \frac{h}{2} \right), \quad i = 1, 2, \dots, n.$$

Hence,

$$\int_a^b f(x) dx \approx h \sum_{i=1}^n f\left(x_{i-1} + \frac{h}{2}\right), \quad (12.7)$$

and $f(x)$ is approximated in the interval $[x_{i-1}, x_i]$ by

$$p_i^*(x) = f\left(x_{i-1} + \frac{h}{2}\right) + \left(x - x_{i-1} - \frac{h}{2}\right) f'\left(x_{i-1} + \frac{h}{2}\right).$$

By applying Taylor's theorem (Theorem 4.3.1) to $f(x)$ in a neighborhood of $x_{i-1} + h/2$, we obtain

$$\begin{aligned} f(x) &= f\left(x_{i-1} + \frac{h}{2}\right) + \left(x - x_{i-1} - \frac{h}{2}\right) f'\left(x_{i-1} + \frac{h}{2}\right) \\ &\quad + \frac{1}{2!} \left(x - x_{i-1} - \frac{h}{2}\right)^2 f''(\eta_i) \\ &= p_i^*(x) + \frac{1}{2!} \left(x - x_{i-1} - \frac{h}{2}\right)^2 f''(\eta_i), \end{aligned}$$

where η_i lies between $x_{i-1} + h/2$ and x . The error of approximating $\int_{x_{i-1}}^{x_i} f(x) dx$ with A_i^* is then given by

$$\begin{aligned} \int_{x_{i-1}}^{x_i} [f(x) - p_i^*(x)] dx &= \frac{1}{2!} \int_{x_{i-1}}^{x_i} \left(x - x_{i-1} - \frac{h}{2}\right)^2 f''(\eta_i) dx \\ &\leq \frac{M_2}{2!} \int_{x_{i-1}}^{x_i} \left(x - x_{i-1} - \frac{h}{2}\right)^2 dx \\ &= \frac{h^3 M_2}{24}. \end{aligned}$$

Consequently, the absolute value of the total error in this case has an upper bound of the form

$$\begin{aligned} \left| \sum_{i=1}^n \int_{x_{i-1}}^{x_i} [f(x) - p_i^*(x)] dx \right| &\leq \frac{nh^3 M_2}{24} \\ &= \frac{(b-a)^3 M_2}{24n^2}. \end{aligned} \quad (12.8)$$

We note that the upper bound in (12.8) is half as large as the one in (12.6). This alternative procedure is therefore slightly more precise than the original one. Both procedures produce an approximating error that is $O(1/n^2)$. It should be noted that this error does not include the roundoff errors in the computation of the areas of the approximating trapezoids.

EXAMPLE 12.1.1. Consider approximating the integral $\int_1^2 dx/x$, which has an exact value equal to $\log 2 \approx 0.693147$. Let us divide the interval $[1, 2]$ into $n = 10$ subintervals of length $h = \frac{1}{10}$. Hence, $x_0 = 1, x_1 = 1.1, \dots, x_{10} = 2.0$. Using the first trapezoidal method (formula 12.2), we obtain

$$\begin{aligned}\int_1^2 \frac{dx}{x} &\approx \frac{1}{20} \left[1 + \frac{1}{2} + 2 \sum_{i=1}^9 \frac{1}{x_i} \right] \\ &= 0.69377.\end{aligned}$$

Using now the second trapezoidal method [formula (12.7)], we get

$$\begin{aligned}\int_1^2 \frac{dx}{x} &\approx \frac{1}{10} \sum_{i=1}^{10} \frac{1}{x_{i-1} + 0.05} \\ &= 0.69284.\end{aligned}$$

12.2. SIMPSON'S METHOD

Let us again consider the integral $\int_a^b f(x) dx$. Let $a = x_0 < x_1 < \dots < x_{2n-1} < x_{2n} = b$ be a sequence of equally spaced points that partition the interval $[a, b]$ such that $x_i - x_{i-1} = h, i = 1, 2, \dots, 2n$. Simpson's method is based on approximating the graph of the function $f(x)$ over the interval $[x_{i-1}, x_{i+1}]$ by a parabola which agrees with $f(x)$ at the points x_{i-1}, x_i , and x_{i+1} . Thus, over $[x_{i-1}, x_{i+1}]$, $f(x)$ is approximated with a Lagrange interpolating polynomial of degree 2 of the form [see formula (9.14)]

$$\begin{aligned}q_i(x) &= f(x_{i-1}) \frac{(x - x_i)(x - x_{i+1})}{(x_{i-1} - x_i)(x_{i-1} - x_{i+1})} \\ &\quad + f(x_i) \frac{(x - x_{i-1})(x - x_{i+1})}{(x_i - x_{i-1})(x_i - x_{i+1})} + f(x_{i+1}) \frac{(x - x_{i-1})(x - x_i)}{(x_{i+1} - x_{i-1})(x_{i+1} - x_i)} \\ &= f(x_{i-1}) \frac{(x - x_i)(x - x_{i+1})}{2h^2} - f(x_i) \frac{(x - x_{i-1})(x - x_{i+1})}{h^2} \\ &\quad + f(x_{i+1}) \frac{(x - x_{i-1})(x - x_i)}{2h^2}.\end{aligned}$$

It follows that

$$\begin{aligned}
 \int_{x_{i-1}}^{x_{i+1}} f(x) dx &\approx \int_{x_{i-1}}^{x_{i+1}} q_i(x) dx \\
 &= \frac{f(x_{i-1})}{2h^2} \int_{x_{i-1}}^{x_{i+1}} (x-x_i)(x-x_{i+1}) dx \\
 &\quad - \frac{f(x_i)}{h^2} \int_{x_{i-1}}^{x_{i+1}} (x-x_{i-1})(x-x_{i+1}) dx \\
 &\quad + \frac{f(x_{i+1})}{2h^2} \int_{x_{i-1}}^{x_{i+1}} (x-x_{i-1})(x-x_i) dx \\
 &= \frac{f(x_{i-1})}{2h^2} \left(\frac{2h^3}{3} \right) - \frac{f(x_i)}{h^2} \left(\frac{-4h^3}{3} \right) + \frac{f(x_{i+1})}{2h^2} \left(\frac{2h^3}{3} \right) \\
 &= \frac{h}{3} [f(x_{i-1}) + 4f(x_i) + f(x_{i+1})], \quad i = 1, 3, \dots, 2n-1.
 \end{aligned} \tag{12.9}$$

By adding up all the approximations in (12.9) for $i = 1, 3, \dots, 2n-1$, we obtain

$$\int_a^b f(x) dx \approx \frac{h}{3} \left[f(x_0) + 4 \sum_{i=1}^n f(x_{2i-1}) + 2 \sum_{i=1}^{n-1} f(x_{2i}) + f(x_{2n}) \right]. \tag{12.10}$$

As before, in the case of the trapezoidal method, the accuracy of the approximation in (12.10) can be figured out by using formula (9.15). Courant and John (1965, page 487), however, stated that the error of approximation can be improved by one order of magnitude by using a cubic interpolating polynomial which agrees with $f(x)$ at x_{i-1}, x_i, x_{i+1} , and whose derivative at x_i is equal to $f'(x_i)$. Such a polynomial gives a better approximation to $f(x)$ over $[x_{i-1}, x_{i+1}]$ than the quadratic one, and still provides the same approximation formula (12.10) for the integral. If $q_i(x)$ is chosen as such, then the error of interpolation resulting from approximating $f(x)$ with $q_i(x)$ over $[x_{i-1}, x_{i+1}]$ is given by $f(x) - q_i(x) = (1/4!)f^{(4)}(\xi_i)(x-x_{i-1})(x-x_i)^2(x-x_{i+1})$, where $x_{i-1} < \xi_i < x_{i+1}$, provided that $f^{(4)}(x)$ exists and is continuous on $[a, b]$. This is equivalent to using formula (9.15) with $n = 3$ and with two of the interpolation points coincident at x_i . We then have

$$\left| \int_{x_{i-1}}^{x_{i+1}} [f(x) - q_i(x)] dx \right| \leq \frac{M_4}{4!} \int_{x_{i-1}}^{x_{i+1}} (x-x_{i-1})(x-x_i)^2(x-x_{i+1}) dx,$$

$$i = 1, 3, \dots, 2n-1, \tag{12.11}$$

where M_4 is an upper bound on $|f^{(4)}(x)|$ for $a \leq x \leq b$. By computing the integral in (12.11) we obtain

$$\left| \int_{x_{i-1}}^{x_{i+1}} [f(x) - q_i(x)] dx \right| \leq \frac{M_4 h^5}{90}, \quad i = 1, 3, \dots, 2n - 1.$$

Consequently, the total error of approximation in (12.10) is less than or equal to

$$\frac{nM_4 h^5}{90} = \frac{M_4 (b-a)^5}{2880n^4}, \quad (12.12)$$

since $h = (b-a)/2n$. Thus the error of approximation with Simpson's method is $O(1/n^4)$, where n is half the number of subintervals into which $[a, b]$ is divided. Hence, Simpson's method yields a much more accurate approximation than the trapezoidal method.

As an example, let us apply Simpson's method to the calculation of the integral in Example 12.1.1 using the same division of $[1, 2]$ into 10 subintervals, each of length $h = \frac{1}{10}$. By applying formula (12.10) we obtain

$$\begin{aligned} \int_1^2 \frac{dx}{x} &\approx \frac{0.10}{3} \left[1 + 4 \left(\frac{1}{1.1} + \frac{1}{1.3} + \frac{1}{1.5} + \frac{1}{1.7} + \frac{1}{1.9} \right) \right. \\ &\quad \left. + 2 \left(\frac{1}{1.2} + \frac{1}{1.4} + \frac{1}{1.6} + \frac{1}{1.8} \right) + \frac{1}{2} \right] \\ &= 0.69315. \end{aligned}$$

12.3. NEWTON–COTES METHODS

The trapezoidal and Simpson's methods are two special cases of a general series of approximate integration methods of the so-called Newton–Cotes type. In the trapezoidal method, straight line segments were used to approximate the graph of the function $f(x)$ between a and b . In Simpson's method, the approximation was carried out using a series of parabolas. We can refine this approximation even further by considering a series of cubic curves, quartic curves, and so on. For cubic approximations, four equally spaced points are used to subdivide each subinterval of $[a, b]$ (instead of two points for the trapezoidal method and three points for Simpson's method), whereas five points are needed for quartic approximation, and so on. All such approximations are of the Newton–Cotes type.

12.4. GAUSSIAN QUADRATURE

All Newton–Cotes methods require the use of equally spaced points, as was seen in the cases of the trapezoidal method and Simpson’s method. If this requirement is waived, then it is possible to select the points in a manner that reduces the approximation error.

Let $x_0 < x_1 < x_2 < \cdots < x_n$ be $n + 1$ distinct points in $[a, b]$. Consider the approximation

$$\int_a^b f(x) dx \approx \sum_{i=0}^n \omega_i f(x_i), \quad (12.13)$$

where the coefficients, $\omega_0, \omega_1, \dots, \omega_n$, are to be determined along with the points $x_0, x_1, \dots, x_n$. The total number of unknown quantities in (12.13) is $2n + 2$. Hence, $2n + 2$ conditions must be specified. According to the so-called Gaussian integration rule, the ω_i ’s and x_i ’s are chosen such that the approximation in (12.13) will be exact for all polynomials of degrees not exceeding $2n + 1$. This is equivalent to requiring that the approximation (12.13) be exact for $f(x) = x^j$, $j = 0, 1, 2, \dots, 2n + 1$, that is,

$$\int_a^b x^j dx = \sum_{i=0}^n \omega_i x_i^j, \quad j = 0, 1, 2, \dots, 2n + 1. \quad (12.14)$$

This process produces $2n + 2$ equations to be solved for the ω_i ’s and x_i ’s. In particular, if the limits of integration are $a = -1$, $b = 1$, then it can be shown (see Phillips and Taylor, 1973, page 140) that the x_i -values will be the $n + 1$ zeros of the Legendre polynomial $p_{n+1}(x)$ of degree $n + 1$ (see Section 10.2). The ω_i -values can be easily found by solving the system of equations (12.14), which is linear in the ω_i ’s.

For example, for $n = 1$, the zeros of $p_2(x) = \frac{1}{2}(3x^2 - 1)$ are $x_0 = -1/\sqrt{3}$, $x_1 = 1/\sqrt{3}$. Applying (12.14), we obtain

$$\begin{aligned} \int_{-1}^1 dx &= \omega_0 + \omega_1 \Rightarrow \omega_0 + \omega_1 = 2, \\ \int_{-1}^1 x dx &= \omega_0 x_0 + \omega_1 x_1 \Rightarrow \frac{1}{\sqrt{3}}(-\omega_0 + \omega_1) = 0, \\ \int_{-1}^1 x^2 dx &= \omega_0 x_0^2 + \omega_1 x_1^2 \Rightarrow \frac{1}{3}(\omega_0 + \omega_1) = \frac{2}{3}, \\ \int_{-1}^1 x^3 dx &= \omega_0 x_0^3 + \omega_1 x_1^3 \Rightarrow \frac{1}{3\sqrt{3}}(-\omega_0 + \omega_1) = 0. \end{aligned}$$

We note that the last two equations are identical to the first two. Solving the

latter for ω_0 and ω_1 , we get $\omega_0 = \omega_1 = 1$. Hence, we have the approximation

$$\int_{-1}^1 f(x) dx \approx f\left(-\frac{1}{\sqrt{3}}\right) + f\left(\frac{1}{\sqrt{3}}\right),$$

which is exact if $f(x)$ is a polynomial of degree not exceeding $2n + 1 = 3$.

If the limits of integration are not equal to $-1, 1$, we can easily convert the integral $\int_a^b f(x) dx$ to one with the limits $-1, 1$ by making the change of variable

$$z = \frac{2x - (a + b)}{b - a}.$$

This converts the general integral $\int_a^b f(x) dx$ to the integral $[(b - a)/2] \int_{-1}^1 g(z) dz$, where

$$g(z) = f\left[\frac{(b - a)z + b + a}{2}\right].$$

We therefore have the approximation

$$\int_a^b f(x) dx \approx \frac{b - a}{2} \sum_{i=0}^n \omega_i g(z_i), \quad (12.15)$$

where the z_i 's are the zeros of the Legendre polynomial $p_{n+1}(z)$.

It can be shown that (see Davis and Rabinowitz, 1975, page 75) that when $a = -1$, $b = 1$, the error of approximation in (12.13) is given by

$$\int_a^b f(x) dx - \sum_{i=0}^n \omega_i f(x_i) = \frac{(b - a)^{2n+3} [(n + 1)!]^4}{(2n + 3) [(2n + 2)!]^3} f^{(2n+2)}(\xi), \quad a < \xi < b, \quad (12.16)$$

provided that $f^{(2n+2)}(x)$ is continuous on $[a, b]$. This error decreases rapidly as n increases. Thus this Gaussian quadrature provides a very good approximation with a formula of the type given in (12.13).

There are several extensions of the approximation in (12.13). These extensions are of the form

$$\int_a^b \lambda(x) f(x) dx \approx \sum_{i=0}^n \omega_i f(x_i), \quad (12.17)$$

where $\lambda(x)$ is a particular positive weight function. As before, the coefficients $\omega_0, \omega_1, \dots, \omega_n$ and the points $x_0, x_1, \dots, x_n$, which belong to $[a, b]$, are chosen so that (12.17) is exact for all polynomials of degrees not exceeding $2n + 1$. The choice of the x_i 's depends on the form of $\lambda(x)$. It can be shown that the values of x_i are the zeros of a polynomial of degree $n + 1$ belonging to the sequence of polynomials that are orthogonal on $[a, b]$ with respect to $\lambda(x)$ (see Davis and Rabinowitz, 1975, page 74; Phillips and Taylor, 1973, page 142). For example, if $a = -1$, $b = 1$, $\lambda(x) = (1 - x)^\alpha(1 + x)^\beta$, $\alpha > -1$, $\beta > -1$, then the x_i 's are the zeros of the Jacobi polynomial $p_{n+1}^{(\alpha, \beta)}(x)$ (see Section 10.3). Also, if $a = -1$, $b = 1$, $\lambda(x) = (1 - x^2)^{-1/2}$, then the x_i 's are the zeros of the Chebyshev polynomial of the first kind, $T_{n+1}(x)$ (see Section 10.4), and so on. For those two cases, formula (12.17) is called the *Gauss–Jacobi quadrature* formula and the *Gauss–Chebyshev quadrature* formula, respectively. The choice $\lambda(x) = 1$ results in the original formula (12.13), which is now referred to as the *Gauss–Legendre quadrature* formula.

EXAMPLE 12.4.1. Consider the integral $\int_0^1 dx/(1+x)$, which has the exact value $\log 2 = 0.69314718$. Applying formula (12.15), we get

$$\begin{aligned} \int_0^1 \frac{dx}{1+x} &\approx \frac{1}{2} \sum_{i=0}^n \frac{\omega_i}{1 + \frac{1}{2}(z_i + 1)} \\ &= \sum_{i=0}^n \frac{\omega_i}{3 + z_i}, \quad -1 \leq z_i \leq 1, \end{aligned} \quad (12.18)$$

where the z_i 's are the zeros of the Legendre polynomial $p_{n+1}(z)$, $z = 2x - 1$.

Let $n = 1$; then $p_2(z) = \frac{1}{2}(3z^2 - 1)$ with zeros equal to $z_0 = -1/\sqrt{3}$, $z_1 = 1/\sqrt{3}$. We have seen earlier that $\omega_0 = 1$, $\omega_1 = 1$; hence, from (12.18), we obtain

$$\begin{aligned} \int_0^1 \frac{dx}{1+x} &\approx \sum_{i=0}^1 \frac{\omega_i}{3 + z_i} \\ &= \frac{\sqrt{3}}{3\sqrt{3} - 1} + \frac{\sqrt{3}}{3\sqrt{3} + 1} \\ &= 0.692307691. \end{aligned}$$

Let us now use $n = 2$ in (12.18). Then $p_3(z) = \frac{1}{2}(5z^3 - 3z)$ (see Section 10.2). Its zeros are $z_0 = -(\frac{3}{5})^{1/2}$, $z_1 = 0$, $z_2 = (\frac{3}{5})^{1/2}$. To find the ω_i 's, we apply formula (12.14) using $a = -1$, $b = 1$, and z in place of x . For

$j = 0, 1, 2, 3, 4, 5$, we have

$$\int_{-1}^1 dz = \omega_0 + \omega_1 + \omega_2,$$

$$\int_{-1}^1 z dz = \omega_0 z_0 + \omega_1 z_1 + \omega_2 z_2,$$

$$\int_{-1}^1 z^2 dz = \omega_0 z_0^2 + \omega_1 z_1^2 + \omega_2 z_2^2,$$

$$\int_{-1}^1 z^3 dz = \omega_0 z_0^3 + \omega_1 z_1^3 + \omega_2 z_2^3,$$

$$\int_{-1}^1 z^4 dz = \omega_0 z_0^4 + \omega_1 z_1^4 + \omega_2 z_2^4,$$

$$\int_{-1}^1 z^5 dz = \omega_0 z_0^5 + \omega_1 z_1^5 + \omega_2 z_2^5.$$

These equations can be written as

$$\omega_0 + \omega_1 + \omega_2 = 2,$$

$$\left(\frac{3}{5}\right)^{1/2}(-\omega_0 + \omega_2) = 0,$$

$$\frac{3}{5}(\omega_0 + \omega_2) = \frac{2}{3},$$

$$\left(\frac{3}{5}\right)^{3/2}(-\omega_0 + \omega_2) = 0,$$

$$\frac{9}{25}(\omega_0 + \omega_2) = \frac{2}{5},$$

$$\frac{9}{25}\left(\frac{3}{5}\right)^{1/2}(-\omega_0 + \omega_2) = 0.$$

The above six equations can be reduced to only three that are linearly independent, namely,

$$\omega_0 + \omega_1 + \omega_2 = 2,$$

$$-\omega_0 + \omega_2 = 0,$$

$$\omega_0 + \omega_2 = \frac{10}{9},$$

the solution of which is $\omega_0 = \frac{5}{9}$, $\omega_1 = \frac{8}{9}$, $\omega_2 = \frac{5}{9}$. Substituting the ω_i 's and z_i 's

in (12.18), we obtain

$$\begin{aligned}\int_0^1 \frac{dx}{1+x} &\approx \frac{\omega_0}{3+z_0} + \frac{\omega_1}{3+z_1} + \frac{\omega_2}{3+z_2} \\ &= \frac{\frac{5}{9}}{3 - (\frac{3}{5})^{1/2}} + \frac{\frac{8}{9}}{3} + \frac{\frac{5}{9}}{3 + (\frac{3}{5})^{1/2}} \\ &= 0.693121685.\end{aligned}$$

For higher values of n , the zeros of Legendre polynomials and the corresponding values of ω_i can be found, for example, in Shoup (1984, Table 7.5) and Krylov (1962, Appendix A).

12.5. APPROXIMATION OVER AN INFINITE INTERVAL

Consider an integral of the form $\int_a^\infty f(x) dx$, which is improper of the first kind (see Section 6.5). It can be approximated by using the integral $\int_a^b f(x) dx$ for a sufficiently large value of b , provided, of course, that $\int_a^\infty f(x) dx$ is convergent. The methods discussed earlier in Sections 12.1–12.4 can then be applied to $\int_a^b f(x) dx$.

For improper integrals of the first kind, of the form $\int_0^\infty \lambda(x)f(x) dx$, $\int_{-\infty}^\infty \lambda(x)f(x) dx$, we have the following Gaussian approximations:

$$\int_0^\infty \lambda(x)f(x) dx \approx \sum_{i=0}^n \omega_i f(x_i), \quad (12.19)$$

$$\int_{-\infty}^\infty \lambda(x)f(x) dx \approx \sum_{i=0}^n \omega_i f(x_i), \quad (12.20)$$

where, as before, the x_i 's and ω_i 's are chosen so that (12.19) and (12.20) are exact for all polynomials of degrees not exceeding $2n+1$. For the weight function $\lambda(x)$ in (12.19), the choice $\lambda(x) = e^{-x}$ gives the *Gauss–Laguerre quadrature*, for which the x_i 's are the zeros of the Laguerre polynomial $L_{n+1}(x)$ of degree $n+1$ and $\alpha = 0$ (see Section 10.6). The associated error of approximation is given by (see Davis and Rabinowitz, 1975, page 173)

$$\int_0^\infty e^{-x}f(x) dx - \sum_{i=0}^n \omega_i f(x_i) = \frac{[(n+1)!]^2}{(2n+2)!} f^{(2n+2)}(\xi), \quad 0 < \xi < \infty.$$

Choosing $\lambda(x) = e^{-x^2}$ in (12.20) gives the *Gauss–Hermite quadrature*, and the corresponding x_i 's are the zeros of the Hermite polynomial $H_{n+1}(x)$ of degree $n+1$ (see Section 10.5). The associated error of approximation is of the form (see Davis and Rabinowitz, 1975, page 174)

$$\int_{-\infty}^\infty e^{-x^2}f(x) dx - \sum_{i=0}^n \omega_i f(x_i) = \frac{(n+1)! \sqrt{\pi}}{2^{n+1}(2n+2)!} f^{(2n+2)}(\xi), \quad -\infty < \xi < \infty.$$

We can also use the Gauss-Laguerre and the Gauss-Hermite quadrature formulas to approximate convergent integrals of the form $\int_0^\infty f(x) dx$, $\int_{-\infty}^\infty f(x) dx$:

$$\begin{aligned}\int_0^\infty f(x) dx &= \int_0^\infty e^{-x} e^x f(x) dx \\ &\approx \sum_{i=0}^n \omega_i e^{x_i} f(x_i), \\ \int_{-\infty}^\infty f(x) dx &= \int_{-\infty}^\infty e^{-x^2} e^{x^2} f(x) dx \\ &\approx \sum_{i=0}^n \omega_i e^{x_i^2} f(x_i).\end{aligned}$$

EXAMPLE 12.5.1. Consider the integral

$$\begin{aligned}I &= \int_0^\infty \frac{e^{-x} x}{1 - e^{-2x}} dx \\ &\approx \sum_{i=0}^n \omega_i f(x_i),\end{aligned}\tag{12.21}$$

where $f(x) = x(1 - e^{-2x})^{-1}$ and the x_i 's are the zeros of the Laguerre polynomial $L_{n+1}(x)$. To find expressions for $L_n(x)$, $n = 0, 1, 2, \dots$, we can use the recurrence relation (10.37) with $\alpha = 0$, which can be written as

$$L_{n+1}(x) = (x - n - 1)L_n(x) - x \frac{dL_n(x)}{dx}, \quad n = 0, 1, 2, \dots \tag{12.22}$$

Recall that $L_0(x) = 1$. Choosing $n = 1$ in (12.21), we get

$$I \approx \omega_0 f(x_0) + \omega_1 f(x_1). \tag{12.23}$$

From (12.22) we have

$$\begin{aligned}L_1(x) &= (x - 1)L_0(x) \\ &= x - 1, \\ L_2(x) &= (x - 2)L_1(x) - x \frac{d(x - 1)}{dx} \\ &= (x - 2)(x - 1) - x \\ &= x^2 - 4x + 2.\end{aligned}$$

The zeros of $L_2(x)$ are $x_0 = 2 - \sqrt{2}$, $x_1 = 2 + \sqrt{2}$. To find ω_0 and ω_1 , formula (12.19) must be exact for all polynomials of degrees not exceeding

$2n + 1 = 3$. This is equivalent to requiring that

$$\begin{aligned}\int_0^\infty e^{-x} dx &= \omega_0 + \omega_1, \\ \int_0^\infty e^{-x} x dx &= \omega_0 x_0 + \omega_1 x_1, \\ \int_0^\infty e^{-x} x^2 dx &= \omega_0 x_0^2 + \omega_1 x_1^2, \\ \int_0^\infty e^{-x} x^3 dx &= \omega_0 x_0^3 + \omega_1 x_1^3.\end{aligned}$$

Only two equations are linearly independent, the solution of which is $\omega_0 = 0.853553$, $\omega_1 = 0.146447$. From (12.23) we then have

$$\begin{aligned}I &\approx \frac{\omega_0 x_0}{1 - e^{-2x_0}} + \frac{\omega_1 x_1}{1 - e^{-2x_1}} \\ &= 1.225054.\end{aligned}$$

Let us now calculate (12.21) using $n = 2, 3, 4$. The zeros of Laguerre polynomials $L_3(x)$, $L_4(x)$, $L_5(x)$, and the values of ω_i for each n are shown in Table 12.1. These values are given in Ralston and Rabinowitz (1978, page 106) and also in Krylov (1962, Appendix C). The corresponding approximate values of I are given in Table 12.1. It can be shown that the exact value of I is $\pi^2/8 \approx 1.2337$.

Table 12.1. Zeros of Laguerre Polynomials (x_i), Values of ω_i , and the Corresponding Approximate Values^a of I

n	x_i	ω_i	I
1	0.585786	0.853553	1.225054
	3.414214	0.146447	
2	0.415775	0.711093	1.234538
	2.294280	0.278518	
	6.289945	0.010389	
3	0.322548	0.603154	1.234309
	1.745761	0.357419	
	4.536620	0.038888	
	9.395071	0.000539	
4	0.263560	0.521756	1.233793
	1.413403	0.398667	
	3.596426	0.075942	
	7.085810	0.003612	
	12.640801	0.000023	

^aSee (12.21).

12.6. THE METHOD OF LAPLACE

This method is used to approximate integrals of the form

$$I(\lambda) = \int_a^b \varphi(x) e^{\lambda h(x)} dx, \quad (12.24)$$

where λ is a large positive constant, $\varphi(x)$ is continuous on $[a, b]$, and the first and second derivatives of $h(x)$ are continuous on $[a, b]$. The limits a and b may be finite or infinite. This integral was used by Laplace in his original development of the central limit theorem (see Section 4.5.1). More specifically, if $X_1, X_2, \dots, X_n, \dots$ is a sequence of independent and identically distributed random variables with a common density function, then the density function of the sum $S_n = \sum_{i=1}^n X_i$ can be represented in the form (12.24) (see Wong, 1989, Chapter 2).

Suppose that $h(x)$ has a single maximum in the interval $[a, b]$ at $x = t$, $a \leq t < b$, where $h'(t) = 0$ and $h''(t) < 0$. Hence, $e^{\lambda h(x)}$ is maximized at t for any $\lambda > 0$. Suppose further that $e^{\lambda h(x)}$ becomes very strongly peaked at $x = t$ and decreases rapidly away from $x = t$ on $[a, b]$ as $\lambda \rightarrow \infty$. In this case, the major portion of $I(\lambda)$ comes from integrating the function $\varphi(x)e^{\lambda h(x)}$ over a small neighborhood around $x = t$. Under these conditions, it can be shown that if $a < t < b$, and as $\lambda \rightarrow \infty$,

$$I(\lambda) \sim \varphi(t) e^{\lambda h(t)} \left[\frac{-2\pi}{\lambda h''(t)} \right]^{1/2}, \quad (12.25)$$

where $\sim$ denotes asymptotic equality (see Section 3.3). Formula (12.25) is known as *Laplace's approximation*.

A heuristic derivation of (12.25) can be arrived at by replacing $\varphi(x)$ and $h(x)$ by the leading terms in their Taylor's series expansions around $x = t$. The integration limits are then extended to $-\infty$ and ∞ , that is,

$$\begin{aligned} \int_a^b \varphi(x) e^{\lambda h(x)} dx &\approx \int_a^b \varphi(t) \exp \left[\lambda h(t) + \frac{\lambda}{2} (x-t)^2 h''(t) \right] dx \\ &\approx \int_{-\infty}^{\infty} \varphi(t) \exp \left[\lambda h(t) + \frac{\lambda}{2} (x-t)^2 h''(t) \right] dx \\ &= \varphi(t) e^{\lambda h(t)} \int_{-\infty}^{\infty} \exp \left[\frac{\lambda}{2} (x-t)^2 h''(t) \right] dx \end{aligned} \quad (12.26)$$

$$= \varphi(t) e^{\lambda h(t)} \left[\frac{-2\pi}{\lambda h''(t)} \right]^{1/2}. \quad (12.27)$$

Formula (12.27) follows from (12.26) by making use of the fact that $\int_0^\infty e^{-x^2} dx = \frac{1}{2} \Gamma(\frac{1}{2}) = \sqrt{\pi}/2$, where $\Gamma(\cdot)$ is the gamma function (see Example 6.9.6), or by simply evaluating the integral of a normal density function.

If $t = a$, then it can be shown that as $\lambda \rightarrow \infty$,

$$I(\lambda) \sim \varphi(a) e^{\lambda h(a)} \left[\frac{-\pi}{2\lambda h''(a)} \right]^{1/2}. \quad (12.28)$$

Rigorous proofs of (12.27) and (12.28) can be found in Wong (1989, Chapter 2), Copson (1965, Chapter 5), Fulks (1978, Chapter 18), and Lange (1999, Chapter 4).

EXAMPLE 12.6.1. Consider the gamma function

$$\Gamma(n+1) = \int_0^\infty e^{-x} x^n dx, \quad n > -1. \quad (12.29)$$

Let us find an approximation for $\Gamma(n+1)$ when n is large and positive, but not necessarily an integer. Let $x = nz$; then (12.29) can be written as

$$\begin{aligned} \Gamma(n+1) &= n \int_0^\infty e^{-nz} \exp[n \log(nz)] dz \\ &= n \int_0^\infty e^{-nz} \exp[n \log n + n \log z] dz \\ &= n^{n+1} \int_0^\infty \exp[n(-z + \log z)] dz. \end{aligned} \quad (12.30)$$

Let $h(z) = -z + \log z$. Then $h(z)$ has a unique maximum at $z = 1$ with $h'(1) = 0$ and $h''(1) = -1$. Applying formula (12.27) to (12.30), we obtain

$$\begin{aligned} \Gamma(n+1) &\sim n^{n+1} e^{-n} \left[\frac{-2\pi}{n(-1)} \right]^{1/2} \\ &= e^{-n} n^n \sqrt{2\pi n}, \end{aligned} \quad (12.31)$$

as $n \rightarrow \infty$. Formula (12.31) is known as *Stirling's formula*.

EXAMPLE 12.6.2. Consider the integral,

$$I_n(\lambda) = \frac{1}{\pi} \int_0^\pi \exp(\lambda \cos x) \cos nx dx,$$

as $\lambda \rightarrow \infty$. This integral looks like (12.24) with $h(x) = \cos x$, which has a single maximum at $x = 0$ in $[0, \pi]$. Since $h'(0) = -1$, then by applying (12.28)

we obtain, as $\lambda \rightarrow \infty$,

$$\begin{aligned} I_n(\lambda) &\sim \frac{1}{\pi} e^\lambda \left[\frac{-\pi}{2\lambda(-1)} \right]^{1/2} \\ &= \frac{e^\lambda}{\sqrt{2\pi\lambda}}. \end{aligned}$$

EXAMPLE 12.6.3. Consider the integral

$$I(\lambda) = \int_0^\pi \exp[\lambda \sin x] dx$$

as $\lambda \rightarrow \infty$. Here, $h(x) = \sin x$, which has a single maximum at $x = \pi/2$ in $[0, \pi]$, and $h''(\pi/2) = -1$. From (12.27), we get

$$\begin{aligned} I(\lambda) &\sim e^\lambda \left[\frac{-2\pi}{\lambda(-1)} \right]^{1/2} \\ &= e^\lambda \sqrt{\frac{2\pi}{\lambda}} \end{aligned}$$

as $\lambda \rightarrow \infty$.

12.7. MULTIPLE INTEGRALS

We recall that integration of a multivariable function was discussed in Section 7.9. In the present section, we consider approximate integration formulas for an n -tuple Riemann integral over a region D in an n -dimensional Euclidean space R^n .

For example, let us consider the double integral $I = \iint_D f(x_1, x_2) dx_1 dx_2$, where $D \subset R^2$ is the region

$$D = \{(x_1, x_2) | a \leq x_1 \leq b, \psi(x_1) \leq x_2 \leq \phi(x_1)\}.$$

Then

$$\begin{aligned} I &= \int_a^b \left[\int_{\psi(x_1)}^{\phi(x_1)} f(x_1, x_2) dx_2 \right] dx_1 \\ &= \int_a^b g(x_1) dx_1, \end{aligned} \tag{12.32}$$

where

$$g(x_1) = \int_{\psi(x_1)}^{\phi(x_1)} f(x_1, x_2) dx_2. \tag{12.33}$$

Let us now apply a Gaussian integration rule to (12.32) using the points $x_1 = z_0, z_1, \dots, z_m$ with the matching coefficients $\omega_0, \omega_1, \dots, \omega_m$,

$$\int_a^b g(x_1) dx_1 \approx \sum_{i=0}^m \omega_i g(z_i).$$

Thus

$$I \approx \sum_{i=0}^m \omega_i \int_{\psi(z_i)}^{\phi(z_i)} f(z_i, x_2) dx_2. \quad (12.34)$$

For the i th of the $m+1$ integrals in (12.34) we can apply a Gaussian integration rule using the points $y_{i0}, y_{i1}, \dots, y_{in}$ with the corresponding coefficients $v_{i0}, v_{i1}, \dots, v_{in}$. We then have

$$\int_{\psi(z_i)}^{\phi(z_i)} f(z_i, x_2) dx_2 \approx \sum_{j=0}^n v_{ij} f(z_i, y_{ij}).$$

Hence,

$$I = \sum_{i=0}^m \sum_{j=0}^n \omega_i v_{ij} f(z_i, y_{ij}).$$

This procedure can obviously be generalized to higher-order multiple integrals. More details can be found in Stroud (1971) and Davis and Rabinowitz (1975, Chapter 5).

The method of Laplace in Section 12.6 can be extended to an n -dimensional integral of the form

$$I(\lambda) = \int_D \varphi(\mathbf{x}) e^{\lambda h(\mathbf{x})} d\mathbf{x},$$

which is a multidimensional version of the integral in (12.24). Here, D is a region in R^n , which may be bounded or unbounded, λ is a large positive constant, and $\mathbf{x} = (x_1, x_2, \dots, x_n)$. As before, it is assumed that:

- a. $\varphi(\mathbf{x})$ is continuous in D .
- b. $h(\mathbf{x})$ has continuous first-order and second-order partial derivatives with respect to $x_1, x_2, \dots, x_n$ in D .
- c. $h(\mathbf{x})$ has a single maximum in D at $\mathbf{x} = \mathbf{t}$.

If $\mathbf{t}$ is an interior point of D , then it is also a stationary point of $h(\mathbf{x})$, that is, $\partial h / \partial x_i|_{\mathbf{x}=\mathbf{t}} = 0$, $i = 1, 2, \dots, n$, since $\mathbf{t}$ is a point of maximum and the

partial derivatives of $h(\mathbf{x})$ exist. Furthermore, the Hessian matrix,

$$\mathbf{H}_h(\mathbf{t}) = \left(\frac{\partial^2 h(\mathbf{x})}{\partial x_i \partial x_j} \right)_{\mathbf{x}=\mathbf{t}},$$

is negative definite. Then, for large λ , $I(\lambda)$ is approximately equal to

$$I(\lambda) \sim \left(\frac{2\pi}{\lambda} \right)^{n/2} \varphi(\mathbf{t}) \{ -\det[\mathbf{H}_h(\mathbf{t})] \}^{-1/2} e^{\lambda h(\mathbf{t})}.$$

A proof of this approximation can be found in Wong (1989, Section 9.5). We note that this expression is a generalization of Laplace's formula (12.25) to an n -tuple Riemann integral.

If $\mathbf{t}$ happens to be on the boundary of D and still satisfies the conditions that $\partial h / \partial x_i|_{\mathbf{x}=\mathbf{t}} = 0$ for $i = 1, 2, \dots, n$, and $\mathbf{H}_h(\mathbf{t})$ is negative definite, then it can be shown that for large λ ,

$$I(\lambda) \sim \frac{1}{2} \left(\frac{2\pi}{\lambda} \right)^{n/2} \varphi(\mathbf{t}) \{ -\det[\mathbf{H}_h(\mathbf{t})] \}^{-1/2} e^{\lambda h(\mathbf{t})},$$

which is one-half of the previous approximation for $I(\lambda)$ (see Wong, 1989, page 498).

12.8. THE MONTE CARLO METHOD

A new approach to approximate integration arose in the 1940s as part of the Monte Carlo method of S. Ulam and J. von Neumann (Haber, 1970). The basic idea of the Monte Carlo method for integrals is described as follows: suppose that we need to compute the integral

$$I = \int_a^b f(x) dx. \quad (12.35)$$

We consider I as the expected value of a certain stochastic process. An estimate of I can be obtained by random sampling from this process, and the estimate is then used as an approximation to I . For example, let X be a continuous random variable that has the uniform distribution $U(a, b)$ over the interval $[a, b]$. The expected value of $f(X)$ is

$$\begin{aligned} E[f(X)] &= \frac{1}{b-a} \int_a^b f(x) dx \\ &= \frac{I}{b-a}. \end{aligned}$$

Let $x_1, x_2, \dots, x_n$ be a random sample from $U(a, b)$. An estimate of $E[f(X)]$ is given by $(1/n)\sum_{i=1}^n f(x_i)$. Hence, an approximate value of I , denoted by $\hat{I}_n$, can be obtained as

$$\hat{I}_n = \frac{b-a}{n} \sum_{i=1}^n f(x_i). \quad (12.36)$$

The justification for using $\hat{I}_n$ as an approximation to I is that $\hat{I}_n$ is a consistent estimator of I , that is, for a given $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} P(|\hat{I}_n - I| \geq \epsilon) = 0.$$

This is true because $(1/n)\sum_{i=1}^n f(x_i)$ converges in probability to $E[f(X)]$ as $n \rightarrow \infty$, according to the law of large numbers (see Sections 3.7 and 5.6.3). In other words, the probability that $\hat{I}_n$ will be different from I can be made arbitrarily close to zero if n is chosen large enough. In fact, we even have the stronger result that $\hat{I}_n$ converges strongly, or almost surely, to I , as $n \rightarrow \infty$, by the strong law of large numbers (see Section 5.6.3).

The closeness of $\hat{I}_n$ to I depends on the variance of $\hat{I}_n$, which is equal to

$$\begin{aligned} \text{Var}(\hat{I}_n) &= (b-a)^2 \text{Var}\left[\frac{1}{n} \sum_{i=1}^n f(x_i)\right] \\ &= \frac{(b-a)^2 \sigma_f^2}{n}, \end{aligned} \quad (12.37)$$

where σ_f^2 is the variance of the random variable $f(X)$, that is,

$$\begin{aligned} \sigma_f^2 &= \text{Var}[f(X)] \\ &= E[f^2(X)] - \{E[f(X)]\}^2 \\ &= \frac{1}{b-a} \int_a^b f^2(x) dx - \left(\frac{I}{b-a}\right)^2. \end{aligned} \quad (12.38)$$

By the central limit theorem (see Section 4.5.1), if n is large enough, then $\hat{I}_n$ is approximately normally distributed with mean I and variance $(1/n)(b-a)^2\sigma_f^2$. Thus,

$$\frac{\hat{I}_n - I}{(b-a)\sigma_f/\sqrt{n}} \xrightarrow{d} Z,$$

where Z has the standard normal distribution, and the symbol $\xrightarrow{d}$ denotes

convergence in distribution (see Section 4.5.1). It follows that for a given $\tau > 0$,

$$P\left[|\hat{I}_n - I| \leq \frac{\tau}{\sqrt{n}}(b-a)\sigma_f\right] \approx \frac{1}{\sqrt{2\pi}} \int_{-\tau}^{\tau} e^{-x^2/2} dx. \quad (12.39)$$

The right-hand side of (12.39) is the probability that a standard normal distribution attains values between $-\tau$ and τ . Let us denote this probability by $1 - \alpha$. Then $\tau = z_{\alpha/2}$, which is the upper $(\alpha/2)$ 100th percentile of the standard normal distribution. If we denote the error of approximation, $\hat{I}_n - I$, by E_n , then formula (12.39) indicates that for large n ,

$$|E_n| \leq \frac{1}{\sqrt{n}}(b-a)\sigma_f z_{\alpha/2} \quad (12.40)$$

with an approximate probability equal to $1 - \alpha$, which is called the confidence coefficient. Thus, for a fixed α , the error bound in (12.40) is proportional to σ_f and is inversely proportional to $\sqrt{n}$. For example, if $1 - \alpha = 0.90$, then $z_{\alpha/2} = 1.645$, and

$$|E_n| \leq \frac{1.645}{\sqrt{n}}(b-a)\sigma_f$$

with an approximate confidence coefficient equal to 0.90. Also, if $1 - \alpha = 0.95$, then $z_{\alpha/2} = 1.96$, and

$$|E_n| \leq \frac{1.96}{\sqrt{n}}(b-a)\sigma_f$$

with an approximate confidence coefficient equal to 0.95.

In order to compute the error bound in (12.40), an estimate of σ_f^2 is needed. Using (12.38) and the random sample $x_1, x_2, \dots, x_n$, an estimate of σ_f^2 is given by

$$\hat{\sigma}_{fn}^2 = \frac{1}{n} \sum_{i=1}^n f^2(x_i) - \left[\frac{1}{n} \sum_{i=1}^n f(x_i) \right]^2. \quad (12.41)$$

12.8.1. Variance Reduction

In order to increase the accuracy of the Monte Carlo approximation of I , the error bound in (12.40) should be reduced for a fixed value of α . We can achieve this by increasing n . Alternatively, we can reduce σ_f by considering a distribution other than the uniform distribution. This can be accomplished by using the so-called method of *importance sampling*, a description of which follows.

Let $g(x)$ be a density function that is positive over the interval $[a, b]$. Thus, $g(x) > 0$, $a \leq x \leq b$, and $\int_a^b g(x) dx = 1$. The integral in (12.35) can be written as

$$I = \int_a^b \frac{f(x)}{g(x)} g(x) dx. \quad (12.42)$$

In this case, I is the expected value of $f(X)/g(X)$, where X is a continuous random variable with the density function $g(x)$. Using now a random sample, $x_1, x_2, \dots, x_n$, from this distribution, an estimate of I can be obtained as

$$\hat{I}_n^* = \frac{1}{n} \sum_{i=1}^n \frac{f(x_i)}{g(x_i)}. \quad (12.43)$$

The variance of $\hat{I}_n^*$ is then given by

$$\text{Var}(\hat{I}_n^*) = \frac{\sigma_{fg}^2}{n},$$

where

$$\begin{aligned} \sigma_{fg}^2 &= \text{Var} \left[\frac{f(X)}{g(X)} \right] \\ &= E \left\{ \left[\frac{f(X)}{g(X)} \right]^2 \right\} - \left\{ E \left[\frac{f(X)}{g(X)} \right] \right\}^2 \\ &= \int_a^b \frac{f^2(x)}{g^2(x)} g(x) dx - I^2. \end{aligned} \quad (12.44)$$

As before, the error bound can be derived on the basis of the central limit theorem, using the fact that $\hat{I}_n^*$ is approximately normally distributed with mean I and variance $(1/n)\sigma_{fg}^2$ for large n . Hence, as in (12.40),

$$|E_n^*| \leq \frac{1}{\sqrt{n}} \sigma_{fg} z_{\alpha/2}$$

with an approximate probability equal to $1 - \alpha$, where $E_n^* = \hat{I}_n^* - I$. The density $g(x)$ should therefore be chosen so that an error bound smaller than

Table 12.2. Approximate Values of $I = \int_1^2 x^2 dx$ Using Formula (12.36)

n	$\hat{I}_n$
50	2.3669
100	2.5087
150	2.2227
200	2.3067
250	2.3718
300	2.3115
350	2.3366

the one in (12.40) can be achieved. For example, if $f(x) > 0$, and if

$$g(x) = \frac{f(x)}{\int_a^b f(x) dx}, \quad (12.45)$$

then $\sigma_{fg}^2 = 0$ as can be seen from formula (12.44).

Unfortunately, since the exact value of $\int_a^b f(x) dx$ is the one we seek to compute, formula (12.45) cannot be used. However, by choosing $g(x)$ to behave approximately as $f(x)$ [assuming $f(x)$ is positive], we should expect a reduction in the variance. Note that the generation of random variables from the $g(x)$ distribution is more involved than just using a random sample from the uniform distribution $U(a, b)$.

EXAMPLE 12.8.1. Consider the integral $I = \int_1^2 x^2 dx = \frac{7}{3} \approx 2.3333$. Suppose that we use a sample of 50 points from the uniform distribution $U(1, 2)$. Applying formula (12.36), we obtain $\hat{I}_{50} = 2.3669$. Repeating this process several times using higher values of n , we obtain the values in Table 12.2.

EXAMPLE 12.8.2. Let us now apply the method of importance sampling to the integral $I = \int_0^1 e^x dx \approx 1.7183$. Consider the density function $g(x) = \frac{2}{3}(1+x)$ over the interval $[0, 1]$. Using the method described in Section 3.7, a random sample, $x_1, x_2, \dots, x_n$, can be generated from this distribution as follows: the cumulative distribution function for $g(x)$ is $y = G(x) = P[X \leq x]$, that is,

$$\begin{aligned} y = G(x) &= \int_0^x g(t) dt \\ &= \frac{2}{3} \int_0^x (1+t) dt \\ &= \frac{2}{3} \left(x + \frac{x^2}{2} \right), \quad 0 \leq x \leq 1. \end{aligned}$$

Table 12.3. Approximate Values of $I = \int_0^1 e^x \, dx$

n	$\hat{I}_n^*$ [Formula (12.46)]	$\hat{I}_n$ [Formula (12.36)]
50	1.7176	1.7156
100	1.7137	1.7025
150	1.7063	1.6854
200	1.7297	1.7516
250	1.7026	1.6713
300	1.7189	1.7201
350	1.7093	1.6908
400	1.7188	1.7192

The only solution of $y = G(x)$ in $[0, 1]$ is

$$x = -1 + (1 + 3y)^{1/2}, \quad 0 \leq y \leq 1.$$

Hence, the inverse function of $G(x)$ is

$$G^{-1}(y) = -1 + (1 + 3y)^{1/2}, \quad 0 \leq y \leq 1.$$

If $y_1, y_2, \dots, y_n$ form a random sample of n values from the uniform distribution $U(0, 1)$, then $x_1 = G^{-1}(y_1), x_2 = G^{-1}(y_2), \dots, x_n = G^{-1}(y_n)$ will form a sample from the distribution with the density function $g(x)$. Formula (12.43) can then be applied to approximate the value of I using the estimate

$$\hat{I}_n^* = \frac{3}{2n} \sum_{i=1}^n \frac{e^{x_i}}{1 + x_i}. \tag{12.46}$$

Table 12.3 gives $\hat{I}_n^*$ for several values of n . For the sake of comparison, values of $\hat{I}_n$ from formula (12.36) [with $f(x) = e^x, a = 0, b = 1$] were also computed using a sample from the uniform distribution $U(0, 1)$. The results are shown in Table 12.3. We note that the values of $\hat{I}_n^*$ are more stable and closer to the true value of I than those of $\hat{I}_n$.

12.8.2. Integrals in Higher Dimensions

The Monte Carlo method can be extended to multidimensional integrals. Consider computing the integral $I = \int_D f(\mathbf{x}) \, d\mathbf{x}$, where D is a bounded region in R^n , the n -dimensional Euclidean space, and $\mathbf{x} = (x_1, x_2, \dots, x_n)$. As before, we consider I as the expected value of a stochastic process having a certain distribution over D . For example, we may take $\mathbf{X}$ to be a continuous random vector uniformly distributed over D . By this we mean that the probability of $\mathbf{X}$ being in D is $1/\nu(D)$, where $\nu(D)$ denotes the volume of D ,

and the probability of $\mathbf{X}$ being outside D is zero. Hence, the expected value of $f(\mathbf{X})$ is

$$\begin{aligned} E[f(\mathbf{X})] &= \frac{1}{v(D)} \int_D f(\mathbf{x}) d\mathbf{x} \\ &= \frac{I}{v(D)}. \end{aligned}$$

The variance of $f(\mathbf{X})$, denoted by σ_f^2 , is

$$\begin{aligned} \sigma_f^2 &= E[f^2(\mathbf{X})] - \{E[f(\mathbf{X})]\}^2 \\ &= \frac{1}{v(D)} \int_D f^2(\mathbf{x}) d\mathbf{x} - \left[\frac{I}{v(D)} \right]^2. \end{aligned}$$

Let us now take a sample of N independent observations on $\mathbf{X}$, namely, $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N$. Then a consistent estimator of $E[f(\mathbf{X})]$ is $(1/N)\sum_{i=1}^N f(\mathbf{X}_i)$, and hence, an estimate of I is given by

$$\hat{I}_N = \frac{v(D)}{N} \sum_{i=1}^N f(\mathbf{X}_i),$$

which can be used to approximate I . The variance of $\hat{I}_N$ is

$$\text{Var}(\hat{I}_N) = \frac{\sigma_f^2}{N} v^2(D). \quad (12.47)$$

If N is large enough, then by the central limit theorem, $\hat{I}_N$ is approximately normally distributed with mean I and variance as in (12.47). It follows that

$$|E_N| \leq \frac{v(D)}{\sqrt{N}} \sigma_f z_{\alpha/2}$$

with an approximate probability equal to $1 - \alpha$, where $E_N = I - \hat{I}_N$ is the error of approximation. This formula is analogous to formula (12.40).

The method of importance sampling can also be applied here to reduce the error of approximation. The application of this method is similar to the case of a single-variable integral as seen earlier.

12.9. APPLICATIONS IN STATISTICS

Approximation of integrals is a problem of substantial concern for statisticians. The statistical literature in this area has grown significantly in the last 20 years, particularly in connection with integrals that arise in Bayesian

statistics. Evans and Swartz (1995) presented a survey of the major techniques and approaches available for the numerical approximation of integrals in statistics. The proceedings edited by Flournoy and Tsutakawa (1991) includes several interesting articles on statistical multiple integration, including a detailed description of available software to compute multidimensional integrals (see the article by Kahaner, 1991, page 9).

12.9.1. The Gauss-Hermite Quadrature

The Gauss-Hermite quadrature mentioned earlier in Section 12.5 is often used for numerical integration in statistics because of its relation to Gaussian (normal) densities. We recall that this quadrature is defined in terms of integrals of the form $\int_{-\infty}^{\infty} e^{-x^2} f(x) dx$. Using formula (12.20), we have approximately

$$\int_{-\infty}^{\infty} e^{-x^2} f(x) dx \approx \sum_{i=0}^n \omega_i f(x_i),$$

where the x_i 's are the zeros of the Hermite polynomial $H_{n+1}(x)$ of degree $n+1$, and the ω_i 's are suitably corresponding weights. Tables of x_i and ω_i values are given by Abramowitz and Stegun (1972, page 924) and by Krylov (1962, Appendix B).

Liu and Pierce (1994) applied the Gauss-Hermite quadrature to integrals of the form $\int_{-\infty}^{\infty} g(t) dt$, which can be expressed in the form

$$\int_{-\infty}^{\infty} g(t) dt = \int_{-\infty}^{\infty} f(t) \phi(t, \mu, \sigma) dt,$$

where $\phi(t, \mu, \sigma)$ is the normal density

$$\phi(t, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(t - \mu)^2\right],$$

and $f(t) = g(t)/\phi(t, \mu, \sigma)$. Thus,

$$\begin{aligned} \int_{-\infty}^{\infty} g(t) dt &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} f(t) \exp\left[-\frac{1}{2\sigma^2}(t - \mu)^2\right] dt \\ &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} f(\mu + \sqrt{2}\sigma x) e^{-x^2} dx \\ &\approx \frac{1}{\sqrt{\pi}} \sum_{i=0}^n \omega_i f(\mu + \sqrt{2}\sigma x_i), \end{aligned} \tag{12.48}$$

where the x_i 's are the zeros of the Hermite polynomial $H_{n+1}(x)$ of degree $n + 1$. We may recall that the x_i 's and ω_i 's are chosen so that this approximation will be exact for all polynomials of degrees not exceeding $2n + 1$. For this reason, Liu and Pierce (1994) recommend choosing μ and σ in (12.48) so that $f(t)$ is well approximated by a low-order polynomial in the region where the values of $\mu + \sqrt{2} \sigma x_i$ are taken. More specifically, the $(n + 1)$ th-order Gauss-Hermite quadrature in (12.48) will be highly effective if the ratio of $g(t)$ to the normal density $\phi(t, \mu, \sigma^2)$ can be well approximated by a polynomial of degree not exceeding $2n + 1$ in the region where $g(t)$ is substantial. This arises frequently, for example, when $g(t)$ is a likelihood function [if $g(t) > 0$], or the product of a likelihood function and a normal density, as was pointed out by Liu and Pierce (1994), who gave several examples to demonstrate the usefulness of the approximation in (12.48).

12.9.2. Minimum Mean Squared Error Quadrature

Correlated observations may arise in some experimental work (Piegorsch and Bailer, 1993). Consider, for example, the model

$$y_{qr} = f(t_q) + \epsilon_{qr},$$

where y_{qr} represents the observed value from experimental unit r at time t_q ($q = 0, 1, \dots, m$; $r = 1, 2, \dots, n$), $f(t_q)$ is the underlying response function, and ϵ_{qr} is a random error term. It is assumed that $E(\epsilon_{qr}) = 0$, $\text{Cov}(\epsilon_{pr}, \epsilon_{qr}) = \sigma_{pq}$, and $\text{Cov}(\epsilon_{pr}, \epsilon_{qs}) = 0$ ($r \neq s$) for all p, q . The area under the response curve over the interval $t_0 < t < t_m$ is

$$A = \int_{t_0}^{t_m} f(t) dt. \quad (12.49)$$

This is an important measure for a variety of experimental situations, including the assessment of chemical bioavailability in drug disposition studies (Gibaldi and Perrier, 1982, Chapters 2, 7) and other clinical settings. If the functional form of $f(\cdot)$ is unknown, the integral in (12.49) is estimated by numerical methods. This is accomplished using a quadrature approximation of the integral in (12.49). By definition, a quadrature estimator of this integral is

$$\hat{A} = \sum_{q=0}^m \phi_q \hat{f}_q, \quad (12.50)$$

where $\hat{f}_q$ is some unbiased estimator of $f_q = f(t_q)$ with $\text{Cov}(\hat{f}_p, \hat{f}_q) = \sigma_{pq}/n$, and the ϕ_q 's form a set of quadrature coefficients.

The expected value of $\hat{A}$,

$$E(\hat{A}) = \sum_{q=0}^m \phi_q f_q,$$

may not necessarily be equal to A due to the quadrature approximation employed in calculating $\hat{A}$. The bias in estimating A is

$$\text{bias} = E(\hat{A}) - A,$$

and the mean squared error (MSE) of $\hat{A}$ is given by

$$\begin{aligned} \text{MSE}(\hat{A}) &= \text{Var } \hat{A} + [E(\hat{A}) - A]^2 \\ &= \sum_{p=0}^m \sum_{q=0}^m \phi_p \phi_q \frac{\sigma_{pq}}{n} + \left(\sum_{q=0}^m \phi_q f_q - A \right)^2, \end{aligned}$$

which can be written as

$$\text{MSE}(\hat{A}) = \frac{1}{n} \boldsymbol{\phi}' \mathbf{V} \boldsymbol{\phi} + (\mathbf{f}' \boldsymbol{\phi} - A)^2,$$

where $\mathbf{V} = (\sigma_{pq})$, $\mathbf{f} = (f_0, f_1, \dots, f_m)'$, and $\boldsymbol{\phi}' = (\phi_0, \phi_1, \dots, \phi_m)$. Hence,

$$\text{MSE}(\hat{A}) = \boldsymbol{\phi}' \left[\frac{1}{n} \mathbf{V} + \mathbf{f} \mathbf{f}' \right] \boldsymbol{\phi} - 2A \mathbf{f}' \boldsymbol{\phi} + A^2. \quad (12.51)$$

Let us now seek an optimum value of $\boldsymbol{\phi}$ that minimizes $\text{MSE}(\hat{A})$ in (12.51). For this purpose, we equate the gradient of $\text{MSE}(\hat{A})$ with respect to $\boldsymbol{\phi}$, namely $\nabla_{\boldsymbol{\phi}} \text{MSE}(\hat{A})$, to zero. We obtain

$$\nabla_{\boldsymbol{\phi}} \text{MSE}(\hat{A}) = 2 \left[\left(\frac{1}{n} \mathbf{V} + \mathbf{f} \mathbf{f}' \right) \boldsymbol{\phi} - A \mathbf{f} \right] = \mathbf{0}. \quad (12.52)$$

In order for this equation to have a solution, the vector $A \mathbf{f}$ must belong to the column space of the matrix $(1/n) \mathbf{V} + \mathbf{f} \mathbf{f}'$. Note that this matrix is positive semidefinite. If its inverse exists, then it will be positive definite, and the only solution, $\boldsymbol{\phi}^*$, to (12.52), namely,

$$\boldsymbol{\phi}^* = A \left(\frac{1}{n} \mathbf{V} + \mathbf{f} \mathbf{f}' \right)^{-1} \mathbf{f}, \quad (12.53)$$

yields a unique minimum of $\text{MSE}(\hat{A})$ (see Section 7.7). Using $\boldsymbol{\phi}^*$ in (12.50),

we obtain the following estimate of A :

$$\begin{aligned}\hat{A}^* &= \hat{\mathbf{f}}' \boldsymbol{\Phi}^* \\ &= A \hat{\mathbf{f}}' \left[\frac{1}{n} \mathbf{V} + \mathbf{f} \mathbf{f}' \right]^{-1} \mathbf{f}.\end{aligned}\quad (12.54)$$

Using the Sherman–Morrison formula (see Exercise 2.14), (12.54) can be written as

$$\hat{A}^* = nA \left[\hat{\mathbf{f}}' \mathbf{V}^{-1} \mathbf{f} - \frac{n \hat{\mathbf{f}}' \mathbf{V}^{-1} \mathbf{f} \mathbf{f}' \mathbf{V}^{-1} \mathbf{f}}{1 + n \mathbf{f}' \mathbf{V}^{-1} \mathbf{f}} \right], \quad (12.55)$$

where $\hat{\mathbf{f}} = (\hat{f}_0, \hat{f}_1, \dots, \hat{f}_m)'$. We refer to $\hat{A}^*$ as an MSE-optimal estimator of A . Replacing $\mathbf{f}$ with its unbiased estimator $\hat{\mathbf{f}}$, and A with some initial estimate, say $A_0 = \hat{\mathbf{f}}' \boldsymbol{\Phi}_0$, where $\boldsymbol{\Phi}_0$ is an initial value of $\boldsymbol{\Phi}$, yields the approximate MSE-optimal estimator

$$\hat{A}^{**} = \frac{c}{1+c} A_0, \quad (12.56)$$

where c is the quadratic form,

$$c = n \hat{\mathbf{f}}' \mathbf{V}^{-1} \hat{\mathbf{f}}.$$

This estimator has the form $\hat{\mathbf{f}}' \tilde{\boldsymbol{\Phi}}$, where $\tilde{\boldsymbol{\Phi}} = [c/(1+c)] \boldsymbol{\Phi}_0$. Since $c \geq 0$, $\hat{A}^{**}$ provides a shrinkage of the initial estimate in A_0 toward zero. This creates a biased estimator of A with a smaller variance, which results in an overall reduction in MSE. A similar approach is used in ridge regression to estimate the parameters of a linear model when the columns of the corresponding matrix $\mathbf{X}$ are multicollinear (see Section 5.6.6).

We note that this procedure requires knowledge of $\mathbf{V}$. In many applications, it may not be possible to specify $\mathbf{V}$. Piegorsch and Bailer (1993) used an estimate of $\mathbf{V}$ when $\mathbf{V}$ was assumed to have certain particular patterns, such as $\mathbf{V} = \sigma^2 \mathbf{I}$ or $\mathbf{V} = \sigma^2[(1-\rho)\mathbf{I} + \rho\mathbf{J}]$, where $\mathbf{I}$ and $\mathbf{J}$ are, respectively, the identity matrix and the square matrix of ones, and σ^2 and ρ are unknown constants. For example, under the equal variance assumption $\mathbf{V} = \sigma^2 \mathbf{I}$, the ratio $c/(1+c)$ is equal to

$$\frac{c}{1+c} = \left(1 + \frac{\sigma^2}{n \hat{\mathbf{f}}' \hat{\mathbf{f}}} \right)^{-1}.$$

If $\hat{\mathbf{f}}$ is normally distributed, $\hat{\mathbf{f}} \sim N[\mathbf{f}, (1/n)\mathbf{V}]$ —as is the case when $\hat{\mathbf{f}}$ is a vector of means, $\hat{\mathbf{f}} = (\bar{y}_0, \bar{y}_1, \dots, \bar{y}_m)'$ with $\bar{y}_q = (1/n) \sum_{r=1}^n y_{qr}$ —then an unbiased

estimate of σ^2 is given by

$$s^2 = \frac{1}{(m+1)(n-1)} \sum_{q=0}^m \sum_{r=1}^n (y_{qr} - \bar{y}_q)^2.$$

Substituting s^2 in place of σ^2 in (12.56), we obtain the area estimator

$$\hat{A}^{**} = A_0 \left(1 + \frac{s^2}{n \hat{\mathbf{f}}' \hat{\mathbf{f}}} \right)^{-1}.$$

A similar quadrature approximation of A in (12.49) was considered earlier by Katz and D'Argenio (1983) using a trapezoidal approximation to A . The quadrature points were selected so that they minimize the expectation of the square of the difference between the exact integral and the quadrature approximation. This approach was applied to simulated pharmacokinetic problems.

12.9.3. Moments of a Ratio of Quadratic Forms

Consider the ratio of quadratic forms,

$$Q = \frac{\mathbf{y}'\mathbf{A}\mathbf{y}}{\mathbf{y}'\mathbf{B}\mathbf{y}}, \quad (12.57)$$

where $\mathbf{A}$ and $\mathbf{B}$ are symmetric matrices, $\mathbf{B}$ is positive definite, and $\mathbf{y}$ is an $n \times 1$ random vector. Ratios such as Q are frequently encountered in statistics and econometrics. In general, their exact distributions are mathematically intractable, especially when the quadratic forms are not independent. For this reason, the derivation of the moments of such ratios, for the purpose of approximating their distributions, is of interest. Sutradhar and Bartlett (1989) obtained approximate expressions for the first four moments of the ratio Q for a normally distributed $\mathbf{y}$. The moments were utilized to approximate the distribution of Q . This approximation was then applied to calculate the percentile points of a modified F -test statistic for testing treatment effects in a one-way model under correlated observations. Morin (1992) derived exact, but complicated, expressions for the first four moments of Q for a normally distributed $\mathbf{y}$. The moments are expressed in terms of confluent hypergeometric functions of many variables.

If $\mathbf{y}$ is not normally distributed, then no tractable formulas exist for the moments of Q in (12.57). Hence, manageable and computable approximations for these moments would be helpful. Lieberman (1994) used the method of Laplace to provide general approximations for the moments of Q without making the normality assumption on $\mathbf{y}$. Lieberman showed that if $E(\mathbf{y}'\mathbf{B}\mathbf{y})$ and $E[(\mathbf{y}'\mathbf{A}\mathbf{y})^k]$ exist for $k \geq 1$, then the Laplace approximation of

$E(Q^k)$, the k th noncentral moment of Q , is given by

$$E_L(Q^k) = \frac{E[(\mathbf{y}'\mathbf{A}\mathbf{y})^k]}{[E(\mathbf{y}'\mathbf{B}\mathbf{y})]^k}. \quad (12.58)$$

In particular, if $\mathbf{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then the Laplace approximation for the mean and second noncentral moment of Q are written explicitly as

$$\begin{aligned} E_L(Q) &= \frac{\text{tr}(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}}{\text{tr}(\mathbf{B}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\mathbf{B}\boldsymbol{\mu}} \\ E_L(Q^2) &= \frac{E[(\mathbf{y}'\mathbf{A}\mathbf{y})^2]}{[E(\mathbf{y}'\mathbf{B}\mathbf{y})]^2} \\ &= \frac{\text{Var}(\mathbf{y}'\mathbf{A}\mathbf{y}) + [E(\mathbf{y}'\mathbf{A}\mathbf{y})]^2}{[E(\mathbf{y}'\mathbf{B}\mathbf{y})]^2} \\ &= \frac{2\text{tr}[(\mathbf{A}\boldsymbol{\Sigma})^2] + 4\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}\boldsymbol{\mu} + [\text{tr}(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}]^2}{[\text{tr}(\mathbf{B}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\mathbf{B}\boldsymbol{\mu}]^2} \end{aligned}$$

(see Searle, 1971, Section 2.5 for expressions for the mean and variance of a quadratic form in normal variables).

EXAMPLE 12.9.1. Consider the linear model, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\mathbf{X}$ is $n \times p$ of rank p , $\boldsymbol{\beta}$ is a vector of unknown parameters, and $\boldsymbol{\epsilon}$ is a random error vector. Under the null hypothesis of no serial correlation in the elements of $\boldsymbol{\epsilon}$, we have $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$. The corresponding Durbin-Watson (1950, 1951) test statistic is given by

$$d = \frac{\boldsymbol{\epsilon}'\mathbf{P}\mathbf{A}_1\mathbf{P}\boldsymbol{\epsilon}}{\boldsymbol{\epsilon}'\mathbf{P}\boldsymbol{\epsilon}},$$

where $\mathbf{P} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, and $\mathbf{A}_1$ is the matrix

$$\mathbf{A}_1 = \begin{bmatrix} 1 & -1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & 2 & -1 & 0 \\ 0 & 0 & \cdots & 0 & -1 & 2 & -1 \\ 0 & 0 & \cdots & 0 & 0 & -1 & 1 \end{bmatrix}.$$

Then, by (12.58), the Laplace approximation of $E(d)$ is

$$E_L(d) = \frac{E(\boldsymbol{\epsilon}' \mathbf{P} \mathbf{A}_1 \mathbf{P} \boldsymbol{\epsilon})}{E(\boldsymbol{\epsilon}' \mathbf{P} \boldsymbol{\epsilon})}. \quad (12.59)$$

Durbin and Watson (1951) showed that d is distributed independently of its own denominator, so that the moments of the ratio d are the ratios of the corresponding moments of the numerator and denominator, that is,

$$E(d^k) = \frac{E[(\boldsymbol{\epsilon}' \mathbf{P} \mathbf{A}_1 \mathbf{P} \boldsymbol{\epsilon})^k]}{E[(\boldsymbol{\epsilon}' \mathbf{P} \boldsymbol{\epsilon})^k]}. \quad (12.60)$$

From (12.59) and (12.60) we note that the Laplace approximation for the mean, $E(d)$, is exact. For $k \geq 2$, Lieberman (1994) showed that

$$\frac{E_L(d^k)}{E(d^k)} = 1 + O\left(\frac{1}{n}\right).$$

Thus, the relative error of approximating higher-order moments of d is $O(1/n)$, regardless of the matrix $\mathbf{X}$.

12.9.4. Laplace's Approximation in Bayesian Statistics

Suppose (Kass, Tierney, and Kadane, 1991) that a data vector $\mathbf{y} = (y_1, y_2, \dots, y_n)'$ has a distribution with the density function $p(\mathbf{y}|\theta)$, where θ is an unknown parameter. Let $L(\theta)$ denote the corresponding likelihood function, which is proportional to $p(\mathbf{y}|\theta)$. In Bayesian statistics, a prior density, $\pi(\theta)$, is assumed on θ , and inferences are based on the posterior density $q(\theta|\mathbf{y})$, which is proportional to $L(\theta)\pi(\theta)$, where the proportionality constant is determined by requiring that $q(\theta|\mathbf{y})$ integrate to one. For a given real-valued function $g(\theta)$, its posterior expectation is given by

$$E[g(\theta)|\mathbf{y}] = \frac{\int g(\theta) L(\theta) \pi(\theta) d\theta}{\int L(\theta) \pi(\theta) d\theta}. \quad (12.61)$$

Tierney, Kass, and Kadane (1989) expressed the integrands in (12.61) as follows:

$$\begin{aligned} g(\theta) L(\theta) \pi(\theta) &= b_N(\theta) \exp[-nh_N(\theta)], \\ L(\theta) \pi(\theta) &= b_D(\theta) \exp[-nh_D(\theta)], \end{aligned}$$

where $b_N(\theta)$ and $b_D(\theta)$ are smooth functions that do not depend on n and $h_N(\theta)$ and $h_D(\theta)$ are constant-order functions of n , as $n \rightarrow \infty$. Formula (12.61) can then be written as

$$E[g(\theta)|\mathbf{y}] = \frac{\int b_N(\theta) \exp[-nh_N(\theta)] d\theta}{\int b_D(\theta) \exp[-nh_D(\theta)] d\theta}. \quad (12.62)$$

Applying Laplace's approximation in (12.25) to the integrals in (12.62), we obtain, if n is large,

$$E[g(\theta)|\mathbf{y}] \approx \left[\frac{h''_D(\hat{\theta}_D)}{h''_N(\hat{\theta}_N)} \right]^{1/2} \frac{b_N(\hat{\theta}_N) \exp[-nh_N(\hat{\theta}_N)]}{b_D(\hat{\theta}_D) \exp[-nh_D(\hat{\theta}_D)]}, \quad (12.63)$$

where $\hat{\theta}_N$ and $\hat{\theta}_D$ are the locations of the single maxima of $-h_N(\theta)$ and $-h_D(\theta)$, respectively. In particular, if we choose $h_N(\theta) = h_D(\theta) = -(1/n)\log[L(\theta)\pi(\theta)]$, $b_N(\theta) = g(\theta)$ and $b_D(\theta) = 1$, then (12.63) reduces to

$$E[g(\theta)|\mathbf{y}] \approx g(\hat{\theta}), \quad (12.64)$$

where $\hat{\theta}$ is the point at which $(1/n)\log[L(\theta)\pi(\theta)]$ attains its maximum.

Formula (12.64) provides a first-order approximation of $E[g(\theta)|\mathbf{y}]$. This approximation is often called the *modal approximation* because θ is the mode of the posterior density. A more accurate second-order approximation of $E[g(\theta)|\mathbf{y}]$ was given by Kass, Tierney, and Kadane (1991).

12.9.5. Other Methods of Approximating Integrals in Statistics

There are several major techniques and approaches available for the numerical approximation of integrals in statistics that are beyond the scope of this book. These techniques, which include the *saddlepoint approximation* and *Markov chain Monte Carlo*, have received a great deal of attention in the statistical literature in recent years.

The saddlepoint method is designed to approximate integrals of the Laplace type in which both the integrand and contour of integration are allowed to be complex valued. It is a powerful tool for obtaining accurate expressions for densities and distribution functions. A good introduction to the basic principles underlying this method was given by De Bruijn (1961, Chapter 5). Daniels (1954) is credited with having introduced it in statistics in the context of approximating the density of a sample mean of independent and identically distributed random variables.

Markov chain Monte Carlo (MCMC) is a general method for the simulation of stochastic processes having probability densities known up to a

constant of proportionality. It generally deals with high-dimensional statistical problems, and has come into prominence in statistical applications during the past several years. Although MCMC has potential applications in several areas of statistics, most attention to date has been focused on Bayesian applications.

For a review of these techniques, see, for example, Geyer (1992), Evans and Swartz (1995), Goutis and Casella (1999), and Strawderman (2000).

FURTHER READING AND ANNOTATED BIBLIOGRAPHY

- Abramowitz, M., and I. A. Stegun, eds. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Wiley, New York. (This volume is an excellent source for a wide variety of numerical tables of mathematical functions. Chap. 25 gives zeros of Legendre, Hermite, and Laguerre polynomials along with their corresponding weight factors.)
- Copson, E. T. (1965). *Asymptotic Expansions*. Cambridge University Press, London. (The method of Laplace is discussed in Chap. 5.)
- Courant, R., and F. John (1965). *Introduction to Calculus and Analysis*, Volume 1. Wiley, New York. (The trapezoidal and Simpson's methods are discussed in Chap. 6.)
- Daniels, H. (1954). "Saddlepoint approximation in statistics." *Ann. Math. Statist.*, **25**, 631–650.
- Davis, P. J., and P. Rabinowitz (1975). *Methods of Numerical Integration*. Academic Press, New York. (This book presents several useful numerical integration methods, including approximate integrations over finite or infinite intervals as well as integration in two or more dimensions.)
- De Bruijn, N. G. (1961). *Asymptotic Methods in Analysis*, 2nd ed. North-Holland, Amsterdam. (Chap. 4 covers the method of Laplace, and the saddlepoint method is the topic of Chap. 5.)
- Durbin, J., and G. S. Watson (1950). "Testing for serial correlation in least squares regression, I." *Biometrika*, **37**, 409–428.
- Durbin, J., and G. S. Watson (1951). "Testing for serial correlation in least squares regression, II." *Biometrika*, **38**, 159–178.
- Evans, M., and T. Swartz (1995). "Methods for approximating integrals in statistics with special emphasis on Bayesian integration problems." *Statist. Sci.*, **10**, 254–272.
- Flournoy, N., and R. K. Tsutakawa, eds. (1991). *Statistical Multiple Integration, Contemporary Mathematics* **115**. Amer. Math. Soc., Providence, Rhode Island. (This volume contains the proceedings of an AMS–IMS–SIAM joint summer research conference on statistical multiple integration, which was held at Humboldt University, Arcata, California, June 17–23, 1989.)
- Fulks, W. (1978). *Advanced Calculus*, 3rd ed. Wiley, New York. (Section 18.3 of this book contains proofs associated with the method of Laplace.)
- Geyer, C. J. (1992). "Practical Markov chain Monte Carlo." *Statist. Sci.*, **7**, 473–511.

- Ghazal, G. A. (1994). "Moments of the ratio of two dependent quadratic forms." *Statist. Prob. Letters*, **20**, 313–319.
- Gibaldi, M., and D. Perrier (1982). *Pharmacokinetics*, 2nd ed. Dekker, New York.
- Goutis, C., and G. Casella (1999). "Explaining the saddlepoint approximation." *Amer. Statist.*, **53**, 216–224.
- Haber, S. (1970). "Numerical evaluation of multiple integrals." *SIAM Rev.*, **12**, 481–526.
- Kahaner, D. K. (1991). "A survey of existing multidimensional quadrature routines." In *Statistical Multiple Integration*, Contemporary Mathematics **115**, N. Flournoy and R. K. Tsutakawa, eds. Amer. Math. Soc., Providence, pp. 9–22.
- Kass, R. E., L. Tierney, and J. B. Kadane (1991). "Laplace's method in Bayesian analysis." In *Statistical Multiple Integration*, Contemporary Mathematics **115**, N. Flournoy and R. K. Tsutakawa, eds. Amer. Math. Soc., Providence, pp. 89–99.
- Katz, D., and D. Z. D'Argenio (1983). "Experimental design for estimating integrals by numerical quadrature, with applications to pharmacokinetic studies." *Biometrics*, **39**, 621–628.
- Krylov, V. I. (1962). *Approximate Calculation of Integrals*. Macmillan, New York. (This book considers only the problem of approximate integration of functions of a single variable. It was translated from the Russian by A. H. Stroud.)
- Lange, K. (1999). *Numerical Analysis for Statisticians*. Springer, New York. (This book contains a wide variety of topics on numerical analysis of potential interest to statisticians, including recent topics such as bootstrap calculations and the Markov chain Monte Carlo method.)
- Lieberman, O. (1994). "A Laplace approximation to the moments of a ratio of quadratic forms." *Biometrika*, **81**, 681–690.
- Liu, Q., and D. A. Pierce (1994). "A note on Gauss–Hermite quadrature." *Biometrika*, **81**, 624–629.
- Morin, D. (1992). "Exact moments of ratios of quadratic forms." *Metron*, **50**, 59–78.
- Morland, T. (1998). "Approximations to the normal distribution function." *Math. Gazette*, **82**, 431–437.
- Nonweiler, T. R. F. (1984). *Computational Mathematics*. Ellis Horwood, Chichester, England. (Numerical quadrature is covered in Chap. 5.)
- Phillips, C., and B. Cornelius (1986). *Computational Numerical Methods*. Ellis Horwood, Chichester, England. (Numerical integration is the subject of Chap. 6.)
- Phillips, G. M., and P. J. Taylor (1973). *Theory and Applications of Numerical Analysis*. Academic Press, New York. (Gaussian quadrature is covered in Chap. 6.)
- Piegorsch, W. W., and A. J. Bailer (1993). "Minimum mean-square error quadrature." *J. Statist. Comput. Simul*, **46**, 217–234.
- Ralston, A., and P. Rabinowitz (1978). *A First Course in Numerical Analysis*. McGraw-Hill, New York. (Gaussian quadrature is covered in Chap. 4.)
- Reid, W. H., and S. J. Skates (1986). "On the asymptotic approximation of integrals." *SIAM J. Appl. Math.*, **46**, 351–358.
- Roussas, G. G. (1973). *A First Course in Mathematical Statistics*. Addison-Wesley, Reading, Massachusetts.
- Searle, S. R. (1971). *Linear Models*. Wiley, New York.

- Shoup, T. E. (1984). *Applied Numerical Methods for the Microcomputer*. Prentice-Hall, Englewood Cliffs, New Jersey.
- Stark, P. A. (1970). *Introduction to Numerical Methods*. Macmillan, London.
- Strawderman, R. L. (2000). “Higher-order asymptotic approximation: Laplace, saddlepoint, and related methods.” *J. Amer. Statist. Assoc.*, **95**, 1358–1364.
- Stroud, A. H. (1971). *Approximate Calculation of Multiple Integrals*. Prentice-Hall, Englewood Cliffs, New Jersey.
- Sutradhar, B. C., and R. F. Bartlett (1989). “An approximation to the distribution of the ratio of two general quadratic forms with application to time series valued designs.” *Comm. Statist. Theory Methods*, **18**, 1563–1588.
- Tierney, L., R. E. Kass, and J. B. Kadane (1989). “Fully exponential Laplace approximations to expectations and variances of nonpositive functions.” *J. Amer. Statist. Assoc.*, **84**, 710–716.
- Wong, R. (1989). *Asymptotic Approximations of Integrals*. Academic Press, New York. (This is a useful reference book on the method of Laplace and Mellin transform techniques for multiple integrals. All results are accompanied by error bounds.)

EXERCISES

In Mathematics

12.1. Consider the integral

$$I_n = \int_1^n \log x \, dx.$$

It is easy to show that

$$I_n = n \log n - n + 1.$$

- (a) Approximate I_n by using the trapezoidal method and the partition points $x_0 = 1$, $x_1 = 2, \dots, x_n = n$, and verify that

$$I_n \approx \log(n!) - \frac{1}{2} \log n.$$

- (b) Deduce from (a) that $n!$ and $n^{n+1/2}e^{-n}$ are of the same order of magnitude, which is essentially what is stated in Stirling’s formula (see Example 12.6.1)

12.2. Obtain an approximation of the integral $\int_0^1 dx/(1+x)$ by Simpson’s method for the following values of n : 2, 4, 8, 16. Show that when $n = 8$, the error of approximation is less than or equal to 0.000002.

12.3. Use Gauss–Legendre quadrature with $n = 2$ to approximate the value of the integral $\int_0^{\pi/2} \sin \theta \, d\theta$. Give an upper bound on the error of approximation using formula (12.16).

12.4. (a) Show that

$$\int_m^\infty e^{-x^2} dx < \frac{1}{m} e^{-m^2}, \quad m > 0.$$

[Hint: Use the inequality $e^{-x^2} < e^{-mx}$ for $x > m$.]

(b) Find a value of m so that the upper bound in (a) is smaller than 10^{-4} .

(c) Use part (b) to find an approximation for $\int_0^\infty e^{-x^2} dx$ correct to three decimal places.

12.5. Obtain an approximate value for the integral $\int_0^\infty x(e^x + e^{-x} - 1)^{-1} dx$ using the Gauss–Laguerre quadrature. [Hint: Use the tables in Appendix C of the book by Krylov (1962, page 347) giving the zeros of Laguerre polynomials and the corresponding values of ω_i .]

12.6. Consider the indefinite integral

$$I(x) = \int_0^x \frac{dt}{1+t^2} = \text{Arctan } x.$$

(a) Make an appropriate change of variables to show that $I(x)$ can be written as

$$I(x) = 2x \int_{-1}^1 \frac{du}{4 + x^2(u+1)^2}$$

(b) Use a five-point Gauss–Legendre quadrature to provide an approximation for $I(x)$.

12.7. Investigate the asymptotic behavior of the integral

$$I(\lambda) = \int_0^1 (\cos x)^\lambda dx$$

as $\lambda \rightarrow \infty$.

12.8. (a) Use the Gauss–Laguerre quadrature to approximate the integral $\int_0^\infty e^{-6x} \sin x \, dx$ using $n = 1, 2, 3, 4$, and compare the results with the true value of the integral.

- (b) Use the Gauss-Hermite quadrature to approximate the integral $\int_{-\infty}^{\infty} |x| \exp(-3x^2) dx$ using $n = 1, 2, 3, 4$, and compare the results with the true value of the integral.

12.9. Give an approximation to the double integral

$$\int_0^1 \int_0^{(1-x_1^2)^{1/2}} (1-x_1^2-x_2^2)^{1/2} dx_1 dx_2$$

by applying the Gauss-Legendre rule to formulas (12.32) and (12.33).

- 12.10.** Show that $\int_0^{\infty} (1+x^2)^{-n} dx$ is asymptotically equal to $\frac{1}{2}(\pi/n)^{1/2}$ as $n \rightarrow \infty$.

In Statistics

- 12.11.** Consider a sample, $X_1, X_2, \dots, X_n$, of independent and identically distributed random variables from the standard normal distribution. Suppose that n is odd. The sample median is the $(m+1)$ th order statistic $X_{(m+1)}$, where $m = (n-1)/2$. It is known that $X_{(m+1)}$ has the density function

$$f(x) = n \binom{n-1}{m} \Phi^m(x) [1 - \Phi(x)]^m \phi(x),$$

where

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

is the standard normal density function, and

$$\Phi(x) = \int_{-\infty}^x \phi(t) dt$$

(see Roussas, 1973, page 194). Since the mean of $X_{(m+1)}$ is zero, the variance of $X_{(m+1)}$ is given by

$$\text{Var}[X_{(m+1)}] = n \binom{n-1}{m} \int_{-\infty}^{\infty} x^2 \Phi^m(x) [1 - \Phi(x)]^m \phi(x) dx.$$

Obtain an approximation for this variance using the Gauss-Hermite quadrature for $n = 11$ and varying numbers of quadrature points.

12.12. (Morland, 1998.) Consider the density function

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

for the standard normal distribution. Show that if $x \geq 0$, then the cumulative distribution function,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt,$$

can be represented as the sum of the series

$$\begin{aligned} \Phi(x) = \frac{1}{2} + \frac{x}{\sqrt{2\pi}} & \left[1 - \frac{x^2}{6} + \frac{x^4}{40} - \frac{x^6}{336} \right. \\ & \left. + \frac{x^8}{3456} - \cdots + (-1)^n \frac{x^{2n}}{(2n+1)2^n n!} + \cdots \right]. \end{aligned}$$

[*Note:* By truncating this series, we can obtain a polynomial approximation of $\Phi(x)$. For example, we have the following approximation of order 11:

$$\Phi(x) \approx \frac{1}{2} + \frac{x}{\sqrt{2\pi}} \left[1 - \frac{x^2}{6} + \frac{x^4}{40} - \frac{x^6}{336} + \frac{x^8}{3456} - \frac{x^{10}}{42240} \right].]$$

12.13. Use the result in Exercise 12.12 to show that there exist constants a, b, c, d , such that

$$\Phi(x) \approx \frac{1}{2} + \frac{x}{\sqrt{2\pi}} \frac{1 + ax^2 + bx^4}{1 + cx^2 + dx^4}.$$

12.14. Apply the method of importance sampling to approximate the value of the integral $\int_0^2 e^{x^2} dx$ using a sample of size $n = 150$ from the distribution whose density function is $g(x) = \frac{1}{4}(1+x)$, $0 \leq x \leq 2$. Compare your answer with the one you would get from applying formula (12.36).

12.15. Consider the density function

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{1}{2}(n+1)}, \quad -\infty < x < \infty,$$

for a t -distribution with n degrees of freedom (see Exercise 6.28). Let $F(x) = \int_{-\infty}^x f(t) dt$ be its cumulative distribution function. Show that for large n ,

$$F(x) \approx \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt,$$

which is the cumulative distribution function for the standard normal. [Hint: Apply Stirling's formula.]