Chapter 2. Numerical Quadrature

1 Introduction

Manu applications in statistical computing require the evaluation of a cumulative distribution function (cdf), $F(x)$, to determine the probability p corresponding to a given percentage point x_p. The first of two major problem types, considered in this chapter, is to compute p as $p = F(x_p)$ using given F and x_p. In most of the cases we will consider, $F(x)$ is not expressible in closed form, so we must somehow approximate an integral $F(x) = \int_{\infty}^{x} f(t)dt$.

Numerical integration (quadrature) methods are sometimes used to obtain approximate values for given cdf's of continuous distributions. The general problem which we will consider is evaluation of the integral

$$\int_a^b w(x)f(x)dx$$

where the function $w(x)$ is the integrand is called the *weight function.*

The second major type of problem is that of evaluating the inverse of a given cdf. This problem occurs whenever we want to determine the percentage point x_p which produces a given probability p. Thus the problem is to compute x_p as $x_p = F^{-1}(p)$. It is seldom possible to determine a simple closed form for $F^{-1}(p)$, so the usual situation is that we have only $F(x)$ and p, and must use these quantities to find x_p.

Since neither of these problems generally has a closed form solution, some type of mathematical approximation must be derived to provide a usable computational method.

2 One Dimensional Quadrature Rules and Formulas

Gaussian quadrature calls for approximating the given integral using the form

$$\begin{aligned} \int_a^b w(x)f(x)dx &= \sum_{j=1}^{n} m_j f(a_j) + E \\ &= \hat{I} + E \end{aligned}$$

where E is the error involved when $\hat{I}$ is taken as the integral value. The integer n is called the *order* of the quadrature formula. The m_j and a_j are defined by the specific type of Gaussian quadrature formula used.

In order to define the quantities m_j and a_j given above, consider a sequence of polynomials $\{p_k(x), k = 0, 1, \ldots\}$. We require that $p_k(x)$ have degree k and that the polynomials be orthogonal with respect to the weight function $w(x)$ over the interval $[a, b]$; in other words,

$$\int_a^b w(x)p_j(x)p_i(x)dx = 0 \quad I \neq j.$$

Assuming that we have a sequence $\{p_k(x)\}$ of orthogonal polynomials, let c_k be the coefficient of x^k in $p_k(x)$ and define

$$\delta_k = \int_a^b w(x)p_k^2(x)dx.$$

Under this definition it can be shown that the m_j are given as

$$m_j = \frac{-c_{n+1}\delta_n}{c_n p_{n+1}(a_j)p_n'(a_j)} \quad j = 1, 2, \ldots, n$$

where the a_j are the n zeros of $p_n(x)$, and $p_n'(a_j)$ denotes the first derivative of $p_n(x)$ evaluated at a_j. The a_j are called *abscissas* and the m_j are called *weights* in Gaussian quadrature. The error E has the form

$$E = \frac{\delta_n}{c_n^2(2n)!}f^{(2n)}(\xi)$$

where $\xi \in (a, b)$ and $f^{(2n)}(x)$ denotes the $2n$th derivative of $f(x)$.

3 Commonly Used Quadrature Formulas

1. Legendre-Gauss
   - Interval: $[-1, 1]$
   - Weight function: $w(x) = 1$
   - Polynomials $p_j(x)$: Legendre polynomials
   - $m_j = \frac{-2}{(n+1)p_{n+1}(a_j)p_n'(a_j)}$
   - $E = \frac{2^{2n+1}(n!)^4}{(2n+1)[(2n)!]^3}f^{(2n)}(\xi)$
2. Jacobi-Gauss
   - Interval: $[-1, 1]$
   - Weight function: $w(x) = (1-x)^\alpha(1+x)^\beta$, $\alpha, \beta > -1$
   - Polynomials $p_j(x)$: Jacobi polynomials
   - $m_j = \frac{-(2n+\alpha+\beta+2)\Gamma(n+\alpha+1)\Gamma(n+\beta+1)2^{\alpha+\beta}}{(n+\alpha+\beta+1)\Gamma(n+\alpha+\beta+1)(n+1)!p_n'(a_j)p_{n+1}(a_j)}$

- $E = \frac{\Gamma(n+\alpha+1)\Gamma(n+\beta+1)\Gamma(n+\alpha+\beta+1)n!2^{2n+\alpha+\beta+1}}{(2n+\alpha+\beta+1)[\Gamma(2n+\alpha+\beta+1)]^2(2n)!} f^{(2n)}(\xi)$

3. Laguerre-Gauss
   - Interval: $(0, \infty)$
   - Weight function: $w(x) = \exp(-x)$
   - Polynomials $p_j(x)$: Laguerre polynomials
   - $m_j = \frac{(n!)^2}{p_{n+1}(a_j)p'_n(a_j)}$
   - $E = \frac{(n!)^2}{(2n)!} f^{(2n)}(\xi)$
4. Hermite-Gauss
   - Interval: $(-\infty, \infty)$
   - Weight function: $w(x) = \exp(-x^2)$
   - Polynomials $p_j(x)$: Hermite polynomials
   - $m_j = \frac{2^{n+1}n!\pi^{1/2}}{p_{n+1}(a_j)p'_n(a_j)}$
   - $E = \frac{n!\pi}{2^n(2n)!} f^{(2n)}(\xi)$
5. Chebyshev-Gauss
   - Interval: $(-1, 1)$
   - Weight function: $w(x) = (1-x^2)^{-1/2}$
   - Polynomials $p_j(x)$: Chebyshev polynomials
   - $p_j(x) = \cos[j \cos^{-1}(x)]$

$$\int_{-1}^{1} (1-x^2)^{-1/2} f(x)dx = \frac{\pi}{n} \sum_{j=1}^{n} f(a_j) + E$$

   - $a_j = \cos \frac{(2j-1)\pi}{2n}$
   - $E = \frac{2\pi}{2^{2n}(2n)!} f^{(2n)}(\xi)$

Example. Consider the integral

$$(2\pi)^{-1/2} \int_{\infty}^{\infty} x \exp(-x^2/2)dx = (\frac{2}{\pi})^{1/2} \int_{\infty}^{\infty} y \exp(-y^2)dy.$$

The Hermite-Gauss formula is appropriate in this case because the interval of integration $(-\infty, \infty)$ is as required by the formula. Suppose we use the second-order Hermite Gauss formula. We then have $a_j = 0.707107$ or -0.707107 and $m_j = 0.886227$. These give

$$(\frac{2}{\pi})^{1/2} \int_{\infty}^{\infty} y \exp(-y^2)dy \approx (\frac{2}{\pi})^{1/2}[(0.886227)(0.707107) - (0.886227)(0.707107)] = 0.$$

One fundamental problem which occurs in many variants, is to approximate a function f by a member f^* of a class of functions which is easy to work with mathematically (for example, polynomials, rational functions, or trigonometric polynomials), where each particular function in the class is specified by the numerical values of a number of parameters. Here we shall mainly be concerned with the problem of **linear approximation**; i.e., a function f is to be approximated using a function f^* that can be expressed as a **linear** combination,

$$f^*(x) = c_0\phi_0(x) + c_1\phi_1(x) + \cdots + c_n\phi_n(x),$$

of $n+1$ functions, $\phi_0, \phi_1, \ldots, \phi_n$, which are chosen in advance; $c_0, c_1, \ldots, c_n$ are parameters whose value is to be determined. For example, if $\phi_i(x) = x^i$, the class of possible f^* is the class of polynomials of degree n. The set $\{1, x, x^2, \ldots, x^n\}$ is said to be a **basis** of the set of all polynomials of degree n. Usually a complicate function can be approximated by polynomials which is simpler in form and which to some extent models the behavior of the given function. In fact, the Weierstrass Approximation Theorem states that

> For every continuous function f defined on a clase, bounded interval it holds that $\lim_{n\to\infty} \min_P \|f - P\|_\infty = 0$ where P varies over the set of all nth-degree polynomials.

Many approximation methods are based on the principle of minimizing some norm or seminorm of the error function $y = f^* - f$, where f^* is to approximate f and has some predetermined form. Suppose a function f is to be approximated using an nth degree of polynomial and the error function $\|f - f^*\|^2$ is the weighted Euclidean norm $\int_a^b |f(x) - f^*(x)|^2 w(x)dx$. This problem is called the **least-squares approximation problem.**

We now introduce a formalism related to the geometric idea: the shortest distance from a point to a linear subspace is the length of the vector (between the point and a point in the subspace). First we need to define the inner product which is $< f, g >= \int_a^b w(x)f(x)g(x)dx$ where $\int_a^b w(x)f^2(x)dx < \infty$ and $\int_a^b w(x)g^2(x)dx < \infty$. Recall that the class of possible f^* is the class of polynomials of degree n. In order to get a well-conditioned representation of a function and easy manipulation, we consider families of orthogonal polynomials. To get the orthogonal polynomial under the above described inner product, they can be obtained by using Gram-Schmidt orthogonalization procedure to the polynomials $\{x^n, n = 0, 1, \cdots\}$. In particular,

1. Legendre polynomials
   - Interval: $[-1, 1]$
   - Weight function: $w(x) = 1$
   - $P_n(t) = \sqrt{\frac{2n+1}{2}} \frac{1}{(2n)!} \frac{d^n}{dt^n}[(t^2-1)^n]$
2. Jacobi polynomials
   - Interval: $[-1, 1]$
   - Weight function: $w(x) = (1-x)^\alpha (1+x)^\beta$, $\alpha, \beta > -1$
   - $J_n^{(\alpha,\beta)}(x) = k_n (1-x)^\alpha (1+x)^\beta \frac{d^n}{dt^n}[(1-x)^{\alpha+n}(1+x)^{\beta+n}]$ where $k_n = (-1)^n \sqrt{2^{-(\alpha+\beta+2n+1)} \frac{\Gamma(\alpha+\beta+n+1)}{\Gamma(\alpha+n+1)\Gamma(\beta+n+1)} \frac{\alpha+\beta+2n+1}{n!}}$.

     If $\alpha = \beta = -1/2$, it leads to Chebyshev polynomials. Note that $\cos(n\theta)$ can be written as a n nth degree of polynomial in $\cos\theta$. We then have $T_0(x) = 1, T_0(x) = x, T_0(x) = 2x^2 - 1, T_0(x) = 4x^3 - 3x$. In general, we have a recursion formula $T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x)$ for $n \geq 1$.
3. Laguerre polynomials
   - Interval: $(0, \infty)$
   - Weight function: $w(x) = \exp(-x)$
   - $L_n(x) = \exp(x) \frac{(-1)^n}{n!} \frac{d^n}{dx^n}[\exp(-x)x^n]$
4. Hermite polynomials
   - Interval: $(-\infty, \infty)$
   - Weight function: $w(x) = \exp(-x^2)$
   - $H_n(x) = \exp(x^2) \frac{(-1)^n}{\sqrt{2^n n! (\pi)^{1/2}}} \frac{d^n}{dx^n}[\exp(-x^2)]$

4 Compound Quadrature Rules and Error Estimates

In each quadrature formula the error term involves the derivative of $f(x)$ evaluated at an unknown point ξ in the interval of integration. Sometimes $f^{(2n)}(x)$ is such a complicated function that we cannot obtain a bound for it. Instead we usually apply them by combining them together, as we now illustrate. Let $[a,b]$ be divided into s subintervals, called **panels**, by the points $a = t_0 < t_1 < t_2 < \cdots < t_{s-1} < t_s = b$. Since

$$\int_a^b w(x)f(x)dx = \sum_{i=0}^{s-1} \int_a^b w(x)f(x)dx,$$

we can estimate the integral by applying a quadrature rule to each panel. If the same rule is used on each, and if the points t_i are uniformly spaced on $[a,b]$ the result is called a **compound** quadrature rule. If, for example, we apply a Legendre-Gauss nth-order formula to approximate each integral, then, under fairly general regularity conditions, the error in this approximation has the form

$$E = (\frac{1}{m})^{2n}(\frac{b-a}{2})^{2n+1}\frac{2^{2n+1}(n!)^4}{(2n+1)[(2n)!]^3}f^{(2n)}(\eta) \quad \eta \in (a,b).$$

Notice the factor $(1/m)^{2n}$ which will force E toward zero as m increases.

5 Newton-Cotes Quadrature

In contrast with the Gaussian quadrature formula which use abscissas that are zeros of orthogonal polynomials, the Newton-Cotes formulas are derived by constraining the abscissas so that they are equally spaced in the (finite-length) interval of integration. In the case $w(x) = 1$, the Newton-Cotes formula is

$$\int_a^b f(x)dx = \sum_{j=0}^{n} m_j f(a+hj) + E$$

where $h = (b-a)/n$, and n is the *order* of the formula.

Consider the use of composite quadrature with the first-order Newton-Cotes formula on each of m subintervals having length $h = (b-a)/m$. An algebraic simplification yields

$$\int_a^b f(x)dx = h(\frac{1}{2}f_0 + \cdots + f_{m-1} + \frac{1}{2}f_m) + E$$

where $f_j = f(a+hj)$. The error in this case can be shown to have the form

$$E = -\frac{(b-a)^3}{12m^2}f^{(2)}(\xi) \quad \xi \in (a,b).$$

These two equations define what is called the *trapezoidal rule* for numerical integration over a finite length interval.

Use of the second-order Newton-Cotes formula in the composite mode with m subintervals and $h = a_{i+1} - a_i$ yields

$$\int_a^b f(x)dx = \frac{h}{3}(f_0 + 4f_1 + 2f_2 + 4f_3 + \cdots + 4f_{2m-3} + 2f_{2m-2} + 4f_{2m-1} + f_{2m}) + E.$$

The error term in this case is

$$E = -\frac{(b-a)^5}{180m^4} f^{(4)}(\xi) \quad \xi \in (a,b).$$

This composite form defines *Simpson's rule*, which is a frequently used method for numerical integration.

There are several issues associated with developing a reliable computer program using this technique.

- The trapezoid and Simpson formula are *closed*, requiring $f(a)$ and $f(b)$. They cannot be applied to integrals which have non-removable singularities, such as

  $$\int_0^1 \log x \cos x^2 dx \quad \text{or} \int_0^1 \frac{\exp(\sin x)}{\sqrt{x}} dx.$$

  Both integrals exist even though the integrands become infinite at the left endpoint.

- Look back at the derivation of the trapezoid or Simpson's rule. The argument hinged on the existence of a Taylor expansion. If f has a singularity on $[a,b]$, such as $f = \sqrt{x}$ on $[0,1]$, a Taylor expansion will not be valid on the whole interval and the derivation of the error, i.e., the trapezoid formula is unjustified. In general, a Riemann sum quadrature rule such as the trapezoid rule leads to a Riemann sum compound rule. Such rules can be made arbitrarily accurate as we increase the number of nodes. But the rate of convergence depends on the smoothness of f.

6 Data Integration

Suppose we are given a sequence of (x, y) pairs

$$(x_i, y_i), \quad x_1 < x_2 < \cdots < x_n$$

which are thought of as accurate values of some (unknown) function $f(x)$, $f(x_i) = y_i$. The problem is to compute

$$I = \int_{x_1}^{x_n} f(x)dx.$$

As the data are accurate, a reasonable approach is to interpolate to the data. That is, to construct a known function $g(x)$ such that $g(x_i) = y_i$ and then compute

$$I \approx I^{'} = \int_{x_1}^{x_n} g(x)dx.$$

The most heavily used interpolants are polynomials and piecewise polynomials, such as piecewise linear polynomials, Hermite cubics or splines. Piecewise polynomials lead to simple expressions for the quadrature rule. These are worked out below. However, if the data contain substantial amounts of experimental or computational error, then interpolation is inappropriate and the smoothing techniques are recommended instead.

Let $g(x)$ be the function that connects each pair of data points with a line segment. This is the **piecewise linear interpolant**. Then

$$I \approx I^{'} = \int_{x_1}^{x_n} g(x)dx = \sum_{i=1}^{n-1} \int_{x_i}^{x_{i+1}} g(x)dx.$$

Since g is linear between x_i and x_{i+1} it can be integrated exactly, giving

$$\begin{aligned} I \approx I^{'} &= \frac{1}{2}\sum_{i=1}^{n-1}(x_{i+1} - x_i)(y_i + y_{i+1}) \\ &= \frac{1}{2}\left[(x_2 - x_1)y_1 + (x_3 - x_1)y_2 + (x_4 - x_2)y_3 + \cdots + (x_n - x_{n-2})y_{n-1} + (x_{n-1} - x_n)y_n\right]. \end{aligned}$$

Note that if the x_i are equally spaced this is the compound trapezoid rule. We can interprete this rule as giving the *exact* integral of the function $g(x)$ which is the piecewise linear interpolant through the data.

7 Infinite and Semiinfinite Intervals

The most popular method is the use of weight function such as the Gauss Laguerre quadrature. In this section, we consider two other methods.

7.1 Truncation

If either endpoints a, b or both can be infinite, a common approach is to truncate the limits and use a standard method on the finite part. Sometimes it is possible to analyze the tails of

the integrand and show that their contribution is negligible, but more often several different intervals are selected and it is verified computationally that the finite integral estimate is not changing. This can give wrong answers in a general situation, but often works if some physical insight can be brought to the problem.

As an example, consider $I = \int_0^\infty \exp(-x)\cos^2(x^2)dx$. If we do this infinite integral by truncating at A, the error in neglecting the tail is no more than $\exp(-A)$.

7.2 Transformation

By setting $x = p(t)$ we can transform the interval. Typical $p(t)$'s for the interval $[0, \infty)$ are $x = \ln t$ or $x = t/(1-t)$. For the former we get

$$\int_0^\infty f(x)dx = \int_0^1 f(-\ln t)\frac{dt}{t}.$$

For $I = \int_0^\infty \exp(-x)\cos^2(x^2)dx$, this is

$$I = \int_0^1 \cos^2(\ln^2 t)dt.$$

This integrand oscillates infinitely often between the values 0 and 1 on any interval containing the origin. If instead we had used the transformation $x = p(t) = -2\ln t$, the integrand would be $2t\cos^2(4\ln^2 t)$. This also oscillates, but it is modulated by the line $y = t$ and hence goes to zero at the origin. Ignoring a small interval $[0, \epsilon]$ leaves us with a more tractable problem.

8 Double Integrals

$$I = \int\int_D f(x,y)dA.$$

This is a much more difficult problem than the evaluation of a one-dimensional integral. It is still an active research topic and many gaps remain in our knowledge.

If D is a rectangle or a trinagle, we can use the following product rule. If

$$\int_\alpha^\beta f(x)dx = \sum_{i=1}^n w_i f(x_i) + R_1$$

and

$$\int_a^b g(y)dy = \sum_{j=1}^m p_j g(y_j) + R_2,$$

then

$$\begin{aligned}
\int_a^b \int_\alpha^\beta f(x,y)dxdy &= \int_a^b \left[\sum_{i=1}^n w_i f(x_i,y) + R_1\right] dy \\
&= \sum_{i=1}^n w_i \int_a^b f(x_i,y)dy + \int_a^b R_1 dy \\
&= \sum_{i=1}^n w_i \left[\sum_{j=1}^m p_j f(x_i,y_j) + R_2\right] + \int_a^b R_1 dy \\
&= \sum_{i=1}^n \sum_{j=1}^m w_i p_j f(x_i,y_j) + \sum_{i=1}^n w_i R_2 + \int_a^b R_1 dy \\
&= \sum_{i=1}^n \sum_{j=1}^m w_i p_j f(x_i,y_j) + \mathcal{R}.
\end{aligned}$$

This is called a **product formula** because it uses nm integrand evaluation points.

A surprising number of regions can be transformed into a rectangle. Sometimes other difficulties are introduced but the technique is still valuable. To illustrate, suppose we want to compute the integral on a triangle

$$\triangle = \{(x,y) : 0 \le x \le 1, 0 \le y \le x\}.$$

Let us make a change of variable $u = x$, $v = y/x$. Then the original integral becomes

$$I = \int_0^1 \int_0^x f(x,y)dydx = \int_0^1 \int_0^1 f(u,uv)udvdu.$$

We can now apply a product rule for the integral on a square.