STA6171: Statistical Computing for DS 1

Numerical Integration

Ick Hoon Jin

Yonsei University, Department of Statistics and Data Science

2020.10.21

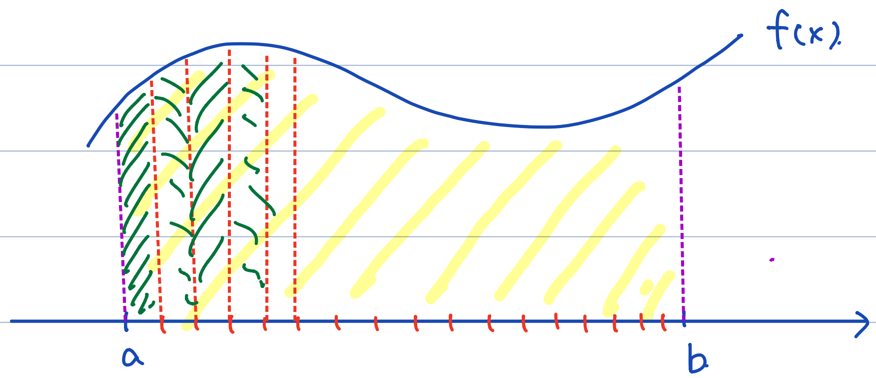
Introduction

$\int_a^b f(x)dx$ ↗ a few function ⇒ derive integration analytically
↘ numerical approximation of integrals.

- Consider a one-dimensional integral of the form

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

- The value of the integral can be derived analytically for only a few functions f. For the rest, numerical approximations of the integral are often useful. $\pi(\theta|y) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d\theta}$ → need to evaluate. ⇒ numerical integration.
- Approximation of integrals is frequently required for Bayesian inference since a posterior distribution may not belong to a familiar distribution.
- Integral approximation is also useful in some ML inference problem when the likelihood is a function of one or more integrals.



(1) Partition interval $[a, b]$ into subintervals

(2). Calculate the area of subinterval

(3) Sum all areas of subinterval.

$[a, b]$ partition into n subintervals.

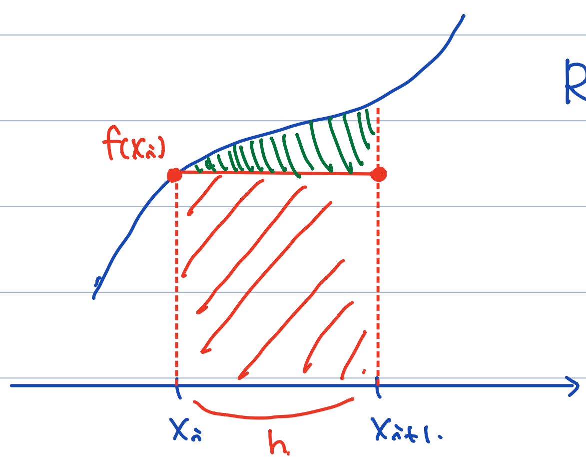
$$[x_i, x_{i+1}] \quad \text{for} \quad i = 0, \cdots, n-1$$

$$x_0 = a \qquad x_n = b.$$

※. All subintervals have equal lengths.

$$\int_a^b f(x)\,dx = \sum_{i=0}^{n-1} \underbrace{\int_{x_i}^{x_{i+1}} f(x)\,dx}$$

↳ how to calculate $\int_{x_i}^{x_{i+1}} f(x)\,dx$.



Riemann's Rule. (0-degree order)

$(x_{i+1} - x_i)\, f(x_i)$ ↳ line from $f(x_i)$

$$\int_a^b f(x)\,dx = \sum_{i=0}^{n-1} (x_{i+1} - x_i)\, f(x_i)$$

Suppose that x_i are equally spaced. $h = (b-a)/n$.

$$\int_a^b f(x)\,dx = \sum_{i=0}^{n-1} h \cdot f(a+ih) = \hat{R}(n)$$

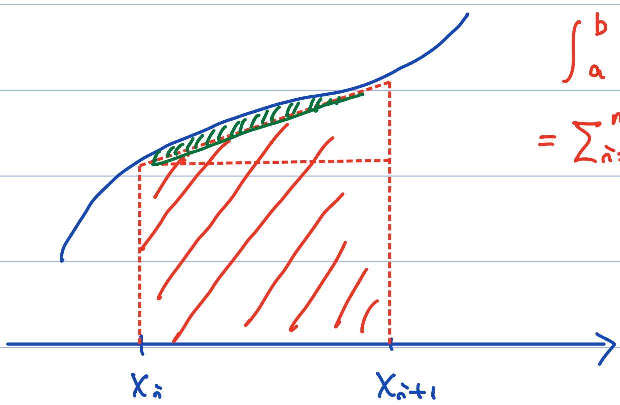
$i=0$ x_0

$i=1$ x_1

$\vdots$

$i=n-1$ x_{n-1}

As $n \to \infty$, the numerical integration converges to the true value.



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

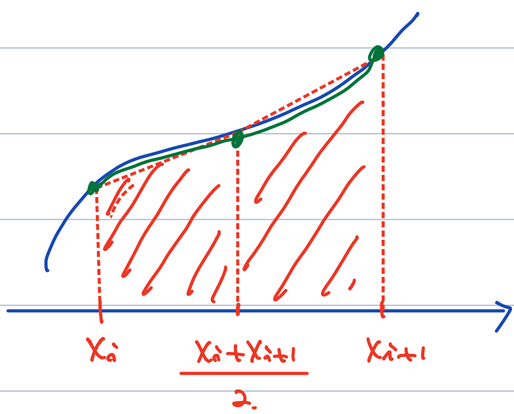
$$= \sum_{i=0}^{n-1} \frac{1}{2} (x_{i+1} - x_i)\left(f(x_i) + f(x_{i+1}) \right)$$

Trapezoidal Rule. (1-order degree)

$$\int_a^b f(x)\,dx = \frac{1}{2} \sum_{i=0}^{n-1} h\left(f(x_i) + f(\cancel{a}\, x_{i+1}) \right)$$

$$= \frac{h}{2} f(a) + h \sum_{i=1}^{n-1} f(a+ih) + \frac{h}{2} f(b) = \hat{T}(n)$$

Simpson's Rule. (2nd-order degree)

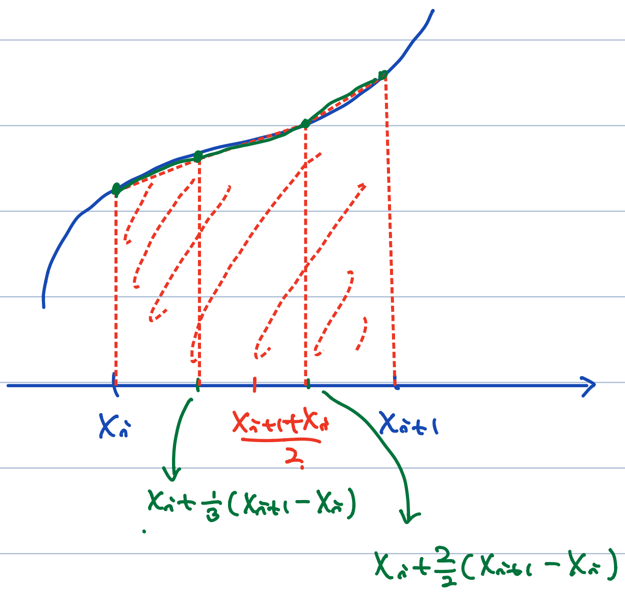


$$\Big\langle \int_a^b f(x)\,dx \approx \sum_{i=0}^{n-1} \frac{1}{6} (x_{i+1} - x_i)\left[f(x_i) + 4f\left(\frac{x_i + x_{i+1}}{2}\right) + f(x_{i+1}) \right]$$

$$= \sum_{i=0}^{n-1} \frac{1}{6} h \left[f(x_i) + 4f\left(\frac{x_i + x_{i+1}}{2}\right) + f(x_{i+1}) \right]$$

$$= \frac{h}{3} \sum_{i=1}^{n/2} \left(f(x_{2i-2}) + 4f(x_{2i-1}) + f(x_{2i}) \right) \Big\rangle$$

Numerical Interation with Higher-order degree.



Introduction

- To initiate an approximation of $\int_a^b f(x)dx$, partition the interval $[a, b]$ into n subintervals, $[x_i, x_{i+1}]$ for $i = 0, \cdots, n-1$, with $x_0 = a$ and $x_n = b$. Then,

$$\int_a^b f(x)dx = \sum_{i=1}^{n-1} \int_{x_i}^{x_{i+1}} f(x)dx.$$

- This composite rule breaks the whole integral into may smaller parts, but postpones the question of how to approximate any single part.
- The approximation of a single part will be made using a simple rule. Within the interval $[x_i, x_{i+1}]$, insert $m+1$ nodes, x_{ij}^* for $j = 0, \cdots, m$.

Introduction

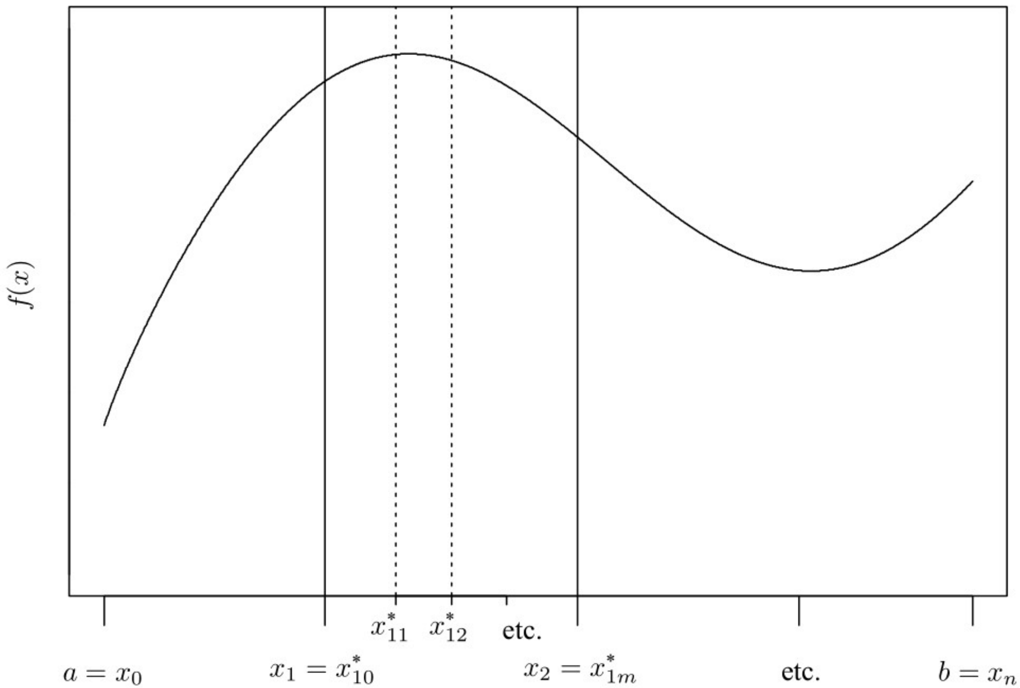


FIGURE 5.1 To integrate f between a and b, the interval is partitioned into n subintervals, $[x_i,\ x_{i+1}]$, each of which is further partitioned using $m+1$ nodes, $x^*_{i0}, \ldots, x^*_{im}$. Note that when $m = 0$, the subinterval $[x_i, x_{i+1}]$ contains only one interior node, $x^*_{i0} = x_i$.

Introduction

- In general, numerical integration methods require neither equal spacing of subintervals or nodes nor equal number of nodes within each subintervals.
- A simple rule will rely on the approximation

$$\int_{x_i}^{x_{i+1}} f(x)dx \approx \sum_{j=0}^{m} A_{ij} f\left(x_{ij}^*\right)$$

for some set of constants A_{ij}.

Newton-Cotes Quadrature

- A simple and flexible class of integration methods consists of the Newton–Cotes rules. In this case, the nodes are equally spaced in $[x_i, x_{i+1}]$, and the same number of nodes is used in each subintervals.
- The Newton–Cotes approach replaces the true integrand with a polynomial approximation on each subinterval.
- The constants A_{ij} are selected so that $\sum_{j=}^{m} A_{ij} f\left(x_{ij}^*\right)$ equals the integrals of an interpolating polynomial on $[x_i, x_{i+1}]$ that matches the value of f at the nodes within this subinterval.

Newton-Cotes Quadrature

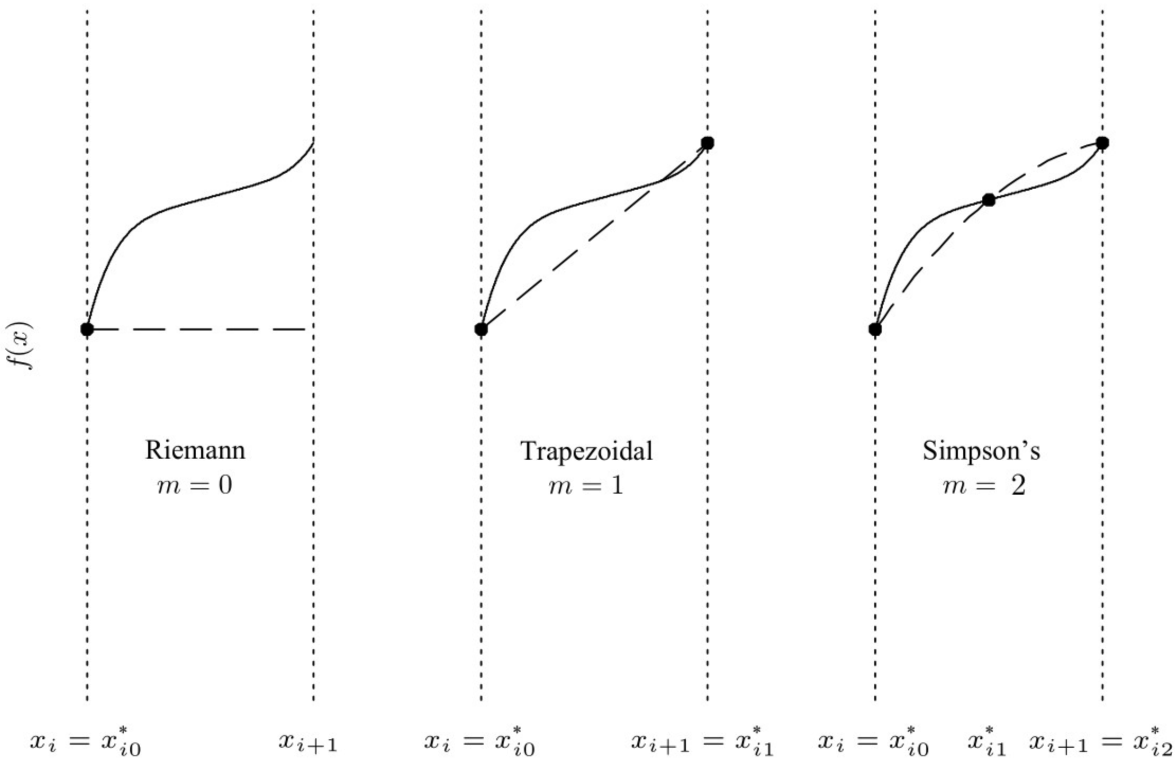


FIGURE 5.2 Approximation (dashed) to f (solid) provided on the subinterval $[x_i, x_{i+1}]$, for the Riemann, trapezoidal, and Simpson's rules.

Riemann Rule

- The simple Riemann rule amounts to approximating f on each subinterval by a constant function, $f(x_i)$, whose value matches that of f at one point on the interval.

$$\int_{x_i}^{x_{i+1}} f(x)dx \approx \int_{x_i}^{x_{i+1}} f(x_i)dx = (x_{i+1} - x_i)f(x_i).$$

  The composite rule sums n such terms to provide an approximation to the integral over $[a, b]$.

Riemann Rule

- Suppose the x_i are equally spaced so that each subinterval has the same length $h = (b - a)/n$. Then, we may write $x_i = a + ih$, and the composite rule is

$$\int_{x_i}^{x_{i+1}} f(x)dx \approx h \sum_{i=0}^{n-1} f(a + ih) = \hat{R}(n).$$

  The approximation converges to the true value of the integral as $n \to \infty$ by definition of the Riemann integral of an integrable function.

- If f is a polynomial of zero degree (i.e., a constant function), then f is constant on each subinterval, so the Riemann rule is exact.

Riemann Rule

- When applying the composite Riemann rule, it makes sense to calculate a sequence of approximations, say $\hat{R}(n_k)$, for an increasing sequence of numbers of subintervals, n_k, as $k = 1, 2, \cdots$.
- Convergence of $\hat{R}(n_k)$ can be monitored using an absolute or relative convergence criterion.

Riemann Rule

TABLE 5.2 Estimates of the integral in (5.7) using the Riemann rule with various numbers of subintervals. All estimates are multiplied by a factor of 10^5. Errors for use in a relative convergence criterion are given in the final column.

Subintervals	Estimate	Relative Error
2	3.49388458186769	
4	1.88761005959780	−0.46
8	1.72890354401971	−0.084
16	1.72889046749119	−0.0000076
32	1.72889038608621	−0.000000047
64	1.72889026784032	−0.000000068
128	1.72889018400995	−0.000000048
256	1.72889013551548	−0.000000028
512	1.72889010959701	−0.000000015
1024	1.72889009621830	−0.0000000077

Trapezoidal Rule

- Although the simple Riemann rule is exact if f is constant on $[a, b]$, it can be quite slow to converge to adequate precision in general.
- An obvious improvement would be to replace the piecewise constant approximation by a piecewise mth-degree polynomial approximation.
- We begin by introducing a class of polynomials that can be used for such approximations. This permits the Riemann rule to be cast as the simplest member of a family of integration rules having increased precision as m increases.

Trapezoidal Rule

- Let the fundamental polynomials be

$$p_{ij}(x) = \prod_{k=0,k\neq j}^{m} \frac{x - x_{jk}^*}{x_{ij}^* - x_{ik}^*}$$

  for $j = 0, \cdots, m$.

- The function $p_i(x) = \sum_{j=0}^{m} f\left(x_{ij}^*\right) p_{ij}(x)$ is an mth-degree polynomial that interpolates f at all the nodes $x_{i0}^*, \cdots, x_{1m}^*$ in $[x_i, x_{i+1}]$.

Trapezoidal Rule

- This approximation replaces integration of an arbitrary function f with polynomial integration. The resulting composite rule is

$$\int_a^b f(x)dx \approx \sum_{i=0}^{n-1} \sum_{j=0}^{m} A_{ij} f\left(x_{ij}^*\right)$$

when there are m nodes on each subinterval.

Trapezoidal Rule

- Letting $m = 1$ with $x_{i0}^* = x_i$ and $x_{i1}^* = x_{i+1}$ yields the trapezoidal rule.
- In this case,

$$p_{i0}(x) = \frac{x - x_{i+1}}{x_i - x_{i+1}} \qquad \text{and} \qquad p_{i1}(x) = \frac{x - x_{i+1}}{x_i - x_{i+1}}.$$

- Integrating these polynomials yields $A_{i0} = A_{i1} = \frac{x_{i+1} - x_i}{2}$.

Trapezoidal Rule

- The trapezoidal rule amounts to

$$\int_a^b f(x)dx \approx \sum_{i=0}^{n-1} \left(\frac{x_{i+1} - x_i}{2}\right) \left(f(x_i) + f(x_{i+1})\right).$$

- When $[a, b]$ is partitioned into n subintervals of equal length $h = (b - a)/n$, then the trapezoidal rule estimate is

$$\int_a^b f(x)dx \approx \frac{h}{2} f(a) + h \sum_{i=1}^{n-1} f(a + ih) + \frac{h}{2} f(b) = \hat{T}(n).$$

Trapezoidal Rule

TABLE 5.3 Estimates of the integral in (5.7) using the Trapezoidal rule with various numbers of subintervals. All estimates are multiplied by a factor of 10^5. Errors for use in a relative convergence criterion are given in the final column.

Subintervals	Estimate	Relative Error
2	3.49387751694744	
4	1.88760652713768	−0.46
8	1.72890177778965	−0.084
16	1.72888958437616	−0.0000071
32	1.72888994452869	0.00000021
64	1.72889004706156	0.000000059
128	1.72889007362057	0.000000015
256	1.72889008032079	0.0000000039
512	1.72889008199967	0.00000000097
1024	1.72889008241962	0.00000000024

Simpson's Rule

- Letting $m = 2$, $x_{i0}^* = x_i$, $x_{i1}^* = \frac{x_i + x_{i+1}}{2}$, and $x_{i2}^* = x_{i+1}$, we obtain Simpson's rule.
- $A_{i0} = A_{i2} = \frac{x_{i+1} - x_i}{6}$ and $A_{i1} = 2(A_{i0} A_{i2})$. This yield the approximation

$$\int_{x_i}^{x_{i+1}} f(x)dx \approx \frac{x_{i+1} - x_i}{6} \left[f(x_i) + 4f\left(\frac{x_i + x_{i+1}}{2}\right) + f(x_{i+1}) \right].$$

for the $(i+1)$-th subinterval.

Simpson's Rule

- Suppose the interval $[a, b]$ has been partitioned into n subintervals of equal length $h = (b - a)/n$, where n is even.
- To apply Simpson's rule, we need an interior node in each $[x_i, x_{i+1}]$ Since n is even, we may adjoin pairs of adjacent subintervals, with the shared endpoint serving as the interior node of the larger intervals. This provides $n/2$ subintervals of length $2h$, for which

$$\int_a^b f(x)dx \approx \frac{h}{3}\sum_{i=1}^{n/2}(f(x_{2i-2}) + 4f(x_{2i-1} + f(x_{2i})) = \hat{S}\left(\frac{n}{2}\right).$$

Simpson's Rule

TABLE 5.4 Estimates of the integral in (5.7) using Simpson's rule with various numbers of subintervals (and two nodes per subinterval). All estimates are multiplied by a factor of 10^5. Errors for use in a relative convergence criterion are given in the final column.

Subintervals	Estimate	Relative Error
2	1.35218286386776	
4	1.67600019467364	0.24
8	1.72888551990500	0.032
16	1.72889006457954	0.0000026
32	1.72889008123918	0.0000000096
64	1.72889008247358	0.00000000071
128	1.72889008255419	0.000000000047
256	1.72889008255929	0.0000000000029
512	1.72889008255961	0.00000000000018
1024	1.72889008255963	0.000000000000014

General kth-Degree Rule

- The proceeding discussion raises the general question about how to determine a Newton-Cotes rule that is exact for polynomials of degree k. This would require constants $c_0, \cdots, c_k$ that satisfy

$$\int_a^b f(x)dx = c_0 f(a) + c_1 f\left(a + \frac{b-a}{k}\right) + \cdots + c_i f\left(a + \frac{i(b-a)}{k}\right) + \cdots + c_k f(b)$$

for any polynomial f

Romberg Integration

- In general, low-degree Newton-Cotes methods are slow to converge. However, there is very efficient mechanism to improve upon a sequence of trapezoidal rule estimates.
- Let $\hat{T}(n)$ denote the trapezoidal rule estimate of $\int_a^b f(x)dx$ using n subintervals of equal length $h = (b-a)/n$. Without loss of generality, suppose $a = 0$ and $b = 1$. Then,

$$\hat{T}(1) = \frac{1}{2}f(0) + \frac{1}{2}f(1),$$

$$\hat{T}(2) = \frac{1}{4}f(0) + \frac{1}{2}f\left(\frac{1}{2}\right) + \frac{1}{4}f(1),$$

$$\hat{T}(4) = \frac{1}{8}f(0) + \frac{1}{4}\left[f\left(\frac{1}{4}\right) + f\left(\frac{2}{4}\right) + f\left(\frac{3}{4}\right)\right] + \frac{1}{8}f(1),$$

and so forth.

$\hat{T}(n)$: approximate $\int_a^b f(x)\,dx$ using n subintervals. $h = \frac{(b-a)}{n}$

W/o loss of generality, suppose $a=0$ and $b=1$.

$$\hat{T}(1) = \frac{1}{2} f(0) + \frac{1}{2} f(1)$$

$$\hat{T}(2) = \frac{1}{4} f(0) + \frac{1}{2} f(\tfrac{1}{2}) + \frac{1}{4} f(1)$$

$$= \frac{1}{2} \hat{T}(1) + \frac{1}{2} f(\tfrac{1}{2})$$

$$\hat{T}(4) = \frac{1}{8} f(0) + \frac{1}{4} f(\tfrac{1}{4}) + \frac{1}{4} f(\tfrac{2}{4}) + \frac{1}{4} f(\tfrac{3}{4}) + \frac{1}{8} f(1)$$

$$= \frac{1}{2} \hat{T}(2) + \frac{1}{4} [f(\tfrac{1}{4}) + f(\tfrac{3}{4})]$$

$$\hat{T}(8) = \frac{1}{16} f(0) + \frac{1}{8} \left(f(\tfrac{1}{8}) + \cdots + f(\tfrac{7}{8}) \right) + \frac{1}{8} f(1)$$

$$= \frac{1}{2} \hat{T}(4) + \frac{1}{8} [f(\tfrac{1}{8}) + f(\tfrac{3}{8}) + f(\tfrac{5}{8}) + f(\tfrac{7}{8})]$$

Make recursive formula

$$\Rightarrow \hat{T}(2n) = \frac{1}{2} \hat{T}(n) + \frac{h}{2} \sum_{i=1}^{n} f\left(a + (i - \tfrac{1}{2}) h\right)$$

Define $\hat{T}_{i,0} = \hat{T}(2^i)$ for $i = 0, \cdots, m$.

$$\begin{array}{cccc}
\hat{T}_{0,0} & & & \\
\hat{T}_{1,0} & \hat{T}_{1,1} & & \\
\hat{T}_{2,0} & \hat{T}_{2,1} & \hat{T}_{2,2} & \\
\hat{T}_{3,0} & \hat{T}_{3,1} & \hat{T}_{3,2} & \hat{T}_{3,3} \\
\vdots & & &
\end{array}$$

using $\displaystyle \hat{T}_{ij} = \frac{4^j \hat{T}_{i,j-1} - \hat{T}_{i-1,j-1}}{4^j - 1}.$

$$\hat{T}_{1,1} = \frac{4^1 \hat{T}_{1,0} - \hat{T}_{0,0}}{4^1 - 1}$$

$$\hat{T}_{2,1} = \frac{4^1 \hat{T}_{2,0} - \hat{T}_{1,0}}{4^1 - 1} \qquad \hat{T}_{2,2} = \frac{4^2 \hat{T}_{2,1} - \hat{T}_{1,1}}{4^2 - 1}$$

Romberg Integration

- Noting that

$$
\begin{aligned}
\hat{T}(2) &= \frac{1}{2}\hat{T}(1) + \frac{1}{2} f\left(\frac{1}{2}\right), \\
\hat{T}(4) &= \frac{1}{2}\hat{T}(2) + \frac{1}{4}\left[f\left(\frac{1}{4}\right) + f\left(\frac{3}{4}\right)\right],
\end{aligned}
$$

and so forth suggests the general recursion relationship

$$
\hat{T}(2n) = \frac{1}{2}\hat{T}(n) + \frac{h}{2}\sum_{i=1}^{n} f\left(a + \left(i - \frac{1}{2}\right) h\right).
$$

Romberg Integration

- Begin by defining $\hat{T}_{i,0} = \hat{T}(2^i)$ for $i = 0, \cdots, m$. Then, define a triangular array of estimates like

$$\begin{array}{cccccc} \hat{T}_{0,0} & & & & & \\ \hat{T}_{1,0} & \hat{T}_{1,1} & & & & \\ \hat{T}_{2,0} & \hat{T}_{2,1} & \hat{T}_{2,2} & & & \\ \hat{T}_{3,0} & \hat{T}_{3,1} & \hat{T}_{3,2} & \hat{T}_{3,3} & & \\ \hat{T}_{4,0} & \hat{T}_{4,1} & \hat{T}_{4,2} & \hat{T}_{4,3} & \hat{T}_{4,3} & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{array}$$

using the relationship

$$\hat{T}_{i,j} = \frac{4^j \hat{T}_{i,j-1} - \hat{T}_{i-1,j-1}}{4^j - 1}$$

for $j = 1, \cdots, i$ and $i = 1, \cdots, m$.

Romberg Integration

- It is important to check that the Romberg calculations do not deteriorate as m is increased. To do this, consider the quotient

$$Q_{ij} = \frac{\hat{T}_{i,j} - \hat{T}_{i-1,j}}{\hat{T}_{i+1,j} - \hat{T}_{i,j}}$$

$$Q_{ij} = \frac{\hat{T}_{i,j} - \hat{T}_{i-1,j}}{\hat{T}_{i+1,j} - \hat{T}_{i,j}}.$$

numerical approximation error.

- The error in $\hat{T}_{i,j}$ is attributable partially to the approximation strategy itself and partially to numerical imprecision introduced by computer roundoff.

error from computer calculation

numerical approximation error >>> computer error.

- As long as the former source dominates, the Q_{ij} values should approach 4^{j+1} as i increases.
- However, when computer roundoff error is substantial relative to approximate error, the Q_{ij} values will become erratic.

Romberg Integration

$\hat{T}_{0,0}$

$\hat{T}_{1,0}$ $\hat{T}_{1,1}$

$\hat{T}_{2,0}$ $\hat{T}_{2,1}$ $\hat{T}_{2,2}$

$\vdots$

To find the best numerical integration value, we use Q_{ij}

TABLE 5.5 Estimates of the integral in (5.7) using Romberg integration. All estimates and differences are multiplied by a factor of 10^5. The final two columns provide performance evaluation measures discussed in the text.

i	j	Subintervals	$\hat{T}_{i,0}$	$\hat{T}_{i,j} - \hat{T}_{i-1,j}$	Q_{ij}
1	0	2	3.49387751694744		
2	0	4	1.88760652713768	−1.60627098980976	
3	0	8	1.72890177778965	−0.15870474934803	10.12
4	0	16	1.72888958437616	−0.00001219341349	13015.61
5	0	32	1.72888994452869	0.00000036015254	−33.86
6	0	64	1.72889004706156	0.00000010253287	3.51
7	0	128	1.72889007362057	0.00000002655901	3.86
8	0	256	1.72889008032079	0.00000000670022	3.96
9	0	512	1.72889008199967	0.00000000167888	3.99
10	0	1024	1.72889008241962	0.00000000041996	4.00
1	1	2			
2	1	4	1.35218286386776		
3	1	8	1.67600019467364	0.32381733080589	
4	1	16	1.72888551990500	0.05288532523136	6.12
5	1	32	1.72889006457954	0.00000454467454	11636.77
6	1	64	1.72889008123918	0.00000001665964	272.80
7	1	128	1.72889008247358	0.00000000123439	13.50
8	1	256	1.72889008255420	0.00000000008062	15.31
9	1	512	1.72889008255929	0.00000000000510	15.82
10	1	1024	1.72889008255961	0.00000000000032	16.14
1	2	2			
2	2	4			
3	2	8	1.69758801672736		
4	2	16	1.73241120825375	0.03482319152639	
5	2	32	1.72889036755784	−0.00352084069591	−9.89
6	2	64	1.72889008234983	−0.00000028520802	12344.82
7	2	128	1.72889008255587	0.00000000020604	−1384.21
8	2	256	1.72889008255957	0.00000000000370	55.66
9	2	512	1.72889008255963	0.00000000000006	59.38
10	2	1024	1.72889008255963	<0.00000000000001	−20.44

4^{j+1}

$Q_{ij} = 4^{0+1} = 4$

$Q_{ij} = 4^{1+1} = 16$

$Q_{ij} = 64$

$\hat{T}_{9.0}$ $\hat{T}_{9.1}$ $\hat{T}_{9.2}$ $\hat{T}_{9.3}$ ... $\hat{T}_{9.9}$

$\hat{T}_{10.0}$ $\hat{T}_{10.1}$ $\hat{T}_{10.2}$

round off. error. tends to dominate error.

We may take. $\hat{T}_{9.2}$ as an estimate of numerical integration