4.6 Numerical Methods of Integrations

Experience in previous chapters told us that find the exact value of an integral could be very difficult. While it is easy to integrate $\int_a^b xe^{-x^2}\,dx$, no one has even managed to find the exact value of $\int_a^b e^{-x^2}\,dx$ even though this integral is important in statistics (normal distribution). In view of this, mathematicians have developed various workable way of find the *approximated* values of a definite integral.

4.6.1 Left-Hand, Mid-Point, and Right-Hand Sums

The key idea behind left-hand, mid-point, right-hand sums is to approximate the region under the graph $y = f(x)$ by bar-charts (i.e. rectangles). Consider a function $f : [a, b] \to \mathbb{R}$. We define P_n to be the uniform partition of $[a, b]$:

$$P_n : a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$$

where $x_i = a + \frac{b-a}{n} i$. We then define

$$\begin{aligned}
L_n &:= \frac{b-a}{n} \sum_{i=1}^{n} f(x_{i-1}) \\
M_n &:= \frac{b-a}{n} \sum_{i=1}^{n} f\left(\frac{x_{i-1} + x_i}{2}\right) \\
R_n &:= \frac{b-a}{n} \sum_{i=1}^{n} f(x_i)
\end{aligned}$$

to be the n-th left-hand sum, mid-point sum, and right-hand sum respectively.

Practically, we could choose n to be a large integer (say 100) and compute L_{100}, M_{100} and R_{100} directly using, for instance, a spreadsheet app. We leave it for readers to play around with Excel on computing these sums. Our emphasis in this section is to determine how accuracy are these approximations.

Proposition 4.23 Let f be a C^1 function on $[a, b]$. Then, the error between left-hand sum L_n and right-hand sum R_n (defined previously) and the actual integral $\int_a^b f(x)\,dx$ is bounded by:

$$\begin{aligned}
\left| \int_a^b f(x)\,dx - L_n \right| &\le \frac{(b-a)^2}{2n} \sup_{[a,b]} |f'| \\
\left| \int_a^b f(x)\,dx - R_n \right| &\le \frac{(b-a)^2}{2n} \sup_{[a,b]} |f'|
\end{aligned}$$

Proof. Consider the uniform partition P_n of $[a, b]$ and denote the partition points by x_i's where $i = 0, 1, \cdots, n$. By the Newton-Leibniz's Formula, we get

$$f(x) = f(x_{i-1}) + \int_{x_{i-1}}^{x} f'(t)\,dt, \qquad \forall x \in [x_{i-1}, x_i].$$

Then, we have

$$\begin{aligned}
\int_a^b f(x)\,dx = \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} f(x)\,dx &= \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} \left(f(x_{i-1}) + \int_{x_{i-1}}^{x} f'(t)\,dt \right) dx \\
&= \underbrace{\sum_{i=1}^{n} f(x_{i-1})(x_i - x_{i-1})}_{=L_n} + \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} \left(\int_{x_{i-1}}^{x} f'(t)\,dt \right) dx.
\end{aligned}$$

The first term is exactly L_n, so the second double integral term gives the error between $\int_a^b f(x)\,dx$ and L_n. Next we estimate:

$$\left|\int_{x_{i-1}}^{x} f'(t)\,dt\right| \le \int_{x_{i-1}}^{x} |f'(t)|\,dt \le \int_{x_{i-1}}^{x} \sup_{[a,b]}|f'|\,dt = \sup_{[a,b]}|f'|\cdot(x - x_{i-1}).$$

This shows

$$\begin{aligned}
\left|\int_a^b f(x)\,dx - L_n\right| &= \left|\sum_{i=1}^{n}\int_{x_{i-1}}^{x_i}\left(\int_{x_{i-1}}^{x} f'(t)\,dt\right)dx\right| \\
&\le \sum_{i=1}^{n}\left|\int_{x_{i-1}}^{x_i}\left(\int_{x_{i-1}}^{x} f'(t)\,dt\right)dx\right| \le \sum_{i=1}^{n}\int_{x_{i-1}}^{x_i}\left|\int_{x_{i-1}}^{x} f'(t)\,dt\right|dx \\
&\le \sum_{i=1}^{n}\int_{x_{i-1}}^{x_i}(x - x_{i-1})\sup_{[a,b]}|f'|\,dx \\
&= \sum_{i=1}^{n}\frac{(x_i - x_{i-1})^2}{2}\sup_{[a,b]}|f'| = \sum_{i=1}^{n}\frac{(b-a)^2}{2n^2}\sup_{[a,b]}|f'| \\
&= \frac{(b-a)^2}{2n}\sup_{[a,b]}|f'|
\end{aligned}$$

where the we used the fact that P_n is a uniform partition so that $x_i - x_{i-1} = \frac{b-a}{n}$.

The proof for the right-hand sum is similar, *mutatis mutandis*. ■

In the above proof, you may use instead the mean-value theorem instead of integral remainder of Taylor series. We use the latter because it may result in a sharper estimate in some other error estimations.

■ **Exercise 4.57** Write up the proof of Proposition 4.23 using mean-value theorem instead.

■ **Exercise 4.58** Write up the proof of the right-hand sum part in Proposition 4.23. Clearly point out what are the essential differences from the proof of the left-hand sum.

■ **Example 4.15** To see how large n need to be in order estimate $\int_1^3 e^{-x^2}\,dx$ up to 4 decimal places, we need to find an n so that

$$\frac{(3-1)^2}{2n}\sup_{[1,3]}\left|(e^{-x^2})'\right| < 0.00001.$$

By straight-forwarding differentiation, we get

$$\frac{d}{dx}e^{-x^2} = -2xe^{-x^2} \implies \left|2xe^{-x^2}\right| \le 2\times 3\times e^{-1}\ \ \forall x\in[1,3] \implies \sup_{[1,3]}\left|(e^{-x^2})'\right| \le \frac{6}{e}.$$

So we need an n such that

$$\frac{4}{2n}\cdot\frac{6}{e} < 0.00001.$$

It can be achieved when $n \ge 441456$.

■ **Exercise 4.59** Find an n such that the left-hand sum L_n gives an approximation of $\int_{-2}^{1} \sin(x^2)\,dx$ with accuracy up to 5 decimal places.

Note that the n that we find above may not be the least possible n.

Next we discuss the error estimation of the mid-point sum. We want to prove a more general result, that is taking the sample point x_i^* in any sub-interval $[x_{i-1}, x_i]$ of a uniform partition so that it makes the ratio of $1-\lambda : \lambda$ with the end points x_{i-1} and x_i.

Proposition 4.24 Let $f(x) : [a,b] \to \mathbb{R}$ be a C^1 function defined on a bounded interval $[a,b]$. Fix a constant $\lambda \in [0,1]$ and a large positive integer n. Consider the uniform partition of $[a,b]$:

$$\{a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b\}$$

and define a numerical approximation of $\int_a^b f(x)\,dx$ by:

$$A_n := \sum_{i=1}^{n} f(x_i^*) \cdot (x_i - x_{i-1}), \quad \text{where } x_i^* := (1-\lambda)x_{i-1} + \lambda x_i.$$

Then we have:

$$\left| \int_a^b f(x)\,dx - A_n \right| \leq \frac{(1-2\lambda+2\lambda^2)(b-a)^2}{2n} \sup_{[a,b]} |f'|\,.$$

Proof. The proof is a modification of that of Proposition 4.23. Since we have demonstrated the use of integral remainder when proving Proposition 4.23, we use mean-value theorem this time.

For any $x \in [x_{i-1}, x_i]$, the mean-value theorem shows there exists c_i (depending on i and x) between x and x_i^* such that

$$f(x) = f(x_i^*) + f'(c_i)(x - x_i^*).$$

Then,

$$\int_a^b f(x)\,dx = \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} f(x)\,dx = \underbrace{\sum_{i=1}^{n} f(x_i^*)(x_i - x_{i-1})}_{=A_n} + \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} f'(c_i)(x - x_i^*)\,dx.$$

The first term is A_n, so the second term gives the error between the integral $\int_a^b f(x)\,dx$ and A_n. Next we estimate the second integral:

$$\begin{aligned} &\left| \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} f'(c_i)(x - x_i^*)\,dx \right| \\ &\leq \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} |f'(c_i)|\,|x - x_i^*|\,dx \\ &\leq \sum_{i=1}^{n} \sup_{[a,b]} |f'| \int_{x_{i-1}}^{x_i} |x - x_i^*|\,dx. \end{aligned}$$

Then we need to compute the integral of $|x - x_i^*|$ over $[x_{i-1}, x_i]$. Note that $x - x_i^* \leq 0$ on

$[x_{i-1}, x_i^*]$ and $x - x_i^* \geq 0$ on $[x_i^*, x_i]$, so we have

$$\begin{aligned}\int_{x_{i-1}}^{x_i} |x - x_i^*|\, dx &= \int_{x_{i-1}}^{x_i^*} (x_i^* - x)\, dx + \int_{x_i^*}^{x_i} (x - x_i^*)\, dx \\ &= \frac{1}{2}(x_i^* - x_{i-1})^2 + \frac{1}{2}(x_i - x_i^*)^2 \\ &= \frac{1}{2}\lambda^2 (x_i - x_{i-1})^2 + \frac{1}{2}(1-\lambda)^2 (x_i - x_{i-1})^2 \\ &= \frac{(b-a)^2}{2n^2}(2\lambda^2 - 2\lambda + 1).\end{aligned}$$

That shows

$$\left| \int_a^b f(x)\, dx - A_n \right| \leq \sum_{i=1}^n \sup_{[a,b]} |f'| \cdot \frac{(b-a)^2}{2n^2}(2\lambda^2 - 2\lambda + 1) = \sup_{[a,b]} |f'| \cdot \frac{(b-a)^2}{2n}(2\lambda^2 - 2\lambda + 1)$$

as desired. ■

The quadratic function $2\lambda^2 - 2\lambda + 1$ achieves its minimum at $\lambda = \frac{1}{2}$. Therefore, the mid-point sum tends to give a slightly better estimate among all other λ-sums.

■ **Exercise 4.60 — Source: MATH1024 Spring 2018 Midterm.** Let $f : [a, b] \to \mathbb{R}$ be a C^2 function on $[a, b]$, and let A_n be as in Proposition 4.24. Show that:

$$\left| \int_a^b f(x)\, dx - A_n \right| \leq \frac{|1 - 2\lambda|\,(b-a)^2}{2n} \sup_{[a,b]} |f'| + \frac{(1 - 3\lambda + 3\lambda^2)(b-a)^3}{6n^2} \sup_{[a,b]} |f''|.$$

[Hint: Consider second-order Taylor's approximation and its remainder.]

4.6.2 Trapezoidal Rule

The trapezoidal rule, as the name implies, approximates the area under the graph of a function by trapeziums. It typically give a better approximation than left-hand and right-hand sums because the trapeziums form a piecewise linear graph that fix the function better than step functions.

Given a function $f : [a, b] \to \mathbb{R}$, we again consider uniform partitions $\{x_i = a + i\Delta x\}_{i=0}^n$ where $\Delta x = \frac{b-a}{n}$. Then, the total area of these trapeziums (as show in Figure **??**) is given by

$$\begin{aligned} T_n &= \left(\frac{f(x_0) + f(x_1)}{2} + \frac{f(x_1) + f(x_2)}{2} + \frac{f(x_2) + f(x_3)}{2} + \cdots + \frac{f(x_{n-1}) + f(x_n)}{2} \right) \cdot \Delta x \\ &= \left(\frac{f(x_0) + f(x_n)}{2} + f(x_1) + f(x_2) + \cdots + f(x_{n-1}) \right) \cdot \Delta x \\ &= \left(\frac{f(a) + f(b)}{2} + f(x_1) + f(x_2) + \cdots + f(x_{n-1}) \right) \cdot \frac{b-a}{n}. \end{aligned}$$

With this formula, the trapezium sum T_n can be easily computed using spreadsheet apps. As before, we are more interested in its error estimation:

Proposition 4.25 Let $f : [a, b] \to \mathbb{R}$ be a C^2 function on $[a, b]$, and T_n be the n-th trapezoidal sum of the integral $\displaystyle\int_a^b f(x)\, dx$ then we have

$$\left| \int_a^b f(x)\, dx - T_n \right| \leq \frac{(b-a)^3}{12n^2} \sup_{[a,b]} |f''|.$$

Proof. Denote the partition points by $x_i = a + i\Delta x$ where $\Delta = \frac{b-a}{n}$. First, we denote

$$A_i = -\frac{x_{i+1} + x_i}{2} \quad \text{and} \quad B = -\left(\frac{x_{i+1} - x_i}{2} \right)^2 = -\frac{(\Delta x)^2}{4}.$$

It can be verified easily that:

$$\Big[(x+A_i)f(x)\Big]_{x_i}^{x_{i+1}} = \frac{f(x_i)+f(x_{i+1})}{2}\cdot\Delta x$$
$$\Big[\big((x+A_i)^2+B\big)f'(x)\Big]_{x_i}^{x_{i+1}} = 0$$

for any $i=0,1,2,\dots,n-1$. Then, by the fact that

$$\begin{aligned}
&\frac{d}{dx}\left((x+A_i)f(x)-\frac{1}{2}((x+A_i)^2+B)f'(x)\right)\\
&=f(x)+(x+A_i)f'(x)-\frac{1}{2}\cdot 2(x+A_i)f'(x)-\frac{1}{2}((x+A_i)^2+B)f''(x)\\
&=f(x)-\frac{1}{2}(x+A_i)^2+B)f''(x).
\end{aligned}$$

Hence, using the Newton-Leibniz's Formula and by our choice of A_i and B, we get

$$\begin{aligned}
&\int_{x_i}^{x_{i+1}} f(x)dx-\int_{x_i}^{x_{i+1}}\frac{1}{2}((x+A_i)^2+B)f''(x)\,dx\\
&=\Big[(x+A_i)f(x)-\frac{1}{2}((x+A_i)^2+B)f'(x)\Big]_{x_i}^{x_{i+1}}\\
&=\frac{f(x_i)+f(x_{i+1})}{2}\cdot\Delta x.
\end{aligned}$$

Note that $\dfrac{f(x_i)+f(x_{i+1})}{2}\cdot\Delta x$ is the area of the i-th trapezium, so we conclude that

$$\begin{aligned}
&\int_a^b f(x)\,dx\\
&=\sum_{i=0}^{n-1}\int_{x_i}^{x_{i+1}} f(x)\,dx=\sum_{i=0}^{n-1}\frac{f(x_i)+f(x_{i+1})}{2}\cdot\Delta x+\sum_{i=0}^{n-1}\int_{x_i}^{x_{i+1}}\frac{1}{2}((x+A_i)^2+B)f''(x)\,dx\\
&=T_n+\sum_{i=0}^{n-1}\int_{x_i}^{x_{i+1}}\frac{1}{2}((x+A_i)^2+B)f''(x)\,dx
\end{aligned}$$

Hence, the integral terms give the error between $\displaystyle\int_a^b f(x)\,dx$ and T_n. It worths noting that

$$(x+A_i)^2+B=(x-x_{i+1})(x-x_i)\le 0 \quad \text{on } [x_i,x_{i+1}].$$

Therefore the error term is given by

$$\begin{aligned}
\left|\int_a^b f(x)\,dx-T_n\right| &\le \frac{1}{2}\sum_{i=0}^{n-1}\int_{x_i}^{x_{i+1}}(x_{i+1}-x)(x-x_i)\left|f''(x)\right|\,dx\\
&\le \frac{1}{2}\sum_{i=0}^{n-1}\int_{x_i}^{x_{i+1}}(x_{i+1}-x)(x-x_i)\,dx\cdot\sup_{[a,b]}|f''|\\
&=\frac{1}{2}\sum_{i=0}^{n-1}\frac{1}{6}(\Delta x)^3\cdot\sup_{[a,b]}|f''|=\frac{(b-a)^3}{12n^2}\sup_{[a,b]}|f''|.
\end{aligned}$$

The third step follows from direct computation of the integral $\displaystyle\int_{x_i}^{x_{i+1}}(x_{i+1}-x)(x-x_i)\,dx$ (left as an exercise for readers). It completes the proof. ■

■ **Exercise 4.61** From the above proof, what can you say about the integral $\int_a^b f(x)\,dx$ and T_n when $f'' > 0$ on $[a,b]$?

4.6.3 Simpson's Rule

The Simpson's rule approximates the graph of a function by quadratic curves. Given a continuous function $f : [a,b] \to \mathbb{R}$, we consider a sequence of uniform partitions $\{P_{2n}\}$ of $[a,b]$ with $2n$ subintervals. Denote the partition points to be $\{x_i\}_{i=0}^{2n}$, then on each interval $[x_{2i}, x_{2i+2}]$ where $i = 0, 1, \cdots, n-1$, we approximate f by a quadratic function $Q_i(x)$ so that $Q_i(x_{2i}) = f(x_{2i})$, $Q_i(x_{2i+1}) = f(x_{2i+1})$ and $Q_i(x_{2i+2}) = f(x_{2i+2})$. One can find that that such an $Q_i(x)$ can be written as:

$$\begin{aligned} Q_i(x) = f(x_{2i}) \cdot \frac{(x - x_{2i+1})(x - x_{2i+2})}{(x_{2i} - x_{2i+1})(x_{2i} - x_{2i+2})} + f(x_{2i+1}) \cdot \frac{(x - x_{2i})(x - x_{2i+2})}{(x_{2i+1} - x_{2i})(x_{2i+1} - x_{2i+2})} \\ + f(x_{2i+2}) \cdot \frac{(x - x_{2i})(x - x_{2i+1})}{(x_{2i+2} - x_{2i})(x_{2i+2} - x_{2i+1})}. \end{aligned}$$

One easy way to see this is to observe that

$$\frac{(x - x_{2i+1})(x - x_{2i+2})}{(x_{2i} - x_{2i+1})(x_{2i} - x_{2i+2})}$$

equals 1 when $x = x_{2i}$, and equals 0 when $x = x_{2i+1}$ or x_{2i+2}. Similar for the second and third terms.

■ **Exercise 4.62** Given n distinct numbers $x_1 < x_2 < \cdots < x_n$, and a set of n numbers $y_1, \cdots, y_n$ (not necessarily distinct), find an n-th degree polynomial $P(x)$ such that $P(x_i) = y_i$ for any $i = 1, 2, \cdots, n$.

Q_i is simply a quadratic function, so the integral below can be found easily:

$$\int_{x_{2i}}^{x_{2i+2}} Q_i(x)\,dx = \frac{b-a}{6n}\left(f(x_{2i}) + 4f(x_{2i+1}) + f(x_{2i+2})\right)$$

using the fact that P_{2n} is a uniform partition so that $x_{2i+2} - x_{2i+1} = x_{2i+1} - x_{2i} = \frac{b-a}{2n}$. This is left as an exercise for readers.

■ **Exercise 4.63** Compute $\int_{x_{2i}}^{x_{2i+2}} Q_i(x)\,dx$.

Summing up the area of Q_i's, we get the following approximated value of $\int_a^b f(x)\,dx$:

$$S_n := \sum_{i=0}^{n-1} \int_{x_{2i}}^{x_{2i+2}} Q_i(x)\,dx = \frac{b-a}{6n}\left(f(a) + f(b) + 2\sum_{i=1}^{n-1} f(x_{2i}) + 4\sum_{i=0}^{n-1} f(x_{2i+1})\right)$$

For the error estimation of the Simpson's rule, it can be shown to be of order $O(1/n^4)$ provided that f is C^4 on $[a,b]$:

Proposition 4.26 Let $f : [a,b] \to \mathbb{R}$ be a C^4 function, then there exists a universal constant $C > 0$ such that

$$\left|\int_a^b f(x)\,dx - S_n\right| \le \frac{(b-a)^5}{Cn^4} \sup_{[a,b]} \left|f^{(4)}\right|.$$

Outline of Proof. The key idea is to use the Lagrange's remainder theorem, which asserts that for each $i = 0, 1, \cdots, n-1$ and any $x \in [x_{2i}, b]$, there exists $h_1(x) \in [x_{2i}, x]$ such that:

$$f(x) = \underbrace{f(x_{2i}) + f'(x_{2i})(x - x_{2i}) + \frac{f''(x_{2i})}{2}(x - x_{2i})^2 + \frac{f'''(x_{2i})}{3!}(x - x_{2i})^3}_{=:P_i(x)} + \frac{f^{(4)}(h_1(x))}{4!}(x - x_{2i})^4.$$

That would give the error estimation of

$$\left| \int_{x_{2i}}^{x_{2i}+2\Delta x} f(x)\,dx - \int_{x_{2i}}^{x_{2i}+2\Delta x} P_i(x)\,dx \right|$$

in terms of $\sup_{[a,b]} \left| f^{(4)} \right|$. Here $\Delta x = \frac{b-a}{2n}$.

Next we estimate the error of

$$\left| \int_{x_{2i}}^{x_{2i}+2\Delta x} P_i(x)\,dx - \int_{x_{2i}}^{x_{2i}+2\Delta x} Q_i(x)\,dx \right|.$$

Recall that

$$\int_{x_{2i}}^{x_{2i+2}} Q_i(x)\,dx = \frac{b-a}{6n}(f(x_{2i}) + 4f(x_{2i} + \Delta x) + f(x_{2i} + 2\Delta x)).$$

Writing

$$f(x_{2i} + \Delta x) = P_i(x_{2i} + \Delta x) + \frac{f^{(4)}(h_1(x_{2i} + \Delta x))}{4!}(\Delta x)^4$$

and similarly for $f(x_{2i} + 2\Delta x)$, one can see there is a lot of cancellations within

$$\int_{x_{2i}}^{x_{2i}+2\Delta x} P_i(x)\,dx - \int_{x_{2i}}^{x_{2i}+2\Delta x} Q_i(x)\,dx,$$

and the only terms left are the 4th derivatives of f.

By considering

$$\begin{aligned}
&\left| \int_{x_{2i}}^{x_{2i}+2\Delta x} f(x)\,dx - \int_{x_{2i}}^{x_{2i}+2\Delta x} Q_i(x)\,dx \right| \\
\leq &\left| \int_{x_{2i}}^{x_{2i}+2\Delta x} f(x)\,dx - \int_{x_{2i}}^{x_{2i}+2\Delta x} P_i(x)\,dx \right| + \left| \int_{x_{2i}}^{x_{2i}+2\Delta x} P_i(x)\,dx - \int_{x_{2i}}^{x_{2i}+2\Delta x} P_i(x)\,dx \right|,
\end{aligned}$$

we get the error estimate on each subinterval $[x_{2i}, x_{2i+2}]$ in terms of 4th derivatives of f. One could sum up these error to yield the desired result.

■

■ **Exercise 4.64** Fill in the detail of the above outline of proof.

ⓘ It is interesting to note that if $f(x)$ is a cubic polynomial, then $f^{(4)} \equiv 0$ so the Simpson's rule indeed gives the exact value of $\int_a^b f(x)\,dx$. Of course, practically speaking we wouldn't integrate a cubic polynomial in this way.