Approximate Integration

1. **Quote.** "All exact science is dominated by the idea of approximation."

   (Bertrand Russell, English Logician and Philosopher 1872-1970)

2. **Quote.** "And now that you don't have to be perfect, you can be good."

   (John Steinbeck, American Author, 1902-1968)

3. **Problem.** Evaluate $\frac{1}{\sqrt{2\pi}} \int_{-3}^{3} e^{-x^2/2}\, dx$.

   The integrand is the probability density function for a normal distribution (Gaussian) with mean $\mu = 0$ and standard deviation $\sigma = 1$. The integral represents the probability of a normally distributed random variable to be observed within three standard deviations from the mean. This probability is approximately equal to 0.9973.

4. **Reminder.**

If f is continuous on $[a, b]$ and if $[a, b]$ is divided into n subintervals

$$[a = x_0, x_1], [x_1, x_2], \ldots, [x_{n-1}, x_n = b]$$

of equal length $\Delta x = \frac{b-a}{n}$ then

$$\int_a^b f(x)dx \approx \sum_{i=1}^{n} f(x_i^*)\Delta x$$

where x_i^* is any point in $[x_{i-1}, x_i]$.

5. **Ways of choosing the sample points x_i^*:**

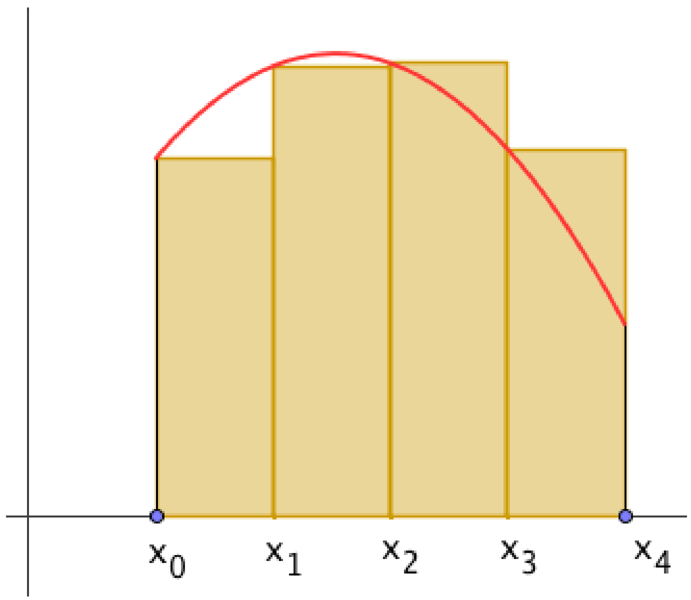
Endpoint Approximation.

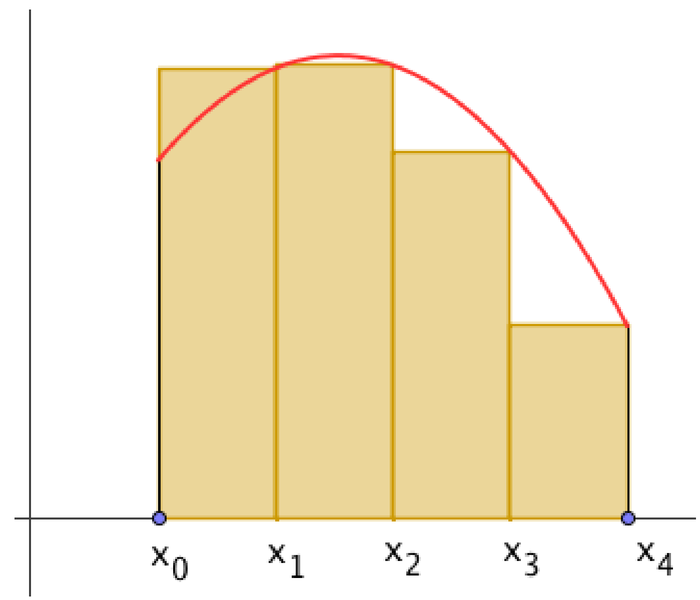
The left-point approximation L_n and the right-point approximation R_n to $\int_a^b f(x)dx$ with
$\Delta x = \frac{b-a}{n}$ are

$$L_n = \sum_{i=1}^{n} f(x_{i-1})\Delta x$$

and

$$R_n = \sum_{i=1}^{n} f(x_i)\Delta x.$$





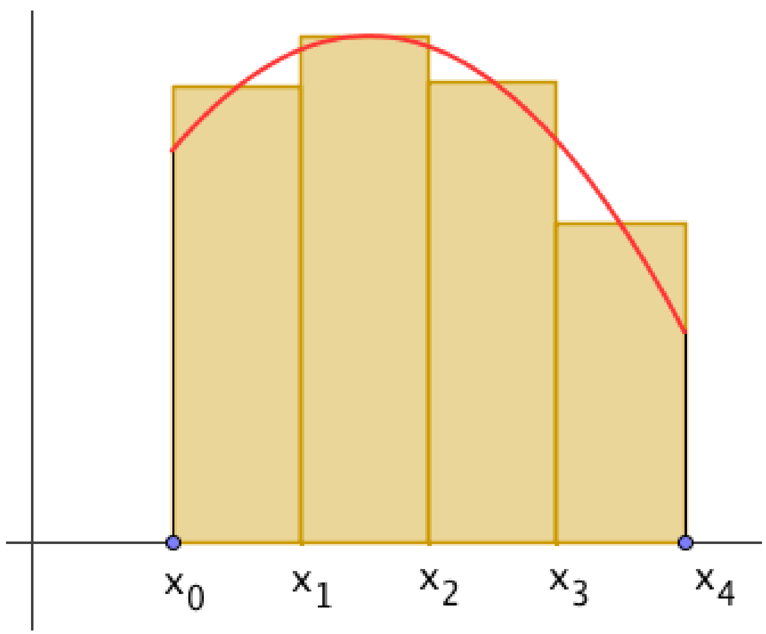
Midpoint Approximation.

The midpoint approximation M_n with $\Delta x = \frac{b-a}{n}$ is

$$M_n = \sum_{i=1}^{n} f(\overline{x}_i)\Delta x$$

where

$$\overline{x}_i = \frac{x_{i-1} + x_i}{2}.$$



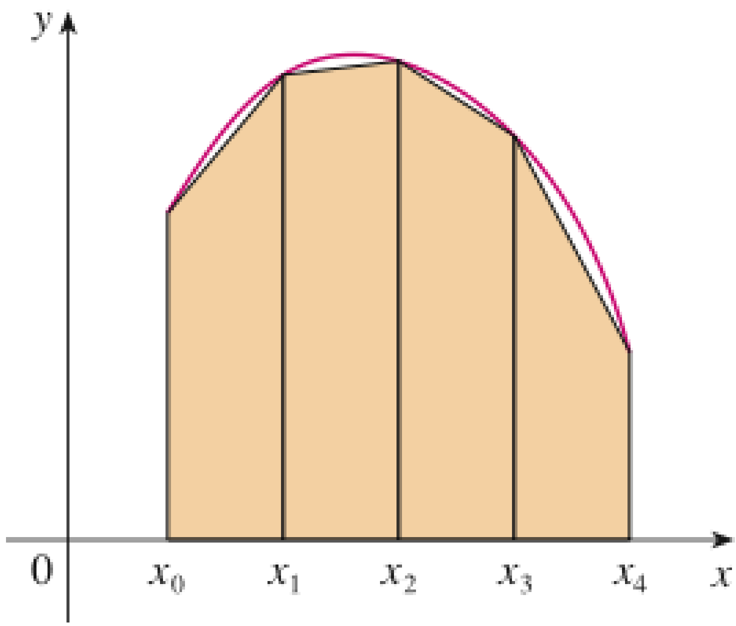
Trapezoid Rule.

The trapezoidal approximation to

$$\int_a^b f(x)dx \text{ with } \Delta x = \frac{b-a}{n}$$

is

$$T_n = \frac{\Delta x}{2}[f(x_0) + 2f(x_1) + 2f(x_2) + \ldots + 2f(x_{n-1}) + f(x_n)].$$



6. **Systematic way to derive formulas**

Integration (quadrature) formulas are generally of the form

$$\int_a^b f(x)dx \approx Q_n = \sum_{j=0}^{n} w_j f(x_j).$$

The x_j are the abscissae, and the w_j are the weights of the formula.

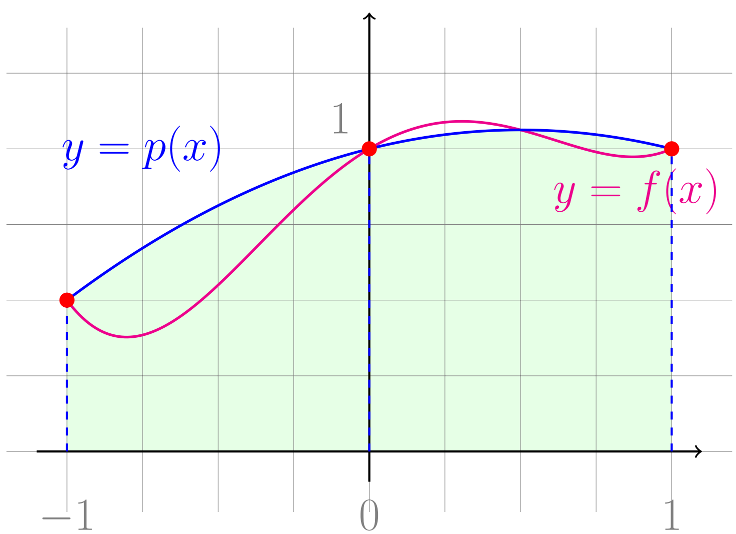
We know how to integrate polynomials; so the idea behind numerical integration is to *replace* the function f by a polynomial that interpolates that function at the abscissae. Find a polynomial (of degree $\leq n$), such that

$$p(x_j) = f(x_j), \qquad j = 0, 1, \ldots, n$$

and

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

Note that there are $n+1$ of the x_j.



7. **Equidistant abscissae**

We are now deriving fomulas for the interval $[a, b] = [-1, 1]$. At first we assume the abscissae are equispaced.

(a) $n = 0$. Clearly, $x_0 = 0$, and the requirement that we get the exact value for $p(x) = 1$ leads to $w_0 = 2$. We've rediscovered the midpoint rule!

However, we get something for free! Because of symmetry, the midpoint rule also gives the exact answer for $p(x) = x$:

$$\int_{-1}^{+1} x \, dx = 0 = 2 \cdot 0.$$

So the midterm rule is actually exact for polynomials of degree ≤ 1; note, it does not give the exact value for a quadratic term:

$$\frac{2}{3} = \int_{-1}^{1} x^2 \, dx \neq 0 = \text{midpoint rule}$$

(b) $n = 1$. Rule = $w_0 f(x_0) + w_1 f(x_1)$.

The points are $x_0 = -1$, $x_1 = 1$. Let $f_0 = f(x_0), f_1 = f(x_1)$. The polynomial of degree ≤ 1 (linear) interpolating the function f at the endpoints is

$$p(x) = f_0 + \frac{f_1 - f_0}{x_1 - x_0}(x - x_0) = f_0 + \frac{1}{2}(f_1 - f_0)(x - 1).$$

$$\int_{-1}^{1} p(x)dx = \int_{-1}^{1} f_0 + \frac{1}{2}(f_1 - f_0)(x - 1)\ dx = 2f_0 + \left[\frac{1}{4}(f_1 - f_0)(x - 1)^2\right]_{-1}^{1} = f_0 + f_1.$$

Alternatively, solve for the weights w_0 and w_1, requiring integration to be exact for $p(x) = 1$, and $p(x) = x$:

$$\int_{-1}^{1} 1dx = 2 = w_0 + w_1, \qquad \int_{-1}^{1} xdx = 0 = w_1 - w_0$$

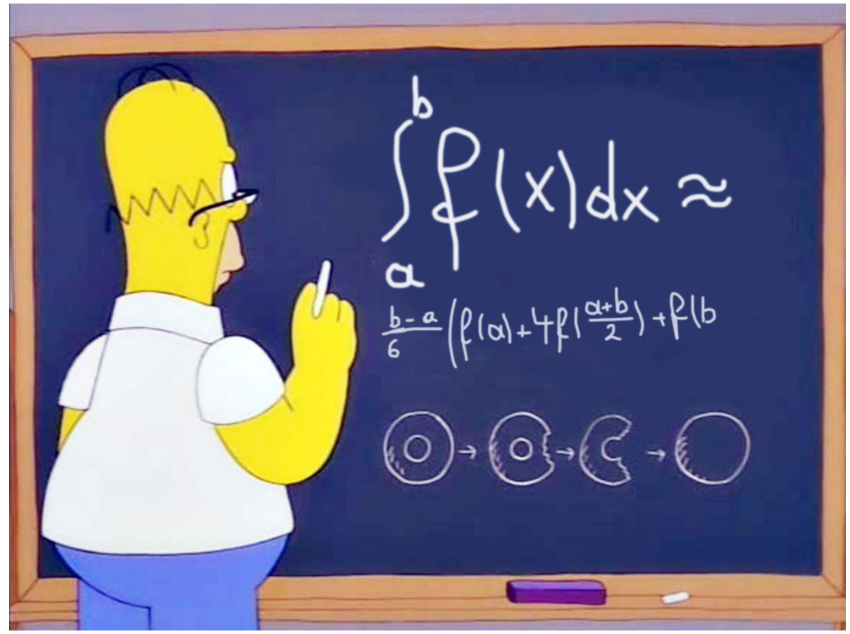
hence $w_0 = w_1 = 1$.

Either way, this gives us the trapezoidal rule, $2\frac{1}{2}(f(-1) + f(1))$, exact for polynomials of degree ≤ 1.

Again, we do not get the correct value if we integrate the function x^2.

(c) $n = 2$. Rule = $w_0 f(x_0) + w_1 f(x_1) + w_2 f(x_2)$.
The points are $x_0 = -1$, $x_1 = 0$, and $x_2 = 1$. Solve for the weights w_0, w_1 and w_2, requiring integration to be exact for $p(x) = 1$, $p(x) = x$ and $p(x) = x^2$:

$$\int_{-1}^{1} 1\, dx = 2 = w_0 + w_1 + w_2$$
$$\int_{-1}^{1} x dx = 0 = w_2 - w_0$$
$$\int_{-1}^{1} x^2 dx = \frac{2}{3} = w_0 + w_2$$



hence

$$w_0 = w_2 = \frac{1}{3}, \qquad w_1 = \frac{4}{3}.$$

And that's Simpson's rule!

By design it is exact for polynomials of degree ≤ 2.

BONUS: Because of symmetry, Simpson's rule is also exact for $p(x) = x^3$.

8. Composite rules

Divide the interval of integration $[a,b]$ into n subintervals of equal length, $h=\frac{b-a}{n}$, $x_k=a+kh$, $k=0,\dots,n$, and write the integral as

$$\int_a^b f(x)dx=\sum_{k=0}^{n-1}\int_{x_k}^{x_k+h} f(x)dx.$$

Apply simple quadrature formulas to each of the integrals in the summation.

$$\begin{aligned}
M(h) &= h\sum_{k=0}^{n-1} f(x_k+\tfrac{h}{2}) \\
T(h) &= h\left(\frac{1}{2}f(x_0)+f(x_1)+\cdots+f(x_{n-1})+\frac{1}{2}f(x_n)\right) \\
S(h) &= \frac{h}{3}\left(\frac{1}{2}f(x_0)+\sum_{k=1}^{n-1} f(x_k)+\frac{1}{2}f(x_n)+2\sum_{k=0}^{n-1} f(x_k+\tfrac{h}{2})\right).
\end{aligned}$$

We can see easily that we have the relationships

$$\begin{aligned}
T(\tfrac{h}{2}) &= \frac{1}{2}\left(T(h)+M(h)\right), \\
S(h) &= \frac{1}{3}T(h)+\frac{2}{3}M(h) \\
&= \frac{1}{3}T(h)+\frac{2}{3}\left(2T(\tfrac{h}{2})-T(h)\right) \\
&= \frac{4}{3}T(\tfrac{h}{2})-\frac{1}{3}T(h).
\end{aligned}$$

9. **Computation** We approximate the integral $\int_0^1 xe^x \, dx$ with the trapezoidal rule and with Simpson's rule. We know the exact answers in this case, and can therefore observe the errors.

n	Trapezoid $T(1/n)$	Simpson $S(1/n)$
1	1.3591409142	-
2	1.0917507748	1.002620728310
4	1.0230644791	1.000169047140
8	1.0057741074	1.000010650140
16	1.0014440271	1.000000666968
32	1.0003610380	1.000000041706
64	1.0000902615	1.000000002607
128	1.0000225655	1.000000000163

10. **Example.** Calculate an approximation to the integral

$$\int_0^3 x^2 dx$$

with $n = 6$ and $\Delta x = 0.5$ by using

(a) midpoint approximation

(b) trapezoidal approximation

11. **Errors in Approximation:**

The **error** E in using an approximation is defined to be the difference between the actual value and the approximation A. That is,

$$E = \int_a^b f(x)\,dx - A$$

It turns out that the size of the error depends on the second derivative of the function f, which measures how much the graph is curved.

The following fact is usually proved in a course on *numerical analysis* (MACM316), so we just state it here.

Error bounds. Suppose that $|f''^{(x)}| \le K$ for x in the interval $[a,b]$. If E_T and E_M are the errors in the Trapezoidal and Midpoint Rules then

$$|E_T| \le \frac{K(b-a)^3}{12n^2} \quad \text{and} \quad |E_M| \le \frac{K(b-a)^3}{24n^2}.$$

$$|E_T| \le \frac{K(b-a)}{12}h^2 \quad \text{and} \quad |E_M| \le \frac{K(b-a)}{24}h^2.$$

Note: $|f''(x)|$ "measures" how far the function is away from being a line; the length of the interval is part of the estimate; and, in both cases the error (bound) is quadratic, i.e., if you double the number of points the error should be divided by approximately **4**.

12. **Example.** Since

$$\int_1^2 \frac{dx}{x} = \ln 2$$

the Trapezoidal and Midpoint Rules could be used to approximate $\ln 2$. Estimate the errors in the the Trapezoidal and Midpoint approximations of this integral by using $n = 10$ intervals.

13. **Example.**

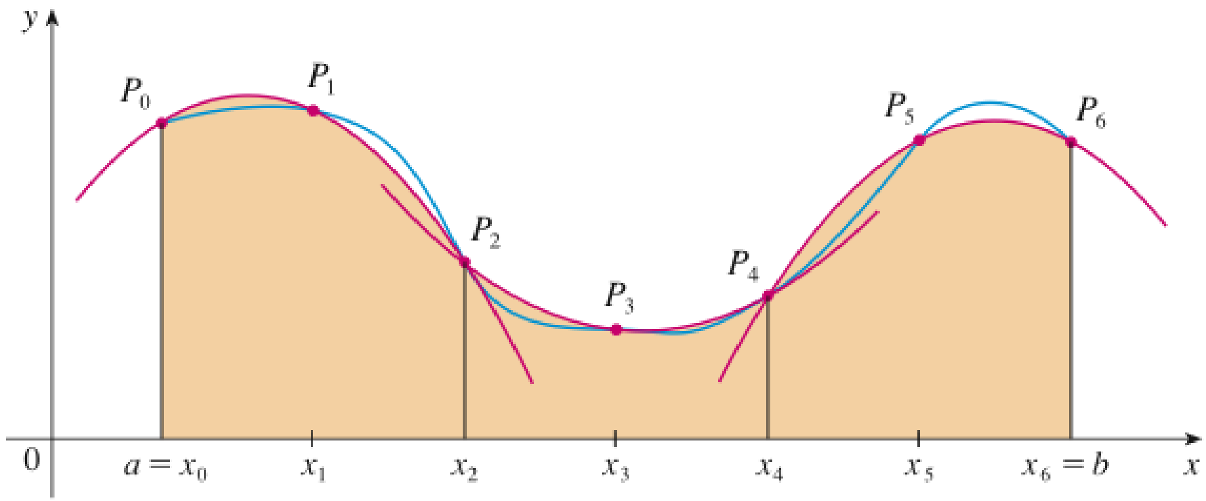
(a) Use the Midpoint Rule with $n = 10$ to approximate the integral $\displaystyle\int_0^1 e^{-x^2}\,dx$.

(b) Give an upper bound for the error involved in this approximation.

(c) How large do we have to choose n so that the approximation M_n to the integral in part (a) is accurate to within 0.00001?

14. **Approximation using parabolic segments – Simpson revisited:**

Let f be continuous on $[a,b]$ and divide the interval into an *even number* n subintervals of equal length $\Delta x = \frac{b-a}{n}$. Suppose the endpoints of these subintervals are, as usual, $a = x_0, x_1, x_2, \ldots, x_n = b$.



Let P_i be the point $(x_i, f(x_i))$. For each even number $i < n$ we approximate the area under the curve $y = f(x)$ over the interval $[x_i, x_{i+2}]$ by the area under the unique parabola that passes through the points P_i, P_{i+1}, and P_{i+2} over the same interval.

15. **Simpson's Rule.**

Let f be continuous on $[a, b]$ and $\Delta x = \frac{b-a}{n}$ with n even.

Then we can approximate $\displaystyle\int_a^b f(x)\,dx$ by the sum

$$S_n = \frac{\Delta x}{3}[f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + \ldots + 2f(x_{n-2}) + 4f(x_{n-1}) + f(x_n)].$$

16. **Example.** Approximate $\displaystyle\int_0^3 \frac{dx}{1+x^4}$ by Simpson's Rule with $n = 6$.

17. **Error in Simpson's Rule.**

Error Bound in Simpson's Rule.

Suppose that $|f^{(4)}(x)| \leq K$ for all x in the interval $[a, b]$. If E_S is the error in using Simpson's Rule, then

$$|E_S| \leq \frac{K(b-a)^5}{180n^4}.$$

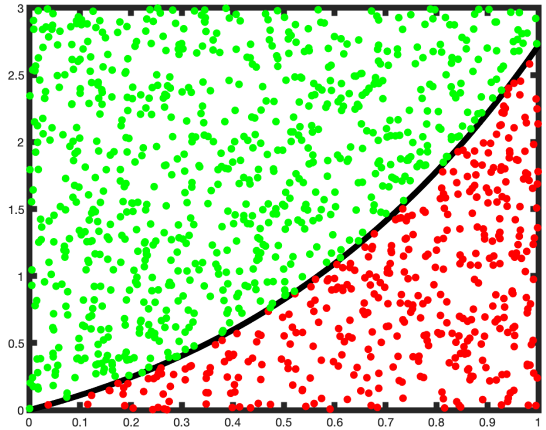
For Simpson, the size of the error depends on the *fourth* derivative of f; It measures how far away f is from being a cubic polynomial.

If you double n, the number of subintervals, the error will be divided by approximately **16** now!

18. **Example.** How large should we take n in order to guarantee that the Simpson's Rule approximation to $\int_1^2 (1/x)dx$ is accurate within 0.0001?

19. How random is this? – Monte Carlo Methods.

Let's compute $\int_0^1 xe^x dx$ again. This time we create random points in the rectangle $0 \leq x \leq 1$, $0 \leq y \leq 3$ (you can think of throwing darts into this rectangle), and count the number of points below the curve. The ratio of hits over total throws will tell us the ratio of the area under the curve and that of the rectangle.



Some results (exact = 1):

n	Monte Carlo	error
100	0.930	-7e-02
1000	0.9660	3.4e-02
10000	0.99210	-7.9e-03
100000	0.996870	3.1e-03
1000000	1.0004640	4.6e-04

Matlab Code:

```
t=rand(n,2);
f1=t(:,1).*exp(t(:,1));  f2=3*t(:,2);
s=f2<f1;
hits=sum(s), relhit=hits/n;
Integral=relhit*3,
Error=Integral-1,
hold off; plot(x,f,'k'); hold;
i1=find(s);i2=find(~s);
plot(t(i1,1),f2(i1),'r.');
plot(t(i2,1),f2(i2),'g.');
```

It looks to get closer to the correct value, but is painfully slow.

20. **Final thoughts**

- In general, higher order rules (Simpson), will achieve the same accuracy with far fewer points than say the trapezoidal rule.
- If f is not "smooth", for example, if its third derivative does not exist, Simpson's rule will do no better than the trapezoidal rule.
- We could free ourselves from the requirement that the abscissae of our quadrature formula have to be spaced evenly. This leads to the Gaussian quadrature formulas.
- In many applications of numerical integration, the function f is not defined by a simple function formula, but may be the result of a complicated calculations. Keeping the number of function evaluations at a minimum while maintaining accuracy is important!

Notes.