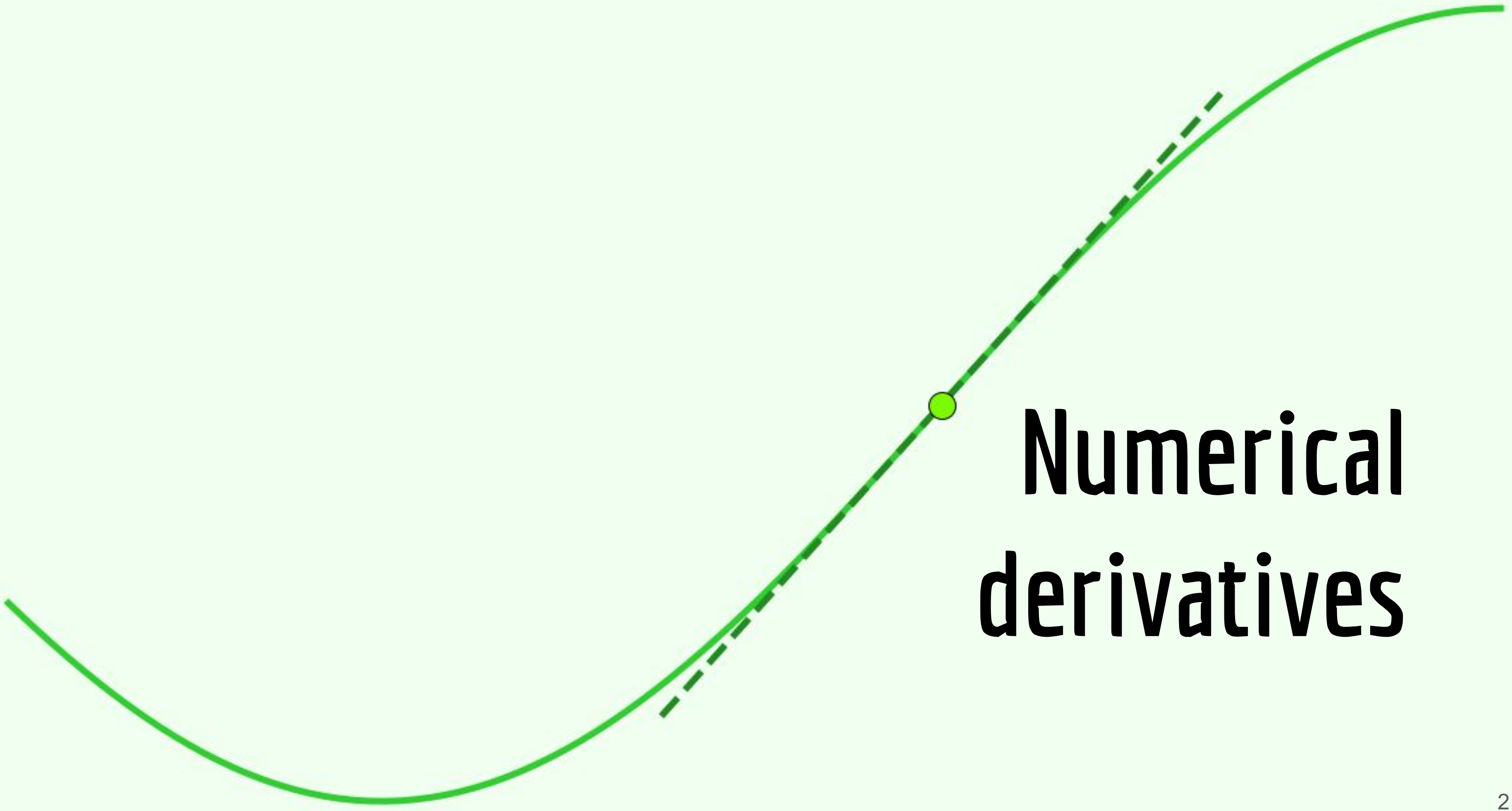
Numerical derivation and integration methods

Kristina Kislyakova
University of Vienna - 05.11.2025



Numerical derivatives

Motivation

Why do we need to compute derivatives numerically? There might be several reasons why need to approximate derivatives, for example:

- We might know the values of the function we need to differentiate only at a sampled data set without knowing the function itself.

Motivation

Why do we need to compute derivatives numerically? There might be several reasons why need to approximate derivatives, for example:

- We might know the values of the function we need to differentiate only at a sampled data set without knowing the function itself.
- In some cases, it may not be obvious that an underlying function exists and all that we have is a discrete data set. We may still be interested in studying changes in the data, which are related, of course, to derivatives.

Motivation

Why do we need to compute derivatives numerically? There might be several reasons why need to approximate derivatives, for example:

- We might know the values of the function we need to differentiate only at a sampled data set without knowing the function itself.
- In some cases, it may not be obvious that an underlying function exists and all that we have is a discrete data set. We may still be interested in studying changes in the data, which are related, of course, to derivatives.
- Sometimes exact formulas are available but they are very complicated, to the point that an exact computation of the derivative requires a lot of function evaluations. It might be significantly simpler to approximate the derivative instead of computing its exact value.

Numerical derivatives (and caveats)

Imagine that you have a function *f(x)*, of which you want to compute the derivative *f'(x)*. The definition of the derivative, namely the limit as $h \to 0$ of

$$f'(x) \approx \frac{f(x+h) - f(x)}{h}$$

suggests a possible numerical procedure to compute this: pick a small value *h*, evaluate *f(x + h)*, if you do not have *f(x)* already evaluated, do that too, and apply the equation above.

Numerical derivatives (and caveats)

Imagine that you have a function *f(x)*, of which you want to compute the derivative *f'(x)*. The definition of the derivative, namely the limit as *h→0* of

$$f'(x) \approx \frac{f(x+h) - f(x)}{h}$$

suggests a possible numerical procedure to compute this: pick a small value *h*, evaluate *f(x + h)*, if you do not have *f(x)* already evaluated, do that too, and apply the equation above.

What can go wrong? *A lot, of course!*

Applied uncritically, this procedure is almost guaranteed to produce inaccurate results. Applied properly, it can be the right way to compute a derivative only when the function *f* is fiercely expensive to compute, when you already have invested in computing *f(x)*, and when, therefore, you want to get the derivative in no more than a single additional function evaluation.

Sources of errors

There are two sources of error when applying the equation

$$f'(x) \approx \frac{f(x+h) - f(x)}{h}$$

to compute the derivative of *f(x)*.

(1) The **truncation error** comes from higher terms in the Taylor series expansion (which we truncate to compute the derivative above):

$$f(x+h) = f(x) + hf'(x) + \frac{1}{2}h^2 f''(x) + \frac{1}{6}h^3 f'''(x) + \cdots$$

so that

$$\frac{f(x+h) - f(x)}{h} = f' + \frac{1}{2}hf'' + \cdots$$

Sources of errors

(2) The **roundoff error** has various contributions.

There is a roundoff error in h: for example, imagine that you are at a point $x = 10.3$ and you choose $h = 0.0001$. Both x and $x + h$ are numbers represented with some fractional error characteristic of the floating-point format ε_m (may be ~ 10^{-7} in single precision). The difference between $x + h$ and x as represented in the machine is of the order of $\varepsilon_m x$, which implies a fractional error in h of order ~ $\varepsilon_m x/h \sim 10^{-2}$.

Sources of errors

(2) The **roundoff error** has various contributions.

There is a roundoff error in h: for example, imagine that you are at a point $x = 10.3$ and you choose $h = 0.0001$. Both x and $x + h$ are numbers represented with some fractional error characteristic of the floating-point format ε_m (may be ~ 10^{-7} in single precision). The difference between $x + h$ and x as represented in the machine is of the order of $\varepsilon_m x$, which implies a fractional error in h of order ~ $\varepsilon_m x/h \sim 10^{-2}$.

With h an "exact" number, the roundoff error in the equation to compute the derivative is $\varepsilon_r \sim \varepsilon_f\,|f(x)/h|$.
Here, ε_f is the fractional accuracy with which f is computed; for a simple function this may be comparable to the machine accuracy, but for a complicated calculation with additional sources of inaccuracy it may be larger.

Sources of errors

The truncation error in the equation for the derivative is of the order of ε_t ~ $|hf''(x)|$.

If you can afford two function evaluations for each derivative calculation, then it is significantly better to use the symmetrized form

$$f'(x) \approx \frac{f(x+h) - f(x-h)}{2h}$$

In this case, with the ap ... error is ~ $\varepsilon_f^{2/3}$, typically 1 or 2 orders of magnitude better than with the non-symmetrized form.

Symbolic derivatives in python

simpy.Derivative

```
# import sympy
from sympy import *

x, y = symbols('x y')
expr = x**2 + 2 * y + y**3
print("Expression : {}".format(expr))

# Use sympy.Derivative() method
expr_diff = Derivative(expr, x)

print("Derivative of expression with respect to x : {}".format(expr_diff))
print("Value of the derivative : {}".format(expr_diff.doit()))
```

```
Expression : x**2 + y**3 + 2*y
Derivative of expression with respect to x : Derivative(x**2 + y**3 + 2*y, x)
Value of the derivative : 2*x
```

Numerical integration methods

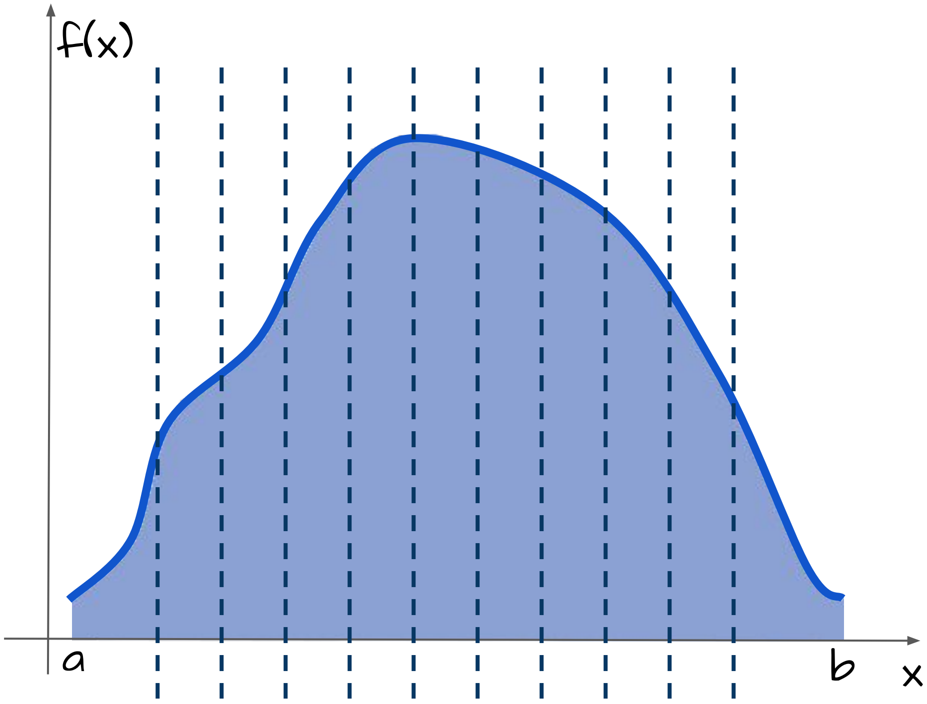
Introduction

Numerical integration (also called *quadrature*) is the procedure used to evaluate:

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

If we divide the interval *[a,b]* into *N* slices, we can compute the integral as a sum of integrals calculated over an individual slice:

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



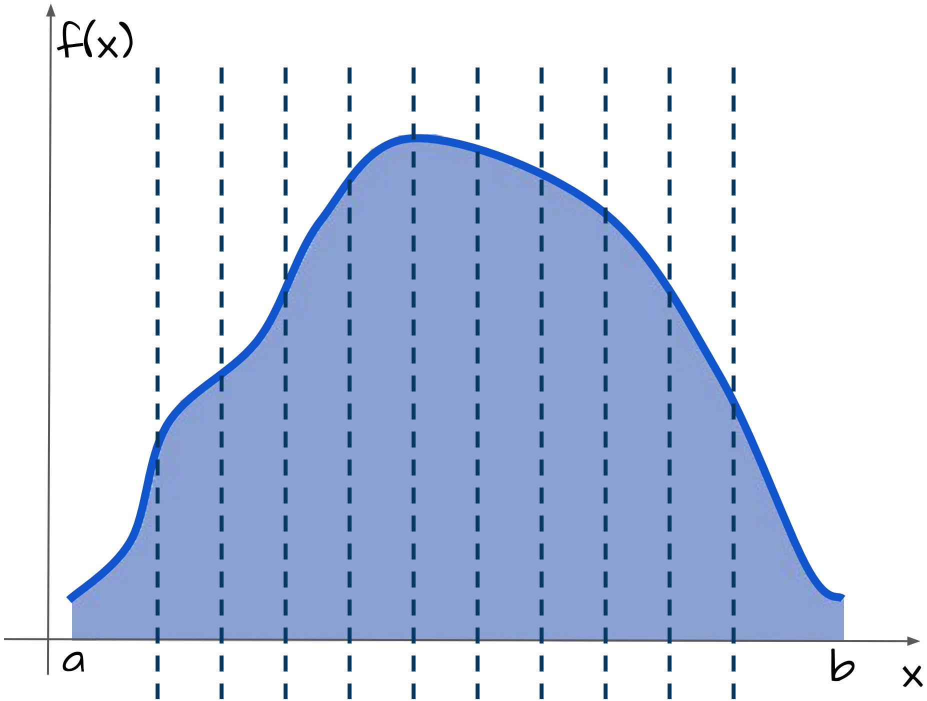
Introduction

Numerical integration (also called *quadrature*) is the procedure used to evaluate:

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

We will look at the following techniques:

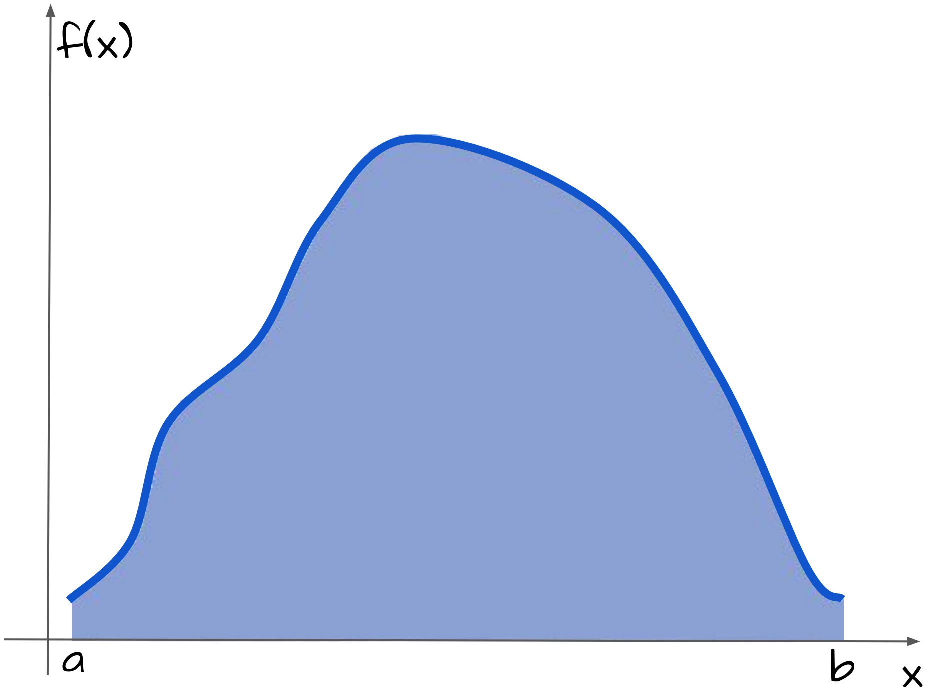
1. **Newton-Cotes formulas**
2. **Gaussian quadrature**
3. **Monte Carlo integration**



1. Newton-Cotes formulas

Idea: approximate the integrand function with a polynomial over a small range of *x*, and compute its integral (easy!).

Attention: polynomials of higher order are not necessarily providing a more accurate result.

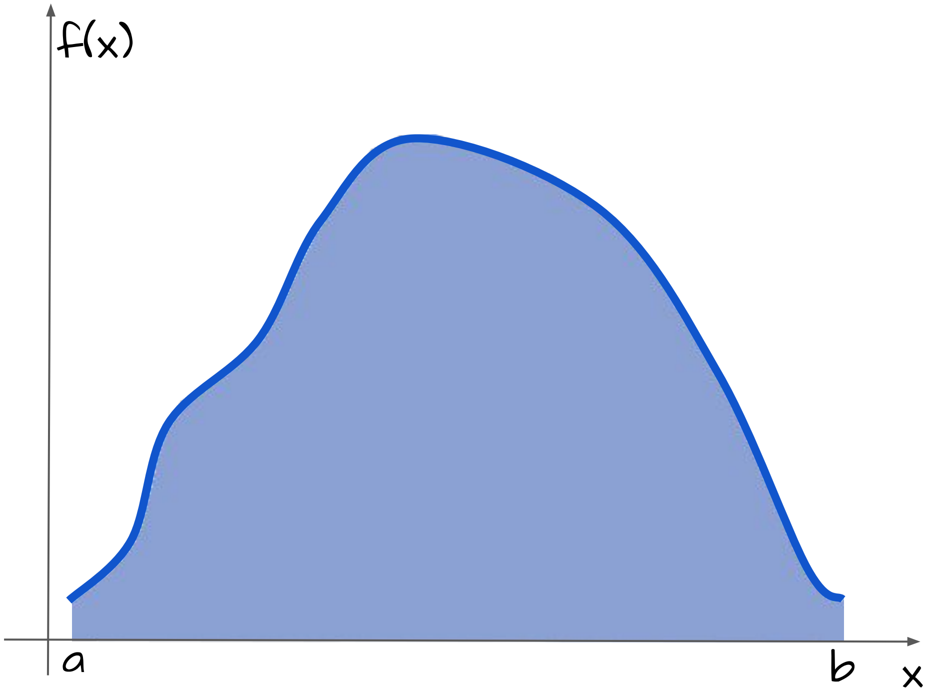


1. Newton-Cotes formulas

Idea: approximate the integrand function with a polynomial over a small range of *x*, and compute its integral (easy!).

Attention: polynomials of higher order are not necessarily providing a more accurate result.

To apply this method, we need to have a sequence of equally spaced abscissas $x_0, x_1, \dots, x_N$, at which we know (or can easily compute) the values assumed by the function: $f(x_0) = f_0, f_1, \dots, f_N$.

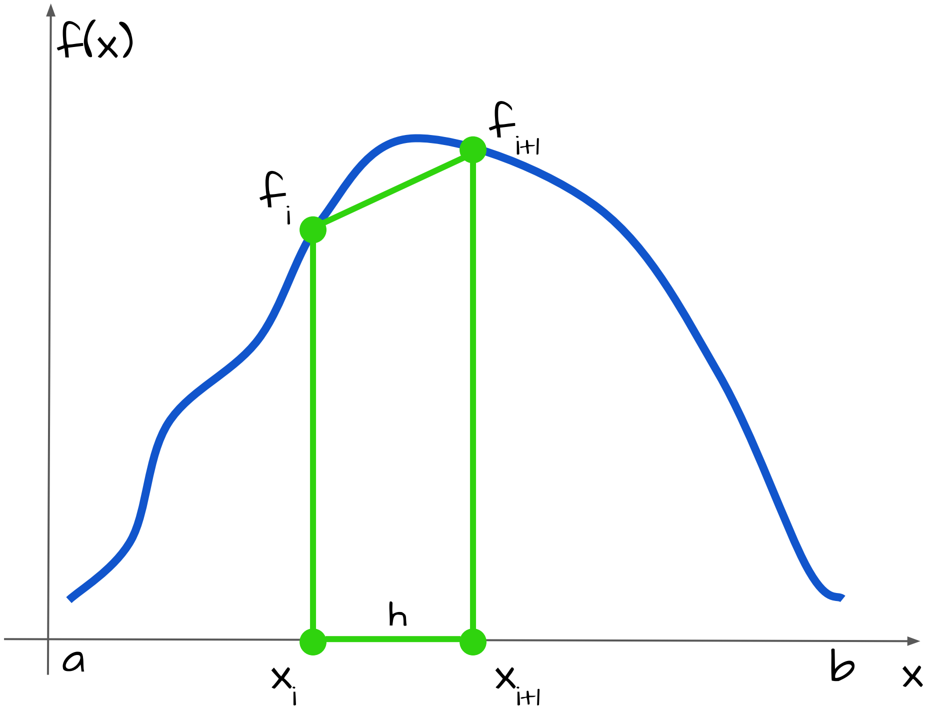


1. Newton-Cotes formulas

Trapezoid rule

The simplest application, using straight lines! (degree 1 polynomial) We consider 2 of the equally spaced abscissas at the time, and approximate the function in that interval with a straight line. The integral of the approximating function in that interval is easily calculated as the area of the resulting trapezoid:

$$\int_{x_i}^{x_{i+1}} f(x)dx = \frac{h(f_i + f_{i+1})}{2}$$



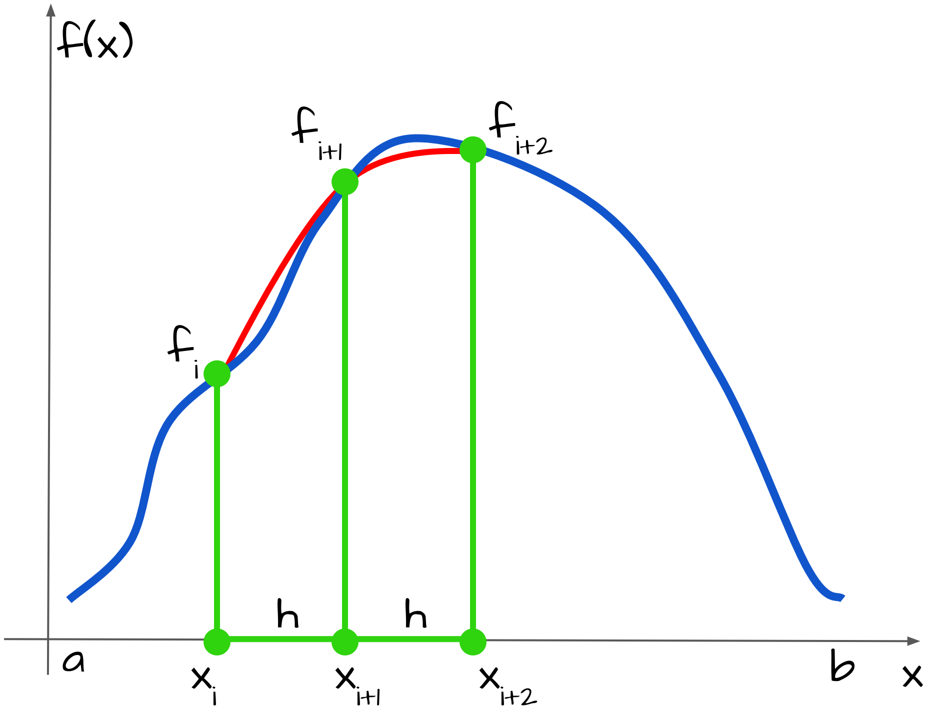
1. Newton-Cotes formulas

Simpson's rule

We can also use a degree 2 polynomial, consider 3 of the equally spaced abscissas at the time, and approximate the function in that interval with a parabola.

$$f(x) = A(x - x_0)^2 + B(x - x_0) + C$$

We determine the 3 coefficients by taking advantage of the fact that the parabola passes by the 3 points (x_i, f_i), (x_{i+1}, f_{i+1}), and (x_{i+2}, f_{i+2}).



1. Newton-Cotes formulas

Simpson's rule

We can also use a degree 2 polynomial, consider 3 of the equally spaced abscissas at the time, and approximate the function in that interval with a parabola.

$$f(x) = A(x - x_0)^2 + B(x - x_0) + C$$

We determine the 3 coefficients by taking advantage of the fact that the parabola passes by the 3 points *(x_i,f_i)*, *(x_{i+1},f_{i+1})*, and *(x_{i+2},f_{i+2})*.

$$A = \frac{f_i - 2f_{i+1} + f_{i+2}}{2h^2}$$

$$B = -\frac{3f_i - 4f_{i+1} + f_{i+2}}{2h}$$

$$C = f_i$$

Now we can easily compute the integral of the parabola:

$$\int_{x_i}^{x_{i+2}} f(x)dx = \frac{h}{3}\left(f_i + 4f_{i+1} + f_{i+2}\right)$$

1. Newton-Cotes formulas

By using these prescriptions for the computation of the integral in small intervals, and by adding them up, we can evaluate it over our interval of interest, *[a,b]*.

1. Newton-Cotes formulas

By using these prescriptions for the computation of the integral in small intervals, and by adding them up, we can evaluate it over our interval of interest, *[a,b]*.

Trapezoid rule

$$\int_{x_i}^{x_{i+1}} f(x)dx = \frac{h(f_i + f_{i+1})}{2} \longrightarrow \int_a^b f(x)dx = \frac{(b-a)}{N}\left[\frac{1}{2}f_0 + f_1 + f_2 + \dots + f_{N-1} + \frac{1}{2}f_N\right] + O\left(\frac{(b-a)^3}{N^3}f''\right)$$

1. Newton-Cotes formulas

By using these prescriptions for the computation of the integral in small intervals, and by adding them up, we can evaluate it over our interval of interest, *[a,b]*.

Trapezoid rule

$$\int_{x_i}^{x_{i+1}} f(x)dx = \frac{h(f_i + f_{i+1})}{2} \longrightarrow \int_a^b f(x)dx = \frac{(b-a)}{N}\left[\frac{1}{2}f_0 + f_1 + f_2 + \dots + f_{N-1} + \frac{1}{2}f_N\right] + O\left(\frac{(b-a)^3}{N^3}f''\right)$$

Simpson's rule

$$\int_{x_i}^{x_{i+2}} f(x)dx = \frac{h}{3}\left(f_i + 4f_{i+1} + f_{i+2}\right)$$

$$\longrightarrow \int_a^b f(x)dx = \frac{(b-a)}{N}\left[\frac{1}{3}f_0 + \frac{4}{3}f_1 + \frac{2}{3}f_2 + \frac{4}{3}f_3 + \dots + \frac{2}{3}f_{N-2} + \frac{4}{3}f_{N-1} + \frac{1}{3}f_N\right] + O\left(\frac{1}{N^4}\right)$$

1. Newton-Cotes formulas

Simpson's 3/8 rule

We can also move to a degree 3 polynomial, consider 4 of the equally spaced abscissas at the time, and approximate the function in that interval with a cubic:

$$f(x) = A(x - x_0)^3 + B(x - x_0)^2 + C(x - x_0) + D$$

1. Newton-Cotes formulas

Simpson's 3/8 rule

We can also move to a degree 3 polynomial, consider 4 of the equally spaced abscissas at the time, and approximate the function in that interval with a cubic:

$$f(x) = A(x-x_0)^3 + B(x-x_0)^2 + C(x-x_0) + D$$

With a procedure similar to the one we used before, we can compute the coefficients and we can easily calculate the integral of the cubic over the small interval:

$$\int_{x_i}^{x_{i+3}} f(x)dx = \frac{3h}{8}\left(f_i + 3f_{i+1} + 3f_{i+2} + f_{i+3}\right)$$

1. Newton-Cotes formulas

Simpson's 3/8 rule

We can also move to a degree 3 polynomial, consider 4 of the equally spaced abscissas at the time, and approximate the function in that interval with a cubic:

$$f(x) = A(x-x_0)^3 + B(x-x_0)^2 + C(x-x_0) + D$$

With a procedure similar to the one we used before, we can compute the coefficients and we can easily calculate the integral of the cubic over the small interval:

$$\int_{x_i}^{x_{i+3}} f(x)dx = \frac{3h}{8}\left(f_i + 3f_{i+1} + 3f_{i+2} + f_{i+3}\right)$$

By adding these up, we can evaluate the integral over our interval of interest, *[a,b]*.

$$\int_a^b f(x)dx = \frac{(b-a)}{N}\left[\frac{3}{8}f_0 + \frac{9}{8}f_1 + \frac{9}{8}f_2 + \frac{6}{8}f_3 + \frac{9}{8}f_4 + \ldots \right.$$
$$\left. + \frac{6}{8}f_{N-3} + \frac{9}{8}f_{N-2} + \frac{9}{8}f_{N-1} + \frac{3}{8}f_N\right] + O\left(\frac{(b-a)^4}{N^4}f^{(4)}\right)$$

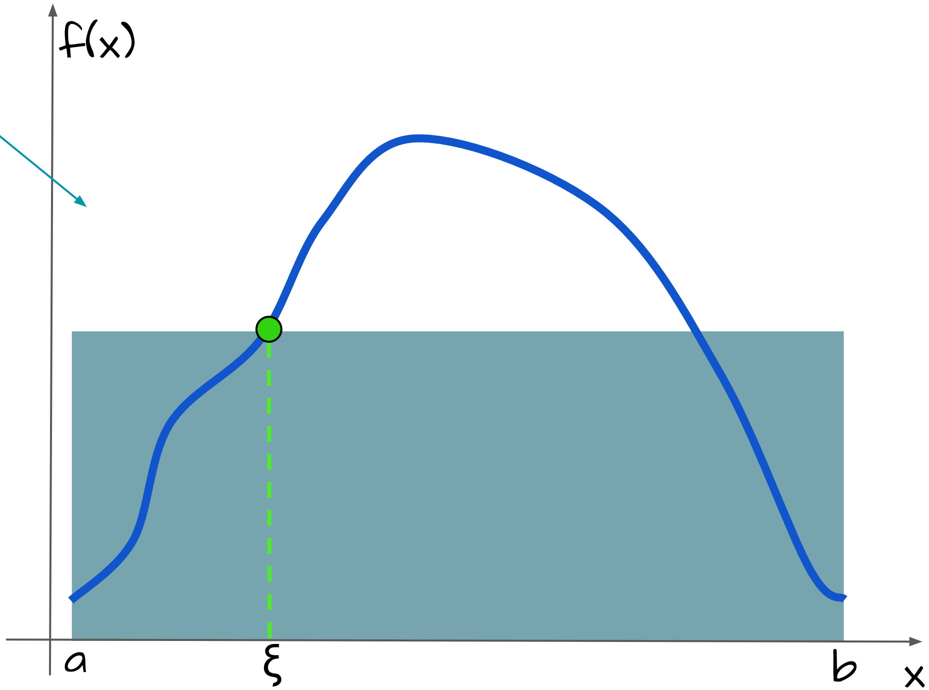
2. Gaussian quadrature

The mean value theorem states that:

$$\int_a^b f(x)dx = (b-a)f(\xi)$$

Therefore, by using this, we could approximate our integral as:

$$\int_a^b f(x)dx = c_0 f(a) + c_1 f(b)$$



2. Gaussian quadrature

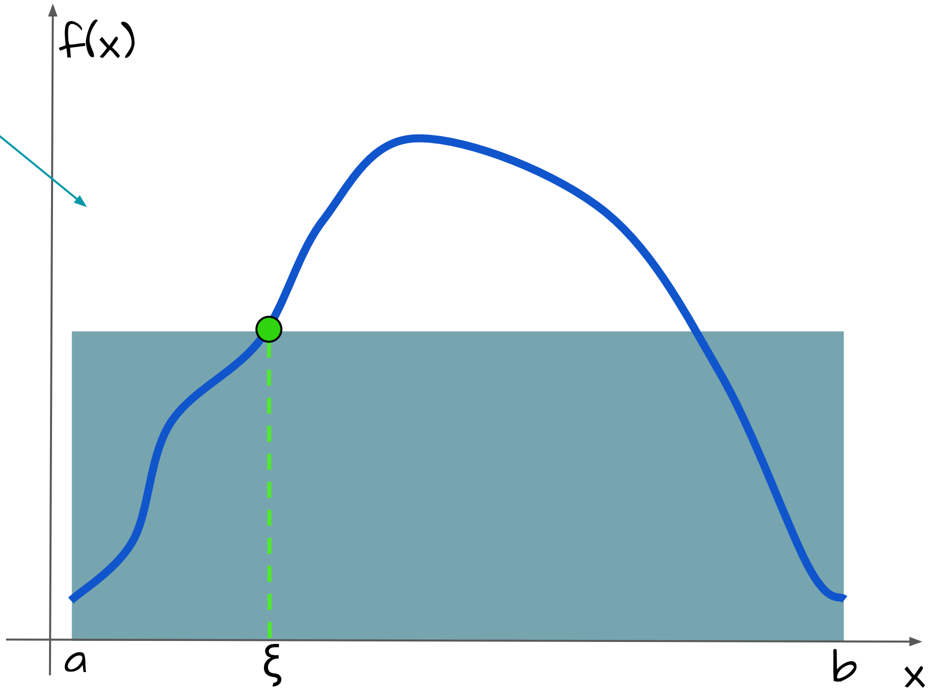
The mean value theorem states that:

$$\int_a^b f(x)dx = (b-a)f(\xi)$$

Therefore, by using this, we could approximate our integral as:

$$\int_a^b f(x)dx = c_0 f(a) + c_1 f(b)$$

The idea of Gaussian quadratures is to choose not only the weight coefficients, but also the location of the abscissas at which the function is to be evaluated:



$$\int_a^b W(x)f(x)dx \sim \sum_{i=1}^{N} w_i f(x_i)$$

careful: high order is not the same as high accuracy!

2. Gaussian quadrature

How does it work?

We can find a set of polynomials that includes exactly one polynomial of order j, $p_j(x)$, for $j = 0, 1, \ldots$, all of which are mutually orthogonal over the weight function $W(x)$.

Such a set can be constructed by means of a recurrence relation:

$$\begin{aligned} p_{-1}(x) &\equiv 0 \\ p_0 &\equiv 1 \\ p_{j+1} &= (x - c_j)p_j(x) - d_j p_{j-1}(x) \\ & \qquad j = 0, 1, 2, \ldots \end{aligned}$$

2. Gaussian quadrature

How does it work?

We can find a set of polynomials that includes exactly one polynomial of order *j*, $p_j(x)$, for $j = 0, 1, \ldots$, all of which are mutually orthogonal over the weight function $W(x)$.
Such a set can be constructed by means of a recurrence relation:

$$p_{-1}(x) \equiv 0$$
$$p_0 \equiv 1$$
$$p_{j+1} = (x - c_j)p_j(x) - d_j p_{j-1}(x)$$
$$j = 0, 1, 2, \ldots$$

where

$$c_j = \frac{\langle x\, p_j | p_j \rangle}{\langle p_j | p_j \rangle} \qquad j = 0, 1, 2, \ldots$$
$$d_j = \frac{\langle p_j | p_j \rangle}{\langle p_{j-1} | p_{j-1} \rangle} \qquad j = 1, 2, 3, \ldots$$

The polynomial $p_j(x)$ can be shown to have exactly *j* distinct roots in the interval (a,b), and it can be shown that there is exactly one root of the former polynomial in between each two adjacent roots of the latter one.

2. Gaussian quadrature

How does it work?

The abscissas of the *N*-point Gaussian quadrature formula

$$\int_a^b W(x)f(x)dx \sim \sum_{i=1}^{N} w_i f(x_i)$$

are precisely the roots of the orthogonal polynomial $p_N(x)$ for the same interval and weighting function. Once the abscissas $x_1, \dots, x_N$ are known, we find the weights w_i such that the equation above gives the correct answer for the integral of the first *N* orthogonal polynomials.

2. Gaussian quadrature

How does it work?

The abscissas of the *N*-point Gaussian quadrature formula

$$\int_a^b W(x)f(x)dx \sim \sum_{i=1}^{N} w_i f(x_i)$$

are precisely the roots of the orthogonal polynomial $p_N(x)$ for the same interval and weighting function. Once the abscissas $x_1, \dots, x_N$ are known, we find the weights w_i such that the equation above gives the correct answer for the integral of the first *N* orthogonal polynomials.

Weights can be found by solving:

$$\begin{bmatrix} p_0(x_1) & \dots & p_0(x_N) \\ p_1(x_1) & \dots & p_1(x_N) \\ \vdots & & \vdots \\ p_{N-1}(x_1) & \dots & p_{N-1}(x_N) \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix} = \begin{bmatrix} \int_a^b W(x)p_0(x)dx \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

or alternatively by using the formula

$$w_i = \frac{\langle p_{N-1} | p_{N-1} \rangle}{p_{N-1}(x_i)p'_N(x_i)}$$

2. Gaussian quadrature

To summarize:

The computation of the Gaussian quadrature rule involves two phases:

1. generation of the polynomials and computation of the coefficients c_j and d_j
2. determination of zeros of $p_n(x)$ and computation of the associated weights

2. Gaussian quadrature

To summarize:

The computation of the Gaussian quadrature rule involves two phases:

1. generation of the polynomials and computation of the coefficients c_j and d_j
2. determination of zeros of $p_n(x)$ and computation of the associated weights

For the case of the “classical” orthogonal polynomials, the coefficients c_j and d_j are explicitly known (see next slide for an example).

If you are confronted with a “nonclassical” weight function $W(x)$, and you don’t know the coefficients c_j and d_j, the construction of the associated set of orthogonal polynomials is not trivial.

2. Gaussian quadrature

Legendre-Gauss quadrature

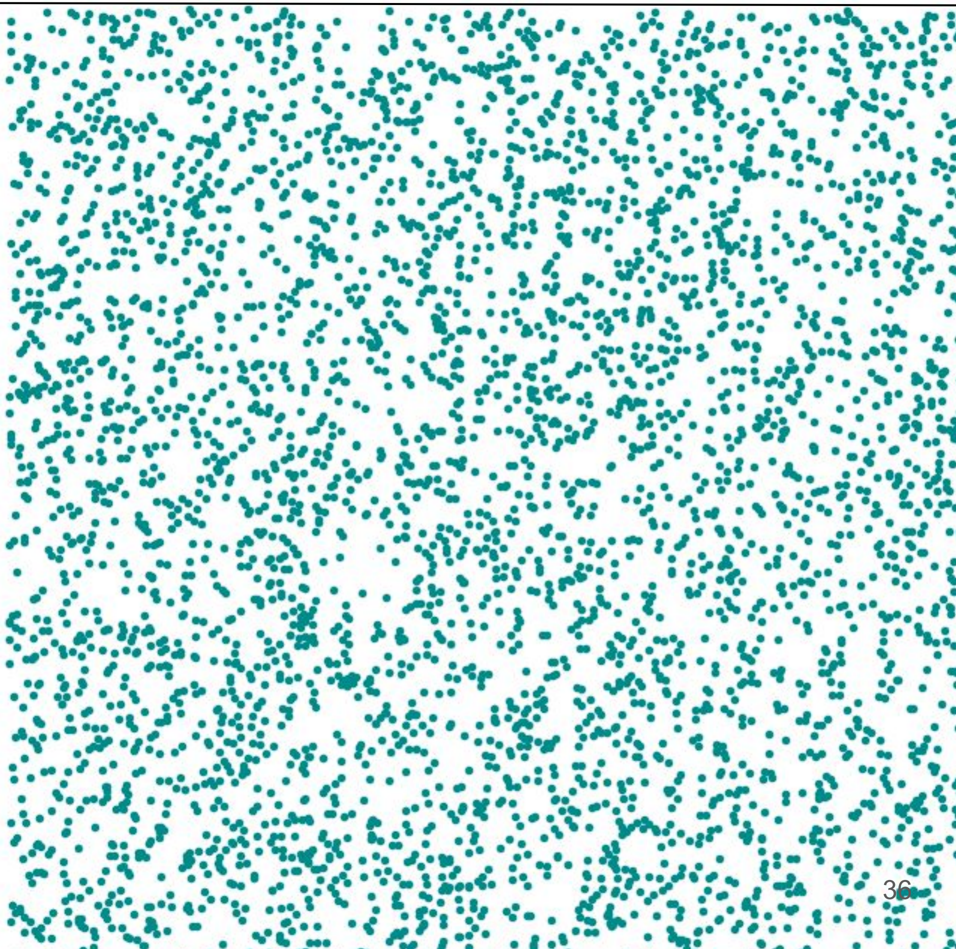
- Weight function → $W(x) = 1$
- Interval → $-1 < x < 1$
- Recurrence relation → $(j+1)P_{j+1} = (2j+1)xP_j - jP_{j-1}$
- Weights → $w_i = \dfrac{2}{(1-x_i^2)[P_N'(x_i)]^2}$

3. Monte Carlo integration

Monte Carlo integration can be carried out with different techniques. Here we will explore the following:

1. **mean values**
2. **importance sampling**
3. **control variates**
4. **antithetic variates**.

These techniques are particularly indicated to carry out multidimensional integrals.
However, they are not always the best choice, and can be extremely slow...



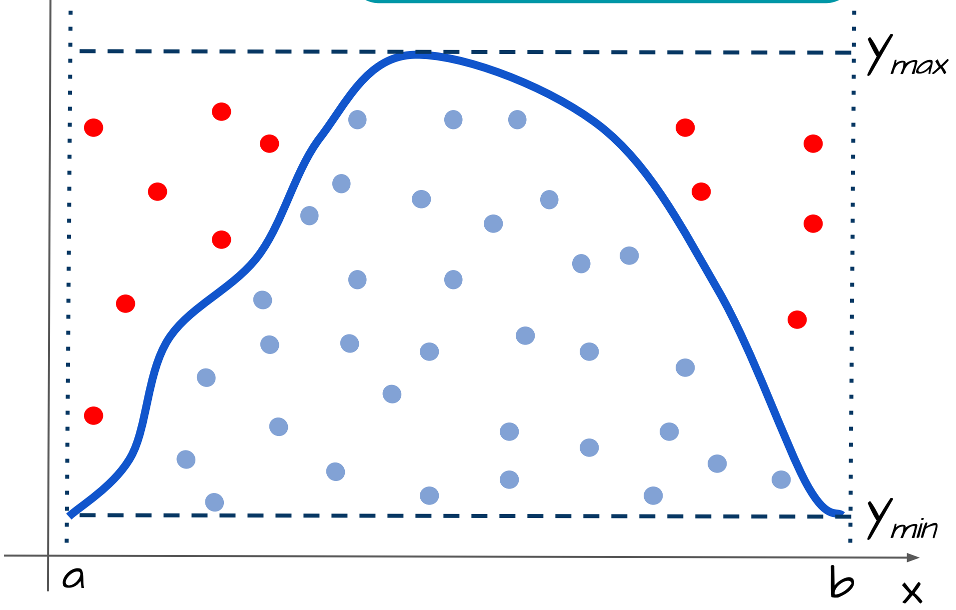
3. Monte Carlo integration

To compute integrals using Monte Carlo methods:

- generate *N* random points (x_i, y_i) uniformly distributed in the rectangle (of known area) defined by the integration interval *[a,b]* and the range of values $[y_{min}, y_{max}]$
- compare the values y_i and $f(x_i)$, and compute the integral as:

$$I = (b-a)(y_{\max} - y_{\min})\frac{N_{[y_i < f(x_i)]}}{N}$$

A challenge: can you find the value of π by using this type of integration?



Attention: this method only works on finite intervals!

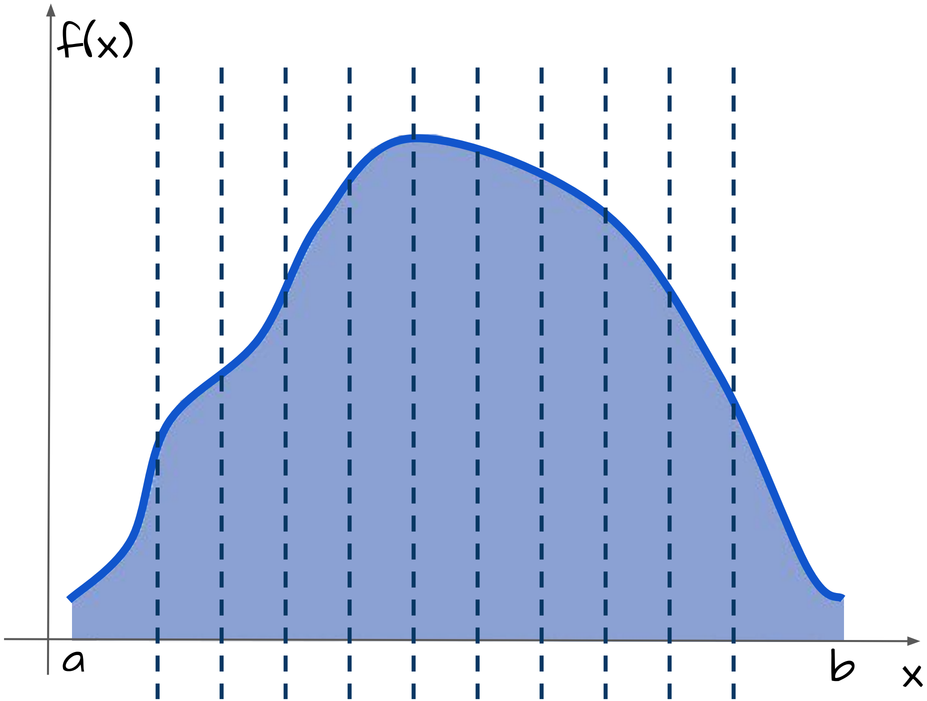
3.1 Monte Carlo integration - Mean Value

There is also a better way to do this. Let's consider the integral

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

and a partition of the interval *[a,b]*; the integral can be evaluated as follows:

$$\begin{aligned} I &= \lim_{N\to\infty} \sum_{i=0}^{N-1} h f(x_i) \\ &= \lim_{N\to\infty} \sum_{i=0}^{N-1} \frac{(b-a)}{N} f(x_i) \\ &= (b-a) \lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} f(x_i) \\ &= (b-a)\langle f \rangle_{[a,b]} \end{aligned}$$



3.1 Monte Carlo integration - Mean Value

There is also a better way to do this.
Let's consider the integral

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

and a partition of the interval *[a,b]*; the integral can be evaluated as follows:

$$\begin{aligned} I &= \lim_{N\to\infty} \sum_{i=0}^{N-1} h f(x_i) \\ &= \lim_{N\to\infty} \sum_{i=0}^{N-1} \frac{(b-a)}{N} f(x_i) \\ &= (b-a) \lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} f(x_i) \\ &= (b-a)\langle f \rangle_{[a,b]} \end{aligned}$$

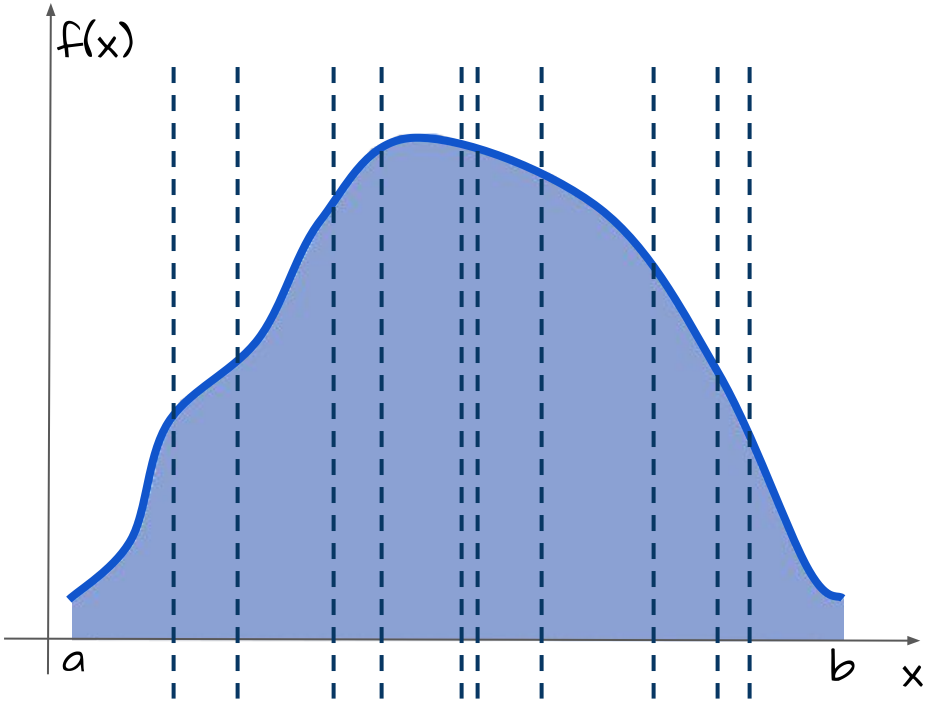
An estimation of $\langle f \rangle$ can be obtained by considering random variables x_i uniformly distributed in the interval *[a,b]*:

$$\begin{aligned} I &= (b-a)\langle f \rangle_{[a,b]} \\ &\simeq (b-a)\frac{1}{N}\sum_{i=0}^{N-1} f(x_i) \end{aligned}$$

3.1 Monte Carlo integration - Mean Value

The central idea of Monte Carlo quadrature is that an integral may be estimated by a sum:

$$I = \int_{-\infty}^{\infty} f(x)p(x)dx = \langle f(x) \rangle$$

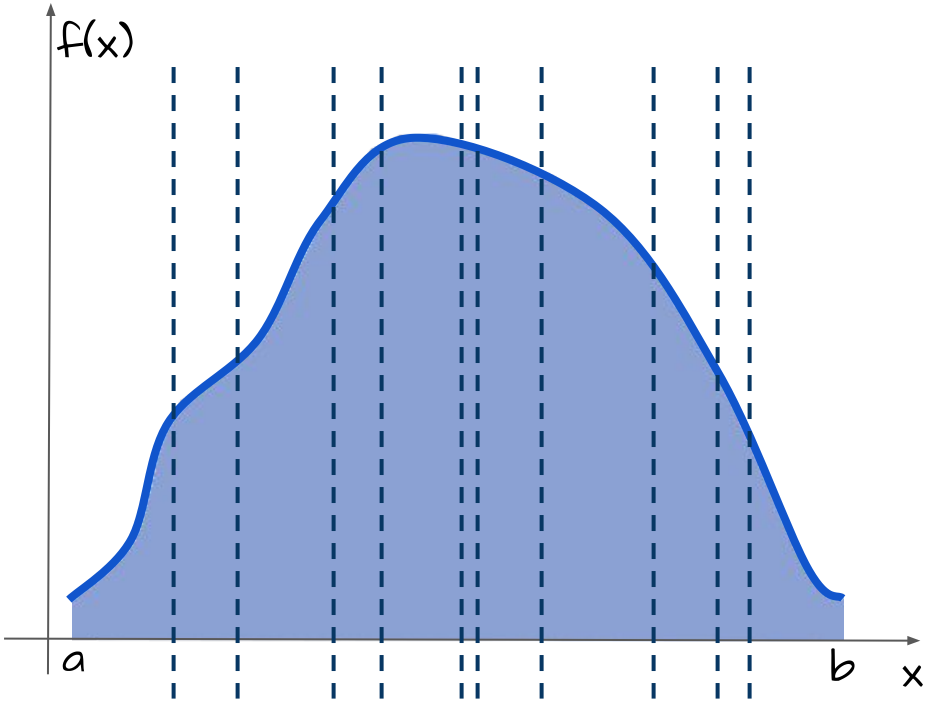


3.1 Monte Carlo integration - Mean Value

The central idea of Monte Carlo quadrature is that an integral may be estimated by a sum:

$$I = \int_{-\infty}^{\infty} f(x)p(x)dx = \langle f(x) \rangle$$

Draw a series of random variables, x_i, from $p(x)$, evaluate $f(x)$ for each x_i. The mean of all the $f(x_i)$ is an estimate of the integral, and its variance decreases as the number of terms increases:



$$\mathrm{var}[I] \simeq \frac{1}{N}\mathrm{var}[f(x)] = \frac{1}{N}\left[\langle f^2(x) \rangle - \langle f(x) \rangle^2\right]$$

3.1 Monte Carlo integration - Mean Value

So, if the integral we want to evaluate is in the form

$$I = \int_{\Omega} f(x)p(x)dx$$

we can estimate its value through the following function, which is an estimator of *I*:

$$F_N = \frac{1}{N} \sum_{i=0}^{N-1} f(x_i)$$

If the integral exists, it is equal to the mean of this function. We can therefore write:

$$F_N = I + \varepsilon$$

3.1 Monte Carlo integration - Mean Value

So, if the integral we want to evaluate is in the form

$$I = \int_{\Omega} f(x)p(x)dx$$

we can estimate its value through the following function, which is an estimator of *I*:

$$F_N = \frac{1}{N}\sum_{i=0}^{N-1} f(x_i)$$

If the integral exists, it is equal to the mean of this function. We can therefore write:

$$F_N = I + \varepsilon$$

The error *ε* is characterized by

$$\varepsilon = \frac{\sigma_I}{\sqrt{N}}$$

where

$$\sigma_I^2 = \int_{\Omega} f^2(x)p(x)dx - I^2$$

This may be inverted to show the number of samples needed to yield a desired error: $N = \sigma^2/\varepsilon^2$.

The integral does not need to exhibit explicitly a function *p(x)* satisfying the properties of a probability distribution. One can simply use $p(x)=1/\Omega$ and $f(x)= \Omega\cdot integrand$.

3.2 Monte Carlo integration - Importance Sampling

So far, we have considered a uniform distribution to sample the abscissas to compute our integral

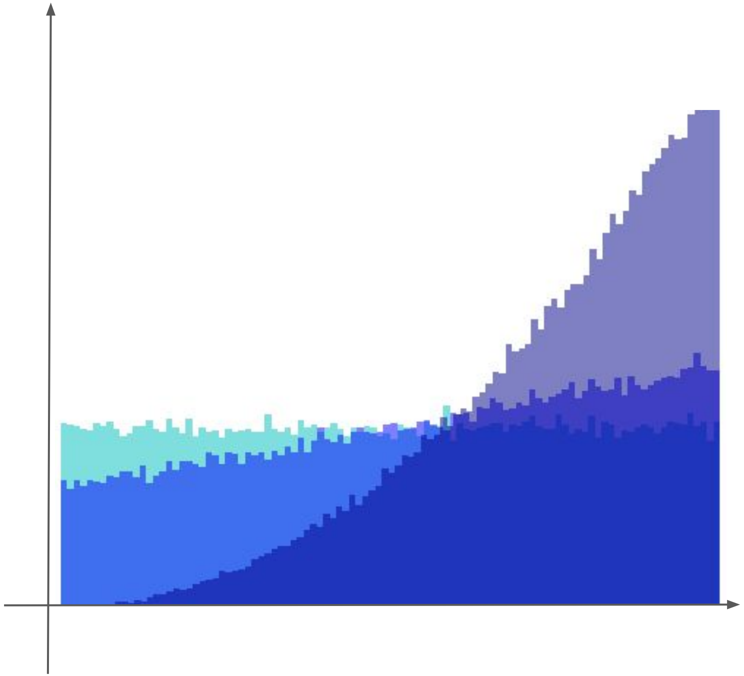
$$I = \int f(x)p(x)dx \ ,$$

but this is not always the best choice! A different probability distribution *w(x)*, satisfying

$$w(x) \geq 0$$
$$\int w(x)dx = 1$$

can also be introduced in the integral:

$$I = \int \frac{f(x)p(x)}{w(x)} w(x)dx$$



3.2 Monte Carlo integration - Importance Sampling

The variance of *I* when using *w(x)* is: $$\mathrm{var}[I]_w = \int \left[\frac{f(x)p(x)}{w(x)}\right]^2 w(x)dx - I^2$$

We want a function *w(x)* that will allow us to minimize the first term of this equation.

3.2 Monte Carlo integration - Importance Sampling

The variance of *I* when using *w(x)* is: $$\mathrm{var}[I]_w = \int \left[\frac{f(x)p(x)}{w(x)}\right]^2 w(x)dx - I^2$$

We want a function *w(x)* that will allow us to minimize the first term of this equation. The function *w(x)* satisfying this can be found by using a Lagrange multiplier *λ*:

$$L(w) = \left\{ \int \left[\frac{f(x)p(x)}{w(x)}\right]^2 w(x)dx + \lambda \int w(x)dx \right\}$$

3.2 Monte Carlo integration - Importance Sampling

The variance of *I* when using *w(x)* is: $$\mathrm{var}[I]_w = \int \left[\frac{f(x)p(x)}{w(x)}\right]^2 w(x)dx - I^2$$

We want a function *w(x)* that will allow us to minimize the first term of this equation. The function *w(x)* satisfying this can be found by using a Lagrange multiplier *λ*:

$$L(w) = \left\{ \int \left[\frac{f(x)p(x)}{w(x)}\right]^2 w(x)dx + \lambda \int w(x)dx \right\}$$

L(w) has to be minimized. So we compute:

$$\frac{d}{dw}\{\ldots\} = 0 \longrightarrow -\left[\frac{f(x)p(x)}{w(x)}\right] + \lambda = 0 \longrightarrow w(x) = \lambda |f(x)p(x)|$$

3.2 Monte Carlo integration - Importance Sampling

The variance of *I* when using *w(x)* is: $$\mathrm{var}[I]_w = \int \left[\frac{f(x)p(x)}{w(x)}\right]^2 w(x)dx - I^2$$

We want a function *w(x)* that will allow us to minimize the first term of this equation. The function *w(x)* satisfying this can be found by using a Lagrange multiplier *λ*:

$$L(w) = \left\{ \int \left[\frac{f(x)p(x)}{w(x)}\right]^2 w(x)dx + \lambda \int w(x)dx \right\}$$

L(w) has to be minimized. So we compute:

$$\frac{d}{dw}\{\ldots\} = 0 \longrightarrow -\left[\frac{f(x)p(x)}{w(x)}\right] + \lambda = 0 \longrightarrow w(x) = \lambda |f(x)p(x)|$$

A Monte Carlo algorithm to evaluate the integral would be to sample a series of x_i from *w(x)* and compute:

$$I = \frac{1}{N}\sum_{i=0}^{N-1} \frac{f(x_i)p(x_i)}{w(x_i)}$$

We expect that "similar" functions will reduce the variance

3.3 Monte Carlo integration - Control Variates

In this alternative technique, our integral

$$I = \int f(x)p(x)dx$$

is rewritten as

$$I = \int \left[f(x) - h(x)\right] p(x)dx + \int h(x)p(x)dx$$

where the integral of *h(x)* is known analytically.

Therefore, to estimate the value of the integral, we calculate:

$$I = \int h(x)p(x)dx + \frac{1}{N}\sum_{i=0}^{N-1}\left[f(x_i) - h(x_i)\right]$$

The technique is advantageous when $\mathrm{var}[f - h] \ll \mathrm{var}[f]$ and this occurs when *h(x)* is very similar to *f(x)*.

This technique is useful when *I* resembles a known integral.

3.4 Monte Carlo integration - Antithetic Variates

This method exploits the decrease in variance that occurs when random variables are negatively correlated. In this case, if the first point gives a value of the integrand that is larger than average, the next point will be likely to give a value that is smaller than average, and the average of the two values will be closer to the actual mean. For example, if we have an integral

$$I = \int_0^1 f(x)dx$$

with *f(x)* linear, we can rewrite it as

$$I = \int_0^1 \frac{1}{2}\left[f(x) + f(1-x)\right]dx$$

and we can evaluate it with Monte Carlo techniques as

$$I = \sum_{i=0}^{N-1} \frac{1}{2}\left[f(x_i) + f(1-x_i)\right]$$

Examples

EXAMPLE 1 - Gaussian 3-point quadrature

$$\int_a^b f(x)dx = \int_a^b f(x)W(x)dx \sim \sum_{i=1}^{3} w_i f(x_i)$$

Legendre polynomials:

- $W(x) = 1$
- we need the interval of integration to be: $[-1, 1]$

$$\begin{aligned} x &= \frac{(b+a)}{2} + \frac{(b-a)}{2} z \\ z &= \frac{x - \frac{(b+a)}{2}}{\frac{(b-a)}{2}} \\ dz &= \frac{2}{(b-a)} dx \end{aligned}$$

- polynomial of order 3:

$$P_3(x) = \frac{5}{2}x^3 - \frac{3}{2}x$$

$$P_3(x) = 0 \longrightarrow x(5x^2 - 3) = 0$$

$$\begin{aligned} x_2 &= 0 \\ x_{1,3} &= \pm\sqrt{3/5} \end{aligned}$$

EXAMPLE 1 - Gaussian 3-point quadrature

$$\int_a^b f(x)dx = \int_{-1}^1 f(z)dz \frac{(b-a)}{2}$$

$$\int_{-1}^1 f(z)dz \sim w_1 \, f\left(-\sqrt{\frac{3}{5}}\right) + w_2 \, f(0) + w_3 \, f\left(\sqrt{\frac{3}{5}}\right)$$

Now we need to find the **weights**, using the fact that the proper choice of weights makes the integration exact for the first 3 orthogonal polynomials...

EXAMPLE 1 - Gaussian 3-point quadrature

- $P_0(x) = 1 \qquad \int_{-1}^{1} 1\, dx = 2$

$$\begin{aligned}\int_{-1}^{1} P_0(x)dx &= w_1\, f\left(-\sqrt{\frac{3}{5}}\right) + w_2\, f(0) + w_3\, f\left(\sqrt{\frac{3}{5}}\right) = \\ &= w_1 + w_2 + w_3 = 2\end{aligned}$$

- $P_1(x) = x \qquad \int_{-1}^{1} x\, dx = 0$

$$\begin{aligned}\int_{-1}^{1} P_1(x)dx &= w_1\, f\left(-\sqrt{\frac{3}{5}}\right) + w_2\, f(0) + w_3\, f\left(\sqrt{\frac{3}{5}}\right) = \\ &= -w_1\sqrt{\frac{3}{5}} + w_3\sqrt{\frac{3}{5}} = 0 \qquad \longrightarrow \qquad w_1 = w_3\end{aligned}$$

EXAMPLE 1 - Gaussian 3-point quadrature

- $P_2(x) = \frac{1}{2}(3x^2 - 1)$ $\int_{-1}^{1} \frac{1}{2}(3x^2 - 1)\, dx = 0$

$$\begin{aligned}\int_{-1}^{1} P_2(x)dx &= w_1\, f\left(-\sqrt{\frac{3}{5}}\right) + w_2\, f(0) + w_3\, f\left(\sqrt{\frac{3}{5}}\right) = \\ &= \frac{2}{5}w_1 - \frac{1}{2}w_2 + \frac{2}{5}w_3 = 0\end{aligned}$$

$$\begin{cases} w_1 = w_3 \\ w_2 = 2 - 2w_1 \\ \frac{2}{5}w_1 - 1 + w_1 + \frac{2}{5}w_1 = 0 \end{cases} \longrightarrow \boxed{\begin{aligned} w_1 &= \frac{5}{9} \\ w_3 &= \frac{5}{9} \\ w_2 &= 2 - \frac{2 \cdot 5}{9} = \frac{8}{9} \end{aligned}}$$

EXAMPLE 1 - Gaussian 3-point quadrature

$$\boxed{\int_{-1}^{1} f(z)dz \sim \frac{5}{9} f\left(-\sqrt{\frac{3}{5}}\right) + \frac{8}{9} f(0) + \frac{5}{9} f\left(\sqrt{\frac{3}{5}}\right)}$$

$$\begin{cases} w_1 = w_3 \\ w_2 = 2 - 2w_1 \\ \frac{2}{5}w_1 - 1 + w_1 + \frac{2}{5}w_1 = 0 \end{cases}$$

$$\boxed{\begin{aligned} w_1 &= \frac{5}{9} \\ w_3 &= \frac{5}{9} \\ w_2 &= 2 - \frac{2 \cdot 5}{9} = \frac{8}{9} \end{aligned}}$$

EXAMPLE 1 - Gaussian 3-point quadrature

Let's see this applied to a specific case:

$$I = \int_0^1 e^x \, dx = e - 1 \sim 1.7183$$

$$\left(x = \frac{1}{2} + \frac{z}{2} \quad \longrightarrow \quad \begin{aligned} z &= 2x - 1 \\ dz &= 2dx \end{aligned} \right)$$

$$\int_{-1}^{1} \frac{e^{\frac{1}{2}+\frac{z}{2}}}{2} dz = \int_{-1}^{1} f(z) dz \sim \frac{5}{9} f\left(-\sqrt{\frac{3}{5}}\right) + \frac{8}{9} f(0) + \frac{5}{9} f\left(\sqrt{\frac{3}{5}}\right)$$

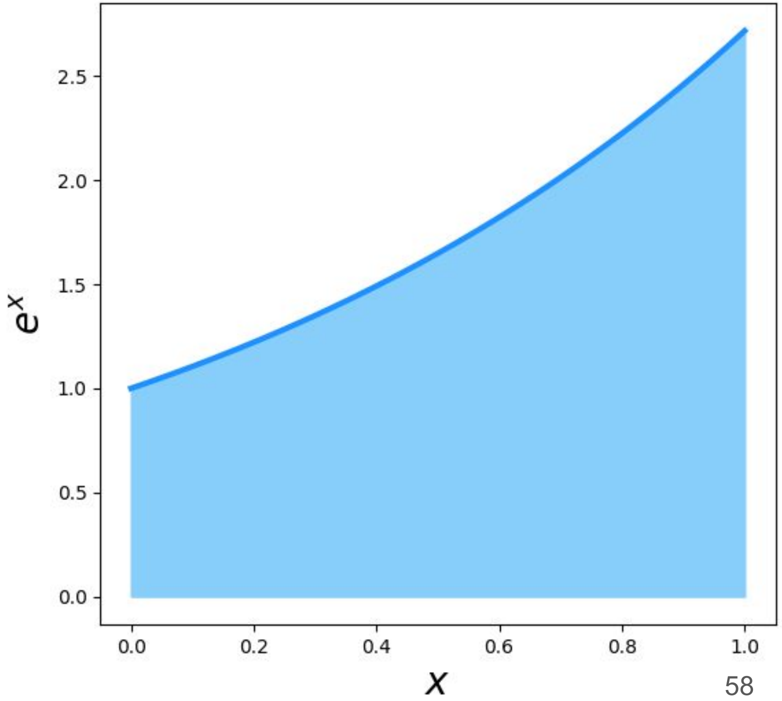
$$= \frac{e^{1/2}}{2} \left(\frac{5}{9} e^{-\frac{\sqrt{3/5}}{2}} + \frac{8}{9} \cdot 1 + \frac{5}{9} e^{-\frac{\sqrt{3/5}}{2}} \right) =$$

$$\sim \quad 1.7183$$

EXAMPLE 2 - Monte Carlo integration

Let’s compute the Monte Carlo Integration of the same function with all the four different methods presented in class.

$$I = \int_0^1 e^x \, dx = e - 1 \sim 1.7183$$



We will think about how to sample the points, how to compute the integral, and we will especially focus on the variance.

RECAP - Monte Carlo integration (Mean Value)

If the integral we want to evaluate is in the form

$$I = \int_{\Omega} f(x)p(x)dx$$

we can estimate its value through the following function, which is an estimator of *I*:

$$F_N = \frac{1}{N}\sum_{i=0}^{N-1} f(x_i)$$

If the integral exists, it is equal to the mean of the function *f*. We can therefore write:

$$F_N = I + \varepsilon$$

The error *ε* is characterized by

$$\varepsilon = \frac{\sigma_I}{\sqrt{N}}$$

where

$$\sigma_I^2 = \int_{\Omega} f^2(x)p(x)dx - I^2$$

This may be inverted to show the number of samples needed to yield a desired error: $N = \sigma^2/\varepsilon^2$.

The integral does not need to exhibit explicitly a function *p(x)* satisfying the properties of a probability distribution. One can simply use $p(x)=1/\Omega$ and $f(x)= \Omega \cdot \mathit{integrand}$.

EXAMPLE 2 - Monte Carlo integration

1. Uniform sampling - Mean values

$$I = \int_0^1 f(x)\,p(x)\,dx = \int_0^1 e^x dx \quad \longleftarrow \quad \left\{ \begin{array}{lcl} f(x) & = & e^x \\ p(x) & = & 1 \end{array} \right.$$

$\{x_i\}$ sampled uniformly

$$\boxed{I \sim \frac{1}{N}\sum_{i=1}^{N} e^{x_i}}$$

$$\boxed{\sigma_I^2} = \int_0^1 f^2(x)\,p(x)\,dx - I^2 =$$

$$= \int_0^1 e^{2x} dx - (e-1)^2 = \frac{e^2-1}{2} - (e-1)^2 = \frac{-e^2+4e-3}{2} \boxed{\sim 0.242}$$

EXAMPLE 2 - Monte Carlo integration

2. Importance sampling

Here we consider a probability distribution that is more similar to the function we want to integrate.

$$e^x \Longrightarrow 1 + x + \dots$$

$$w(x) \propto 1 + x \rightarrow w(x) = A(1+x)$$

$$\begin{cases} w(x) \geq 0 \text{ in } [0,1] \\ \displaystyle\int_0^1 w(x)\,dx = 1 \; = \int_0^1 A(1+x)\,dx = A\left(x + \frac{x^2}{2}\right)\Bigg|_0^1 = \frac{3}{2}A = 1 \end{cases}$$

$$\longrightarrow \quad A = \frac{2}{3} \quad \longrightarrow \quad w(x) = \frac{2}{3}(1+x)$$

EXAMPLE 2 - Monte Carlo integration

2. Importance sampling

$$I = \int_0^1 \frac{f(x)\,p(x)}{w(x)} \cdot w(x) dx = \int_0^1 e^x dx \leftarrow \begin{cases} f(x) &= e^x \\ p(x) &= 1 \\ w(x) &= \frac{2}{3}(1+x) \end{cases}$$

$$I = \int_0^1 \underbrace{\frac{3e^x}{2(1+x)}}_{\text{function to integrate}} \cdot \underbrace{\frac{2}{3}(1+x)}_{\text{probability distribution}} dx$$

$\{x_i\}$ sampled from w(x):

$$c(x) = \int_0^x w(x')\,dx' = \frac{2}{3}\left(x + \frac{x^2}{2}\right)$$

$$C_i = \frac{2}{3}\left(x_i + \frac{x_i^2}{2}\right)$$

$$\rightarrow x_i = -1 + \sqrt{1 + 3C_i}$$

EXAMPLE 2 - Monte Carlo integration

2. Importance sampling

$$\boxed{I \sim \frac{1}{N}\sum_{i=1}^{N} \frac{e^{x_i}}{(1+x_i)}\frac{3}{2}}$$

$$\begin{aligned}
\boxed{\sigma_I^2} &= \int_0^1 \left[\frac{3e^x}{2(1+x)}\right]^2 \cdot \frac{2}{3}(1+x)\,dx - I^2 = \\
&= \int_0^1 \frac{9\,e^{2x}}{4(1+x)^2}\,\frac{2}{3}\,(1+x)\,dx - (e-1)^2 = \\
&= \int_0^1 \frac{3}{2}\,\frac{e^{2x}}{(1+x)}\,dx - (e-1)^2 \boxed{\sim 0.0269}
\end{aligned}$$

EXAMPLE 2 - Monte Carlo integration

3. Control Variates

$$I = \int_0^1 [f(x) - h(x)]\, p(x)\, dx + \int_0^1 h(x)\, p(x)\, dx = \int_0^1 e^x dx$$

We want h(x) to be similar to f(x), so we choose it to be a Taylor expansion of f(x):

$$\begin{aligned} f(x) &= e^x \\ h(x) &= 1 + x \\ p(x) &= 1 \end{aligned}$$

$$\longrightarrow \quad I = \int_0^1 [e^x - 1 - x]\, dx + \frac{3}{2}$$

EXAMPLE 2 - Monte Carlo integration

3. Control Variates

$$\boxed{I \sim \frac{1}{N}\sum_{i=1}^{N}[e^{x_i} - 1 - x_i] + \frac{3}{2}}$$

$$\begin{aligned}
\boxed{\sigma_I^2} &= \int_0^1 [f(x) - h(x)]^2\, p(x)\, dx - \left[\int_0^1 [f(x) - h(x)]\, p(x)\, dx\right]^2 = \\
&= \int_0^1 [e^x - (1+x)]^2\, dx - \left(I - \frac{3}{2}\right)^2 = \\
&= \int_0^1 [e^{2x} + 1 + 2x + x^2 - 2e^x - 2xe^x]\, dx - \left(e - 1 - \frac{3}{2}\right)^2 = \\
&= \left[\frac{e^{2x}}{2} + x + x^2 + \frac{x^3}{3} - 2e^x - 2(xe^x - e^x)\right]_0^1 - \left(e - \frac{5}{2}\right)^2 = \\
&= -\frac{e^2}{2} + 3e - \frac{27}{4} + \frac{7}{3}\ \boxed{\sim 0.0437}
\end{aligned}$$

EXAMPLE 2 - Monte Carlo integration

4. Antithetic Variates

$$I = \int_0^1 \frac{1}{2}\left[f(x) + f(1-x)\right] dx = \int_0^1 \frac{1}{2}\left(e^x + e^{1-x}\right) dx$$

$$\boxed{I \sim \frac{1}{N}\sum_{i=1}^{N}\left(\frac{e^{x_i} + e^{1-x_i}}{2}\right)}$$

$$\boxed{\sigma_I^2} = \int_0^1 \left(\frac{e^x + e^{1-x}}{2}\right)^2 dx - I^2 =$$

$$= \int_0^1 \frac{1}{4}\left(e^{2x} + e^{2-2x} + 2e\right) dx - (e-1)^2 =$$

$$= \frac{1}{4}\left(\frac{e^{2x}}{2} + e^2\,\frac{e^{-2x}}{-2} + 2ex\right)\Bigg|_0^1 - (e-1)^2 = -\frac{3}{4}e^2 + \frac{5}{2}e - \frac{5}{4}\,\boxed{\sim 0.0039}$$

Summary

Let's summarize

method	main characteristics
Newton-Cotes formulas	● pretty robust methods ● trapezoid can be done adaptively to get error estimate ● Simpson's rule can have troubles with singularities
Gaussian quadrature	● highly accurate with few points ● no error estimation or ability to do iteration ● need to first compute integration points
Monte Carlo integration	● accurate error estimate ● can be not very efficient