Application areas of high-dimensional integrals:

"$d \gg 1$"

↑ "dimension d much greater than 1"

* Physics; Statistical Mechanics

N gas atoms; each atom has three coordinates (x, y, z)

dimension: $d = 3N$

the "partition function" ("how does the gas depend on temperature T")

$$Z = \iiint \cdots \iiint e^{-\frac{\text{----}}{k_B T}} dx_1 dy_1 dz_1 dx_2 dy_2 dz_2 \cdots dx_N dy_N dz_N$$

for a realistic gas: $d > 10^9$

* Mathematical Finance; "options"

example: 100 stocks, 250 trading days

$$\Rightarrow d = 100 * 250 = 25000$$

Price =

$$\underbrace{\iint \cdots \int}_{25000 \text{ times}} \text{Payoff}(\underbrace{S_{1,1}, \cdots, S_{100,250}}_{25000}) * \text{Probability}(\underbrace{S_{1,1}, \cdots, S_{100,250}}_{25000})$$

$$\underbrace{dS_{1,1} \cdots dS_{100,250}}_{25000}$$

* Molecular Modelling (Chemistry/Biology)

average binding energy between a medicine and a virus:

$$\langle E \rangle = \underbrace{\iint \cdots \int}_{\text{volume}} \text{Energy}(N \text{ position vectors}) * \text{Probability}(N \text{ positions})$$

$d = 3N$ $\quad d\vec{r}_1, d\vec{r}_2 \cdots d\vec{r}_N$

1 molecule with N atoms $\Rightarrow d = 3000$ (if $N = 1000$)

* Modern application : Machine learning & Image generation

"AI"

image with noise (blurred) $\xrightarrow{\text{transform}}$ image of a cat (for example)

digital image of 512 x 512 pixels
with 3 colour channels (red, green, blue) $\Rightarrow$ 512 x 512 x 3
$= 786432$ numbers

AI sees the image as one single point
in a 786432 dimensional space.

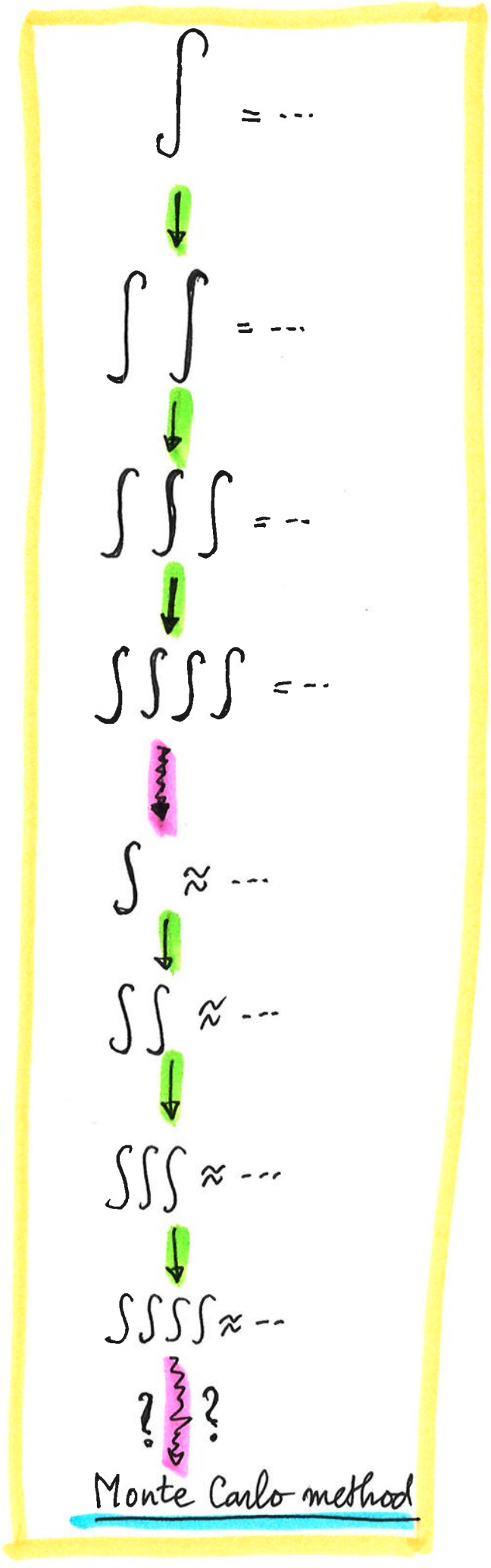
To find the image of the cat, AI needs to calculate the probability that a certain pixel combination a real cat forms: integration over a gigantic space.

$$P(\text{cat}) = \int\int \cdots \int P(\text{image}_{\text{with noise}}) * P(\text{noise}) \; d\text{pixel}_1 \cdots d\text{pixel}_{786432}$$

P(cat): probability

$\int\int \cdots \int$: 786432

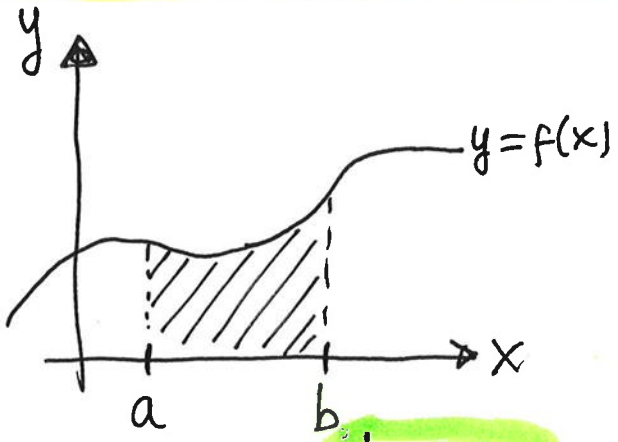
$d\text{pixel}_1 \cdots d\text{pixel}_{786432}$


= ...
= ...
= ...
= ...
≈ ...
≈ ...
≈ ...
≈ ...
≈ ...
? ?
Monte Carlo method

1 d

$y = f(x)$

a curve in the x-y plane



the area below the curve $y=f(x)$:

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

$= a$ number

interval $[a,b]$

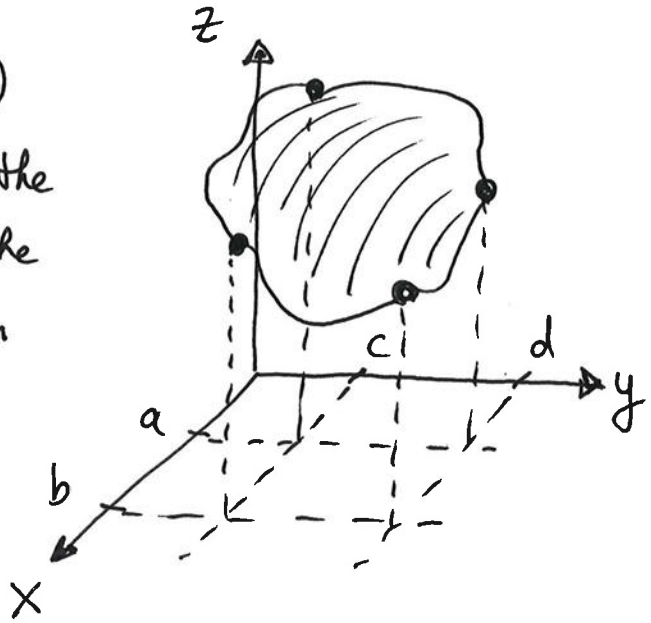
one-dimensional

$F(x)$ is the primitive or the anti-derivative of the function $f(x)$

2 d

$z = f(x,y)$

a surface of the function f in the x-y-z domain



the volume below the surface $z=f(x,y)$:

$$\int_c^d \int_a^b f(x,y)\,dx\,dy = \int_a^b \int_c^d f(x,y)\,dy\,dx$$

$= a$ number

rectangle $[a,b] \times [c,d]$

notation

two-dimensional

3d $W = f(x,y,z)$

a three-dimensional object in the space

x,y,z,w (4d) !!!

$$\underbrace{\iiint}_{R} f(x,y,z)\,dx\,dy\,dz$$

R ← three-dimensional

d dimensions

$$W = f(x_1, x_2, x_3, \cdots, x_d)$$

$$\underbrace{\iiint \cdots \iint}_{d \text{ integrals}} f(x_1, x_2, \cdots, x_d)\,dx_1\,dx_2 \cdots dx_d$$

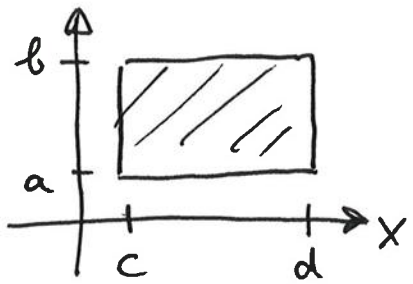
→ the triple integral $\underbrace{\iiint}_{3d \text{ object}} 1\,dx\,dy\,dz$

= the volume of the 3d object

[the double integral $\underbrace{\iint}_{2d \text{ area}} 1\,dx\,dy$ = the area of the 2d area]

[the single integral $\int_a^b 1\,dx = b-a$ = the length of the 1d interval]

The double integral $\int_a^b \int_c^d 1 \, dx \, dy$



$$= \int_a^b [x]_{x=c}^{x=d} \, dy$$

$$= \int_a^b (d-c) \, dy$$

$$= (d-c) \cdot \int_a^b 1 \, dy$$

$$= (d-c)(b-a) = \text{area of } \boxed{}$$

Example: $z = f(x,y) = x^2 y$

$$\int_{-2}^{4} \left\{ \int_1^3 f(x,y) \, dx \right\} dy = \int_{-2}^{4} \left\{ \frac{1}{3} x^3 y \Big|_{x=1}^{x=3} \right\} dy$$

$$= \int_{-2}^{4} \left[9y - \frac{1}{3} y \right] dy$$

$$= \int_{-2}^{4} \frac{26}{3} y \, dy = \frac{26}{3} \cdot \frac{1}{2} y^2 \Big|_{y=-2}^{y=4}$$

$$= \frac{26}{3} \cdot \frac{1}{2} (16-4)$$

$$= \frac{13}{3} 12 = 13 \cdot 4 = 52$$

and the other way round:

$$\int_1^3 \left\{ \int_{-2}^{4} f(x,y) \, dy \right\} dx = \int_1^3 \left\{ \frac{1}{2} x^2 y^2 \Big|_{y=-2}^{y=4} \right\} dx$$

$$= \int_1^3 (8x^2 - 2x^2) \, dx$$

$$= \int_1^3 6x^2 \, dx = \frac{6}{3} x^3 \Big|_{x=1}^{x=3} = \frac{6 \cdot 3^3}{3} - \frac{6 \cdot 1^3}{3}$$

$$= 52$$

the same result

Example: $f(x_1, x_2, x_3, x_4) = x_1 x_2 x_3 x_4$

over the "hypercube"

$$[0,1]^4 = [0,1] \times [0,1] \times [0,1] \times [0,1]$$

(x_1, x_2, x_3, x_4)

we can not visualize this !!

$$\int_0^1 \int_0^1 \int_0^1 \int_0^1 f(x_1, x_2, x_3, x_4)\, dx_1 dx_2 dx_3 dx_4$$

$$= \int_0^1 \int_0^1 \int_0^1 \int_0^1 x_1 x_2 x_3 x_4 \, dx_1 dx_2 dx_3 dx_4$$

because of simple product

$$= \int_0^1 x_1 dx_1 * \int_0^1 x_2 dx_2 * \int_0^1 x_3 dx_3 * \int_0^1 x_4 dx_4$$

$$= \frac{1}{2} x_1^2 \Big|_0^1 * \frac{1}{2} x_2^2 \Big|_0^1 * \frac{1}{2} x_3^2 \Big|_0^1 * \frac{1}{2} x_4^2 \Big|_0^1$$

$$= \frac{1}{2} * \frac{1}{2} * \frac{1}{2} * \frac{1}{2} = \frac{1}{2^4} = \frac{1}{16}$$

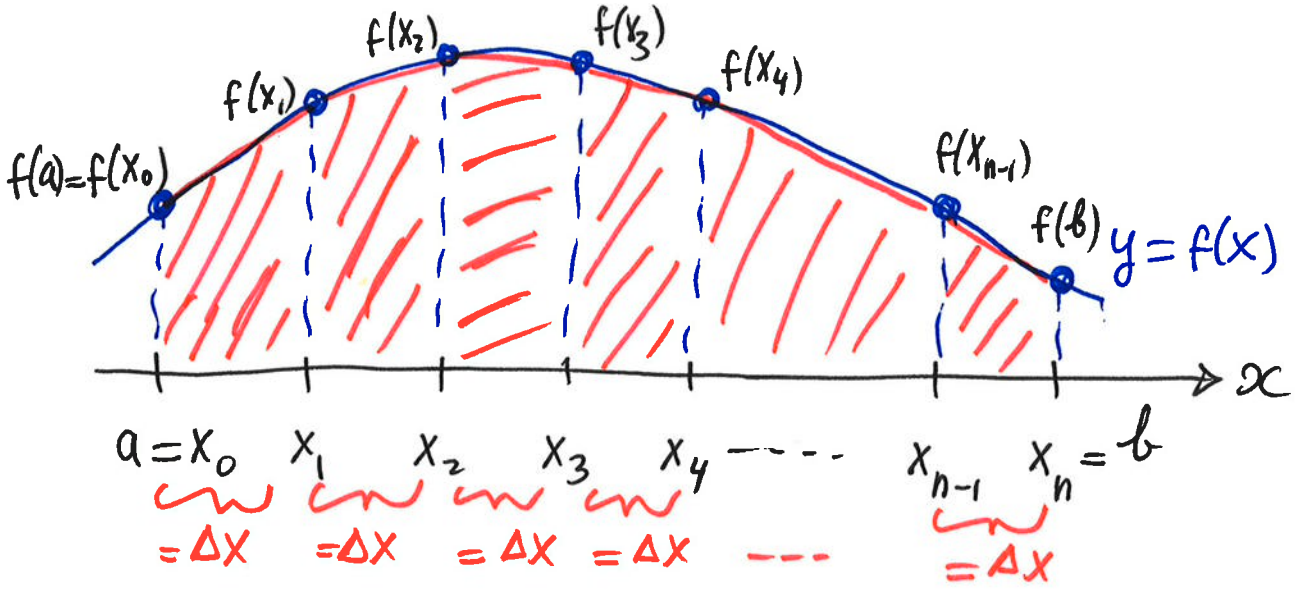
How to apply integration if $f(x)$ or $f(x,y)$ or ... are too complicated to calculate exactly ?

We need to apply Numerical Integration ! "Quadrature"

Example: what is the value of $\int_a^b e^{-x^2} dx$?

No primitive (anti-derivative) is available !

$\Rightarrow$ approximate the integral



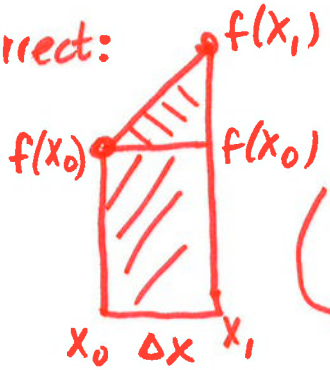
- subdivide the interval $[a,b]$ into n parts of length Δx

$$\Rightarrow \frac{b-a}{n} = \Delta x \quad (\text{or } n\Delta x = b-a)$$

- approximate $\int_a^b f$ by the area of the red marked pieces "trapezoids"

$$\Rightarrow \int_a^b f(x)\,dx \approx \frac{1}{2}\Delta x\,(f(x_0)+f(x_1)) + \frac{1}{2}\Delta x\,(f(x_1)+f(x_2)) + \frac{1}{2}\Delta x\,(f(x_2)+f(x_3)) + \dots + \frac{1}{2}\Delta x\,(f(x_{n-1})+f(x_n))$$

why is this correct:



area $\square = \Delta x * f(x_0)$

area $\triangle = \frac{1}{2}\Delta x\,(f(x_1)-f(x_0))$

$\Delta x\, f(x_0) + \frac{1}{2}\Delta x\, f(x_1) - \frac{1}{2}\Delta x f(x_0)$

$= \frac{1}{2}\Delta x\, f(x_0) + \frac{1}{2}\Delta x\, f(x_1)$

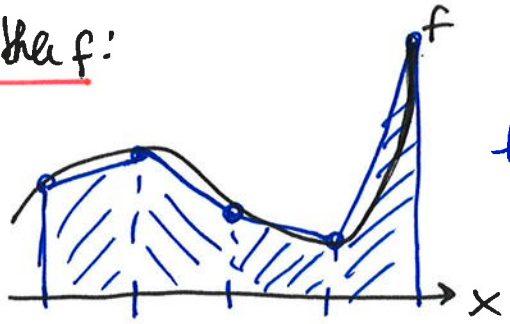
and also for the pieces

$$= \frac{1}{2}\Delta x\, f(x_0) + \underbrace{\left(\tfrac{1}{2}+\tfrac{1}{2}\right)}_{=1}\Delta x f(x_1) + \underbrace{\left(\tfrac{1}{2}+\tfrac{1}{2}\right)}_{=1}\Delta x f(x_2) + \dots + \underbrace{\left(\tfrac{1}{2}+\tfrac{1}{2}\right)}_{=1}\Delta x f(x_{n-1}) + \frac{1}{2}\Delta x f(x_n)$$

TRAPEZOIDAL RULE

approximation looks very accurate (from the picture)

another f:



less accurate ---

the accuracy depends on f'

this can be derived from a Taylor expansion

error in trapezoid rule:

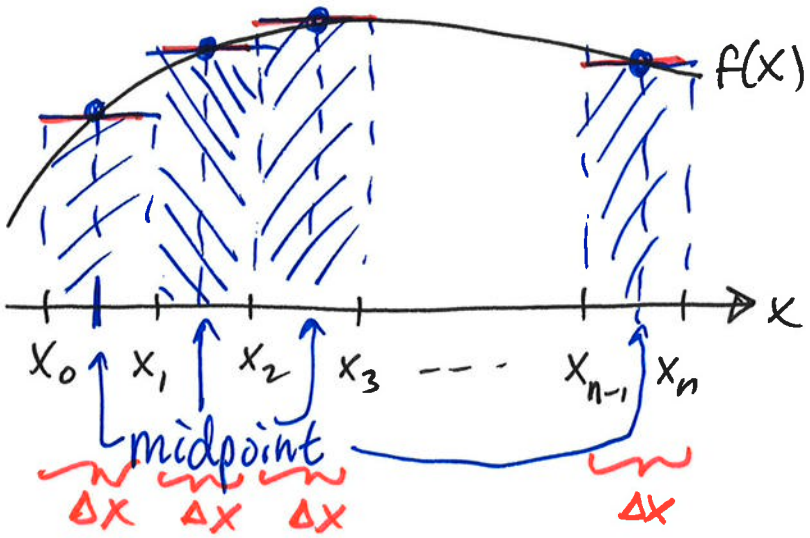
$$\approx -\frac{(b-a)^2(\Delta x)^2}{12}\langle f''\rangle$$

fixed number

smaller Δx ⇓ error quadratically smaller.

the average value of f'' over the whole interval

Another approach:



Midpoint rule (rectangular rule)

$$\int_a^b f(x)\,dx \approx \Delta x\, f\left(\frac{a+x_1}{2}\right) + \Delta x\, f\left(\frac{x_1+x_2}{2}\right) + \Delta x\, f\left(\frac{x_2+x_3}{2}\right) + \cdots + \Delta x\, f\left(\frac{x_{n-1}+b}{2}\right)$$

$$\text{error (via Taylor --)} \approx \frac{(b-a)}{3}(\Delta x)^2\langle f''\rangle$$

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

Summarized:

Trapezoidal rule: $I \approx \frac{\Delta x}{2} \sum_{i=1}^{n} \left(f(x_{i-1}) + f(x_i) \right)$

error $\sim (\Delta x)^2$

Midpoint rule: $I \approx \Delta x \sum_{i=1}^{n} f\left(\frac{x_{i-1} + x_i}{2} \right)$

error $\sim (\Delta x)^2$

another (more accurate) option:

Simpson's rule: $I \approx \frac{\Delta x}{3} \left[f(x_0) + 4 \sum_{\text{odd indices}} f(x_i) + 2 \sum_{\text{even indices}} f(x_i) + f(x_n) \right]$

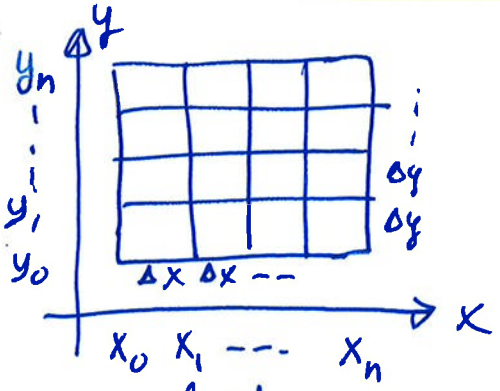
error $\sim (\Delta x)^4$

Note that Δx is small

$\Rightarrow (\Delta x)^4 \ll (\Delta x)^2$

↑ much smaller

In 2d:



2d Trapezoidrule: $\int_a^b \int_c^d f(x,y)\,dx\,dy$

$\approx \frac{\Delta x \Delta y}{4} \left\{ \text{corner points} + \text{--- sides} + 4 \sum_{j=1}^{n} \sum_{i=1}^{n} f(x_i, x_j) \right\}$

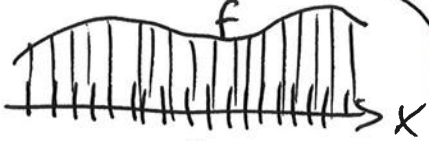
complicated ("expensive") formula!

What about an integral in d dimensions ?

for Trapezoid, midpoint, Simpson, ...

in 1d : N points
in 2d : N^2 points
in 3d : N^3 points
⋮
in d dimensions : N^d points !

example $N = 100$

1d 

⇒ in 6 dimensions :

100^6 calculations

$= 10^{12}$

This phenomenon is called "the Curse of Dimensionality"

$N = 1000 \Rightarrow 1000^{10}$

in 10 dimensions $= (10^3)^{10}$

$= 10^{30}$ calculations



Error dependency on the dimension d

let M = total number of evaluations of f : $M = N^d$ $(\Leftrightarrow N = M^{1/d})$

⇒ Trapezoidal error $\sim N^{-2} \sim M^{-2/d}$

Simpson's rule error $\sim N^{-4} \sim M^{-4/d}$

Why this fails for large dimension d?

if dimension $d=8$: Simpson's rule error $\sim M^{-1/2} = \frac{1}{\sqrt{M}}$

"factor of 10" smaller

to gain one extra decimal place of accuracy,
you need to multiply your computation size
by a factor of 100!

What to do --- ??

Monte Carlo Integration

in 1d:

The exact average (mean) of a function $f(x)$ over an interval $[a,b]$: $\langle f \rangle = \frac{1}{b-a}\int_a^b f(x)\,dx$

$$\Rightarrow \int_a^b f(x)\,dx = (b-a) * \langle f \rangle$$

and $\langle f \rangle \approx \frac{\sum_{i=1}^{n} f(x_i)}{n}$

$$\Rightarrow \int_a^b f(x)\,dx \approx \frac{b-a}{n}\sum_{i=1}^{n} f(x_i)$$

in d dimensions:

$$\underbrace{\iint\cdots\int}_{d \text{ times}} f(x_1, x_2, \cdots, x_d)\,dx_1 dx_2 \cdots dx_n$$

$$\approx \frac{V}{M}\sum_{i=1}^{M} f(x_1^{(i)}, x_2^{(i)}, \cdots, x_d^{(i)})$$

V: volume of d dimensional domain

M: total number of points in d dimensions

The error in Monte Carlo

follows from statistics ("Central limit Theorem"):

$$\text{error} \sim \frac{1}{\sqrt{M}} \quad \text{for all dimensions } d \quad !!!$$

(no d in this formula)

the speed of error decay is completely independant of the dimension d !

Example $f(x_1, x_2, \dots, x_d) = \sin(x_1 + x_2 + \dots + x_6)$

on the six dimensional domain

$$[0,\pi]^6 = [0,\pi] \times \dots \times [0,\pi]$$

(x_1 … x_6)

Simpson's rule: 20 points per axis

$\Rightarrow 20^6 = 64\,000\,000$ calculations

from 6 → 7 dimensions: $20^7 = 1\,280\,000\,00$ --- !!!

(breaks computer memory)

Monte Carlo: take $M = 10^6$

$$\Rightarrow \text{error} \sim \frac{1}{\sqrt{10^6}} = 10^{-3}$$

Comparison: for which dimension d we take which method?

The so-called "break-even point"

Monte Carlo "$\gtrsim$" Trapezoidal rule (in terms of efficiency)

$\sim \frac{1}{\sqrt{M}}$ (error) $\qquad \sim \frac{1}{M^{2/d}}$ (error)

when "equal"? $\Rightarrow$ $d = 4$ (four dimensions)

$\Rightarrow$ for $d > 4$ Monte Carlo more efficient than Trapezoidal rule

Simpson's rule $\rightarrow$ error $\sim \frac{1}{M^{4/d}}$

$\Rightarrow$ for $d > 8$ Monte Carlo more efficient than Simpson

