Comments on singularities

Open-interval formulas can be used

- singular point(s) should be at end(s); divide up interval in parts if needed
- but convergence with number of points n may be very slow

Divergent part can some times be subtracted and solved analytically

More sophisticated methods exist for difficult cases

Other methods

Gaussian quadrature:

- non-uniform grid points; n+1 points → exact result for polynomial of order n
- several Julia packages, e.g., FastGaussQuadrature.jl

Gauss-Kronrod quadrature:

- uses two Gaussian quad. evaluations for different n, similarly to Romberg
- package QuadGK.jl uses a version of this method

Adaptive grid (adaptive mesh):

- dynamically adapted to be more dense where most needed

Infinite integration range

Change variables to make range finite

Multi-Dimensional integration

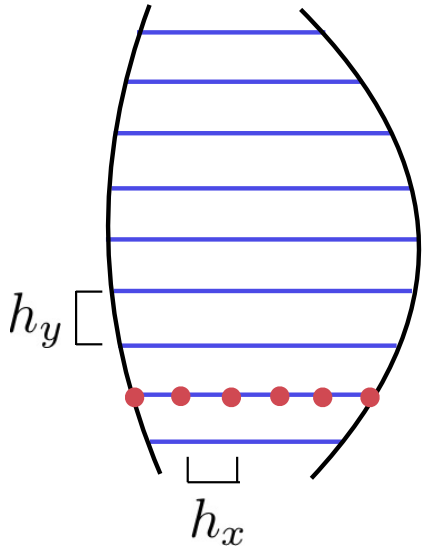
$$I = \int_{a_n}^{b_n} dx_n \cdots \int_{a_2}^{b_2} dx_2 \int_{a_n}^{b_n} dx_1 f(x_1, x_2, \ldots, x_n),$$

Can be carried out numerically dimension-by-dimension
Example, function of two variables

$$I = \int_{a_y}^{b_y} dy \int_{a_x(y)}^{b_x(y)} dx f(x, y)$$

Integrating numerically over x first, gives a function of y:

$$F(y) = \int_{a_x(y)}^{b_x(y)} dx f(x, y)$$



This has to be done for values of y on a grid, to be used in

$$I = \int_{a_y}^{b_y} dy F(y)$$

Very time consuming for large dimensionality D; scaling M^D of effort
- M represents mean (geom) number of grid points for 1D integrals

Monte Carlo Integration

An integral over a finite volume V:

- is (by definition) the mean value of the function times the volume

$$I = \int_a^b f(x)dx = (b-a)\langle f \rangle$$

The mean value <f> can be estimated by sampling

- generate N random (uniformly distributed) x values x_i in the range, then

$$\bar{f} = \frac{1}{N}\sum_{i=1}^{N} f(x_i) \to \langle f \rangle, \quad \text{when } N \to \infty$$

For finite N, there is a statistical error:

$$\left\langle \bar{f} - \langle f \rangle \right\rangle \propto \frac{1}{\sqrt{N}}$$

interepretation of the mean error:
If the "simulation" is repeated many times, the averaged squared error (variance) tends to a value ***a/N***, for with ***a*** some constant

The statistical result for the integral should be expressed as

$$I = \bar{I} \pm \sigma = V(\bar{f} + \sigma/V) \qquad \sigma \propto N^{-1/2}$$

Computing the "error bar" σ is an important aspect of the sampling method

Standard ilustration of MC integration; estimate of π

Consider a circle of radius 1, centered at (x,y)=0. Define a function:

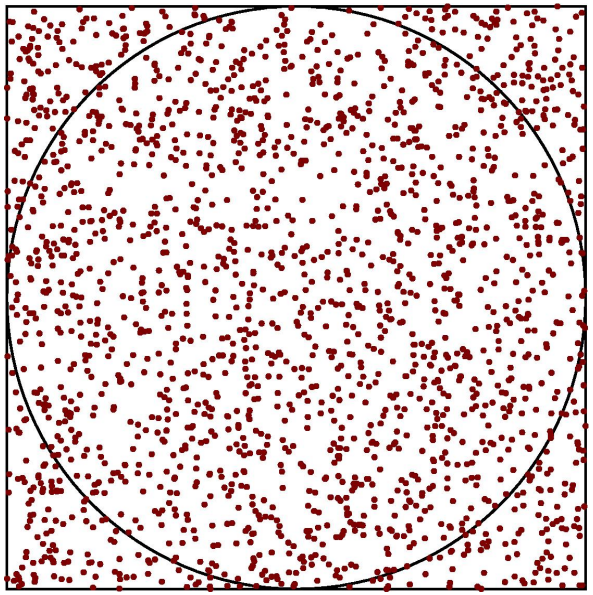
$$f(x,y) = \begin{cases} 1, & \text{if } x^2+y^2 \le 1 \\ 0, & \text{if } x^2+y^2 > 1 \end{cases}$$

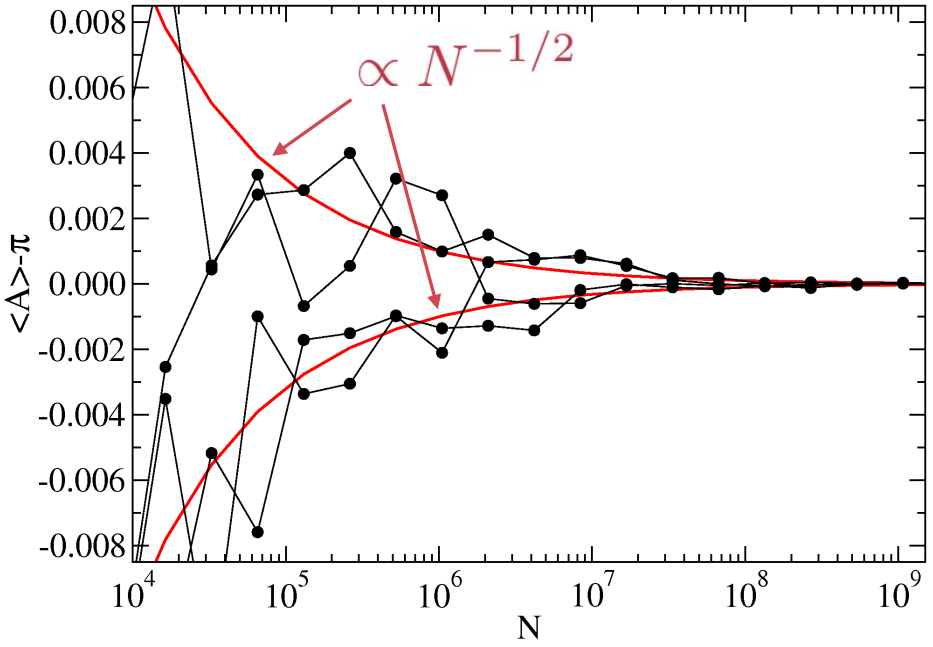
mean value inside the surrounding box

Use MC sampling to compute:

$$A = \int_{-1}^{1} dy \int_{-1}^{1} dx f(x,y) = \pi \; = 4\langle f \rangle_{\square}$$

Expected fraction of "hits" inside circle = π/4





The error after N steps

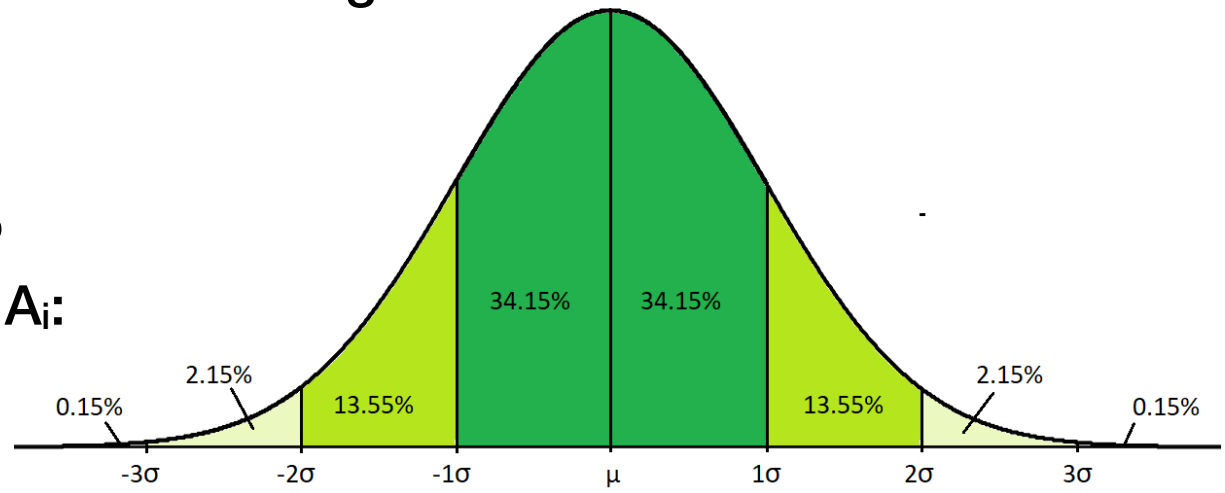
Four repetitions of a simulation, dots showing partial results as the mean value evolves

We should compute the statistical error properly

Statistical errors

Expressing a statistical estimate as A ± σ, the meaning normally is
- σ represents one standard deviation of the computed mean value A
- under the assumption of normal-distributed fluctuations

Then, the probability of the true value being
- within [A-σ,A+σ] is 68%
- within [A-2σ,A+2σ] is 95%
- within [A-3σ,A+3σ] is 99.7%



For M independent samples A_i:

$$\bar{A} = \frac{1}{M}\sum_{i=1}^{M} A_i$$

$$\sigma_A = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(A_i - \bar{A})^2} = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(A_i^2 - \bar{A}^2)}$$

$$= \sqrt{\overline{A^2} - (\bar{A})^2}$$

This is the standard deviation of the distribution of values $\{A_i\}$

But the "error bar" is the standard deviation of the mean of $\{A_i\}$

The mean value fluctuates less than the width σ_A of the distribution

- imagine taking the number of samples M to infinity:

$$\sigma_A = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(A_i - \bar{A})^2}$$

will approach a constant value
- the standard deviation of the distribution

$$\bar{A} = \frac{1}{M}\sum_{i=1}^{M} A_i$$

will approach a constant value
- the actual value <A> of A

→ σ_A cannot be the proper statistical error of A

Variances add: variance of the sum $\sum_{i=1}^{M} A_i$ is $M\sigma_A^2$

- standard deviation of the sum is $\sqrt{M}\sigma_A$
- divide by M; standard deviation of the mean is $\sigma_A/\sqrt{M}$
- here M should be replaced by M-1 (reflecting infinite uncertainty if M=1)

$$\sigma = \sqrt{\frac{1}{M(M-1)}\sum_{i=1}^{M}(A_i - \bar{A})^2} = \sqrt{\frac{1}{M(M-1)}\sum_{i=1}^{M}(A_i^2 - \bar{A}^2)} = \sqrt{\frac{\overline{A^2} - (\bar{A})^2}{M-1}}$$

Data binning

The statistical error ("error bar") has its conventional meaning only if the values $\{A_i\}$ are normal distributed

- typically they obey some completely different distribution

Apply **central limit theorem** to obtain normal distributed "bin averages"

A **bin average** is based on M samples as before, but now B of them

- B different mean values (estimates of A): $\bar{A}_1, \bar{A}_2, \ldots, \bar{A}_B$

$$\bar{A}_b = \frac{1}{M}\sum_{i=1}^{M} A_{b,i}$$

$A_{b,i}$ is value #i belonging to bin b

Regardless of the distribution of individual values

- if M is large enough, the bin averages are normal-distributed

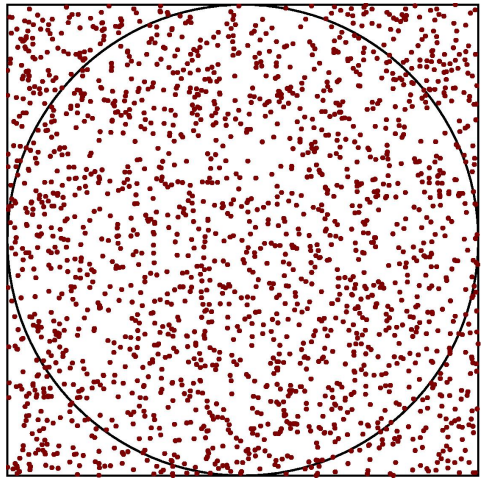
Use standard formulas with the bin data:

$$\bar{A} = \frac{1}{B}\sum_{b=1}^{B} \bar{A}_b \qquad \sigma = \sqrt{\frac{1}{B(B-1)}\sum_{b=1}^{B}(\bar{A}_b - \bar{A})^2} = \sqrt{\frac{1}{B(B-1)}\sum_{b=1}^{B}(\bar{A}_b^2 - \bar{A}^2)} = \sqrt{\frac{\overline{A^2} - (\bar{A})^2}{B-1}}$$

Emergence of normal distribution

- example: sampling f=1circle in square
- lets just consider the estimate of the mean <f>

$$f(x,y) = \begin{cases} 1, & \text{if } x^2+y^2 \le 1 \\ 0, & \text{if } x^2+y^2 < 1 \end{cases}$$



For each sample, the probabilities of f=0,1 are:

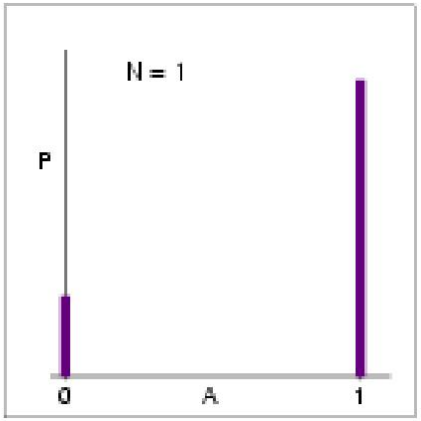
$$P(f=1) = \pi/4, \quad P(f=0) = 1-\pi/4$$

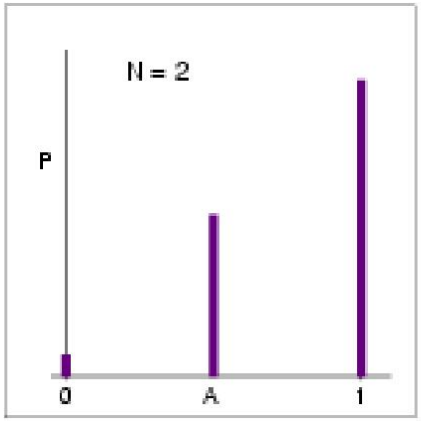
For N samples, the possible average values A are

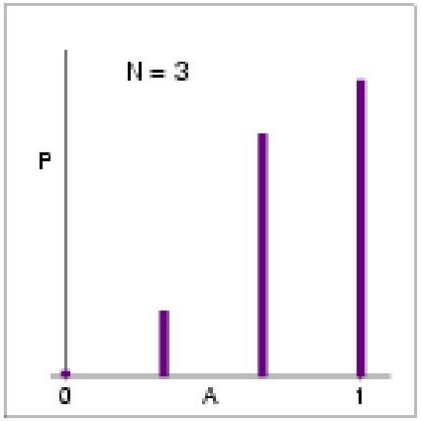
$$A \in \left\{0, \frac{1}{N}, \frac{2}{N}, \ldots, \frac{N-1}{N}, 1\right\}$$

the probabilities of these averages are

$$P\left(A = \frac{m}{N}\right) = \frac{N!}{m!(N-m)!}\left(\frac{\pi}{4}\right)^m \left(1-\frac{\pi}{4}\right)^{N-m}$$





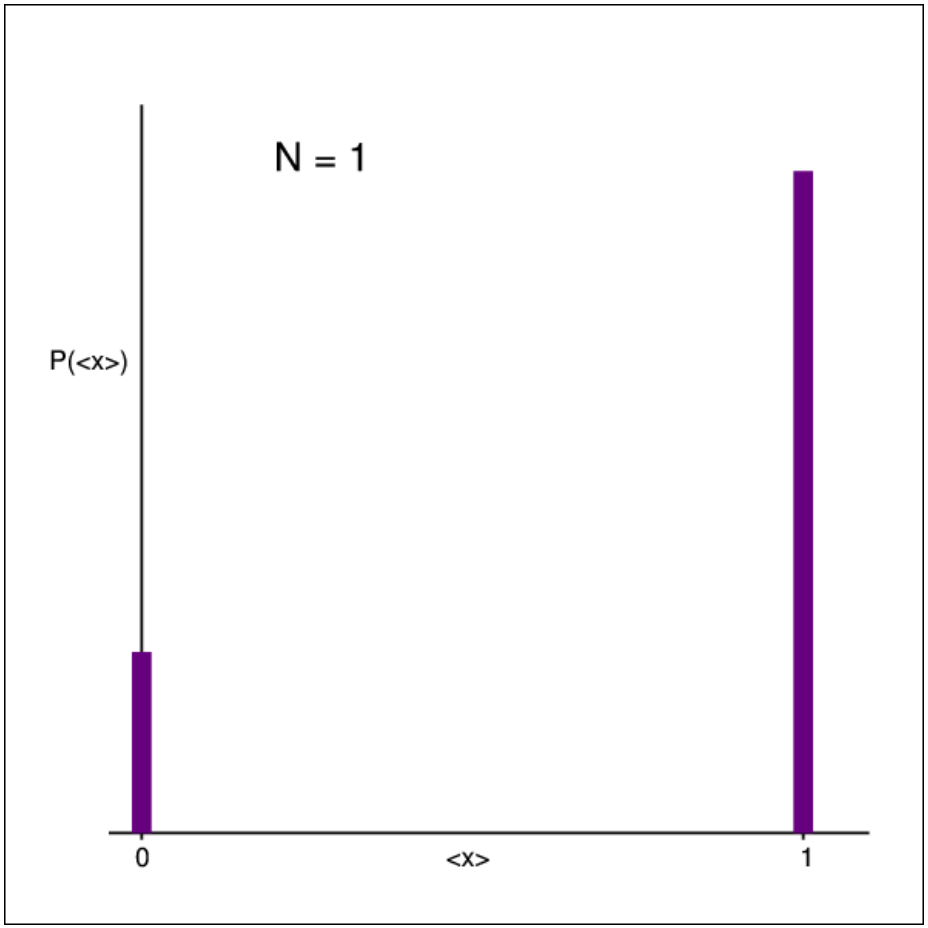


Evolution of P(A)
from N=1 to 100

Note: We can think of the probability distribution of a continuum of A values

P(A) is a sum of delta-functions; reflects discrete set of possible values

For large N, a small broadening of the deltas (e.g., bars or Gaussians) give a continuous distribution



$$P(A) = \sum_{m=0}^{N} \frac{N!}{m!(N-m)!} \left(\frac{\pi}{4}\right)^{m} \left(1-\frac{\pi}{4}\right)^{N-m} \delta(A - m/N)$$