Monte Carlo Integration

"Monte Carlo" methods use random numbers for sampling

Although somewhat less glamorous than gambling devices (dice, roulette, cards, etc.) random number generators on the computer are more efficient (>10^8 random numbers / s)

Fortran 90 intrinsic subroutine: `random_number(r)`

-generates random floats in the range [0,1)

- other (external) generators may be better

An integral can be written as an average:

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

Estimate of average based on N random points $a \leq x_i \leq b$

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

Fortran 90 implementation:

```
average=0.d0
do i=1,n
   call random_number(r)
   average=average+func(a+r*(b-a))
enddo
average=average/n
```

Fluctuations (statistical error; “error bar”): $\sigma \sim 1/\sqrt{N}$

Time-scaling (number of ops) for D dimensional integration

Numerical integration on regular mesh:

$$t \sim M_1^D$$

Monte Carlo integration to desired accuracy (error = σ):

$$t \sim N \sim \frac{1}{\sigma^2}$$ (D independent)

Monte Carlo is more efficient for large D

Example: Statistical mechanics of many-particle systems (liquids, gases,...) D = 3N (N = # of particles)

Simple Monte Carlo integration problematic if integrand has sharp peaks (rarely visited); Use **importance sampling** methods

First: Error analysis in simple Monte Carlo integration

- carries over to importance sampling discussed later

Standard illustration: the area of a circle

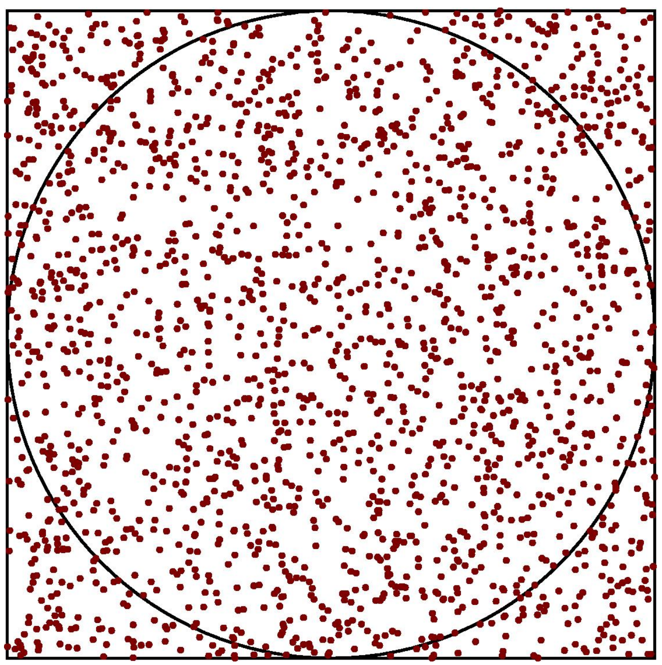
$$A = \int_{-1}^{1} dy \int_{-1}^{1} dx f(x,y), \quad f(x,y)=1, \text{ if } x^2+y^2 \le 1, \quad f(x,y)=0, \text{ if } x^2+y^2 > 1.$$

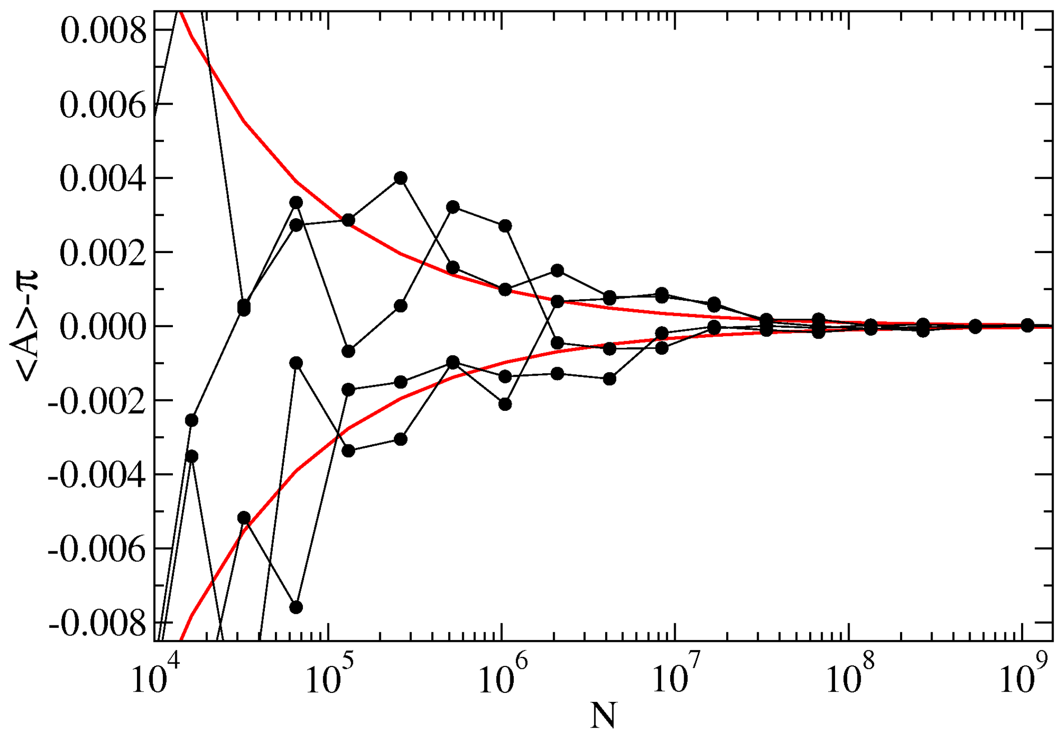
Example using random generator (function) ran():

```
do i=1,n
  x=2.d0*ran()-1.d0; y=2.d0*ran()-1.d0
  if (x**2+y**2 < 1.d0) a=a+1.d0
Enddo
A=4.d0*a/n
```

Gives an estimate of $A = \pi$

4 independent runs





Statistical errors ("error bars")

Calculation based on N points. What is the error?

Consider M independent calculations (each with N points)

Statistically independent averages $\bar{A}_i, \; i = 1, \ldots, M$

Average

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

Standard deviation

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

But, we want the standard deviation of the average

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

The M independent calculations are often referred to as "bins"

The standard deviation (error) has a well defined meaning if

- the bin averages follow a Gaussian distribution

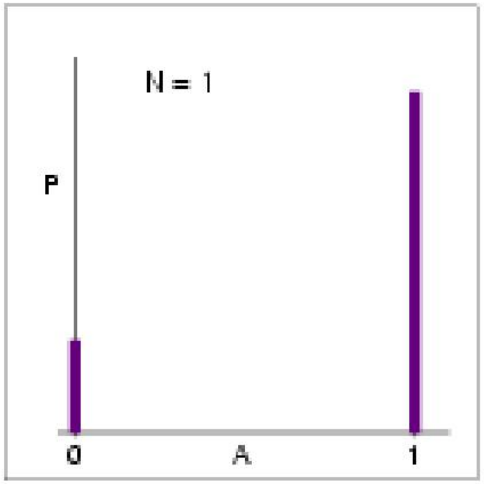
Law of large numbers

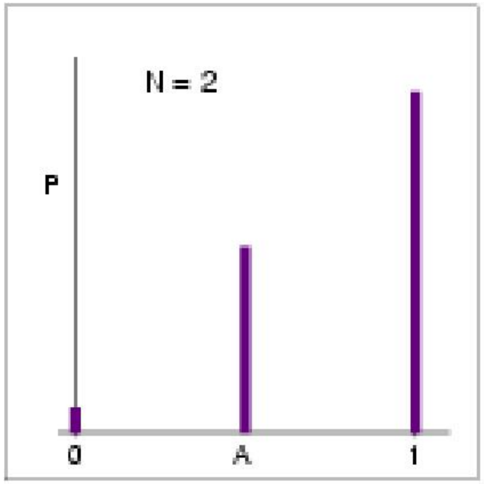
- Gaussian distribution approached for large N (points/bin)

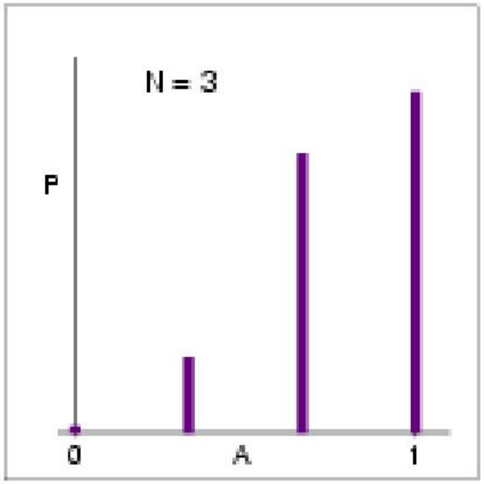
In the circle area calculation; binomial distribution:

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

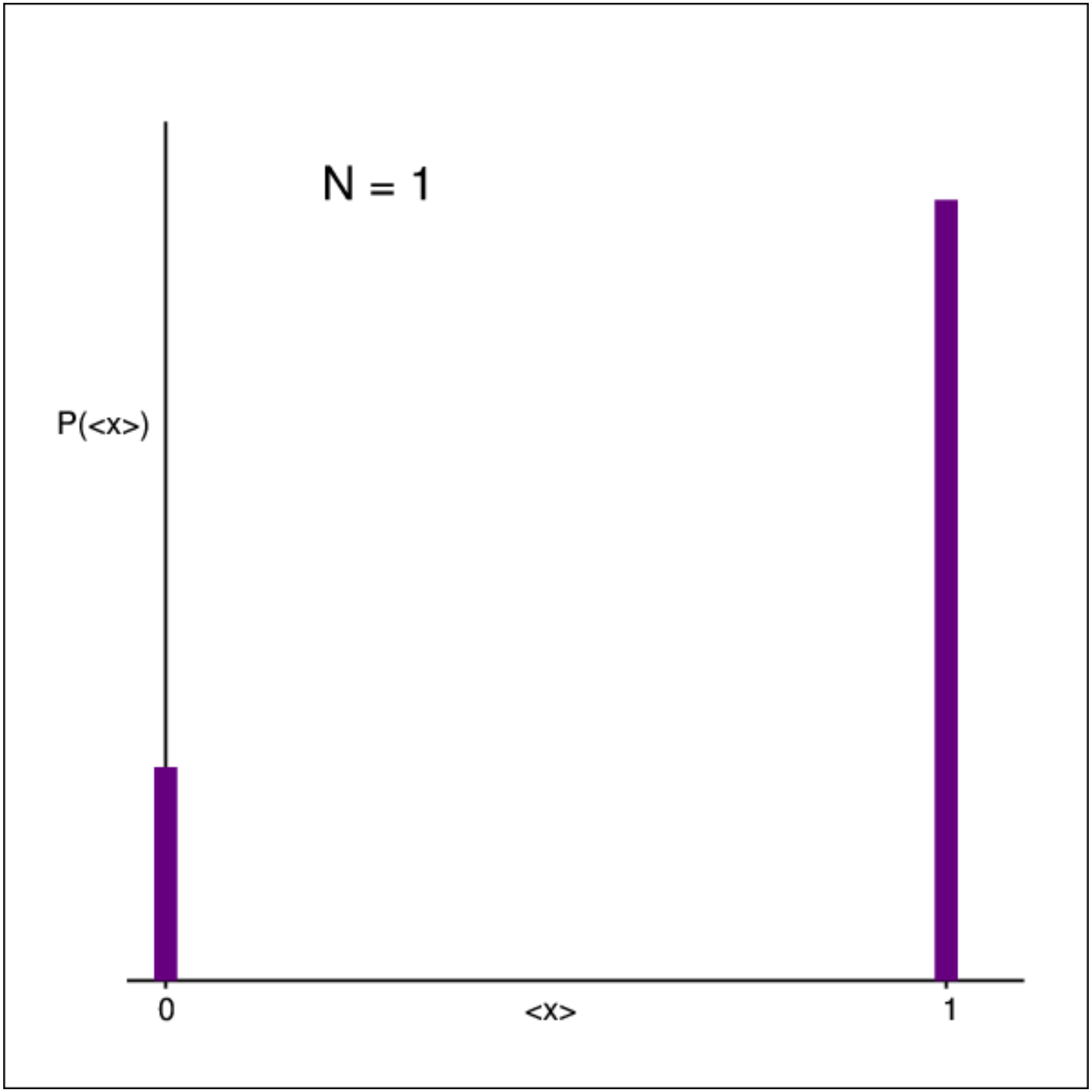
N=1 (A=0,1), N=2 (A=0,1/2, 1), N=3 (A=0,1/3,2/3,1),...







Let’s look at the distribution for N=1,...,100!



Program “circle.f90”; computing error bars

```
do j=1,nbin
   sum=0.d0
   do i=1,n
      call random_number(r)
      x2=(dble(r-0.5))**2
      call random_number(r)
      y2=(dble(r-0.5))**2
      if (x2+y2 <= 0.25d0) sum=sum+1.d0
   enddo
   sum=4.d0*sum/dble(n)
   av=av+sum
   er=er+sum**2
   write(*,'(A,I3,A,F11.6)')'Bin: ',j,' Result: ',sum
enddo
av=av/dble(nbin)
er=er/dble(nbin)
er=sqrt((er-av**2)/dble(nbin-1))
```

Monte Carlo integration less efficient if

- the integrand has sharp peaks (visited infrequently)

Example: Modified circle integration; singularity at r=0

$$f(r) = r^{-\alpha}, \quad r = \sqrt{x^2 + y^2}$$ Integrable if $\alpha < 2$

$$I = \int_{-1}^{1} dy \int_{-1}^{1} dx f(x,y), \quad f(x,y) = r^{-\alpha}, \text{ if } r \le 1, \quad f(x,y) = 0, \text{ if } r > 1$$

Distribution of r inside circle: $P(r) = 2r$

Distribution of function values inside circle:

$$P(f)df = P(r)\left|\frac{dr}{df}\right| df = \frac{2}{\alpha} f^{-1-2/\alpha} df$$

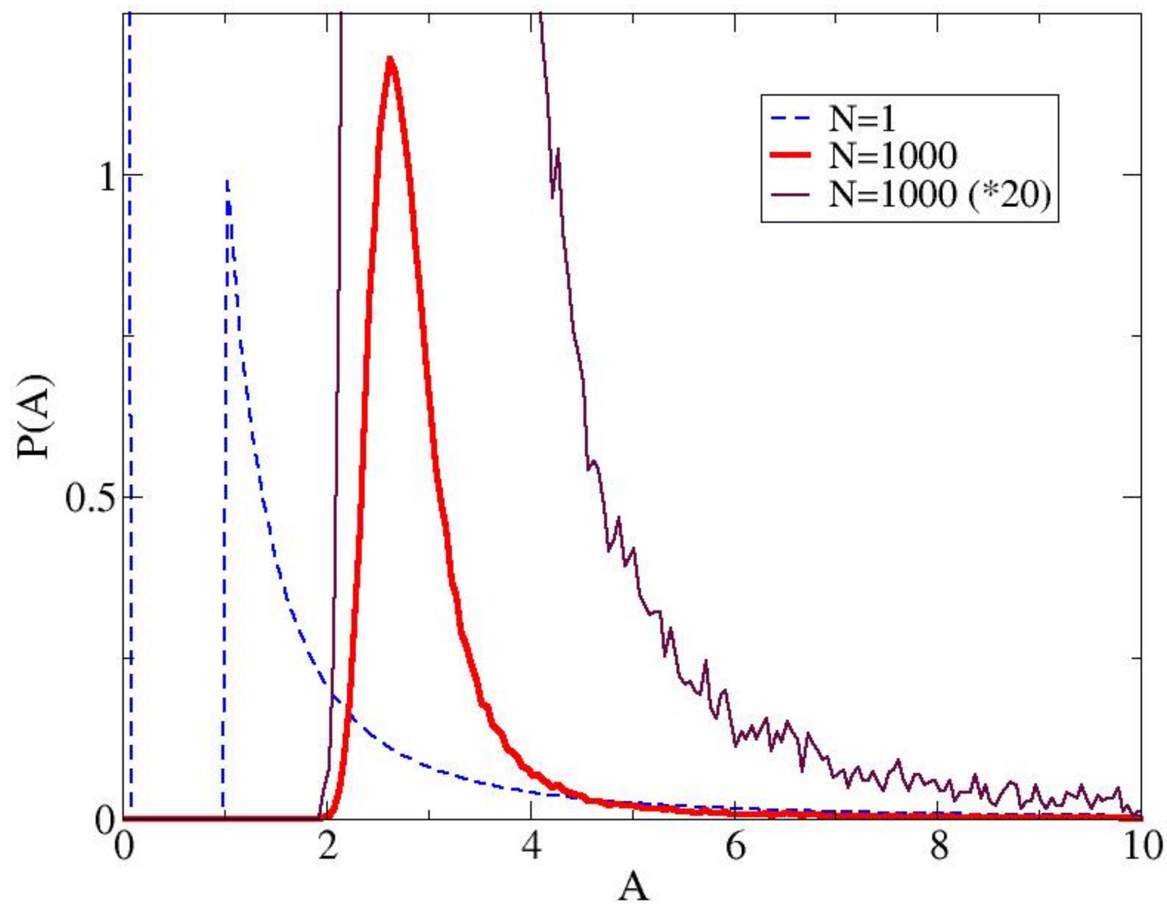
Probability of point falling inside circle: $\pi/4$

$$P(f) = (1 - \pi/4)\delta(f) + \frac{\pi}{4}\frac{2}{\alpha} f^{-1-2/\alpha}\Theta(f-1)$$

Distribution of average over N samples:

$$P(A) = \int_0^\infty df_N \cdots \int_0^\infty df_1 P(f_N) \cdots P(f_1) \delta[A - (f_1 + \cdots + f_N)/N]$$

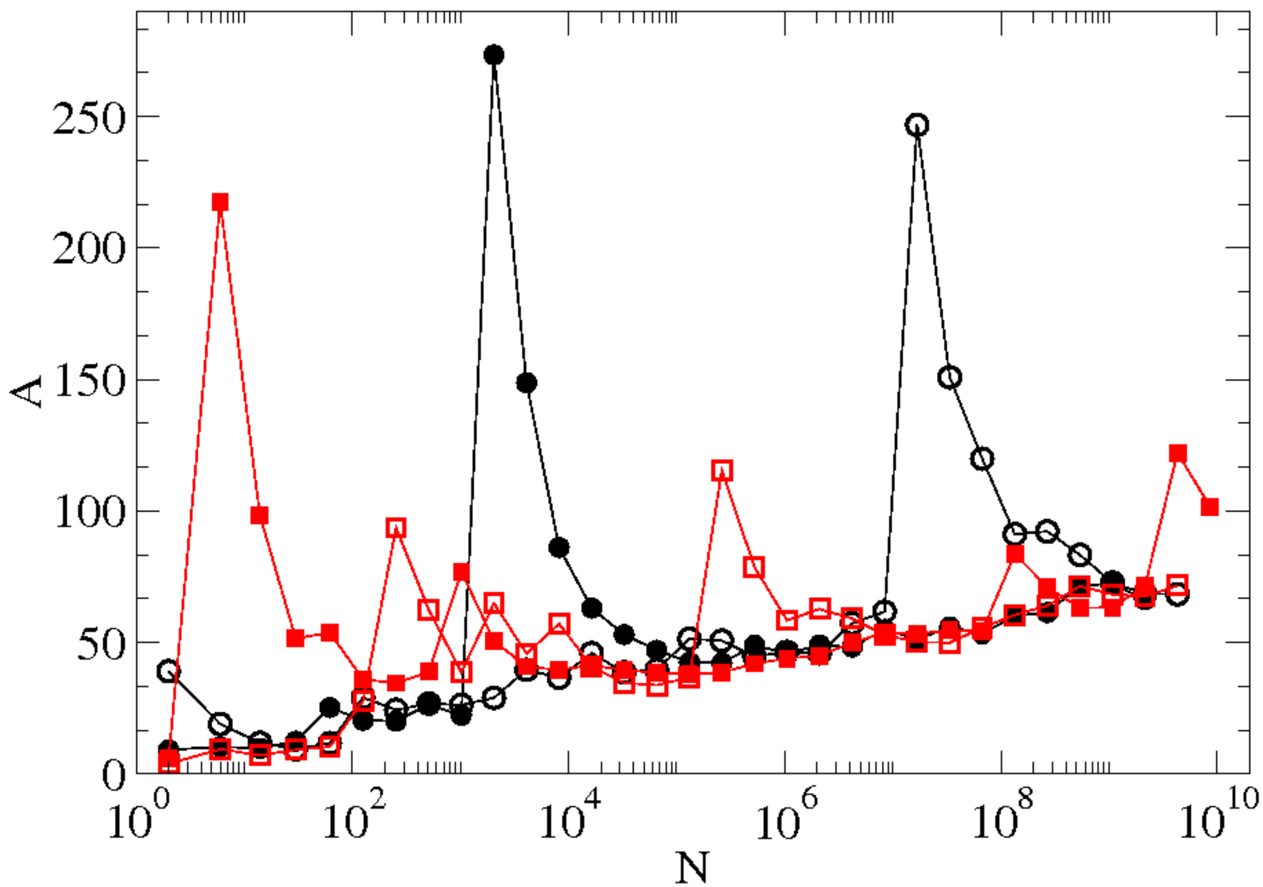
Monte Carlo calculation of P(A) for N=1000, $\alpha = 3/2$



- Still far from Gaussian distribution
- Broad distribution implies large error bars
- Bin size >> 1000 required to compute error bars reliably

What happens if we attempt to calculate a divergent integral?

Example: Same integral as before with $f(r) = r^{-\alpha}, \alpha = 2$



3 independent runs

- Erratic behavior seen
- Arbitrarily large fluctuations upward
- Well-defined distribution of A for finite N
- Log-divergent typical value vs N
- Average not defined (infinite) for N=∞