past week:

3. Intrinsic generators

(b) For a quantitative test of *uniformity* consider the moment of order k:

$$\langle x^k \rangle^{calc} = \frac{1}{N} \sum_{i=1}^N x_i^k; \qquad \langle x^k \rangle^{th} = \int_0^1 dx \ x^k \ P(x)$$

For the uniform distribution $p_u(x)$ in [0,1[, i.e. for

$$p_u(x) = \begin{cases} 1 \ for \ 0 \le x \le 1\\ 0 \ outside \end{cases}$$

we have $\langle x^k \rangle^{th} = 1/(k+1)$. Consider the error

$$\Delta_N(k) = \left| \langle x^k \rangle^{calc} - \langle x^k \rangle^{th} \right| = \left| \frac{1}{N} \sum_{i=1}^N x_i^k - \frac{1}{k+1} \right|$$

for the expected moment of order k and study its asymptotic behaviour for large N. If the behaviour is $\sim 1/\sqrt{N}$, then the distribution is random and uniform. Do the test for k=1, 3, 7, and N=100, 10.000, 100.000.

how to calculate the sum of the series for increasing N?

no need of recalculating again the sum from scratch; print out **partial** sums:

```
ALLOCATE(RND(N))
CALL RANDOM_NUMBER(RND)
SUM=0
```

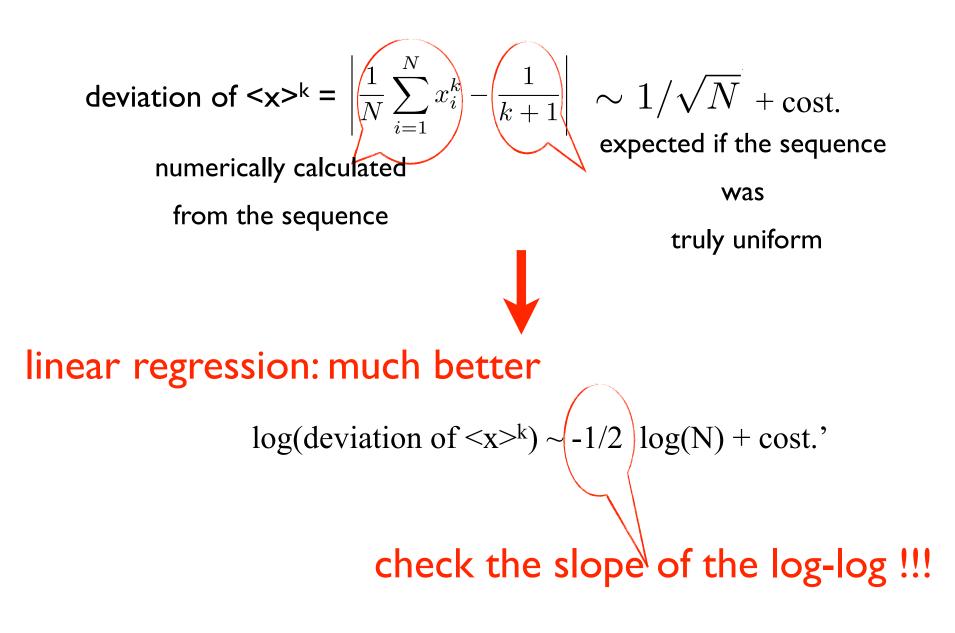
```
DO I=1, N

SUM=SUM+RND(I)

PRINT*, I, SUM/I \longrightarrow print out the result as a

END DO function of "i"
```

do you want to check a power law?



do you want to fit with gnuplot?

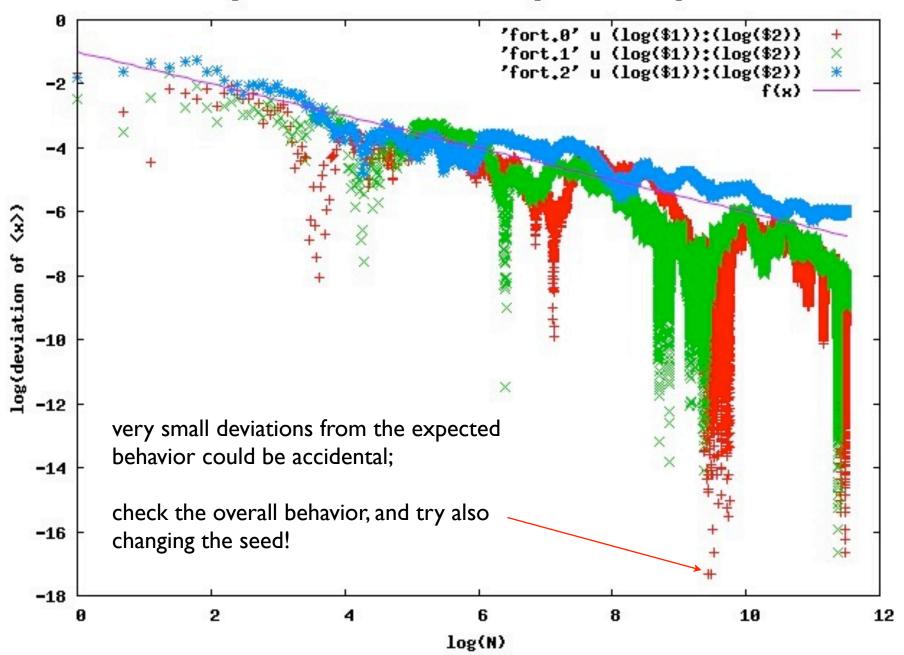
Suppose you have the data in two columns, x and y, and you suspect a power low $y = x^a + const$

Consider that: log(y) = a * log(x) + b

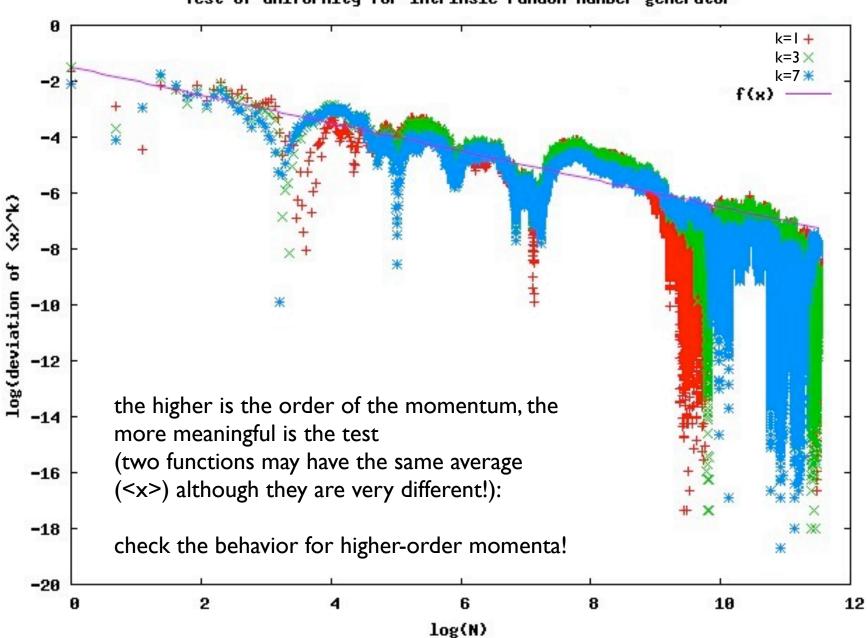
gnuplot> f(x) = a * x + b

gnuplot> fit f(x) 'data.dat' u (log(\$1)):(log(\$2)) via a,b

gnuplot> plot f(x), 'data.dat'







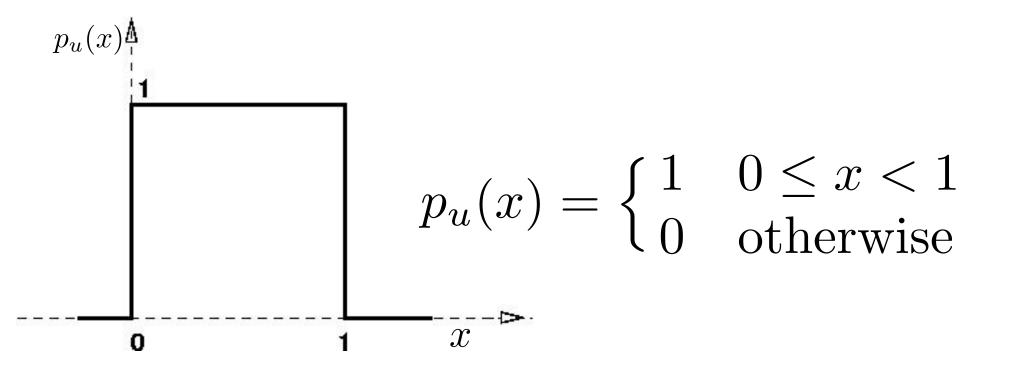
Test of uniformity for intrinsic random number generator

 I) Random numbers with non uniform distributions and
 II) random processes

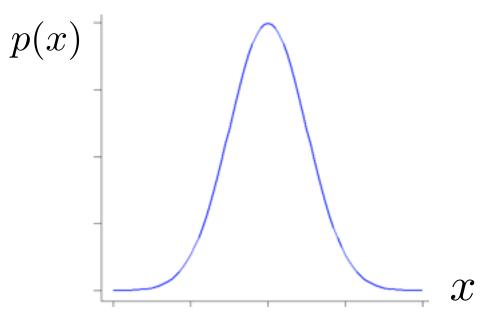
> M. Peressi - UniTS - Laurea Magistrale in Physics Laboratory of Computational Physics - Unit III

last lecture:

generation of real (pseudo)random numbers with uniform distribution in [0;1[

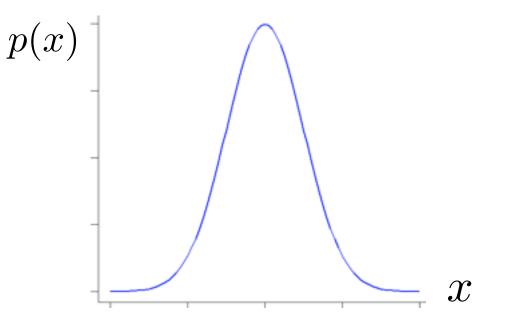


Part I - Random numbers with non uniform distributions:



How can we generate random numbers with a given distribution p(x)?

Part I - Random numbers with non uniform distributions:



 inverse transformation method (general)
 rejection method (general)
 some "ad hoc" methods: the Box-Muller algorithm for the gaussian distribution Non uniform random numbers distribution: 1) inverse transformation method (general)

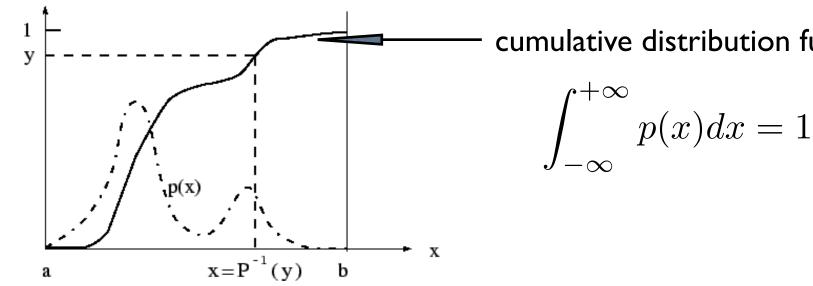
Problem: Generate sample of a random variable (or *variate*) x with a given distribution p.

Solution: 2-step process

 \bullet Generate a random variate uniformly distributed in [0,1] .. also called a $\mathit{random\ number}$

• Use an appropriate transformation to convert the random number to a random variate of the correct distribution

Non uniform random numbers distribution: I) inverse transformation method - algorithm P(x)



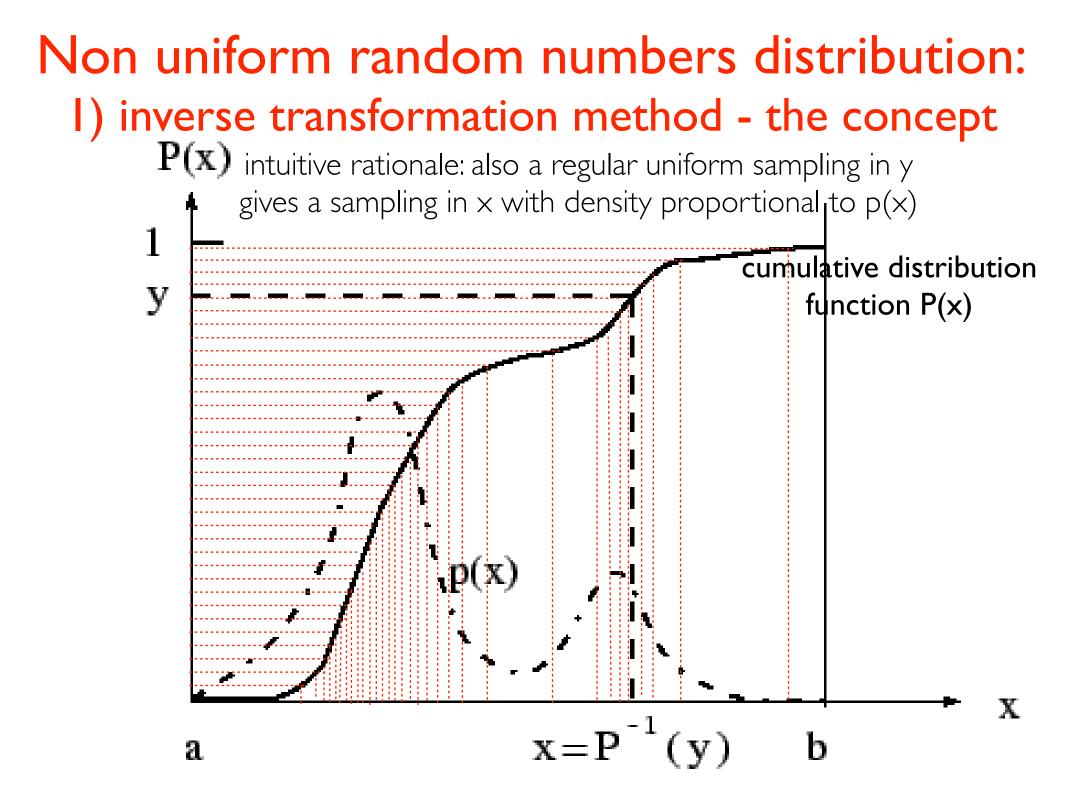
cumulative distribution function P(x)

Let p(x) be a desired distribution, and $y = P(x) = \int p(x')dx'$ the corresponding *cumulative distribution*. Assume that $P^{-1}(y)$ is known.

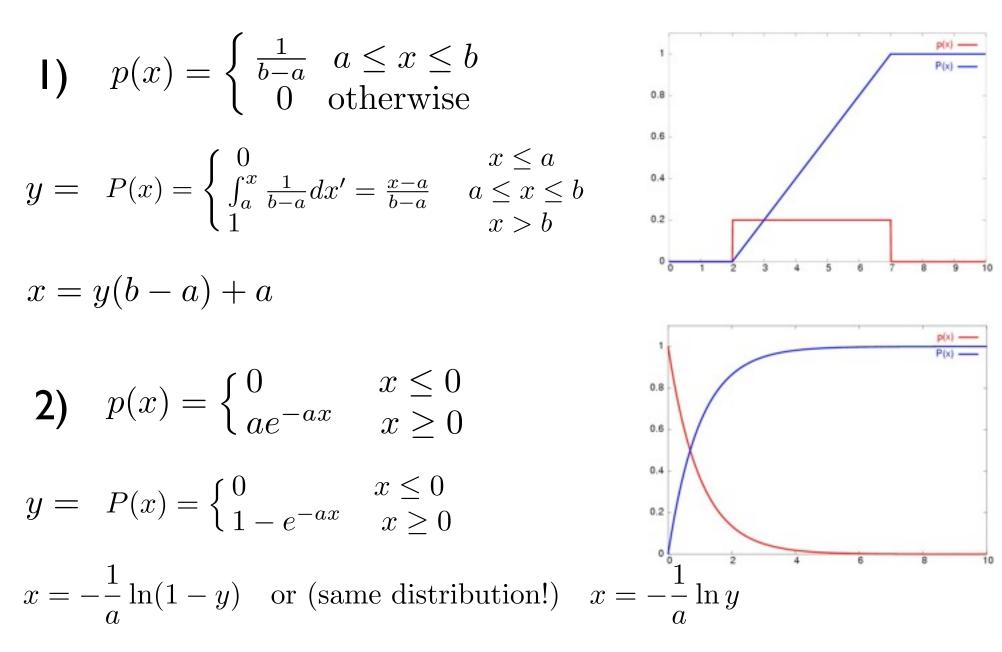
- Sample y from an equidistribution in the interval (0,1). (i.e., use $p_u(y)$)
- Compute $x = P^{-1}(y)$.

The variable x then has the desired probability density p(x).

 $y = P(x) \Longrightarrow dy = dP(x) \Longrightarrow p_u(y)dy = dP(x)$ (since $p_u(y) = 1$ for $0 \le y \le 1$) But : dP(x) = p(x)dx, therefore $p(x)dx = p_u(y)dy$



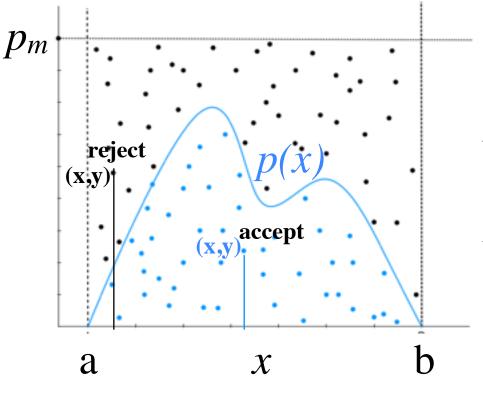
Non uniform random numbers distribution: 1) inverse transformation method - examples



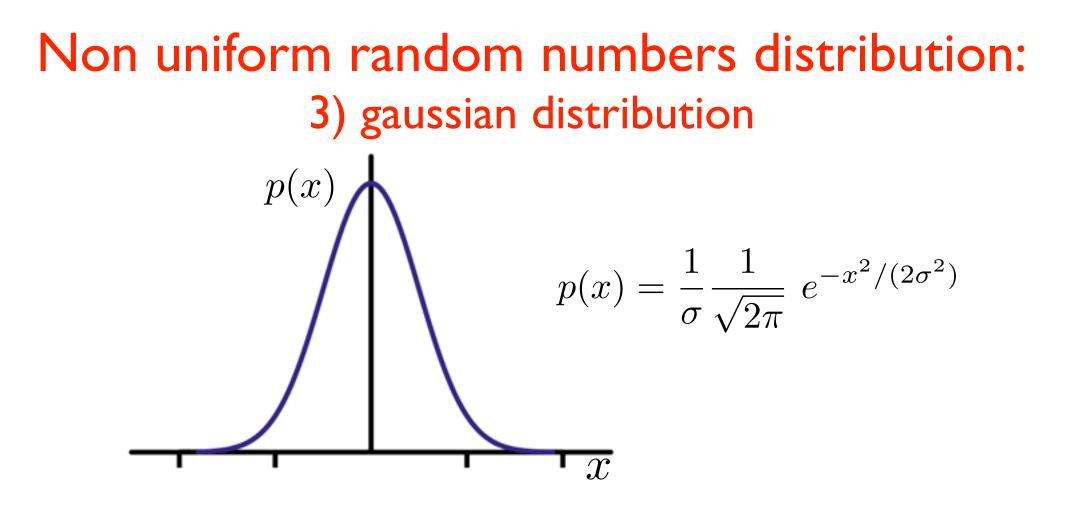
Non uniform random numbers distribution: 2) rejection method (general)

Let [a, b] be the allowed range of values of the variate x, and p_m the maximum of the distribution p(x).

- 1. Sample a pair of equidistributed random numbers, $x \in [a, b]$ and $y \in [0, p_m]$.
- 2. If $y \le p(x)$, accept x as the next random number, otherwise return to step 1.



Due to Von Newmann (1947). Applicable to almost all distributions. Can be inefficient if the area of the rectangle $[a,b] \otimes [0,p_m]$ is large compared to the area below the curve p(x)



How to produce numbers with gaussian distribution?

- Inverse transformation method: impossible The cumulative distribution function P(x) cannot be analytically calculated!
- Rejection method: inefficient

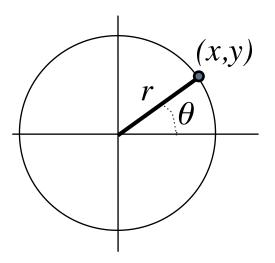
Hint: consider the distribution in 2D instead of ID (here $\sigma = I$):

$$p(x)p(y)dxdy = (2\pi)^{-1} e^{-(x^2+y^2)/2} dxdy$$

Hint: consider the distribution in 2D instead of ID (here $\sigma = I$):

$$p(x)p(y)dxdy = (2\pi)^{-1} e^{-(x^2+y^2)/2} dxdy$$

Use polar coordinates: $r=\sqrt{x^2+y^2}$, $heta=rctan\,(y/x)$; def.: $ho\equiv r^2/2$



Hint: consider the distribution in 2D instead of ID (here $\sigma = I$):

$$p(x)p(y)dxdy = (2\pi)^{-1} e^{-(x^2+y^2)/2} dxdy$$

Use polar coordinates: $r=\sqrt{x^2+y^2}$, $heta=rctan\,(y/x)$; def.: $ho\equiv r^2/2$

$$\rightarrow dxdy = r \ dr \ d\theta = d\rho \ d\theta$$

and therefore:

$$p(x)p(y) \ dx \ dy = p(\rho,\theta) \ d\rho \ d\theta = (2\pi)^{-1} \ e^{-\rho} \ d\rho \ d\theta$$

Hint: consider the distribution in 2D instead of ID (here $\sigma = I$):

$$p(x)p(y)dxdy = (2\pi)^{-1} e^{-(x^2+y^2)/2} dxdy$$

Use polar coordinates: $r=\sqrt{x^2+y^2}$, $heta=rctan\,(y/x)$; def.: $ho\equiv r^2/2$

$$\rightarrow dxdy = r \ dr \ d\theta = d\rho \ d\theta$$

and therefore:

$$p(x)p(y) \ dx \ dy = p(\rho,\theta) \ d\rho \ d\theta = (2\pi)^{-1} e^{-\rho} \ d\rho \ d\theta$$

 $f \quad \begin{cases} \rho \text{ exponentially distributed} \\ \theta \text{ uniformly distributed in}[0, 2\pi] \end{cases} \rightarrow \begin{cases} x = r \cos \theta = \sqrt{2\rho} \cos \theta \\ y = r \sin \theta = \sqrt{2\rho} \sin \theta \\ x, y \text{ have gaussian distribution} \\ \text{with } \langle x \rangle = \langle y \rangle = 0 \text{ and } \sigma = 1 \end{cases}$

Non uniform random numbers distribution: 3) gaussian distribution - Box-Muller recipe #1

If $\begin{cases} \rho \text{ exponentially distributed} \\ \theta \text{ uniformly distributed in}[0, 2\pi] \end{cases} \rightarrow \begin{cases} x = r \cos \theta = \sqrt{2\rho} \cos \theta \\ y = r \sin \theta = \sqrt{2\rho} \sin \theta \\ x, y \text{ have gaussian distribution} \\ \text{with } \langle x \rangle = \langle y \rangle = 0 \text{ and } \sigma = 1 \end{cases}$

Recipe #1 (BASIC FORM):

$$\begin{cases} X, Y \text{ unif. distrib. in } [0, 1[\\ \rho = -\ln(X) \text{ distributed with } p(\rho) = e^{-\rho} \\ \theta = 2\pi Y \text{ distributed with } (2\pi)^{-1} p_u \end{cases} \begin{cases} x = r \cos \theta = \sqrt{-2\ln X} \cos(2\pi Y) \\ y = r \sin \theta = \sqrt{-2\ln X} \sin(2\pi Y) \end{cases}$$

NOTE:

x, y are normally distributed and statistically independent. Gaussian variates with given variances σ_x , σ_v are obtained by multiplying x and y by σ_x and σ_v respectively

Non uniform random numbers distribution: 3) gaussian distribution - Box-Muller recipe #2

If $\begin{cases} \rho \text{ exponentially distributed} \\ \theta \text{ uniformly distributed in}[0, 2\pi] \end{cases} \rightarrow \begin{cases} x = r \cos \theta = \sqrt{2\rho} \cos \theta \\ y = r \sin \theta = \sqrt{2\rho} \sin \theta \\ x, y \text{ have gaussian distribution} \\ \text{with } \langle x \rangle = \langle y \rangle = 0 \text{ and } \sigma = 1 \end{cases}$

Recipe #2 (POLAR FORM) (implemented in **boxmuller.f90**) :

 $x = \sqrt{-2\ln R^2} \ \frac{X}{R}$ X, Y uniformly distributed in [-1,1]; take (X, Y) only within the unitary circle; $\Rightarrow R^2 = X^2 + Y^2$ is $y = \sqrt{-2\ln R^2} \ \frac{Y}{R}$ **uniformly** distributed in [0,1]since: $\cos \theta = \frac{X}{R}, \quad \sin \theta = \frac{Y}{R}$ Advantages: avoids the calculations of sin and cos functions

Some programs:

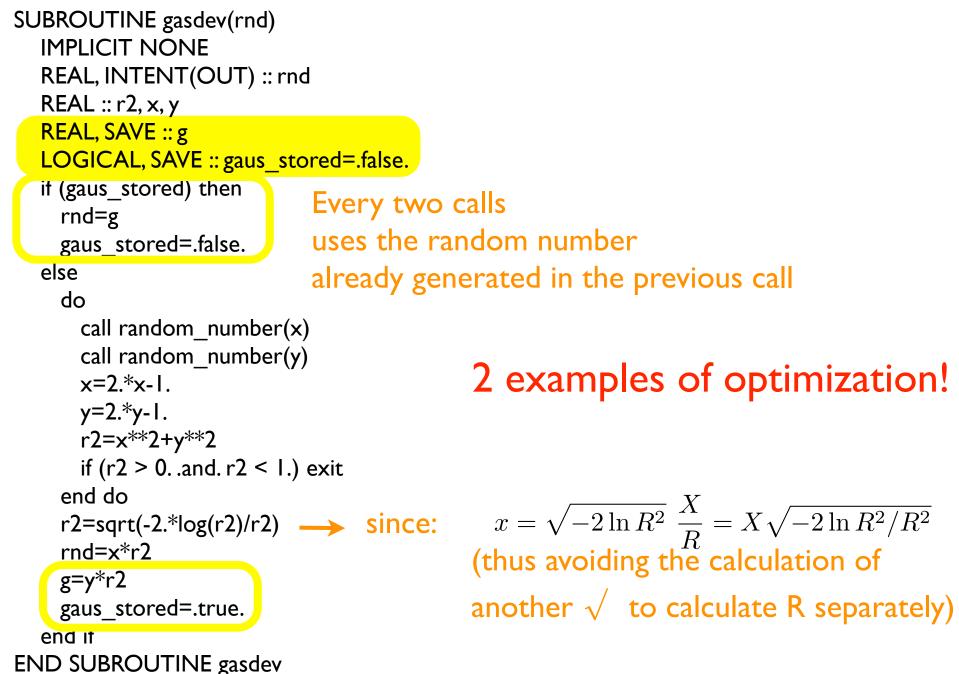
on moodle2 or on INFIS account: \$/home/peressi/comp-phys/III-random-nonuniform-and-processes/f90 [do: \$cp /home/peressi/... .../f90/* .]

expdev.f90 boxmuller.f90

A look at the expdev.f90 code

```
subroutine expdev(x)
   REAL, intent (out) :: x
   REAL :: r
  do
     call random_number(r)
     <mark>i</mark>f(r > 0) exit
                               r is generated in [0,1[;
   end do
                                but r=0 has to be discarded;
                               if r=0, generate another random number;
   x = -\log(r)
                               if not, exit from the unbounded loop
 END subroutine expdev
                               and calculate its log
```

A look at the boxmuller.f90 code



A look at the gasdev.c code

```
#include <math.h>
float gasdev(long *idum)
{
    float ran1(long *idum);
                                Every two calls
    static int iset=0;
    static double gset;
                                uses the random number
    double fac,rsq,v1,v2;
                                already generated in the previous call
    if (iset == 0) {
        do {
                                         2 examples of optimization!
             v1=2.0*ran1(idum)-1.0;
             v2=2.0*ran1(idum)-1.0;
             rsq=v1*v1+v2*v2;
        } while (rsq >= 1.0 || rsq == 0.0);
        fac=sqrt(-2.0*log(rsq)/rsq); \longrightarrow since: x = \sqrt{-2 \ln R^2} \frac{X}{R} = X \sqrt{-2 \ln R^2/R^2}
        gset=v1*fac;
                                           (thus avoiding the calculation of
        iset=1;
        return (float)(v2*fac);
                                           another \sqrt{} to calculate R separately)
    } else {
        iset=0;
        return (float)gset;
    }
}
```

Other programs:

in the same directories indicated before:

(optional, but useful!)

random.f90 (is a module) t_random.f90

to compile: \$gfortran random.f90 t_random.f90 (the module first!) Part II -Using random numbers to simulate random processes

Random processes: radioactive decay

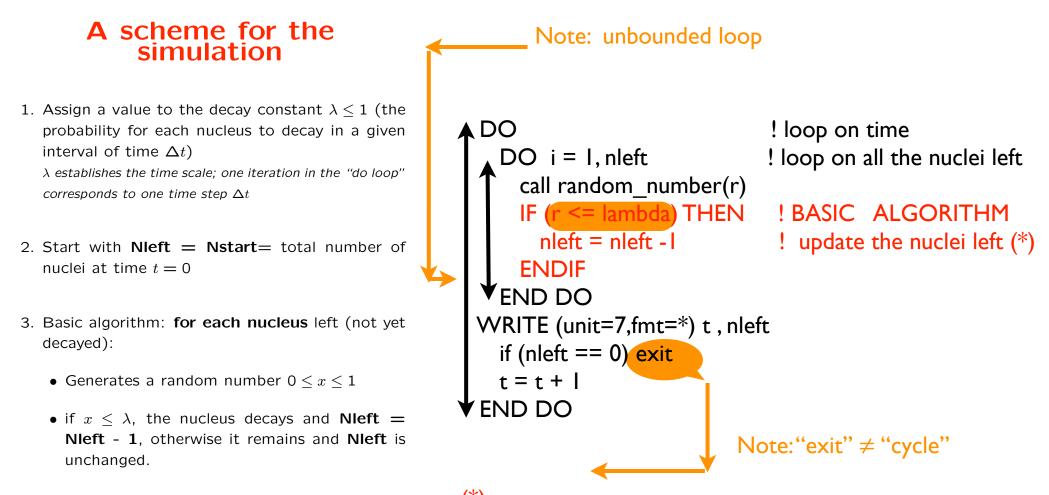
N(t) Atoms present at time t λ Probability for each atom to decay in Δt $\Delta N(t)$ Atoms which decay between t and $t+\Delta t$

$$\Delta N(t) = -\lambda N(t)\Delta t$$

we use the probability λ of decay of each atom to simulate the behavior of the number of atoms left; we should be able to obtain (on average):

$$N(t) = N(t=0)e^{-\lambda t}$$

Radioactive decay: numerical simulation



- 4. Repeat for each nucleus
- 5. Repeat the cycle for the next time step

(*) Notice that the upper bound of the inner loop (nleft) is changed within the execution of the loop; but with most compilers, in the execution the loop goes on up to the initial value of nleft; this ensures that the implementation of the algorithm is correct. The program checkloop.f90 is a test for the behavior of the loop. Look also at decay_checkloop.f90. If nleft would be changed (decreased) during the execution, the effect would be an overestimate of the decay rate. CHECK with your compiler!

Programs:

in the same directory indicated before:

decay_f90 decay_checkloop.f90

checkloop.f90

Details on Fortran: unbounded loops [name:] DO

exit [name]

or [name:] DO END DO [name]

(name is useful in case of nested loops for explicitly indicating from which loop to exit)

possible forms of "do while":

DO if (condition)exit END DO

or:

DOWHILE (.not. condition)

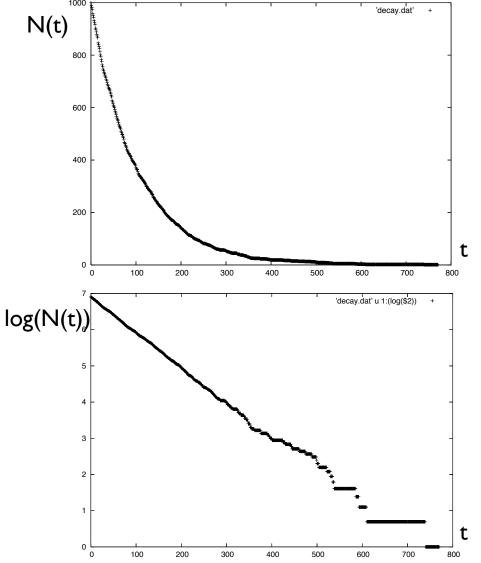
•••

END DO

NOTE: first is better ("if () ..exit" can be placed everywhere in the loop, whereas DO WHILE must execute the loop up to the end)

- Additional note: Difference between EXIT and CYCLE

Radioactive decay: results of numerical simulation

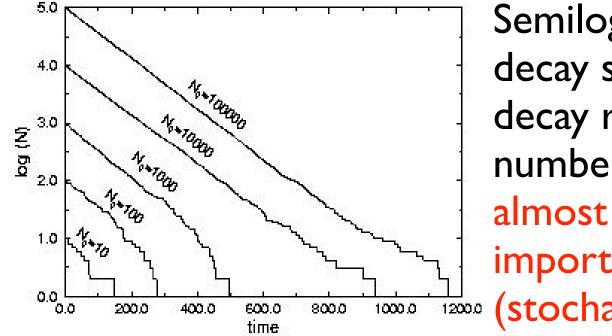


plot of the results of decay simulation (N vs t) with N=1000

 $N(t) \sim N_0 \exp(-a t)$

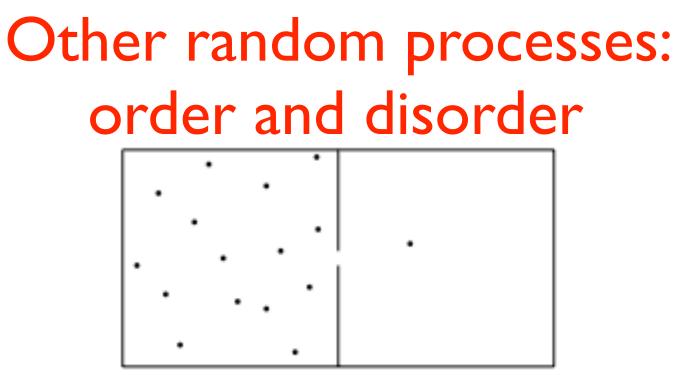
semilog plot (log(N) vs t)
=> log(N(t)) = log N₀ - a t
=> slope is -a

Radioactive decay: results of numerical simulation



Semilog plot of the results of decay simulation for the same decay rate and different initial number of atoms: almost a straight line, but with important deviations

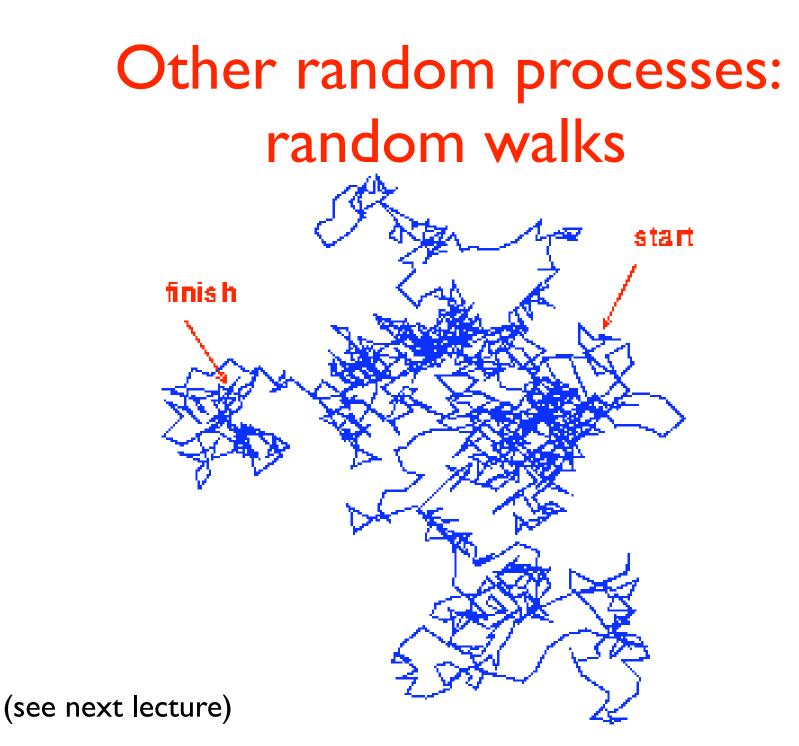
numerical simulations: OK on average and for large numbers



A box is divided into two parts communicating through a small hole. One particle randomly can pass through the hole per unit time, from the left to the right or viceversa.

 $N_{left}(t)$: number of particles present at time t in the left side Given $N_{left}(0)$, what is $N_{left}(t)$?

(see later, lectures on the statistical ensembles)



Part III - Fitting data

Least-square method

Suppose to have N_D data (independent measurements of the variable y which is function of the variable x):

$$(x_i, y_i \pm \sigma_i), \quad i = 1, N_D$$

with $\pm \sigma_i$ error associated to the *i* value of *y*.

 Purpose: determine the function y = f(x) which better described these data. If the analytic form of the function is known, a part from a set M_P of parameters {a₁, a₂..., a_{M_P}}, i.e., f(x) = f(x; {a_m}), the goal is to find the best set of parameters. To test whether the data *fit* via *f(x)* is good or not calculate the quantity

$$\chi^{2} := \sum_{i=1}^{N_{D}} \left(\frac{y_{i} - f(x_{i}; \{a_{m}\})}{\sigma_{i}} \right)^{2}$$

Note that by dividing by σ_i , data with larger errors have smaller weight in this weighted average.

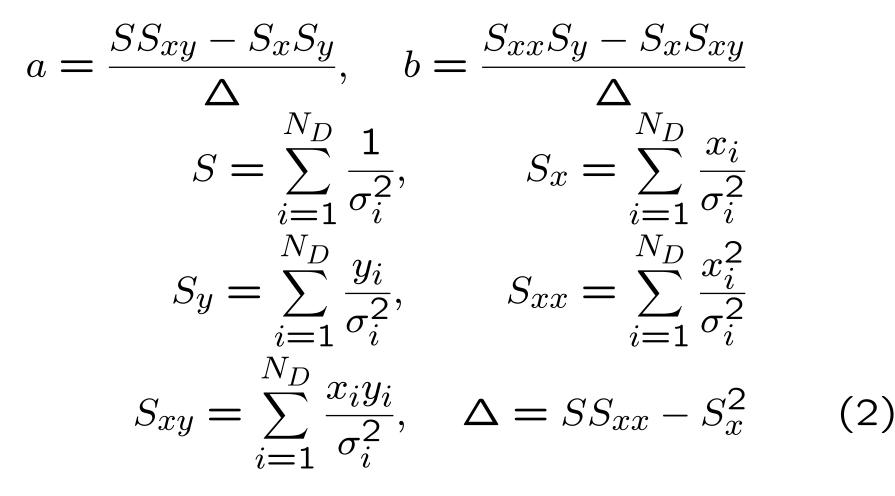
• The smallest χ^2 , the better the fit is.

• Least-squares fitting: The parameters M_P that minimize χ^2 are found by:

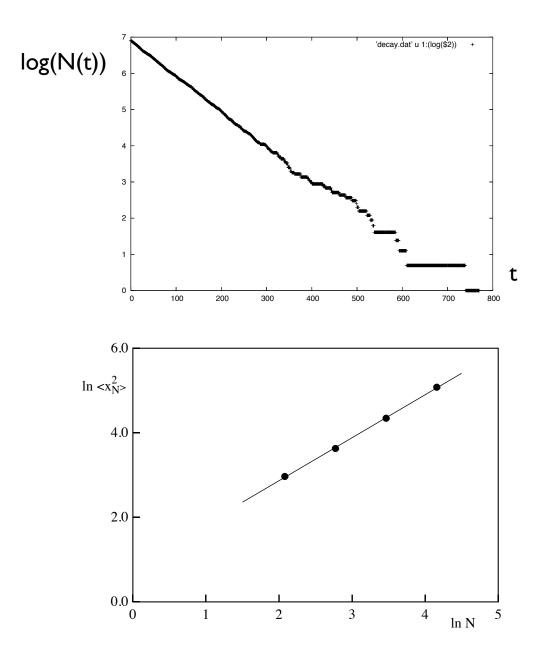
$$\frac{\partial \chi^2}{\partial a_m} = 0 \quad (m = 1, M_P)$$
$$\implies \sum_{i=1}^{N_D} \frac{y_i - f(x_i)}{\sigma_i^2} \frac{\partial f(x; \{a_m\})}{\partial a_m} = 0 \qquad (1)$$

example: see program fit.f90

• If f(x; a, b) = ax + b (linear regression), the equations giving χ^2 minimum reduce to:



Examples - linear regression



radioactive decay: N(t) ~ N₀ exp(- a t)

we can fit with the exp. but it is better to fit:

 $log(N(t)) = log N_0 - a t$

Random walk: <x²_N> ~ N^a

but it is better to fit:

 $\log < x^2_N > = a \log N$

Example: fit using gnuplot - I

Suppose you want to fit your data (say, 'data.dat') with an exponential function. You have to give: 1) the functional form ; 2) the name of the parameters

gnuplot> f(x) = a * exp(-x*b)

Then we have to recall these informations together with the data we want to fit: it can be convenient to inizialize the parameters:

gnuplot> a=0.; b=1. (for example)

gnuplot> fit f(x) 'data.dat' via a,b

On the screen you will have something like:

```
Final set of parameters Asymptotic Standard Error
```

a = 1 +/- 8.276e-08 (8.276e-06%) b = 10 +/- 1.23e-06 (1.23e-05%)

correlation matrix of the fit parameters:

a b a 1.000 b 0.671 1.000

It's convenient to plot together the original data and the fit:

gnuplot> plot f(x), 'data.dat'

Example: fit using gnuplot - II

If you prefer to use linear regression, **use logarithmic data in the data file**, **or** directly fit the log of the original data using **gnuplot**:

gnuplot> f(x) = a + b*x

. . .

Then we have to recall these informations together with the data we want to fit (in the following example: $x=\log of$ the first column; $y=\log of$ the second column):

gnuplot> fit f(x) 'data.dat' u (log(\$1)):(log(\$2)) via a,b

Also in this case it will be convenient to plot together the original data and the fit:

gnuplot> plot f(x), 'data.dat' u (log(\$1)):(log(\$2))

In case of needs, we can limit the set of data to fit in a certain range **[x_min:x_max]**:

gnuplot> fit [x_min:x_max] f(x) 'data.dat' u ... via ...

Part IV more on fortran

A few notes on Fortran

related to the exercises

Intrinsic functions:

LOGARITHM

log returns the natural logarithm
log 10 returns the common (base 10) logarithm
(NOTE: also in gnuplot, log and log 10 are defined with the same meaning)

INTEGER PART

nint(x) and the others, similar but different (see Lect. II) =>
ex. II requires histogram for negative and positive data values

Arrays:

possible to label the elements from a negative number or 0: **dimension array(-n:m)** (e.g., useful for making histograms) [default in Fortran: n=1; in c and c++: n=0]

what is int() ? similar intrinsic functions? how to choose?

AINT(A[,KIND])

 \cdot Real elemental function

 \cdot Returns A truncated to a whole number. AINT(A) is the largest integer which is smaller than |A|, with the sign of A. For example, AINT(3.7) is 3.0, and AINT(-3.7) is -3.0.

 \cdot Argument A is Real; optional argument KIND is Integer

ANINT(A[,KIND])

 \cdot Real elemental function

 \cdot Returns the nearest whole number to A. For example, ANINT(3.7) is 4.0, and AINT(-3.7) is -4.0.

 \cdot Argument A is Real; optional argument KIND is Integer

FLOOR(A,KIND)

- Integer elemental function
- Returns the largest integer \leq A. For example, FLOOR(3.7) is 3, and FLOOR(-3.7) is -4.
- Argument A is Real of any kind; optional argument KIND is Integer
- Argument KIND is only available in Fortran 95

INT(A[,KIND])

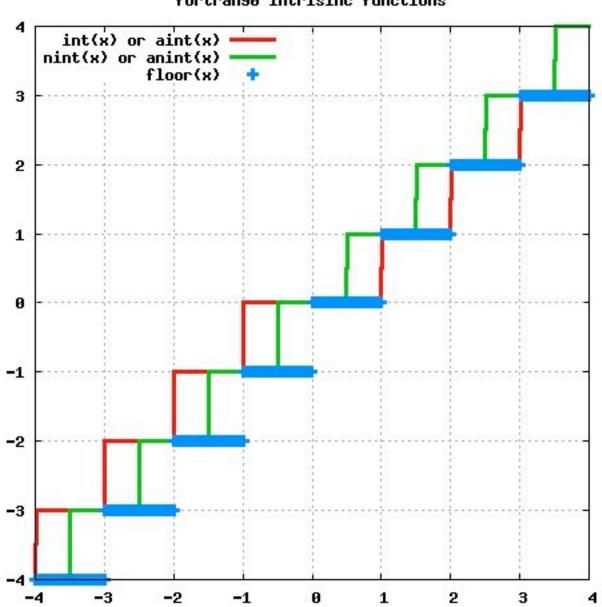
Integer elemental function

• This function truncates A and converts it into an integer. If A is complex, only the real part is converted. If A is integer, this function changes the kind only.

• A is numeric; optional argument KIND is Integer.

NINT(A[,KIND])

- Integer elemental function
- Returns the nearest integer to the real value A.
- A is Real



fortran90 intrisinc functions

Array dimension:

default : dimension array([1:]n)
but also using other dimensions e.g.: dimension array(-n:m)

Important to **check dimensions** of the array when compiling or during execution ! If not done, it is difficult to interpret error messages (typically: "segmentation fault"), or even possible to obtain unpredictable results!

Default in g95 and gfortran: boundaries not checked; use compiler option:

gfortran -fbounds-check myprogram.f90

Print:

man gfortran

and scroll the pages to see the possible options of compilation

Structure of a main program with one function

program name_program implicit none (*) <declaration of variables> <executable statements> (see: expdev.f90 or boxmuller.f90)

```
contains
```

```
subroutine ... (or function)
...
end subroutine
```

end program

(*) General suggestion for variable declaration: Use "implicit none" + explicit declaration of variables

See also the use of **module** in Lect. II and III.