VI - Random numbers

1. Introduction

(*) Problem: Generate a set of random numbers in a computer

(*) Applications: Simulation of stochastic processes, Monte-Carlo integration

2 - Uniform deviates

2.1. Generalities

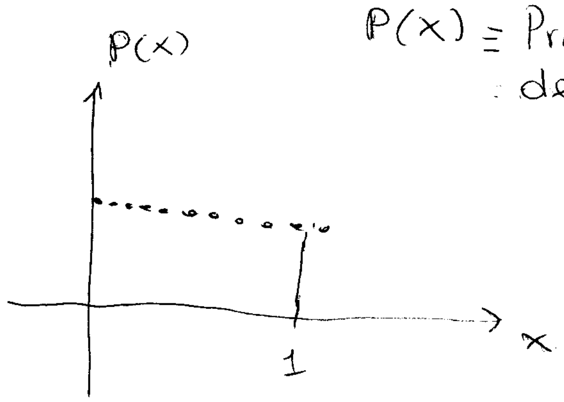
The probability of generating a number between x and $x+dx$ is

$$P(x)\,dx = \begin{cases} dx & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\int_{-\infty}^{\infty} P(x)\,dx = 1$$

In practice : x (random # $\in \underbrace{[0,1]}_{\text{"default interval"}}$

$P(x) \equiv$ Probability density



2.2- System-supplied generators (linear congruential generators)

Generate a PERIODIC sequence of integers $I_1, I_2, I_3 \dots$, each between 0 and $m-1$, so that

$$I_{j+1} = \underbrace{a}_{\text{multiplier}} I_j + \underbrace{c}_{\text{increment}} \underbrace{(\text{mod } m)}_{\text{"modulo" } m}$$

$10 \text{ mod } 3 = 1$

$15 \text{ mod } 5 = 0$

$\Rightarrow$ Output: $\frac{I_i}{m}$, $i = 0, 1, \dots, m$

- Advantages: very fast
- Drawback: They are not random (they are called "pseudo-random numbers")

Examples: x = ran(iseed) (FORTRAN)

Random[ ] (Mathematica)

x = rand (MatLab)

x = rand(n) (MatLab) $\Rightarrow$ returns an $n \times n$ matrix with such numbers

"iseed" $\Rightarrow$ gives the initial value I_1 of the sequence

(*) In practice:

- The "seed" is chosen so that m is as large as possible
- One tries to break the above correlations (reference: Numerical Recipes; see web list)

(*) How to go beyond uniform deviates?

3 - Transformation method

3.1 - Generalities (non-uniform deviates)

- Step 1: One generates a uniform deviate x
- Step 2: One takes a prescribed function $y(x)$

$$\underbrace{|P(y)\,dy|}_{\substack{\text{Probability}\\\text{distribution}\\\text{of } y}} = \underbrace{|p(x)\,dx|}_{\text{probability distr. of } x} \Rightarrow P(y) = p(x)\left|\frac{dx}{dy}\right|$$

3.2 - Exponential deviates

$$y(x) = -\ln x \quad \Rightarrow \quad x(y) = e^{-y} \quad \Rightarrow \quad \left|\frac{dx}{dy}\right| = e^{-y}$$

$$p(y)\,dy = \underbrace{p(x)}_{\substack{\shortparallel\\ 1 \text{ (uniform deviate)}}}\left|\frac{dx}{dy}\right| dy \quad \Rightarrow \quad \boxed{p(y)\,dy = e^{-y}\,dy}$$

Procedure:

- Step 1: generate a uniformly distributed random number : $x = \text{ran(iseed)}$
- Step 2: take $y(x) = -\ln(x)$
  (y has an exponential deviate)

(For a sequence of random numbers: one puts this into a do loop, etc.)

(*) Please note: This procedure can be, in principle, generalized to any function.
(One takes $y(x) = F^{-1}(x)$ with

$$x = F(y) = \int_0^y P(y)\,dy \;)$$

3.3- Normal (Gaussian) Deviates

More than 1D: $x_1, x_2, \cdots x_n$ are random deviates with a joint probability distribution $P(x_1, x_2, \ldots)$

$$p(y_1, y_2, \ldots)\,dy_1\,dy_2 \ldots = p(x_1, x_2, \ldots)\underbrace{\left|\frac{\partial(x_1, x_2, \ldots)}{\partial(y_1, y_2, \ldots)}\right|}_{\text{Jacobian determinant}} dy_1\,dy_2 \ldots$$

(*) Application: Box-Muller method

Generates random numbers with a normal (Gaussian) distribution

1D : $p(y)\,dy = \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$

2D : $P(y_1, y_2)\, dy_1\, dy_2 = \frac{1}{2\pi} e^{-\frac{(y_1^2+y_2^2)}{2}} dy_1 dy_2$

Procedure:

- One picks out two uniform deviates

$$x_1 \in [0,1]$$
$$x_2 \in [0,1]$$

They are related to 2 normally-distributed quantities y_1, y_2 by

$$y_1 = \sqrt{-2 \ln x_1} \cos(2\pi x_2) \qquad (*)$$
$$y_2 = \sqrt{-2 \ln x_1} \sin(2\pi x_2) \qquad (**)$$

and

$$x_1 = \exp\left[-\frac{1}{2}(y_1^2 + y_2^2)\right]$$
$$x_2 = \frac{1}{2\pi} \text{ArcTan}\left[\frac{y_2}{y_1}\right]$$

Check: $(*)^2 + (**)^2 = -2 \ln x_1 \left[\underbrace{\cos^2 2\pi x_2 + \sin^2 2\pi x_2}_{1}\right]$

$$\Rightarrow x_1 = \exp\left[-\frac{1}{2}(y_1^2 + y_2^2)\right]$$

$$\frac{(**)}{(*)} = \tan(2\pi x_2) = \frac{y_2}{y_1}$$

Jacobian:

$$\frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} = \frac{\partial x_1}{\partial y_1} \cdot \frac{\partial x_2}{\partial y_2} - \frac{\partial x_2}{\partial y_1} \cdot \frac{\partial x_1}{\partial y_2}$$

$$\frac{\partial x_1}{\partial y_1} = -y_1 \exp\left[-\frac{1}{2}(y_1^2 + y_2^2)\right]$$

$$\frac{\partial x_2}{\partial y_2} = \frac{1}{2\pi} \frac{1}{y_1} \frac{1}{1 + \left[\frac{y_2}{y_1}\right]^2}$$

$$\frac{\partial x_2}{\partial y_1} = \frac{1}{2\pi}\left(\frac{-y_2}{y_1^2}\right) \frac{1}{1+\left[\frac{y_2^2}{y_1^2}\right]}$$

$$\frac{\partial x_1}{\partial y_2} = -y_2 \exp\left[-\frac{1}{2}(y_2^2+y_1^2)\right]$$

$$\frac{\partial x_1}{\partial y_1}\cdot\frac{\partial x_2}{\partial y_2} = \frac{-1}{2\pi}\left[\frac{1}{\left[1+\frac{y_2^2}{y_1^2}\right]} e^{-\frac{1}{2}(y_1^2+y_2^2)}\right] \quad (*)$$

$$\frac{\partial x_2}{\partial y_1}\cdot\frac{\partial x_1}{\partial y_2} = \frac{1}{(2\pi)}\left[\frac{1}{1+\frac{y_2^2}{y_1^2}}\left(\frac{y_2^2}{y_1^2}\right) e^{-\frac{1}{2}(y_1^2+y_2^2)}\right] \quad (**)$$

$$(*)-(**) = \frac{-1}{2\pi}\left[\frac{1+y_2^2/y_1^2}{1+y_2^2/y_1^2} e^{-\frac{1}{2}(y_1^2+y_2^2)}\right]$$

$$\Rightarrow P(y_1,y_2)\,dy_1\,dy_2 = \underbrace{P(x_1,x_2)}_{=1} \frac{1}{2\pi} e^{-\frac{1}{2}(y_1^2+y_2^2)}\,dy_1\,dy_2$$

(*) One generates two random (uniform deviates) numbers obeying gaussian distributions from two uniformly distributed numbers.

(*) Please note: $P(y_1,y_2) = P(y_1)\,P(y_2)$ with

$$P(y_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_i^2} \quad (i=1,2)$$

(each y_i is independently distributed!)

How to implement this numerically?

Possibility 1: • generate 2 uniformly distributed random numbers

$x_1 = \text{ran(seed)}$

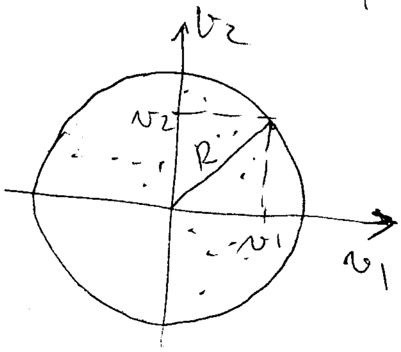
$x_2 = \text{ran(seed)}$

• If $x_1 \neq 0$

$$y_1 = \sqrt{-2\ln x_1}\,\cos 2\pi x_2$$

$$y_2 = \sqrt{-2\ln x_1}\,\cos 2\pi x_2$$

Possibility 2 : we pickout the ordinate/ abscissa of a random point inside a unit circle



$R^2 = v_1^2 + v_2^2$ is a uniform deviate (can be used for x_1).

$\theta = \text{Arc Tan}\left(\frac{v_2}{v_1}\right) = 2\pi x_2$

Advantage :

$$\left.\begin{aligned} \cos(2\pi x_2) &= \frac{v_1}{R} \\ \sin(2\pi x_2) &= \frac{v_2}{R} \end{aligned}\right\} \text{ no need for trigonometric functions}$$

- Algorithm

Input : seed
Output: Gaussian deviates

Step 1: Generates two uniformly distributed random numbers in the square from -1 to 1, in each direction

$v_1 = 2\,\text{ran(seed)} - 1$
$v_2 = 2\,\text{ran(seed)} - 1$

Step 2: Compute R :

$R^2 = v_1^2 + v_2^2$

If $R^2 > 1$ or $R^2 = 0$ then repeat step 1
(in this case either both numbers are OUTSIDE the unit circle or R vanishes, which will be problematic when computing the log)

else

$$\left.\begin{aligned} y_1 &= \sqrt{-2\ln(R^2)}\,\frac{v_1}{R} \\ y_2 &= \sqrt{-2\ln(R^2)}\,\frac{v_2}{R} \end{aligned}\right\} \text{ make Box-Muller transformation}$$

(*) Please note: for the transformation method to be applicable, it is necessary that $\int p(x)dx$ be computable and invertible (This is not always possible)

4. Rejection method

(*) **Task**: generate random deviates whose distribution function $p(x)dx$ is known and computable

(This method DOES NOT require that

$P(x) = \int p(x)\,dx$ is readily computable, and that

P^{-1} exists).

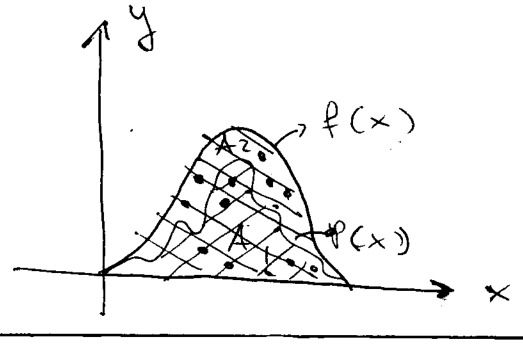
Key idea:

- Step 1: take a function $f(x)$ ("comparison function"), which:
  - i) lies above $g(x)$, $\forall x$
  - ii) encloses a finite area
  - iii) Has a readily computable and invertible integral

$$F(x) = \int f(x)\,dx$$

- Step 2: Generate a random point uniformly distributed under $f(x)$ (x_0, y_0)
- Step 3: Check: point lies "below" $p(x)$ → accept it
  point lies "above" $p(x)$ → reject it

(*) **Graphical representation**:



Example:

Obtain random numbers with

$$p(x) = \begin{cases} 4/[\pi(1+x^2)] & \text{if } 0<x<1 \\ 0 & \text{otherwise} \end{cases}$$

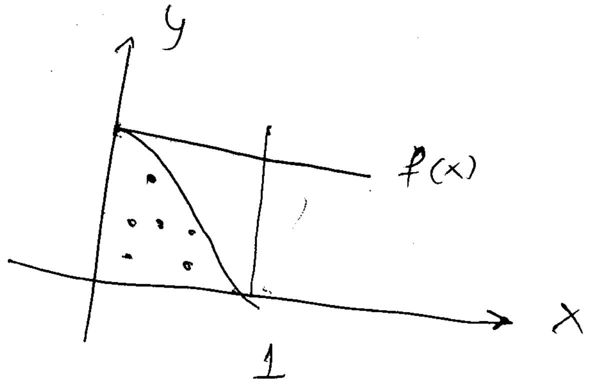
from rectangularly distributed random numbers using the rejection method

- Rectangularly distributed random numbers:

$$f(x) = \frac{4}{\pi} \quad \text{(comparison function)}$$

→ we generate random number between 0 and 1 and accept it with the probability

$$\frac{p(x)}{f(x)} = \frac{1}{1+x^2}$$



5 - Principles of Monte Carlo integration

⊛ **Task**: Perform a multidimensional integral

$$\int_{a_1}^{b_1}\int_{a_2}^{b_2}\int_{a_3}^{b_3}\cdots f(x_1,\dots,x_n)\, dx_1 dx_2 \cdots dx_n$$

⊛ Problems with conventional methods (e.g. quadratures):

- Too large loops need to be performed (inefficient, there may be a large propagation of round-off errors, etc.)
- Many times the volume V ($dV = dx_1 dx_2 \cdots dx_n$) is not easy to sample.

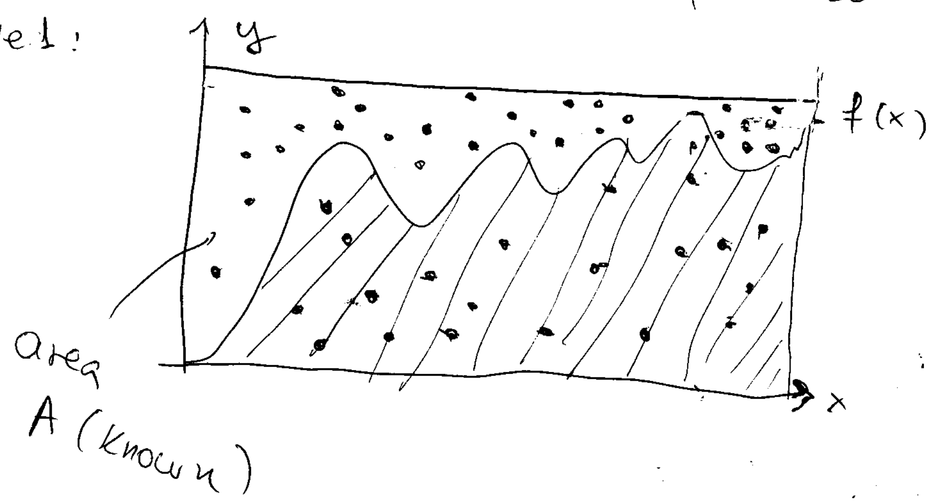
⊛ Procedure:

- Choose a volume W which encloses the volume V
- Generate a set of n random points within this volume (W)

  Check: If a point lies within V, take this into account
  else discard the point / try again

$\therefore \int f(v)\, dV \sim W *$ fraction of random points within f

Example 1:
(10)



Example 2 :

Compute the weight of the intersection of a torus with the edge of a large box

- weight : $\iiint \underbrace{e}_{\substack{\text{density}\\ \text{of the}\\ \text{torus}}} \underbrace{dx\,dy\,dz}_{\text{volume of the torus}}$

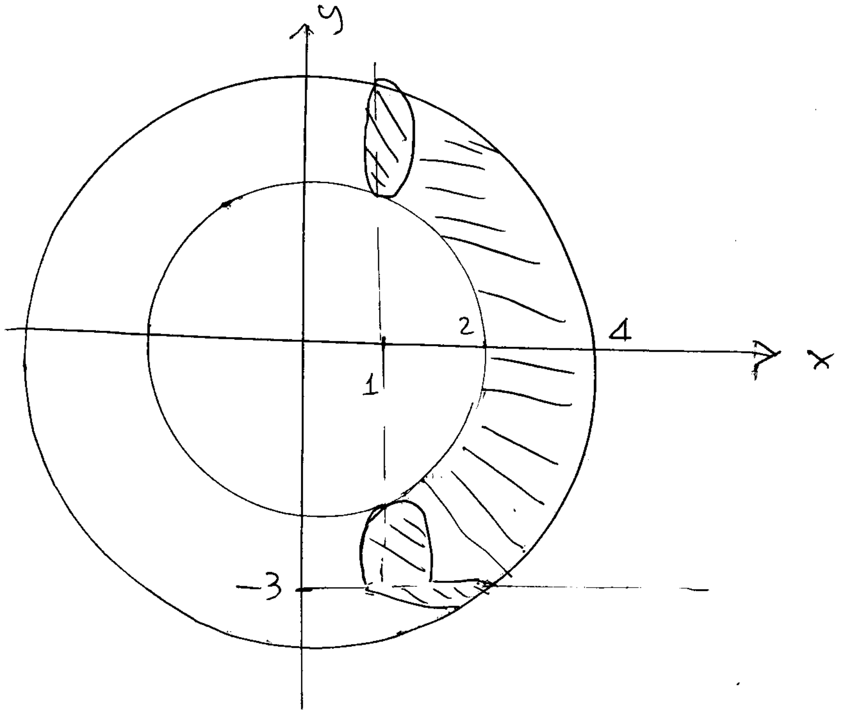
- The object in question is defined by the conditions

(a) $z^2 + \left(\sqrt{x^2+y^2} - 3\right)^2 \leq 1$

(Torus centered at the origin with major radius $=4$; minor radius $= 2$)

(b) $x \geq 1$
(c) $y \geq -3$ } two faces of the box (i.e., two planes "cutting" the torus)

-



$V \Rightarrow$ rectangular box with $1 \leq x \leq 4$

$-3 \leq y \leq 4$

$-1 \leq z \leq 1$

$W \Rightarrow$ Volume of the truncated torus

Flow chart: