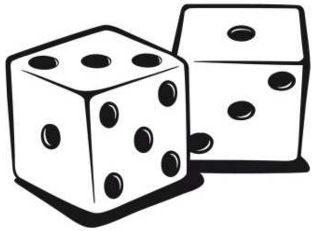
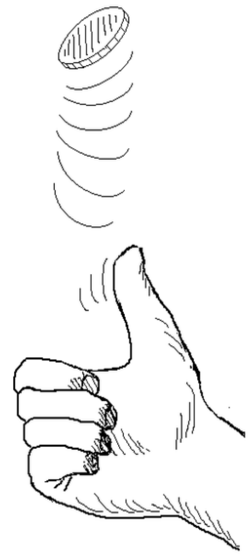
Alea Jacta Est !

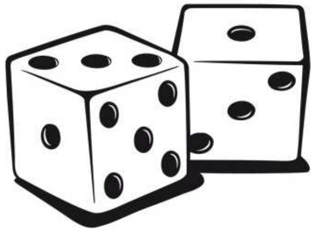


Didier CERTAIN

BREIZH DATA CLUB



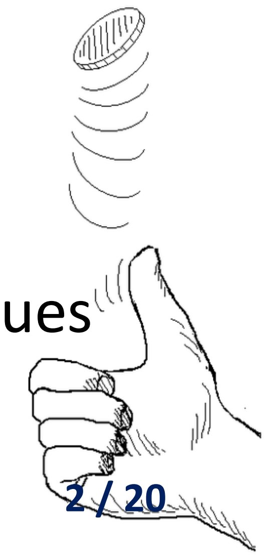
Alea Jacta Est !



Quelques **trucs amusants** avec le **hasard**, les **probabilités** et les **statistiques**.

Les *Big Data* sont, par définition, des données volumineuses qui proviennent de notre environnement usuel : métiers, usages, sciences, réseaux sociaux, etc.

Besoin de données fictives pour faire des tests ou des expériences ?
On peut générer ces données plus ou moins aléatoirement.
Et cela provoque des résultats parfois surprenants :
artistiques, oniriques, littéraires, scientifiques, etc.



Dans ce talk, on va **se promener parmi quelques fantaisies** mathématiques disponibles autour de nous en **informatique**.

Pachinko Flipper & Machine à sous

Pachinko Flipper & Machine à sous

Planche de Galton (*1822 - 1911*)

Quand le nombre de billes devient suffisamment grand, on voit apparaître une distribution gaussienne

À l'envers *À l'endroit*

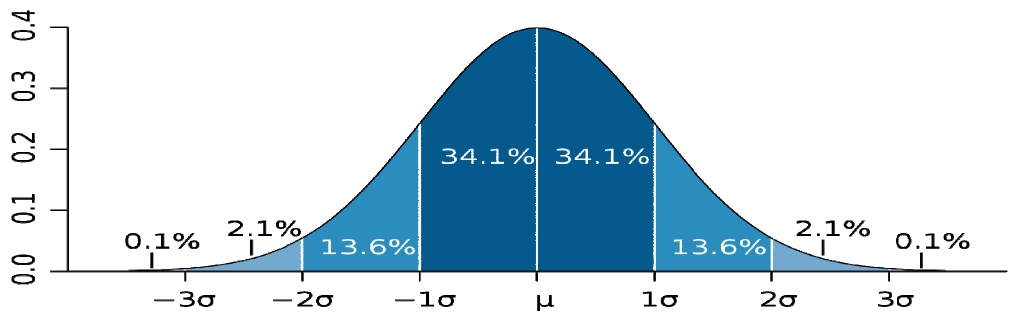
Planche de Galton

Gauss (1777 – 1855)

Quand le nombre d'expériences aléatoires tend vers l'infini,
il y a convergence de la loi binomiale vers la loi normale (distribution gaussienne)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

avec μ l'espérance et σ l'écart-type.



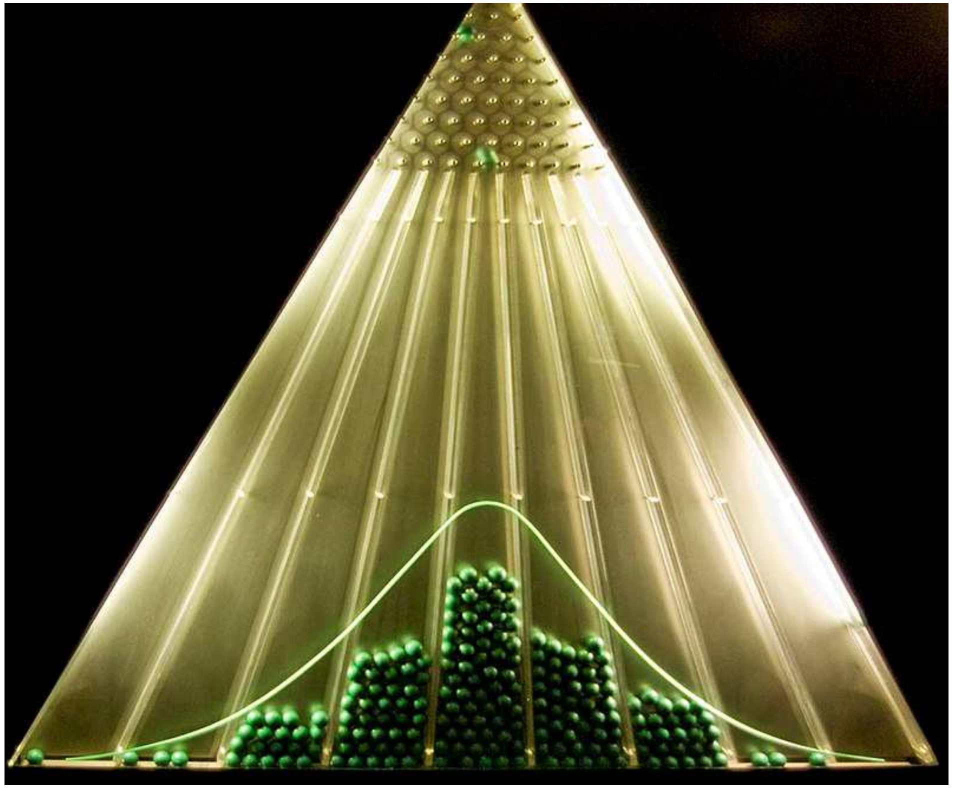
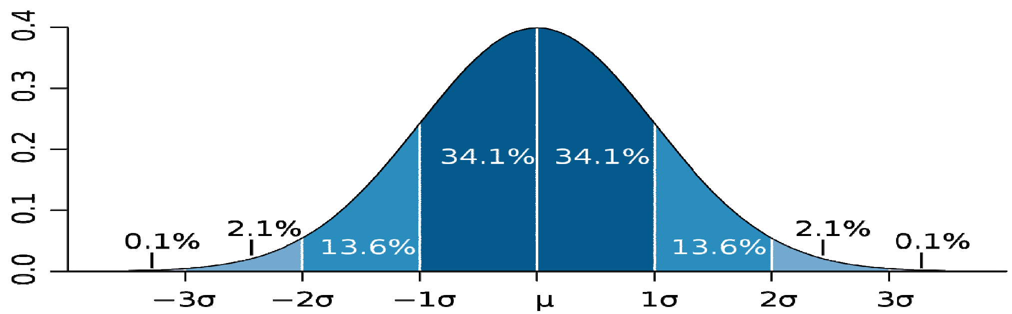
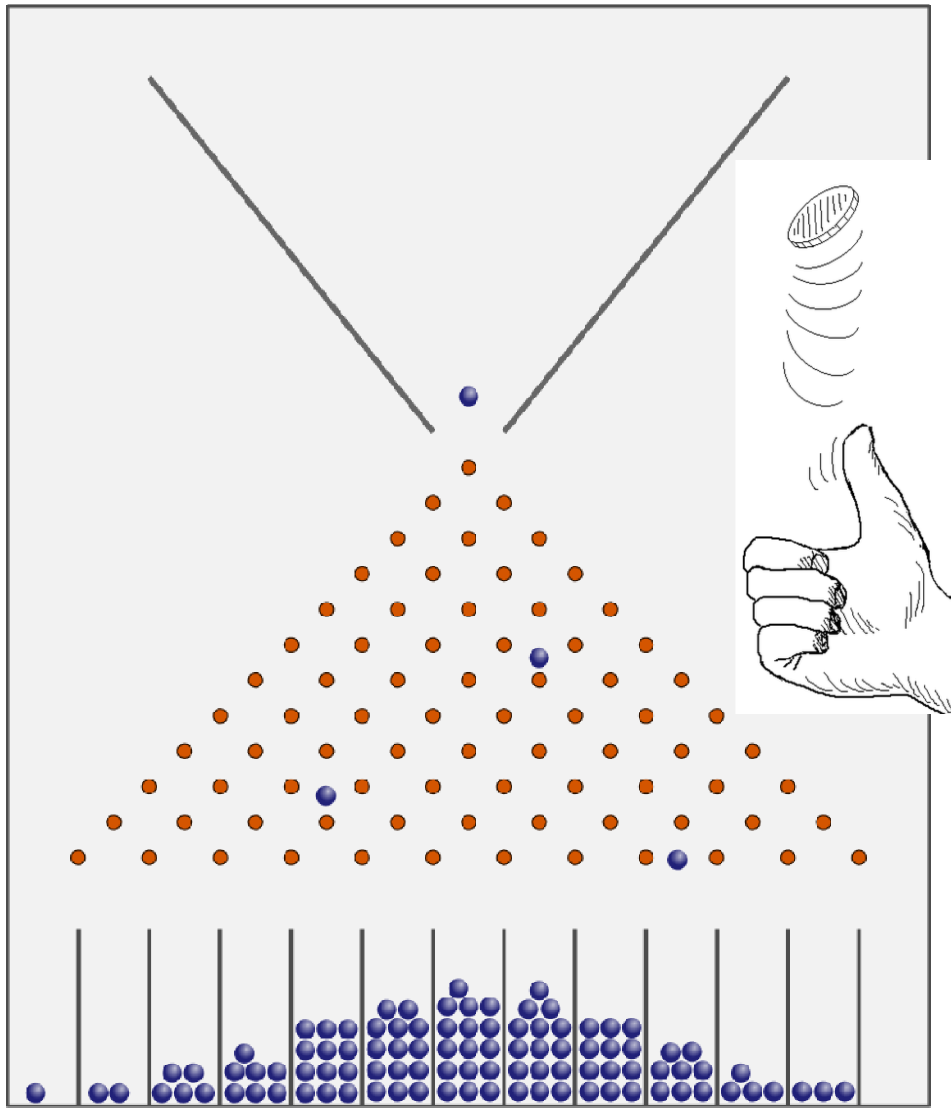


Planche de Galton

Convergence de la loi binomiale
vers la loi normale (distribution gaussienne)
quand le nbre d'expériences tend vers l'infini.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\,\mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

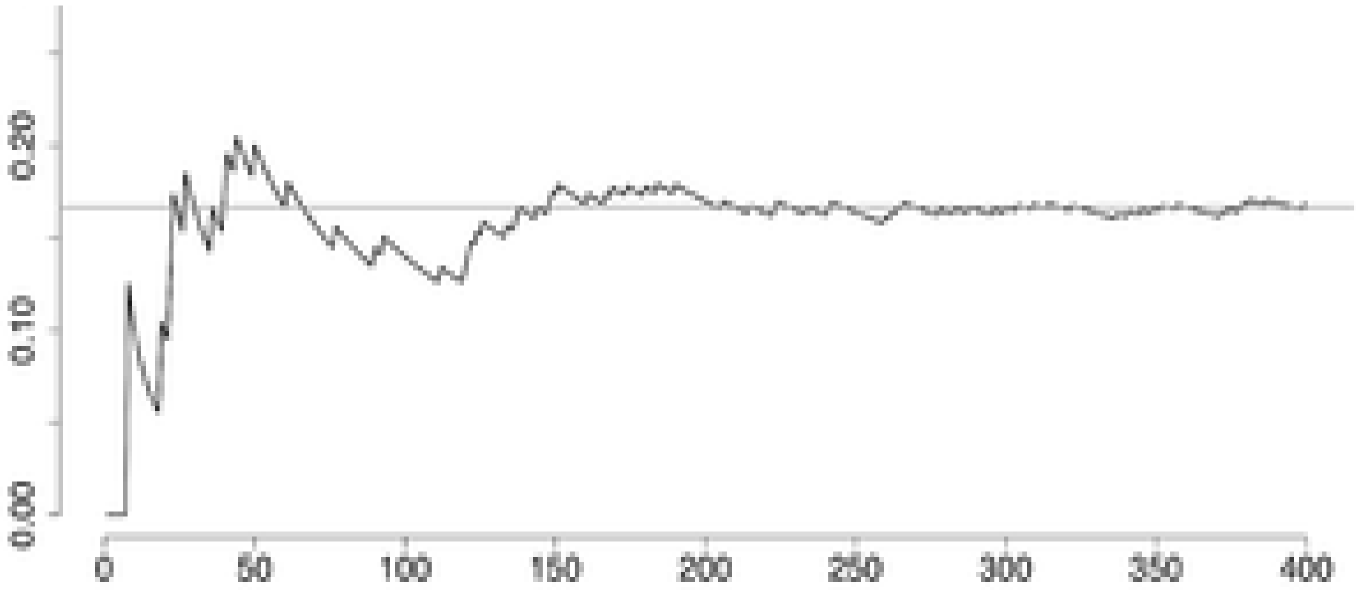




Loi des grands nombres

La moyenne des N premiers termes d’une suite de variables aléatoires converge vers l’espérance probabiliste quand N tend vers l’infini.

Moyenne
empirique



μ (espérance)

Nb. d’itérations

Aiguilles de Buffon (1707 – 1788)

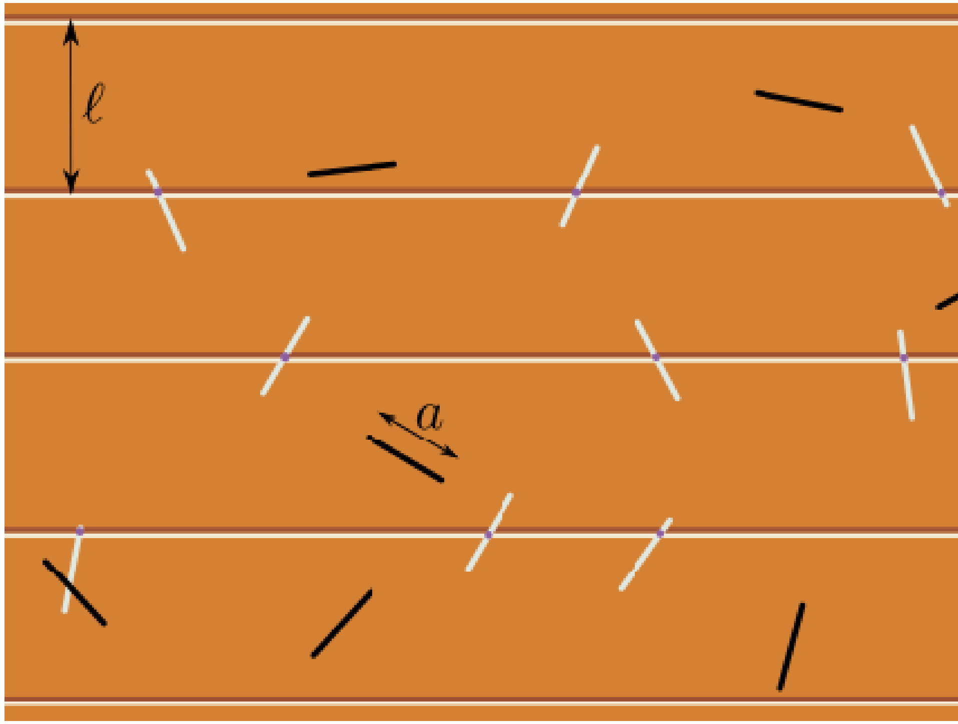
Dans ma chambre, dont le parquet est formé de lames parallèles,
je laisse tomber une aiguille sur le plancher.
Quelle est la probabilité pour que cette aiguille
tombe sur l'une des rainures séparant les lames du parquet ?

- Soit ℓ la largeur d'une lame de parquet
- Soit a la longueur de l'aiguille

alors la probabilité **P** (ℓ , a) est égale à

$$P = 2 . a / \pi . \ell$$

et si $\ell = a$ alors **P** = $2 / \pi$



Aiguilles de Buffon

Le parquet est formé de lames parallèles de largeur ℓ

je laisse tomber une aiguille de longueur a sur le plancher.

La probabilité **P (ℓ , a)** pour que l'aiguille tombe sur l'une des rainures est égale à

$$\mathbf{P} = 2.\, a \,/\, \pi.\, \ell$$

Et donc, après un grand nombre d'essais, je peux calculer la moyenne empirique qui tend vers l'espérance probabiliste **Esp(P)**

Je peux en déduire $\boxed{\pi = 2.\, a \,/\, \mathbf{Esp(\, P\,)} \,.\, \ell}$

Jeu de fléchettes

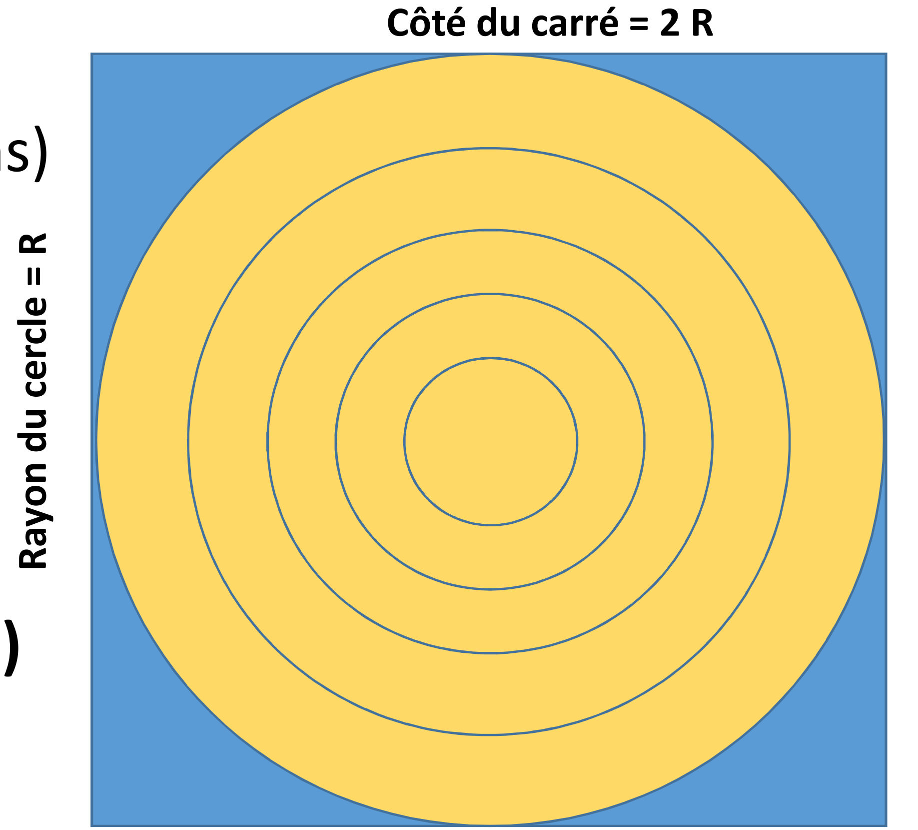
Dans ma chambre, j'ai une cible de rayon **R** inscrite dans un carré de côté **2.R**
je lance des fléchettes de manière aléatoire sur ma cible.

La probabilité **P (R)** pour qu'une fléchette tombe dans la cible (et non pas dans les coins) est égale à

$$\mathbf{P\ (R)} = \pi\ \mathbf{R.R} \ / \ 4\ \mathbf{R.R} = \pi \ / \ 4$$

Et donc, si je fais de nombreux tirs,
je peux calculer la moyenne empirique
qui tend vers l'espérance probabiliste **Esp(P)**

Je peux en déduire $\pi = 4.\ \mathbf{Esp(\ P\ )}$



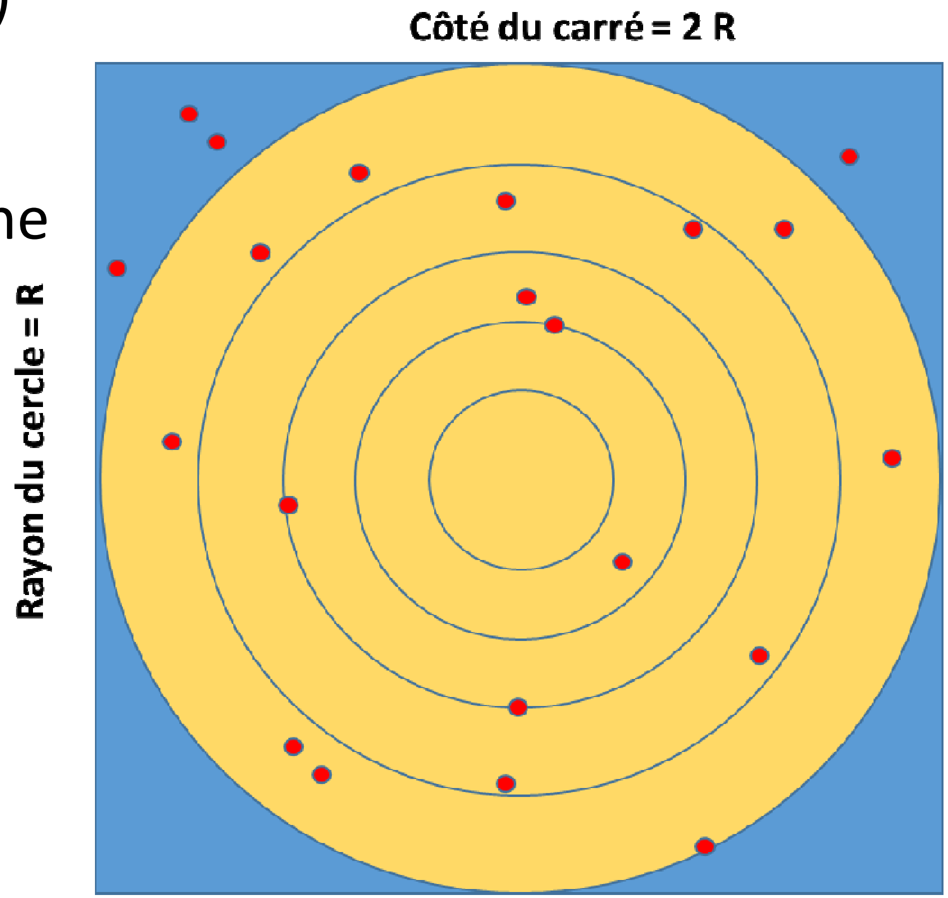
Jeu de fléchettes

La probabilité **P (R)** pour qu'une fléchette tombe dans la cible (et non pas dans les coins) converge vers **π / 4**

En répétant les tirs, je peux calculer la moyenne empirique qui tend vers l'espérance **Esp(P)**

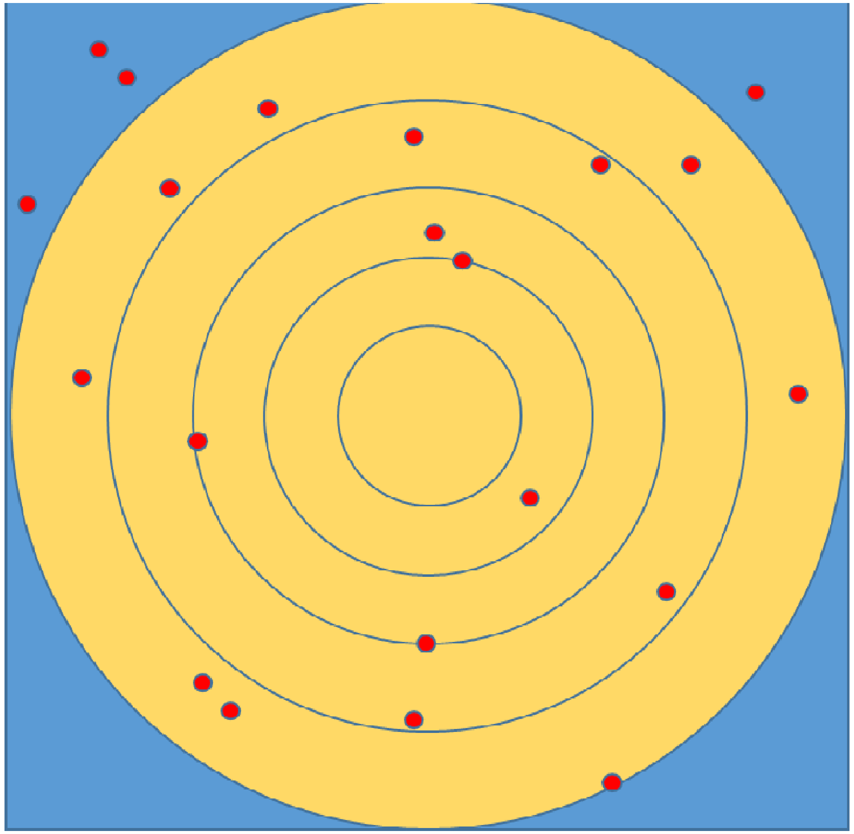
Je peux alors en déduire π = 4. **Esp(P)**

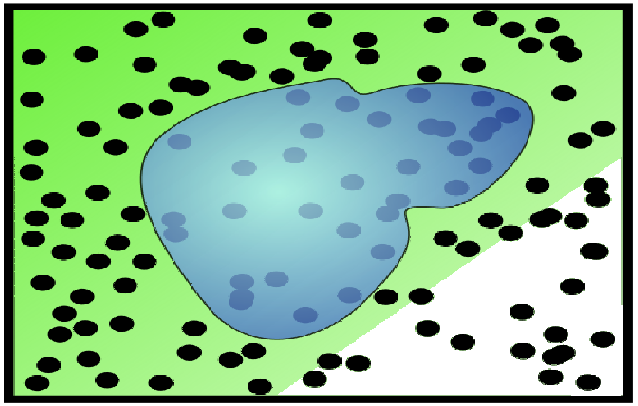
- pour **100 tirs** → π # 3,320
- pour **1 000 tirs** → π # 3,096
- pour **10 000 tirs** → π # 3,154
- pour **100 000 tirs** → π # 3,148

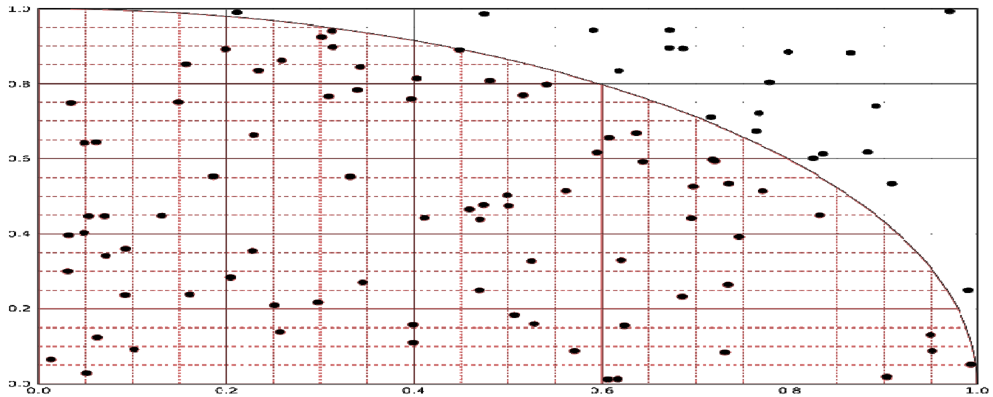


Méthode de Monte-Carlo par Nicholas Metropolis (1915 - 1999)

Sous réserve de disposer un générateur aléatoire (ou pseudo-aléatoire), cette méthode permet de calculer des surfaces complexes (à un faible coût de calcul).

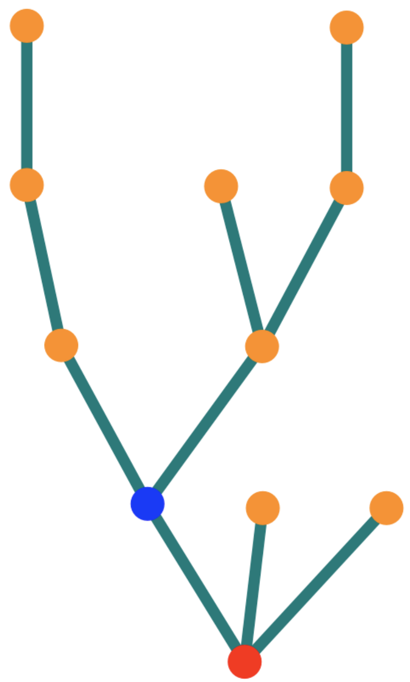


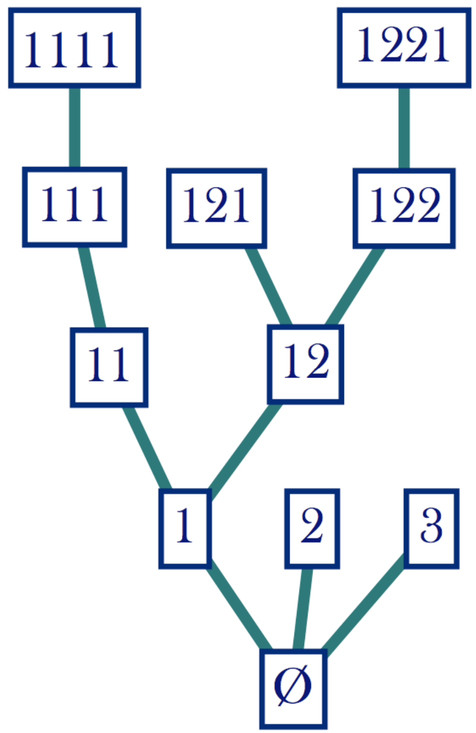


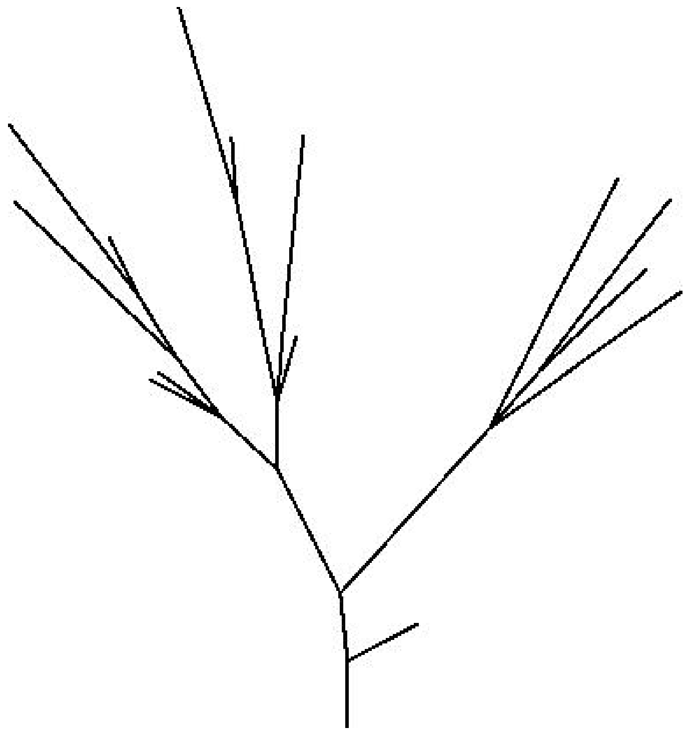


Arbres aléatoires

Avec un générateur aléatoire (ou pseudo-aléatoire),
on peut aussi créer des arbres ou d'autres structures complexes



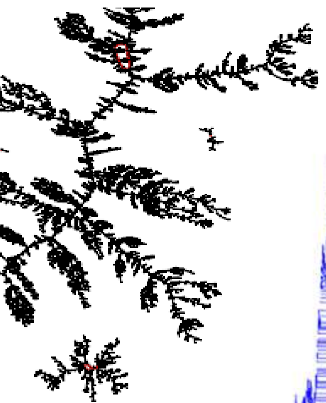


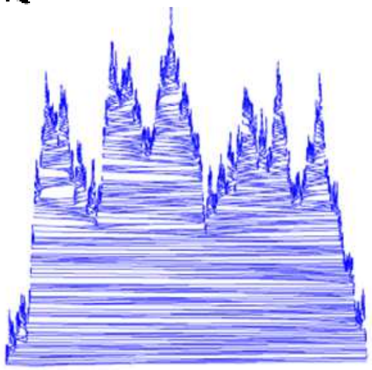


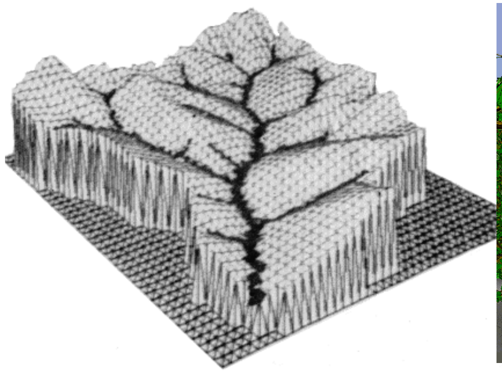
Arbres aléatoires

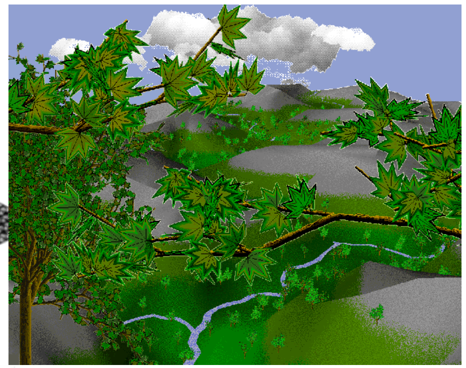
On peut même **générer des paysages complexes** qui imitent la nature dans le vrai monde (minéraux, végétaux, animaux, etc.)

Cela peut servir pour imiter ou simuler le réel dans les Arts, les Sciences, les Technologies, la Culture, etc.









Générateurs pseudo-aléatoires

Question : Comment on fait pour créer un aléa (ou un pseudo-aléa) ?

- **Aléa** : pile ou face, dés, roulette, cartes, tirage dans un chapeau, etc.
  mais aussi chronomètre, radioactivité, bruits électromagnétiques,
  bruits thermiques, bruits sismiques, etc.

- **Pseudo-aléa** : ALEA() en Excel,
  RANDOM() en Python,
  SAMPLE() en R,
  RAND() en C++, etc.

Mais . . . Comment ça marche une fonction pseudo-aléatoire 

Fonctions pseudo-aléatoires

Soit une suite U_n telle que $U_{n+1} = (U_n * A) \bmod (B)$

U_n est un entier appartenant à l'intervalle [0 . . B-1]

U_n / **B** est un rationnel appartenant à l'intervalle [0 . . 1 [

On choisit **A** et **B** parmi les nombres de **Mersenne (1588 – 1648).**
Ce sont des nombres premiers de la forme **N = 2**p – 1** où **p** est premier.

(pour les informaticiens, ce sont des nombres du style 0x0…FFFFFF)

Fonctions pseudo-aléatoires

Un exemple

Soit la suite U_n telle que $U_{n+1} = (U_n * A) \bmod (B)$

Soit la suite V_n telle que $V_{n+1} = (V_n * A) \bmod (B)$

Avec **A** = 131 071 (0x 01 FF FF) et **B** = 8 388 607 (0x 7F FF FF)
qui sont des nombres de Mersenne. **A = 2**17 – 1** et **B = 2**23 – 1**

i

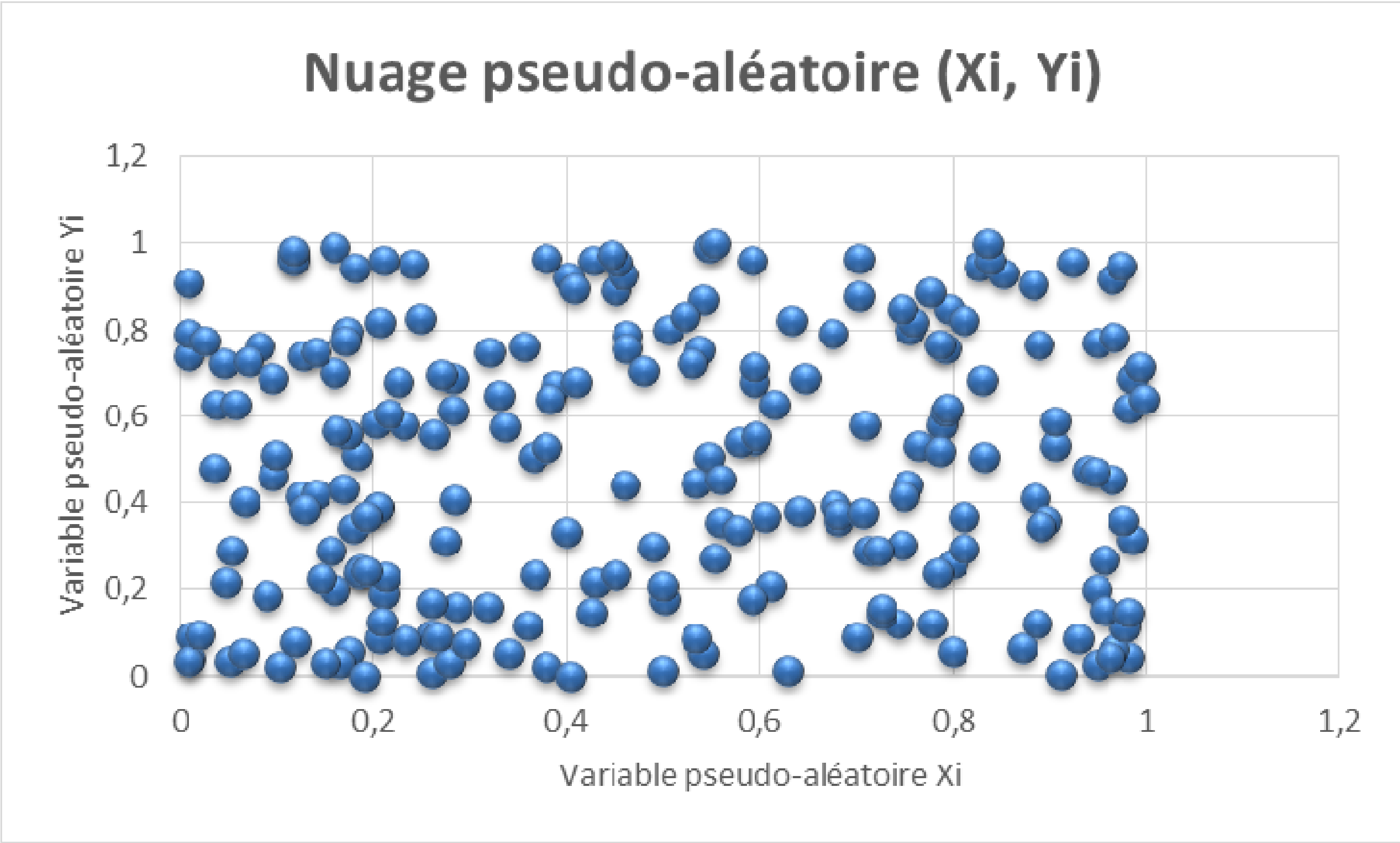
On initialise U_n avec U_0 = 123456 et V_n avec V_0 = 654321

Puis on génère un nuage des points (X_i, Y_i)

avec $X_i = U_i / B$ et $Y_i = V_i / B$ définis sur [0 . . 1 [

Fonctions pseudo-aléatoires

Un exemple



En conclusion

Moi, ce qui me fascine dans cette promenade ludique,

c’est qu’on rencontre des scientifiques de la Renaissance, du Siècle des Lumières, du XIXème ou du XXème siècle qui s’intéressaient à des **problèmes complexes sans aucun moyen de calcul sophistiqué**.

Au lieu de faire tourner des ordis,
ils étudiaient, ils réfléchissaient, ils exploraient, ils cogitaient, …
Et ils faisaient avancer la Science.

Cela nous laisse encore **beaucoup d’espoir pour l’avenir . . .**