

# **Introduction à la statistique et aux probabilités**

**Mathématiques Générales B  
Université de Genève**

Sylvain Sardy

21 février 2008

# 0. Organisation du cours

**Page web** : A visiter régulièrement.

**Horaires du cours** : jeudis de 10h15 à 12h00.

**Horaires des TD/TP** : jeudis de 8h15 à 10h00.

**Software statistique** : R disponible gratuitement

**Assistants et leurs heures de réception** : voir page web.

**Examen** : voir page web.

**Objectifs** : apprendre concepts clefs en statistique et probabilités, tels que :

- Analyse exploratoire et utilisation du software statistique R.
- Analyse combinatoire
- Calculs élémentaires de probabilités
- Calculs de probabilités conditionnelles, formule de Bayes, indépendance
- Distributions univariées, discrètes et continues. En particulier, distributions Bernoulli, Binomial et Poisson, ainsi que Uniforme, Gaussienne et Exponentielle.
- Distribution de plusieurs variables, conjointes, marginales, conditionnelles
- Espérance, variance, quantile
- Introduction à la régression.

## Livres :

- E. Cantoni, Ph. Huber & E. Ronchetti (2006), Maîtriser l'aléatoire : exercices résolus de probabilités et statistique, ed. Springer-France.
- Baillargeon, G. (1995). Probabilités, statistique et technique de régression. Les Editions SMG.
- Dagnelie, P. (1998). Statistique théorique de appliquées, Tome 1. Bibliothèques des Universités : statistique. De Boeck et Larcier.
- Delmas, B. (1996). Statistique Descriptive. Editions Nathan, Paris.
- Dodge (1999). Analyse de régression appliquée.
- Erikson, B.H. et Nosanchuk, T.A. (1985). Understanding Data. The Open University Press, Philadelphia.
- Freedman, D., Pisani, R., Purves, R. et Adhikari, A. (1991). Statistics. W. W. Norton & Co., New York.
- Goldfarb, B. et Pardoux, C. (2004). Introduction à la méthode statistique. Dunod.

- Klatzmann, J. (1985). Attention Statistiques! Comment en déjouer les pièges. Cahiers libres 405, Editions La Découverte, Paris.
- Sanders, D.H., Murph, A.F. et Eng, R.G. (1984). Les Statistiques : une Approche Nouvelle. McGraw-Hill, Montréal.
- Tukey, J.W. (1977). Exploratory Data Analysis. Addison-Wesley, Reading, Massachusetts, London.
- Wonnacott, R.J. et Wonnacott, T.H. (1985). Introductory Statistics. Wiley, New York.

# 1. Motivations

La statistique est une science qui consiste à :

- recueillir
- analyser
- et interpréter

des mesures d'information (les données) dans le but de :

- mieux comprendre
- influencer, contrôler
- prédire
- tester
- optimiser
- détecter

des phénomènes complexes et aléatoires/stochastiques/random.

L'aléatoire et les applications de la statistique sont partout :

**Jeux de hasard** : pile ou face.

Quelle est la probabilité de gagner ?

$$P(\text{gagner}) = \frac{1}{2}.$$

**Jeux de hasard** : euro million.

Quelle est la probabilité de gagner au loto/euro million ?

$$P(\text{gagner}) = \frac{5}{50} \frac{4}{49} \frac{3}{48} \frac{2}{47} \frac{1}{46}$$

soit 1 chance sur 76'000'000 de trouver les 5 bons chiffres entre 1 et 50.

Peut-on augmenter ses chances de gagner au loto/euro million ?



**Votation/Election/Referendum** : les sondages tentent de prédire l'opinion d'une population.

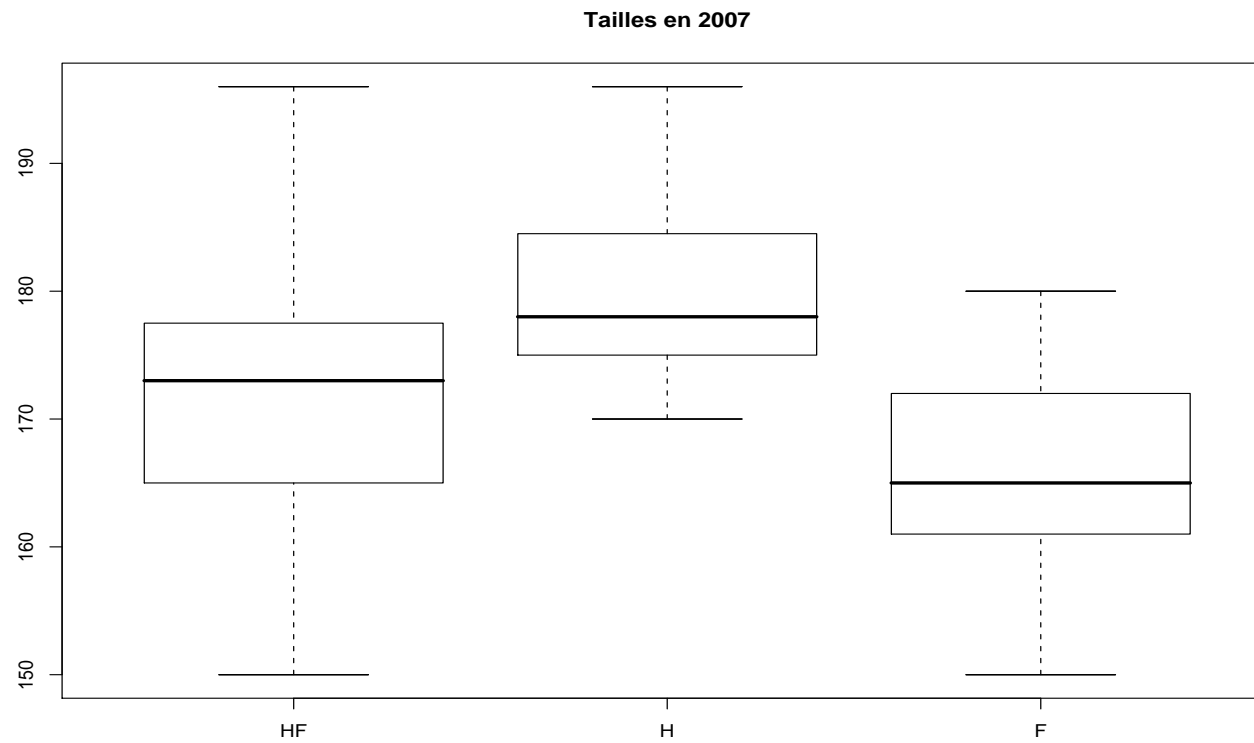
Supposons un referendum entre OUI et NON. Comment feriez-vous pour prédire lequel va passer ?

## Nos données

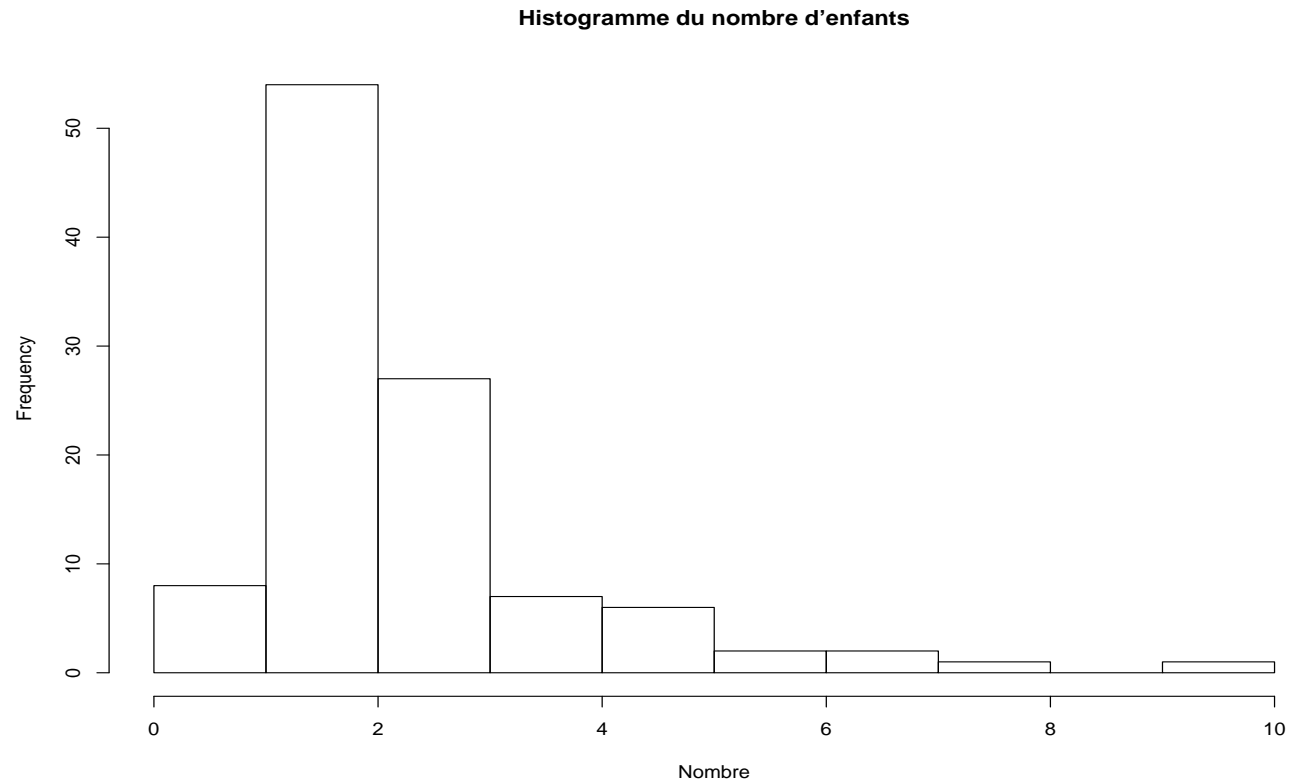
Taille (cm)	Sexe (H/F)	Branche <sup>†</sup>	MoisJour	Nbr frères/soeurs+moi	Intérêt math (0-10)
170	H	5	0227	3	***

<sup>†</sup> 1 : Pharmacie, 2 : Sciences de la terre, 3 : Biologie, 4 : (Bio)chimie, 5 : Autre.

Analyse des données Taille de l'année dernière à l'aide du boxplot :



Analyse des données Nbr enfants/famille à l'aide de l'histogramme :



## **Birthday problem.**

D'après vous, quelle est la probabilité qu'au moins deux personnes soient nées le même jour dans un groupe de 20 personnes ?

Simulation Monte Carlo.

Courbe de probabilité.

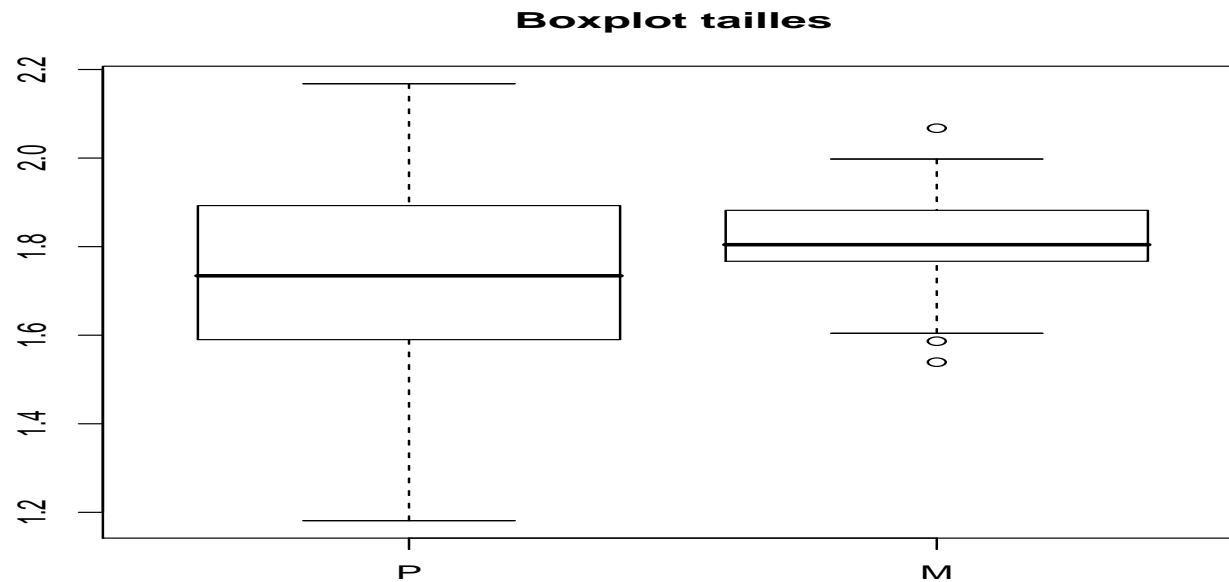
**Pharmaceutique** : tester un nouveau médicament. Comment feriez-vous pour tester si une hormone de croissance est efficace ?

Recueillir des données :

Placebo	1.65m	1.80m	...	1.63m
Médicament	1.47m	...	1.92m	

Faire une analyse exploratoire :

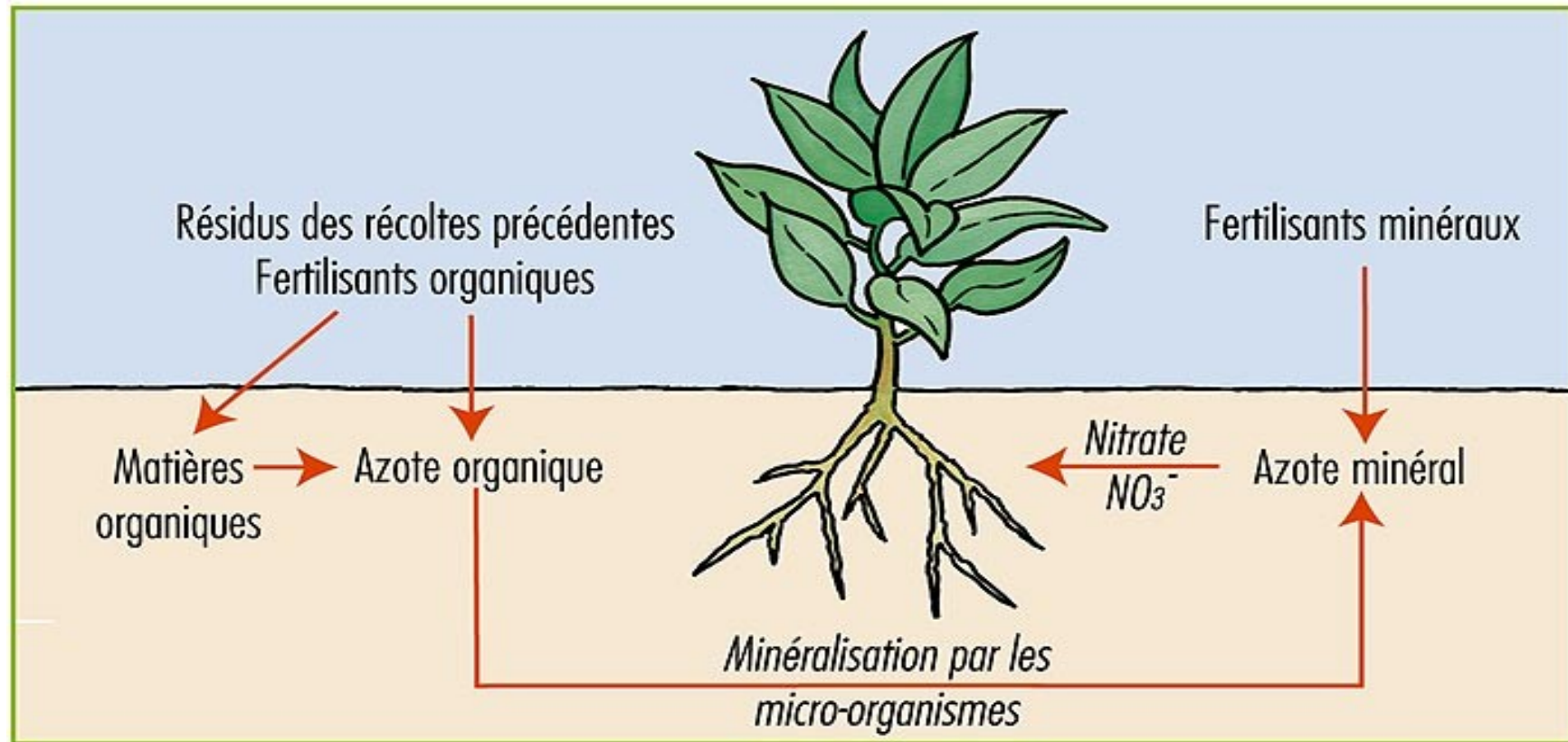
```
> boxplot(data1, data2)
```



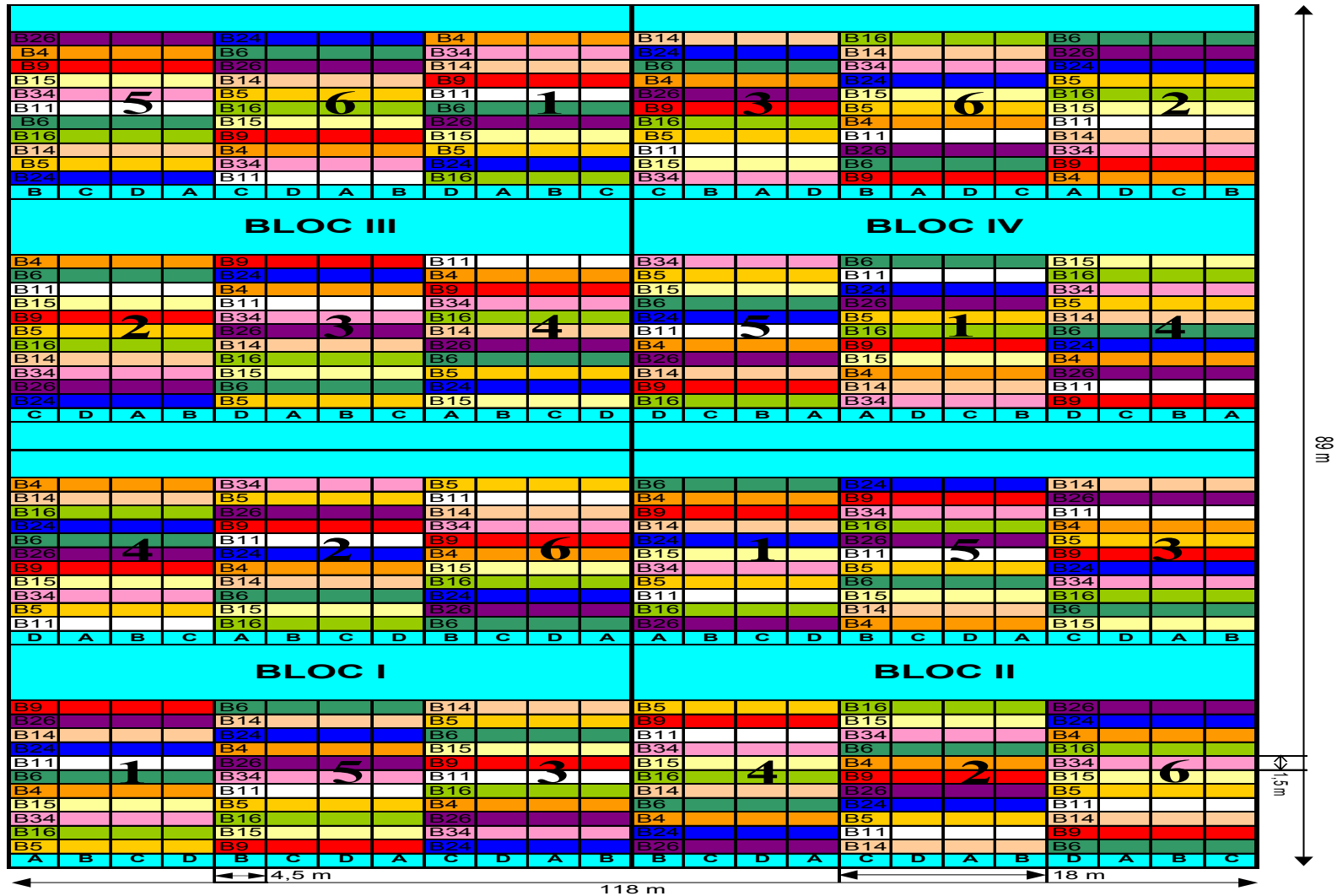
Faire un test statistique approprié.

Dangers : bien choisir la population, qui fait les mesures, avec quoi, etc.

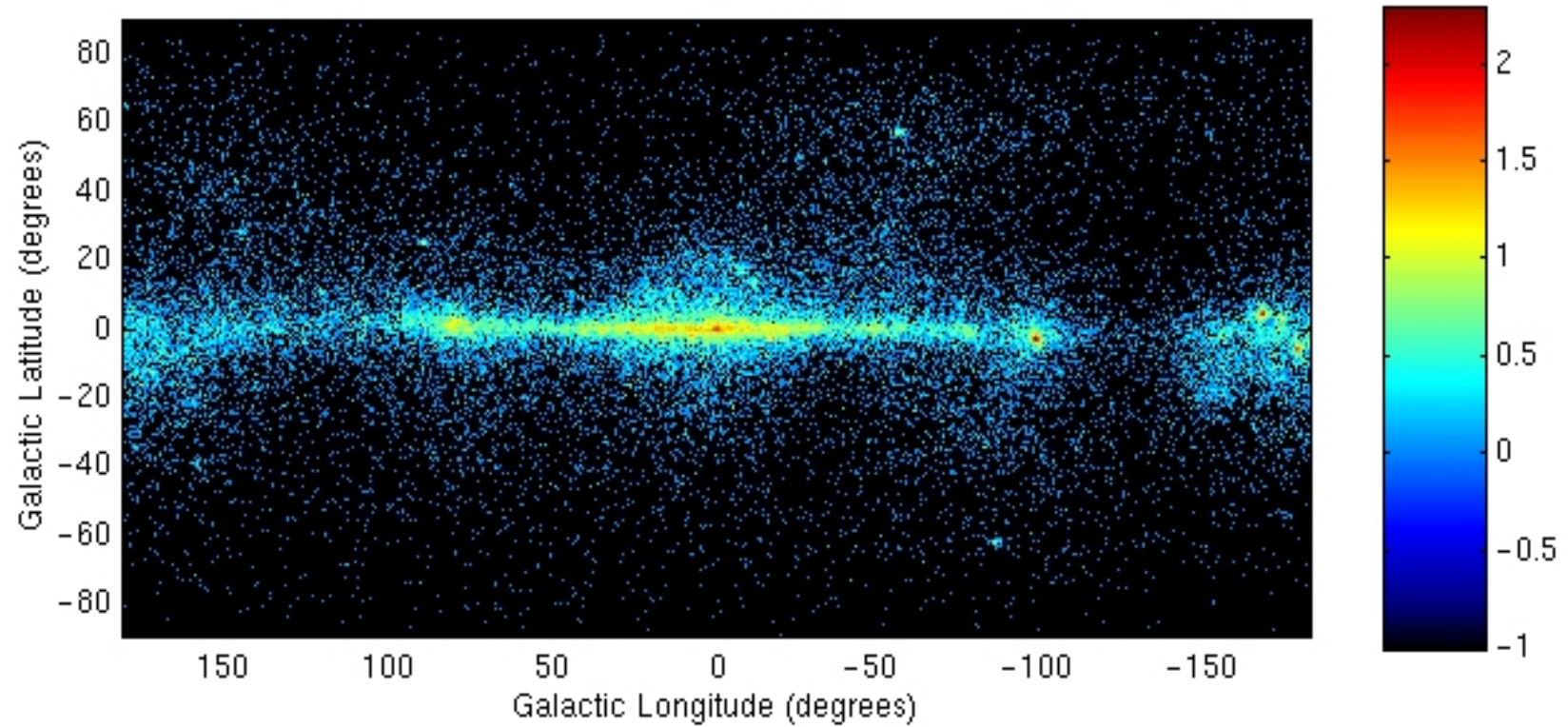
**Agriculture** : améliorer les blés en faisant une expérience dans un champ.



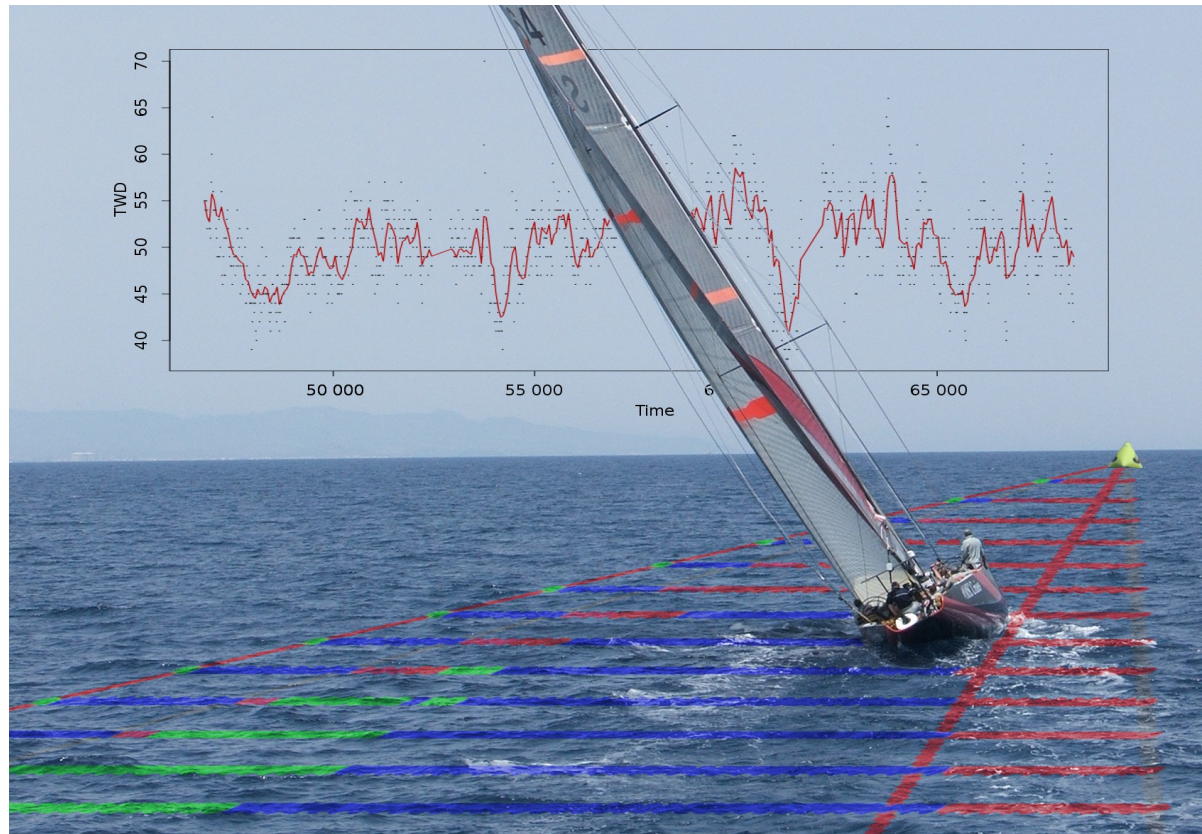




**Astronomie** :  $720 \times 360$  capteurs comptent un nombre *aléatoire* de particules.



## Sport : optimiser les performances d'Alinghi.

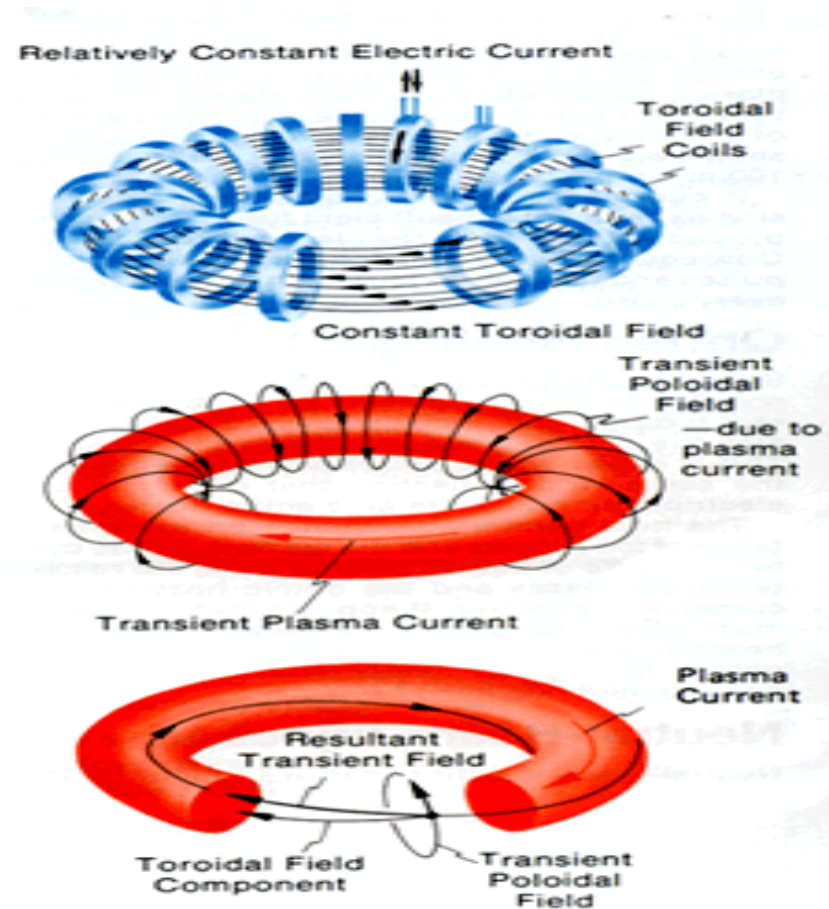


**Economie/Social** : observer la santé d'un pays par son inflation, sa production, sa consommation, son développement durable, etc.

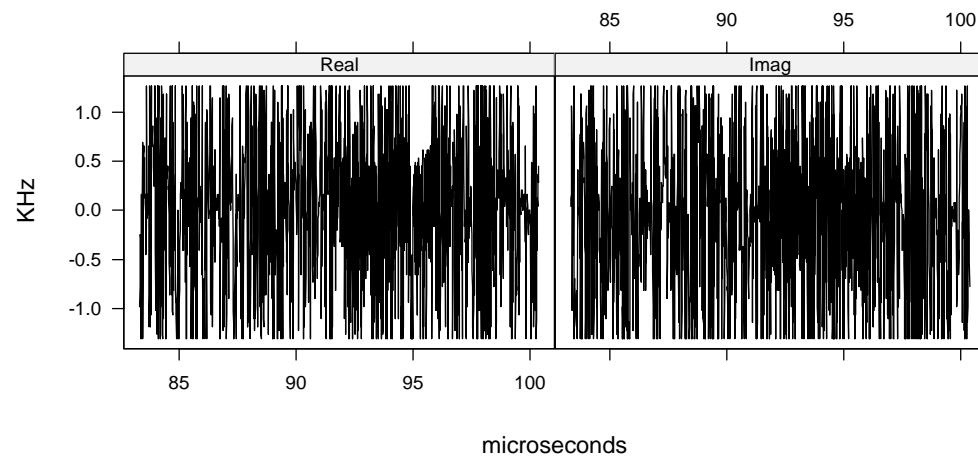
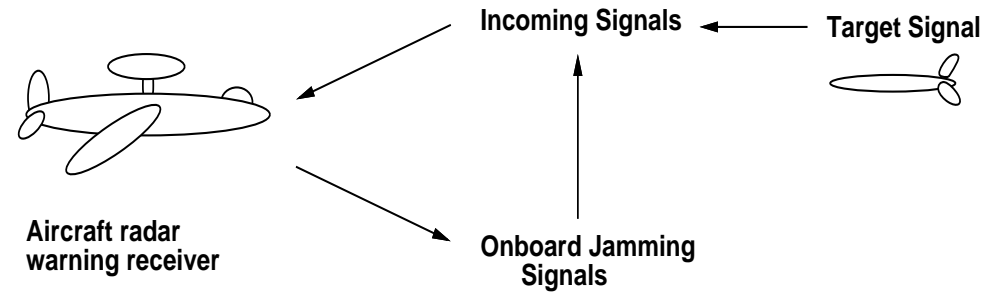
Avoir des statistiques comparables entre pays européens, e.g., baisser l'émission de  $CO_2$  de 20% !

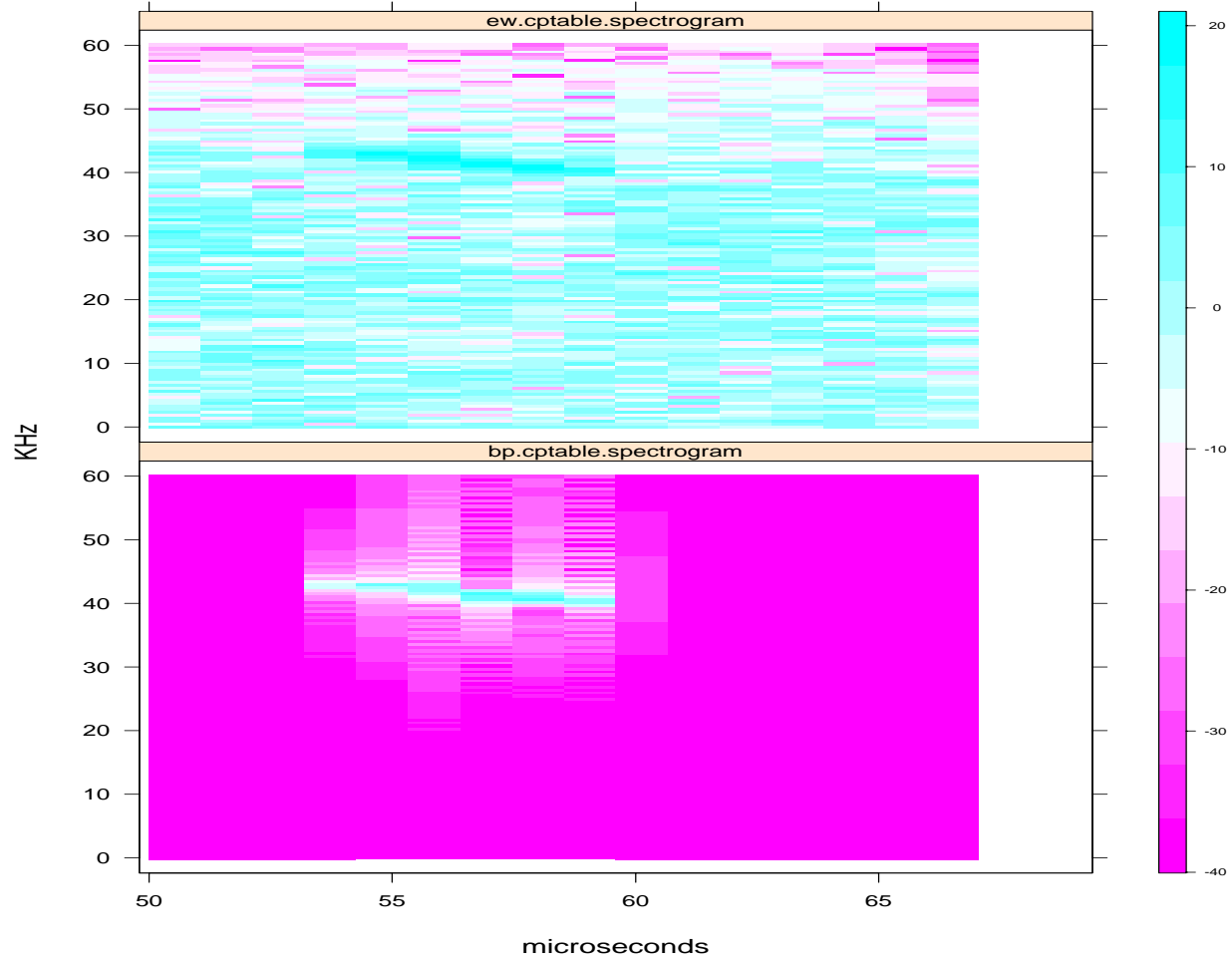
**Chimie** : mieux comprendre et contrôler les processus chimiques.

**Physique** : mieux comprendre la fusion dans le soleil ou dans un tokamak.

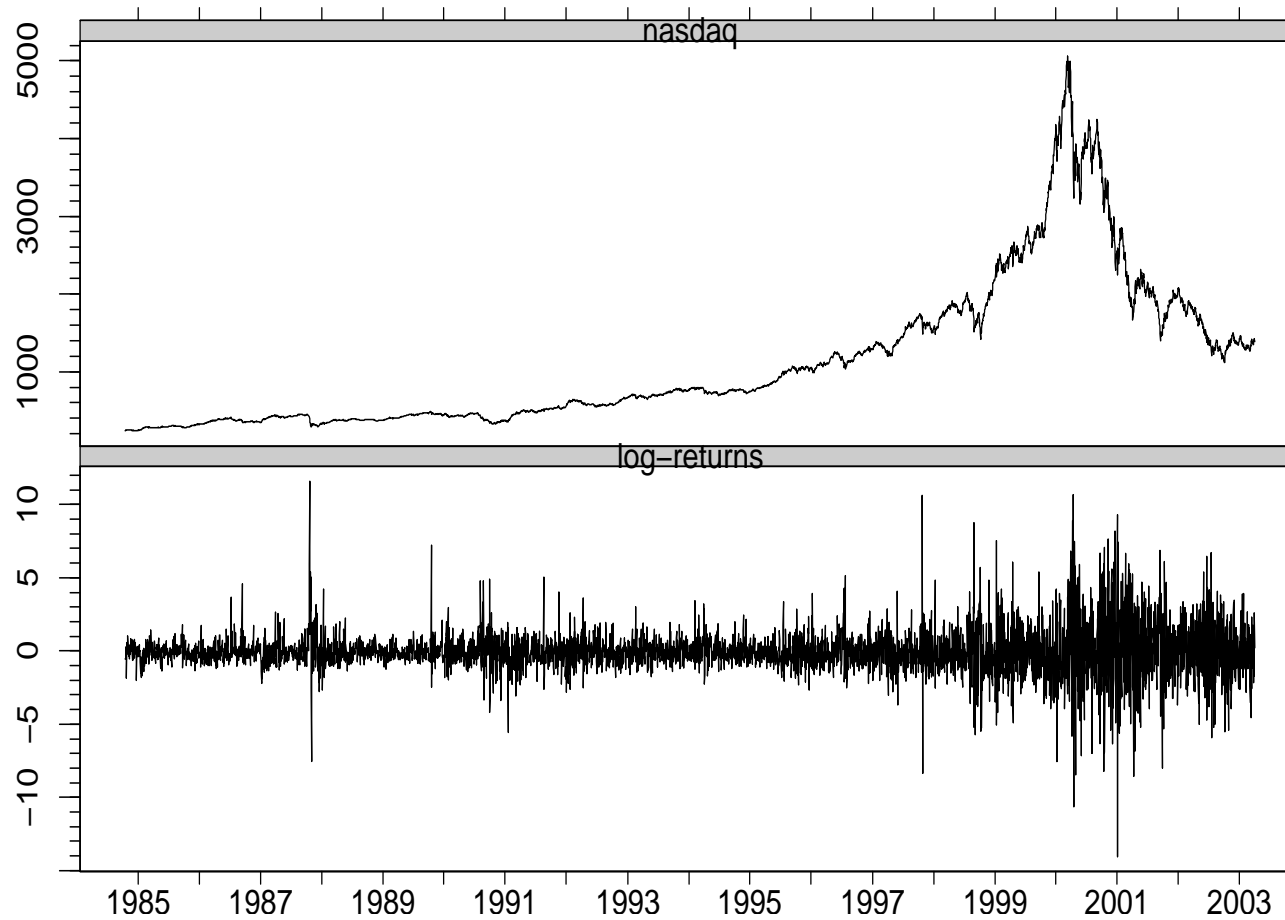


## Militaire : détecter un missile.



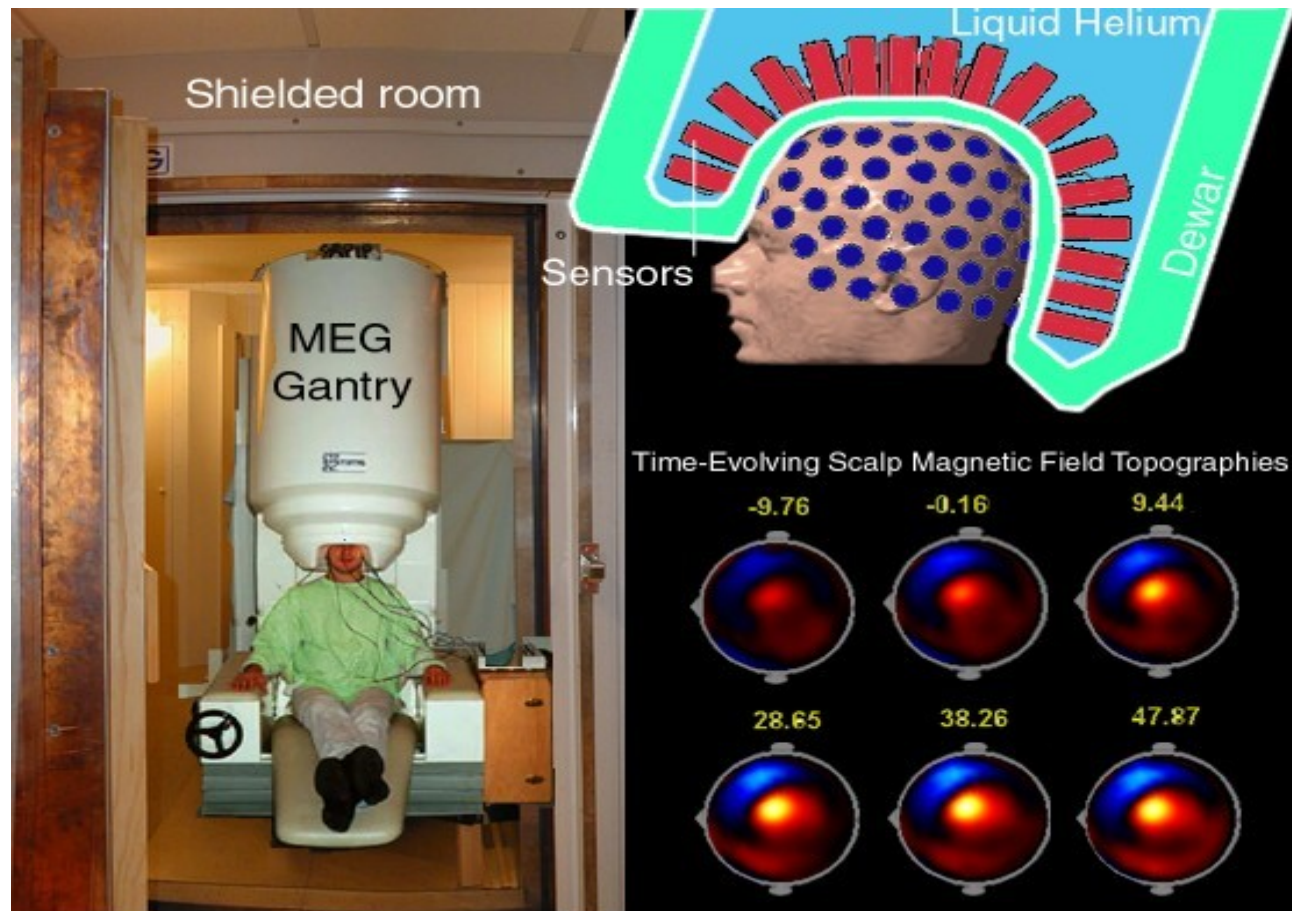


**Finance** : optimiser les portefeuilles boursiers tout en contrôlant le risque.





**Imagerie médicale** : agir sur les crises d'épilepsie, détecter une tumeur.



Vous avez vu en mathématiques des outils variés :

- Nombres : entier, rationnel, réel, etc.
- Figures géométriques : triangle, sphère, etc.
- Fonctions : bijective, convexe, dérivable, etc.

Nous allons voir en statistiques un nouvel outil : la variable aléatoire.

Définition de v.a. réelle : une fonction définie sur l'ensemble des résultats possibles, appelé ensemble fondamental ou univers  $\Omega$ , d'une expérience aléatoire

$$X : \Omega \longrightarrow X(\Omega) \subset \mathbb{R}.$$

Deux types de variables aléatoires :

- discrète quand  $X(\Omega) = \{x_1, x_2, x_3, \dots\}$  est un ensemble fini ou dénombrable.
- continue quand  $X(\Omega)$  est un intervalle de  $\mathbb{R}$ .

Exemples de processus aléatoires

– **Pile ou face :**

$$P(X = \text{pile}) = 1 - P(X = \text{face}).$$

– **Nombre de faces de deux P/F successifs :**

– **Euro million :**

- **Pharmaceutique** :  $\Omega =$  taille d'une personne avec Placebo ou Médicament.  $X(\Omega) = [0, \infty)$ . On modélise  $X$  comme une v.a. continue.

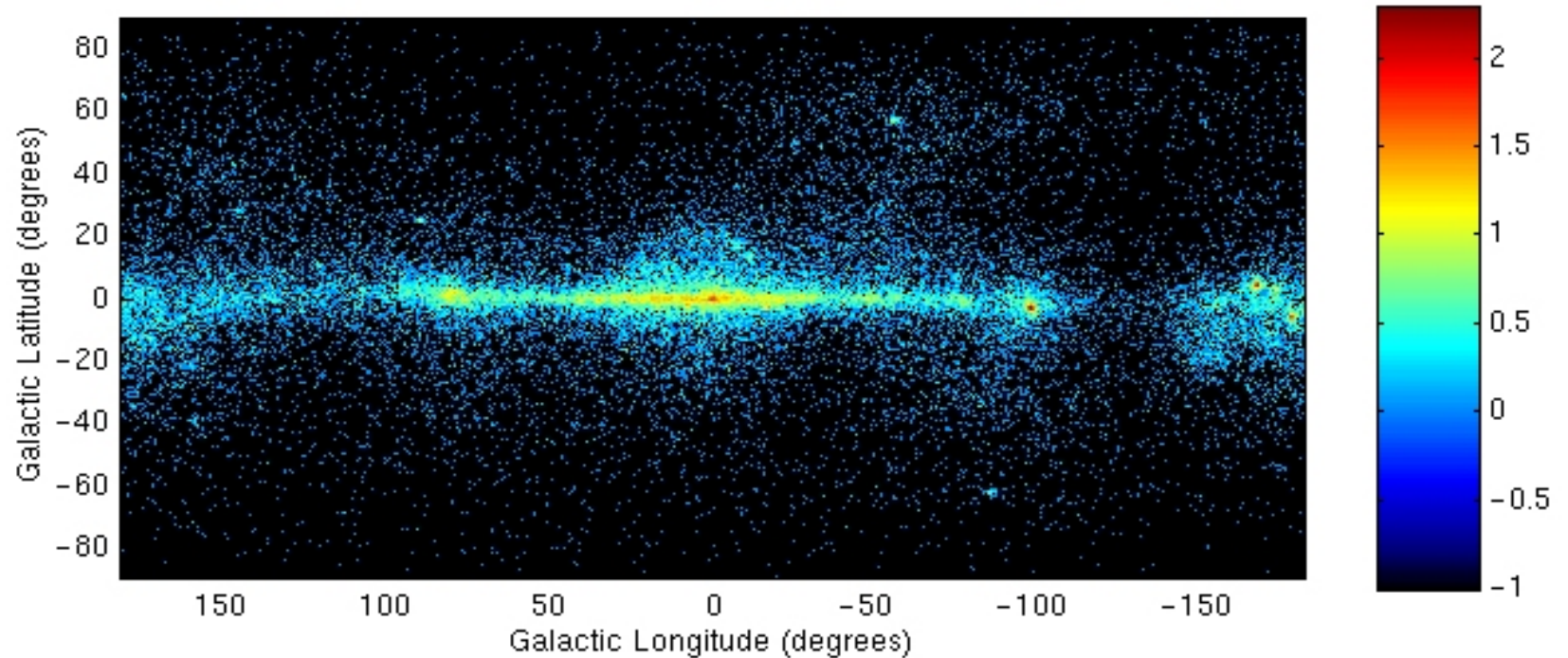
Questions : Quelle est la probabilité qu'une personne avec placebo mesure moins de 180cm ?

$$P(X_P \leq 180) = ?$$

Si on arrive à déterminer une fonction de densité  $f_{X_P}(\cdot)$ , on verra que répondre à la question consiste à calculer une intégrale

$$P(X_P \leq 180) = \int_0^{180} f_{X_P}(u) du.$$

- **Astronomie** : Comptages possibles  $\Omega = \{0, 1, 2, \dots\}$



La loi de Poisson est un modèle possible pour modéliser l'aspect aléatoire de cette image :  $P(X_{0,-20} = 3) = \exp(-\lambda) \frac{\lambda^3}{3!}$ .