

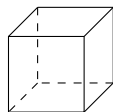
La "malédiction" de la dimension

Alain Franc

INRA BioGeCo & INRIA Equipe Pleiade

Saint-Pierre d'Oléron
ANF "Données massives"
2017

$$C_n = [0, a]^n \in \mathbb{R}^n$$



$$\text{Vol } C_n = a^n$$

$$n = 3$$

En grande dimension

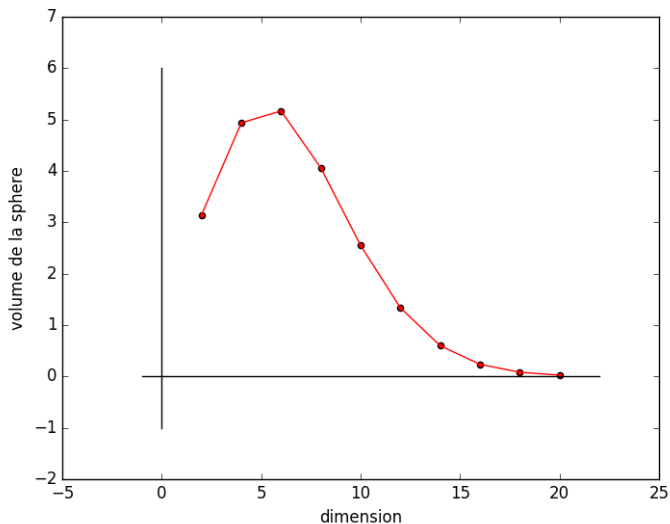
$$\lim_{n \rightarrow \infty} \text{Vol } C_n = \begin{cases} 0 & \text{si } a < 1 \\ 1 & \text{si } a = 1 \\ +\infty & \text{si } a > 1 \end{cases}$$

$$\mathbb{S}^n = \{x \in \mathbb{R}^n : \|x\| \leq 1\}$$

$$\text{Vol } \mathbb{S}^n = \frac{\pi^{n/2}}{\Gamma\left(\frac{n}{2} + 1\right)}$$

$$\text{Vol } \mathbb{S}^{2k} = \frac{\pi^k}{k!} \approx \left(\frac{e\pi}{k}\right)^k$$

Volume de la sphère à n dimensions



Surface de la sphère

Volume d'une sphère de rayon r

$$\text{Vol } \mathbb{S}^n(r) = \frac{\pi^{n/2}}{\Gamma(\frac{n}{2} + 1)} r^n$$

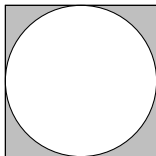
On considère deux sphères emboîtées de rayons $r < r'$.

$$\frac{\text{Vol } \mathbb{S}^n(r') - \text{Vol } \mathbb{S}^n(r)}{\text{Vol } \mathbb{S}^n(r')} = 1 - \left(\frac{r}{r'}\right)^n \rightarrow 1$$

Lemme

Tout le volume d'une sphère est concentré à sa surface en grande dimension.

Comparaison de la sphère et du cube

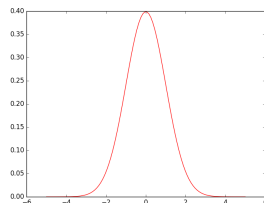


En grande dimension

Tous les points sont dans les coins

Indication : le diamètre de la sphère est toujours $D = 1$, alors que la diagonale du cube a pour longueur \sqrt{n}

$$\mathbb{P}\{X \in [x, x + dx]\} = \frac{1}{\sqrt{2\pi}} \exp -\frac{x^2}{2}$$



Taille de la queue de distribution pour $x \in \mathbb{R}^n$

n	1	2	5	10	20	100
$\mathbb{P}\{\ x\ > 2\}$	0.045	0.135	0.549	0.947	0.999	1

1

¹J. Wang, Geometric structure of High-Dimensional Data and dimensionality reduction, 2012

Concentration de la moyenne et de la variance

On se donne une variable aléatoire $X \in \mathbb{R}$ et n observations (x_1, \dots, x_n) iid.

$$\|x\| = \sqrt{\sum_i x_i^2}$$

Concentration

Alors, $\exists a, b$ tels que

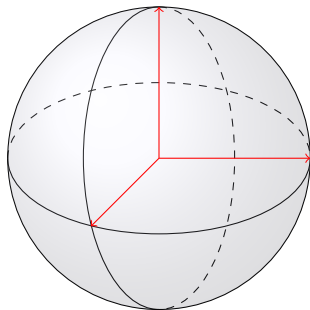
$$\mu(\|x\|) = \sqrt{an - b} + O(1/n)$$

$$\sigma(\|x\|) = b + O(1/\sqrt{n})$$

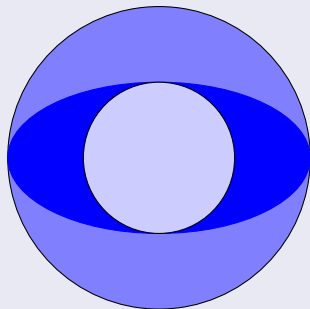
$n \rightarrow +\infty \implies$

Tous les points sont sur la peau de la sphère de rayon \sqrt{an}

Concentration de la mesure sur la sphère



Cercles inscrits et exinscrits à une ellipse



Paraphrasé, source Wikipédia

On ne peut pas avoir de désordre complet dans une structure assez grande, ou plutôt qu'une telle structure contient nécessairement des sous-structures ayant un certain ordre.

Exemple

Le théorème de Ramsey énonce que si l'on impose un tracé en un nombre fixé de couleurs et une taille fixée (par exemple 100), un « dessin » arbitraire suffisamment grand contiendra nécessairement un réseau de cette taille, donc formé de 100 traits adjacents, tous colorés de la même couleur.

