

Chapitre 4 - Fléau de la dimension

24 février 2026

Plan du chapitre

Malédiction de la dimension

Problèmes géométrique

Problèmes probabilistes

La Malédiction de la Dimension

Terme introduit par **Richard Bellman en 1961**. Dans des espaces de grande dimension, de nouvelles difficultés apparaissent car l'espace devient "vide".

L'exemple de Leo Breiman

- En dimension 1 : 100 observations suffisent à couvrir l'intervalle $[0, 1]$ pour faire un histogramme fiable.
- En dimension 10 : Ces 100 observations dans $[0, 1]^{10}$ sont des points perdus dans un vaste espace vide.
- **Conséquence** : Pour obtenir la même densité de couverture qu'en 1D, il faudrait 10^{20} observations ! L'apprentissage devient impossible par manque d'exemples.

Pas de "plus proche voisin" en grande dimension

Le volume d'une sphère de rayon r en dimension d s'effondre lorsque d grandit. Pour une sphère de rayon 1 :

$$V(1) = \frac{\pi^{d/2}}{\Gamma(1 + d/2)} \sim \left(\frac{2\pi e}{d}\right)^{d/2} (d\pi)^{-1/2}$$

Ce volume décroît vers zéro en $d^{-\frac{d+1}{2}}$.

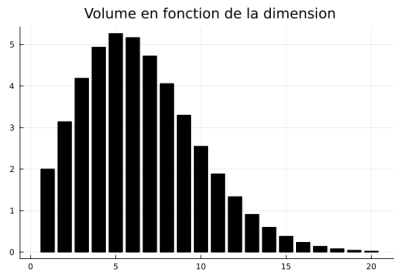


Figure – Volume de la boule unité en fonction de d .

L'illusion de la couverture spatiale

Combien faut-il de boules de rayon 1 pour couvrir $[0, 1]^d$? Il faut que $nV(1) \geq 1$, ce qui implique :

$$n \geq \frac{\Gamma(1 + d/2)}{\pi^{d/2}} \sim \left(\frac{d}{2\pi e} \right)^{d/2} \sqrt{d\pi}$$

Quelques valeurs édifiantes :

- $d = 20 \rightarrow n = 39$ points.
- $d = 50 \rightarrow n = 5 \times 10^{12}$ points.
- $d = 100 \rightarrow n = 4 \times 10^{39}$ points.

Conclusion : Les méthodes basées sur le voisinage (k -NN) échouent en grande dimension.

Répartition de la masse (La coquille vide)

En grande dimension, presque tout le volume d'une boule se trouve dans son "écorce" (sa surface extérieure).

Prenons une écorce C_d représentant seulement 1% du rayon extérieur :

$$\frac{\text{Volume de } C_d(0, r)}{\text{Volume de } B_d(0, r)} = 1 - 0.99^d$$

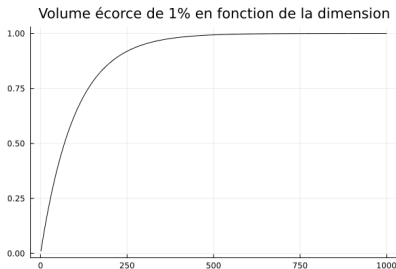
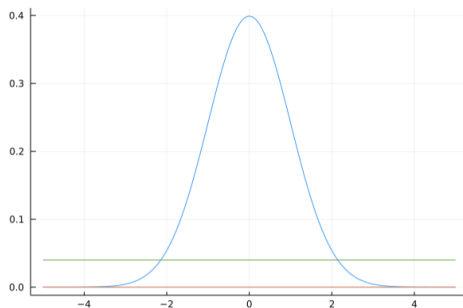


Figure – La majeure partie de la masse se trouve dans l'écorce.

Comportement Gaussien : 1D vs Grande Dimension

En petite dimension (ex : 1D)

Pour $X \sim \mathcal{N}(0, 1)$, l'essentiel de la masse de probabilité est concentré autour de la moyenne (la cloche).



Comportement Gaussien : 1D vs Grande Dimension

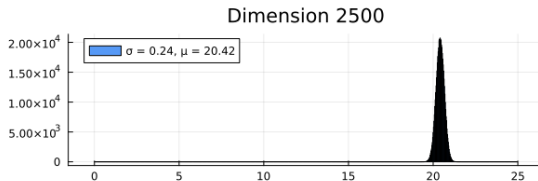
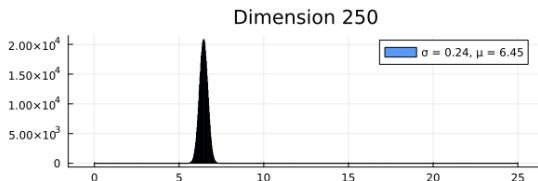
En grande dimension

Pour $X \sim \mathcal{N}(0, I_d)$, la masse fuit vers les queues de distribution.
La probabilité d'être proche du centre s'effondre :

$$\mathbb{P}(\exp(-\|X\|^2/2) \geq \delta) \leq \frac{1}{\delta 2^{d/2}}$$

Éloignement des points

Les points sont uniformément éloignés les uns des autres.



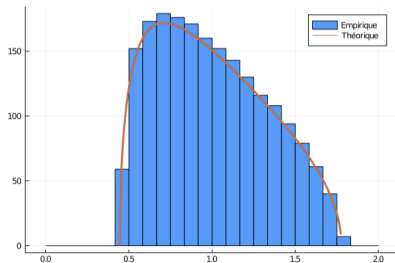
Distances entre paires (5000 points)

La distance moyenne croît avec la dimension, mais l'écart-type reste constant. Les points semblent tous équidistants.

Estimation de la matrice de covariance (Marcenko-Pastur)

$S = \frac{1}{n}XX^T \in \mathbb{R}^{d \times d}$. Si $n \gg d$: loi des grands nombres.

Si d est du même ordre que n ($d/n \rightarrow c \in (0, 1]$) :



Spectre des valeurs propres (Loi de Marcenko-Pastur).

L'étalement des valeurs propres vaut $4\sqrt{c}$. Une covariance très mal estimée ruine les performances des modèles sophistiqués.