

# Risque réel

Bruno Bouzy

[bruno.bouzy@u-paris.fr](mailto:bruno.bouzy@u-paris.fr)

Cours apprentissage machine

Janvier 2022

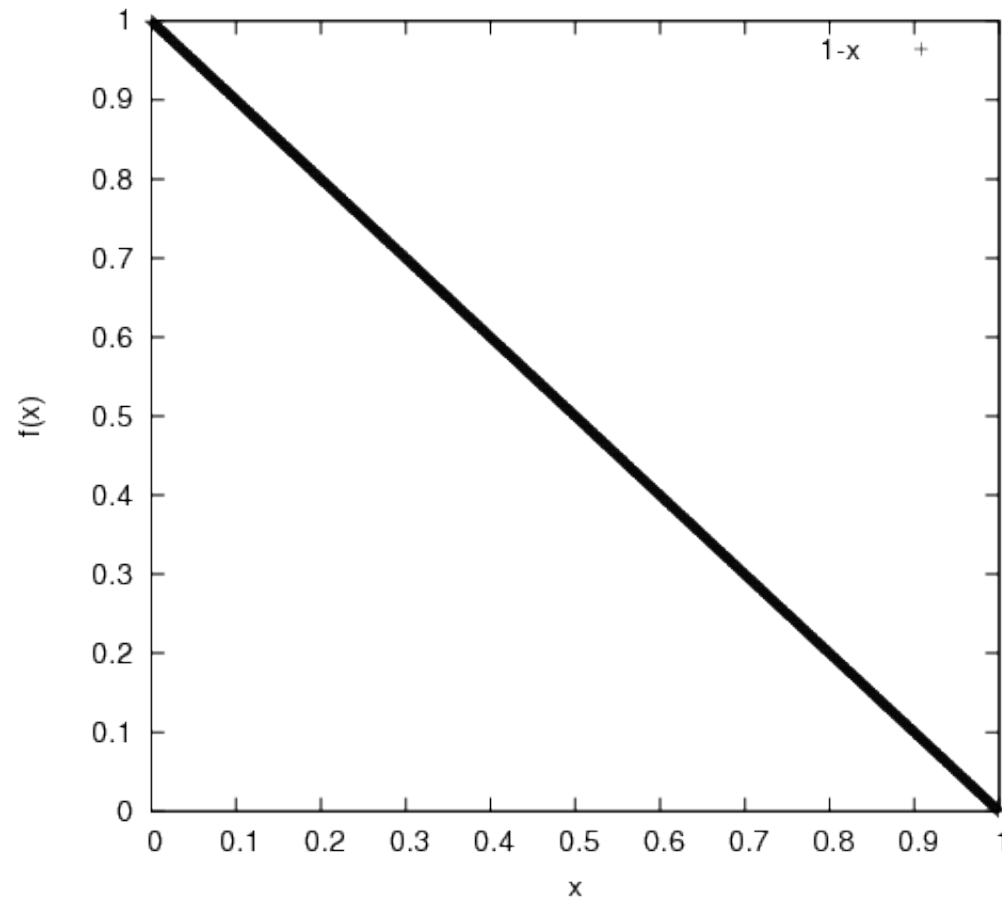
# Objectif

- Visualiser le **risque réel** de l'approximation de 4 fonctions cibles de « complexité » croissante
- Le risque réel correspond à l'erreur commise dans la réalité.
- On ne connaît pas la réalité, on l'approxime par l'espace des hypothèses.
- Théoriquement, le risque réel est inconnu.
- On connaît et on travaille dans l'espace des hypothèses.
- Le risque empirique est mesuré effectivement dans l'espace des hypothèses et il correspond à l'erreur commise dans cet espace.
- Ici, pour visualiser le risque réel, on triche : on suppose que la réalité est une discrétisation haute de l'intervalle  $[0, 1]$  avec  $M=100\ 000$  points.
- Et que l'espace des hypothèses utilise des discrétisations avec  $N$  points et  $N = M/2, M/4, M/8$ , etc tant que  $N > 2$
- On veut montrer que le risque réel diminue lorsque  $N$  augmente

# Les 4 fonctions cibles

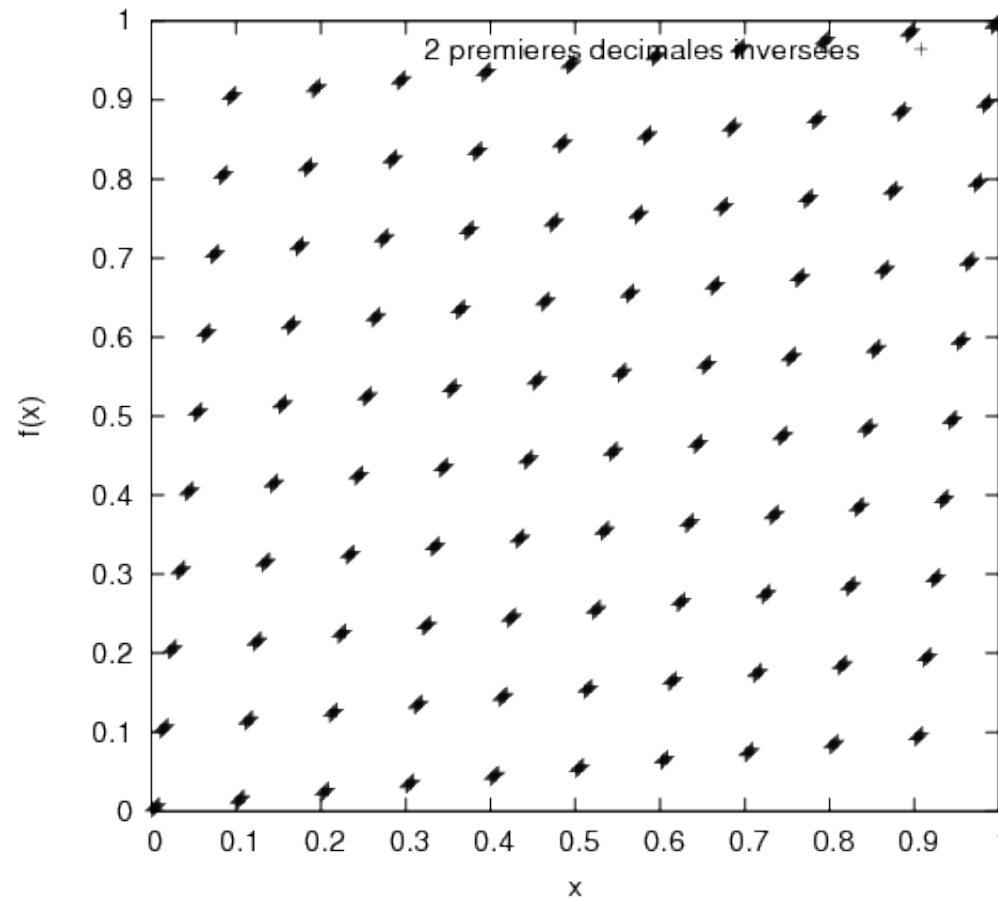
- $f : [0, 1] \rightarrow [0, 1]$ 
  - $f(x) = 1-x$
  - $f(x) = x$  avec 2 premiers chiffres inversés
  - $f(x) = \text{fdecale}(x)$ 
    - La  $i$ ème décimale de  $f(x)$  est la  $j$ ème décimale de  $x$
    - Avec  $j = i + d_i/3$
    - Avec  $d_i$  :  $i$ ème décimale de  $x$
  - $f(x) = \text{random}()$

$$f(x) = 1-x$$



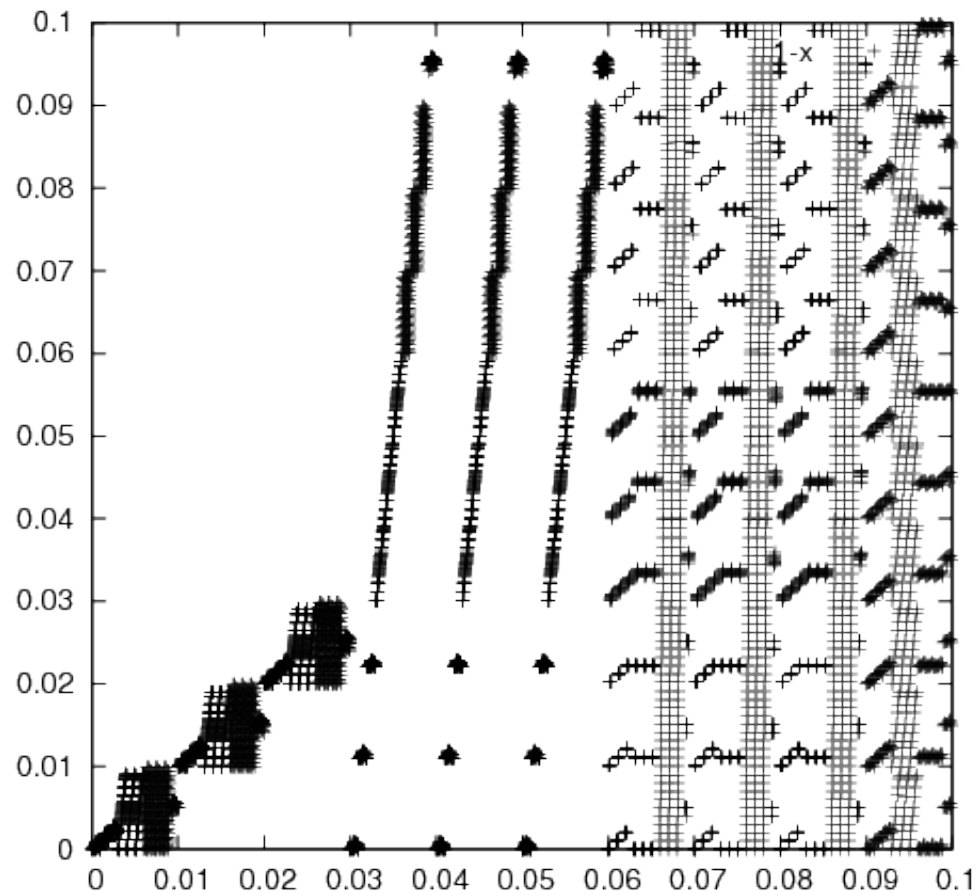
Risque Réel

# inverser les 2 premiers chiffres



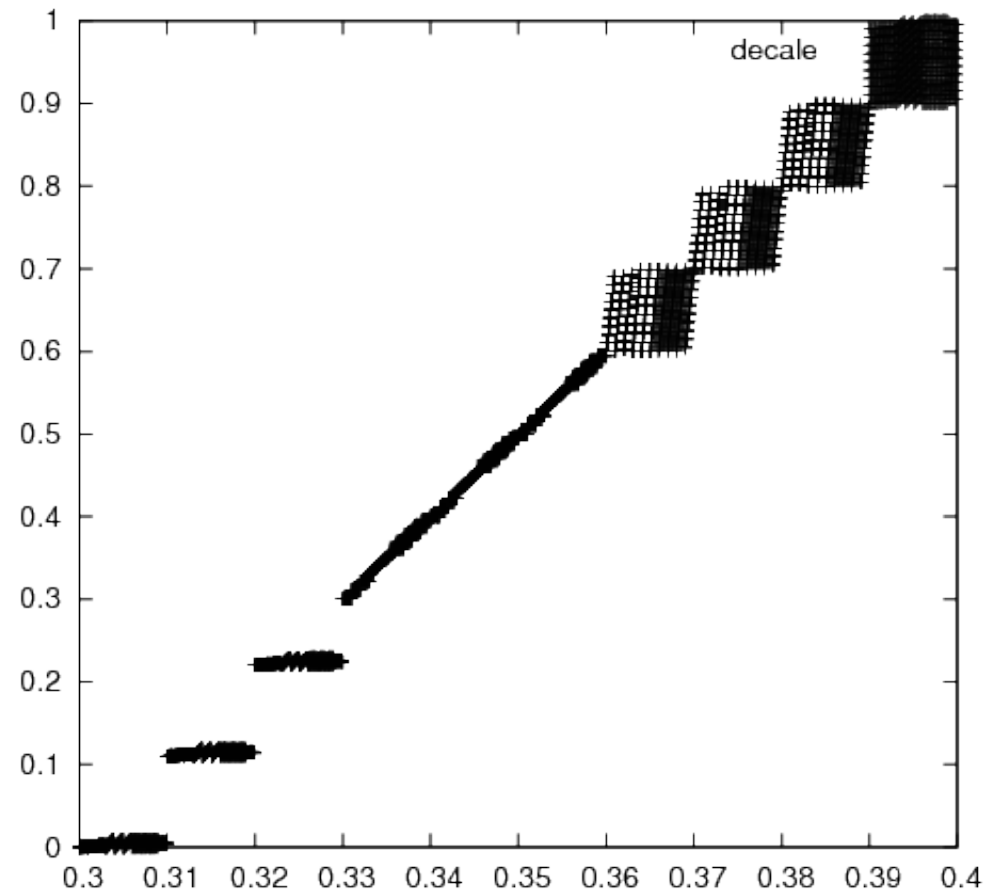
Risque Réel

# $f_{dec}(x)$



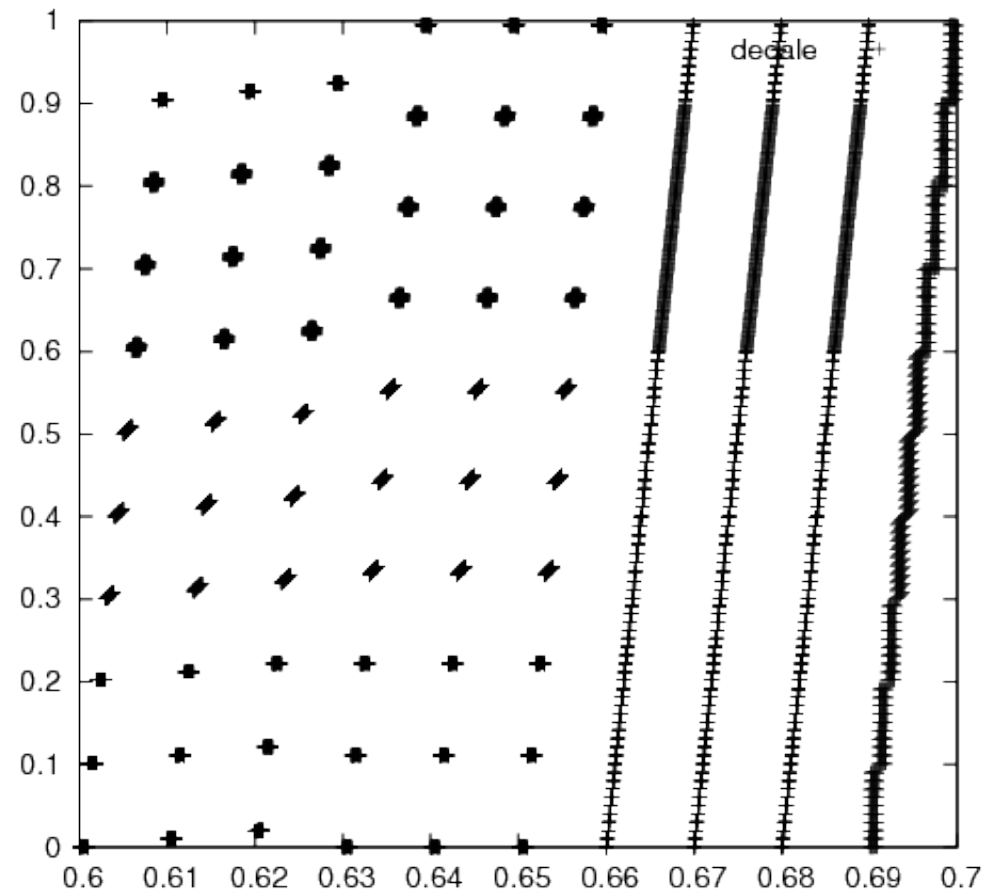
Risque Réel

# Fdecale sur [0.3, 0.4]



Risque Réel

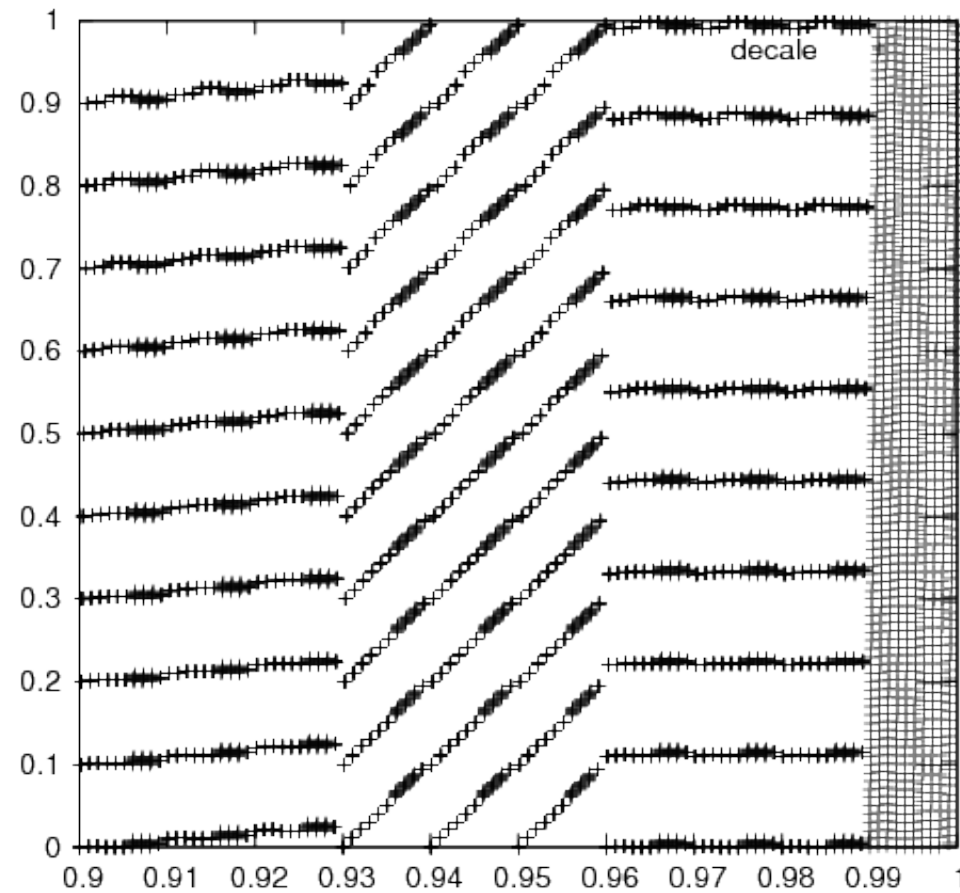
# Fdecale sur [0.6, 0.7]



Risque Réel

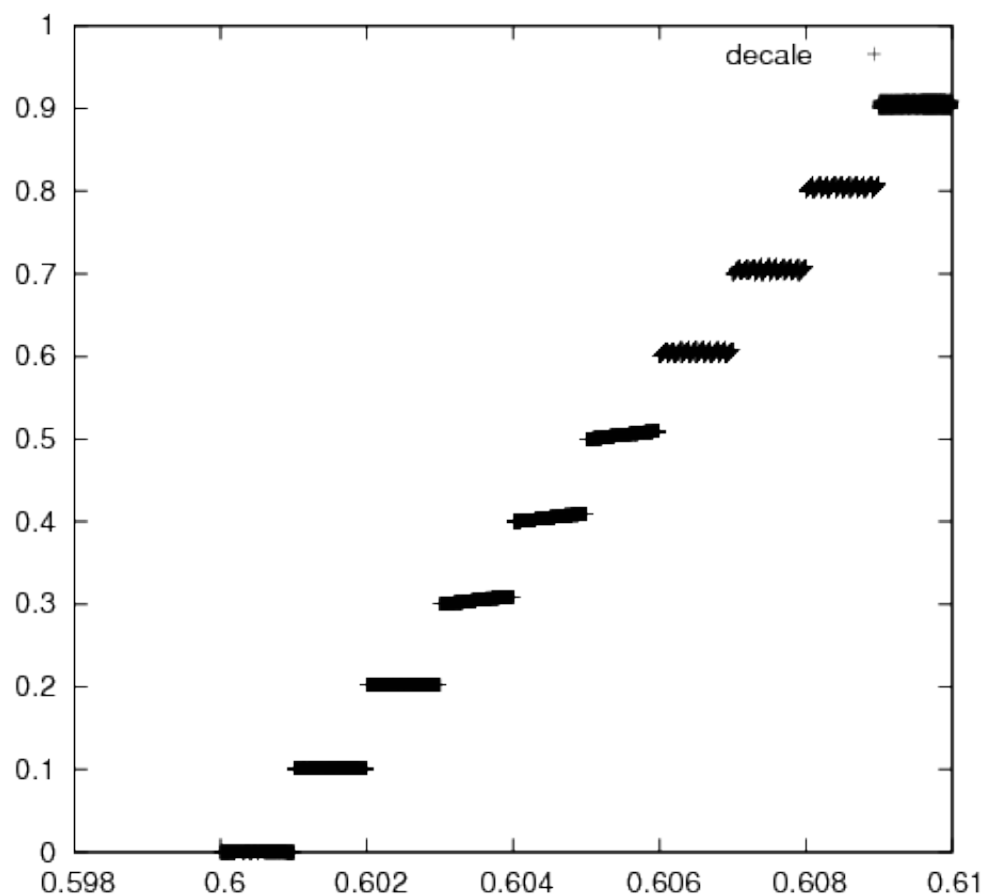


# Fdecale sur $[0.9, 1]$



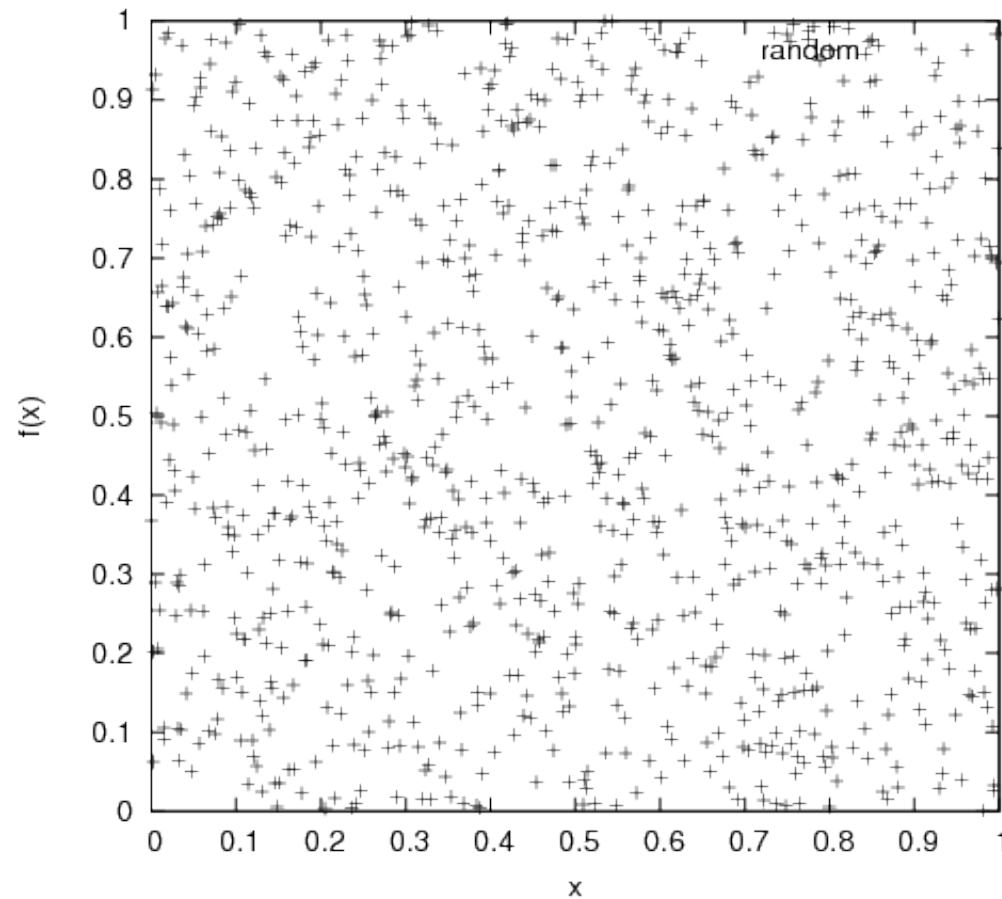
Risque Réel

# Fdecale sur [0.60, 0.61]



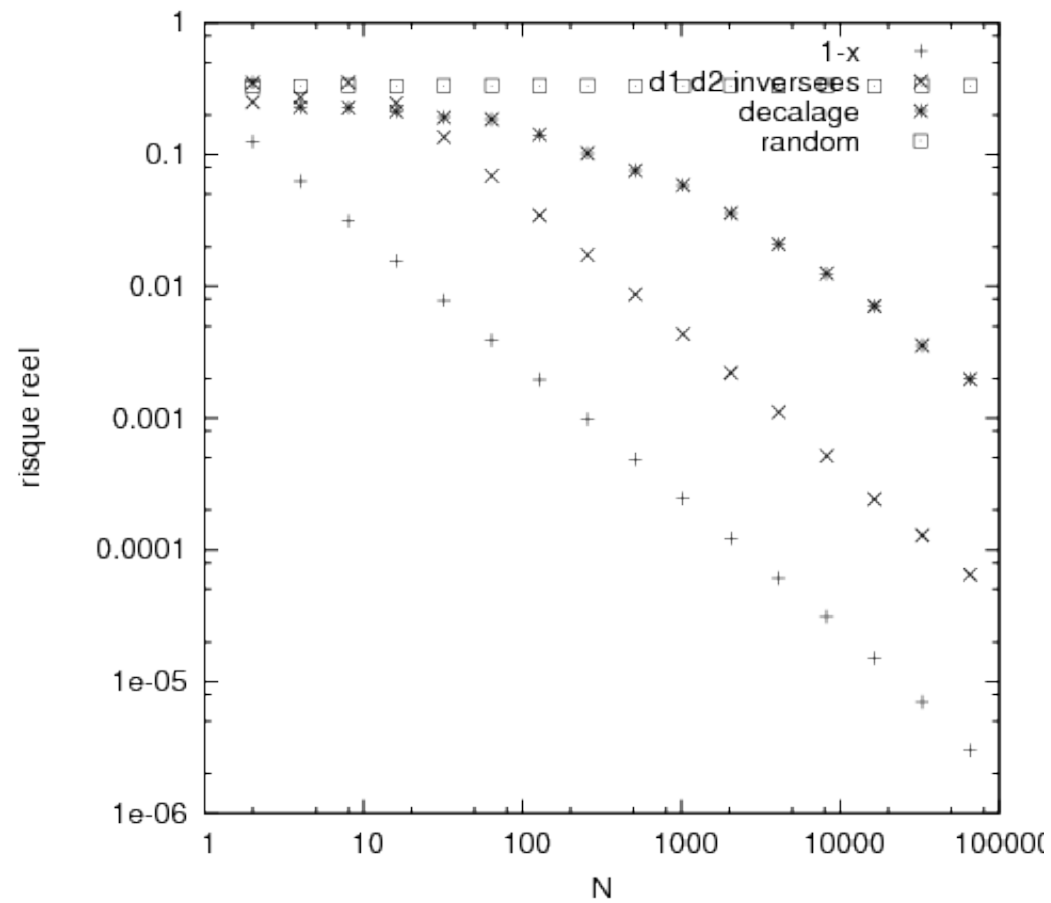
Risque Réel

$$f(x)=\text{random}()$$



Risque Réel

# Risque réel



Risque Réel

# Conclusion

- Pour les 3 premières fonctions
  - Biais = 0
  - Le risque réel tend vers 0 lorsque  $N$  tend vers l'infini
  - Plus ou moins vite selon la complexité de  $f$
- Pour  $f_{\text{decade}}$  :
  - $f_{\text{decade}}$  est une fonction cible très « compliquée visuellement »
  - Mais le biais d'approximation avec des fonctions constantes par intervalles est cependant nul
- Pour la fonction  $\text{random}()$ 
  - Biais = 0.3
  - Le risque réel est constant