

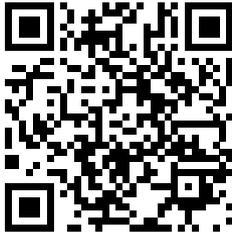
# BENCHMARKING

Loïc Rouquette

# Introduction du cours

## Licence

Ce cours est dérivé du cours de « Experimental Analysis of Algorithms » de Christine Solnon (INSA Lyon) et du mooc fun « Recherche reproductible : principes méthodologiques pour une science transparente » sous Licence [Creative Commons CC-BY 3.0](#).



[Experimental Analysis of Algorithms](#)



[Recherche reproductible : principes méthodologiques pour une science transparente.](#)

# Acquis d'apprentissages visés & Évaluation

## AAVs

- **Rédiger** un document computationnel ;
- **Mettre en œuvre** une architecture replicable ;
- **Utiliser** des outils statistiques et des métriques adaptés aux données analysées ;
- **Présenter** des données de manière transparente.

## Évaluation

Projet : réalisation d'une analyse expérimentale sur l'impact de la lecture des différentes couches mémoire dans le calcul matriciel.

# Programme

## Séance 1

- Choix des critères des performances logicielles
- Importance de l'analyse des performances
- Choix des facteurs, des points de conception et des mesures de performance
- Principes de conception expérimentale

## Séance 2

- Analyse statistique des performances
- Visualisation des données de performance
- Méthodes de validation des mesures de performance
- Vérification de la reproductibilité des expériences
- Évaluation de la robustesse des conclusions

### Séance 3

- Définition de la réproductibilité
- Importance de la réproductibilité dans la recherche informatique
- Problèmes courants de réplication des résultats
- Principes de la science reproductible
- Reproductibilité vs. reproductibilité exacte

### Séance 4

- Gestion des environnements de développement (ex : Docker, Conda)
- Contrôle de version avec Git
- Automatisation des workflows de recherche
- Étude des pratiques émergentes en matière de réproductibilité
- Exploration de sujets avancés (ex : conteneurisation, workflow-as-code)
- Défis actuels et futures directions de recherche

# Choix des critères des performances logicielles

## Exactitude vs Précision

Aa

### Définition

**Exactitude** : proche de la réalité

**Précision** : qualité de la mesure

!

### Attention

Il ne faut pas confondre avec l' *accuracy* et la *precision* en *Machine Learning* (ML).

## Le temps

## Le temps

### *Dominant Cost*

Équivalent empirique de l'analyse de complexité.

**Concept** : partir des fonctions simples vers les fonctions complexes.

**Difficulté** : évaluer correctement les coûts.

**Exactitude** : élevée

**Précision** : faible

## Le temps

### *CPU Time*

Calcul du temps *CPU* nécessaire à réaliser la tâche.

**Concept** : faire des sondages sur les fonctions exécutés par l'ordonnanceur afin de pouvoir calculer le temps d'exécution de l'algorithme (statistique).

**Exactitude** : faible (pas de généralisation)

**Précision** : précis sur les calculs long ( $> 1s$ )

## Le temps

### *Real Time*

Calcul du temps écoulé entre le début et la fin de la tâche.

**Concept** : on lance un chronomètre.

**Exactitude** : faible (pas de généralisation)

**Précision** : ça dépend !

## La mémoire

Principalement deux métriques :

- La mémoire maximum utilisée
- La bande passante (débit)

# La qualité

## La qualité

Principalement pour les problèmes d'optimisation ou d'apprentissages.

## La qualité

Principalement pour les problèmes d'optimisation ou d'apprentissages.

### Problèmes d'optimisations

## La qualité

Principalement pour les problèmes d'optimisation ou d'apprentissages.

### Problèmes d'optimisations

??

- Recherche exhaustive
- Approximation par recherche locale
- Approximation par bornes théoriques
- Approche *anytime*

## La qualité

Principalement pour les problèmes d'optimisation ou d'apprentissages.

### Problèmes d'optimisations

??

- Recherche exhaustive
- Approximation par recherche locale
- Approximation par bornes théoriques
- Approche *anytime*

### Modèles physiques

- Exactitude et Précision

## La qualité en ML

### Classificateur binaire

Le **recall** permet de savoir le **pourcentage de positifs bien prédit par le modèle**.

$$recall = \frac{\text{vrai positif}}{\text{vrai positif} + \text{faux négatif}}$$

		Classes réelles	
		Positive	Négative
Classes prédites	Positive	Vrai positif	Faux positif
	Négative	Faux négatif	Vrai négatif

## La qualité en ML

### Classificateur binaire

La **précision** permet de savoir le **nombre de prédictions positifs bien effectuées**.

$$precision = \frac{\text{vrai positif}}{\text{vrai positif} + \text{faux positif}}$$

		Classes réelles	
		Positive	Négative
Classes prédites	Positive	Vrai positif	Faux positif
	Négative	Faux négatif	Vrai négatif

## La qualité en ML

### Classificateur binaire

L'**accuracy** permet de savoir le **nombre d'instances de données correctement classées par rapport au nombre total d'instances de données**.

$$accuracy = \frac{\text{vrai positif} + \text{vrai négatif}}{\text{vrai positif} + \text{faux positif} + \text{vrai négatif} + \text{faux négatif}}$$

		Classes réelles	
		Positive	Négative
Classes prédites	Positive	Vrai positif	Faux positif
	Négative	Faux négatif	Vrai négatif

## La qualité en ML

### Classificateur binaire

L'**F1 score** est la **moyenne harmonique** de la **précision** et du *recall*.

$$F1 \text{ score} = 2 \times \frac{\textit{recall} \times \textit{precision}}{\textit{recall} + \textit{precision}}$$

		Classes réelles	
		Positive	Négative
Classes prédites	Positive	Vrai positif	Faux positif
	Négative	Faux négatif	Vrai négatif

## La qualité en ML

### Régression

*Mean Square Error* (MSE) :

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

*Root Mean Square Error* (RMSE) :

$$RMSE = \sqrt{MSE}$$

# Importance de l'analyse des performances

# Pourquoi analyser les performances ?

## Pourquoi analyser les performances ?

- Question de coût : prix des services cloud, électricité, matériel (CPU, GPU, TPU, etc.)

## Pourquoi analyser les performances ?

- Question de coût : prix des services cloud, électricité, matériel (CPU, GPU, TPU, etc.)
- Question d'économie d'énergie :

« Au niveau mondial, les data centers utilisent aujourd'hui **3** de l'électricité

POURCENT

produite et sa part devrait augmenter considérablement dans les années à venir. Pour

2020, elle était ainsi estimée à **650**, soit **une consommation supérieure à celle**

TÉRAWHATHEURES

**de la France.** »

- Commissariat à l'énergie atomique et aux énergies alternatives (CEA)

### Pourquoi analyser les performances ?

- Question de coût : prix des services cloud, électricité, matériel (CPU, GPU, TPU, etc.)
- Question d'économie d'énergie :

« Au niveau mondial, les data centers utilisent aujourd'hui **3** de l'électricité

POURCENT

produite et sa part devrait augmenter considérablement dans les années à venir. Pour

2020, elle était ainsi estimée à **650**, soit **une consommation supérieure à celle**

TÉRAWHATHEURES

**de la France.** »

– Commissariat à l'énergie atomique et aux énergies alternatives (CEA)

- Question de puissance disponible & évolution : taille des gravures **5 nanomètres** !

## Pourquoi analyser les performances ?

- Question de coût : prix des services cloud, électricité, matériel (CPU, GPU, TPU, etc.)
- Question d'économie d'énergie :

« Au niveau mondial, les data centers utilisent aujourd'hui **3** de l'électricité

POURCENT

produite et sa part devrait augmenter considérablement dans les années à venir. Pour

2020, elle était ainsi estimée à **650**, soit **une consommation supérieure à celle**

TÉRAWHATHEURES

**de la France.** »

– Commissariat à l'énergie atomique et aux énergies alternatives (CEA)

- Question de puissance disponible & évolution : taille des gravures **5 nanomètres** !
- IoT, embarqué : contraintes techniques

# Pourquoi analyser les performances ?

## Pourquoi analyser les performances ?

- Pour les modèles : exactitude et précision
  - *e.g* : modèle médical, faut-il opérer un patient ?

## Pourquoi analyser les performances ?

- Pour les modèles : exactitude et précision
  - *e.g* : modèle médical, faut-il opérer un patient ?
- Modèles économiques
  - impact possible sur la politique des états

## Pourquoi analyser les performances ?

- Pour les modèles : exactitude et précision
  - *e.g* : modèle médical, faut-il opérer un patient ?
- Modèles économiques
  - impact possible sur la politique des états
- Modèles physiques
  - compréhension des processus physiques

## Théorie vs Pratique



TP 1 - Partie 1

Choix des facteurs, des  
points de conception et  
des mesures de  
performance

**Facteurs** : choisir les paramètres les plus influençables

- Exploiter la littérature, les connaissances sur les algorithmes et les expérimentations exploratoires

**Niveaux** : identifier les valeurs pertinentes pour chaque facteurs

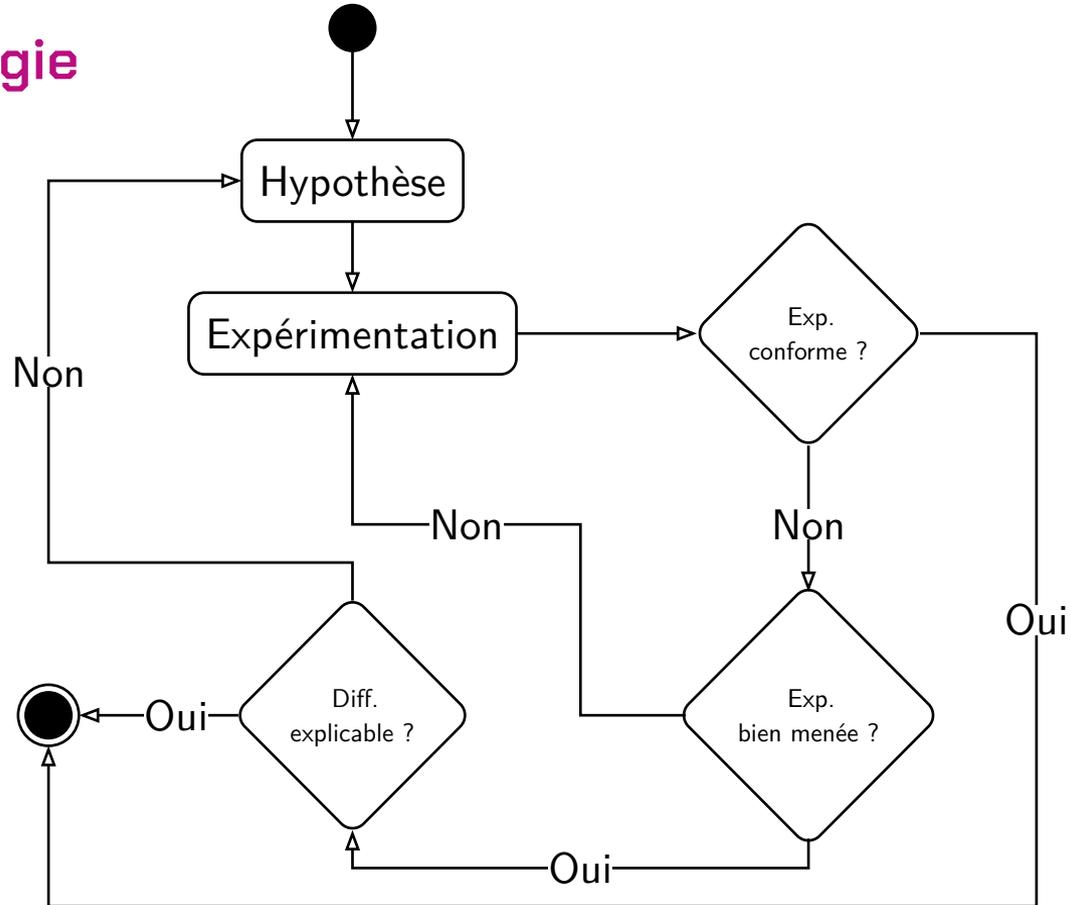
- Facteur symbolique : 1 niveau par valeur
- Facteur numérique :
  - Identifier les intervalles de valeurs pertinentes
  - Échantillonner avec une progression exponentielle
    - 1, 2, 4, 8, 16 ou 1, 10, 100, 1000, ...

### Points de conception :

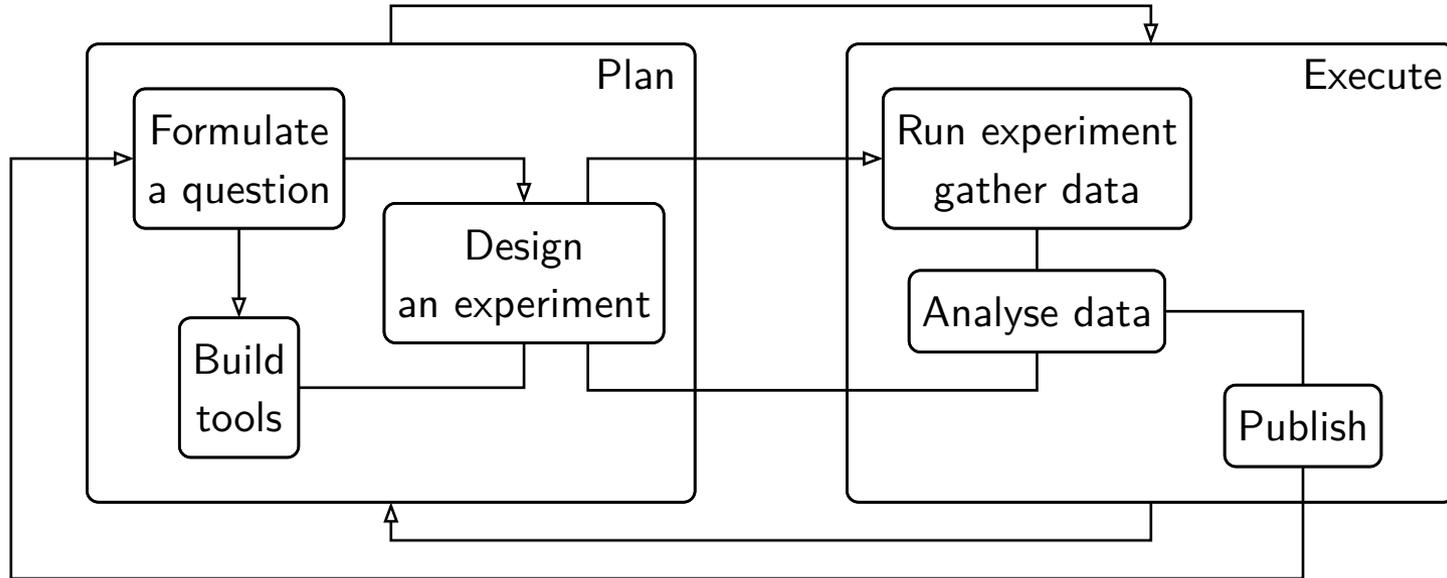
- plan factoriel complet = toutes les combinaisons de facteurs/niveaux possibles
  - Pour: identifie tous les effets des facteurs, en incluant des effets dûs à l'intépendances des facteurs ;
  - Contre: nombre exponentiel des combinaisons au regard du nombre de facteurs
- plan factoriel fractionnel = sélection d'un sous-ensemble de combinaisons
  - il faut choisir un sous-ensemble représentatif et complémentaire de combinaisons

# Principes de conception expérimentale

# Epistémologie



## The Experiment Process (McGeoch 2012)



# Pratique



TP 1 - Partie 2

## Bibliographie

McGeoch C C. 2012. *A Guide to Experimental Algorithmics*. Cambridge University Press