

Présentation du projet **DATALAB** du **LMRS**

Développement de démonstrateurs numériques
IA – mathématiques à visée industrielle

Corentin Lothodé

LMRS

Le 12 mai 2022 à Saint-Étienne du Rouvray



Plan

- 1 Contexte et objectifs
- 2 Description détaillée
 - Description des tâches
 - Description des moyens
- 3 Travaux en cours
 - Méthode PINN
 - Semi-Markov et Drift-Markov
 - Calibration de modèle en finance
- 4 Perspectives

Plan

- 1 Contexte et objectifs
- 2 Description détaillée
 - Description des tâches
 - Description des moyens
- 3 Travaux en cours
 - Méthode PINN
 - Semi-Markov et Drift-Markov
 - Calibration de modèle en finance
- 4 Perspectives

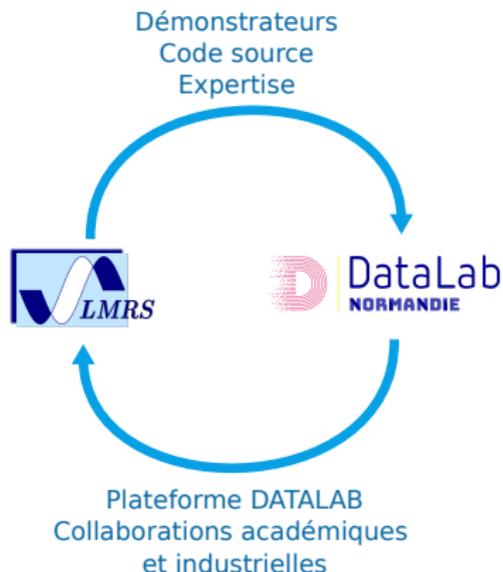
Contexte

- Participation du LMRS au DATALAB depuis sa création en 2019.
- La donnée et l'IA s'inscrivent dans la recherche du laboratoire :
 - forte expertise en **sciences du numérique**,
 - groupe spécialisé en **calcul scientifique**,
 - équipe **Statistique**.



Objectifs

- **Développer les compétences** internes du laboratoire en IA.
- Réaliser deux **démonstrateurs numériques**.
- **Mise à disposition** des démonstrateurs.



Plan

- 1 Contexte et objectifs
- 2 Description détaillée
 - Description des tâches
 - Description des moyens
- 3 Travaux en cours
 - Méthode PINN
 - Semi-Markov et Drift-Markov
 - Calibration de modèle en finance
- 4 Perspectives

Plan

- 1 Contexte et objectifs
- 2 Description détaillée
 - Description des tâches
 - Description des moyens
- 3 Travaux en cours
 - Méthode PINN
 - Semi-Markov et Drift-Markov
 - Calibration de modèle en finance
- 4 Perspectives

Développement de démonstrateurs numériques IA – mathématiques à visée industrielle

Tâche 1

Responsables :

- Ionut Danaila
- Corentin Lothodé
- Francky Luddens

Démonstrateur appliqué au calcul scientifique IA et EDP^a.

^aÉquations aux dérivées partielles

Tâche 2

Responsables :

- Vlad Barbu
- Mohamed Ben Alaya
- Nicolas Vergne

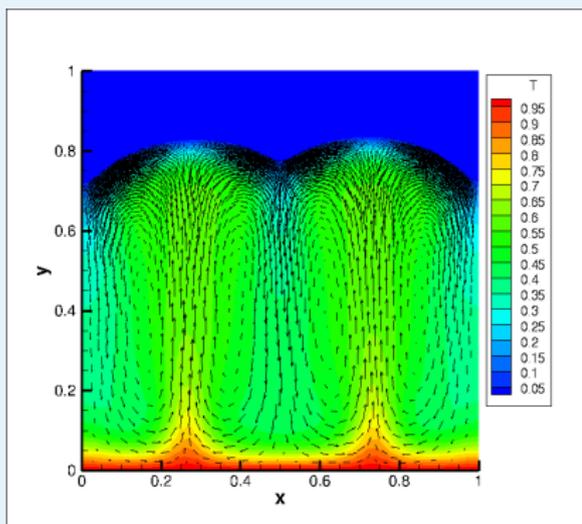
Démonstrateur appliqué aux statistiques IA et processus de Markov.

Tâche 1 : Démonstrateur EDP

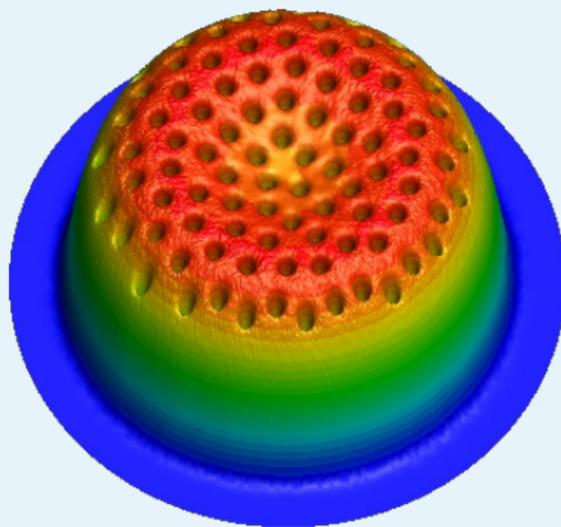
- **Réseaux neuronaux** informés par la **physique** (PINN)
- **Développement** de la méthode PINN pour des applications :
 - Matériaux à changement de phase (MCP),
 - Condensats de Bose-Einstein et hélium superfluide.
- **Intégration** dans le DATALAB et publication en **open source**.

Tâche 1 : Exemples d'applications

Chauffage de MCP



Condensat de Bose-Einstein



Tâche 1 : Exemples d'applications industrielles

Ex. : Datacenter



Applications

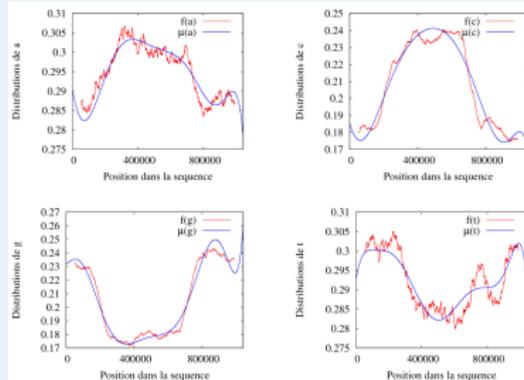
- *MCP* : régulation thermique passive
 - **BTP** : Inclusion dans les bâtiments, réduction chauffage / climatisation,
 - **Électronique** : Amélioration des cycles processeurs.
- *Superfluides et supraconducteurs* :
 - Refroidissement extrême pour les **réacteurs, accélérateurs,**
 - **Mobilité** : train à sustentation magnétique.

Tâche 2 : Démonstrateur Statistique

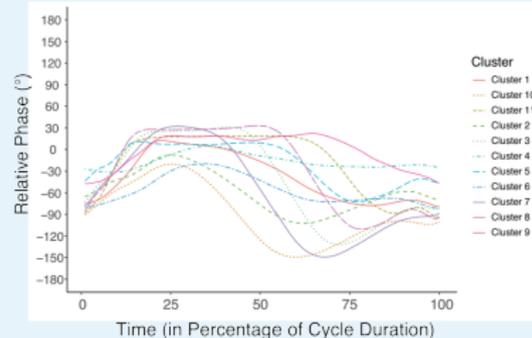
- Développement de **méthodes neuronales supervisées** pour la calibration de modèle en **finance**.
- Développement d'une **toolbox** Markov-Chain (Semi-Markov et Drift-Markov).

Tâche 2 : Exemples d'applications

Bio-statistique

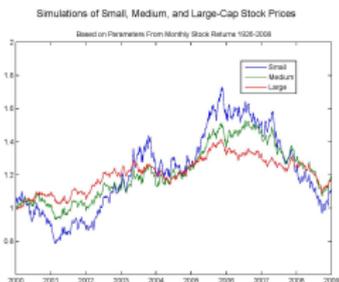


Clustering notation



Tâche 2 : Exemples d'applications industrielles

Marchés financiers



Applications

- Application **finance** avec la calibration de modèle.
- Aide à la recherche de marqueurs **génomiques**.
- Aide à la décision, par exemple en **sport** (JO Paris 2024).

Interaction entre les tâches

- Hybridation de la méthode neuronale appliquée à la finance avec la méthode PINN.
- Groupes de travail.
- Séminaires communs.

Plan

- 1 Contexte et objectifs
- 2 Description détaillée
 - Description des tâches
 - Description des moyens
- 3 Travaux en cours
 - Méthode PINN
 - Semi-Markov et Drift-Markov
 - Calibration de modèle en finance
- 4 Perspectives

Moyens humains

Tâche 1

- Ingénieur de recherche :
Mohammad Issa,
6 mois.
- Stagiaire Master 2 MAM^a :
Bahae-Eddine Madir,
6 mois.

^aModélisation et Analyse
Mathématiques

Tâche 2

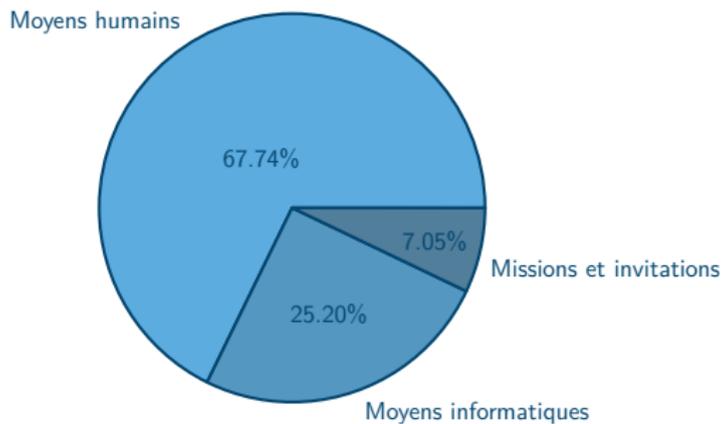
- Ingénieur de recherche :
Ioanis Mavrogiannis,
6 mois.
- Stagiaire Master 2 MFA^a :
Mohamad Ali Noureddine,
6 mois.

^aMathématiques Fondamentales et
Appliquées

Autres moyens

- Informatique :
 - 2 PC portables performants, adaptés au deep learning, pour les ingénieurs de recherche,
 - Serveur de calcul performant adapté au deep learning,
- Missions pour le personnel non-permanent.
- Professeurs invités spécialisés sur les questions d'IA.

Répartition



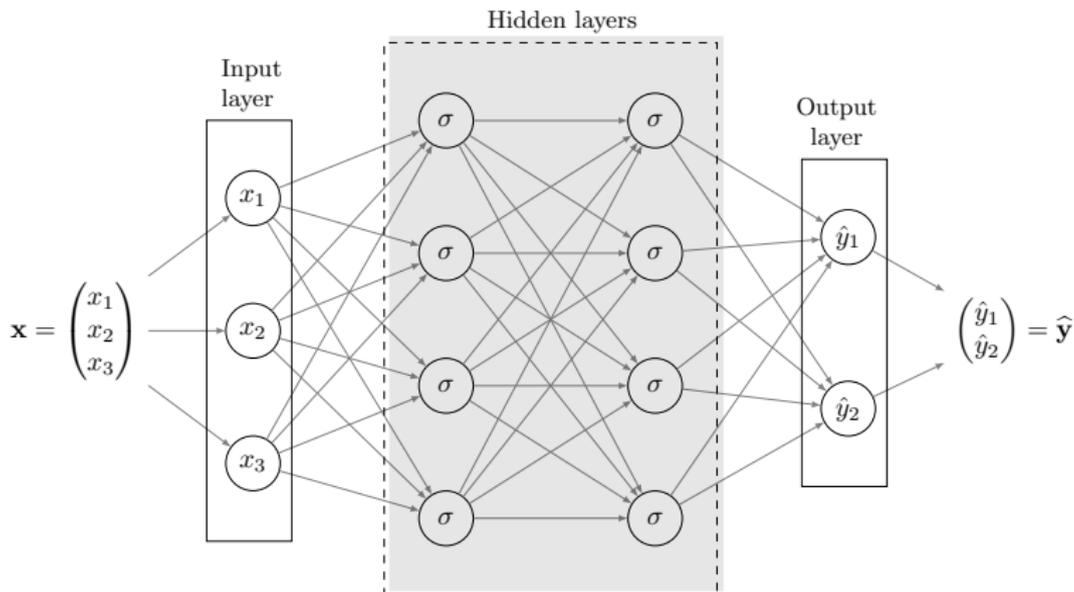
Plan

- 1 Contexte et objectifs
- 2 Description détaillée
 - Description des tâches
 - Description des moyens
- 3 Travaux en cours**
 - Méthode PINN
 - Semi-Markov et Drift-Markov
 - Calibration de modèle en finance
- 4 Perspectives

Plan

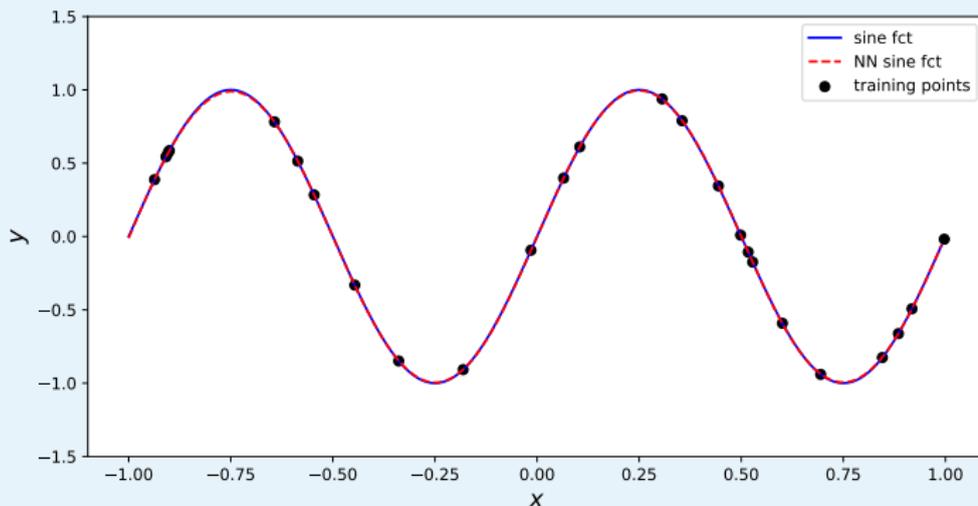
- 1 Contexte et objectifs
- 2 Description détaillée
 - Description des tâches
 - Description des moyens
- 3 Travaux en cours
 - Méthode PINN
 - Semi-Markov et Drift-Markov
 - Calibration de modèle en finance
- 4 Perspectives

Réseau de neurones classique I



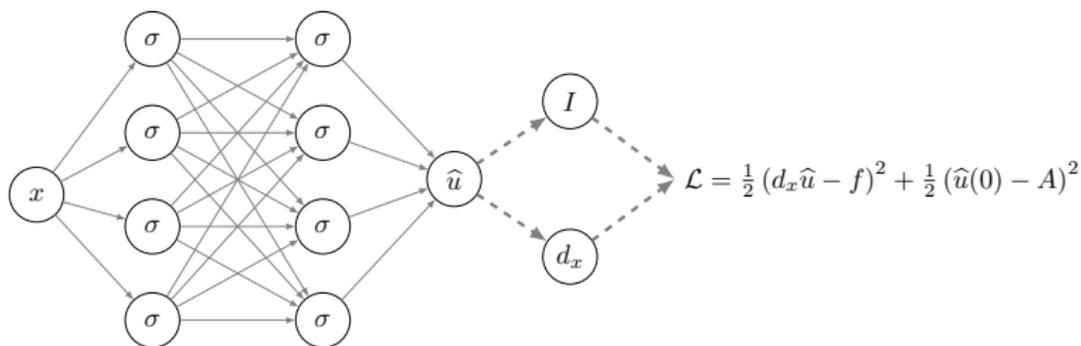
Réseau de neurones classique II

Exemple : $x \mapsto \sin(2\pi x)$ sur $[-1, 1]$



Réseau Physics Informed Neural Network I

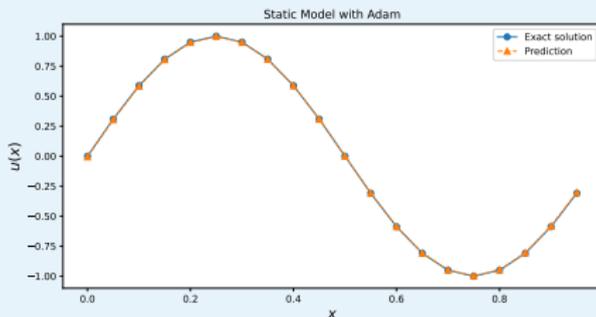
$$\begin{cases} u'(x) = f(x) \\ u(0) = A \end{cases}$$



Réseau Physics Informed Neural Network II

Exemple : équation de la chaleur et solution manufacturée

$$\begin{cases} u''(x) + 4\pi^2 \sin(2\pi x) = 0 & \text{dans }]0, 1[\\ u(0) = u(1) = 0 \end{cases}$$



Plan

- 1 Contexte et objectifs
- 2 Description détaillée
 - Description des tâches
 - Description des moyens
- 3 Travaux en cours
 - Méthode PINN
 - **Semi-Markov et Drift-Markov**
 - Calibration de modèle en finance
- 4 Perspectives

Modèles stochastiques avancés

L'intérêt/idée de base:

- Fournir à la communauté scientifique des **outils automatisés** pour la prise en compte de l'aléatoire dans la modélisation, estimation, aide à la décision.
- Plusieurs challenges: prise en compte de la **hétérogénéité**, différents **types de dépendance**, événements **rares/extrêmes**
- Applications: **fiabilité des systèmes** (fiabilité, disponibilité, temps moyens de fonctionnement, taux de défaillance etc.), modélisation en **génomique**, finance, logistique, etc.

Quelques modèles stochastiques:

- Modèles de **Markov régulés/dérivants** (Drifting Markov models)
- Modèles **semi-markoviens**
- Modèles de **Markov cachés**, modèles **semi-markoviens cachés**

Logiciels dédiés

Contexte : Plusieurs packages R développés dans notre équipe

- Juillet 2017: R package SMM: Simulation and Estimation of Multi-State Discrete-Time Semi-Markov and Markov Models \
<https://cran.r-project.org/web/packages/SMM>
- Mars 2021 : R package smmR : Simulation, Estimation and Reliability of Semi-Markov Models \
<https://cran.r-project.org/web/packages/smmR/index.html>
- Mars 2021 : R package drimmR : Estimation, Simulation and Reliability of Drifting Markov Models\
<https://cran.r-project.org/web/packages/drimmR/index.html>

Projet Datalab et perspectives

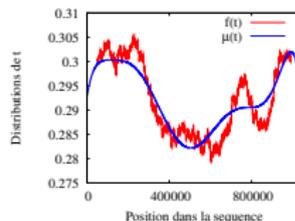
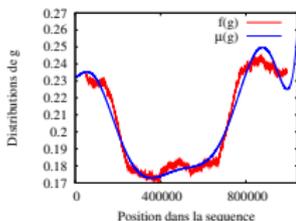
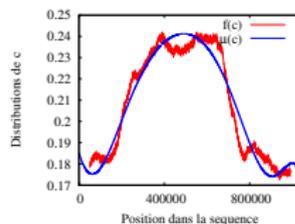
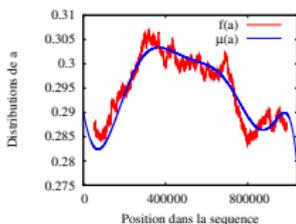
Dans le cadre du projet Datalab:

- Contrat de 6 mois de Iannis Mavrogiannis (Université de l'Égée, Grèce)
- Modèles **semi-markoviens dérivants**
- Construction d'un package R

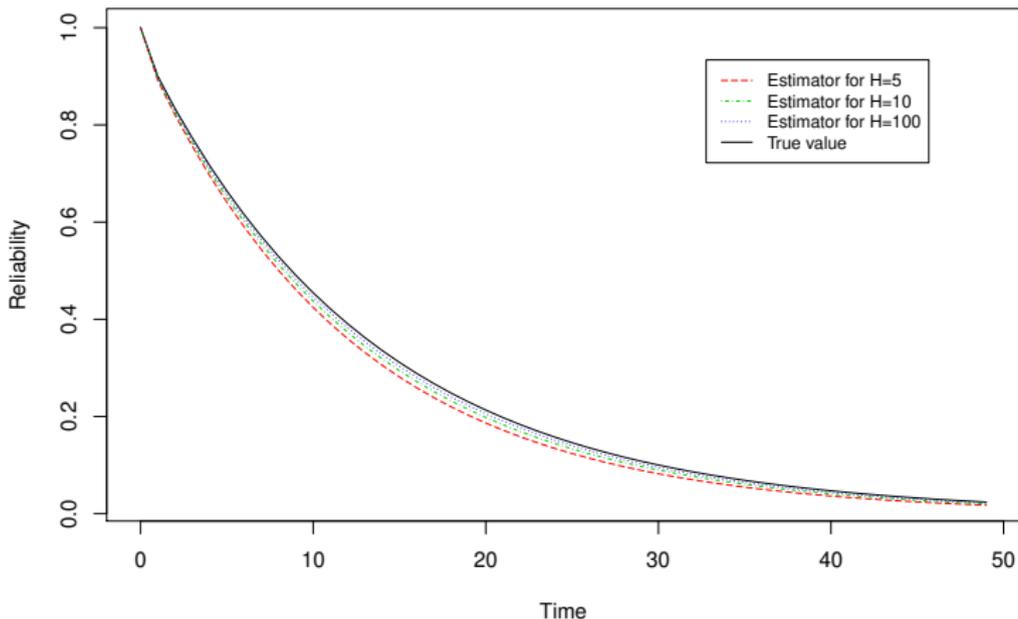
Perspectives:

- Modèles plus complexes et plus réalistes
- Modèles de type caché (HMM), très utilisés en **traitement du signal** (reconnaissance de la parole, de l'écriture, des images) et en **bioinformatique**
- Mise en pratique pour des applications réelles (fiabilité des systèmes, analyse de survie, bioinformatique)
- Construction des outils sous-jacents pour un large public

Fréquences / Lois stationnaires: modèles markoviens dérivants sur *Chlamydia trachomatis*



Fiabilité des systèmes markoviens dérivants



Time

Plan

- 1 Contexte et objectifs
- 2 Description détaillée
 - Description des tâches
 - Description des moyens
- 3 Travaux en cours**
 - Méthode PINN
 - Semi-Markov et Drift-Markov
 - Calibration de modèle en finance**
- 4 Perspectives

Modèle de Black-Scholes: Call option

- Dans le modèle de Black-Scholes, l'actif risqué satisfait l'EDS

$$dS_t = rS_t dt + \sigma S_t dW_t,$$

La solution S suit un mouvement brownien géométrique

$$S_t = S_0 \exp(\sigma W_t + (r - \sigma^2/2)t).$$

- Le prix de l'option call avec le payoff $(S_T - K)_+$ est

$$\begin{aligned} C &= e^{-rT} \mathbb{E}(S_T - K)_+ \\ &= S_0 \Phi\left(\frac{1}{\sigma\sqrt{T}}(\log(S_0/K) + (r + \sigma^2/2)T)\right) \\ &\quad - Ke^{-rT} \Phi\left(\frac{1}{\sigma\sqrt{T}}(\log(S_0/K) + (r - \sigma^2/2)T)\right), \end{aligned}$$

où $\Phi(x) = \mathbb{P}(G \leq x)$, $G \sim \mathcal{N}(0, 1)$.

Calcul du prix de l'option par la méthode de Monte Carlo

- L'approximation de $e^{-rT} \mathbb{E}(S_T - K)_+$ est

$$\hat{C} = \frac{e^{-rT}}{N} \sum_{i=1}^N (S_{T,i} - K)_+$$

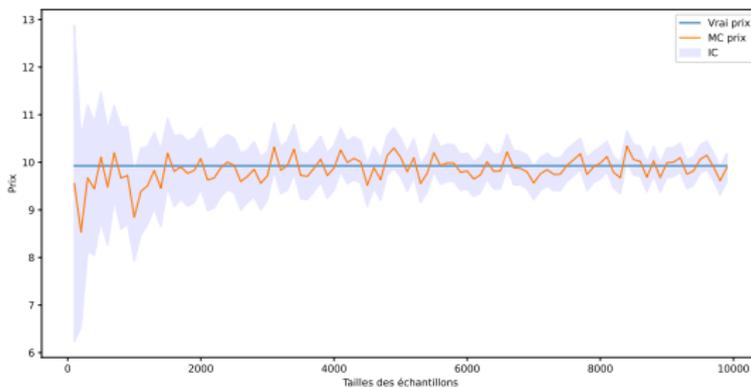
où $(S_{T,i})_{1 \leq i \leq N}$ est une suite de v.a. i.i.d de même loi que S_T .

- Un intervalle de confiance au risque de 5% pour $e^{-rT} \mathbb{E}(S_T - K)_+$ est

$$\left[\hat{C} - \frac{1.96 \hat{\sigma}_N}{\sqrt{N}}, \hat{C} + \frac{1.96 \hat{\sigma}_N}{\sqrt{N}} \right]$$

où $\hat{\sigma}_N$ est l'estimateur de l'écart-type et on rappelle que $\mathbb{P}(|G| \leq 1.96) = 0.95$.

Convergence de la méthode Monte Carlo



Modèle d'Heston

- Dans le modèle d'Heston, l'actif risqué satisfait l'EDS

$$dS_t = rS_t dt + S_t \sqrt{V_t} dW_t,$$

et le carré de la volatilité suit un modèle Cox-Ingersoll-Ross (CIR)

$$dV_t = \kappa(\theta - V_t)dt + \sigma \sqrt{V_t} d\tilde{W}_t.$$

Les deux browniens sont de corrélation ρ et $V_{t=0} = V_0$.

- Il n'y a pas de solution explicite du modèle ni de formule fermée pour le prix de l'option.
- On peut utiliser des méthodes numériques pour calculer le prix des options (par exemple la méthode de Monte Carlo ou résolution d'une EDP, inversion de Fourier).

Calibration et Deep Learning

- Soit $\{\Pi_{T_k}^{K_k}\}$, $1 \leq k \leq M$, les prix observés sur le marché. On cherche les paramètres qui minimisent $\sum_{k=1}^M \|P_{T_k}^{K_k}(\kappa, \theta, \sigma, V_0) - \Pi_{T_k}^{K_k}\|^2$ où $P_{T_k}^{K_k}$ est le prix donné par le modèle.
- Le temps de calcul est très couteux (plusieurs Monte Carlo).
- Utilisation du Deep Learning par reseaux de neurones pour approcher la fonction $P_{T_k}^{K_k}$ notée $\tilde{P}_{T_k}^{K_k}$ et cela en entrainant le système pour plusieurs valeurs de K et T .
- Puis chercher les paramètres qui minimisent la fonction

$$\sum_{k=1}^M \|\tilde{P}_{T_k}^{K_k}(\kappa, \theta, \sigma, V_0) - \Pi_{T_k}^{K_k}\|^2$$

Plan

- 1 Contexte et objectifs
- 2 Description détaillée
 - Description des tâches
 - Description des moyens
- 3 Travaux en cours
 - Méthode PINN
 - Semi-Markov et Drift-Markov
 - Calibration de modèle en finance
- 4 Perspectives

Perspectives

Avec la région :

- Projet LMRS DATALAB en 2023-?.
- Thèse financée ou co-financée par la région.
- Chaire d'excellence, chaire professeur junior.

Avec d'autres partenaires :

- Appels à projet avec l'ANR.
- Projets européens.



Merci !

Merci beaucoup pour votre attention !

