Generative Modeling for Finance

Résumé du séminaire

Alexis VO

CMAP

École polytechnique

11 juin 2025

Introduction

Dans le cadre de mon stage, j'ai eu l'opportunité de pouvoir suivre – en anglais – un séminaire "groupe de travail" en ce mercredi 11 juin à 11h00. Ce document est un bref résumé de ce que j'ai pu comprendre. On étudie l'application des modèles génératifs en finance, notamment via les méthodes de *flow matching* et *conditional flow matching*. L'objectif global est de transformer une distribution de probabilité connue p (source simulée) en une distribution cible q (observée), grâce à des transformations différentiables paramétrées.

Contents

1 Premier exposé : Flow matching et modélisation de densité

1.1 Cadre général

Soit p une distribution source sur $\mathbb{R}^d$, que l'on sait simuler.
Soit q une distribution de données sur $\mathbb{R}^d$, à partir de laquelle on observe des échantillons $(x^i)_{i=1}^N$.
On veut approximer q à partir de p en construisant un transformateur différentiable $\varphi_\theta : \mathbb{R}^d \to \mathbb{R}^d$, tel que la loi de $\varphi_\theta(X_0)$, où $X_0 \sim p$, soit proche de q. On note p^θ la loi induite.

1.2 Objectif

Minimiser la divergence de Kullback-Leibler :

$$\mathcal{L}(\theta) = \mathrm{KL}(q \| p^\theta) = \mathbb{E}_{x\sim q}\left[\log \frac{q(x)}{p^\theta(x)}\right] = -\mathbb{E}_{x\sim q}\left[\log p^\theta(x)\right] + \text{const.}$$

En pratique :

$$\mathcal{L}_N(\theta) = -\frac{1}{N}\sum_{i=1}^N \log p^\theta(x^i).$$

1.3 Modélisation par flows

On cherche à modéliser la densité $p^\theta(x)$ à l'aide de transformations différentiables inversibles (normalizing flows), avec :

$$p^\theta(x) = \left|\det\left(\frac{\partial \psi_\theta^{-1}(x)}{\partial x}\right)\right| p\left(\psi_\theta^{-1}(x)\right).$$

1.4 Approche dynamique par équation différentielle

On considère une famille de transformations ψ_t, telle que $\psi_0(x) = x$, définie par :

$$\frac{d}{dt}\psi_t(x) = u_t(\psi_t(x)),$$

où u_t est un champ de vecteurs (vitesse). La densité associée p_t évolue selon l'équation de conservation de la masse :

$$\frac{d}{dt}p_t(x) = -\nabla \cdot (u_t(x)p_t(x)).$$

D'où une formule d'évolution de la densité :

$$\log p_1(\psi_1(x)) = \log p(x) - \int_0^1 \nabla \cdot u_t(\psi_t(x))dt.$$

1.5 Implémentation

- Le champ de vitesse u_t^θ est paramétré par un réseau de neurones.
- On intègre numériquement l'équation différentielle (dans le sens inverse pour l'évaluation de la densité).
- Pour chaque donnée x^i, on estime $\log p^\theta(x^i)$ via l'équation ci-dessus.

2 Deuxième exposé : Conditional Flow Matching (CFM)

2.1 Motivation

On souhaite modéliser un flow conditionnel $\psi_t(x_0 \mid x_1)$, qui relie la source $X_0 \sim p$ à une cible $X_1 \sim q$, en conditionnant la trajectoire sur l'arrivée.

2.2 Évolution conditionnelle

Soit $P_t(x \mid x_1)$ la loi conditionnelle de X_t sachant $X_1 = x_1$. Elle satisfait aussi une équation de conservation :

$$\frac{d}{dt}P_t(x \mid x_1) + \nabla \cdot (P_t(x \mid x_1)u_t(x \mid x_1)) = 0.$$

Le champ de vitesse marginal $u_t(x)$ s'obtient par intégration :

$$u_t(x) = \int u_t(x \mid x_1)\frac{P_t(x \mid x_1)q(x_1)}{p_t(x)}dx_1.$$

2.3 Apprentissage par Flow Matching

Objectif : apprendre un champ de vitesse u_t^θ approchant u_t. On minimise la perte suivante :

$$\mathcal{L}_{\mathrm{FM}}(\theta) = \int_0^1 \mathbb{E}_{x_t}\left[\|u_t(x_t) - u_t^\theta(x_t)\|^2\right] dt.$$

En version conditionnelle :

$$\mathcal{L}_{\mathrm{CFM}}(\theta) = \int_0^1 \mathbb{E}_{x_t,x_1}\left[\|u_t(x_t \mid x_1) - u_t^\theta(x_t)\|^2\right] dt.$$

Théorème : Les gradients de $\mathcal{L}_{\mathrm{FM}}$ et $\mathcal{L}_{\mathrm{CFM}}$ par rapport à θ sont égaux.

3 Résultats et perspectives

Les méthodes de flow matching, basées sur des EDO et des approximations de champ de vitesse, permettent de transformer des distributions de manière efficace et stable. En finance, elles peuvent être utilisées pour générer des trajectoires réalistes de prix, calibrer des modèles de risque, ou simuler des dynamiques de marché sous contraintes. Des extensions incluent l'ajout de bruit brownien (diffusions stochastiques), des conditionnements complexes, et la régularisation géométrique (transport optimal).

Références que j'ai pu trouver :

- Lipman et al., *Flow Matching for Generative Modeling*, NeurIPS 2022.
- Tzen & Raginsky, *Neural Stochastic Differential Equations*, ICLR 2019.
- Papamakarios et al., *Normalizing Flows for Probabilistic Modeling and Inference*, JMLR 2021.