

Simulation et Monte Carlo

Marche aléatoire auto-évitante

Gabriel Romon

ENSAE

Plan

- 1 Introduction
- 2 Un premier algorithme de simulation
- 3 Simulation par la méthode de rejet
- 4 Estimation de la distance au carré entre point de départ et point d'arrivée
- 5 Simulation par pivot

Introduction

- Marche aléatoire dans \mathbb{Z}^d : suite $(x_k)_{k \in \mathbb{N}}$ de \mathbb{Z}^d telle que $x_0 = 0$ et $\forall k \in \mathbb{N}, \|x_{k+1} - x_k\|_2 = 1$
- Simulation : partir de 0 et choisir une direction parmi les $2d$ possibles au hasard
- Marche aléatoire auto-évitante dans \mathbb{Z}^d (**MAAE**) : marche aléatoire dont tous les éléments sont distincts
- Simulation : sujet de recherche depuis les années 50

Introduction

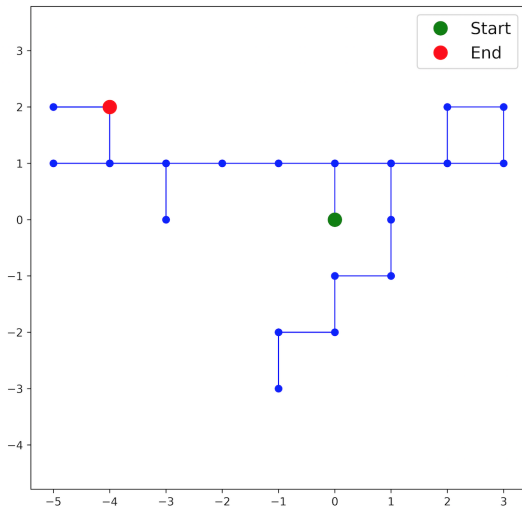


Figure 1 – Marche aléatoire simple de longueur 40

Introduction

- Soit $\text{MAAE}(n)$ l'ensemble des marches évitantes de longueur n
- $|\text{MAAE}(n)|$? Connu pour d et n petits.
Estimation asymptotique : $\sim A\mu^n n^{\gamma-1}$ où A, μ, γ dépendent de d
- Nombre de marches générales de longueur n : $(2d)^n$
Pour $d = 2$, proportion de MAAE de longueurs n :

$$\approx 0.65^n n^{0.34} \rightarrow 0 \quad \text{rapidement}$$

- Dans la suite, $d = 2$

Algorithme myope

- **Algorithme myope** : à chaque étape, l'algorithme choisit la position suivante uniformément parmi les voisins qui n'ont jamais été visités.
- Peu coûteux : complexité $O(n^2)$
- **Mais** : peut rester coincer
- **Mais** : MAAE renvoyées ne sont pas équiprobables

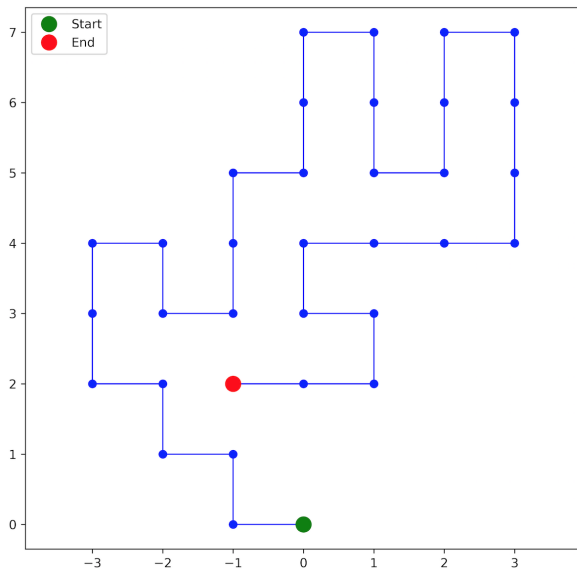


Figure 2 – Marche coincée à l'étape 23

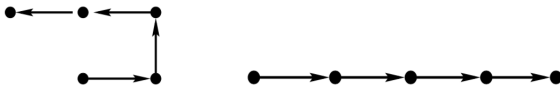


Figure 3 – 2 marches de longueur 4 non équiprobables

- à gauche : $\frac{1}{4} \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{2}$
- à droite : $\frac{1}{4} \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{3}$

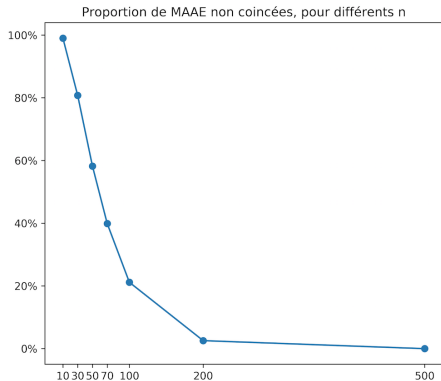
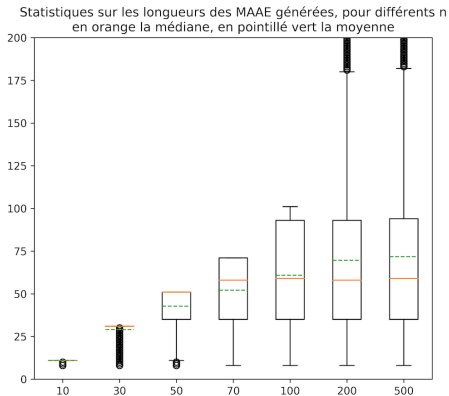


Figure 4 – Statistiques concernant les marches générées par l'algorithme myope, pour $10 \leq n \leq 500$

Méthode de rejet

- MAAE renvoyées sont équiprobables
- Probabilité d'acceptation : $\approx 0.65^n n^{0.34}$.
- Complexité exponentielle $\approx 1.53^n n^{-0.34}$, donc algorithme inefficace quand n grandit

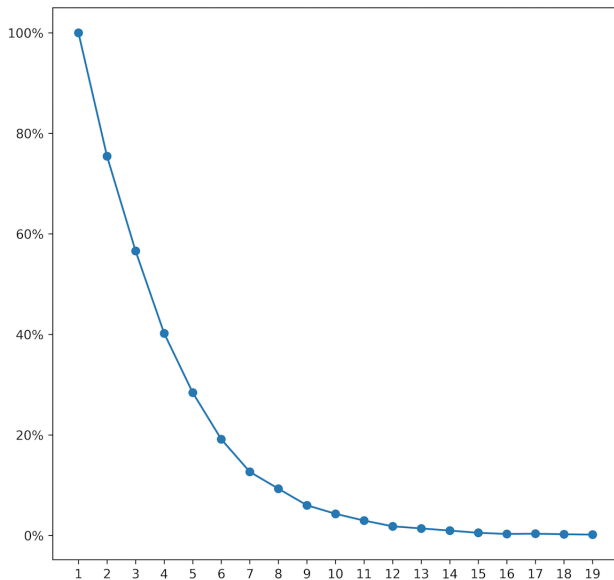


Figure 5 – Taux d'acceptation pour différents n

Importance sampling

- Loi cible : uniforme sur $\text{MAAE}(n)$, loi de proposition : celle induite par l'algorithme myope
- La loi cible est connue à une constante près (le nombre de MAAE), donc importance sampling pondéré est faisable
- Soit S une fonction qui prend en argument une MAAE. Pour estimer $S(X)$, $\hat{S} := \frac{\sum_{i=1}^m w(x_i) S(x_i)}{\sum_{i=1}^m w(x_i)}$ où $w(x_i)$ est l'inverse de la probabilité de la marche x_i
- \hat{S} est biaisé mais consistant

Importance sampling

- $S(X) :=$ distance au carré entre point de départ et point d'arrivée de la marche (classique dans la littérature)
- Vraies valeurs moyennes sont connues. Avec 10.000 échantillons, erreur relative $\leq 2\%$
- Valeurs estimées varient significativement entre les exécutions, donc méthode instable

Importance sampling

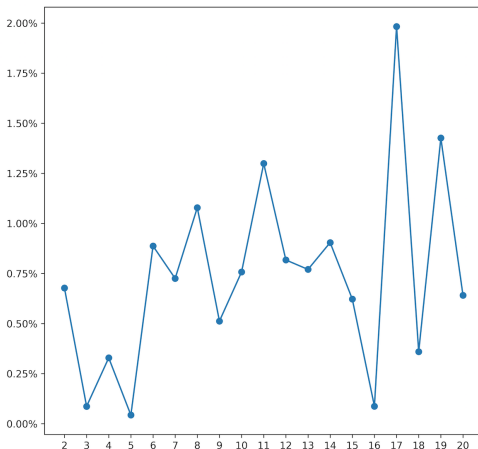


Figure 6 – Erreurs relatives pour $2 \leq n \leq 20$ avec un échantillon de 10.000 dans chaque cas

- Méthode récursive pour générer une MAAE. Pour $n \leq 3$, via algorithme myope . Pour $n \geq 4$, via deux appels récursifs (jusqu'à ce que la concaténation soit une MAAE)
- MAAE renvoyées sont équiprobables
- Complexité : $\Theta(n^{c_1 \log n + c_2})$ avec $c_1 \approx 0.17$ and $c_2 \approx 0.72$
- Méthode bien plus efficace que le rejet. Utilisée en pratique pour générer de très longues chaînes uniformément.

Dimérisation

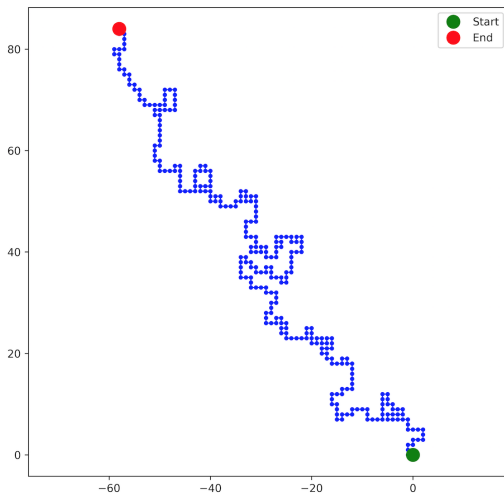


Figure 7 – MAAE de longueur 300 générée par dimérisation

- Construction d'une chaîne de Markov sur $\text{MAAE}(n)$
- La chaîne est irréductible et réversible pour la loi uniforme sur $\text{MAAE}(n)$. Cette loi est donc sa loi stationnaire
- Principe : partir d'une MAAE, couper en deux au hasard et appliquer une transformation du groupe diédral au deuxième bout, puis concaténer. Répéter jusqu'à obtenir une MAAE.
- Initialisation : soit par dimérisation (pas besoin de burn-in), soit avec un bâton (besoin de burn-in, supprimer $\approx 10 \cdot n^{1.19}$ premières observations)

Pivot : dimérization

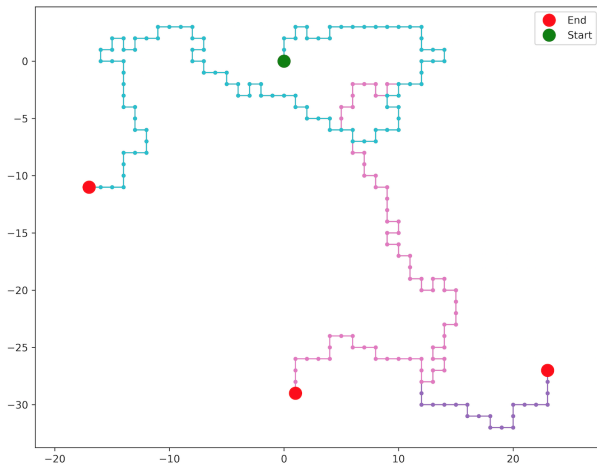


Figure 8 – 20 MAAE de longueur 100 générées par pivot, initialisation par dimérization

Pivot : bâton

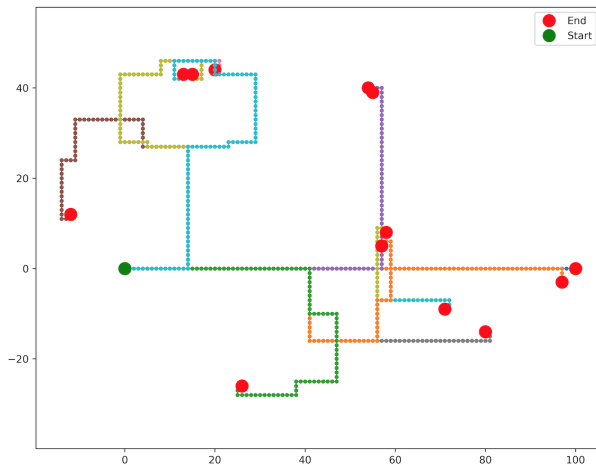


Figure 9 – 20 MAAE de longueur 100 générées par pivot, initialisation par un bâton horizontal

Pivot pour l'estimation

- Théorème ergodique justifie l'utilisation de la moyenne empirique
- Avec 10.000 échantillons, erreur relative moins bonne que l'importance sampling ($\approx 4\%$)
- Avec 200.000 échantillons, erreur relative $\leq 1\%$

Pivot pour l'estimation

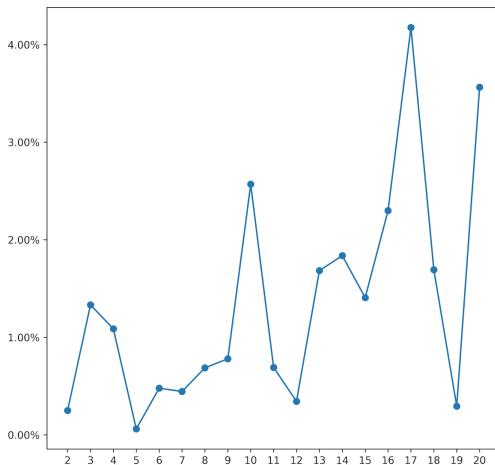


Figure 10 – Erreurs relatives pour $2 \leq n \leq 20$ avec un échantillon de 10.000 dans chaque cas