

Monte Carlo, Sudoku et chaînes de Markov

Rami El Haddad
rami.haddad@usj.edu.lb



- 1 Méthodes de Monte Carlo
 - Approche intuitive
 - Fondements mathématiques
- 2 Stratification
 - Avantages d'une stratification
 - Types de stratification
 - Stratification Sudoku
- 3 Application aux chaînes de Markov
- 4 Simulations numériques
 - Évolution d'un écosystème méditerranéen
 - Occurrences dans un brin d'ADN

- 1 Méthodes de Monte Carlo
 - Approche intuitive
 - Fondements mathématiques
- 2 Stratification
 - Avantages d'une stratification
 - Types de stratification
 - Stratification Sudoku
- 3 Application aux chaînes de Markov
- 4 Simulations numériques
 - Évolution d'un écosystème méditerranéen
 - Occurrences dans un brin d'ADN

Méthodes de Monte Carlo : approche intuitive

La **modélisation mathématique** consiste à représenter un problème complexe par un modèle abstrait pour ensuite l'analyser sans devoir travailler sur un système réel coûteux.

La **simulation** désigne l'exécution d'un programme informatique sur un ordinateur en vue d'imiter artificiellement le phénomène physique.

Une simulation utilisant des nombres aléatoires est dite **de Monte Carlo**.



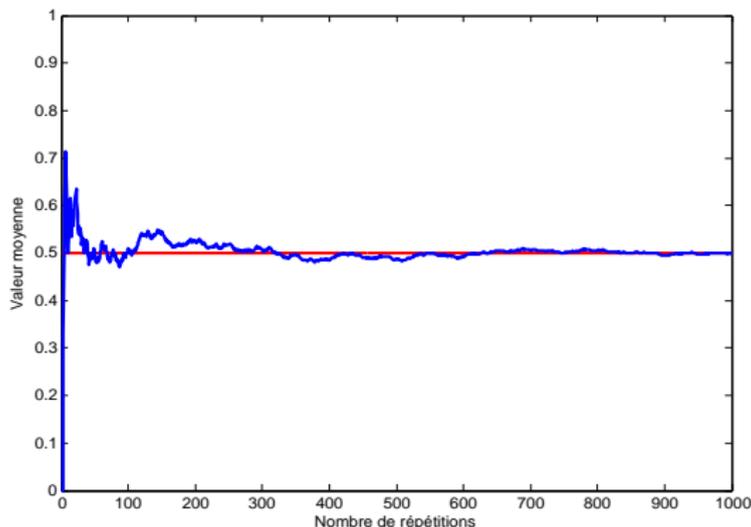
Méthodes de Monte Carlo : approche intuitive

Loi des grands nombres : quand le nombre d'essais augmente, la fréquence observée s'approche de la probabilité théorique.



Méthodes de Monte Carlo : approche intuitive

Loi des grands nombres : quand le nombre d'essais augmente, la fréquence observée s'approche de la probabilité théorique.



Fréquences d'apparition de *Face* en lançant une pièce de monnaie

Nombres aléatoires

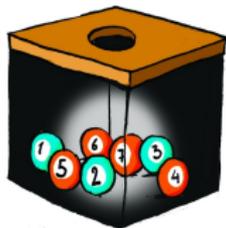
Nombres aléatoires = valeurs réelles ayant les propriétés statistiques de la variable aléatoire et produites par le hasard.

Génération de nombres aléatoires

Procédés physiques



Vrais nombres aléatoires



Calculateurs électroniques

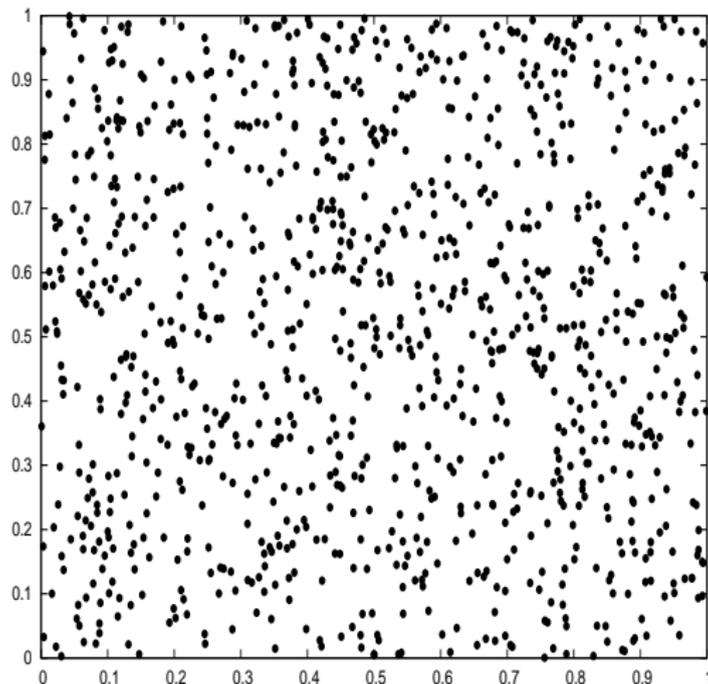


Nombres *pseudo*-aléatoires



Tous les langages de programmation disposent d'un générateur de nombres pseudo-aléatoires : ran, rand, grand, Random,...

Nombres aléatoires



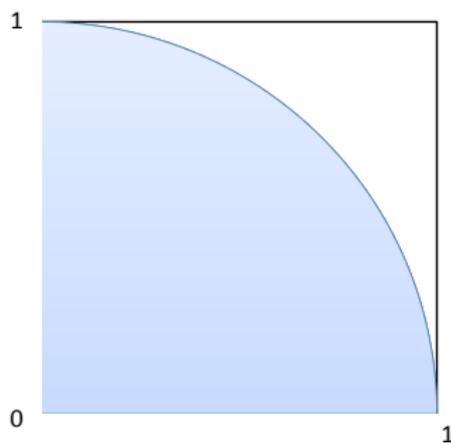
1 024 points pseudo-aléatoires dans $[0, 1]^2$.

Estimation de π

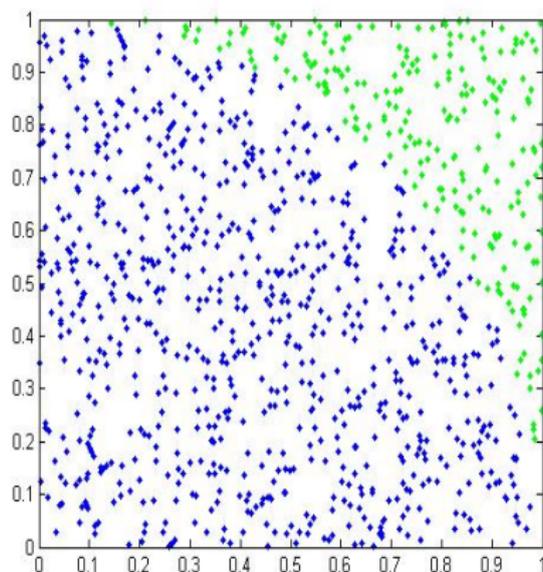
- ▶ On considère le quart du disque de rayon 1 inscrit dans le carré $[0, 1]^2$:

$$\frac{\text{Aire}(\text{quart du disque})}{\text{Aire}(\text{carré})} = \frac{\pi}{4}$$

- ▶ On choisit au hasard N points aléatoires (x, y) dans le carré et on calcule la proportion de ceux qui tombent dans le quart du disque, c.-à-d. tels que $x^2 + y^2 \leq 1$.

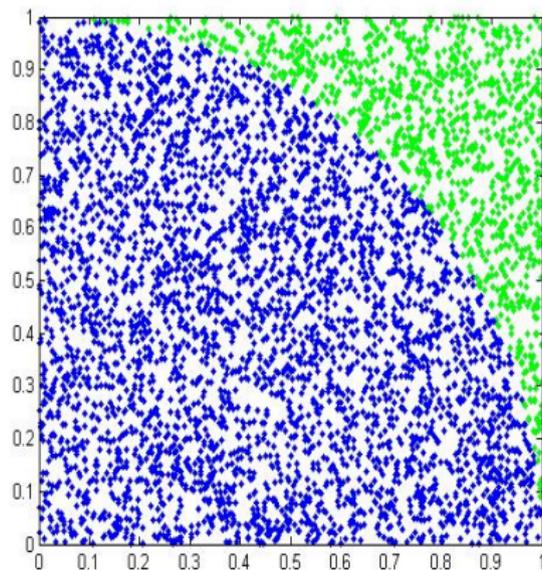


Estimation de π



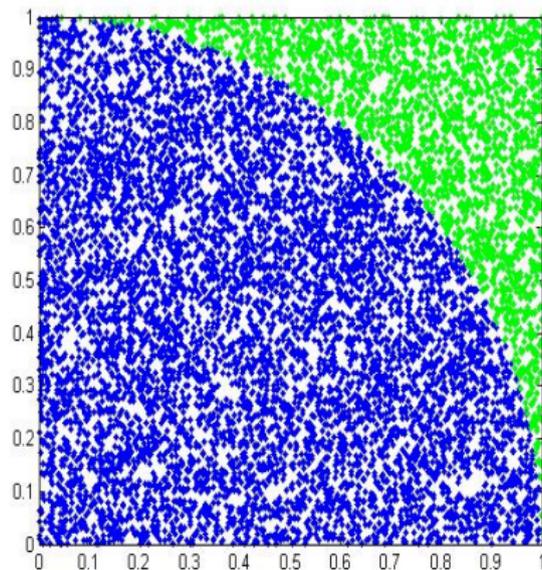
Pour $N = 1\,000$ on obtient $\pi \approx 3,2000$.

Estimation de π



Pour $N = 5\,000$ on obtient $\pi \approx 3,1552$.

Estimation de π



Pour $N = 10\,000$ on obtient $\pi \approx 3,1476$.

Soit

- $I := [0, 1)$,
- $s \geq 2$ un entier,
- $f : I^s \rightarrow \mathbb{R}$ une fonction de carré intégrable.

On veut estimer

$$\mathcal{J} := \int_{I^s} f(x) dx.$$

Si U est une variable aléatoire de loi uniforme sur I^s , alors

$$\mathcal{J} = \mathbb{E}[f(U)].$$

- $\{U_1, \dots, U_N\}$ des v. a. indépendantes et de loi uniforme sur I^s ,
Loi des grands nombres :

$$\mathbb{P}\left(\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{\ell=1}^N f \circ U_\ell = \mathcal{J}\right) = 1.$$

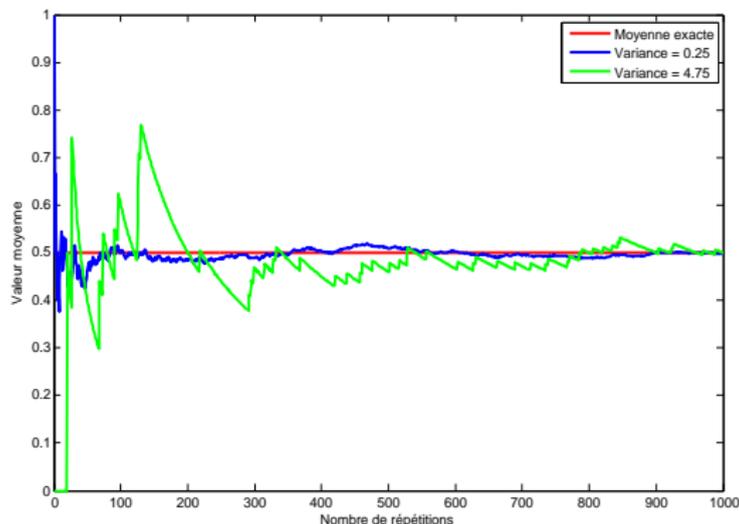
- Estimateur de Monte Carlo de \mathcal{J} :

$$\bar{X}_N := \frac{1}{N} \sum_{\ell=1}^N f(U_\ell).$$

$$\mathbb{E}[\bar{X}_N] = \mathcal{J} \quad \text{et} \quad \text{Var}(\bar{X}_N) = \mathcal{O}\left(\frac{1}{N}\right).$$

Effet de la variance d'un estimateur de Monte Carlo

L'espérance (ou moyenne) seule ne suffit pas !!! Il est aussi important que la variance de l'estimateur soit faible.

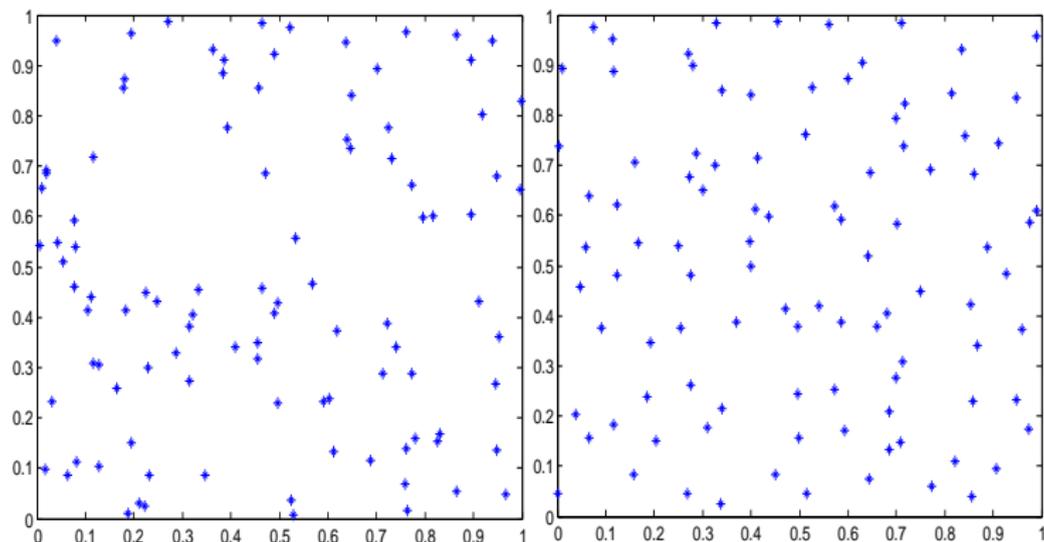


Fréquences d'apparition de *Face* en lançant une pièce de monnaie

- 1 Méthodes de Monte Carlo
 - Approche intuitive
 - Fondements mathématiques
- 2 Stratification
 - Avantages d'une stratification
 - Types de stratification
 - Stratification Sudoku
- 3 Application aux chaînes de Markov
- 4 Simulations numériques
 - Évolution d'un écosystème méditerranéen
 - Occurrences dans un brin d'ADN

Avantages d'une stratification

Objectif : construire des ensembles de points aléatoires ayant une meilleure répartition pour obtenir un estimateur à faible variance.

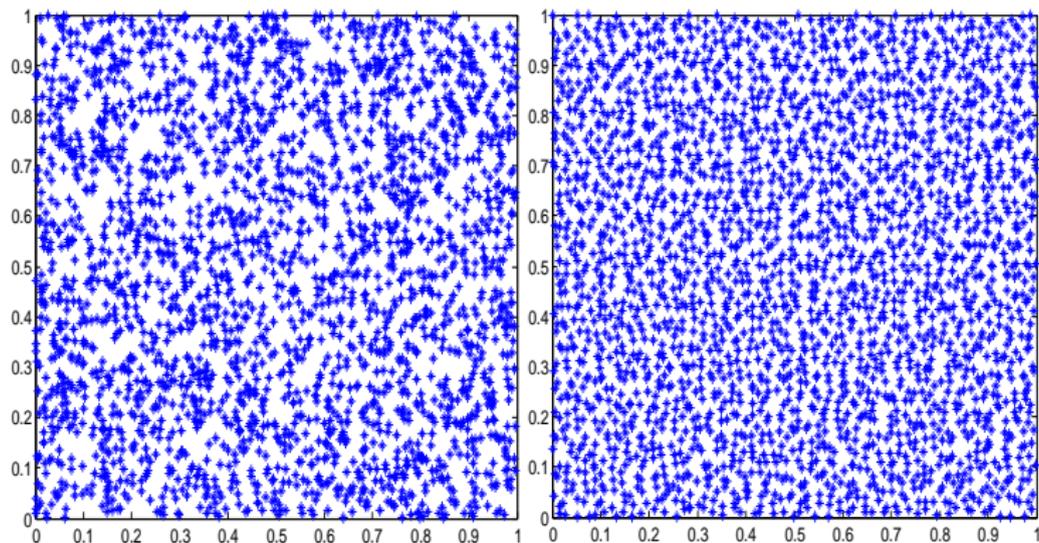


À gauche : ensemble de 100 points aléatoires.

À droite : ensemble stratifié de 100 points.

Avantages d'une stratification

Objectif : construire des ensembles de points aléatoires ayant une meilleure répartition pour obtenir un estimateur à faible variance.

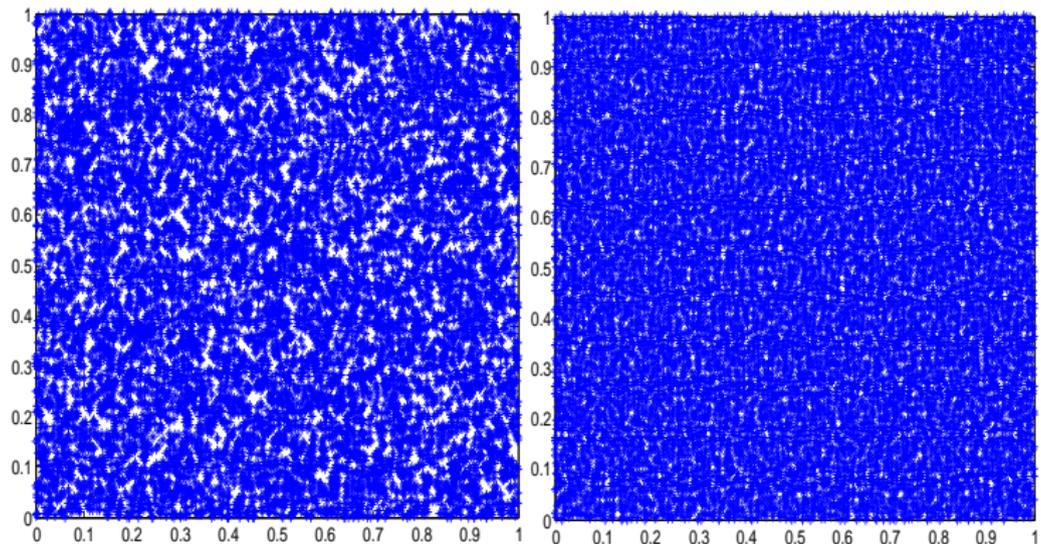


À gauche : ensemble de 2 500 points aléatoires.

À droite : ensemble stratifié de 2 500 points.

Avantages d'une stratification

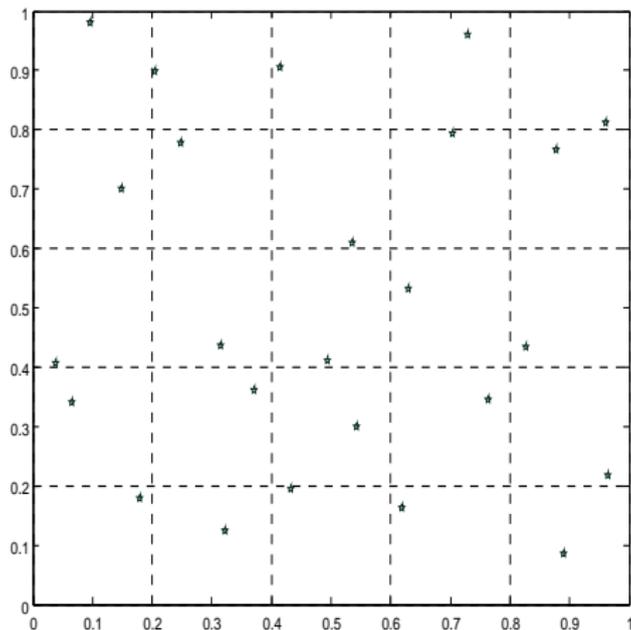
Objectif : construire des ensembles de points aléatoires ayant une meilleure répartition pour obtenir un estimateur à faible variance.



À gauche : ensemble de 10 000 points aléatoires.

À droite : ensemble stratifié de 10 000 points.

Stratification simple (MCS)



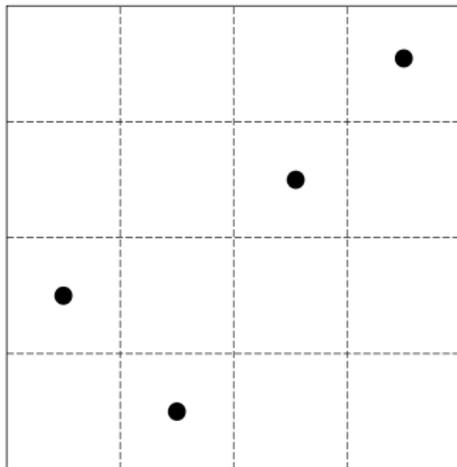
Échantillon MCS de 25 points.

Stratification par hypercubes latins (LHS)

| | | | |
|---|---|---|---|
| 4 | 1 | 2 | 3 |
| 1 | 2 | 3 | 4 |
| 3 | 4 | 1 | 2 |
| 2 | 3 | 4 | 1 |

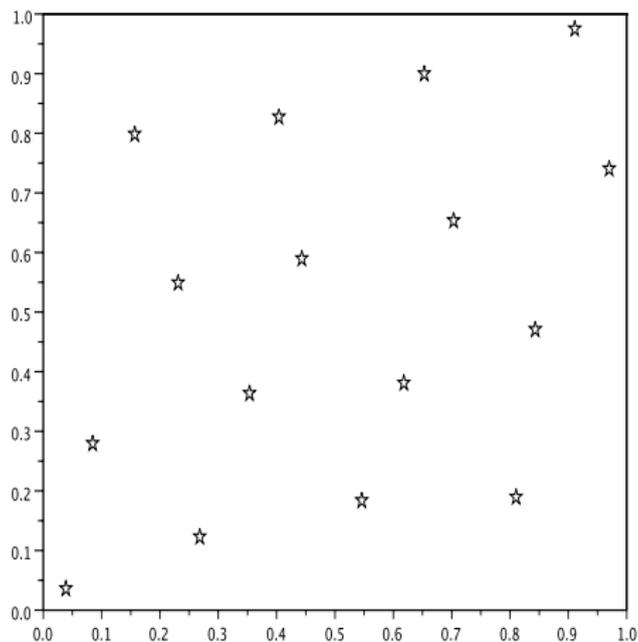
Carré latin 4×4 .

Stratification par hypercubes latins (LHS)



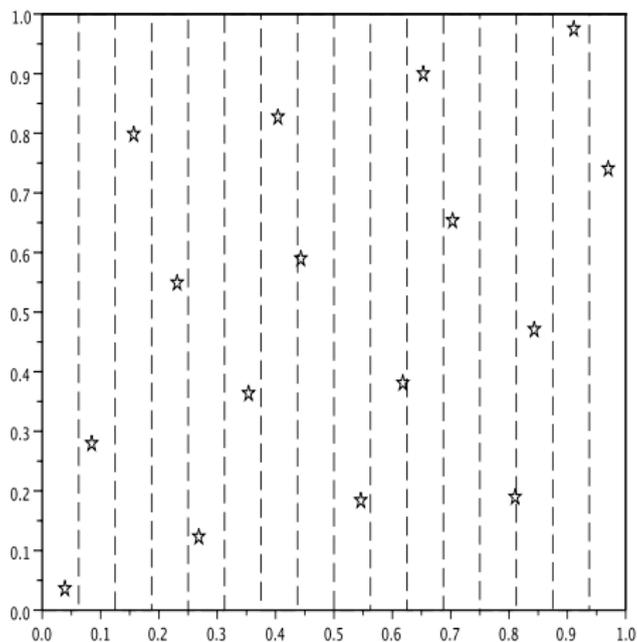
Échantillon LHS de 4 points.

Stratification par hypercubes latins (LHS)



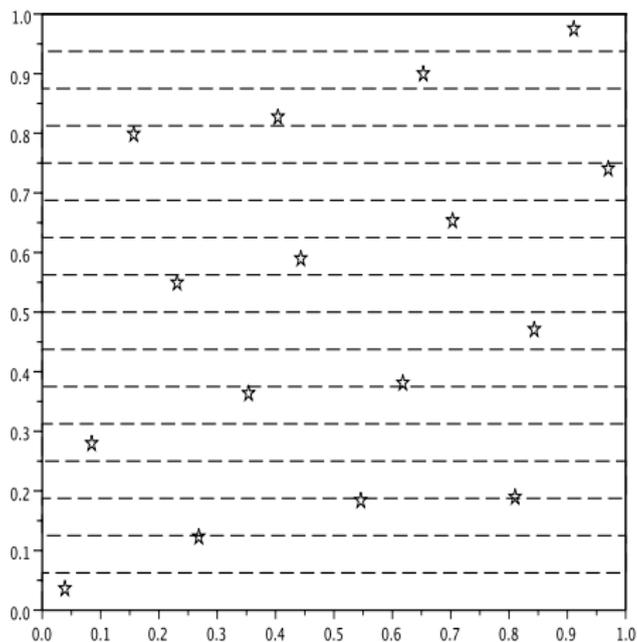
Échantillon LHS de 16 points.

Stratification par hypercubes latins (LHS)



Échantillon LHS de 16 points.

Stratification par hypercubes latins (LHS)



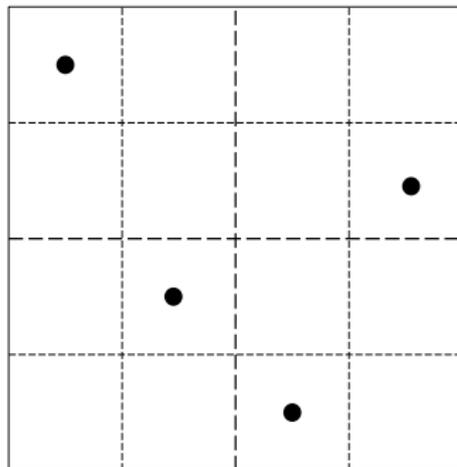
Échantillon LHS de 16 points.

Stratification Sudoku

| | | | |
|---|---|---|---|
| 2 | 1 | 4 | 3 |
| 3 | 4 | 1 | 2 |
| 1 | 2 | 3 | 4 |
| 4 | 3 | 2 | 1 |

Grille de Sudoku 4×4 .

Stratification Sudoku



Échantillon Sudoku de 4 points.

Échantillonnage Sudoku

Soit

- $N = n^s$

- $W_\ell := (W_\ell^1, \dots, W_\ell^s)$

- $W_\ell^i := \frac{\ell^i - 1}{n} + \frac{\sigma^i(\widehat{\ell}^i) - 1}{N} + \frac{U_\ell^i}{N}$

$$\ell := (\ell^1, \dots, \ell^s), 1 \leq \ell^i \leq n$$

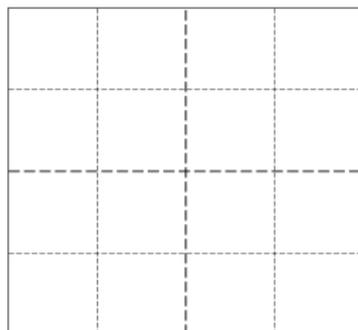
$$\widehat{\ell}^i := (\ell^1, \dots, \ell^{i-1}, \ell^{i+1}, \dots, \ell^s)$$

$\sigma^1, \dots, \sigma^s$ bijections aléatoires

indépendantes

$$\{1, \dots, n\}^{s-1} \rightarrow \{1, \dots, n^{s-1}\}$$

U^1, \dots, U^s variables aléatoires
indépendantes de loi uniforme sur
 $]0, 1[$.



Échantillonnage Sudoku

Soit

- $N = n^s$

- $W_\ell := (W_\ell^1, \dots, W_\ell^s)$

- $W_\ell^i := \frac{\ell^i - 1}{n} + \frac{\sigma^i(\widehat{\ell}^i) - 1}{N} + \frac{U_\ell^i}{N}$

$$\ell := (\ell^1, \dots, \ell^s), 1 \leq \ell^i \leq n$$

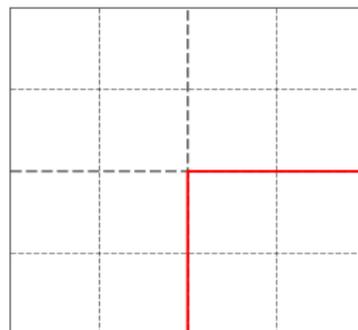
$$\widehat{\ell}^i := (\ell^1, \dots, \ell^{i-1}, \ell^{i+1}, \dots, \ell^s)$$

$\sigma^1, \dots, \sigma^s$ bijections aléatoires

indépendantes

$$\{1, \dots, n\}^{s-1} \rightarrow \{1, \dots, n^{s-1}\}$$

U^1, \dots, U^s variables aléatoires
indépendantes de loi uniforme sur
 $]0, 1[$.



Échantillonnage Sudoku

Soit

- $N = n^s$

- $W_\ell := (W_\ell^1, \dots, W_\ell^s)$

- $W_\ell^i := \frac{\ell^i - 1}{n} + \frac{\sigma^i(\widehat{\ell}^i) - 1}{N} + \frac{U_\ell^i}{N}$

$$\ell := (\ell^1, \dots, \ell^s), 1 \leq \ell^i \leq n$$

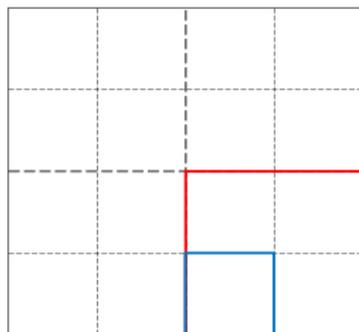
$$\widehat{\ell}^i := (\ell^1, \dots, \ell^{i-1}, \ell^{i+1}, \dots, \ell^s)$$

$\sigma^1, \dots, \sigma^s$ bijections aléatoires

indépendantes

$$\{1, \dots, n\}^{s-1} \rightarrow \{1, \dots, n^{s-1}\}$$

U^1, \dots, U^s variables aléatoires
indépendantes de loi uniforme sur
 $]0, 1[$.



Échantillonnage Sudoku

Soit

- $N = n^s$

- $W_\ell := (W_\ell^1, \dots, W_\ell^s)$

- $W_\ell^i := \frac{\ell^i - 1}{n} + \frac{\sigma^i(\widehat{\ell}^i) - 1}{N} + \frac{U_\ell^i}{N}$

$$\ell := (\ell^1, \dots, \ell^s), 1 \leq \ell^i \leq n$$

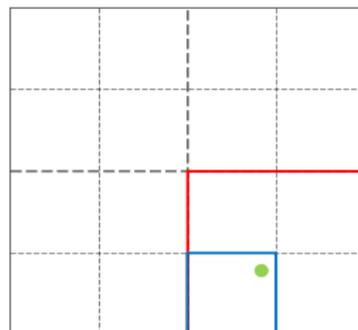
$$\widehat{\ell}^i := (\ell^1, \dots, \ell^{i-1}, \ell^{i+1}, \dots, \ell^s)$$

$\sigma^1, \dots, \sigma^s$ bijections aléatoires

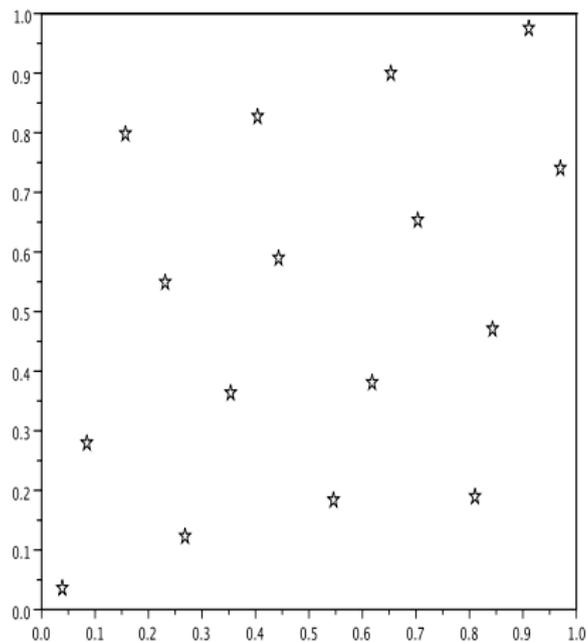
indépendantes

$$\{1, \dots, n\}^{s-1} \rightarrow \{1, \dots, n^{s-1}\}$$

U^1, \dots, U^s variables aléatoires
indépendantes de loi uniforme sur
 I^N .

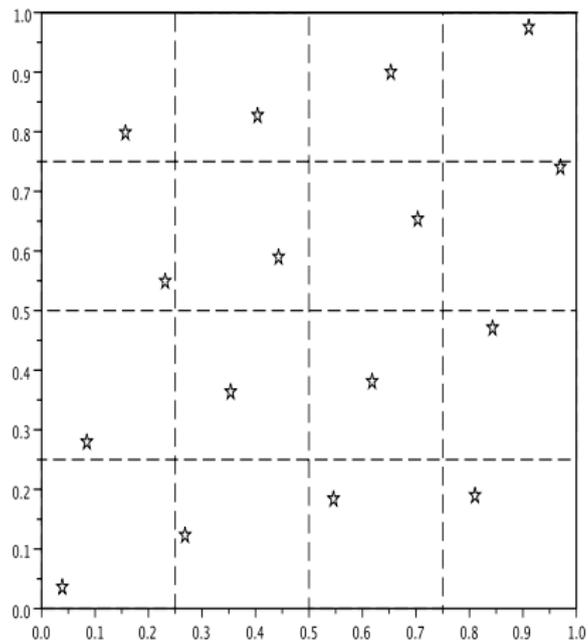


Échantillonnage Sudoku



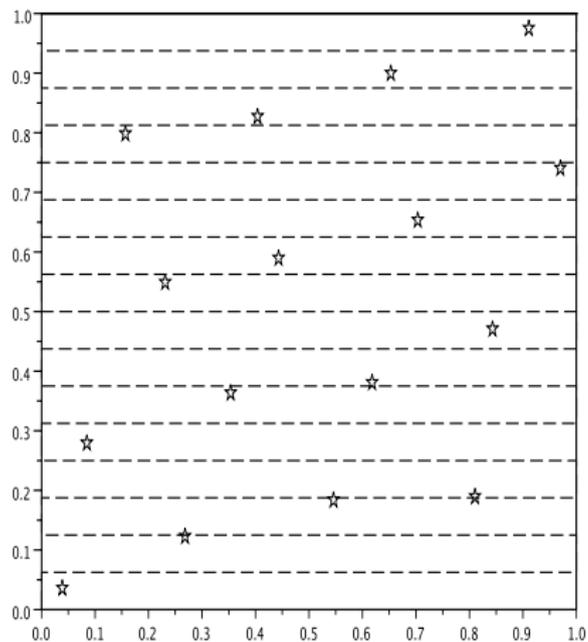
Échantillon Sudoku de 16 points.

Échantillonnage Sudoku



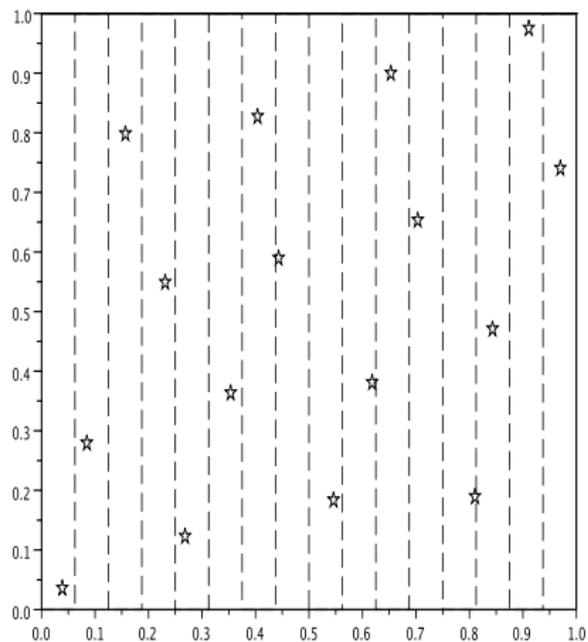
Échantillon Sudoku de 16 points.

Échantillonnage Sudoku



Échantillon Sudoku de 16 points.

Échantillonnage Sudoku



Échantillon Sudoku de 16 points.

Amélioration obtenue par la stratification Sudoku

L'estimateur de Monte Carlo classique \bar{X}_N :

- ▶ est sans biais, i.e. $\mathbb{E}[\bar{X}_N] = \mathcal{J}$;
- ▶ a une variance de l'ordre

$$\text{Var}(\bar{X}_N) = \mathcal{O}\left(\frac{1}{N}\right).$$

L'estimateur de Monte Carlo Sudoku \bar{Z}_N^* :

- ▶ est sans biais ;
- ▶ a une variance de l'ordre

$$\text{Var}(\bar{Z}_N^*) = \mathcal{O}\left(\frac{1}{N^{1+1/s}}\right).$$

En particulier, pour $s = 2$, $\text{Var}(\bar{Z}_N^*) = \mathcal{O}\left(\frac{1}{N^{3/2}}\right)$.

Amélioration obtenue par la stratification Sudoku

L'estimateur de Monte Carlo classique \bar{X}_N :

- ▶ est sans biais, i.e. $\mathbb{E}[\bar{X}_N] = \mathcal{J}$;
- ▶ a une variance de l'ordre

$$\text{Var}(\bar{X}_N) = \mathcal{O}\left(\frac{1}{N}\right).$$

L'estimateur de Monte Carlo Sudoku \bar{Z}_N^* :

- ▶ est sans biais ;
- ▶ a une variance de l'ordre

$$\text{Var}(\bar{Z}_N^*) = \mathcal{O}\left(\frac{1}{N^{1+1/s}}\right).$$

En particulier, pour $s = 2$, $\text{Var}(\bar{Z}_N^*) = \mathcal{O}\left(\frac{1}{N^{3/2}}\right)$.

- 1 Méthodes de Monte Carlo
 - Approche intuitive
 - Fondements mathématiques
- 2 Stratification
 - Avantages d'une stratification
 - Types de stratification
 - Stratification Sudoku
- 3 Application aux chaînes de Markov
- 4 Simulations numériques
 - Évolution d'un écosystème méditerranéen
 - Occurrences dans un brin d'ADN

Application aux chaînes de Markov



with $p_0 = 1$. The general solution is, for $i \geq 1$,

$$x_i = x_0 \frac{p_0 \cdots p_{i-1}}{q_1 \cdots q_i}.$$

The positive recurrence condition $\sum_{i \in E} x_i < \infty$ is

$$1 + \sum_{i \geq 1} \frac{p_0 \cdots p_{i-1}}{q_1 \cdots q_i} < \infty,$$

and if it is satisfied, the stationary distribution π is obtained by normalization of the general solution. This gives

$$\pi(0) = \left(1 + \sum_{i \geq 1} \frac{p_0 \cdots p_{i-1}}{q_1 \cdots q_i} \right)^{-1},$$

and for $i \geq 1$,

$$\pi(i) = \pi(0) \frac{p_0 \cdots p_{i-1}}{q_1 \cdots q_i}.$$

In the special case where $p_i = p$, $q_i = q = 1 - p$, the positive recurrence condition becomes $1 + \frac{1}{q} \sum_{j \geq 0} \left(\frac{p}{q}\right)^j < \infty$, that is to say $p < q$, or equivalently,

$$p < \frac{1}{2}.$$

Application aux chaînes de Markov



with $p_0 = 1$. The general solution is, for $i \geq 1$,

$$x_i = x_0 \frac{p_0 \cdots p_{i-1}}{q_1 \cdots q_i}.$$

The positive recurrence condition $\sum_{i \in E} x_i < \infty$ is

$$1 + \sum_{i \geq 1} \frac{p_0 \cdots p_{i-1}}{q_1 \cdots q_i} < \infty,$$

and if it is satisfied, the stationary distribution π is obtained by normalization of the general solution. This gives

$$\pi(0) = \left(1 + \sum_{i \geq 1} \frac{p_0 \cdots p_{i-1}}{q_1 \cdots q_i} \right)^{-1},$$

and for $i \geq 1$,

$$\pi(i) = \pi(0) \frac{p_0 \cdots p_{i-1}}{q_1 \cdots q_i}.$$

In the special case where $p_i = p$, $q_i = q = 1 - p$, the positive recurrence condition becomes $1 + \frac{1}{q} \sum_{j \geq 0} \left(\frac{p}{q}\right)^j < \infty$, that is to say $p < q$, or equivalently,

$$p < \frac{1}{2}.$$



Application aux chaînes de Markov

Un processus aléatoire (ou stochastique) est une succession d'épreuves aléatoires dont les issues (résultats possibles) sont toujours les mêmes, mais où la répartition de probabilité évolue.

Une chaîne de Markov est un processus aléatoire dans lequel l'évolution future dépend du présent, mais pas du passé. Elles sont caractérisées par leurs matrices de transition.



Andreï Markov (1856-1922)

Application aux chaînes de Markov

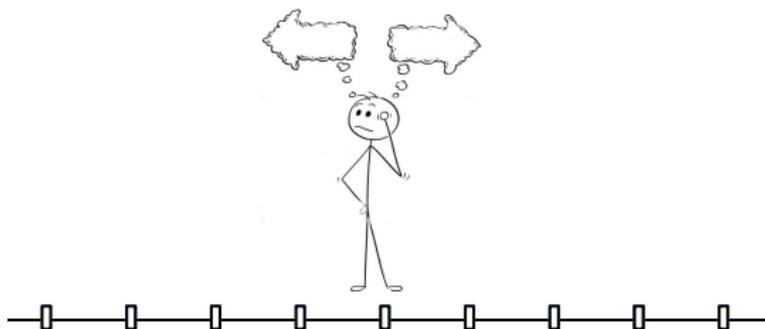
Marche aléatoire sur \mathbb{Z} :

$$S_{k+1} = S_k + W_{k+1}, \quad \text{avec} \quad \mathbb{P}(W_{k+1} = -1) = \frac{1}{2} \quad \text{et} \quad \mathbb{P}(W_{k+1} = +1) = \frac{1}{2}.$$

Application aux chaînes de Markov

Marche aléatoire sur \mathbb{Z} :

$$S_{k+1} = S_k + W_{k+1}, \quad \text{avec} \quad \mathbb{P}(W_{k+1} = -1) = \frac{1}{2} \quad \text{et} \quad \mathbb{P}(W_{k+1} = +1) = \frac{1}{2}.$$



Application aux chaînes de Markov

Marche aléatoire sur \mathbb{Z} :

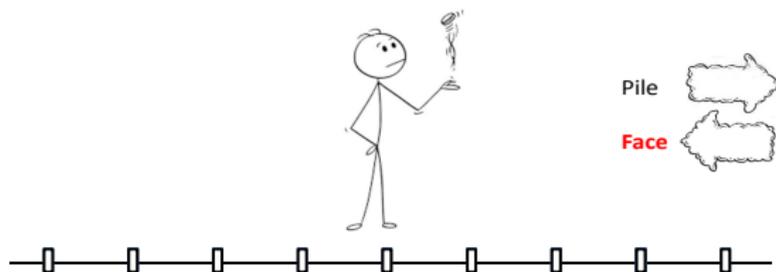
$$S_{k+1} = S_k + W_{k+1}, \quad \text{avec} \quad \mathbb{P}(W_{k+1} = -1) = \frac{1}{2} \quad \text{et} \quad \mathbb{P}(W_{k+1} = +1) = \frac{1}{2}.$$



Application aux chaînes de Markov

Marche aléatoire sur \mathbb{Z} :

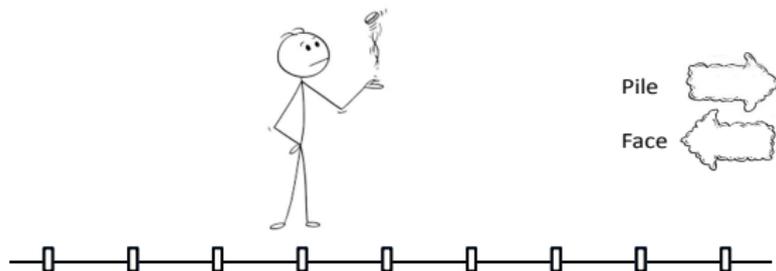
$$S_{k+1} = S_k + W_{k+1}, \quad \text{avec} \quad \mathbb{P}(W_{k+1} = -1) = \frac{1}{2} \quad \text{et} \quad \mathbb{P}(W_{k+1} = +1) = \frac{1}{2}.$$



Application aux chaînes de Markov

Marche aléatoire sur \mathbb{Z} :

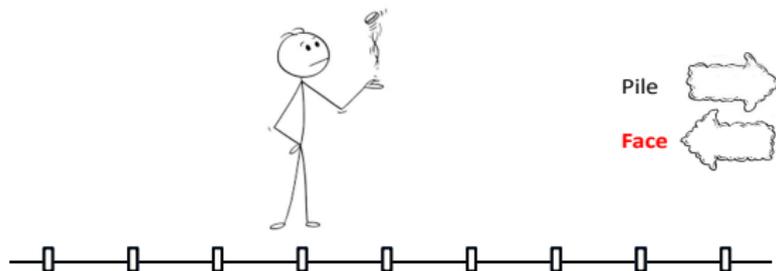
$$S_{k+1} = S_k + W_{k+1}, \quad \text{avec} \quad \mathbb{P}(W_{k+1} = -1) = \frac{1}{2} \quad \text{et} \quad \mathbb{P}(W_{k+1} = +1) = \frac{1}{2}.$$



Application aux chaînes de Markov

Marche aléatoire sur \mathbb{Z} :

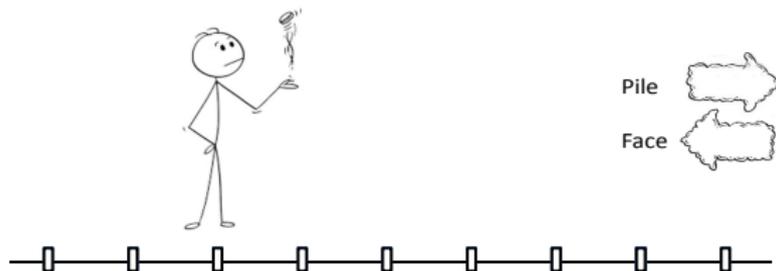
$$S_{k+1} = S_k + W_{k+1}, \quad \text{avec} \quad \mathbb{P}(W_{k+1} = -1) = \frac{1}{2} \quad \text{et} \quad \mathbb{P}(W_{k+1} = +1) = \frac{1}{2}.$$



Application aux chaînes de Markov

Marche aléatoire sur \mathbb{Z} :

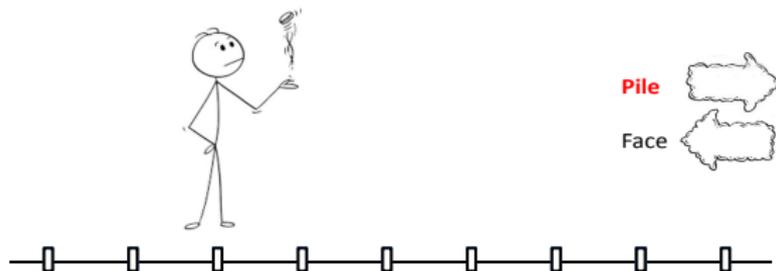
$$S_{k+1} = S_k + W_{k+1}, \quad \text{avec} \quad \mathbb{P}(W_{k+1} = -1) = \frac{1}{2} \quad \text{et} \quad \mathbb{P}(W_{k+1} = +1) = \frac{1}{2}.$$



Application aux chaînes de Markov

Marche aléatoire sur \mathbb{Z} :

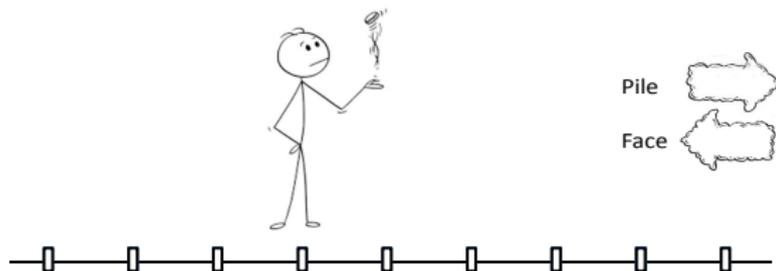
$$S_{k+1} = S_k + W_{k+1}, \quad \text{avec} \quad \mathbb{P}(W_{k+1} = -1) = \frac{1}{2} \quad \text{et} \quad \mathbb{P}(W_{k+1} = +1) = \frac{1}{2}.$$



Application aux chaînes de Markov

Marche aléatoire sur \mathbb{Z} :

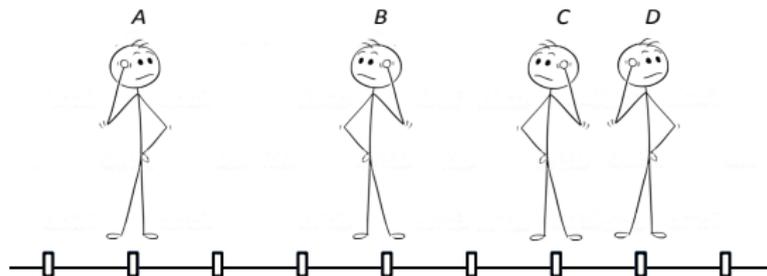
$$S_{k+1} = S_k + W_{k+1}, \quad \text{avec} \quad \mathbb{P}(W_{k+1} = -1) = \frac{1}{2} \quad \text{et} \quad \mathbb{P}(W_{k+1} = +1) = \frac{1}{2}.$$



Application aux chaînes de Markov

Algorithme de Monte Carlo classique :

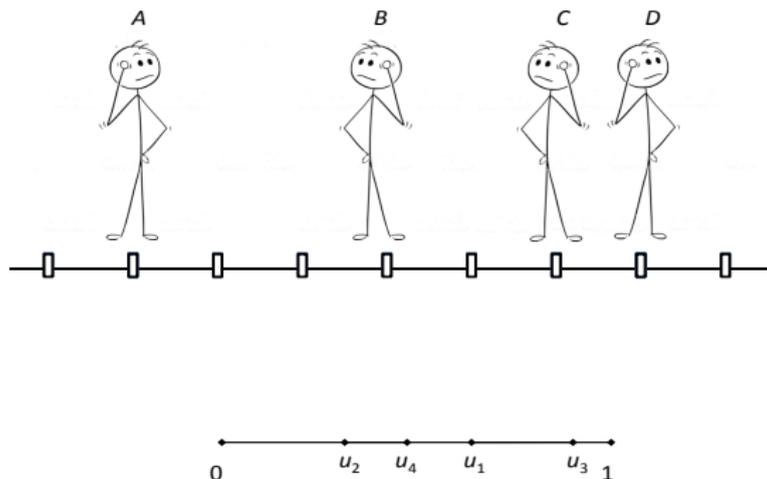
$$S_{k+1} = S_k + W_{k+1}, \quad \text{avec} \quad \mathbb{P}(W_{k+1} = -1) = \frac{1}{2} \quad \text{et} \quad \mathbb{P}(W_{k+1} = +1) = \frac{1}{2}.$$



Application aux chaînes de Markov

Algorithme de Monte Carlo classique :

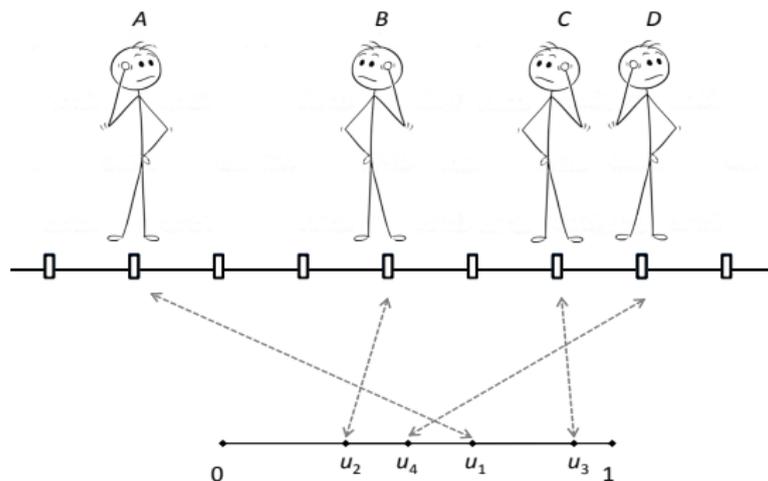
$$S_{k+1} = S_k + W_{k+1}, \quad \text{avec} \quad \mathbb{P}(W_{k+1} = -1) = \frac{1}{2} \quad \text{et} \quad \mathbb{P}(W_{k+1} = +1) = \frac{1}{2}.$$



Application aux chaînes de Markov

Algorithme de Monte Carlo classique :

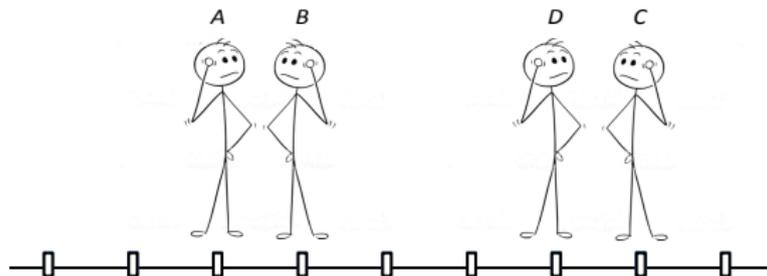
$$S_{k+1} = S_k + W_{k+1}, \quad \text{avec} \quad \mathbb{P}(W_{k+1} = -1) = \frac{1}{2} \quad \text{et} \quad \mathbb{P}(W_{k+1} = +1) = \frac{1}{2}.$$



Application aux chaînes de Markov

Algorithme de Monte Carlo classique :

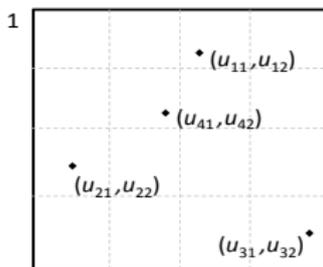
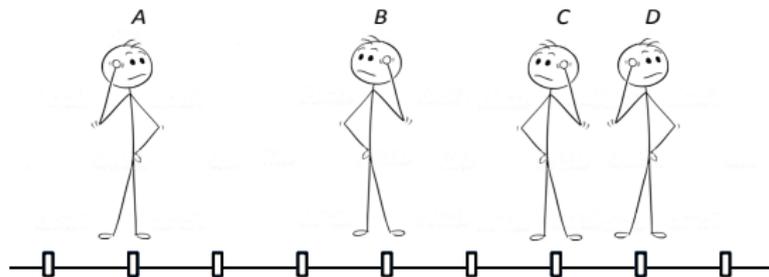
$$S_{k+1} = S_k + W_{k+1}, \quad \text{avec} \quad \mathbb{P}(W_{k+1} = -1) = \frac{1}{2} \quad \text{et} \quad \mathbb{P}(W_{k+1} = +1) = \frac{1}{2}.$$



Application aux chaînes de Markov

Algorithme Sudoku :

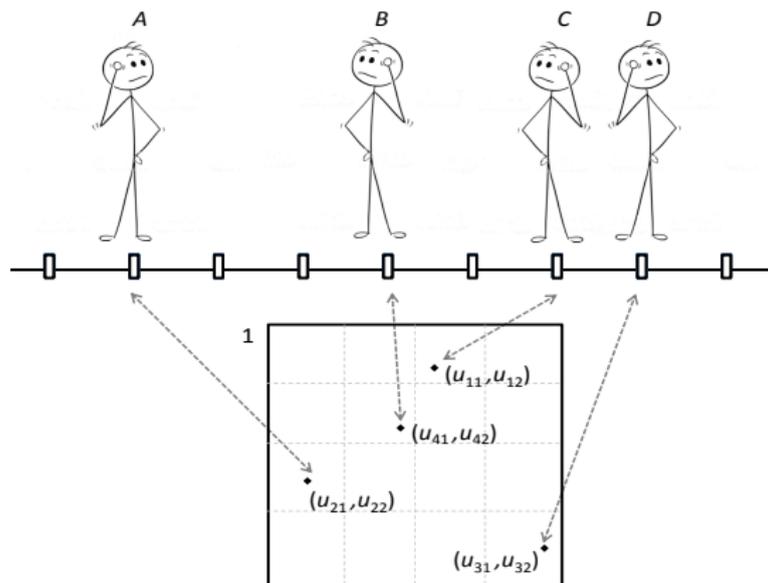
$$S_{k+1} = S_k + W_{k+1}, \quad \text{avec} \quad \mathbb{P}(W_{k+1} = -1) = \frac{1}{2} \quad \text{et} \quad \mathbb{P}(W_{k+1} = +1) = \frac{1}{2}.$$



Application aux chaînes de Markov

Algorithme Sudoku :

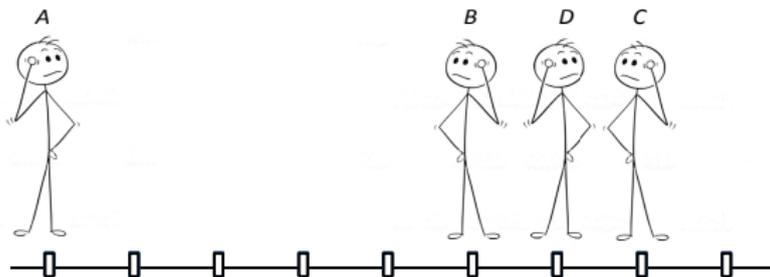
$$S_{k+1} = S_k + W_{k+1}, \quad \text{avec} \quad \mathbb{P}(W_{k+1} = -1) = \frac{1}{2} \quad \text{et} \quad \mathbb{P}(W_{k+1} = +1) = \frac{1}{2}.$$



Application aux chaînes de Markov

Algorithme Sudoku :

$$S_{k+1} = S_k + W_{k+1}, \quad \text{avec} \quad \mathbb{P}(W_{k+1} = -1) = \frac{1}{2} \quad \text{et} \quad \mathbb{P}(W_{k+1} = +1) = \frac{1}{2}.$$



- 1 Méthodes de Monte Carlo
 - Approche intuitive
 - Fondements mathématiques

- 2 Stratification
 - Avantages d'une stratification
 - Types de stratification
 - Stratification Sudoku

- 3 Application aux chaînes de Markov

- 4 Simulations numériques
 - Évolution d'un écosystème méditerranéen
 - Occurrences dans un brin d'ADN

Évolution d'un écosystème méditerranéen

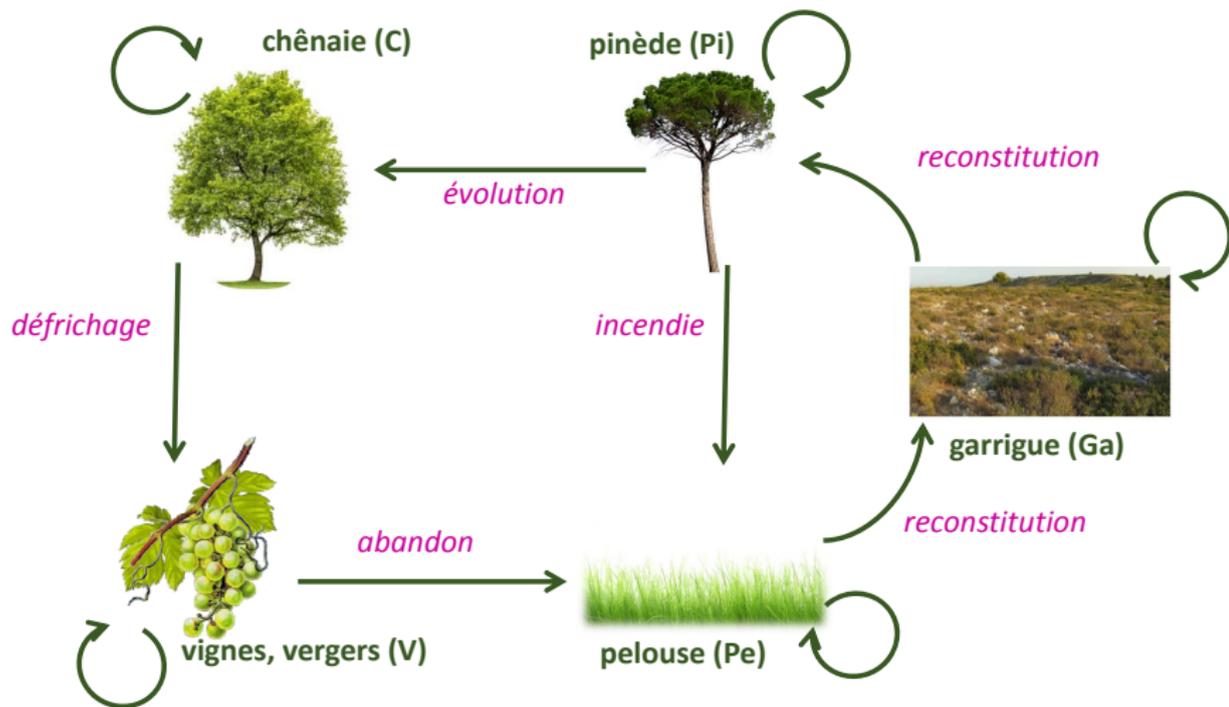
A l'origine la forêt méditerranéenne (sur roche calcaire à faible altitude) était très certainement dominée par des **chênes**.

L'action de l'homme a éradiqué ces forêts primitives pour leur substituer des **vignes**, des vergers...

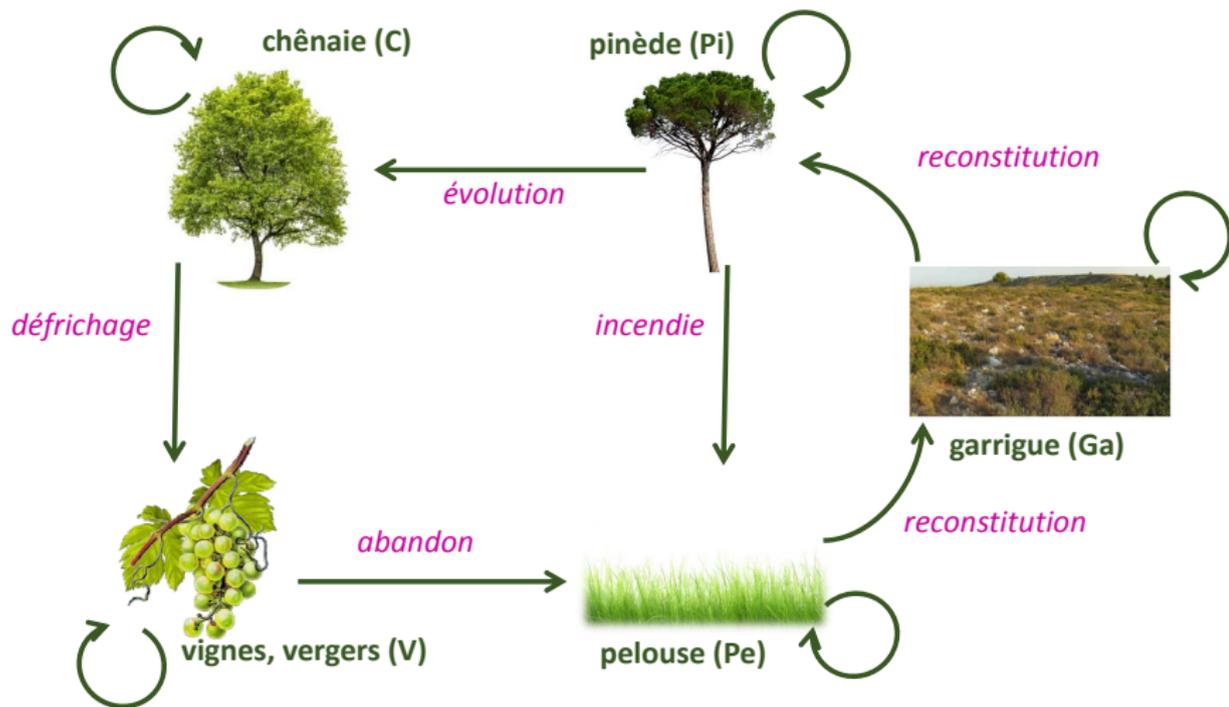
L'abandon de toute activité agricole, au lieu de conduire à la restauration naturelle de ces chênaies, a bien souvent favorisé l'implantation d'une autre espèce, le **pin** d'Alep, après un passage par un état de **garrigue**.

Ces forêts de substitution, hautement inflammables, subissent de manière récurrente le passage du feu (incendies volontaires ou non), le sol mis à nu se recouvre pendant un temps de **pelouses** ; ces forêts sont donc condamnées à une perpétuelle reconstitution.

Évolution d'un écosystème méditerranéen



Évolution d'un écosystème méditerranéen



Suite à un incendie, l'écosystème est constitué d'une pelouse à la date t . Quelle sera sa composition à la date $t + 7$?

Évolution d'un écosystème méditerranéen

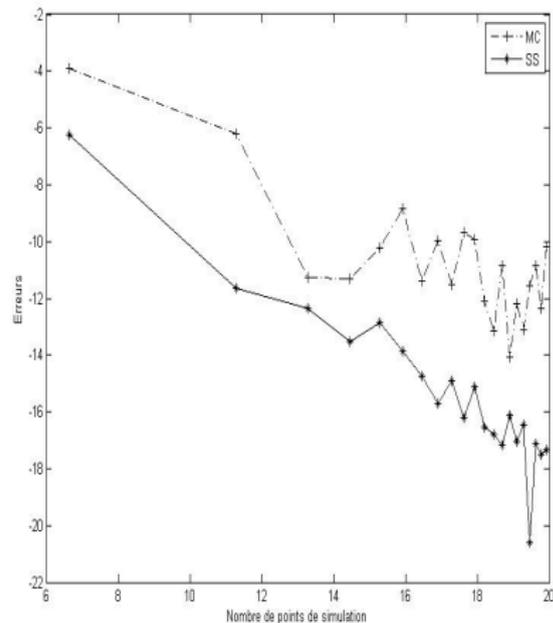
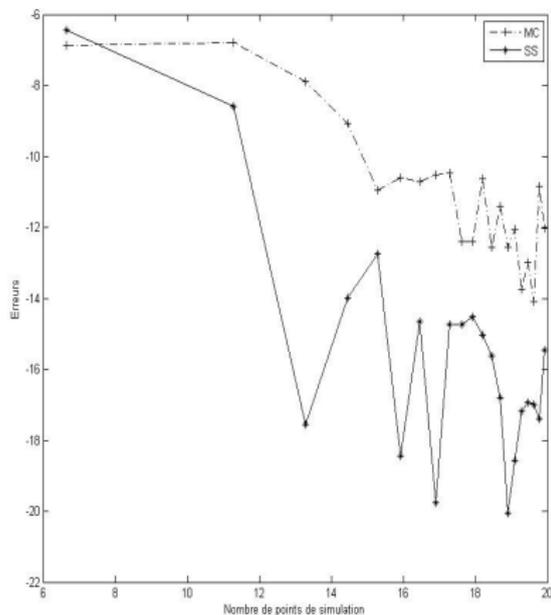
On modélise cette évolution par une chaîne de Markov $(X_n)_{n \in \mathbb{N}}$ à cinq états C, V, Pe, Ga, Pi.

Pour tout $n \in \mathbb{N}$, X_n désignera l'état de cet écosystème à l'instant n .

La matrice de transition de la chaîne est la suivante :

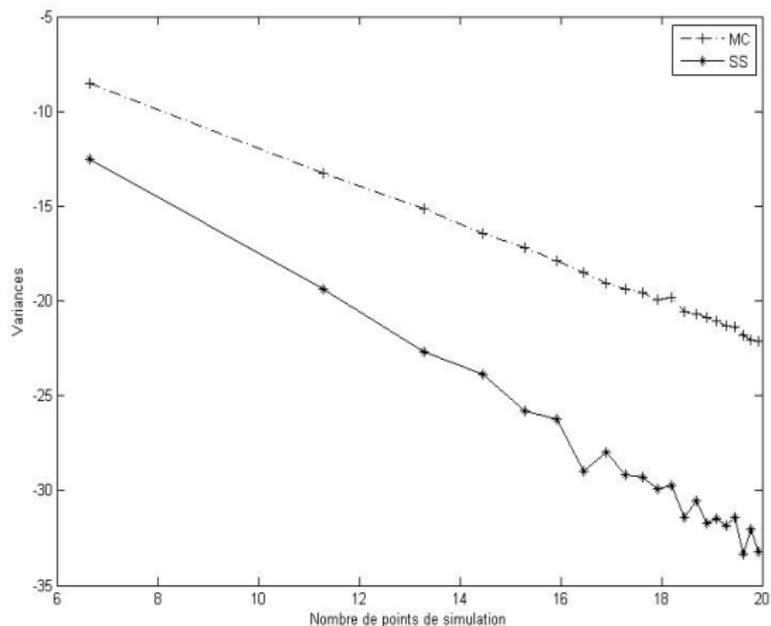
$$\begin{pmatrix} 0,8 & 0,2 & 0 & 0 & 0 \\ 0 & 0,7 & 0,3 & 0 & 0 \\ 0 & 0 & 0,4 & 0,6 & 0 \\ 0 & 0 & 0 & 0,2 & 0,8 \\ 0,1 & 0 & 0,25 & 0 & 0,65 \end{pmatrix}$$

Évolution d'un écosystème méditerranéen



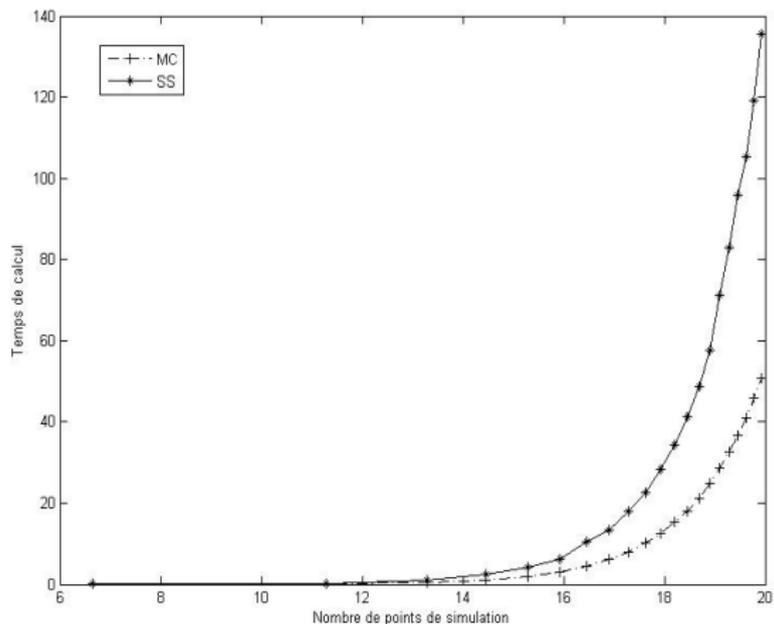
Erreurs pour C et V en fonction du nombre de points de simulation N .

Évolution d'un écosystème méditerranéen



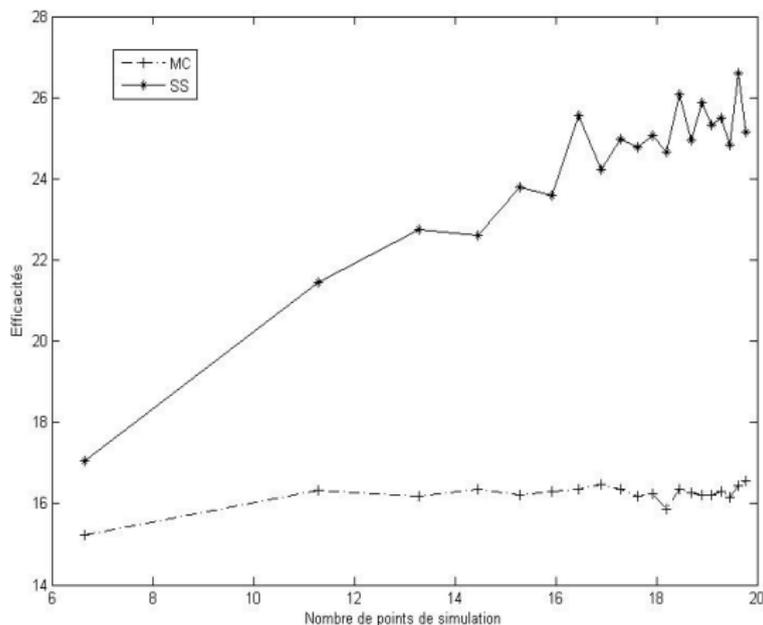
Variances pour π_i en fonction du nombre de points de simulation N .

Évolution d'un écosystème méditerranéen



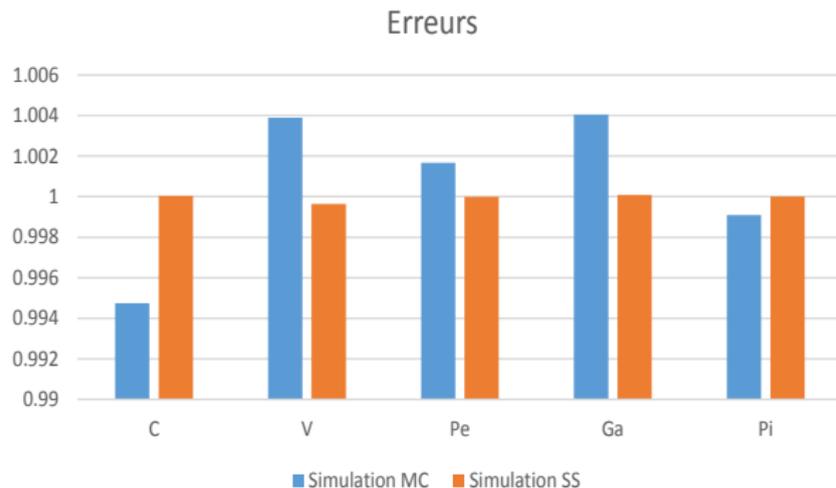
Temps de calcul pour π en fonction du nombre de points de simulation N .

Évolution d'un écosystème méditerranéen



Efficacités pour P_i en fonction du nombre de points de simulation N .

Évolution d'un écosystème méditerranéen



Les pourcentages obtenus sont les suivants :

| C | V | Pe | Ga | Pi |
|--------|-------|--------|--------|--------|
| 16,31% | 6,14% | 20,46% | 16,01% | 41,07% |

Occurrences dans un brin d'ADN

Un brin de l'ADN du virus HIV-1 compte 9 717 nucléotides. Le nombre d'occurrences des 16 dinucléotides sur ce brin est le suivant :

| | | | | | | | |
|-------|-----|-------|-----|-----|-----|-----|-----|
| aa | ac | ag | at | ca | cc | cg | ct |
| 1 112 | 561 | 1 024 | 713 | 795 | 413 | 95 | 470 |
| ga | gc | gg | gt | ta | tc | tg | tt |
| 820 | 457 | 661 | 432 | 684 | 342 | 590 | 548 |



Occurrences dans un brin d'ADN

Un brin de l'ADN du virus HIV-1 compte 9 717 nucléotides. Le nombre d'occurrences des 16 dinucléotides sur ce brin est le suivant :

| | | | | | | | |
|-------|-----|-------|-----|-----|-----|-----|-----|
| aa | ac | ag | at | ca | cc | cg | ct |
| 1 112 | 561 | 1 024 | 713 | 795 | 413 | 95 | 470 |
| ga | gc | gg | gt | ta | tc | tg | tt |
| 820 | 457 | 661 | 432 | 684 | 342 | 590 | 548 |



En considérant un tel brin d'ADN commençant par un a, quelle est la probabilité que le 500^{ième} nucléotide soit un g ?

Occurrences dans un brin d'ADN

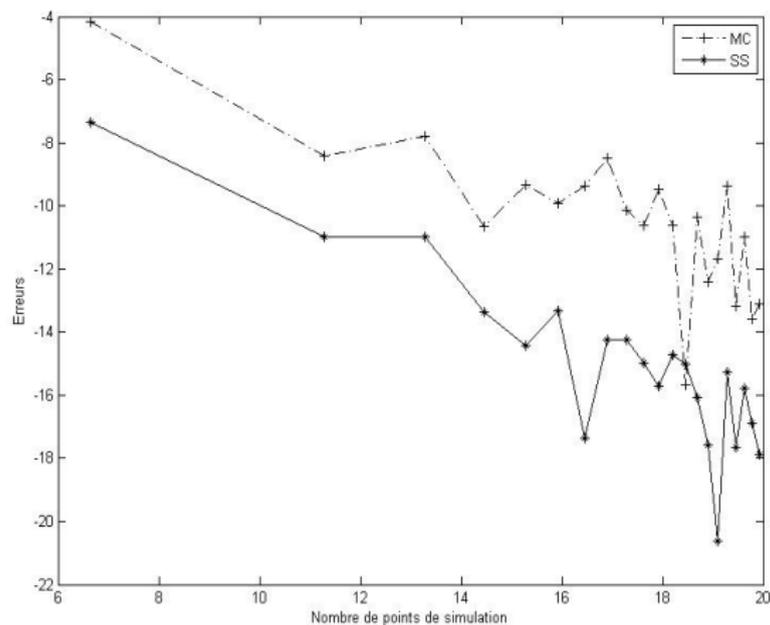
On propose une modélisation d'une telle séquence d'ADN par une chaîne de Markov qui rend compte du comptage des dinucléotides.

On considère la suite de nucléotides comme les 9 717 premiers états d'une chaîne de Markov $(X_n)_{n \in \mathbb{N}}$ dont les états sont : a, c, g, et t.

On prend pour matrice de transition :

$$\begin{pmatrix} \frac{1112}{3410} & \frac{561}{3410} & \frac{1024}{3410} & \frac{713}{3410} \\ \frac{795}{1773} & \frac{413}{1773} & \frac{95}{1773} & \frac{470}{1773} \\ \frac{820}{2370} & \frac{457}{2370} & \frac{661}{2370} & \frac{432}{2370} \\ \frac{684}{2164} & \frac{342}{2164} & \frac{590}{2164} & \frac{548}{2164} \end{pmatrix}$$

Occurrences dans un brin d'ADN



Erreurs en fonction du nombre de points de simulation N (pourcentage exact = 24,39%).







Merci de votre attention

