

Modélisation de phénomènes aléatoires

Cours 4

Thierry Bodineau

École Polytechnique
Département de Mathématiques Appliquées



Résumé du cours 3

Théorème 5.1

$\{X_n\}$ chaîne de Markov irréductible, récurrente positive.

On note π sa mesure de probabilité invariante.

Soit $F : E \rightarrow \mathbb{R}$ telle que $\mathbb{E}_\pi(|F|) < \infty$

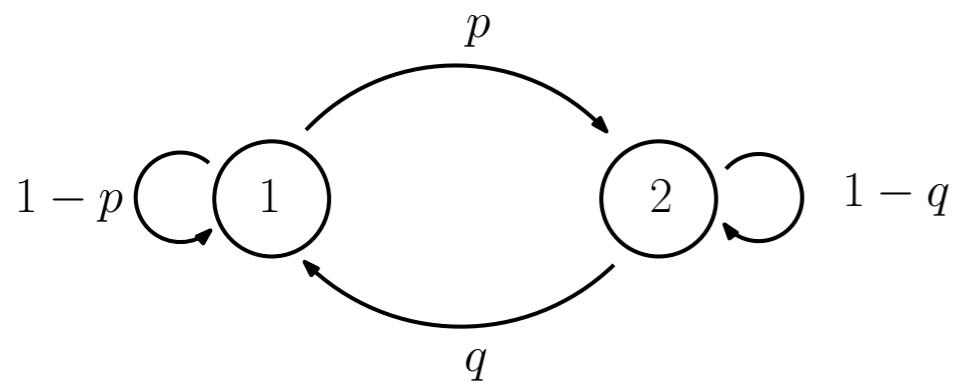
$$\frac{1}{n} \sum_{i=0}^{n-1} F(X_i) \xrightarrow[n \rightarrow \infty]{p.s.} \mathbb{E}_\pi(F) = \sum_{x \in E} \pi(x) F(x)$$

Définition 5.3

Une chaîne de Markov irréductible est *apériodique* si pour tous x, y de E il existe $n(x, y) \in \mathbb{N}$ tel que

$$\forall n \geq n(x, y), \quad \mathbb{P}_x(X_n = y) = P^n(x, y) > 0$$

Graphe de transition



$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}, \quad p+q>0$$

Mesure invariante :

$$\pi(1) = \frac{q}{p+q}, \quad \pi(2) = \frac{p}{p+q}$$

- Si $p = q = 1$: Comportement périodique

$$\forall n \geq 0, \quad \mathbb{P}_1(X_{2n} = 2) = 0, \quad \mathbb{P}_1(X_{2n+1} = 2) = 1$$

- Si $p, q \in]0, 1[$: Convergence exponentielle vers π

$$\forall x, y \in \{1, 2\}, \quad \mathbb{P}_x(X_n = y) = P^n(x, y) \xrightarrow{n \rightarrow \infty} \pi(y)$$

Convergence

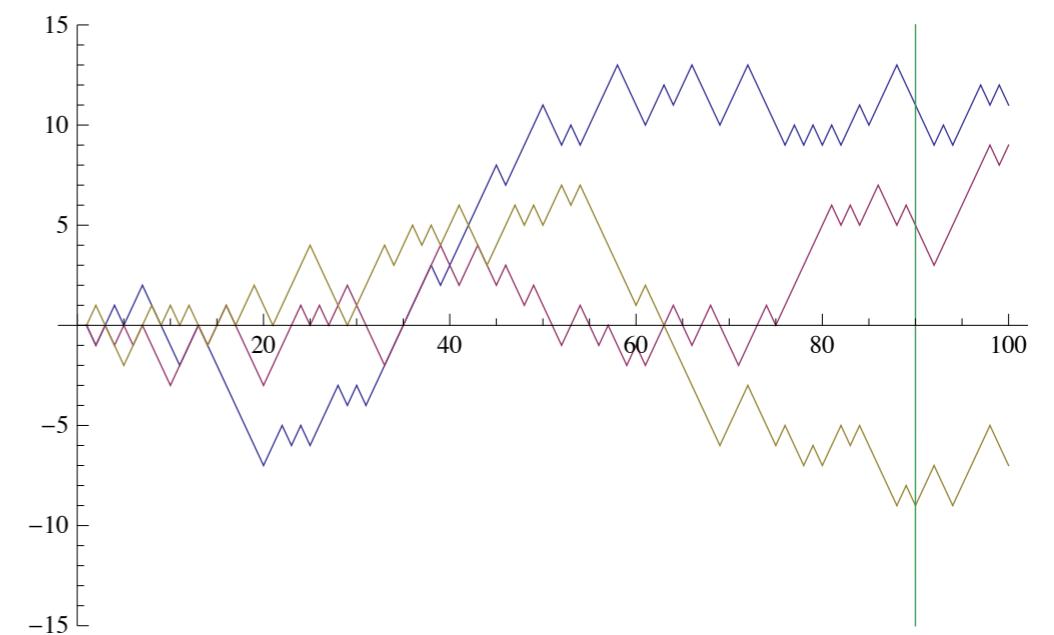
Théorème 5.5

$\{X_n\}_{n \geq 0}$ chaîne de Markov irréductible et apériodique de mesure de probabilité invariante π .

$$\forall x, z \in E, \quad \lim_{n \rightarrow \infty} \mathbb{P}_x(X_n = z) = \pi(z)$$

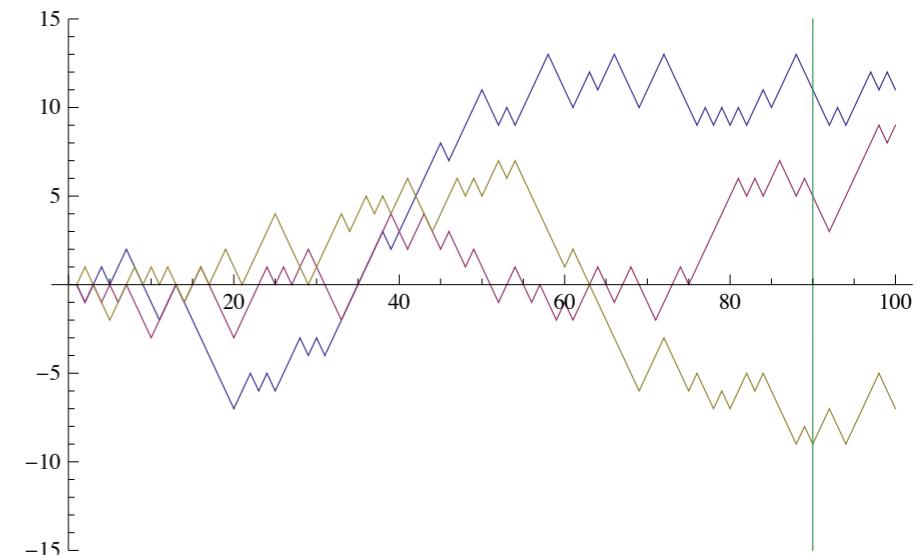
Oubli du point de départ

Pour n grand, X_n a presque la distribution π



Simulation de la mesure π

On simule K réalisations $\{X_n^{(k)}\}_{k \leq K}$ indépendantes de la chaîne de Markov

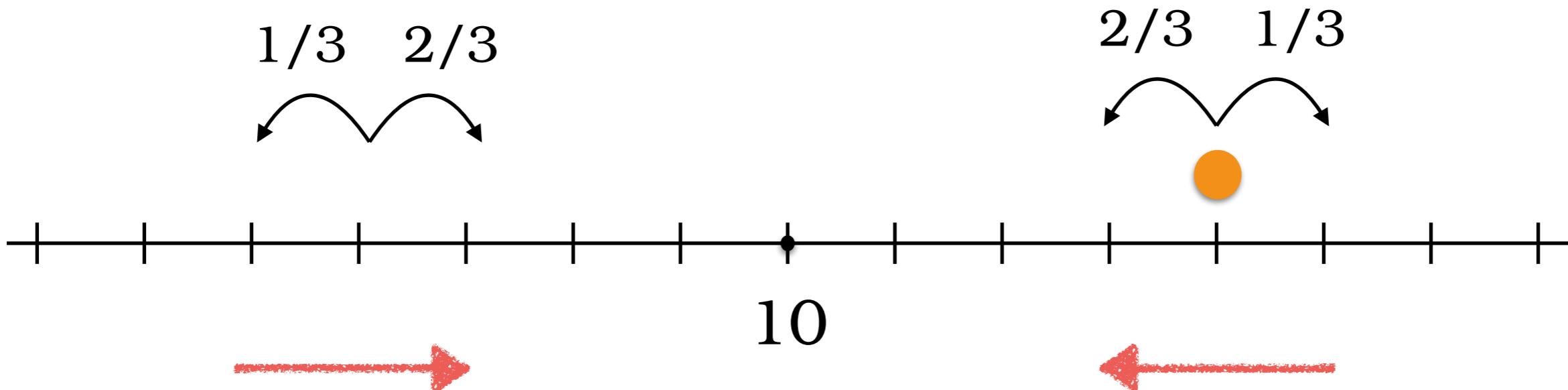
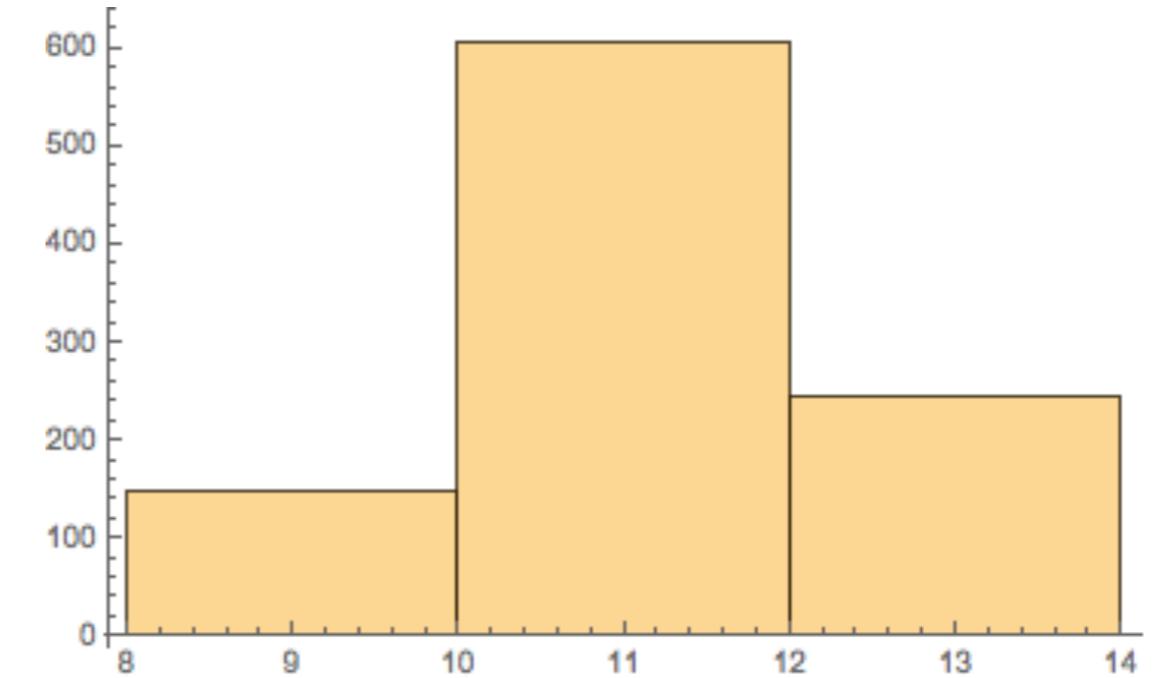
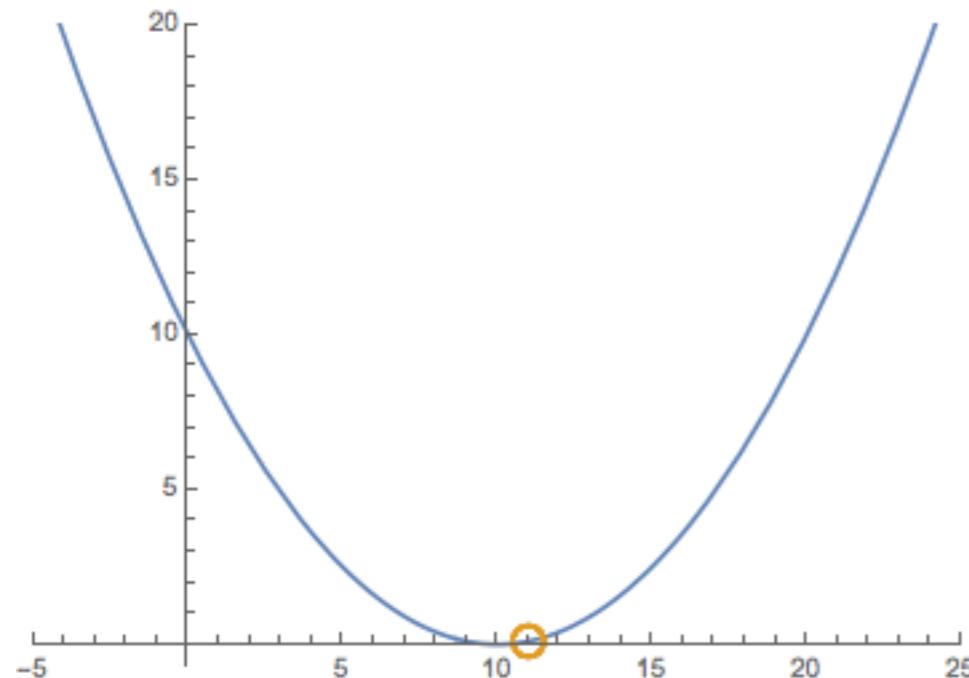


$$\frac{1}{K} \sum_{k=1}^K \mathbf{1}_{\{X_n^{(k)} = z\}} \xrightarrow{K \rightarrow \infty} \mathbb{P}_x(X_n = z) \xrightarrow{n \rightarrow \infty} \pi(z)$$

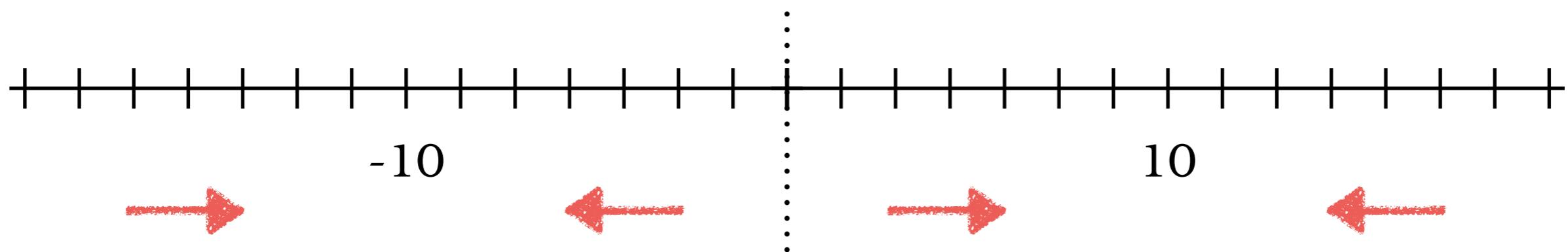
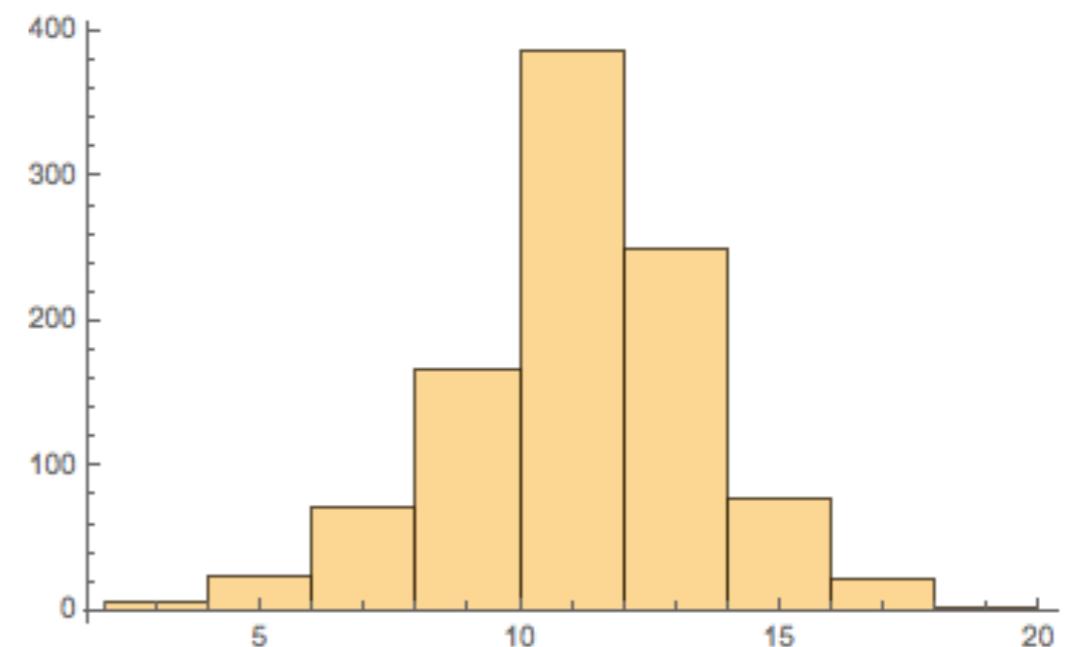
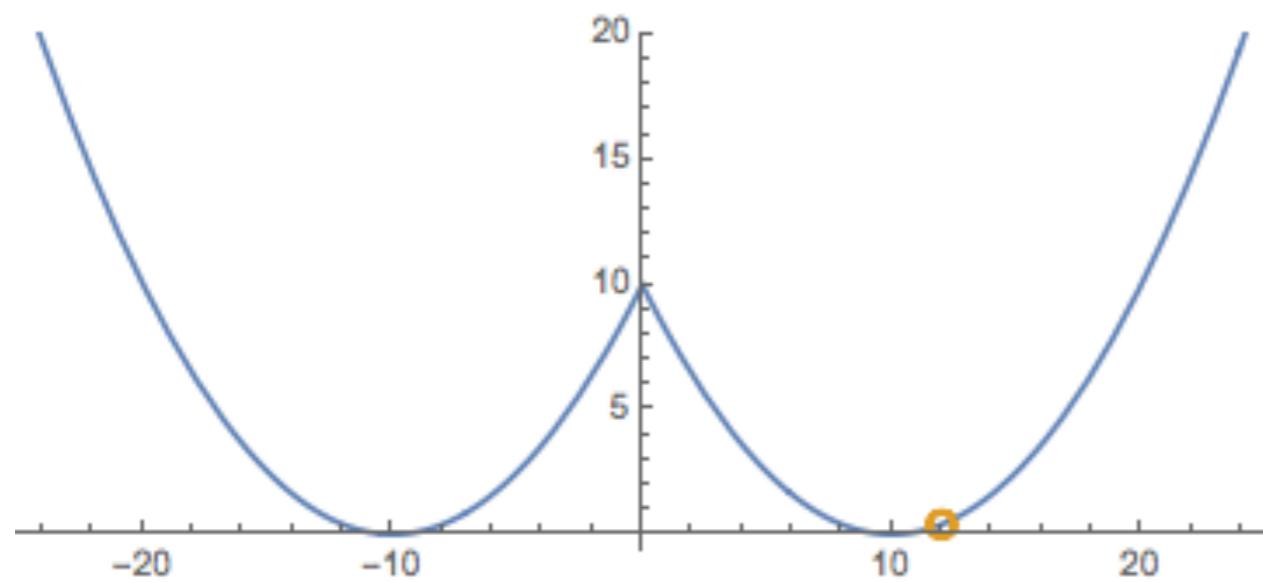


Question: Comment choisir K et n pour simuler π

Particule dans un puits de potentiel



Double puits de potentiel

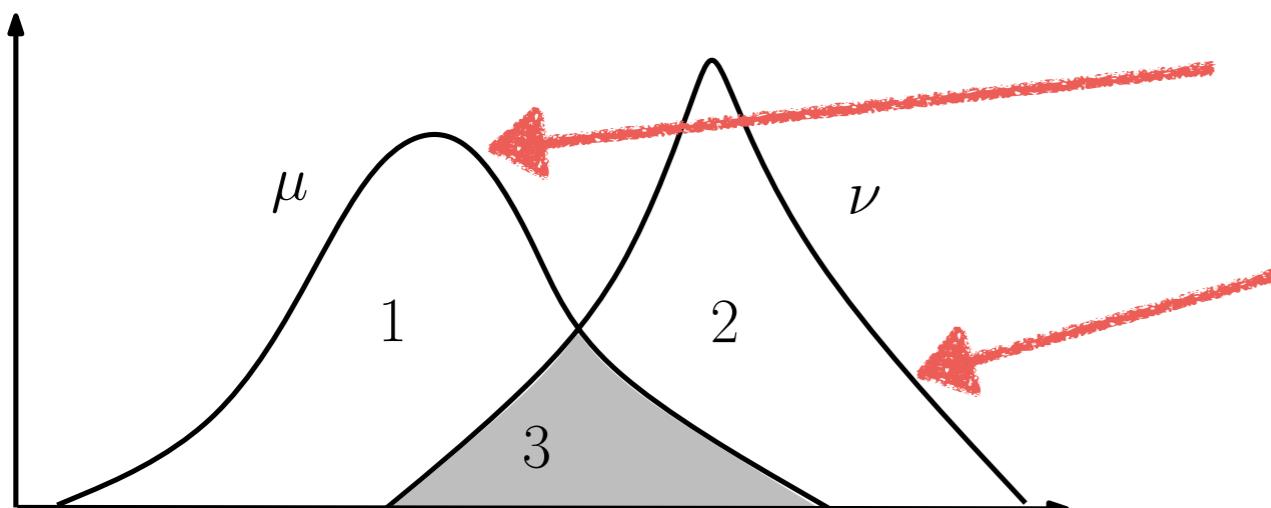


Distance en variation

Définition 5.9 et Lemme 5.10

Soient μ et ν deux mesures de probabilité sur E dénombrable
On définit leur *distance en variation totale* par

$$\|\mu - \nu\|_{VT} = \frac{1}{2} \sum_{x \in E} |\mu(x) - \nu(x)| = \max_{A \subset E} |\mu(A) - \nu(A)|$$



La distance en variation est proportionnelle à l'aire des 2 zones.

Objectif. Quantifier la vitesse de convergence $\varepsilon(n)$ quand $n \rightarrow \infty$

$$\left\| P^n(x, \cdot) - \pi(\cdot) \right\|_{VT} = \frac{1}{2} \sum_{y \in E} |\mathbb{P}_x(X_n = y) - \pi(y)| = \varepsilon(n)$$



Wolfgang Doeblin (1915-1940)

Thèse 1938

Sur l'équation de Kolmogoroff
(1940)

Condition de Doeblin

Il existe $r \geq 1$, $\delta > 0$ et une mesure de probabilité ν sur E tels que

$$P^r(x, z) \geq \delta \nu(z) \quad \text{pour tous } x, z \text{ dans } E$$

Version quantitative
de l'irréductibilité

Théorème 5.11

$\{X_n\}$ chaîne de Markov irréductible, apériodique. Si la **condition de Doeblin** est satisfaite alors : il existe une mesure de probabilité invariante π et

$$\sup_{x \in E} \|P^n(x, \cdot) - \pi\|_{VT} = \frac{1}{2} \sup_{x \in E} \sum_{y \in E} |P^n(x, y) - \pi(y)| \leq (1 - \delta)^{\lceil n/r \rceil}$$

Convergence uniforme par rapport aux états initiaux

vitesse de convergence exponentielle

Si E est fini

Chaîne irréductibilité, apériodique \Rightarrow condition de Doeblin

Apériodique \Rightarrow Pour tous x, y de E il existe $n(x, y) \in \mathbb{N}$ tel que

$$\forall n \geq n(x, y), \quad \mathbb{P}_x(X_n = y) = P^n(x, y) > 0$$

E fini $r = \max_{x, y \in E} n(x, y) \quad \Rightarrow \quad \min_{x, y \in E} P^r(x, y) > 0$

On pose

$$\delta = \sum_{y \in E} \min_{x \in E} P^r(x, y) > 0 \quad \text{et} \quad \nu(y) = \frac{1}{\delta} \min_{x \in E} P^r(x, y)$$

Condition de Doeblin

$$P^r(x, y) \geq \delta \nu(y) \quad \text{pour tous } x, y \text{ dans } E$$

Exemple: convergence de la marche fainéante

Chaîne de Markov sur $E = \{1, \dots, L\}$ (avec conditions périodiques)

$$\forall i \in E, \quad P(i, i+1) = P(i, i-1) = 1/4, \quad P(i, i) = 1/2$$

Mesure de probabilité invariante $\pi(i) = 1/L$

apériodique

- Condition de Doeblin (avec $r = L/2$ et $\delta \simeq \frac{1}{4^{L/2}}$) implique

$$\sup_{x \in E} \|P^n(x, \cdot) - \pi\|_{VT} \leq \left(1 - \frac{1}{4^{L/2}}\right)^{\lceil \frac{n}{L/2} \rceil} \simeq \exp\left(-c \frac{n}{2^L L}\right)$$

- Un couplage spécifique permet d'obtenir

$$\|P^n(x, \cdot) - P^n(y, \cdot)\|_{VT} \leq \frac{L^2}{4n}$$

Matrice de transition modifiée sur $E = \{1, \dots, L\}$

$$\forall i, j \in E, \quad P(i, j) = \frac{1}{4} \mathbb{1}_{|i-j|=1} + \frac{1}{2L}$$



possibilité de saut vers
n'importe quel site

- Condition de Doeblin (avec $r = 1$, $\delta = 1/2$ et $\nu(y) = 1/L$) implique

$$\sup_{x \in E} \|P^n(x, \cdot) - \pi\|_{VT} \leq \frac{1}{2^n}$$

La convergence est accélérée (cf. matrice de transition de Google)

Application aux algorithmes stochastiques

Chapitre 6

Optimisation

Objectifs :

- Minimiser $x \rightarrow V(x)$
- Identifier les valeurs de x où le minimum est atteint

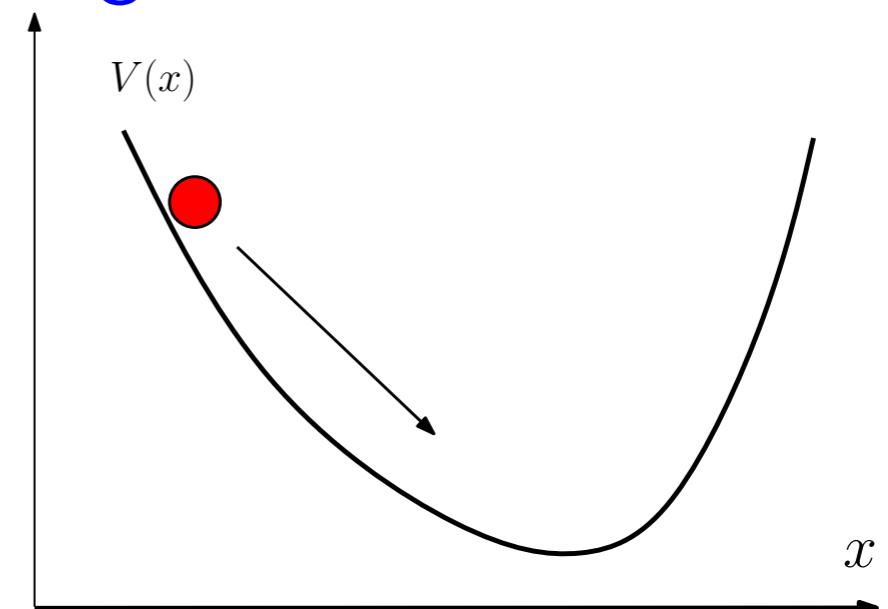
Applications : optimiser un coût, un rendement

Si V est convexe : méthode de descente de gradient

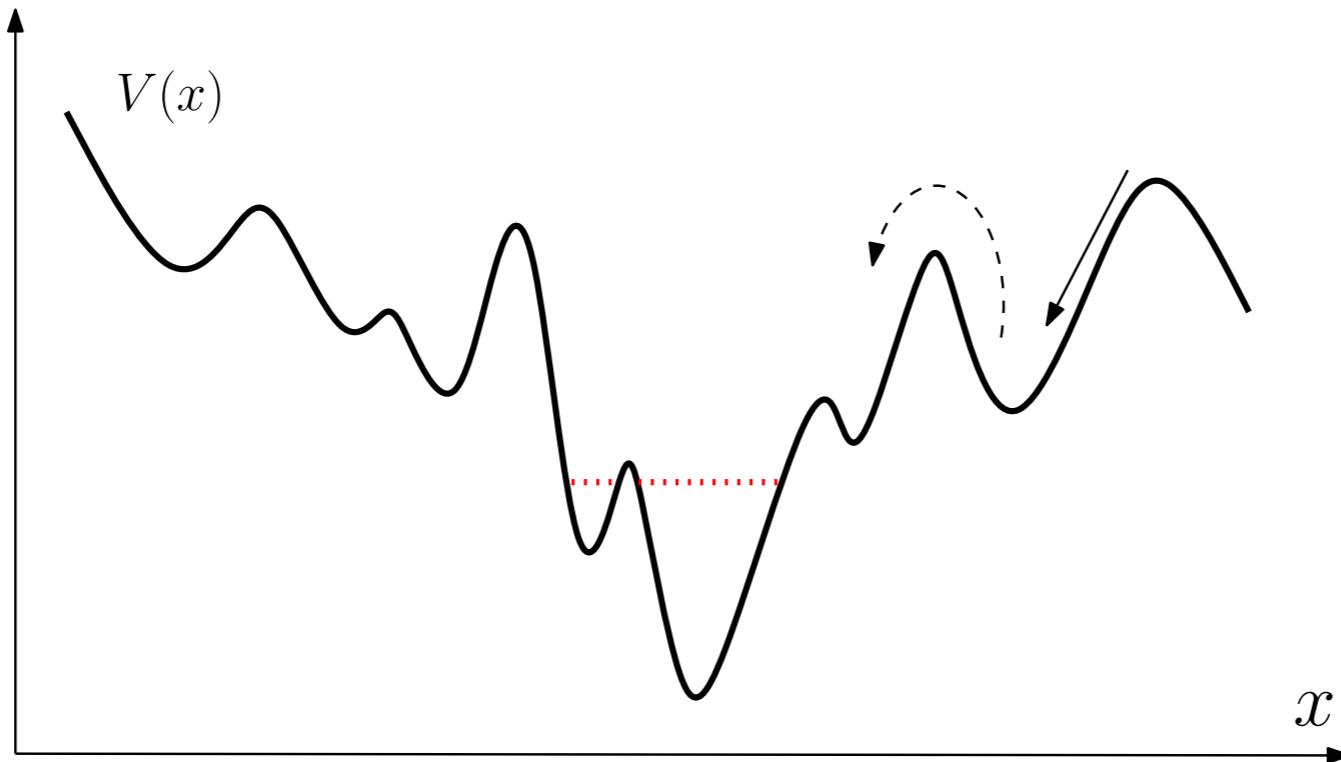
$$\partial_t x(t) = -\nabla V(x_t)$$

$$x_t \xrightarrow[t \rightarrow \infty]{} \text{Argmin}(V)$$

Généralisable dans le cas discret



Une méthode déterministe peut être piégée dans les minima locaux



Questions :

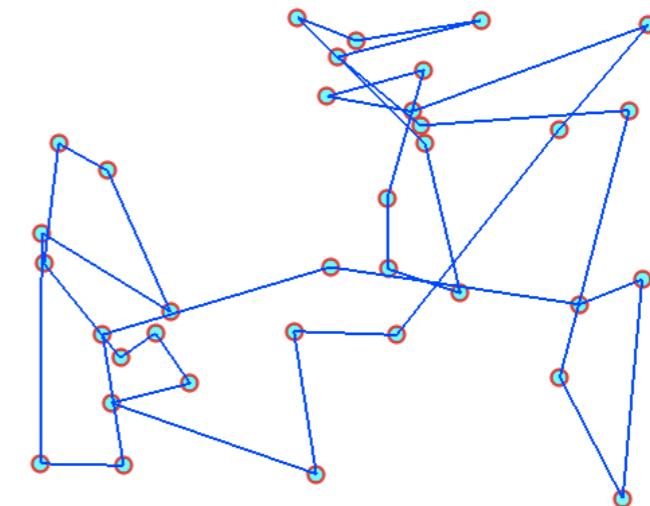
- Si V a de nombreux minima locaux ?
- Si la dimension de l'espace est grande ?

Exemple: problème du voyageur de commerce

Les distances entre K villes $\{d(i,j)\}_{\substack{1 \leq i \leq K \\ 1 \leq j \leq K}}$ sont données

Objectif :

Trajet minimum pour visiter K villes et revenir au point de départ en ne passant par chaque ville qu'une seule fois.



Déterminer σ la permutation de $\{1, \dots, K\}$ telle que

$$\min_{\sigma \in S_K} \{V(\sigma)\} \quad \text{avec} \quad V(\sigma) = \sum_{i=1}^K d(\sigma(i), \sigma(i+1))$$

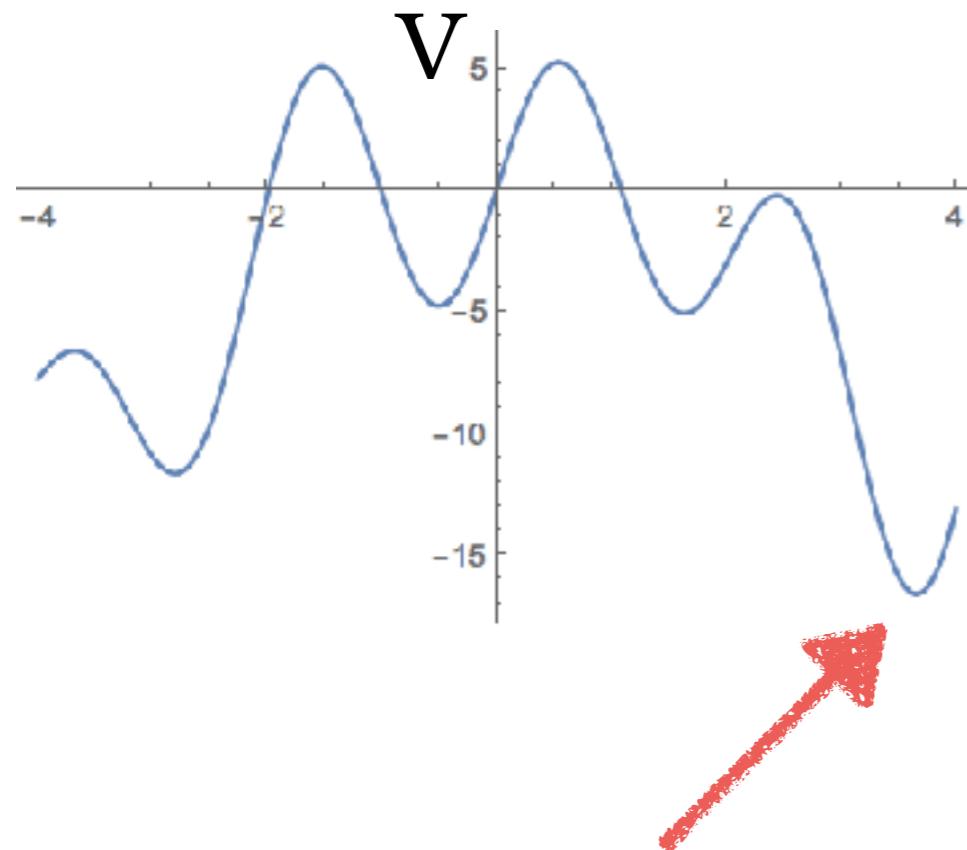
Problème NP complet \Rightarrow résolution très difficile pour K grand

Méthode de Laplace

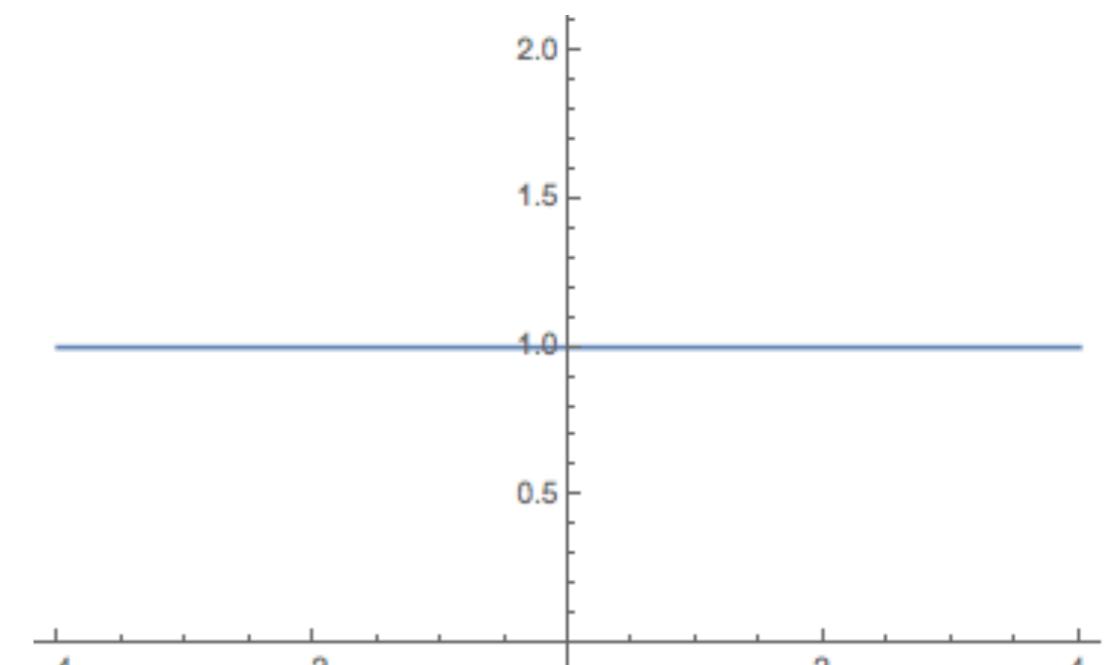
Espace d'états E fini (mais de cardinal très grand).

μ_T : mesure de Gibbs associée à V et au paramètre $T > 0$

$$\forall x \in E, \quad \mu_T(x) = \frac{1}{Z_T} \exp\left(-\frac{1}{T} V(x)\right)$$



minimum



$$\exp\left(-\frac{1}{T} V(x)\right)$$

Méthode de Laplace

Espace d'états E fini (mais de cardinal très grand).

μ_T : mesure de Gibbs associée à V et au paramètre $T > 0$

$$\forall x \in E, \quad \mu_T(x) = \frac{1}{Z_T} \exp\left(-\frac{1}{T} V(x)\right)$$

Lemme 6.1

\mathcal{M} = points de E où V atteint son minimum.

$$\forall x \in E, \quad \lim_{T \rightarrow 0} \mu_T(x) = \begin{cases} \frac{1}{\text{card}\mathcal{M}} & \text{si } x \in \mathcal{M} \\ 0 & \text{si } x \notin \mathcal{M} \end{cases}$$

Résolution approchée du problème de minimisation

Stratégie.

Fixer T petit. Simuler une variable X sous μ_T pour obtenir une valeur proche d'un minimum car

$$\lim_{T \rightarrow 0} \mu_T(\mathcal{M}) = 1.$$

Difficulté pour simuler la mesure de Gibbs

$$\forall x \in E, \quad \mu_T(x) = \frac{1}{Z_T} \exp\left(-\frac{1}{T} V(x)\right)$$

car il faut calculer

$$Z_T = \sum_{x \in E} \exp\left(-\frac{1}{T} V(x)\right)$$

Algorithme de Hastings-Metropolis

Enjeu. Simuler un tirage aléatoire sous une mesure π

Objectif. Construire une chaîne de Markov réversible sous π

Soit Q une matrice de transition irréductible telle que

$$Q(x, y) > 0 \quad \Rightarrow \quad Q(y, x) > 0$$

Soit $h(u) = \frac{u}{1+u}$. Soit $\pi(x) > 0$ une mesure, on pose

$$x \neq y, \quad R(x, y) = \begin{cases} h\left(\frac{\pi(y)Q(y, x)}{\pi(x)Q(x, y)}\right) & \text{si } Q(x, y) \neq 0 \\ 0 & \text{sinon} \end{cases}$$

On définit la matrice de transition P

$$\begin{cases} P(x, y) = Q(x, y)R(x, y) & \text{si } x \neq y \\ P(x, x) = 1 - \sum_{y \neq x} P(x, y) & \end{cases}$$

Recette pour simuler P

Ingrédients : une suite U_n de v.a. i.i.d uniformes sur $[0, 1]$

- *Étape 1 :* Initialiser X_0
- *Étape $n+1$:*
 - Choisir y selon la loi $Q(X_n, y)$
 - Choisir U_{n+1} uniformément dans $[0, 1]$
 - Si $U_{n+1} < R(X_n, y)$ poser $X_{n+1} = y$,
 - sinon poser $X_{n+1} = X_n$.

Théorème 6.2.

P est irréductible, apériodique et réversible pour la mesure π qui est donc l'unique mesure de probabilité invariante.

Preuve de la réversibilité.

$$\begin{aligned}\pi(x)P(x,y) &= \pi(x)Q(x,y)h\left(\frac{\pi(y)Q(y,x)}{\pi(x)Q(x,y)}\right) \\ &= \pi(y)Q(y,x)\frac{\pi(x)Q(x,y)}{\pi(y)Q(y,x)} h\left(\frac{\pi(y)Q(y,x)}{\pi(x)Q(x,y)}\right) \\ &= \pi(y)Q(y,x) h\left(\frac{\pi(x)Q(x,y)}{\pi(y)Q(y,x)}\right) \quad \text{red arrow from } \frac{\pi(y)Q(y,x)}{\pi(x)Q(x,y)} \\ &= \pi(y)P(y,x)\end{aligned}$$

$uh\left(\frac{1}{u}\right) = h(u)$

Conséquence. Convergence de la chaîne vers π

$$\forall x \in E, \quad \pi(x) = \lim_{n \rightarrow \infty} \mathbb{P}(X_n = x)$$

vitesse de convergence ?

Simuler la mesure de Gibbs par Hastings-Metropolis

$$\forall x \in E, \quad \mu_T(x) = \frac{1}{Z_T} \exp\left(-\frac{1}{T} V(x)\right)$$

$$\frac{\mu_T(y)}{\mu_T(x)}$$

Pour $x \neq y$, la matrice de transition s'écrit

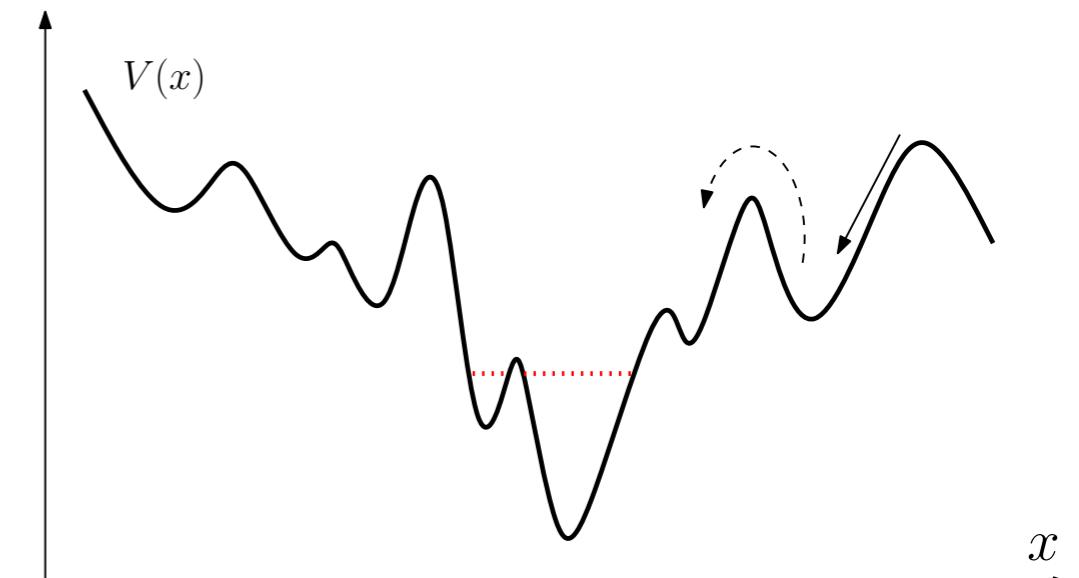
$$P(x, y) = Q(x, y) h \left(\exp \left(\frac{1}{T} (V(x) - V(y)) \right) \frac{Q(y, x)}{Q(x, y)} \right)$$

Le terme Z_T a disparu

Descente de gradient stochastique

Si $V(x) > V(y)$: transition favorisée

Si $V(x) < V(y)$: transition défavorisée



Question.

Comment choisir T pour décrire \mathcal{M} par μ_T ?

Algorithme de recuit simulé

Appliquer Hastings-Metropolis en variant le paramètre T_n à chaque temps n avec $T_n \rightarrow 0$

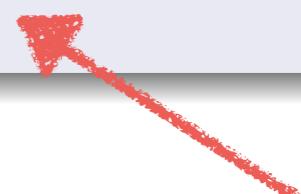
Interprétation. Déplacements initiaux rapides puis fixation dans un minimum quand $n \rightarrow \infty$.

Théorème 6.4

Étant donnée une fonction V , il existe $C(V)$ telle que

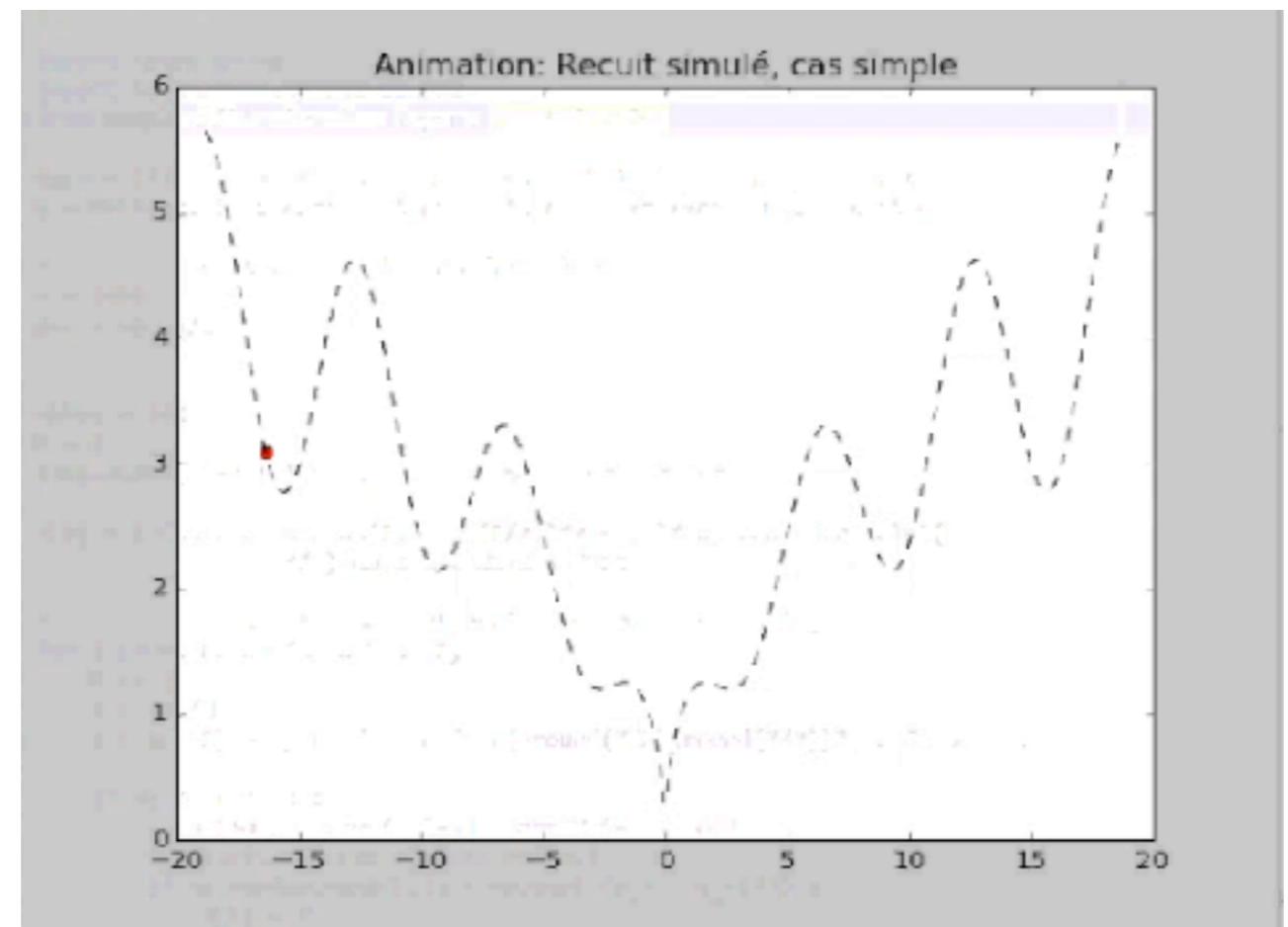
pour $T_n = \frac{C(V)}{\log n}$ alors

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_n \in \mathcal{M}) = 1$$



Minima du potentiel

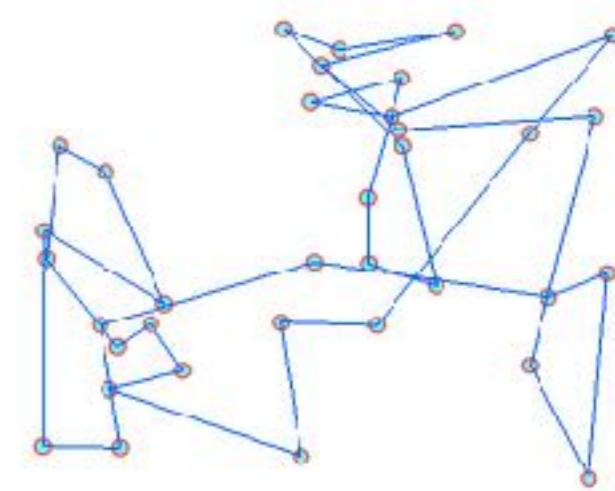
Simulation Python sur moodle



Problème du voyageur de commerce

Les distances entre K villes $\{d(i, j)\}_{\substack{1 \leq i \leq K \\ 1 \leq j \leq K}}$ sont données

Déterminer σ la permutation de $\{1, \dots, K\}$ telle que



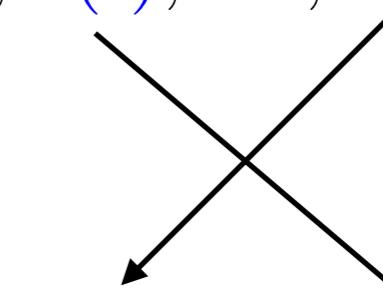
$$\min_{\sigma \in S_K} \{V(\sigma)\} \quad \text{avec} \quad V(\sigma) = \sum_{i=1}^K d(\sigma(i), \sigma(i+1))$$

Recuit simulé :

On choisit 2 villes au hasard et on les permute.

$$Q(\sigma, \sigma^{i,j}) = \frac{2}{K(K-1)}$$

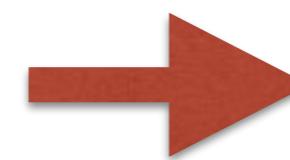
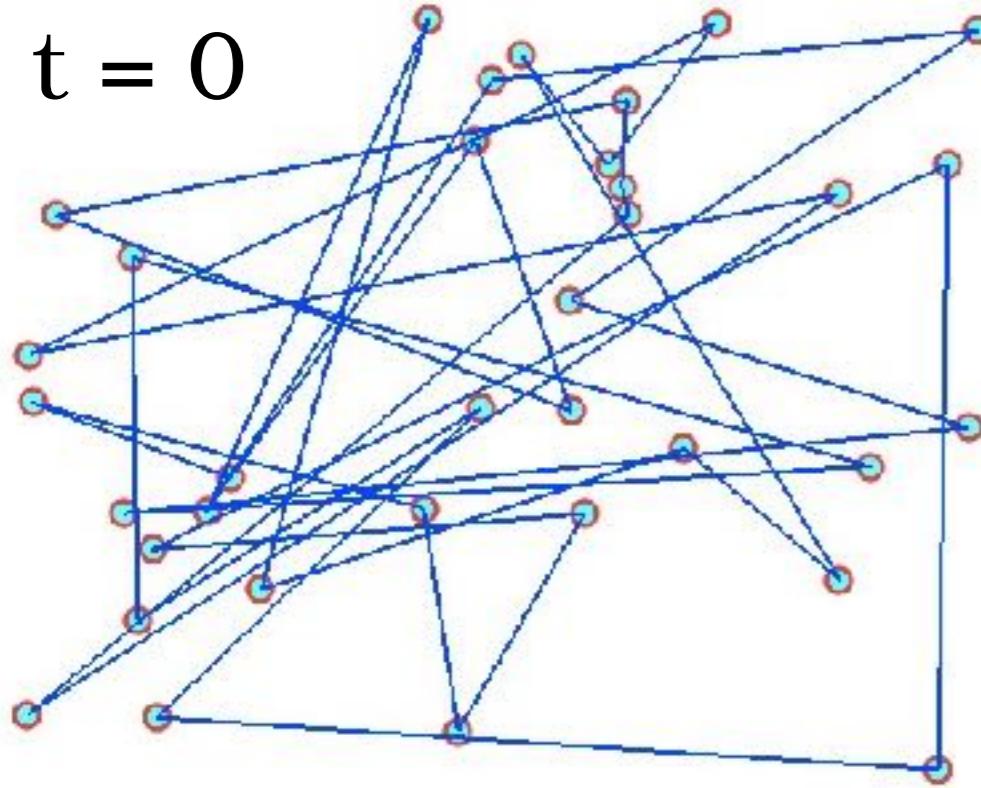
$$\sigma = (\sigma(1), \dots, \sigma(i), \dots, \sigma(j), \dots, \sigma(K))$$



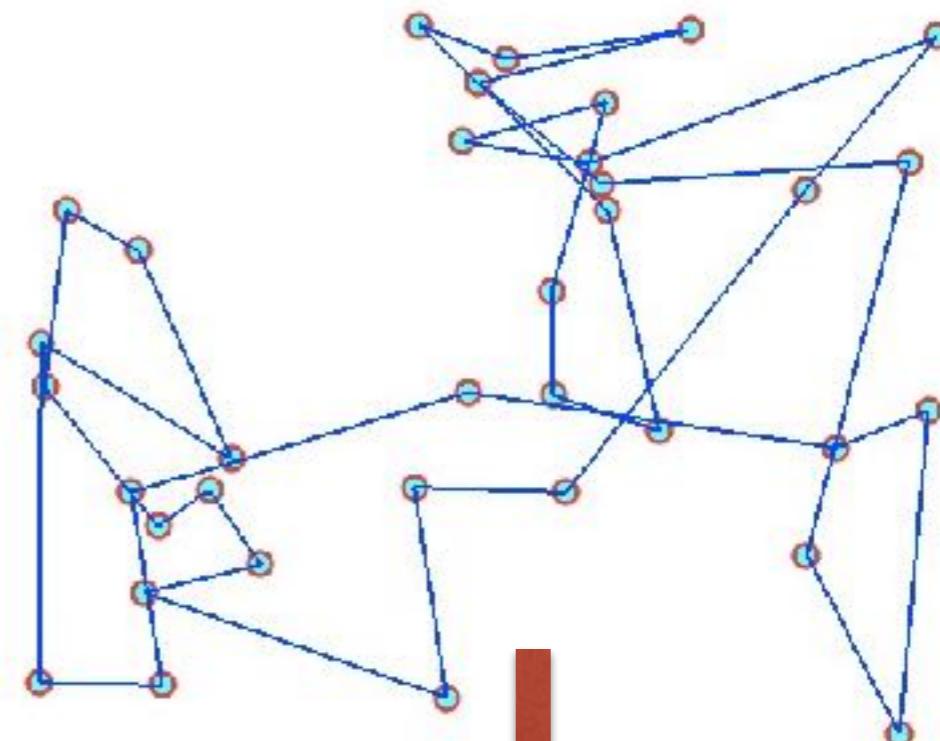
$$\sigma^{i,j} = (\sigma(1), \dots, \sigma(j), \dots, \sigma(i), \dots, \sigma(K))$$

Exemple : Voyageur de commerce

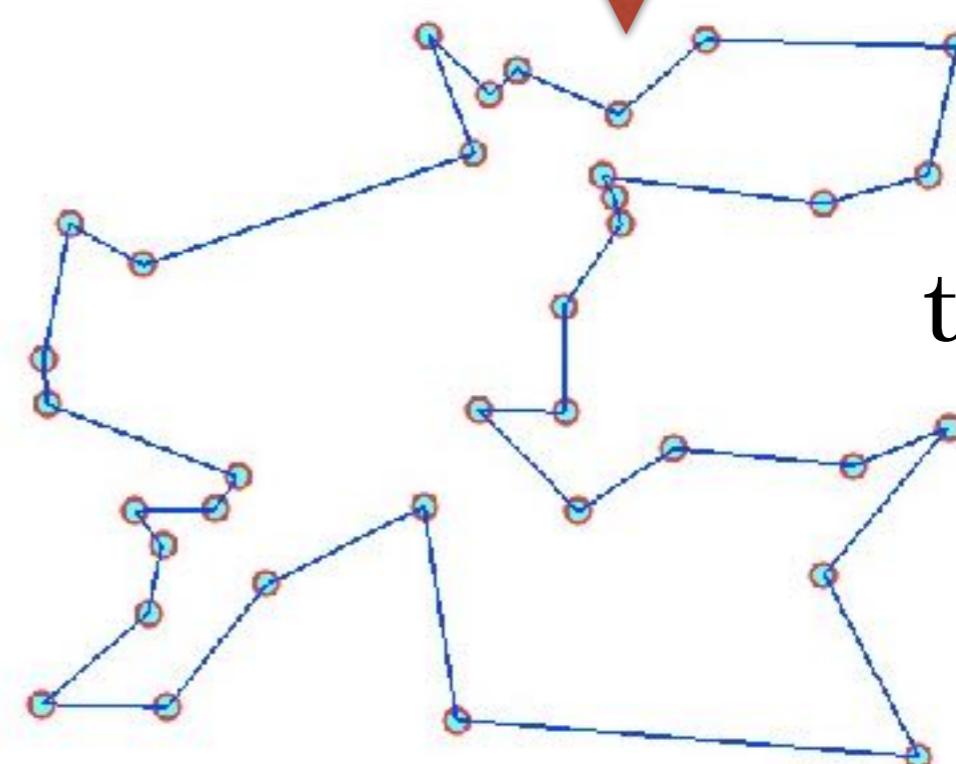
$t = 0$



$t = 2000$



$t = 10^4$



Simulation Python
sur moodle

Modèle d'Ising

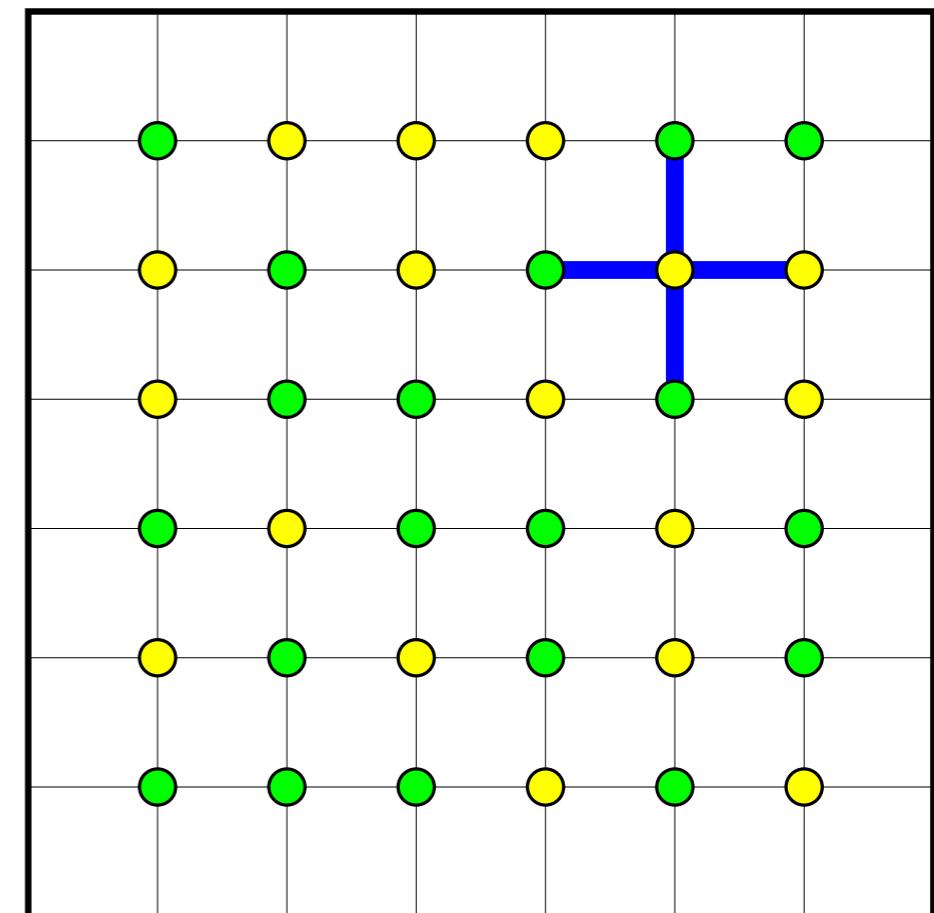
Mesures de Gibbs : Physique statistique

Modèle d'Ising : Transitions de phases

Domaine $\Lambda = \{1, \dots, L\}^d$

Spin au site $i \in \Lambda$: $s_i = \pm 1$

Configuration : $S = \{s_i\}_{i \in \Lambda}$



On associe à la configuration S

L'énergie : $V(S) = - \sum_{\substack{i,j \in \Lambda \\ i \sim j}} s_i s_j$

La probabilité : $\mu_T(S) = \frac{1}{Z_T} \exp \left(-\frac{1}{T} V(S) \right)$

Interprétation : T joue le rôle de la température

Simulation d'Ising par Hastings-Metropolis

$$S = \{s_j\}_{j \in \Lambda} \xrightarrow[\text{du spin en } i]{\text{changement}} S^{(i)} = \begin{cases} S_i^{(i)} = -s_i \\ S_j^{(i)} = s_j, \quad j \neq i \end{cases}$$

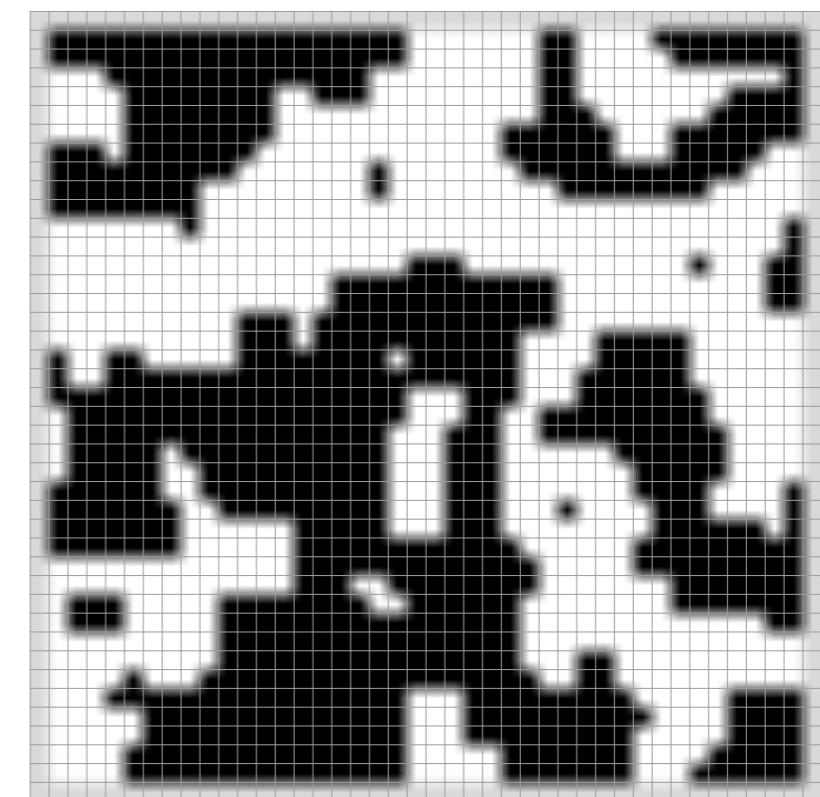
Matrice de transition Q : $\forall i, Q(S, S^{(i)}) = 1/\text{Card}(\Lambda)$

Simulations :

haute température : $T \gg 1$



basse température : $T \ll 1$



Traitement d'images

Image noir et blanc : $L \times L$ pixels pouvant prendre les valeurs ± 1

L'image observée est dégradée : $\Sigma = \{\sigma_i\}_{\Lambda}$ avec $\sigma_i = \pm 1$

On cherche à restaurer l'image Σ en optimisant 2 contraintes :

- Rester proche de l'image initiale
- Supprimer les parasites en forçant les pixels à garder localement la même valeur

L'image restaurée $S = \{s_i\}_{\Lambda}$ minimise

$$V(S) = \beta \sum_{i \in \Lambda} (s_i - \sigma_i)^2 - \alpha \sum_{\substack{i, j \in \Lambda \\ i \sim j}} s_i s_j$$

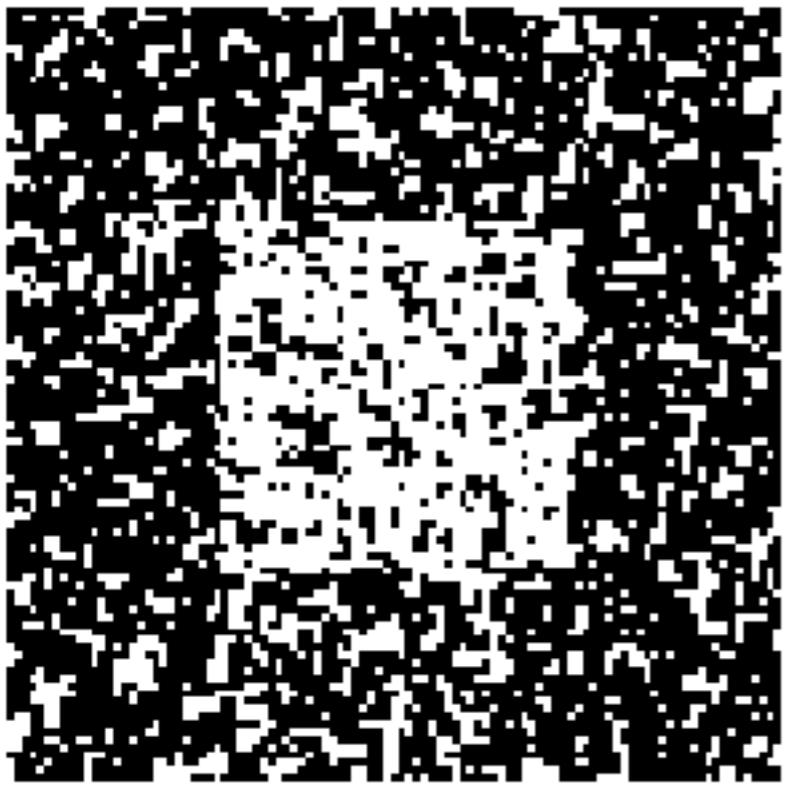
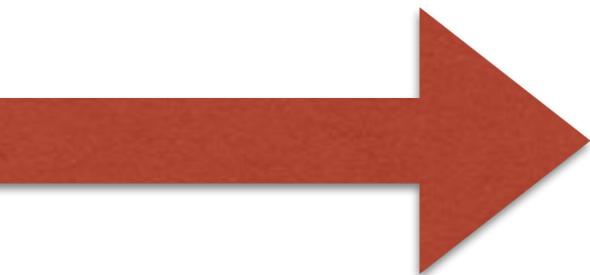
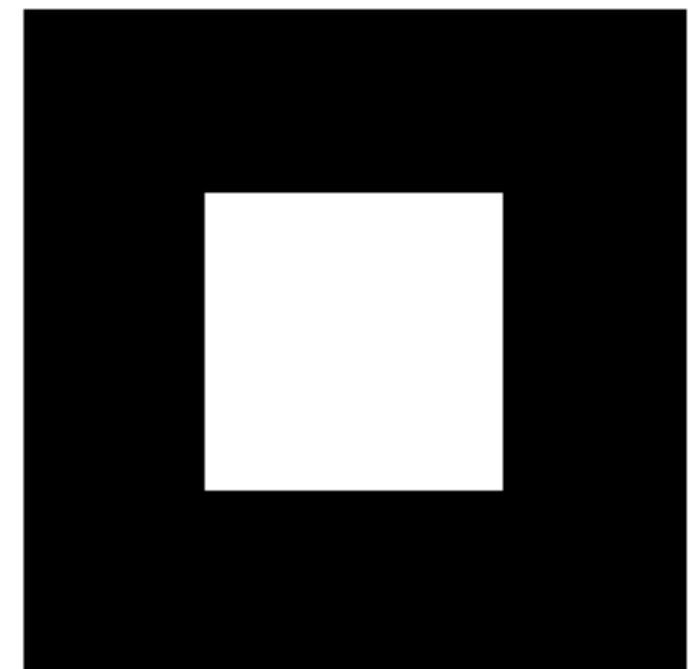


Image bruitée



Après traitement

Image sans bruit



Exemple de segmentation d'images

Cartographie



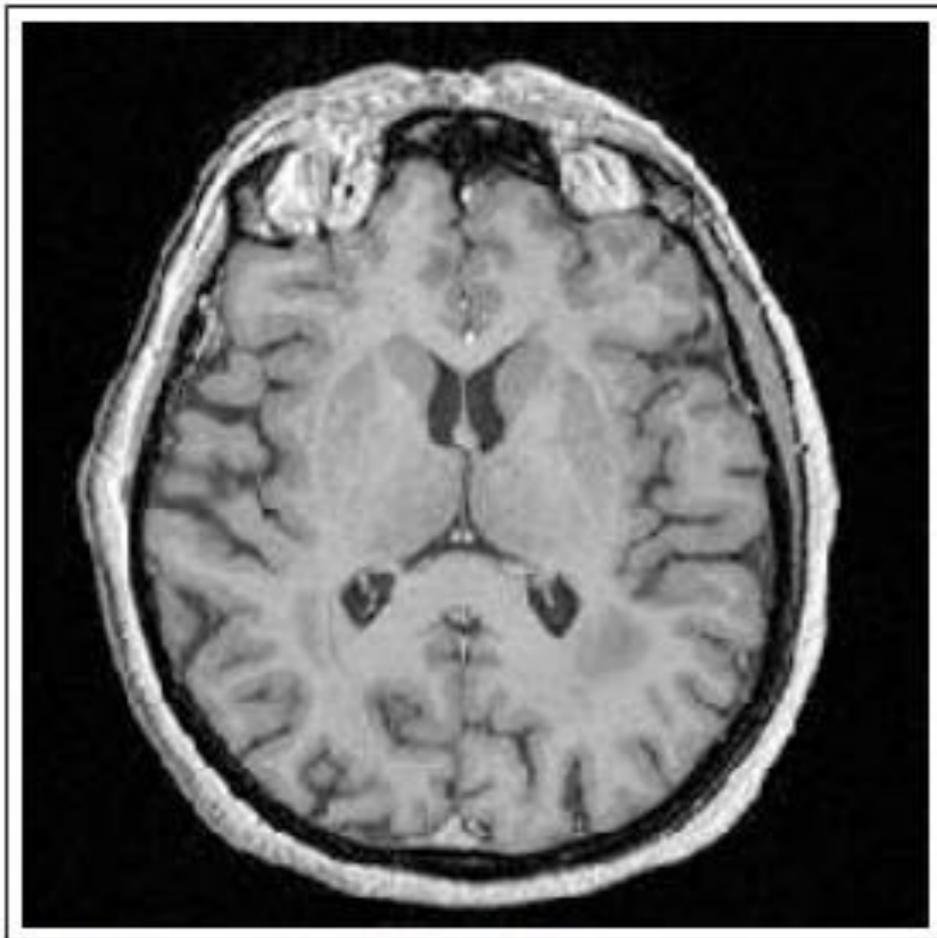
a. Image radar originale ©ERS-1



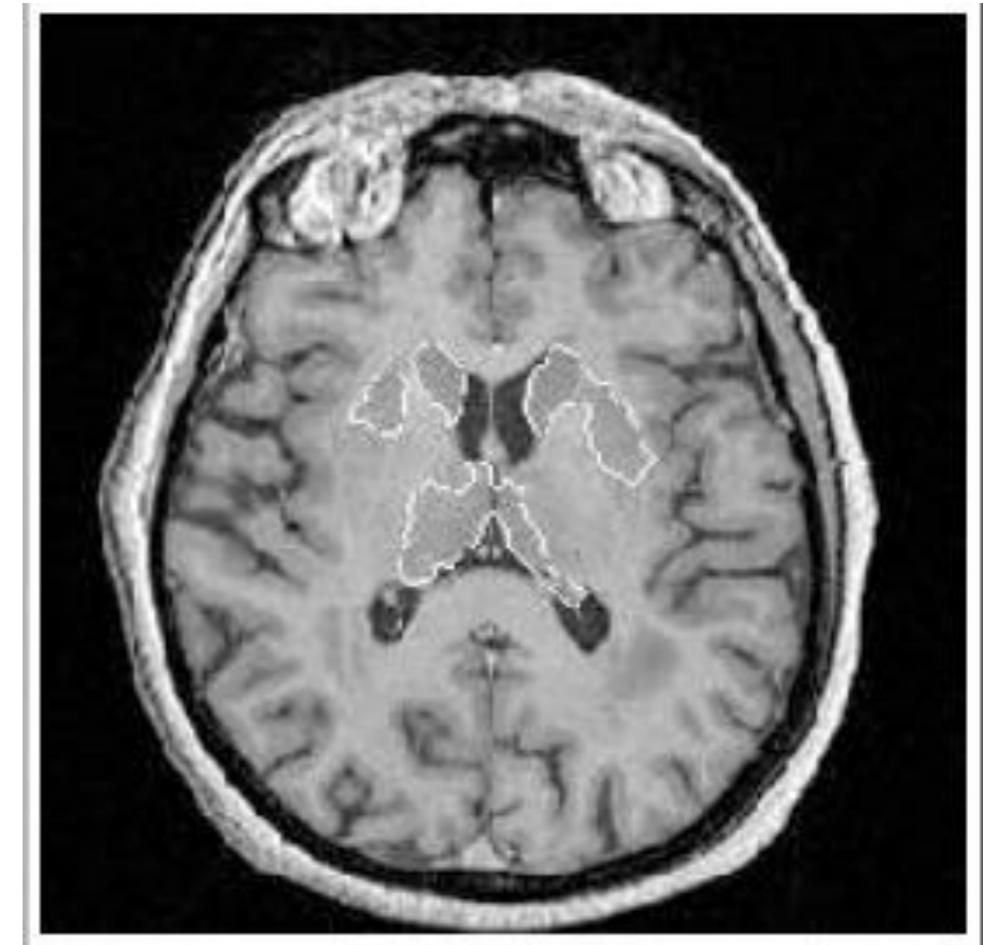
b. Image segmentée en régions

Exemple de segmentation d'images

Imagerie médicale



a. Image cérébrale IRM.



Résultat de la segmentation