

Savez-vous ce qu'est l'échelonnement optimal ?

Échelonnement Optimal de l'Algorithme MALA avec Loi Cible Laplace

Stage de M2, Probabilités et Statistiques

Pablo Jimenez

sous la direction d'Alain Durmus

16 septembre 2019

ENS Paris Saclay, CMLA



Motivations

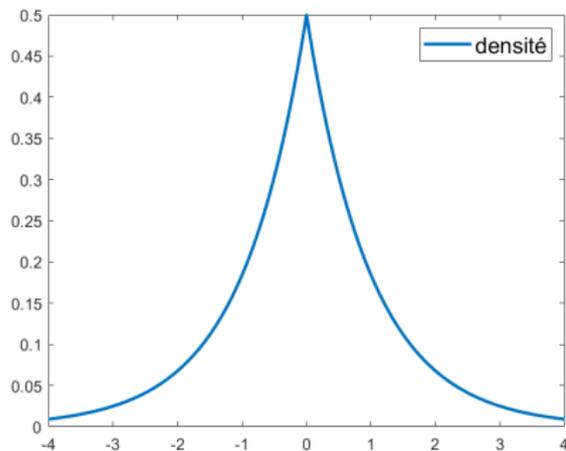
Algorithme de Metropolis-Hastings

Échelonnement optimal

Résultats du stage

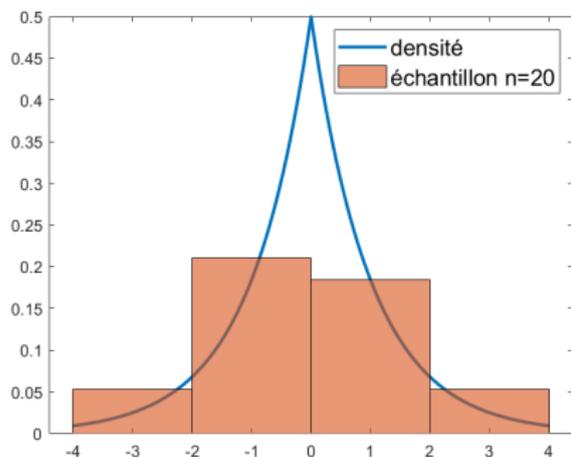
Motivations

Simulation de lois



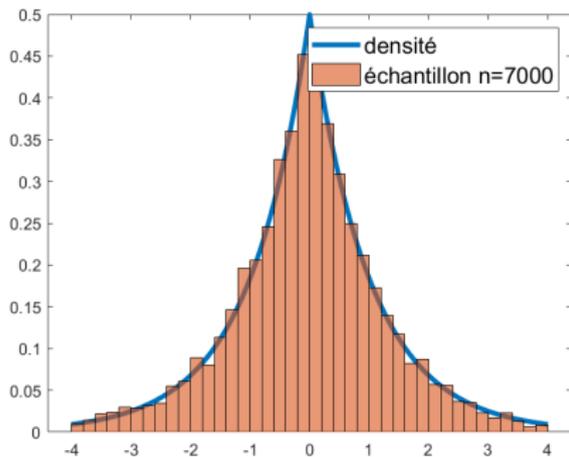
- Donnée : une densité π sur \mathbb{R}^d .
- Objectif : un **échantillon** $(X_1^d, \dots, X_n^d)_n$ i.i.d. de loi π .
- Utilité : approcher des intégrales comme $\int_{\mathbb{R}^d} h(x^d) \pi(x^d) dx^d$.
- [Voir un exemple](#)

Simulation de lois



- Donnée : une densité π sur \mathbb{R}^d .
- Objectif : un **échantillon** $(X_1^d, \dots, X_n^d)_n$ i.i.d. de loi π .
- Utilité : approcher des intégrales comme $\int_{\mathbb{R}^d} h(x^d) \pi(x^d) dx^d$.
- [Voir un exemple](#)

Simulation de lois



- Donnée : une densité π sur \mathbb{R}^d .
- Objectif : un **échantillon** $(X_1^d, \dots, X_n^d)_n$ i.i.d. de loi π .
- Utilité : approcher des intégrales comme $\int_{\mathbb{R}^d} h(x^d) \pi(x^d) dx^d$.
- [Voir un exemple](#)

Méthodes de Monte Carlo : résumé

- Utilité : approcher des intégrales comme $\int_{\mathbb{R}^d} h(x^d)\pi(x^d)dx^d$.
- On utilise $n^{-1} \sum_{k=1}^n h(X_k^d)$.
- Avantageux en **grande dimension** \rightarrow inférence bayésienne.
- Problème : $(X_n^d)_{n \in \mathbb{N}^*}$ i.i.d. de loi $\pi \rightarrow$ coûteux/impossible.

Méthodes de Monte Carlo : résumé

- Utilité : approcher des intégrales comme $\int_{\mathbb{R}^d} h(x^d)\pi(x^d)dx^d$.
- On utilise $n^{-1} \sum_{k=1}^n h(X_k^d)$.
- Avantageux en **grande dimension** \rightarrow inférence bayésienne.
- Problème : $(X_n^d)_{n \in \mathbb{N}^*}$ i.i.d. de loi $\pi \rightarrow$ coûteux/impossible.
- Solution : $(X_n^d)_{n \in \mathbb{N}^*}$ **chaîne de Markov** approchant π .

Méthodes de Monte Carlo par chaîne de Markov (MCMC)



N. Metropolis
(1915-1999)

- Metropolis et al. (1953)
- Hastings (1970)



W.K. Hastings
(1930-2016)

Algorithme de Metropolis-Hastings

Description de l'algorithme de Metropolis-Hastings

1. Partant de $X_n^d \in \mathbb{R}^d$, simuler la **proposition** Y_{n+1}^d de loi $q^d(X_n^d, \cdot)$.
2. Calculer la **probabilité d'acceptation** $\alpha(X_n^d, Y_{n+1}^d)$, où

$$\alpha(x^d, y^d) = 1 \wedge \frac{\pi(y^d) q^d(y^d, x^d)}{\pi(x^d) q^d(x^d, y^d)}.$$

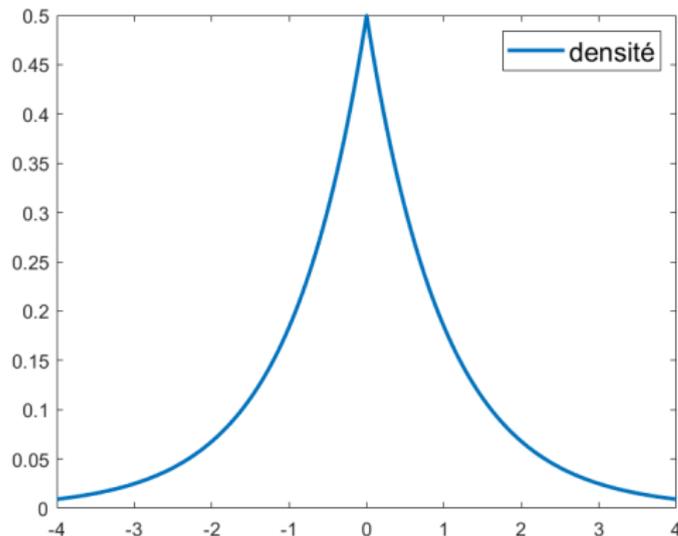
3. Simuler U_{n+1} uniforme sur $[0, 1]$.
4. Évaluer l'**évènement d'acceptation** A_{n+1}^d , défini par

$$A_{n+1}^d = \left\{ U_{n+1} \leq \alpha(X_n^d, Y_{n+1}^d) \right\}.$$

5. Résultat : X_{n+1}^d est donné par

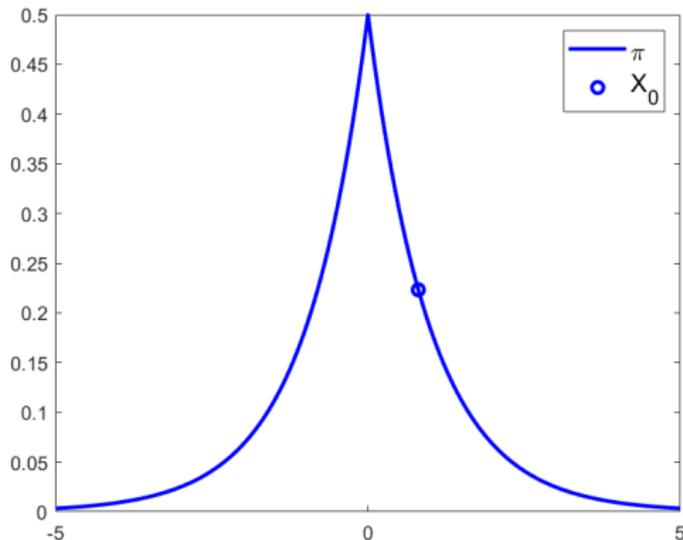
$$X_{n+1}^d = \underbrace{Y_{n+1}^d \mathbb{1}_{A_{n+1}^d}}_{\text{acceptation}} + \underbrace{X_n^d (1 - \mathbb{1}_{A_{n+1}^d})}_{\text{rejet}}.$$

Mise en pratique de l'algorithme



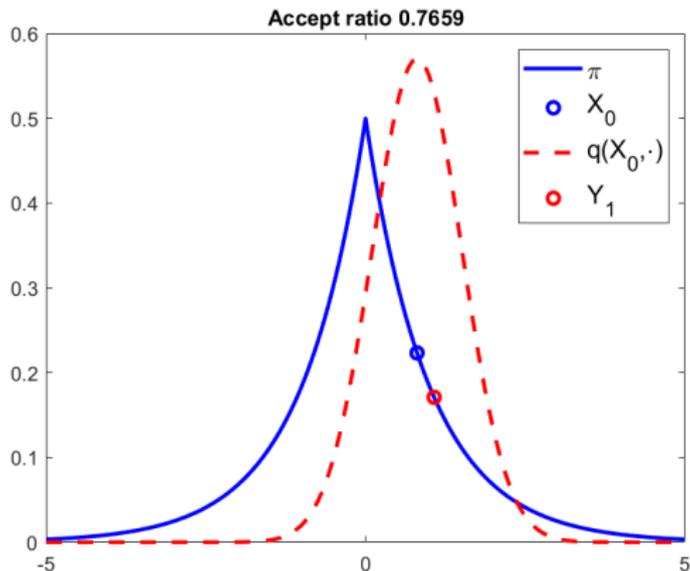
1. Simuler X_0^d .
2. Simuler Y_1^d .
3. Calculer la probabilité d'acceptation $\alpha(X_n^d, Y_{n+1}^d)$.
4. Évaluer si Y_1^d est accepté ou rejeté.

Mise en pratique de l'algorithme



1. Simuler X_0^d .
2. Simuler Y_1^d .
3. Calculer la probabilité d'acceptation $\alpha(X_n^d, Y_{n+1}^d)$.
4. Évaluer si Y_1^d est accepté ou rejeté.

Mise en pratique de l'algorithme

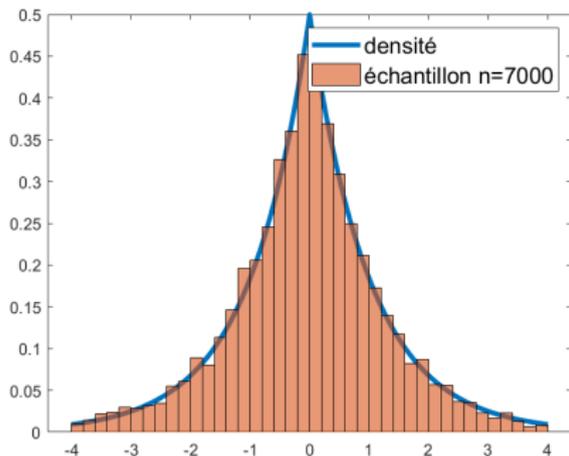


1. Simuler X_0^d .
2. Simuler Y_1^d .
3. Calculer la probabilité d'acceptation $\alpha(X_n^d, Y_{n+1}^d)$.
4. Évaluer si Y_1^d est accepté ou rejeté.

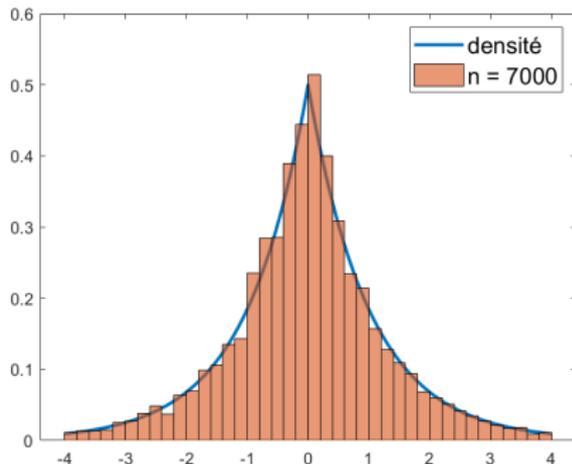
Algorithme de Metropolis-Hastings : résultats

Théorème : Tierney (1994)

Sous des hypothèses raisonnables sur q^d , $(X_n^d)_n$ est une chaîne de Markov **ergodique** admettant π comme loi **invariante**.



Échantillon i.i.d.



Échantillon par chaîne de Markov

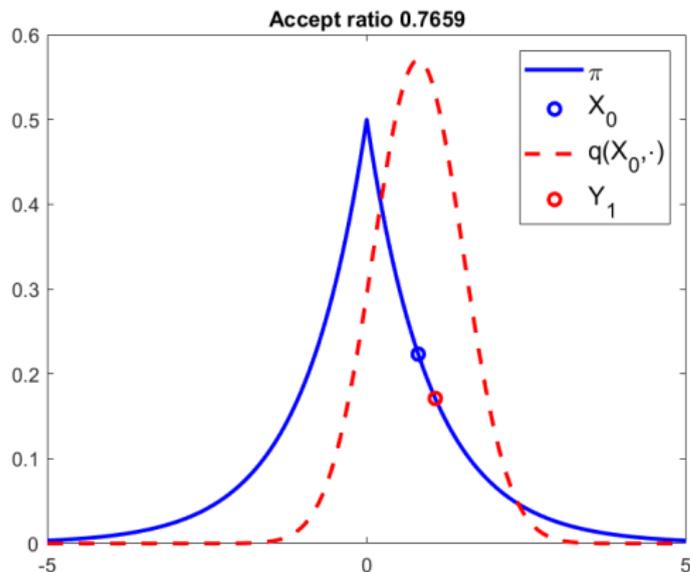
Théorème : Tierney (1994)

Sous des hypothèses raisonnables sur q^d , $(X_n^d)_n$ est une chaîne de Markov **ergodique** admettant π comme loi **invariante**.

Hypothèses: q^d et π **strictement positives** sur \mathbb{R}^d .

Quels choix pour la **proposition** q^d ?

Exemple : Metropolis à marche aléatoire (RWM)



Étape de Metropolis à marche aléatoire.

- $Y_{n+1}^d = X_n^d + Z_{n+1}^d$, avec Z_{n+1}^d gaussienne centrée.
- Proposition $q^d(x^d, \cdot)$ gaussienne centrée $N(x^d, \sigma_d^2 \text{Id}_d)$.
- Calibrer σ_d^2 en fonction de la cible π :
 - assez grand pour limiter les corrélations,
 - assez petit pour une acceptation élevée.

Exemple : dynamique de Langevin

- Supposons $\pi : x^d \mapsto \exp(-V(x^d))/Z$, avec Z la constante de normalisation, et $V : \mathbb{R}^d \rightarrow \mathbb{R}$ régulière.
- Équation de **diffusion** du processus $(\mathbf{X}_t^d)_{t \geq 0}$:

$$d\mathbf{X}_t^d = -\nabla V(\mathbf{X}_t^d) dt + \sqrt{2}d\mathbf{B}_t^d,$$

avec \mathbf{B}^d un mouvement brownien standard sur \mathbb{R}^d .

- Roberts and Tweedie (1996) : $(\mathbf{X}_t^d)_{t \geq 0}$ est **ergodique** avec π comme loi **invariante**.
- Problème: **difficile à simuler**.
- Idée : discrétiser l'EDS par Euler-Murayama.

Exemple : dynamique de Langevin discrétisée

- Équation de **diffusion** du processus $(\mathbf{X}_t^d)_{t \geq 0}$:

$$d\mathbf{X}_t^d = -\nabla V(\mathbf{X}_t^d) dt + \sqrt{2} d\mathbf{B}_t^d,$$

avec \mathbf{B}^d un mouvement brownien standard sur \mathbb{R}^d .

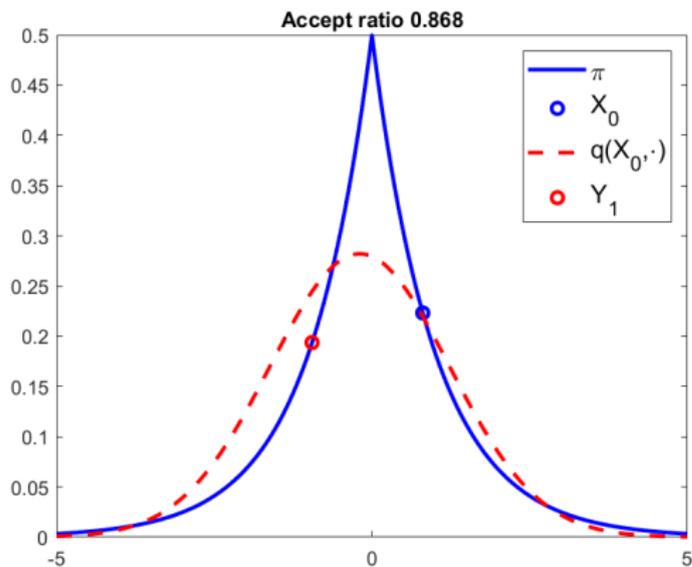
- Discrétisation $(X_k^d)_{k \geq 0}$ d'Euler-Maruyama de pas $\sigma_d^2 > 0$ définit une **chaîne de Markov**:

$$X_{k+1}^d = X_k^d - \sigma_d^2 \nabla V(X_k^d) + \sqrt{2\sigma_d^2} Z_{k+1}^d,$$

avec les suites $\{(Z_k^d)_{k \geq 1} \mid d \in \mathbb{N}^*\}$ indépendantes et $(Z_k^d)_{k \geq 1}$ i.i.d. de loi $N(0, \text{Id}_d)$ indépendants.

- Problème : π n'est **pas invariante** pour $(X_k^d)_{k \geq 1}$.
- Idée : utiliser cette dynamique comme **proposition** dans M-H.

Exemple : algorithme de Langevin ajusté (MALA)



Étape de l'algorithme de Langevin ajusté.

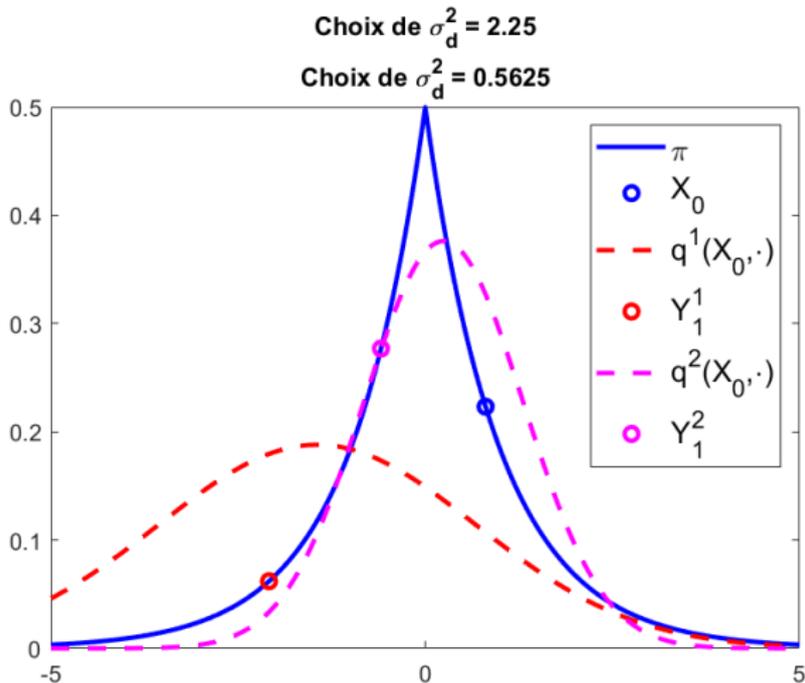
- Propositions $(Y_n^d)_n$ données par discrétisation :

$$Y_{n+1}^d = X_n^d - \sigma_d^2 \nabla V(X_n^d) + \sqrt{2\sigma_d^2} Z_{n+1}^d,$$

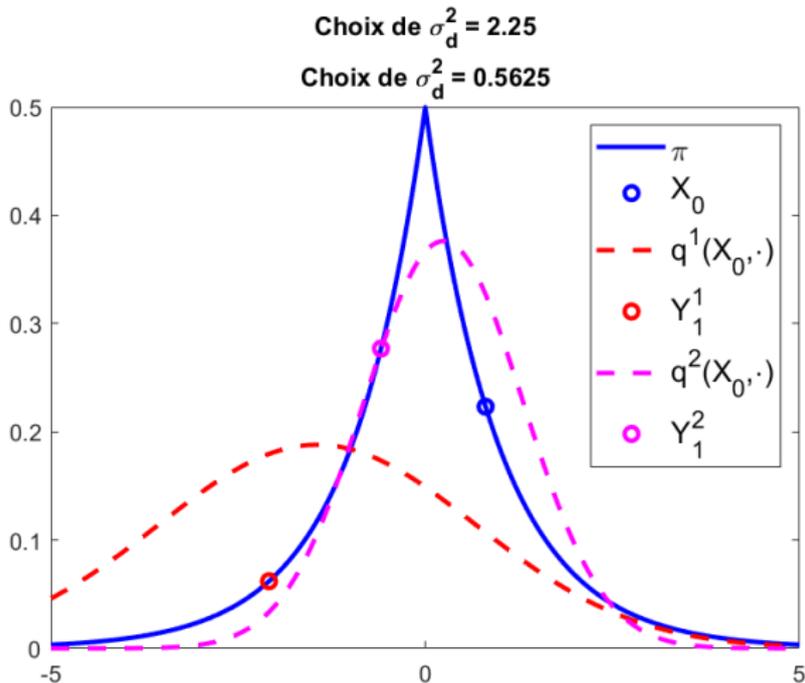
avec Z_{n+1}^d de loi $N(0, \text{Id}_d)$.

- **Proposition gaussienne biaisée** vers fortes densités.
- Calibrer σ_d^2 en fonction de la **cible** π .

Comment faire un **choix optimal** σ_d^2 en fonction de d et π ?



Comment faire un **choix optimal** σ_d^2 en fonction de d et π ?



Avant de répondre, donner du sens à "**choix optimal**".

Échelonnement optimal

Comment choisir la variance de la proposition ?

Dans le cas de RWM et MALA : effectuer un bon choix de σ_d^2 et quantifier l'efficacité des méthodes.

Objectifs :

1. approcher le comportement en **grande dimension** par un objet plus simple,
2. trouver un **critère d'optimalité** pour σ_d^2 ,
3. pouvoir **comparer** les méthodes.

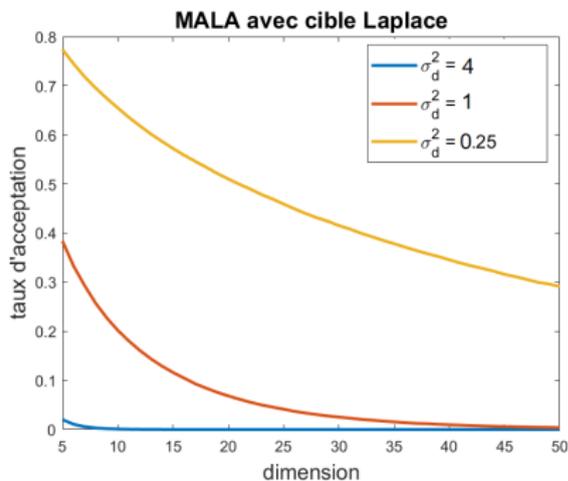
Un cadre asymptotique

- A partir de π densité sur \mathbb{R} , construire une suite $(\pi^d)_{d \geq 1}$

$$\pi^d(x^d) = \prod_{i=1}^d \pi(x_i^d) .$$

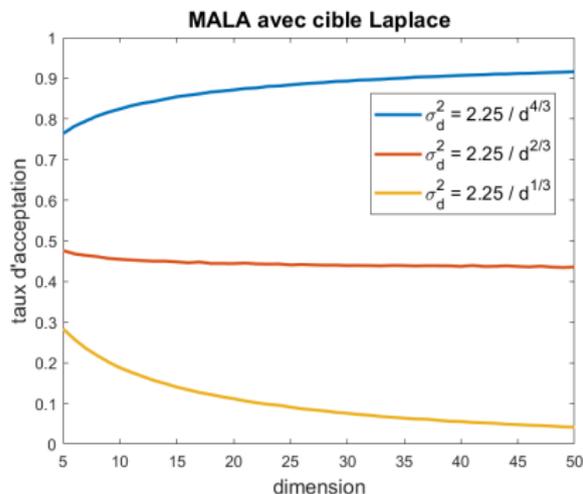
- Construire la suite de chaînes de Markov $\{(X_n^d)_{n \geq 0} \mid d \geq 1\}$, démarrées à l'équilibre, sur un même espace probabilisé.
- But : étudier la limite de $\mathbb{P}(A_1^d)$ lorsque $d \rightarrow +\infty$.

$$\bullet A_1^d = \left\{ U_1 \leq \alpha(X_0^d, Y_1^d) \right\} .$$



- Problème : Si σ_d^2 constant avec d , $\lim_{d \rightarrow +\infty} \mathbb{P}(A_1^d) = 0$.

Un premier critère pour σ_d^2 : l'échelonnement



Exemple d'échelonnement en $\alpha = 2/3$.

- Choix : $\sigma_d^2 = \ell^2 / d^\alpha$ avec $\ell > 0$ et $\alpha > 0$.
- But : chercher α tel que

$$a(\ell) = \lim_{d \rightarrow +\infty} \mathbb{P}(A_1^d) \in]0, 1[$$

pour tout $\ell > 0$.

- On dit alors que l'échelonnement de la chaîne est en $1/d^\alpha$.

Un résultat de l'échelonnement

- Interpoler et échelonner X^d en Y^d , avec

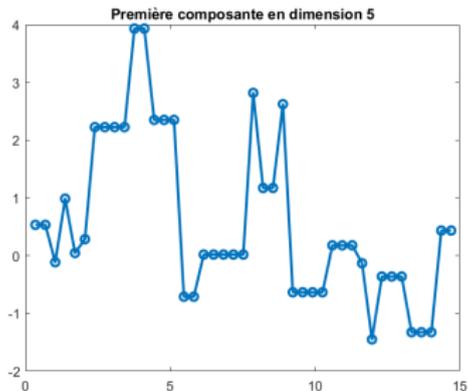
$$Y_t^d = ([d^\alpha t] - d^\alpha t) X_{[d^\alpha t]}^d + (d^\alpha t - [d^\alpha t]) X_{[d^\alpha t]}^d.$$

- Roberts et al. (1997) : il existe une diffusion Y tel que

$$Y_{(1)}^d \xrightarrow[d \rightarrow +\infty]{\text{loi}} Y.$$

- Y vérifie l'EDS :

$$dY_t = h(\ell) (\log \pi)'(Y_t) dt + \sqrt{2h(\ell)} d\mathbf{B}_t.$$



Interpolation et échelonnement de la chaîne.

Un résultat de l'échelonnement

- Interpoler et échelonner X^d en Y^d , avec

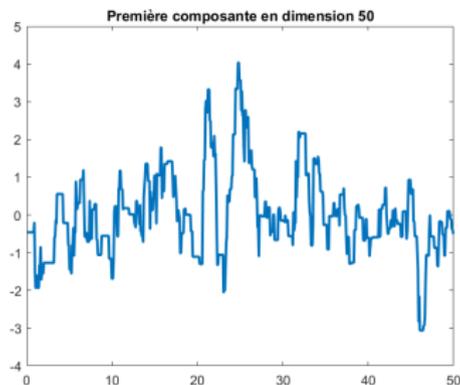
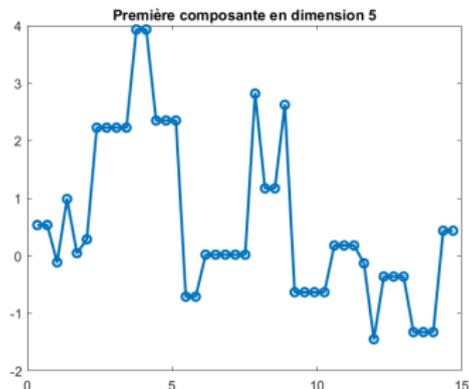
$$Y_t^d = ([d^\alpha t] - d^\alpha t) X_{[d^\alpha t]}^d + (d^\alpha t - [d^\alpha t]) X_{[d^\alpha t]}^d.$$

- Roberts et al. (1997) : il existe une diffusion Y tel que

$$Y_{(1)}^d \xrightarrow[d \rightarrow +\infty]{\text{loi}} Y.$$

- Y vérifie l'EDS :

$$dY_t = h(\ell) (\log \pi)'(Y_t) dt + \sqrt{2h(\ell)} dB_t.$$



Un résultat de l'échelonnement

- Interpoler et échelonner X^d en Y^d , avec

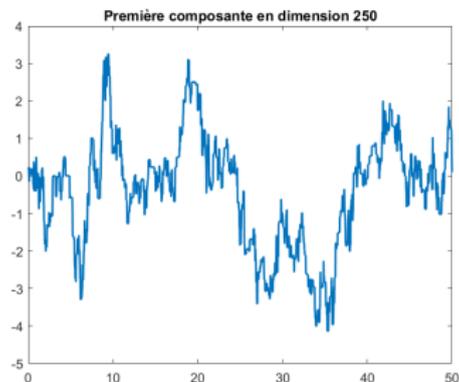
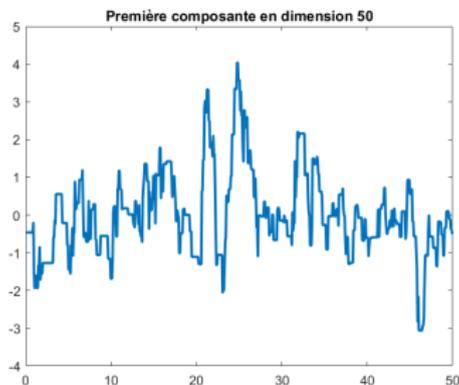
$$Y_t^d = ([d^\alpha t] - d^\alpha t) X_{[d^\alpha t]}^d + (d^\alpha t - [d^\alpha t]) X_{[d^\alpha t]}^d.$$

- Roberts et al. (1997) : il existe une diffusion Y tel que

$$Y_{(1)}^d \xrightarrow[d \rightarrow +\infty]{\text{loi}} Y.$$

- Y vérifie l'EDS :

$$dY_t = h(\ell) (\log \pi)'(Y_t) dt + \sqrt{2h(\ell)} dB_t.$$



Un résultat de l'échelonnement

- Interpoler et **échelonner** X^d en Y^d , avec

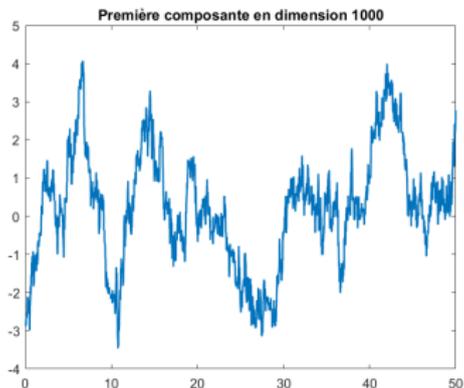
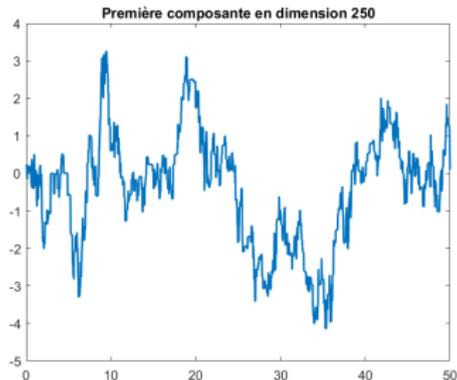
$$Y_t^d = ([d^\alpha t] - d^\alpha t) X_{[d^\alpha t]}^d + (d^\alpha t - [d^\alpha t]) X_{[d^\alpha t]}^d.$$

- Roberts et al. (1997) : il existe une diffusion Y tel que

$$Y_{(1)}^d \xrightarrow[d \rightarrow +\infty]{\text{loi}} Y.$$

- Y vérifie l'EDS :

$$dY_t = h(\ell) (\log \pi)'(Y_t) dt + \sqrt{2h(\ell)} dB_t.$$



Un résultat de l'échelonnement

- Interpoler et échelonner X^d en Y^d , avec

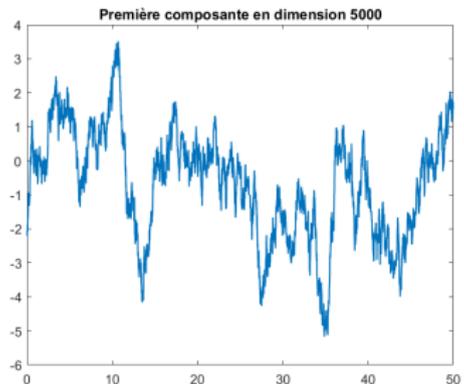
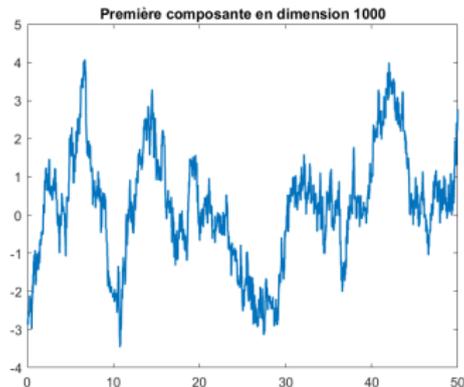
$$Y_t^d = ([d^\alpha t] - d^\alpha t) X_{[d^\alpha t]}^d + (d^\alpha t - [d^\alpha t]) X_{[d^\alpha t]}^d.$$

- Roberts et al. (1997) : il existe une diffusion Y tel que

$$Y_{(1)}^d \xrightarrow[d \rightarrow +\infty]{\text{loi}} Y.$$

- Y vérifie l'EDS :

$$dY_t = h(\ell) (\log \pi)'(Y_t) dt + \sqrt{2h(\ell)} dB_t.$$



Un résultat de l'échelonnement

- Interpoler et **échelonner** X^d en Y^d , avec

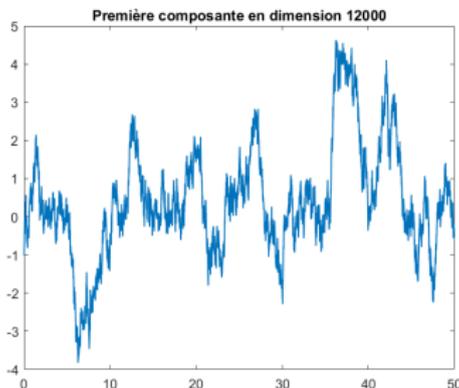
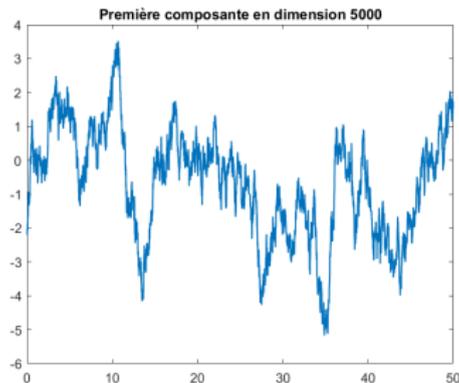
$$Y_t^d = ([d^\alpha t] - d^\alpha t) X_{[d^\alpha t]}^d + (d^\alpha t - [d^\alpha t]) X_{[d^\alpha t]}^d.$$

- Roberts et al. (1997) : il existe une diffusion Y tel que

$$Y_{(1)}^d \xrightarrow[d \rightarrow +\infty]{\text{loi}} Y.$$

- Y vérifie l'EDS :

$$dY_t = h(\ell) (\log \pi)'(Y_t) dt + \sqrt{2h(\ell)} dB_t.$$



Un résultat de l'échelonnement

- Interpoler et **échelonner** X^d en Y^d , avec

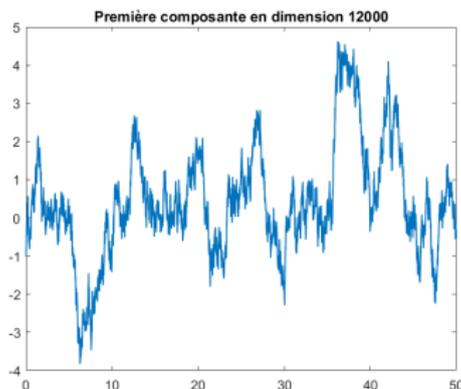
$$Y_t^d = ([d^\alpha t] - d^\alpha t) X_{[d^\alpha t]}^d + (d^\alpha t - [d^\alpha t]) X_{[d^\alpha t]}^d.$$

- Roberts et al. (1997) : il existe une diffusion Y tel que

$$Y_{(1)}^d \xrightarrow[d \rightarrow +\infty]{\text{loi}} Y.$$

- Y vérifie l'EDS :

$$dY_t = h(\ell) (\log \pi)'(Y_t) dt + \sqrt{2h(\ell)} d\mathbf{B}_t.$$



Comment choisir la variance de la proposition ?

Objectifs :

1. approcher le comportement en **grande dimension** par un objet plus simple,
2. trouver un **critère d'optimalité** pour σ_d^2 ,
3. pouvoir **comparer** les méthodes.

Comment choisir la variance de la proposition ?

Objectifs :

1. approcher le comportement en **grande dimension** par un objet plus simple, ✓ **convergence en loi vers une diffusion**
2. trouver un **critère d'optimalité** pour σ_d^2 ,
3. pouvoir **comparer** les méthodes.

Conséquences de la limite en loi

- \mathbf{Y} vérifie l'EDS :

$$d\mathbf{Y}_t = h(\ell) (\log \pi)' (\mathbf{Y}_t) dt + \sqrt{2h(\ell)} d\mathbf{B}_t.$$

- La vitesse h est la limite de l'efficacité d'ordre un :

$$2h(\ell) = \lim_{d \rightarrow +\infty} d^\alpha \mathbb{E} \left[(X_0^d - X_1^d)^2 \right].$$

- Si $\mathbf{Y}^{(1)}$ vérifie la même EDS avec $h = 1$, alors

$$\left(\mathbf{Y}_{h(\ell)t}^{(1)} \right)_t \stackrel{\text{loi}}{=} (\mathbf{Y}_t)_t.$$

- Roberts and Tweedie (1996) : $\mathbf{Y}^{(1)}$ est **ergodique** et admet π comme loi invariante.
- Conséquence : **maximiser h** donne une **meilleure convergence**.
- Échelonnement du temps en d^α : α est un ordre de grandeur de la **complexité**.

Comment choisir la variance de la proposition ?

Objectifs :

1. approcher le comportement en **grande dimension** par un objet plus simple, ✓ **convergence en loi vers une diffusion**,
2. trouver un **critère d'optimalité** pour σ_d^2 ,
3. pouvoir **comparer** les méthodes.

Comment choisir la variance de la proposition ?

Objectifs :

1. approcher le comportement en **grande dimension** par un objet plus simple, ✓ **convergence en loi vers une diffusion**,
2. trouver un **critère d'optimalité** pour σ_d^2 , ✓ $\sigma_d^2 = \ell^2 / d^\alpha$ où ℓ **maximise h** ,
3. pouvoir **comparer** les méthodes.

Comment choisir la variance de la proposition ?

Objectifs :

1. approcher le comportement en **grande dimension** par un objet plus simple, ✓ **convergence en loi vers une diffusion**,
2. trouver un **critère d'optimalité** pour σ_d^2 , ✓ $\sigma_d^2 = \ell^2 / d^\alpha$ où ℓ maximise h ,
3. pouvoir **comparer** les méthodes, ✓ on **compare** la valeur α donnée par l'**échelonnement** (meilleur lorsque α petit).

- **RWM** : échelonnement en $\alpha = 1$ pour les densités régulières par Roberts et al. (1997).
- **MALA** : échelonnement en $\alpha = 1/3$ pour les densités régulières par Roberts and Rosenthal (1997).
- **RWM** : échelonnement en $\alpha = 1$ pour les densités peu régulières (dérivée faible en moyenne L^p) par Durmus et al. (2017).

État de la recherche : comparaison des méthodes

- **RWM** : échelonnement en $\alpha = 1$ pour les densités régulières par Roberts et al. (1997).
- **MALA** : échelonnement en $\alpha = 1/3$ pour les densités régulières par Roberts and Rosenthal (1997).
- **RWM** : échelonnement en $\alpha = 1$ pour les densités peu régulières (dérivée faible en moyenne L^p) par Durmus et al. (2017).
- Question : **MALA** pour densités peu régulières ?
- Étude du cas $\pi : x \mapsto e^{-|x|}/2$.

Résultats du stage

Échelonnement de MALA avec loi cible Laplace

- **Cible** $\pi : x \mapsto e^{-|x|}/2$.
- Mise à jour de la chaîne :

$$X_{n+1}^d = X_n^d + \left(-\frac{\ell^2}{d^\alpha} \operatorname{sgn}(X_n^d) + \sqrt{2} \frac{\ell}{d^{\alpha/2}} Z_{n+1}^d \right) \mathbb{1}_{A_{n+1}^d}.$$

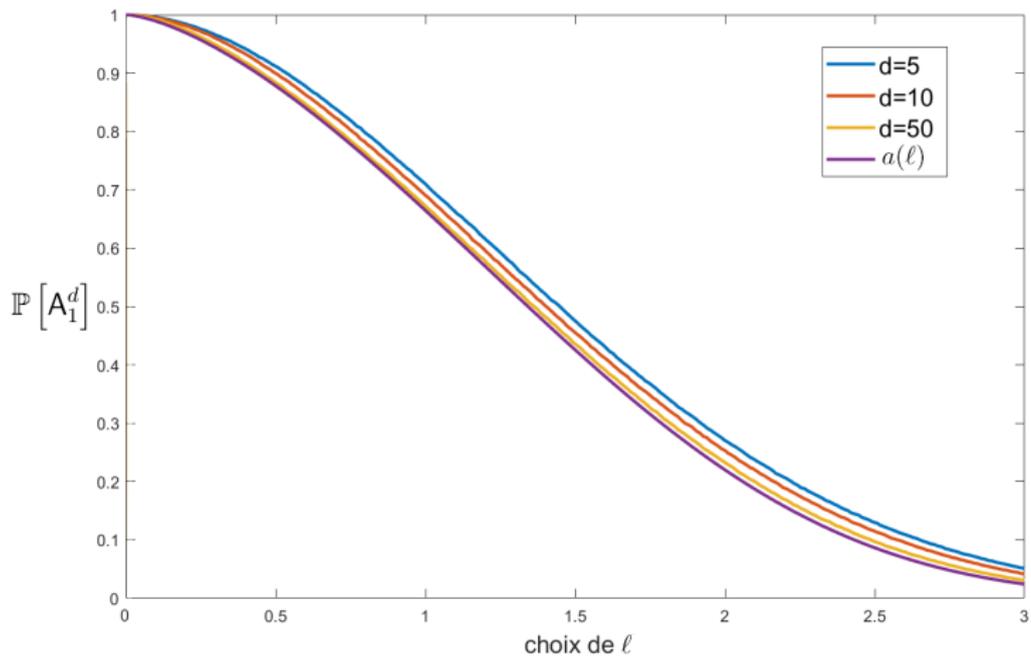
- Rappel : $\mathbb{P}(A_1^d) = \mathbb{E} [\alpha(X_0^d, Y_1^d)]$.

Théorème (valeur d'échelonnement)

Soit $\sigma_d^2 = \ell^2/d^\alpha$, avec $\alpha = 2/3$, et Φ la f.d.r. de la loi $N(0, 1)$.

Alors $\lim_{d \rightarrow +\infty} \mathbb{P}[A_1^d] = a(\ell)$, avec $a(\ell) = 2\Phi(-\ell^{3/2}/(3\pi^{1/2})^{1/2})$.

Illustration



Taux d'acceptation en fonction de ℓ pour MALA avec cible Laplace.

Théorème (diffusion limite et échelonnement optimal)

- $\alpha = 2/3$ et $(\mathbf{Y}_t^d)_{t \geq 0}$ défini par

$$\mathbf{Y}_t^d = \left(\lceil d^{2/3} t \rceil - d^{2/3} t \right) X_{\lfloor d^{2/3} t \rfloor}^d + \left(d^{2/3} t - \lfloor d^{2/3} t \rfloor \right) X_{\lceil d^{2/3} t \rceil}^d.$$

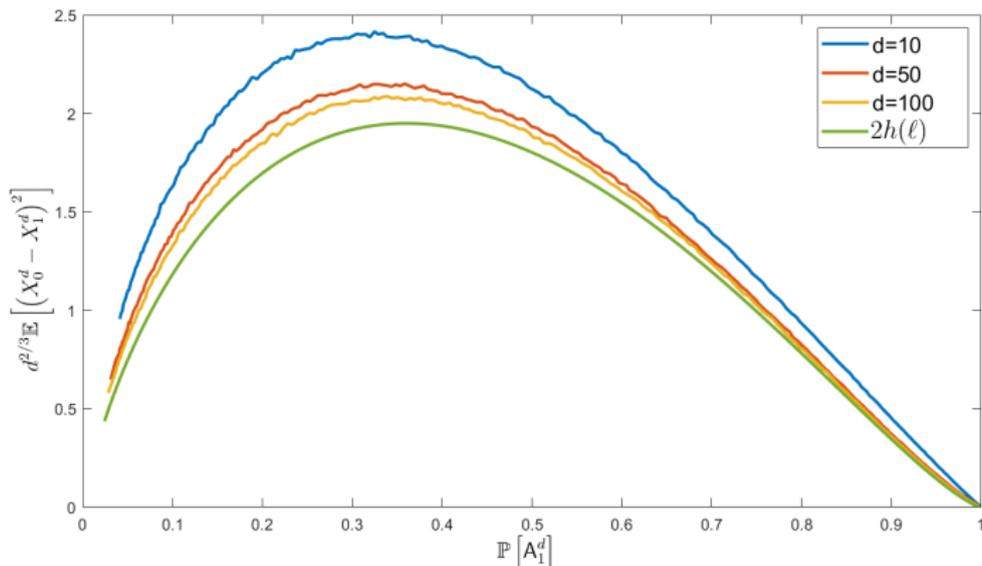
- Soit $(\mathbf{Y}_t)_{t \geq 0}$ solution de l'EDS

$$d\mathbf{Y}_t = h(\ell) \operatorname{sgn}(\mathbf{Y}_t) dt + \sqrt{2h(\ell)} d\mathbf{B}_t,$$

avec \mathbf{B} un mouvement brownien standard sur \mathbb{R} .

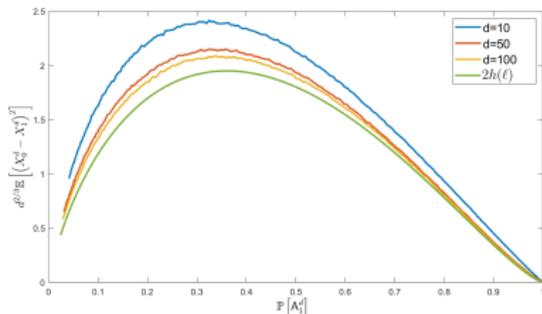
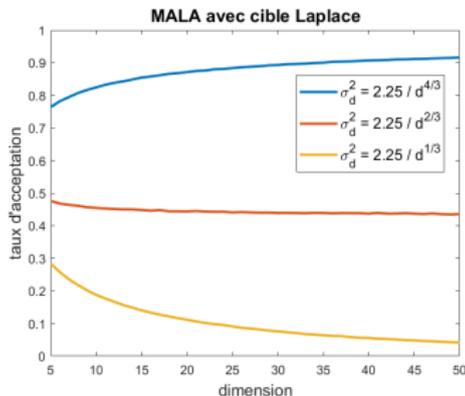
- Alors $\{(\mathbf{Y}_{t,1}^d)_{t \geq 0} \mid d \in \mathbb{N}^*\}$ converge faiblement vers $(\mathbf{Y}_t)_{t \geq 0}$.
- De plus $h(\ell) = \ell^2 \Phi(-\ell^{3/2} / (3\pi^{1/2})^{1/2})$,
- h est maximal pour ℓ tel que $a(\ell) = 0.360$.

Illustration



Efficacité d'ordre un multipliée par $d^{2/3}$ en fonction du taux d'acceptation, pour MALA avec cible Laplace.

Conclusion



Résultats pour l'échelonnement de MALA avec cible Laplace :

- Nouvelle valeur de l'échelonnement $\alpha = 2/3$.
- Convergence en loi de \mathbf{Y}^d vers une diffusion.
- Nouvelle optimisation $a(\ell) = 0.360$.

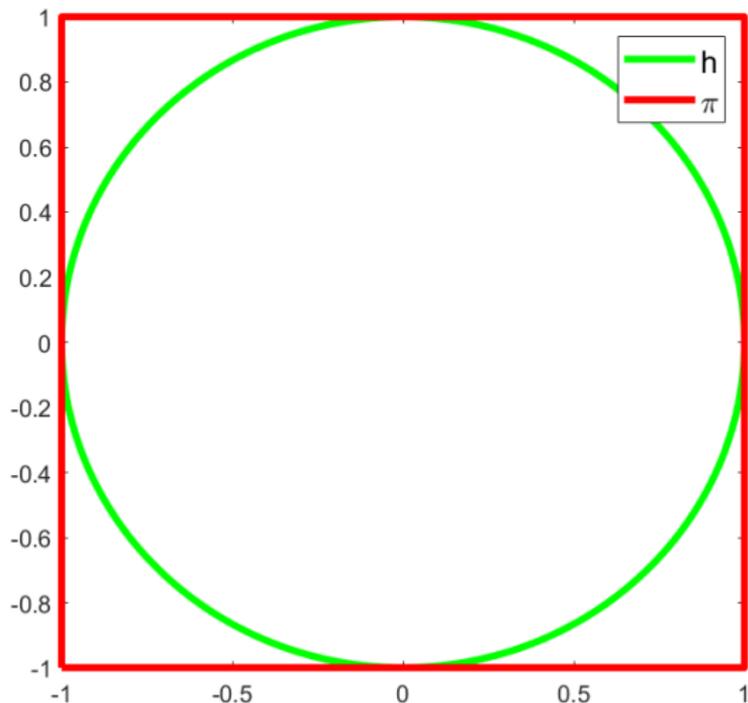
Perspective : généraliser à des densités continues mais pas C^1 .

Merci de votre attention !

References

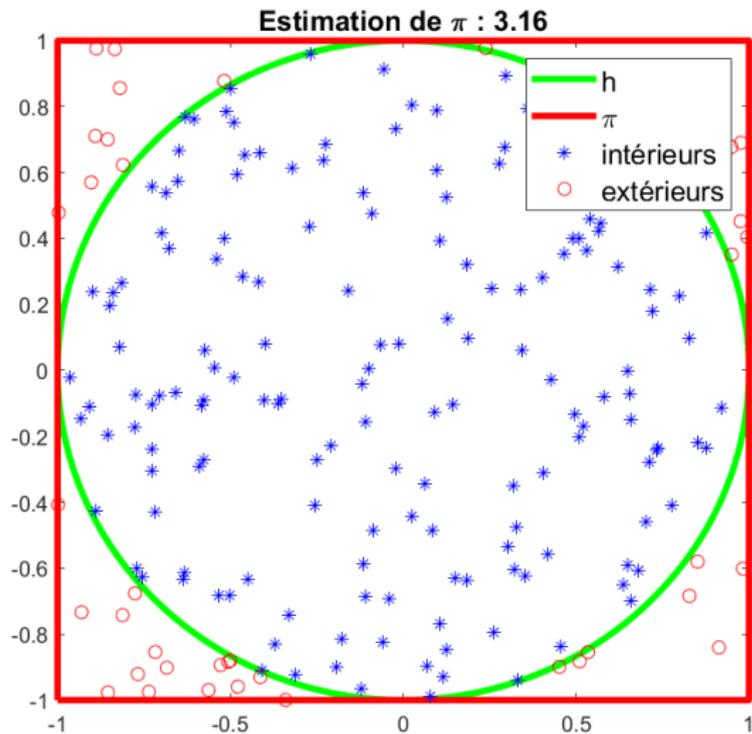
- A. Durmus, S. Le Corff, E. Moulines, and G.O. Roberts. Optimal scaling of the random walk Metropolis algorithm under L_p mean differentiability. *Journal of Applied Probability*, 54(4):1233–1260, 2017. doi: 10.1017/jpr.2017.61.
- W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970. ISSN 00063444. URL <http://www.jstor.org/stable/2334940>.
- N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, and A.H. Teller. Equations of state calculations by fast computing machine. *J. Chem. Phys.*, 21:1087–1091, 1953.
- Gareth O. Roberts and Richard L. Tweedie. Exponential convergence of langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 12 1996. URL <https://projecteuclid.org:443/euclid.bj/1178291835>.
- G.O. Roberts and J.S. Rosenthal. Optimal scaling of discrete approximations to Langevin diffusions. *J. R. Statist. Soc. B*, 60:255–268, 1997.
- G.O. Roberts, A. Gelman, and W.R. Gilks. Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability*, 7(1):110–120, 1997.
- Luke Tierney. Markov chains for exploring posterior distributions. *Ann. Statist.*, 22(4):1701–1728, 12 1994. doi: 10.1214/aos/1176325750. URL <https://doi.org/10.1214/aos/1176325750>.

Méthodes de Monte Carlo : exemple



- On veut estimer π .
- Alors
 $\pi = U([-1, 1]^2)$ et
 $h = \mathbb{1}_{B(0,1)}$.
- $\int_{\mathbb{R}^2} h(x^2)\pi(x^2)dx^2$
 $= \pi/4$.

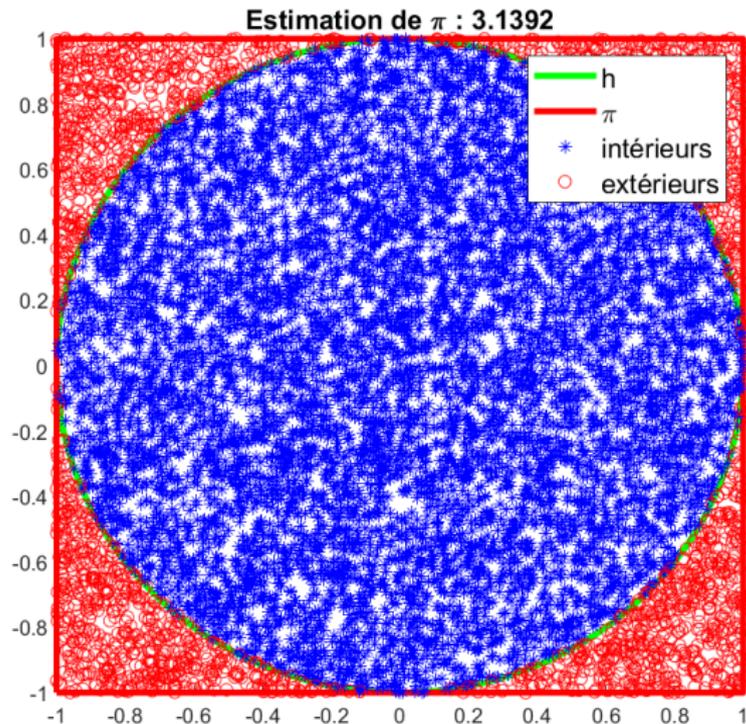
Méthodes de Monte Carlo : exemple



$n = 200$

- On veut estimer π .
- Alors
 $\pi = \mathbb{U}([-1, 1]^2)$ et
 $h = \mathbb{1}_{B(0,1)}$.
- $\int_{\mathbb{R}^2} h(x^2)\pi(x^2)dx^2 = \pi/4$.

Méthodes de Monte Carlo : exemple



$n = 10000$

- On veut estimer π .
- Alors
$$\pi = \mathbb{U}([-1, 1]^2) \text{ et } h = \mathbb{1}_{B(0,1)}.$$
- $$\int_{\mathbb{R}^2} h(x^2)\pi(x^2)dx^2 = \pi/4.$$