
# Filtrage et Modélisation Stochastique

Loïc Maurin - *loic.maurin@meteo.fr*

# 4 - Markov Chain Monte Carlo

# Aux cours précédents

Processus stochastiques et Filtre de Kalman

# Au cours précédent

## Processus stochastiques

**Processus stochastique** : famille de variables aléatoires $(X_t)_{t\in\mathcal{T}}$.

**Temps discret** : $t \in \mathbb{N}$, le processus se modélise avec des **probabilités de transition**.

**Temps continu** : $t \in \mathbb{R}$, le processus se modélise avec des **densités de probabilités** ou **taux de transitions** notés $W(n, m)$ pour la transition entre de l'état $n$ vers l'état $m$.

**Equation Maîtresse** : équation différentielle qui représente l'évolution temporelle de la densité de probabilité. On a vu qu'elle s'établissait de manière analogue à un bilan physique.

**Relations sur la moyenne et la variance** : à partir de l'équation maîtresse, on peut déduire l'évolution temporelle de la moyenne et de la variance du processus. Pour rappel,

$$\frac{d\langle n\rangle}{dt} = \langle W^+(n) - W^-(n)\rangle\,,\ \frac{dV}{dt} = 2\,\langle (n-\langle n\rangle)(W^+(n) - W^-(n))\rangle + \langle W^+(n) + W^-(n)\rangle$$

# Au cours précédent

## Filtre de Kalman

**Construction de l'Estimateur** : $P(x_n|z_n) \sim \mathcal{N}(\hat{x}_n, P_n)$, par récurrence.

Estimateur **assymptotiquement sans biais** sur lequel **la covariance de l'erreur est minimisée**, grâce à la relation entre le gain du filtre $K_n$ et la matrice de covariance $P_n$.

**Hypothèses de modélisation** : on suppose un **bruit d'état** $\phi \sim \mathcal{N}(0, \Phi)$ sur la dynamique du système et un **bruit de mesure** $\psi \sim \mathcal{N}(0, \Psi)$ sur l'observation du système.

**Influence du gain du filtre K** : si $K \to 0$, l'estimateur ne repose que sur la dynamique, et si $K \to \infty$, l'estimateur repose uniquement sur les mesures.

**Linéarité** : les propriétés du Filtre de Kalman sont établies pour un système linéaire perturbé par des bruits blancs gaussiens. Dans le cas d'un système non-linéaire, on introduit le **Tangeant-Linéaire** et l'**Adjoint** du système pour construire un **Filtre de Kalman Etendu**.

*Remarque le filtre Kalman étendu ne minimise pas formellement l'erreur d'estimation.*

# Au cours précédent

Vers l'estimation de Chaînes de Markov par méthodes de Monte-Carlo (MCMC)

**Filtre de Kalman** : la fermeture du système d'équations du filtre s'effectue grâce aux propriétés gaussiennes des bruits d'état et de mesure. On peut ainsi se limiter à construire **l'évolution de la moyenne et de la variance** de l'estimateur.

$\rightarrow$ Comment construire un estimateur quand on ne connaît plus la structure du bruit ?

# Introduction

## Chaînes de Markov et Méthodes de Monte-Carlo

$\rightarrow$ Comment construire un estimateur quand on ne connaît plus la structure du bruit ?

**Méthodes de Monte-Carlo** : estimation d'une densité de probabilité par tirages.

*Remarque : Une fois que l'on a accès à une estimation de la densité de probabilité, nous pouvons déduire les moments de la distribution, notamment la moyenne et la variance.*

**Chaînes de Markov** : processus stochastique dont la transition entre 2 états $x_{n-1} \rightarrow x_n$ ne dépend que de l'état précédent $x_{n-1}$.

*Remarque : Le filtre Kalman entre dans le cadre des chaînes de Markov.*

# Remarque introductive

## Markov Chain Monte Carlo

Ce cours est indépendant de celui sur le Filtrage Kalman.

$\rightarrow$ Le Filtre de Kalman et les Chaînes de Markov sont 2 représentants de Processus Stochastiques

$\rightarrow$ Les 2 sont utilisés en **Assimilation de données** pour établir le filtre de Kalman d'ensemble (EnKF), avec :

1. **Une composante de filtrage** :

   *Comment recaler un modèle par rapport aux observations / mesures* ?

2. **Une composante d'échantillonage** :

   *Comment estimer la dispersion du modèle* ?

# Motivations

## Estimation bayésienne et exploration d'une distribution

En estimation bayésienne, on cherche à construire un estimateur (bayésien), c'est à dire une distribution conditionnée aux données d'apprentissage :

$$P(\theta|data) = \frac{P(\theta) \cdot P(data|\theta)}{P(data)}$$

$\rightarrow$ Le problème principal : estimer **la loi marginale** $P(data)$

# Motivations

## Estimation bayésienne et exploration d'une distribution

$\rightarrow$ Le problème principal : estimer **la loi marginale** $P(data)$

- Cas discret :

$$P(data) = \sum_{\theta} P(data|\theta) \times P(\theta)$$

- Cas continu :

$$P(data) = \int_{\theta} P(data|\theta) \times P(\theta) d\theta$$

Dans le cas continu, calculer $P(data)$ peut devenir intractable.

# Markov Chain Monte-Carlo

MCMC

$\rightarrow$ Pourquoi utiliser les méthodes de Monte-Carlo sur des chaînes de Markov ?

- Apprendre une distribution, un processus stochastique par échantillonage aléatoire.
- Très utile sur des systèmes "en boîte noire" pour lesquels il existe des réalisations (ou des simulations).
- Efficace sur des problèmes de grandes dimensions *(dont la combinatoire est trop grande pour des méthodes classiques)*.

# Markov Chain Monte-Carlo

## Historique et applications

- Aiguilles de Buffon (1733) : Estimation de $\pi$ par le compte de Buffon.
- Projet Manhattan (1940) : utilisation des premiers ordinateurs pour la simulation de processus physiques, lors de la construction de la bombe atomique.

$\rightarrow$ Les Méthodes de Monte-Carlo se développent avec l'augmentation des puissances de calcul, avec, en particulier, quelques applications remarquables :

- PageRank : estimation du poids d'une page web par échantillonage sur ses liens sortants,
- AlphaGo : (Monte-Carlo Tree Search), estimation du meilleur coup par échantillonage des différentes trajectoires possibles.

# Méthodes de Monte-Carlo

# Méthodes de Monte-Carlo

## Construction de l'estimateur

**Estimateur de Monte-Carlo** Pour une suite de variables aléatoires indépendantes, identiquement distribuées, $\{q_1, \ldots, q_N\}$ (N arbitrairement grand), et pour une fonction $f$ à valeurs réelles et mesurables, les estimateur de Monte-Carlo sont défini par les **moyennes d'ensemble** de $f$ :

$$\hat{f}_N^{MC} = \frac{1}{N} \sum_{n=1}^{N} f(q_n)$$

La moyenne de l'ensemble converge alors vers l'espérance de $f$ :

$$\lim_{N \to \infty} \hat{f}_N^{MC} = \mathbb{E}_\pi[f]$$

Formellement, la variable aléatoire définie par la moyenne d'ensemble $\hat{f}_N^{MC}$ tend vers une distribution de Dirac autour de l'espérance de la fonction : $\lim_{N \to \infty} \hat{f}_N^{MC} = \delta_{\mathbb{E}_\pi[f]}$.

# Méthodes de Monte-Carlo

## Erreur de l'estimateur

**Propriété** : L'estimateur défini est assymptotiquement sans biais.

$\rightarrow$ En pratique, on cherche à estimer l'exactitude de l'estimateur pour un petit échantillon *(afin d'utiliser ces méthodes sur des ressources de calcul limitées)*.

**Propriété** : L'estimateur de Monte-Carlo, pour une fonction réelle et de carré intégrable (pour laquelle $\mathbb{E}_\pi[f], \mathbb{E}_\pi[f]$ existent), satisfait le Théorème Central Limite. C'est-à-dire, la suite des estimateurs de Monte-Carlo standardisés converge vers une loi normale centrée réduite.

$$\lim_{N\to\infty} \frac{\hat{f}_N^{MC} - \mathbb{E}_\pi[f]}{SE_N} \sim \mathcal{N}(0,1)$$

où $SE_N$ est l'erreur quadratique de l'estiamteur de Monte-Carlo : $SE_N = \sqrt{\frac{Var_\pi[f]}{N}}$

# Méthodes de Monte-Carlo

## Résumé

Autrement dit, à la limite,

$$\hat{f}_N^{MC} = \mathcal{N}(\mathbb{E}_\pi[f], SE_N[f])$$

En pratique, la quantification de l'erreur par le TCL suppose de connaître la variance $Var_\pi[f]$. Si $f^2$ est de carré intégrable ($\mathbb{E}[f^4]$ existe), on peut approximer la variance par un autre estimateur de Monte-Carlo.

**Avantages** : $SE_N \propto \frac{1}{\sqrt{N}}$, l'erreur quadratique moyenne décroît avec l'inverse de la racine carrée de la taille de l'échantillon. En pratique, cela permet de dimensionnner au préalable le nombre d'échantillons nécessaires pour atteindre une erreur donnée.

**Limitations** : la quantification de l'erreur est probabiliste. Il reste toujours une (mal)-chance que l'estimateur s'échoue dans la queue de la distribution (ex : $\hat{f}_N^{MC} > \mathbb{E}_\pi[f] + 3SE_N[f]$).

# Loi forte des Grands Nombres

## (Rappel, si nécessaire) Théorèmes fondamentaux

**Théorème** Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles **indépendantes** et **identiquement distribuées (i.i.d.)** et définies sur le même espace de probabilité $(\Omega, \mathcal{F}, \mathbf{P})$.

En posant $\mu = \mathbb{E}[X_i]$ et $\sigma^2 = \mathbb{V}(X_i) < \infty$.

La variable aléatoire $S_n = \sum_{i=1}^{n} X_i$ vérifie

$$\lim_{n\to\infty} \frac{S_n}{n} = \mu, p.s.$$

**p.s.** : presque sûr = "à une infinité dénombrable de points près"

**Application** Méthode de Monte-Carlo : calcul d'une intégrale par des tirages de variables aléatoires.

Pour $U$ variable aléatoire uniforme sur $[0, 1]$, $\lim\limits_{n\to\infty} \sum f(U_i) = \mathbb{E}[f(U)] = \int_0^1 f(x)dx$

# Théorème Central Limite

(Rappel, si nécessaire) Théorèmes fondamentaux

**Théorème** Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles **i.i.d.**

On note $\mu = \mathbb{E}[X]$ et $\sigma^2 = \mathbb{V}(X)$. On suppose $0 < \sigma^2 < \infty$.

En posant $S_n = \sum_{i=1}^{n} X_i$,

$$Y_n = \frac{S_n - n\mu}{\sigma\sqrt{(n)}} \underset{n\to\infty}{\longrightarrow} Y \sim \mathcal{N}(0,1)$$

**Remarque** Il s'agit d'une convergence en loi : la suite des lois $Y_n$ tend vers $Y$ qui est aussi une loi de probabilité.

**Application** Le Thèorème Centrale Limite donne des indications sur les vitesses de convergence vers la loi assymptotique.

# Chaînes de Markov

# Chaînes de Markov

## Définition

$\rightarrow$ But : on cherche à explorer une distribution cible $\pi$.

**Définition** : Sur un espace d'états $Q$, une chaîne de Markov est définie comme une suite de transitions entre états $q_1, ..., q_N$ dont la transition entre 2 états $q_n \rightarrow q_{n+1}$ ne dépend que de l'état $q_n$.

$$P(q_{n+1}|q_1, ..., q_n) = P(q_{n+1}|q_n)$$

**Application** : Une chaîne de Markov permet d'échantillonner des chemins discrets sur un espace ambient (à explorer). La mise au point d'une chaîne de Markov sur cet espace permet d'identifier une distribution cible $\pi$.

# Chaînes de Markov

Définition

**Distribution de transitions** : Soit un espace ambient $Q$ équipé d'une tribu (ou $\sigma$-algèbre) $\mathcal{Q}$. On peut spécifier les **transitions de Markov** comme une densité de probabilité conditionnelle :

$$\begin{aligned} T\colon \quad & Q \times Q \to \mathbb{R}^+ \\ & (q, q') \mapsto T(q'|q) \end{aligned}$$

pour une transition de $q$ vers $q'$.

Etant donné un point $q_0$, un tirage aléatoire de $T(\cdot|q_0)$ forme un **saut** ou une **transition**.

$$\bar{q}_1 \sim T(q_1|q_0)$$

# Chaînes de Markov

## Définition

Etant donné un point $q_0$, un tirage aléatoire de $T(\cdot|q_0)$ forme un **saut** ou une **transition**.

$$\bar{q}_1 \sim T(q_1|q_0)$$

En itérant ces tirages aléatoires, on réalise (ou simule) une trajectoire $\{\bar{q}_1, \dots, \bar{q}_n\}$ où

$$\begin{gathered} \bar{q}_1 \sim T(q_1|q_0) \\ \dots \\ \bar{q}_N \sim T(q_N|q_{N-1}) \end{gathered}$$

$\rightarrow$ On génère des séquences de points corrélés.

*(Au contraire des tirages i.i.d. effectués pour les Méthodes de Monte-Carlo).*

# Chaînes de Markov

## Exemple : trajectoire sur un espace à 2 dimensions

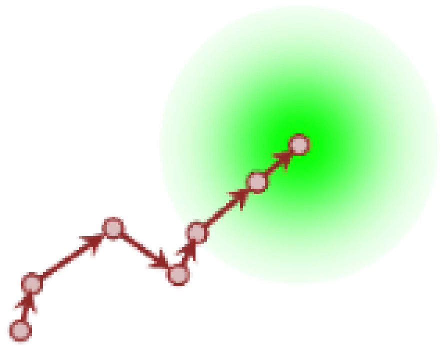
On se dote d'un espace ambient à 2 dimensions $Q = \mathbb{R}^2$, avec 2 fonctions de coordonnées :

$$\omega_1 : \begin{array}{l} Q \to \mathbb{R} \\ q \mapsto q^1 \end{array} \quad \Bigg| \quad \omega_2 : \begin{array}{l} Q \to \mathbb{R} \\ q \mapsto q^2 \end{array}$$

On définit une densité de probabilité de transition de Markov :

$$T(q_1|q_0) = \mathcal{N}(q_1^1|q_0^1, \sigma)\mathcal{N}(q_1^2|q_0^2, \sigma)$$

*A gauche : Réalisation d'une trajectoire d'une chaîne de Markov, Markov Chain Monte Carlo in Practice, M. Betancourt*



# Chaînes de Markov

## Distribution stationnaire

**Résultat empirique** : la réalisation d'une chaîne de Markov converge vers une distribution invariante par transitions de Markov, **la distribution stationnaire**, pour une distribution de transitions donnée $T('q|q')$.

$$\pi = \int dq'\pi(q')T(q|q')$$

**Application**

En reprenant le **problème de marginalisation** $P(data) =?$, si nous arrivons à construire une chaîne de Markov dont la distribtion stationnaire est celle des données ($\pi = P(data)$), alors nous pouvons construire un estimateur de manière analogue à celui de Monte-Carlo.

# Chaînes de Markov

## Stationnarité - Démonstration

Construisons les densités de probabilités rencontrées sur le chemin formé par la chaîne de Markov :

1. $\bar{q}_0 \sim \rho$. On définit la distribution initiale $\rho = \delta_{q_0}$ (distribution de Dirac autour de $q_0$, le point initial est $q_0$ presque sûrement).

2. L'espérance de la position du point 1 est donnée par :

$$(T\rho)(q_1) = \int dq_0 T(q_1|q_0)\rho(q_0)$$

3. En itérant au point 2, puis au point n :

$$(T^2\rho)(q_2) = (T \cdot T\rho)(q_2) = \int dq_1 dq_2 T(q_2|q_1)T(q_1|q_0)\rho(q_0)$$

$$(T^N\rho)(q_N) = (T \cdot T^{N-1}\rho)(q_{N-1}) = (T \cdot ... \cdot T\rho)(q_N)$$

# Chaînes de Markov

## Stationnarité - Démonstration

En consruisant la densité de probabilité au point $N$ :

$$(T^N\rho)(q_N) = (T \cdot T^{N-1}\rho)(q_{N-1}) = (T \cdot \ldots \cdot T\rho)(q_N)$$

En observant la convergence, **si la limite existe**,

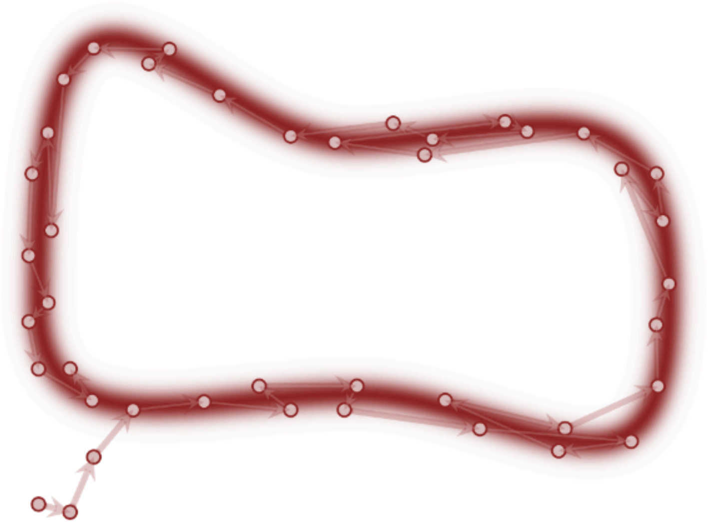
$$\lim_{N\to\infty} T^N\rho = \pi$$

Alors c'est un point fixe :

$$T\pi = \pi$$

$\rightarrow$ *Ce qu'on vient d'énoncer ne présage pas de l'existance de la limite. Simplement, si elle exite alors c'est une distribution stationnaire. En pratique, on construit une chaîne de Markov et ses transitions pour qu'elle tende une distribution limite stationnaire.*

# Chaînes de Markov



# Chaîne de Markov

## (Optionnel) Vitesses de convergence

**Convergence** : Etant donné une distance $\|\cdot\|$, la chaîne de Markov converge si $\forall\epsilon \in \mathbb{R}^+$, $\exists N(\rho) \in \mathbb{N}$ tel que

$$\|T^N\rho - \pi\| \leq \epsilon$$

**Distance en variations totales** : $\|\rho - \pi\|_{TV} = sup_{B\in\mathcal{Q}}|\rho[B] - \pi[B]|$

**Vitesses de convergence**

- Ergodicité polynômiale : $\|\rho - \pi\|_{TV} \leq C(\rho)(N+1)^{-\beta}$
- Ergodicité géométrique : $\|\rho - \pi\|_{TV} \leq C(\rho)r^N$
- Ergodicité uniforme : $\|\rho - \pi\|_{TV} \leq Cr^N$

*Remarque : L'ergodicité uniforme permet la convergence rapuide (en un nombre limité d'opérations), mais est typiqueemnt réservée aux espaces bornés.*

# Chaînes de Markov

## (Optionnel) Spectre de la matrice de transition et Convergence

# Markov Chain Monte-Carlo

# Markov Chain Monte-Carlo

## Construction de l'estimateur

**Estimateur MCMC** Etant donné une suite de points $\{q_1, ..., q_N\}$ qui forment la réalisation d'une chaîne de Markov, l'estimateur **Markov Chain Monte-Carlo** est défini par

$$\hat{f}_N^{MCMC} = \frac{1}{N+1} \sum_{n=0}^{N} f(q_n)$$

Contrairement aux estimateurs de Monte-Carlo (MC), le comportement assymptotique des estimateurs MCMC n'est pas défini

$$\lim_{N \to \infty} \pi_{f_N^{MCMC}} = \delta_{\mathbb{E}_\pi[f]} \text{ ???}$$

*MCMC : Markov Chain Monte-Carlo*

# Markov Chain Monte-Carlo

## Conditions de convergence - Nombre d'itérations infinies

La convergence des estimateurs MCMC est garantie à condition que **la chaîne soit récurrente.** Dans ce cas, la limite existe seulement pour **un nombre fini d'initialisations**.

$$\lim_{N\to\infty} \pi_{f_N^{MCMC}} = \delta_{\mathbb{E}_\pi[f]}$$

Ce résultat peut être généralisé à toute distribution initale de points par la **condition de Harris**.

**Chaîne de Markov récurrente** : La chaîne est irréductible selon des transitions de Markov sont **apériodiques** et **irréductible**.

**Chaîne de Harris** : chaîne de Markov dont la chaîne retourne **un nombre non-borné de fois** dans une partie quelconque de l'espace d'états.

*$\longrightarrow$ En pratique, construire ou utiliser un estimateur **MCMC** nécessite de vérifier précautionneusement les hypothèses sur les transitions de la chaîne de Markov.*

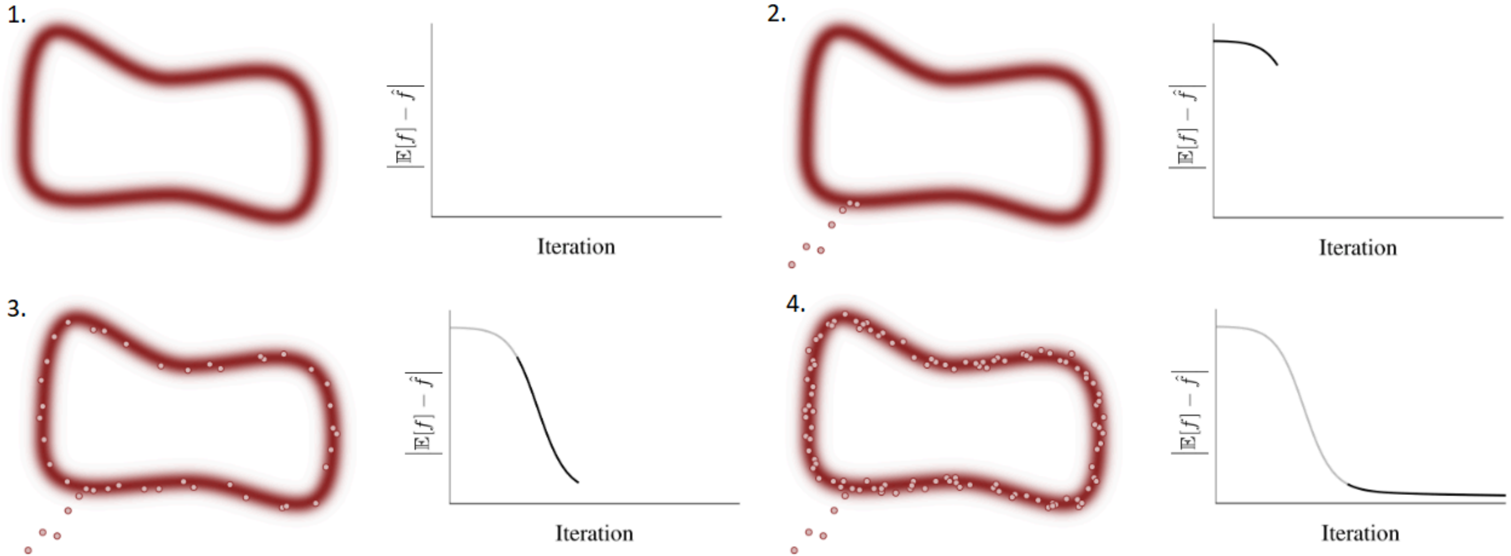
# Markov Chain Monte-Carlo

## Conditions de convergence - Nombre d'itérations fini

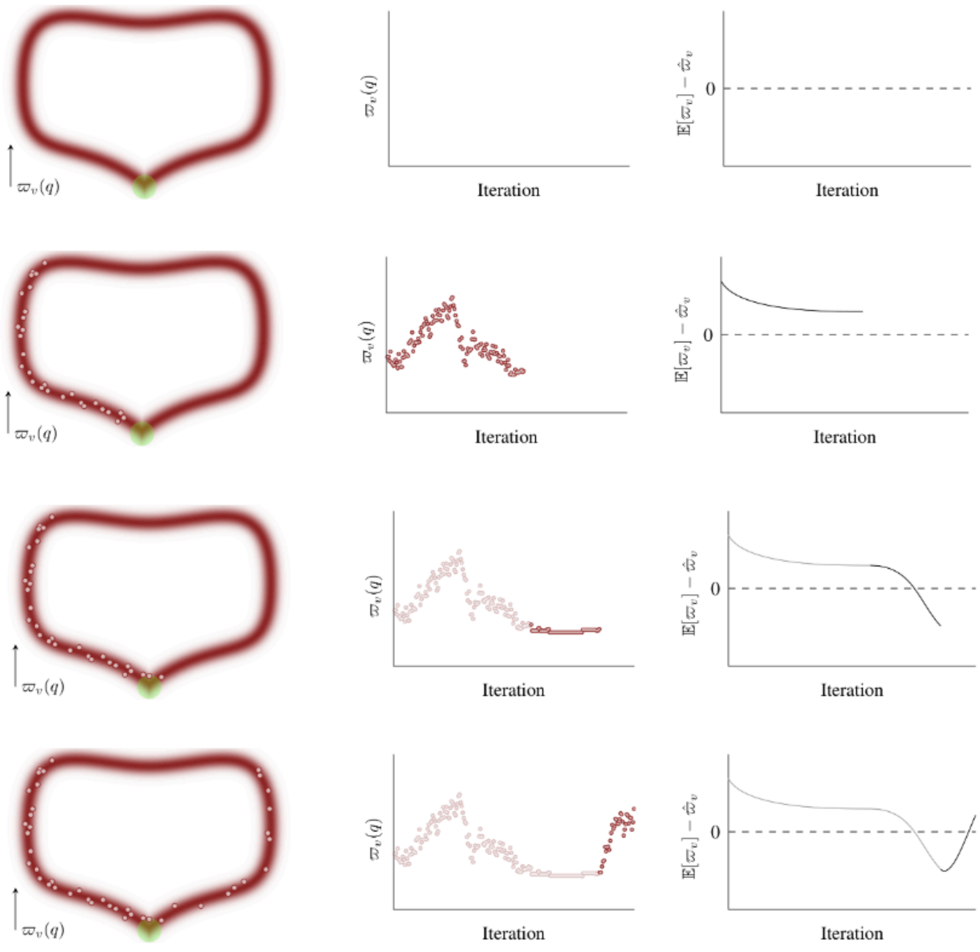
Nous venons d'étudier le comportement assymptotique d'un estimateur **MCMC**, c'est-à-dire en nombre d'itérations infini.

$\rightarrow$ En pratique, qu'en est il de la convergence en nombre d'itérations finies ?

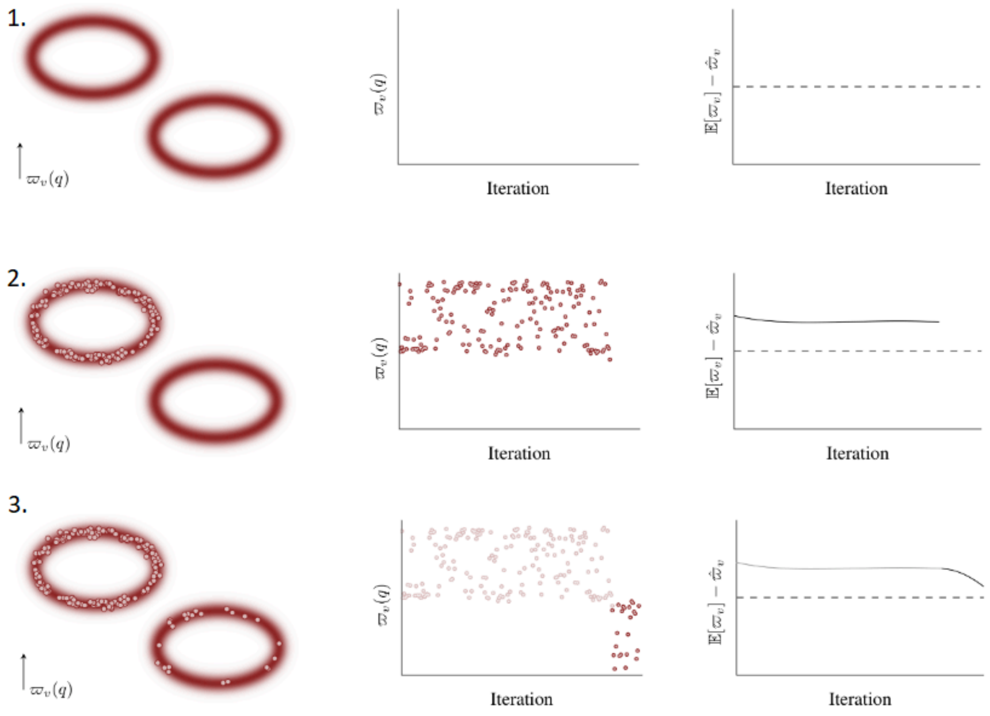
# Markov Chain Monte-Carlo




Etapes de convergence d'une chaîne de Markov, Markov Chain Monte Carlo in Practice, M. Betancourt



$\varpi_v(q)$
$\varpi_v(q)$
Iteration
$\mathbb{E}[\varpi_v] - \hat{\varpi}_v$
0
Iteration
$\varpi_v(q)$
$\varpi_v(q)$
Iteration
$\mathbb{E}[\varpi_v] - \hat{\varpi}_v$
0
Iteration
$\varpi_v(q)$
$\varpi_v(q)$
Iteration
$\mathbb{E}[\varpi_v] - \hat{\varpi}_v$
0
Iteration
$\varpi_v(q)$
$\varpi_v(q)$
Iteration
$\mathbb{E}[\varpi_v] - \hat{\varpi}_v$
0
Iteration




Convergence de l'estimateur MCMC, cas métastable, Markov Chain Monte Carlo in Practice, M. Betancourt

# Markov Chain Monte Carlo

## Convergence

$\rightarrow$ En pratique, la convergence dépend beaucoup de la distribution à explorer.

Ces méthodes, bien qu'utiles, sont moins robustes que les méthodes de Monte-Carlo. Elles dépendent grandement du jeu de données à explorer et nécessitent une mise au point minutieuse.

# Markov Chain Monte-Carlo

(Optionnel) Vitesse de convergence théorique

# Markov Chain Monte-Carlo

## (Optionnel) MCMC et Théorème Central Limite

Le Théorème Central Limite s'applique pour des conditions particulières d'estimateurs MCMC.

Si on considère un fonction de carré intégrable $f : Q \mapsto \mathbb{R}$. On suppose que la distribution de porbabilité des transitions satisfait le Théorème Central Limite. Dans ce cas, à partir d'un certain rang, suffisament long, la chaîne de Markov peut être apporximée à une Gaussienne :

$$\hat{f}_N^{MCMC} \sim \mathcal{N}(\mathbb{E}[f], MCMC - SE[f])$$

où MCMC-SE (*Markov Chain Monte Carlo Standard Error*) est définie par :

$$MCMC - SE[f] = \sqrt{\frac{Var[f]}{\lambda[f] \cdot N}}$$

*Remarque : selon le terme $\lambda[f] \cdot N$, la décroissance de l'erreur pour un estimateur MCMC peut être plus rapide que celle d'un estimateur de Monte-Carlo.*

# Markov Chain Monte-Carlo

## Algorithme de Metropolis-Hastings

**Implémentation** : Comment construire les distributions de transitions pour obtenir une Chaîne de Markov qui converge vers une distribution stationnaire ?

**L'Algorithme de Métropolis-Hastings** propose une approche générique par essais / erreurs.

1. On définit une distribution à priori pour définir les probabilités de transition :

$$K: \quad \begin{aligned} Q \times Q &\to \mathbb{R}^+ \\ (q, q') &\mapsto K(q'|q) \end{aligned}$$

# Markov Chain Monte-Carlo

## Algorithme de Metropolis-Hastings

**L'Algorithme de Métropolis-Hastings** propose une approche générique par essais / erreurs.

2. Pour une transition entre $q$ et $q'$, on définit la probabilité d'acceptation *(acceptance probability)* de Metropolis-Hastings

$$a(q', q) = min(1, \frac{K(q|q')\pi(q')}{K(q'|q)\pi(q)})$$

avec $\frac{\pi(q')}{\pi(q)}$ le *ratio de Metropolis*, et $\frac{K(q|q')}{K(q'|q)}$ la *correction d'Hastings*

3. La *transition de Métropolis* est définie comme la probabilité de sauter vers la proposition $q'$ avec une probabilité $a(q', q)$ et de rester au point initial avec la probabilité $1 - a(q', q)$

# Markov Chain Monte-Carlo

## Algorithme de Metropolis-Hastings

La distribution des transitions de Markov peut être définie comme :

$$T(q'|q) = a(q', q) \cdot Q(q'|q) + (1- \in dq' Q(q'|q) a(q|q')) \cdot \delta(q - q')$$

**Random Walk Metropolis**

Dans ce cas, $Q(q'|q, \Sigma) = \mathcal{N}(q'|q, \Sigma)$ On perturbe le point initial par une gaussienne.

# A retenir

Markov Chain Monte Carlo

- Méthodes de Monte-Carlo
- Chaînes de Markov
- Markov Chain Monte-Carlo

# La semaine prochaine

Spoiler alert !

**TP** : Markov Chain Monte-Carlo

$\rightarrow$ Apportez vos PC, TP sur Google Colab

# Liens du cours

Ressources utiles sur les Monte Carlo Markov Chains

Markov Chain Monte Carlo in Practice, M. Betancourt

Compound extremes in a changing climate – a Markov chain approach

Algorithme de Métropolis-Hastings

Processus Stochastiques, Notes de cours, P. Carmona

CS168: The Modern Algorithmic Toolbox Lecture #14: Markov Chain Monte Carlo

END