

Filtrage stochastique

Loïc MAURIN

November 19, 2025

Contents

0.1	Introduction	1
0.1.1	Objectifs du cours	1
0.1.2	Plan du cours	1
0.1.3	Démystifions l'aspect stochastique !	2
0.1.4	Processus stochastiques - Applications	2
0.1.5	Rappels de Probabilités et Statistiques	3
0.2	Processus stochastiques	7
0.2.1	Définition	7
0.2.2	Processus stochastique - Théorie	8
0.2.3	Exercice	12
0.2.4	A retenir	12
0.2.5	Processus stochastiques - En bref	12
0.2.6	Précisions sur le Mouvement Brownien	13
0.2.7	Sources	13
0.3	Filtrage Kalman	14
0.3.1	Filtre Kalman	14
0.3.2	Filtre de Kalman - Implémentation	17
0.3.3	Dynamique de l'erreur	17
0.3.4	Synthèse	18
0.3.5	Exemple - Estimation d'une constante aléatoire	19
0.3.6	Dynamique Non-Linéaire	21
0.3.7	A retenir	22
0.3.8	Sources	23
0.4	Chaînes de Markov	24
0.4.1	Définition	24
0.4.2	Conditionnements successifs et calcul	24
0.4.3	Etats transients et récurrents	24
0.4.4	Mesures invariantes (ou stationnaires)	25
0.4.5	Théorème ergodique	25
0.4.6	Convergence vers l'équilibre	25
0.4.7	Les propriétés d'une chaîne de Markov	25
0.4.8	Application au PageRank	26
0.5	Markov Chains Monte Carlo	28
0.5.1	Markov Chain Monte Carlo	28
0.5.2	Markov Chain Monte-Carlo	28
0.5.3	Historique et applications	29
0.5.4	Méthodes de Monte-Carlo	29
0.5.5	Chaînes de Markov	30
0.5.6	Markov Chain Monte-Carlo	33
0.5.7	Liens du cours	36
0.6	Equations différentielles stochastiques	37
0.6.1	Définition	37
0.6.2	De l'équation différentielle ordinaire (EDO) à l'équation différentielle stochastique (EDS)	37

0.6.3	Processus de premier ordre	37
0.6.4	Formule d'Itô	37
0.6.5	Intégration numérique - Méthode d'Euler-Maruyama	38
0.6.6	Convergence de la Méthode Euler-Maruyama	38
0.6.7	Applications	39
0.6.8	Références	40
0.7	Filtre Kalman d'Ensemble	41
0.7.1	Assimilation de données	41
0.7.2	Ensemble Kalman Filter (EnKF)	41
0.7.3	Formulation de l'EnKF pour l'analyse	43
0.8	Assimilation de données	45
0.8.1	Description formelle	45
0.9	Modèles de diffusion pour l'assimilation de données	46
0.9.1	Diffusion Models for Data Assimilation	46

0.1 Introduction

Cours de Filtrage Stochastique donné aux élèves du Master HPC-BigData mené conjointement entre ENM et l'ENSEEIHT.

Ce cours présente les fondements théoriques et pratique pour l'assimilation de données en Météorologie.

Le cours présente les processus stochastiques, à savoir, modéliser un système à partir d'une dynamique déterministe et d'une erreur stochastique (issue d'un mouvement brownien, par exemple).

En pratique, le cours introduit le Filtre Kalman, outil essentiel pour corriger un système ou un modèle dynamique à partir d'observations ou de mesures. Puis il présente les méthodes de simulations stochastiques : Méthodes de Monte-Carlo et Chaîne de Markov. Ces méthodes sont utiles dans un troisième temps pour simuler la dispersion (ou l'erreur) d'un modèle chaotique, en perturbant d'une part le modèle et d'autre part ses observations.

Cet ensemble constitue une base pour les méthodes d'assimilations en sciences de l'atmosphère mais également de nombreuses applications en ingénierie, statistiques et sciences des données, ainsi que dans de nombreux domaines où une simulation stochastique est utile (finance, biologie, etc.).

Loïc Maurin - loic.maurin@meteo.fr

0.1.1 Objectifs du cours

Construire une boîte à outils pour modéliser l'incertitude et les phénomènes aléatoires

Applications en ingénierie et science des données :

- Assimiler des données bruitées (*ex : capteurs de vitesse, température, etc.*)
- Modéliser des phénomènes chaotiques (*ex : dynamique atmosphérique*)
- Contrôler et optimiser des systèmes dont la dynamique est inconnue (boîte noire)

Applications en Météo et sciences du climat :

- Estimer les variances d'ébauche, à l'assimilation de données d'observations
 - Construction des ensemblistes
-

0.1.2 Plan du cours

- Introduction
 - Filtre de Kalman
 - TP : Application au filtrage d'un oscillateur harmonique
 - Chaînes et processus de Markov
 - TP : Markov Chains Monte-Carlo avec Page-Rank
 - (Un peu de) théorie sur les processus stochastiques et Ensemble Kalman Filter
 - TP : Assimilation de mesures de température et de salinité *in situ* pour un modèle d'océan
 - Régression par Processus Gaussiens et Optimisation Bayésienne
 - TP : modèle de courant de surface par rapport aux données de bouées
 - TP : optimisation des hyper-paramètres d'un réseau de neurone
 - Devoir
-

0.1.3 Démystifions l'aspect stochastique !

- Dans ce cours, nous nous intéresserons à modéliser la moyenne et la (co)-variance des processus
 - *en sachant que les méthodes peuvent être étendues aux moments d'ordres supérieurs*
 - Ici, Stochastique = Gaussien (dans la mesure du possible)
 - *on modélisera les processus par rapport à la loi normale, en tirant partie du Théorème Central Limite,*
 - *en gardant à l'esprit d'autres lois peuvent supporter des modèles (ex: processus de poisson)*
-

0.1.4 Processus stochastiques - Applications

Notions abordées :

- Marche aléatoire et Mouvement brownien
- Régression par processus gaussiens
- Processus de Markov : Transition d'états sans mémoire

Modèles et techniques développés :

- Optimisation Bayésienne : échantillonage sur un processus gaussien
 - Echantillonage par méthodes Monte-Carlo : Markov Chain Monte-Carlo
 - Filtre Kalman : cas particulier de processus gaussien
-

Filtrage Kalman

Quelques exemples d'applications

Utiliser au mieux la dynamique d'un système pour recaler ses observations

- Localisation / Recalage GPS
 - Filtrage de données robotiques
 - Assimilation de données pour la Météo (Ensemble Kalman Filter - EnKF)
-

Markov Chain Monte-Carlo (MCMC)

Echantillonner et estimer les probabilités de transitions entre états d'un système

- Recherche de pages web (Page Rank)
 - Marche aléatoire sur le graphe des pages webs
- Estimation des récurrences et probabilités d'évènements extrêmes dans un modèle de climat :
 - Marche aléatoire sur des séries temporelles de pluies : comparaison de données historiques et projetées par le modèle de climat.

Application en climatologie

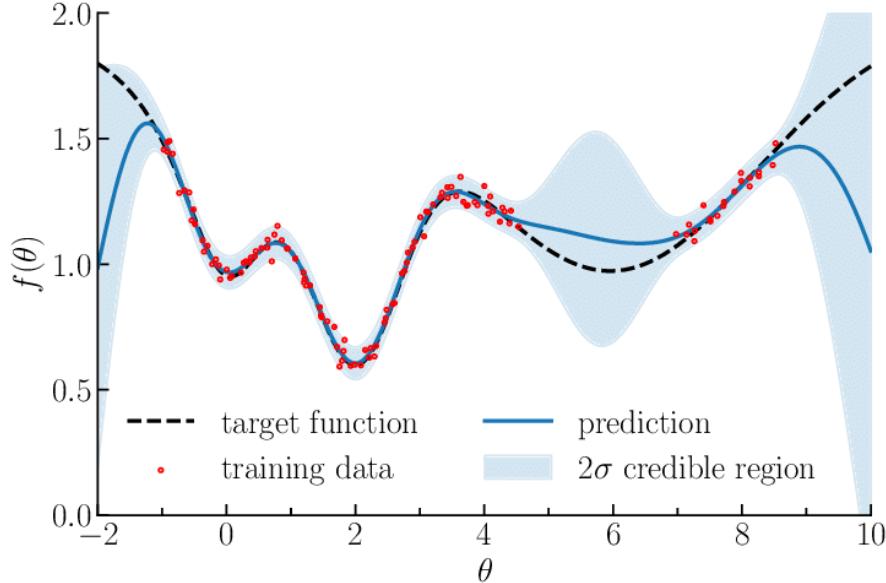
Les estimateurs MCMC sont utilisé en climatologie, pour mesurer les changements de régime de temps d'une projection climatique par rapport à une série de données historiques : Sedlmeier et al. [2016]

Régression par Processus Gaussiens

Modéliser les incertitudes lors de l'exploration d'un champ / d'un espace de paramètres

- Résolution de problèmes inverses :
 - Contrôle de bras robotiques,
 - Cartographie des sols : recherche pétrolière, géophysique
- Optimisation
 - Tuning des hyperparamètres d'un réseau de neurone
 - Réglage de procédés industriels par plans d'expériences

Ajustement d'un processus gaussien



0.1.5 Rappels de Probabilités et Statistiques

Pour être à l'aise sur les notions du cours

- Loi Normale
- Loi jointe, Loi marginale
- Théorème Centrale Limite et Loi des Grand Nombres
- Théorème de Bayes
- Estimateur du Maximum Vraisemblance

Lois usuelles

Loi uniforme : X suit une **loi uniforme** sur $[a, b] \in \mathbb{R}$ si elle admet la densité de probabilité $f : \mathbb{R} \mapsto \mathbb{R}$

$$\forall x \in \mathbb{R}, f(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in [a, b] \\ 0 & \text{sinon} \end{cases} \quad (1)$$

Loi exponentielle : X suit une **loi exponentielle** $\mathcal{E}(\lambda)$ ($\lambda > 0$) si elle admet la densité de probabilité $f : \mathbb{R} \mapsto \mathbb{R}$

$$\forall x \in \mathbb{R}, f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases} \quad (2)$$

On a alors $\mathbb{E}[X] = 1/\lambda$, $\mathbb{V}[X] = 1/\lambda^2$

Loi Normale : X suit une **loi normale** $\mathcal{N}(m, \sigma^2)$, $((m, \sigma) \in \mathbb{R} \times \mathbb{R}_+)$

$$\forall x \in \mathbb{R}, f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \quad (3)$$

On a alors $\mathbb{E}[X] = m, \mathbb{V}[X] = \sigma^2$

Loi Gamma : X suit une **loi gamma** $\gamma(p, \lambda)$, $(p > 0, \lambda > 0)$ si sa densité est

$$\forall x \in \mathbb{R}, f(x) = \begin{cases} \frac{\lambda^p}{\Gamma(p)} (\lambda x)^{p-1} e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases} \quad (4)$$

On a alors $\mathbb{E}[X] = p/\lambda, \mathbb{V}[X] = p/\lambda^2$

Exemple d'application : la loi gamma en microphysique

En microphysique (modélisation des nuages et interactions entre hydrométéores), la distribution des gouttes par rapport à leur diamètre est donnée par une loi exponentielle (Loi de Marshall-Palmer).

Loi de Marshall-Palmer

$$N(D)dD = N_0 e^{-\lambda D} dD \quad (5)$$

, où, $f(D) = e^{-\lambda D}$.

où la densité de probabilité est $f(D) = e^{-\lambda D}$.

On peut alors lier les moments de la distribution avec les caractéristiques physiques des gouttes :

- **Concentration de goutelettes** N_c et **moment d'ordre 0** : $N = \int_0^\infty f(D)dD$
- **Contenu en eau** q_c et **moment d'ordre 3** : $L = \frac{\pi\rho_w}{6} \int_0^\infty D^3 f(D)dD$ (*lié au volume des gouttes*)
- **Réflectivité radar** Σ_c et **moment d'ordre 6** : $R \propto \int_0^\infty D^6 f(D)dD$ (*lié à la diffusion de Rayleigh*)

En modélisation microphysique, “le jeu” est de choisir le nombre de moments à modéliser (**fermeture**), la discréttisation sur l'espace des diamètres (**échantillonage**), pour représenter les interactions possibles entre les gouttes (de pluie, de neige, etc.).

Remarque : Une densité de loi Gamma (plus générique que la loi exponentielle) est choisie sur certains schémas microphysiques.

Loi forte des Grands Nombres

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

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = \mu, p.s. \quad (6)$$

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 $[0, 1]$, $\lim_{n \rightarrow \infty} \sum f(U_i) = \mathbb{E}[f(U)] = \int_0^1 f(x)dx$

Théorème Centrale Limite

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

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

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

$$Y_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow[n \rightarrow \infty]{} Y \sim \mathcal{N}(0, 1) \quad (7)$$

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 Théorème Centrale Limite donne des indications sur les vitesses de convergence vers la loi asymptotique.

LGN vs TCL

Loi des Grand Nombres : convergence des moyennes empiriques vers la moyenne de la loi des v.a.

Théorème Centrale Limite : convergence asymptotique de la loi sur la moyenne empirique (donne des indications sur une loi limite).

Modes de convergences

- Convergence presque sûre (**p.s.**) :

$$P(\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}) = 1 \quad (8)$$

- Convergence dans L^P :

$$\lim_{n \rightarrow \infty} \mathbb{E}[|X_n - X|^p] = 0 \quad (9)$$

- Convergence en probabilité **P**:

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) = 0 \quad (10)$$

Remarque :

- Convergence $L^p \Rightarrow$ Convergence **P**,
 - Convergence **p.s.** \Rightarrow Convergence **P**.
-

Estimation bayésienne

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

$$P(\theta | data) = \frac{P(\theta) \cdot P(data | \theta)}{P(data)} \quad (11)$$

où :

- $P(\theta)$ est la **distribution à priori** des paramètres du modèle, ou **prior**
- $P(data | \theta)$ est la **vraisemblance du modèle** par rapport aux données, ou **likelihood** en anglais.
- $P(\theta | data)$ est la **distribution à posteriori** des paramètres du modèle, ou **posterior**
- $P(data)$ est la loi ou distribution marginale des données.

*Remarque : on utilise souvent le logarithme de ces grandeurs. Ainsi, on ne s'étonnera pas de minimiser la **Negative Log Likelihood** ou (NLL) en Deep Learning, ce qui revient exactement à maximiser la vraisemblance du modèle.*

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

En pratique : on cherche à éviter le calcul de la loi marginale des données.

- Cas discret :

$$P(data) = \sum_{\theta} P(data|\theta) \times P(\theta) \quad (12)$$

- Cas continu :

$$P(data) = \int_{\theta} P(data|\theta) \times P(\theta) d\theta \quad (13)$$

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

0.2 Processus stochastiques

0.2.1 Définition

On considère un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$, un espace mesurable (E, \mathcal{E}) et un ensemble T .

Définition On appelle processus stochastique, ou processus aléatoire, une famille $(X_t)_{t \in T}$ de variables aléatoires à valeurs dans E .

Autrement dit, pour tout $t \in T$, l'application $\omega \mapsto X_t(\omega)$ est une application mesurable de (Ω, \mathcal{F}) dans (E, \mathcal{E}) . On appelle E l'espace d'état du processus.

Remarque

- X_t est souvent l'expression d'une variable à un temps t ,
- T représente l'ensemble des dates possibles

Définition

- Lorsque $T = \mathbb{N}$ ou $T = \mathbb{Z}$, on parle de **processus à temps discret**
- Lorsque $T = \mathbb{R}$, on parle de **processus à temps continu**

Exemple

- Processus à temps discret : évolution du PIB de la France par année où :
 X_t représente le PIB, $t \in \mathbb{N}$ est l'année
- Processus à temps continu : évolution du cours d'une action où :
 X_t est la valeur de l'action, $t \in \mathbb{R}_+$ est le temps, considéré comme continu étant donnée la fréquence de rafraîchissement des valeurs.

Définition On appelle **filtration** une suite $(\mathcal{F}_\sqcup)_{t \in T}$ σ -algèbres vérifiant

$$s \leq t \Rightarrow \mathcal{F}_f \subset \mathcal{F}_\sqcup \subset \mathcal{F} \quad (14)$$

Remarque On utilise la notion de filtration pour représenter l'information disponible à date t .

Quand on observe un processus au cours du temps, on connaît les valeurs de X_s pour $s \leq t$, mais pas encore pour $s > t$. On sera donc souvent amenés à conditionner par les variables $(X_s)_{s \leq t}$.

Définition Le processus $(X_t)_{t \in T}$ est dit adapté à la filtration $(\mathcal{F}_\sqcup)_{t \in T}$, si pour tout $t \in T$, X_t est \mathcal{F}_t -mesurable.

Exemple : les Processus Auto-Régressifs

Processus Auto-Régressifs (AR) Dans cet exemple, $T = \mathbb{N}$. Soit $(\epsilon_t)_{t \in \mathbb{N}}$ une suite de variables aléatoires i.i.d de loi $\mathcal{N}(0, 1)$, $(\alpha, \beta) \in \mathbb{R}^2$, $X_0 = 0$ et

$$X_{t+1} = \alpha X_t + \beta + \epsilon_t \quad (15)$$

On définit $\mathcal{F}_t = \sigma(\epsilon_s, s \leq t)$. On peut vérifier que le processus $(X_t)_{t \in \mathbb{N}}$ est adapté à la filtration $(\mathcal{F}_\sqcup)_{t \in \mathbb{N}}$.

Des processus stochastiques particuliers

- Les **processus de Markov**, dont les transitions ne dépendent que de l'état précédent,

$$\mathcal{P}(X_{t+1} | \mathcal{F}_t) = \mathcal{P}(X_{t+1} | X_t) \quad (16)$$

- Les **processus gaussiens** $(X_t)_{t \in T}$, pour lesquels toutes les lois fini-dimensionnelles $\mathcal{L}(X_{t_1}, \dots, X_{t_n})$ sont gaussiennes. On définit les processus gaussiens par une moyenne au cours du temps $m(t) = \mathbb{E}[X_t]$ et un opérateur de covariance $K(s, t) = \text{Cov}(X_s, X_t)$

- Le **filtre Kalman**, un exemple particulier de processus gaussiens, pour lequel l'état d'un système dynamique est estimé, conditionné à des observations :

$$\hat{X}_k = \mathbb{E}[X_k] \quad (17)$$

$$P_k = \mathbb{E}[(X_k - \hat{X}_k)(X_k - \hat{X}_k)^T] \quad (18)$$

$$p(X_k|Z_k) \sim \mathcal{N}(\hat{X}_k, P_k) \quad (19)$$

0.2.2 Processus stochastique - Théorie

Comment décrire les transitions entre états, y compris en temps continu ?

Marche Aléatoire

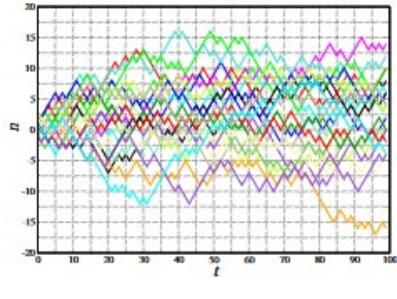


FIGURE 3.1 – 25 réalisations du mouvement brownien discret.

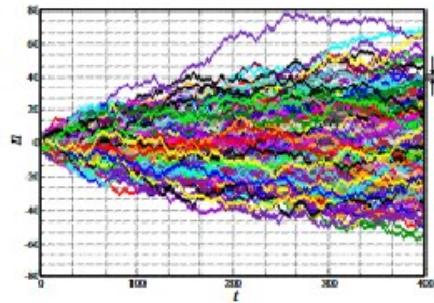


FIGURE 3.2 – 500 réalisations du mouvement brownien discret, pendant 400 tours.
La proportion de trajectoires qui aboutissent à $n = 40$ au temps $t = 400$ est la probabilité

Vers une description formelle des transitions entre états On prend un monde divisé en cases (numérotées chacune par un indice n). A chaque pas de temps, la particule fait un saut sur une case immédiatement à sa droite ou à sa gauche.

- Quelle est la probabilité $P(n, t)$ de trouver la particule dans la case n , au temps t ?
(sachant que la particule début à la case $n = 0$ à $t = 0$)
-

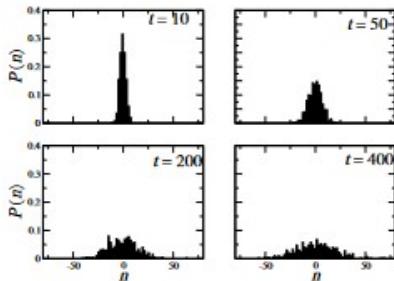


FIGURE 3.3 – Les probabilités $P(n, t_i)$ pour quatre temps $t_i = 10, 50, 200, 400$ obtenu à partir des trajectoires de la figure 3.2.

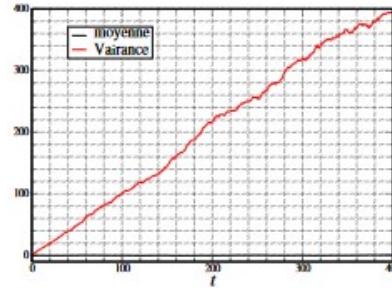


FIGURE 3.4 – Moyenne $\langle n(t) \rangle$ et variance $V(t)$ extraite des coupes temporelles des trajectoires de la figure 3.2.

On cherche à décrire $P(n, t_i)$, pour un temps t_i donné, ou autrement dit, on cherche à faire une coupe temporelle du mouvement brownien au temps t_i .

Pour cela, nous nous intéressons à extraire les moments statistiques, ici, la moyenne et la variance du processus en fonction du temps.

Ici, on remarque que (figure de droite), **la moyenne est nulle** à chaque pas de temps, et **la variance évolue linéairement** en fonction du temps.

Démonstration Soit D la variable aléatoire de déplacement à chaque tour. D prend les valeurs ± 1 avec une probabilité de 0,5.

Sa moyenne est $\mu = \langle D \rangle = 0$ et sa variance $\sigma^2 = \langle (D - \mu)^2 \rangle$.

La position de la particule est une variable aléatoire X qui s'exprime au bout de T pas de temps comme la somme de T déplacements D indépendants.

On obtient :

$$\langle X(T) \rangle = \langle \sum_{t=0}^T D_t \rangle = \sum_{t=0}^T \langle D_t \rangle = 0 \quad (20)$$

$$Var(X(T)) = Var\left(\sum_{t=0}^T D_t\right) = \sum_{t=0}^T Var(D_t) = T\sigma^2 \quad (21)$$

TODO : introduire ici la définition de marche aléatoire

TODO : introduire ici la définition de mouvement brownien

Le mouvement brownien

Généralisation vers le temps continu Vers une description formelle des transitions entre états

Généralisation à d'autres lois Le résultat précédent se généralise aisément à d'autres lois de probabilité de transition, mais le résultat est le même : la variance croît linéairement avec le temps.

- Quelle modélisation quand le pas de temps δt tend vers 0 ?

Généralisation au temps continu On décrit les probabilités de saut comme une densité qui dépend du pas de temps : $W(n)\delta t$.

On note :

- $W^-(n)$: **densité de probabilité ou taux de transition** de saut vers la gauche à partir de la case n .
- $W^+(n)$: **densité de probabilité ou taux de transition** de saut vers la droite.

Cette description peut être étendue à des sauts de plusieurs cases, en considérant $W(n, m)$: densité de probabilité de sauter de n à m . Et, éventuellement, à un espace d'états continus (*non traité formellement dans le cours*).

Vers l'équation Maîtresse

En reprenant les sauts de ± 1 case, nous pouvons effectuer un bilan infinitésimal (analogue aux bilans de physique).

- Probabilité d'aboutir à la case n à $t + \delta t$?

3 possibilités :

1. La particule est en $n + 1$ à t et saute vers n , avec une probabilité de $W^-(n + 1)dt$
2. La particule est en $n - 1$ à t et saute vers n , avec une probabilité de $W^+(n - 1)dt$
3. La particule reste en n avec une probabilité de $1 - (W^+(n) + W^-(n))dt$

En reprenant les sauts de ± 1 case, nous pouvons effectuer un bilan infinitésimal (analogue aux bilans de physique).

- Probabilité d'aboutir à la case n à $t + \delta t$?

On obtient alors (avec des événements indépendants) :

$$P(n, t + dt) = \frac{P(n + 1, t)W^-(n + 1)dt}{+ P(n - 1, t)W^+(n - 1)dt} + P(n, t)(1 - (W^+(n) + W^-(n)))dt \quad (22)$$

En reprenant les sauts de ± 1 case, nous pouvons effectuer un bilan infinitésimal (analogue aux bilans de physique).

- Probabilité d'aboutir à la case n à $t + \delta t$?

Ce qui nous mène à établir l'**équation Maîtresse** :

$$\frac{\partial P(n, t)}{\partial t} = \frac{W^+(n - 1)P(n - 1, t)}{W^-(n)P(n, t)} - \frac{W^-(n + 1, t)P(n + 1, t)}{W^+(n)P(n, t)} - \quad (23)$$

Nous obtenons en quelque sorte un **bilan de flux de probabilités** entre cellules adjacentes, telle qu'elle pourrait être établie en physique statistique, avec un flux sur la face gauche de la cellule, et un flux sur la face droite.

On peut ainsi généraliser l'équation précédente à des transitions entre états m, n quelconques.

$$\frac{\partial P(n, t)}{\partial t} = \sum_m W(m \rightarrow n)P(m, t) - W(n \rightarrow m)P(n, t) \quad (24)$$

Formellement, on écrira :

$$\partial_t P(t) = \mathcal{L}P(t) \quad (25)$$

avec $\mathcal{L} = \mathcal{L}_m^n = W(m \rightarrow n)$ la matrice de transition entre états n, m et $P(t)$ le vecteur colonne des cellules du système.

Remarque *On arrive ici au bout du travail de formalisation. L'exemple bilan sur un mouvement brownien avec des sauts élémentaires ± 1 pouvant être généralisé à d'autres problèmes avec des transitions plus complexes.*

Moyenne et Variance d'un processus stochastique en temps continu Après insertion des moments d'ordre 1, $\langle n(t) \rangle = \sum_n n P(n, t)$, et d'ordre 2, $\langle n^2(t) \rangle = \sum_n n^2 P(n, t)$, dans l'équation maîtresse, nous obtenons,

Moyenne :

$$\frac{dn(t)}{dt} = \langle W^+(n) - W^-(n) \rangle = \sum_n (W^+(n) - W^-(n)) P(n, t) \quad (26)$$

Variance :

$$\frac{dV}{dt} = 2\langle (n - \langle n \rangle)(W^+(n) - W^-(n)) \rangle + \langle W^+(n) + W^-(n) \rangle \quad (27)$$

Remarque Ces formules (un peu ingrates), nous permettent d'extraire une équation différentielle pour chacun des moments de la distribution, et obtenir ainsi la moyenne et la variance du processus au cours du temps.

Processus de Poisson

Exemple de processus stochastique

Le processus de Poisson est fondamental pour comprendre les processus stochastiques. On le rencontre sur plusieurs phénomènes :

- Nombre de particules radioactives qui se désintègrent pendant un temps t ,
- Nombre de communications reçues par un central sur une période ΔT ,
- Réaction chimique.

Exemple d'une particule radioactive :

On note adt la probabilité de désintégration d'une particule sur un temps dt

- Quelle est la probabilité d'observer n événements pendant un temps t ?
- On note : $W^+(n) = \alpha$, et $W^-(n) = 0$, (la probabilité de “ré-intégration”, ou de l'événement contraire est nulle).

Selon l'équation Maîtresse :

$$\frac{\partial P(n, t)}{\partial t} = \alpha(P(n-1, t) - P(n, t)), n \geq 1 \quad (28)$$

$$\frac{\partial P(n, t)}{\partial t} = \alpha P(0, t) \quad (29)$$

Sa résolution donne :

$$P(n, t) = e^{-\alpha t} \frac{(\alpha t)^n}{n!} \quad (30)$$

Exemple d'une particule radioactive :

- Quelle est la probabilité d'observer n événements pendant un temps t ?
- Quelles sont la moyenne et la variance du processus ?

Selon l'équation Maîtresse :

$$\frac{\partial \langle n(t) \rangle}{\partial t} = \sum_n (W^+(n) - W^-(n)) P(n; t) = \alpha, \text{ avec } n(t=0) = 0$$

donc $\langle n(t) \rangle = \alpha t$

et $\frac{\partial \langle n^2(t) \rangle}{\partial t} = 2\langle \alpha n \rangle + \langle \alpha \rangle = 2\alpha \langle n \rangle + \alpha$ avec variance nulle à l'origine $\langle n^2(0) \rangle = 0$
donc $\langle n^2(t) \rangle = \alpha^2 t + \alpha t$.

On obtient la moyenne et la variance d'un processus de Poisson :

$$M = \langle n \rangle = \alpha t, V = \langle n^2 \rangle - \langle n \rangle^2 = \alpha t \quad (31)$$

0.2.3 Exercice

Cinétique chimique

Nous posons $W^+(n) = \alpha$, $W^-(n) = \mu n$. L'équation de cinétique chimique est donnée par

$$\frac{dn}{dt} = \alpha - \mu n \quad (32)$$

Avec $n_0 = 0$, on a $n(t) = (\alpha/\mu)(1 - e^{-\mu t})$

1. Montrer que la variance vaut $V(t) = (\alpha/\mu)(1 - e^{-\mu t})$
2. Vérifier que pour l'état stationnaire, quand $\partial_t P = 0$, la solution est une distribution de Poisson de paramètre $\lambda = \alpha/\mu$

Relations utiles

$$\frac{dn}{dt} = W^+(n) - W^-(n) \quad (33)$$

$$\frac{dn^2(t)}{dt} = 2n(W^+(n) - W^-(n)) + W^+(n) + W^-(n) \quad (34)$$

$$n = \sum_n nP(n, t), n^2(t) = \sum_n n^2 P(n, t), f(n) = \sum_n f(n)P(n, t) \quad (35)$$

0.2.4 A retenir

- Construction d'un processus stochastique à partir des **taux de transition**, ou **densités de probabilités** de transition,
- Analogie avec des bilans physiques,
- Après formulation d'un processus stochastique, les grandeurs qui nous intéressent sont la moyenne du processus (qui souvent correspond à la dynamique classique), et sa variance (ce qui nous permettra de quantifier des incertitudes).

Remarques

Ce cours pose les bases sur processus stochastiques, ce qui permettra d'aller vers les modèles et applications lors des prochains cours : Filtre de Kalman, Processus Gaussiens, Chaînes de Markov.

0.2.5 Processus stochastiques - En bref

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 .

Équation 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,

$$\frac{dn}{dt} = W^+(n) - W^-(n), \frac{dV}{dt} = 2(n - \bar{n})(W^+(n) - W^-(n)) + W^+(n) + W^-(n) \quad (36)$$

0.2.6 Précisions sur le Mouvement Brownien

On a vu que pour une marche aléatoire, avec des sauts de ± 1 équiprobables, la coupe temporelle à un pas de temps donné $t \in \mathbb{N}$, la distribution des positions se rapproche d'une gaussienne.

On peut alors formaliser le mouvement brownien, en temps continu, par rapport à cette observation.

Définition Soit $B = (B_t)_{t \geq 0}$ une famille de variables aléatoires indexées dans le temps. On dit que B est un mouvement brownien si c'est un processus à trajectoires continues telles que

1. $\forall t \geq 0 : B_t \sim \mathcal{N}(0, t)$
2. Pour tout $0 \geq t_1 \geq \dots \geq t_n$, les variables aléatoires $B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_n} - B_{t_{n-1}}$, sont indépendantes.

Remarque La seconde propriété signifie que le mouvement brownien n'a pas de mémoire du passé.

*Note : Wiener a formalisé le mouvement brownien, d'où le nom de **Processus de Wiener** et les notations $W(n, m)$ vues précédemment*

0.2.7 Sources

Liens utiles du pour les processus stochastiques

(Houchmandzadeh, B.: Processus stochastiques,)

(Chopin, N.: Introduction aux processus stochastiques, Notes de cours)

(Sedlmeier et al. [2016])

0.3 Filtrage Kalman

0.3.1 Filtre Kalman

Filtre de Kalman (1960), Rudolf Kalman (mathématicien hongrois) : utilisé pour la première fois pour l'estimation de trajectoire des programmes Apollo.

→ Comment concilier au mieux l'information disponible (capteurs), et les équations de la dynamique pour contrôler un système ?

Applications : Contrôle, Calage GPS, Filtrage et assimilation de données.

Contrôle d'un système dynamique

Système dynamique discret

$$x_{n+1} = Ax_n + Bu_n + \phi_n \quad (37)$$

$$y_{n+1} = Cx_{n+1} + \psi_{n+1} \quad (38)$$

où, à chaque instant n :

- x_n est l'état du système. *Exemple : Température d'un moteur d'avion.*
- u_n est la commande du système. *Exemple : Débit de carburant dans le moteur.*
- y_n est la mesure de l'état du système. *Exemple : Mesure renvoyée par le thermomètre.*

Remarque : En météo, on appelleraient C un **opérateur d'observation**. C'est le lien entre l'**espace des mesures** (ex : la tension au bornes thermomètre), et l'**espace d'état** (ex : la température effective mesurée).

Bruits associés au système

Système dynamique discret

$$x_{n+1} = Ax_n + Bu_n + \phi_n \quad (39)$$

$$y_{n+1} = Cx_{n+1} + \psi_{n+1} \quad (40)$$

Bruits ϕ_n, ψ_n : on modélise un bruit d'état ϕ_n un bruit de mesure ψ_n .

1. **Bruit d'état** ϕ_n : représente notre méconnaissance de la physique du système. *Exemple : Variations de la richesse du mélange.*
2. **Bruit de mesure** ψ_n : représente un bruit de mesure. *Exemple: bruit électronique de la sonde de température.*

Hypothèse : Les bruits ϕ et ψ sont supposés blancs, gaussiens, centrés, stationnaires et indépendants l'un de l'autre. Ces bruit sont chacun associé à une matrice de covariance Φ , et Ψ .

Construction d'un estimateur

On cherche à construire un estimateur qui dépende de l'état estimé à l'instant précédent, de la mesure renoyée par le capteur et de la comande imposée.

Nous construisons alors un estimateur de la forme :

$$\hat{x}_{n+1} = A_f \hat{x}_n + B_f u_n + K_{n+1} y_{n+1} \quad (41)$$

→ Comment construire A_f , B_f , K_{n+1} ?

Note : On cherche à construire un estimateur de la forme générique $\hat{x}_{n+1} = f(\hat{x}_n, y_{n+1}, u_n)$

Nous nous concentrerons sur des systèmes linéaires, et verrons plus tard comment l'étendre à des systèmes non-linéaires.

Estimateur (assymptotiquement) sans biais

Erreur d'estimation : On cherche à minimiser l'erreur d'estimation, définie ci-dessous,

$$e = \hat{x} - x \quad (42)$$

Pour un estimateur parfait, on aurait $\forall n \in \mathbb{N}, e_n = 0$.

Sans accès aux vraies valeurs de x et face à un processus stochastique, annuler l'erreur est généralement impossible. Nous décrivons alors (non sans mal) l'erreur e_n par récurrence:

$$e_{n+1} = (I - K_{n+1}C)Ae_n + (A_f + K_{n+1}CA - A)\hat{x}_n + (B_f + K_{n+1}CB - B)u_n + (K_{n+1}C - I)\phi_n + K_{n+1}\psi_{n+1} \quad (43)$$

Espérance de l'erreur Sachant que les bruits blancs gaussiens sont d'espérance nulle :

$\forall n \in \mathbb{N}, \mathbb{E}[\phi_n] = 0, \mathbb{E}[\psi_n] = 0$, on décrit l'évolution, l'espérance de l'erreur.

$$\mathbb{E}[e_{n+1}] = (I - K_{n+1}C)A\mathbb{E}[e_n] + (A_f + K_{n+1}CA - A)\hat{x}_n + (B_f + K_{n+1}CB - B)u_n \quad (44)$$

Estimateur assymptotiquement sans biais

On cherche à construire un estimateur dont l'espérance de l'erreur tend vers 0.

Définition

$$\lim_{n \rightarrow \infty} \mathbb{E}[e_n] = 0 \quad (45)$$

Espérance de l'erreur Sachant que les bruits blancs gaussiens sont d'espérance nulle :

$\forall n \in \mathbb{N}, \mathbb{E}[\phi_n] = 0, \mathbb{E}[\psi_n] = 0$, on décrit l'évolution, l'espérance de l'erreur.

$$\mathbb{E}[e_{n+1}] = (I - K_{n+1}C)A\mathbb{E}[e_n] + (A_f + K_{n+1}CA - A)\hat{x}_n + (B_f + K_{n+1}CB - B)u_n \quad (46)$$

Estimateur assymptotiquement sans biais $\rightarrow \lim_{n \rightarrow \infty} \mathbb{E}[e_n] = 0$

Conditions suffisantes

1. $B_f + K_{n+1}CB - B = 0$ (*on annule le terme lié à la comande u_n*)
 2. $A_f + K_{n+1}CA - A = 0$ (*on annule le terme dû à l'estimation \hat{x}_n*)
 3. $(I - K_{n+1}C)$ est stable (*module inférieur à 1 pour que l'erreur puisse décroître*)
-

Construction des Matrices du filtre

Filtre de Kalman

Les conditions de stabilité fixent les matrices A_f et B_f .

$$A_f = (I - K_{n+1}C)A \quad (47)$$

$$B_f = (I - K_{n+1}C)B \quad (48)$$

Il reste à régler K_{n+1} pour que $(I - K_{n+1}C)$ soit stable.

Définition : K est appelé le **gain du filtre**.

Forme de l'estimateur : Nous établissons (et rencontrons souvent) la forme suivante du filtre.

$$\hat{x}_{n+1} = A\hat{x}_n + Bu_n + K_{n+1}[y_{n+1} - C(A\hat{x}_n + Bu_n)] \quad (49)$$

Remarque K peut être vu comme un compromis à régler entre la fidélité au modèle numérique ($A\hat{x}_n + Bu_n$) et la fidélité aux valeurs de mesure (y_{n+1}).

Filtre Prédicteur - Correcteur

Structure de prédicteur-correcteur

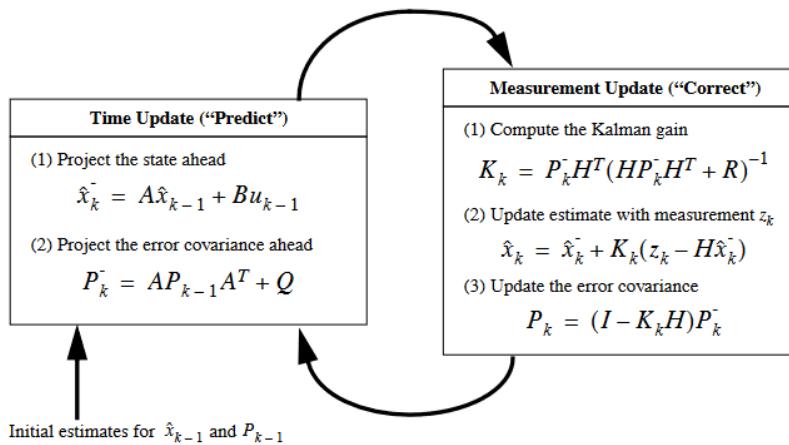
Le filtre de Kalman est un “prédicteur-correcteur”, l'estimation \hat{x} de x se construit en 2 temps :

Mise à jour de l'état :

1. **Prédiction** : Estimation à priori de l'état \hat{x}^- , comme si on n'avait que les équations du système à disposition.
2. **Correction** : Construction de l'estimation à posteriori avec l'information apportée par les mesures.

Mise à jour des matrices de covariance (liées au bruit ajouté à chaque pas) :

1. **Prédiction** : Estimation de la matrice de covariance P_k^- , par rapport au **bruit d'état** Q (ou Φ comme noté précédemment).
2. **Correction** : Construction de la matrice de covariance P_k par rapport au **bruit de mesure** R (ou Ψ comme noté précédemment).



Sur l'image ci-dessus, par cohérence avec les notations du cours : $H \rightarrow C, Q \rightarrow \Phi, R \rightarrow \Psi$ “Schéma du prédicteur-correcteur associé au Filtre de Kalman. Source : Welch & Bishop, Intro to the Kalman Filter”

0.3.2 Filtre de Kalman - Implémentation

La recette de cuisine !

1. On initialise \hat{x} à \hat{x}_0 : on peut par exemple prendre la valeur y_0 renvoyée par le capteur.
2. On initialise P à P_0 : on peut prendre la valeur Ψ de la covariance de bruit du capteur.
3. On fait évoluer K selon :

$$K_{n+1} = (AP_nA^T + \Psi)C^T \times (CAP_nA^TC^T + C\Psi C^T + \Psi)^{-1} \quad (50)$$

4. On fait évoluer \hat{x} selon :

$$\hat{x}_{n+1} = A\hat{x}_n + Bu_n + K_{n+1}[y_{n+1} - C(A\hat{x}_n + Bu_n)] \quad (51)$$

5. On fait évoluer P selon :

$$P_{n+1} = (I - K_{n+1}C)(APA^T + \Phi) \quad (52)$$

Est-ce que ça marche vraiment en pratique ?

Considérations sur l'évolution de la variance

On a établit que $\lim_{n \rightarrow \infty} \mathbb{E}[e_n] = 0$, ainsi que la formule pour P_{n+1} .

→ Est-ce qu'on est obligé d'attendre ∞ pour que le filtre commence à fonctionner ?

→ Pourquoi on se balade avec une matrice de covariance P_{n+1} (si ce n'est pas par pure beauté mathématique) ?

→ En pratique, comment fixer K_{n+1} le gain du filtre ?

Dispersion Nous allons travailler sur la dynamique de l'erreur e_n et sa dispersion, pour

$$P = \mathbb{E}[e_n \times e_n^T] \quad (53)$$

Dans un cas simple à une dimension, $P = \mathbb{E}[(\hat{x}_n - x_n)^2] = \mathbb{V}[e_n]$, s'écrit bien comme la variance de l'erreur.

0.3.3 Dynamique de l'erreur

Choix du gain K

Condition d'optimalité du gain du filtre

On cherche K de telle sorte que la variance de l'erreur $P = \mathbb{E}[e \times e^T] \sim \mathbb{E}[e]$ soit minimale.

Une condition nécessaire est de chercher le gain K tel que $P(k)$ soit extrémal, autrement dit que

$$\forall n \in \mathbb{N}, \frac{\partial P_{n+1}}{\partial K_{n+1}} = 0 \quad (54)$$

Dynamique de l'erreur En manipulant les équations du filtre, on obtient

$$e_{n+1} = (I - K_{n+1}C)Ae_n + (K_{n+1}C - I)\phi_n + K_{n+1}\psi_{n+1} \quad (55)$$

$$\frac{\partial e_{n+1}}{\partial K_{n+1}} = -CAe_n + C\phi_n + \psi_n \quad (56)$$

On cherche le gain K pour établir $\forall n \in \mathbb{N}, \frac{\partial P_{n+1}}{\partial K_{n+1}} = 0$

Condition suffisante

$$\mathbb{E}\left[\frac{\partial e_{n+1}}{\partial K_{n+1}} e_{n+1}^T\right] = 0 \quad (57)$$

Formule de K à partir de P

$$P_{n+1} = (AP_n A^T + \Psi)C^T \times (CAP_n A^T C^T + C\Phi C^T + \Psi)^{-1} \quad (58)$$

Formule de P à partir de K (en prenant en compte $P_0 = \Psi$)

$$P_{n+1} = (I - K_{n+1}C)(AP_n A^T + \Phi) \quad (59)$$

Remarque On ne peut pas obtenir de condition d'optimalité, mais simplement d'une relation de récurrence entre P et K . C'est elle qui nous permet d'implémenter K en pratique.

0.3.4 Synthèse

Estimateur assymptotiquement sans biais construit de manière récurrente

Estimateur sans biais

On cherche à obtenir un estimateur qui satisfait $\lim_{n \rightarrow \infty} \mathbb{E}[e_n] = 0$ (assymptotiquement sans biais)

Gain optimal

On cherche à régler K pour que les variances d'erreurs soit les plus faibles à chaque pas de temps, c'est-à-dire $\frac{\partial P_{n+1}}{\partial K_{n+1}} = 0$

Récurrence

Dans la mesure où nous ne connaissons pas les valeurs vraies x_n , nous tirons parti des relations de récurrence sur l'erreur e_n pour progresser vers une erreur nulle.

C'est cela même qui fait la structure du Filtre Kalman. Et c'est bien pratique dans la mesure où le filtre ne dépend que des valeurs à l'état n pour estimer l'état $n + 1$ (le filtre est robuste et facile à mettre en oeuvre).

Structure probabiliste

→ Quel lien avec les probabilités ?

Point de départ

- Bruit d'état $\phi \sim \mathcal{N}(0, \Phi)$ (ex : Φ donné par la distribution de températures au point de mesure, i.e. la climatologie du lieu)
- Bruit de mesure $\psi \sim \mathcal{N}(0, \Psi)$ (ex : Ψ , précision donnée par la fiche technique du capteur de température)

Point d'arrivée

- On modélise $P(x_n | z_n) \sim \mathcal{N}(\hat{x}_n, P_n)$, avec notre estimateur \hat{x}_n
 - Avec les propriétés de l'estimateur, on progresse vers $\mathbb{E}[\hat{x}_n] = x_k$ en gardant une dispersion minimale $\mathbb{E}[(x_n - \hat{x}_n)(x_n - \hat{x}_n)^T] = P_n$.
-

Implémentation pratique d'un filtre de Kalman

Hypothèses structurantes Il reste en pratique à vérifier que les bruits de mesure et d'état sont effectivement des bruits blancs gaussiens, stationnaires.

1. Qu'on a “suffisamment capté” la dynamique du système et son observation dans les matrices A , B et C , pour que les bruits Φ et Ψ soient effectivement centrés.
2. Que les bruits puissent être assimilés à des bruits blancs gaussiens. Cela implique des tests statistiques, et éventuellement un travail sur les équations pour centrer et réduire le bruit (lien avec le TCL).
3. Qu'il y ait effectivement indépendance entre le bruit de mesure et le bruit d'état. *Exemple : le capteur de température qui influence son environnement en ralentissant le flux d'air dont il mesure la température.*

En pratique

Il n'y a pas forcément de réponse systématique à ces questions, simplement un travail de mise au point du filtre, sur un problème donné → **Coeur du travail de l'ingénieur**

0.3.5 Exemple - Estimation d'une constante aléatoire

Filtre de Kalman

Problème

Nous cherchons à estimer une tension constante. Avec, à disposition, un volt-mètre, dont les mesures sont perturbées par un bruit blanc d'écart-type $\sigma = 0.1V$.

Système linéaire

$$x_{n+1} = x_n + \phi_{n+1} \quad (60)$$

$$z_{n+1} = x_{n+1} + \psi_{n+1} \quad (61)$$

Équations du Filtre Dans ce cas simple, $A_f = B_f = C = I = 1$, nous étudions alors l'estimation \hat{x} , la (co)-variance de l'erreur P_n et le gain K_n , qui sont les caractéristiques essentielles du filtre.

1. Prédicteur :

$$\hat{x}_{n+1}^- = \hat{x}_n \quad (62)$$

$$P_{n+1}^- = P_n + \Phi \quad (= P_n + \mathbb{V}[\phi_{n+1}]) \quad (63)$$

2. Correcteur :

$$K_{n+1} = P_{n+1}^-(P_{n+1}^- + \Psi)^{-1} = \frac{P_n + \Phi}{P_n + \Phi + \mathbb{V}[\psi_n]} \quad (64)$$

$$\hat{x}_{n+1} = \hat{x}_{n+1}^- + K_{n+1}(z_{n+1} - \hat{x}_{n+1}^-) \quad (65)$$

$$P_{n+1} = (1 - K_{n+1})P_{n+1}^- \quad (66)$$

Hypothèses et modélisation

1. Ici, on suppose une variance sur le bruit d'état : $\Phi = 1 \times 10^{-5}$. C'est à dire qu'on se laisse la possibilité d'avoir un petit bruit d'état autour de la constante à mesurer (perturbations électromagnétiques ou autre).
2. Choix des conditions initiales : on suppose que la tension à mesurer est une constante tirée d'une loi normale centrée.

On fixe alors $x_0 = 0$, et on choisit une valeur arbitraire, mais non-nulle ($P_0 \neq 0$) pour P_0 . Ici, $P_0 = 1$.

Vraie valeur, Mesures, et Estimation (50 relevés)

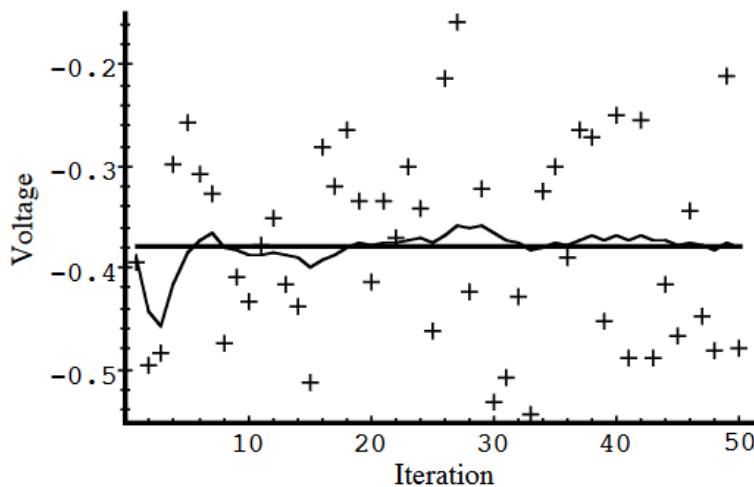


Figure 3-1. The first simulation: $R = (0.1)^2 = 0.01$. The true value of the random constant $x = -0.37727$ is given by the solid line, the noisy measurements by the cross marks, and the filter estimate by the remaining curve.

Source : Welch & Bishop, *Intro to the Kalman Filter*"

Vitesse de convergence - (Co)-Variance de l'erreur P_n

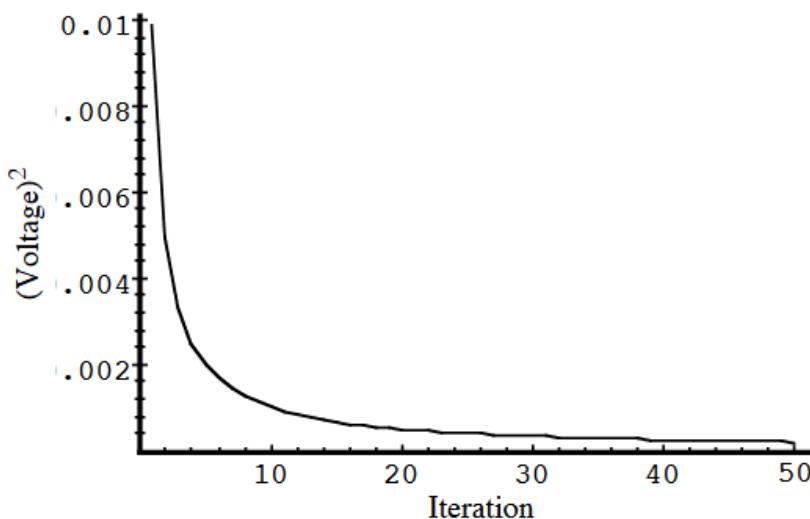


Figure 3-2. After 50 iterations, our initial (rough) error covariance P_k^- choice of 1 has settled to about 0.0002 (Volts²).

Source : Welch & Bishop, *Intro to the Kalman Filter*"

Influence des bruits Ψ et Φ , avec $\Psi = 1$

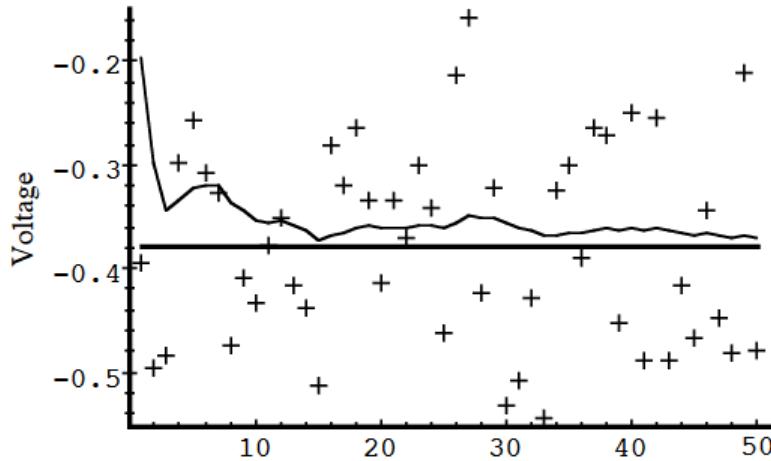


Figure 3.3. Second simulation: $R = 1$. The filter is slower to respond to the measurements, resulting in reduced estimate variance.

Source : Welch & Bishop, *Intro to the Kalman Filter*"

Influence des bruits Ψ et Φ , avec $\Psi = 1.10^{-5}$

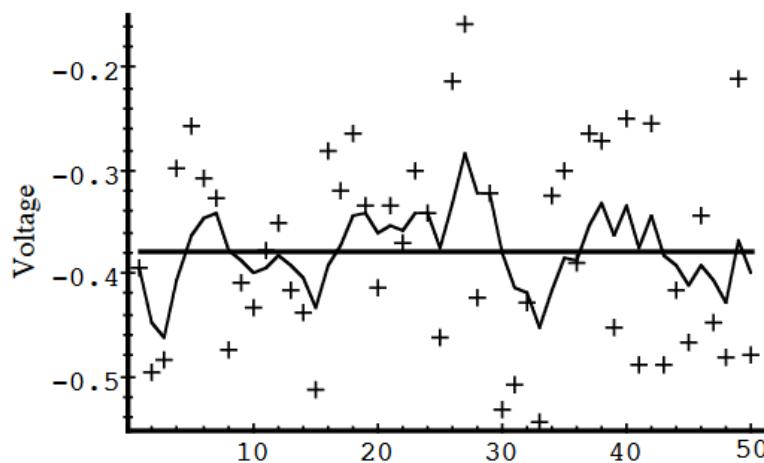


Figure 3.4. Third simulation: $R = 0.0001$. The filter responds to measurements quickly, increasing the estimate variance.

Source : Welch & Bishop, *Intro to the Kalman Filter*"

0.3.6 Dynamique Non-Linéaire

Extended Kalman Filter

Si la dynamique est non linéaire ?

Remplaçons :

$$x_{n+1} = Ax_n + Bu_n + \phi_n \quad (67)$$

$$y_{n+1} = Cx_{n+1} + \psi_{n+1} \quad (68)$$

Par :

$$x_{n+1} = f(x_n, u_n, \phi_n), (\text{modèle non-linaire}) \quad (69)$$

$$y_{n+1} = h(x_{n+1}, \psi_{n+1}), (\text{observateur non-linaire}) \quad (70)$$

Remarque C'est le cas en Météo où le modèle repose sur les équations de Navier-Stokes (non-linéaires), et les observateurs reposent sur les lois non-linéaires (ex : réflectivité radar $R \propto D^6$)

Solution : Linéariser les équations (modèle et observateurs), autour du point de fonctionnement \hat{x}_n estimé.

Outil : les matrices jacobiniennes des opérateurs $y = h(x)$ et modèles $x_{n+1} = f(x_n)$

On note alors :

1. $J_{(i,j)}^A = \frac{\partial f_i}{\partial x_j}$, la jacobienne de f par rapport à x
2. $W_{(i,j)} = \frac{\partial f_i}{\partial \phi_j}$, la jacobienne de f par rapport au bruit d'état ϕ
3. $J_{(i,j)}^H = \frac{\partial h_i}{\partial x_j}$, la jacobienne de h par rapport à x
4. $V_{(i,j)} = \frac{\partial h_i}{\partial \psi_j}$, la jacobienne de h par rapport au bruit de mesure ψ .
5. On utilise directement f , et h directement pour le calcul de \hat{x}_n^- à priori et l'évaluation de la mesure $h(y_n)$.
6. On utilise les matrices jacobiniennes dans le calcul des covariances :

$$\begin{aligned} P_{n+1}^- &= AP_n A^T + \Phi \text{ devient } P_{n+1}^- = J^A P_n (J^A)^T + W \Phi W^T \\ K_{n+1} &\text{ devient } K_{n+1} = P_{n+1}^- (J^H)^T (J^H P_{n+1}^- (J^H)^T + V \Psi V^T) \\ P_{n+1} &\text{ devient } P_{n+1} = (I - K_{n+1} J^H) P_{n+1}^- \\ \text{où } J^A \text{ et } J^H &\text{ ont été évaluées au point } \hat{x}_n \end{aligned}$$

En Météo

On ne s'étonnera pas de trouver les termes de **Tangent-Linéaire** (TL) et **Adjoint** (AD) en assimilation de données, il s'agit respectivement de la **jacobienne** J^A et de sa **transposée** $(J^A)^T$ (*plus exactement du conjugué de sa transposée si on travaille sur un espace complexe*).

En pratique

Là encore, le **coeur du travail d'ingénieur** est d'obtenir les opérateurs adéquats (en respectant de manière empirique les hypothèses sur les distributions).

0.3.7 A retenir

Qu'est-ce qu'un filtre de Kalman déjà ?

Basique

- La recette et le fonctionnement de prédicteur-correcteur pour le filtre Kalman,
- (Théorie), c'est un estimateur assymptotiquement sans biais, et avec une variance d'erreur à minimiser.

En fonction de l'espace de stockage disponible

- La démarche récursive de construction du filtre, *utile pour comprendre le filtre "pas-à-pas"*.

En pratique, avec l'algorithme à disposition

- Comment régler le gain du filtre ?
- Comment respecter au mieux les hypothèses de bruit blanc gaussien ?
- Comment linéariser le modèle et les observateurs de mesure ?

En gardant à l'esprit que le Filtre de Kalman est plutôt un filtre simple et robuste, et que le travail consiste à construire les “bonnes” matrices de covariance, et régler le gain du filtre.

0.3.8 Sources

Liens utiles

- [](Welch, Greg & Bishop, Gary. (2006). An Introduction to the Kalman Filter. Proc. Siggraph Course. 8.)
(G. Welch, G. Bishop, University of North Carolina, An Introduction to the Kalman Filter)
(O. Herscovici-Schiller, ONERA, Introduction au filtrage de Kalman et à la commande optimale)
(G. Chardon, CentraleSupélec, Filtrage de Kalman)
-

0.4 Chaînes de Markov

0.4.1 Définition

Définition : Une chaîne de Markov homogène à valeurs dans S de matrice de transition P et de loi initial λ est une famille de variables aléatoires $(X_n, n \in \mathbb{N})$ telle que :

1. $\mathbb{P}(X_0 = i) = \lambda_i$
2. $\mathbb{P}(X_{n+1} = i_{n+1}|X_n = i_n, \dots, X_0 = i_0) = \mathbb{P}(X_{n+1} = i_{n+1}|X_n = i_n) = p_{i_n, j_n}$

$(X_n)_{n \in \mathbb{N}}$ est une chaîne de Markov de paramètres (λ, P)

Caractérisation : $(X_n)_{0 \leq n \leq N}$ est Markov λ, P ssi pour tous i_0, \dots, i_N

$$\mathbb{P}(X_0 = 0, \dots, X_N = i_N) = \lambda(i_0)p_{i_0, i_1}p_{i_1, i_2} \dots p_{i_{N-1}, i_N} \quad (71)$$

0.4.2 Conditionnements successifs et calcul

Propriété : Si $(X_n)_{0 \leq n \leq N}$ est Markov (λ, P) , alors, conditionnellement à $X_m = i$, le processus $(X_{m+n}, n \geq 0)$ est Markov (δ_i, P) et est indépendant de (X_0, \dots, X_m) .

On note δ_i la masse de Dirac en i , et \mathbb{P}_i la loi d'une chaîne (δ_i, P) . On dit que la chaîne est issue de i car $X_0 = i$ p.s.

Calcul algébrique

On note une mesure positive λ comme un vecteur ligne. λP est la mesure $(\lambda P)(j) = \sum_i \lambda(i)p_{i,j}$.

On note $p_{i,j}^{(n)}$ l'élément i, j de la matrice P^n .

Propriété : Soit $(X_n)_{n \geq 0}$ Markov (λ, P) . Alors, pour tout n , la loi de X_n est λP^n , i.e.

$$\mathbb{P}(X_n = j) = (\lambda P^n)(j) \quad (72)$$

En particulier, pour $\lambda = \delta_i$ et tout $m \geq 0$, on obtient

$$\mathbb{P}_i(X_n = j) = \mathbb{P}(X_{m+n} = j | X_m = i) = p_{i,j}^n \quad (73)$$

0.4.3 Etats transients et récurrents

Définition

- On dit que i est transient si $\sum_n p_{ii}^n < \infty$
- On dit que i est récurrent si $\sum_n p_{ii}^n = +\infty$

On peut alors mettre la matrice de transition sous une forme canonique, avec les états absorbants au début et les états récurrents ensuite :

$$P = \left(\begin{array}{c|c} Q & R \\ \hline 0 & I \end{array} \right) \quad (74)$$

où $Q^n \rightarrow 0$.

Propriété Pour une chaîne de Markov avec un état absorbant, la matrice $I - Q$ est inversible d'inverse $N = \sum_n Q^n$. n_{ij} est le nombre moyen de fois où la chaîne est dans l'état j si elle part de l'état i .

0.4.4 Mesures invariantes (ou stationnaires)

Propriétés

On dit que la mesure positive λ est invariante (ou encore stationnaire) si $\lambda P = \lambda$.

Proposition Si $(X_n)_{n \in \mathbb{N}}$ est Markov (λ, P) et λ est invariante, alors la loi de X_n est constante et vaut λ . En outre, pour tout m , $(X_{m+n}, n \geq 0)$ est Markov (λ, P) .

En particulier, les limites des lois de X_n sont automatiquement des probabilités invariantes sur un espace fini.

Propriété On suppose I fini et que pour un $i_0 \in I$, on ait $\forall j, p_{i_0,j}^{(n)} \rightarrow \pi_j$.

Alors π est une probabilité invariante.

0.4.5 Théorème ergodique

Théorème Soit X une chaîne irréductible positive récurrente. Alors, pour toute fonction f positive ou bornée, presque sûrement :

$$\frac{1}{n} \sum_{k=1}^n f(X_k) \rightarrow \pi(f) \quad (75)$$

0.4.6 Convergence vers l'équilibre

Théorème Soit une chaîne de Markov irréductible positive récurrente de probabilité invariante π et apériodique. Alors pour toute loi initiale

$$\mathbb{P}(X_n = i) \rightarrow \pi(i) \quad (76)$$

en particulier, pour tout i , $p_{ij}^{(n)} \rightarrow \pi(j)$.

0.4.7 Les propriétés d'une chaîne de Markov

Homogénéité

Une chaîne de Markov est homogène si ses probabilités de transitions ne dépendent pas du temps.

Irréductibilité

Une chaîne de Markov est irréductible si chaque état est accessible (en un ou plusieurs pas) depuis chaque état. C'est à dire qu'il n'y a pas d'état absorbant. Pour un état $L'k$ la probabilité d'atteindre cet état depuis un autre $l \neq k$ est non nulle.

Récurrence

Les états **récurrents** d'une chaîne de markov sont les états visités régulièrement par la chaîne. Si l'espérance du temps de retour, pour un état donné, est borné, l'état est **positivement récurrent**, sinon, l'état est **nullement récurrent**.

Une chaîne est récurrente selon la condition de Harris, si tous les états $X \in S$ de cette chaîne sont visités indéfiniment lorsque $t \rightarrow \infty$. Cette propriété est importante pour les espaces d'états continus et non bornés. Quand la chaîne tend vers un ensemble d'états irréductibles et récurrents selon la condition de Harris, elle y reste et visite tous les états avec une probabilité de 1.

Stationnarité

Une chaîne de Markov stationnaire a des probabilités de transitions indépendantes du temps. Si P est une matrice $n \times n$:

$$\pi P = \pi$$

π est alors la distribution limite de la chaîne. Les méthodes MCMC ont vocation à explorer cette distribution limite.

Ergodicité

L'ergodicité est une propriété des chaînes de Markov irréductibles, apériodiques et récurrentes positivement selon la condition de Harris. L'ergodicité est définie comme

$$\forall \theta_i, \theta_j \in \Theta, \lim_{n \rightarrow \infty} P^{(n)}(\theta_i \rightarrow \theta_j) = \pi(\theta_j) \quad (77)$$

L'ergodicité est la propriété selon laquelle, à partir d'un certain rang, la chaîne converge vers la distribution stationnaire.

0.4.8 Application au PageRank

L'algorithme PageRank a émergé à la fin des années 90, sous l'impulsion de Brin et Page, les fondateurs de Google. Contrairement à une recherche thématique ou sémantique classique (bag of words, similarité cosinus, etc.), l'idée de Brin et Page est d'exploiter les liens entre les pages web, **un site web étant important s'il est relié à d'autres sites importants**.

Le Page Rank s'appuie sur une marche aléatoire sur les sites webs afin d'estimer le nombre de visites d'une page web.

Définition

Soit X_t la page web visitée par une marche aléatoire au pas de temps t .

On note $N(i, n)$ le nombre de visites de la page web i pour $t \leq n$. Le **Page Rank** de la page i est définie par sa fréquence de visites, dans la limite d'une marche aléatoire infinie :

$$\rho(i) = \lim_{n \rightarrow \infty} \frac{N(i, n)}{n} \quad (78)$$

La marche aléatoire est définie sur l'ensemble des Pages Web. En partant d'une page i , l'algorithme choisit l'un de ses liens sortants de manière équiprobable.

Introduction d'un surfeur aléatoire

Le Page Rank s'appuie sur l'idée d'un surfeur aléatoire. Le surfeur aléatoire mime le comportement d'un internaute. Le surfeur circule sur les pages web en cliquant sur les liens hypertextes, et, si l'on se retrouve bloqué (page sans liens), il redémarre une recherche depuis la barre URL.

Ainsi, l'algorithme PageRank repart régulièrement d'une page web tirée aléatoirement, afin d'éviter un blocage sur les feuilles du graphe web.

Exemple

On donne ci-dessous un exemple de graphe web G_1 avec sa matrice de transition M :

$$M = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 1/3 & 1/3 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 0 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad (79)$$

- Pouvez-vous décrire les limitations de cette modélisation ?
 - Comment modifier la matrice de transitions dans le cadre de PageRank ?
-

0.5 Markov Chains Monte Carlo

0.5.1 Markov Chain Monte Carlo

Au cours précédent

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.

→ Comment construire un estimateur quand on ne connaît plus la structure du bruit ?

Chaînes de Markov et Méthodes de Monte-Carlo

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.

Vers le Filtre Kalman d'Ensemble

Remarque : Ce cours est indépendant de celui sur le Filtrage Kalman. Mais il va nous permettre de définir une autre méthode de calcul des matrices de covariance pour le filtre Kalman.

→ Le Filtre de Kalman et les Chaînes de Markov sont 2 représentants de Processus Stochastiques

→ 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 ?

0.5.2 Markov Chain Monte-Carlo

MCMC

→ 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*).
-

0.5.3 Historique et applications

- Aiguilles de Buffon (1733) : Estimation de π 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.
- 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.
-

0.5.4 Méthodes de Monte-Carlo

Construction de l'estimateur

Estimateur de Monte-Carlo Pour une suite de variables aléatoires indépendantes, identiquement distribuées, $\{q_1, \dots, q_N\}$ (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 :

$$\hat{f}_N^{MC} = \frac{1}{N} \sum_{n=1}^N f(q_n) \quad (80)$$

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

$$\lim_{N \rightarrow \infty} \hat{f}_N^{MC} = \mathbb{E}_{\pi}[f] \quad (81)$$

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 : $\lim_{N \rightarrow \infty} \hat{f}_N^{MC} = \delta_{\mathbb{E}_{\pi}[f]}$.

Erreur de l'estimateur

Propriété : L'estimateur défini est assymptotiquement sans biais.

→ 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 $\mathbb{E}_{\pi}[f]$, $\mathbb{E}_{\pi}[f^2]$ 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.

$$\lim_{N \rightarrow \infty} \frac{\hat{f}_N^{MC} - \mathbb{E}_{\pi}[f]}{SE_N} \sim \mathcal{N}(0, 1) \quad (82)$$

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

En bref

Autrement dit, à la limite,

$$\hat{f}_N^{MC} = \mathcal{N}(\mathbb{E}_\pi[f], SE_N[f]) \quad (83)$$

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 dimensionner 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 : $\hat{f}_N^{MC} > \mathbb{E}_\pi[f] + 3SE_N[f]$).

0.5.5 Chaînes de Markov

→ But : on cherche à explorer une distribution cible π .

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

$$P(q_{n+1}|q_1, \dots, q_n) = P(q_{n+1}|q_n) \quad (84)$$

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 π .

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

$$T: \begin{aligned} Q \times Q &\rightarrow \mathbb{R}^+ \\ (q, q') &\mapsto T(q'|q) \end{aligned} \quad (85)$$

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**.

$$\bar{q}_1 \sim T(q_1|q_0) \quad (86)$$

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

$$\begin{aligned} \bar{q}_1 &\sim T(q_1|q_0) \\ &\dots \\ \bar{q}_N &\sim T(q_N|q_{N-1}) \end{aligned} \quad (87)$$

→ On génère des séquences de points corrélés.

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

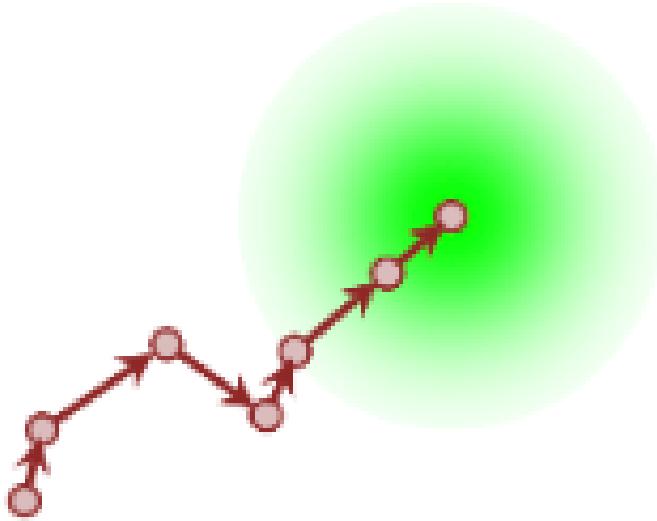
Exemple : trajectoire sur un espace à 2 dimensions

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

$$\begin{aligned}\omega_1: \quad Q &\rightarrow \mathbb{R} & \omega_2: \quad Q &\rightarrow \mathbb{R} \\ q &\mapsto q^1 & q &\mapsto q^2\end{aligned}\tag{88}$$

On définit une densité de probabilité de transition de Markov :

$$T(q_1|q_0) = \mathcal{N}(q_1^1|q_0^1, \sigma) \mathcal{N}(q_1^2|q_0^2, \sigma)\tag{89}$$



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

Distribution stationnaire (ou invariante)

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 $T('q|q')$.

$$\pi = \int dq' \pi(q') T(q|q')\tag{90}$$

Application

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

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 :

$$(T\rho)(q_1) = \int dq_0 T(q_1|q_0)\rho(q_0) \quad (91)$$

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

$$(T^2\rho)(q_2) = (T \cdot T\rho)(q_2) = \int dq_1 dq_2 T(q_2|q_1)T(q_1|q_0)\rho(q_0) \quad (92)$$

$$(T^N\rho)(q_N) = (T \cdot T^{N-1}\rho)(q_{N-1}) = (T \cdot \dots \cdot T\rho)(q_N) \quad (93)$$

En consruisant la densité de probabilité au point N :

$$(T^N\rho)(q_N) = (T \cdot T^{N-1}\rho)(q_{N-1}) = (T \cdot \dots \cdot T\rho)(q_N) \quad (94)$$

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

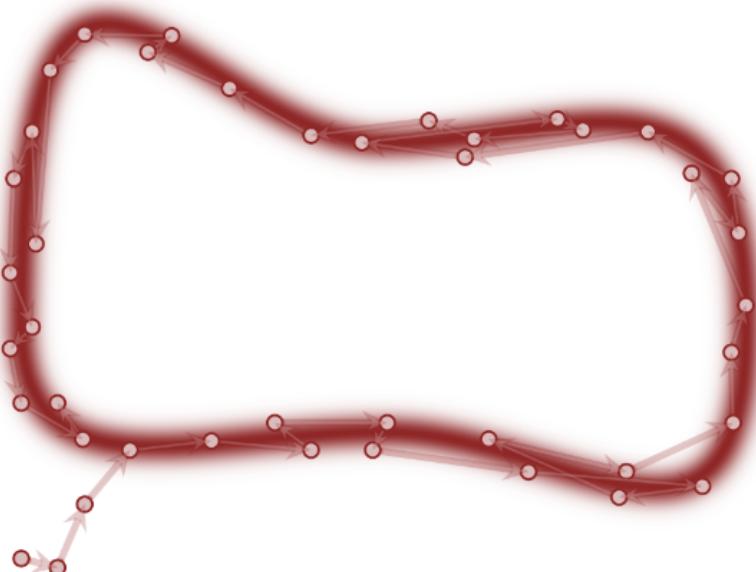
$$\lim_{N \rightarrow \infty} T^N \rho = \pi \quad (95)$$

Alors c'est un point fixe :

$$T\pi = \pi \quad (96)$$

→ 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.

Stationnarité - Illustration



Source : *Markov Chain Monte Carlo in Practice*, M. Betancourt

(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

$$\|T^N \rho - \pi\| \leq \epsilon \quad (97)$$

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 rapide (en un nombre limité d'opérations), mais est typiquement réservée aux espaces bornés.

(Optionnel) Spectre de la matrice de transition et Convergence

0.5.6 Markov Chain Monte-Carlo

Construction de l'estimateur

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

$$\hat{f}_N^{MCMC} = \frac{1}{N+1} \sum_{n=0}^N f(q_n) \quad (98)$$

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

$$\lim_{N \rightarrow \infty} \pi_{f_N^{MCMC}} = \delta_{\mathbb{E}_{\pi}[f]} ??? \quad (99)$$

MCMC : 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.

$$\lim_{N \rightarrow \infty} \pi_{f_N^{MCMC}} = \delta_{\mathbb{E}_{\pi}[f]} \quad (100)$$

Ce résultat peut être généralisé à toute distribution initiale 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.

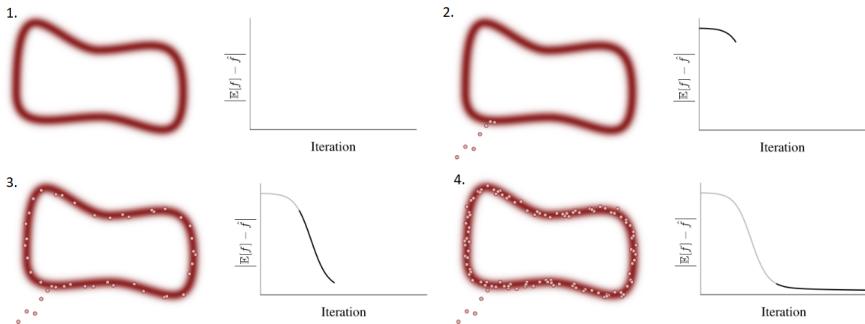
→ En pratique, construire ou utiliser un estimateur **MCMC** nécessite de vérifier précautionneusement les hypothèses sur les transitions de la chaîne de Markov.

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.

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

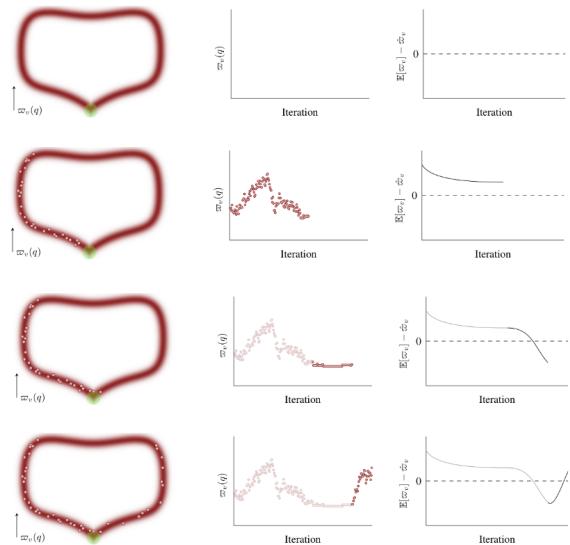
Illustrations - Cas stable



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

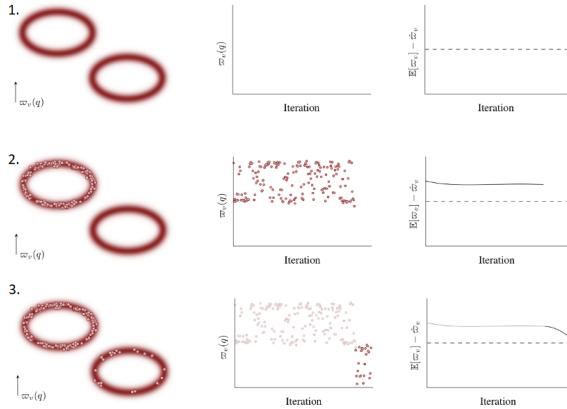
Source : *Markov Chain Monte Carlo in Practice*, M. Betancourt

Illustrations - Cas instables



Source : *Markov Chain Monte Carlo in Practice*, M. Betancourt

Illustrations - Cas instables



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

Source : *Markov Chain Monte Carlo in Practice*, M. Betancourt

Convergence

→ 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.

(Optionnel) Vitesse de convergence théorique

(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 :

$$\hat{f}_N^{MCMC} \sim \mathcal{N}(\mathbb{E}[f], MCMC - SE[f]) \quad (101)$$

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

$$MCMC - SE[f] = \sqrt{\frac{Var[f]}{\lambda[f] \cdot N}} \quad (102)$$

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.

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 :

$$\begin{aligned} K: \quad Q \times Q &\rightarrow \mathbb{R}^+ \\ (q, q') &\mapsto K(q'|q) \end{aligned} \tag{103}$$

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

$$a(q', q) = \min\left(1, \frac{K(q|q')\pi(q')}{K(q'|q)\pi(q)}\right) \tag{104}$$

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 $1 - a(q', q)$

La distribution des transitions de Markov peut être définie comme :

$$T(q'|q) = a(q', q) \cdot Q(q'|q) + (1 - a(q', q)) \cdot \delta(q - q') \tag{105}$$

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

0.5.7 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

0.6 Equations différentielles stochastiques

0.6.1 Définition

Un **processus stochastique** est une collection de variables aléatoires $\{X_t\}_{t \geq 0}$ définies sur un espace probabilisé commun où t représente le temps. Le processus stochastique décrit l'évolution d'un système soumis à l'aléatoire.

Une **équation différentielle stochastique** est la généralisation de la notion d'équation différentielle en prenant en compte un terme de bruit blanc.

0.6.2 De l'équation différentielle ordinaire (EDO) à l'équation différentielle stochastique (EDS)

On considère une équation différentielle ordinaire $\frac{dX_t}{dt} = \mu(X(t))$ à laquelle nous aimeraions rajouter du bruit. Nous considérons alors un **bruit blanc** $\xi(t)$ dépendant du temps et de l'espace.

Motivations : on souhaiterait définir l'équation d'évolution

$$\frac{dX_t}{dt} = \mu(X(t)) + \sigma(X(t))\xi(t) \quad (106)$$

Remarque : on aurait envie que le processus évolue par petits sauts gaussiens, de moyenne $\mu(t)$ et de variance $\sigma(t)$.

→ Encore faut-il qu'une différentiation par rapport à un bruit blanc $\xi(t)dt$ ait un sens !

On définit alors un mouvement brownien W_t , où W_t est un taux de transition, ou une densité de probabilité de transition. Le mouvement brownien définit alors l'équation différentielle suivante :

$$dX(t) = \mu(X(t))dt + \sigma((X(t)))dW_t \quad (107)$$

Et on cherche à donner un sens à l'intégration suivante :

$$X(t) = X(0) + \int_0^t \mu(X(s))ds + \int_0^t \sigma(X(s))dW(s) \quad (108)$$

0.6.3 Processus de premier ordre

Considérons une équation différentielle stochastique (EDS) du premier ordre :

$$dX_t = a(X_t, t)dt + b(X_t, t)dW_t \quad (109)$$

où :

- X_t est le processus stochastique inconnu,
- $a(X_t, t)$ est le terme de dérive
- $b(X_t, t)$ est le terme de diffusion
- W_t est le mouvement brownien standard

0.6.4 Formule d'Itô

La **formule d'Itô** est une formule fondamentale du calcul stochastique. Elle permet de calculer la différentielle df d'une fonction $f(X_t, t)$ où X_t suit une équation différentielle stochastique.

Remarque : la formule d'Itô nous sert de base pour passer d'une étude déterministe du système représenté par ses états $\{X_t\}_{t \geq 0}$ à une étude stochastique de ce même système (par exemple : ajout d'une force aléatoire ou d'un bruit de mesure).

Enoncé : Soit $\{X_t\}_{t \geq 0}$ un processus stochastique qui satisfait :

$$dX_t = a(X_t, t)dt + b(X_t, t)dW_t \quad (110)$$

Alors, pour une fonction $f(X_t, t) \in \mathcal{C}^2$ deux fois continûment différentiable, sa différentielle s'exprime :

$$df = \left(\frac{\partial f}{\partial t} + a(X_t, t) \frac{\partial f}{\partial x} + \frac{1}{2} b^2(X_t, t) \frac{\partial^2 f}{\partial x^2} \right) dt + b(X_t, t) \frac{\partial f}{\partial x} dW_t \quad (111)$$

Exemple :

Soit $X_t = W_t$ (le mouvement brownien standard) et $f(X_t) = X_t^2$. Calculons df :

$$\frac{\partial f}{\partial x} = 2X_t, \frac{\partial^2 f}{\partial x^2} = 2, \frac{\partial f}{\partial t} = 0 \quad (112)$$

Comme $dX_t = dW_t$, avec $a = 0, b = 0$, la formule d'Itô donne :

$$df = (0 + 0 \cdot 2X_t + \frac{1}{2} \cdot 1^2 \cdot 2)dt + 1 \cdot 2X_t dW_t = dt + 2X_t dW_t \quad (113)$$

Ce résultat montre que :

$$d(W_t^2) = dt + 2W_t dW_t \quad (114)$$

Remarque : Il existe une autre interprétation des EDS, par l'intégrale de Stratanovitch, qui permettrait l'équivalent d'une intégration trapézoïdale, là où Itô propose une intégration cohérente numériquement avec Euler explicite.

0.6.5 Intégration numérique - Méthode d'Euler-Maruyama

La **méthode d'Euler-Maruyama** permet d'approximer numériquement les solutions des EDS. En discrétilisant selon un pas de temps Δt , on obtient :

$$X_{n+1} = X_n + a(X_n, t_n) \Delta t + b(X_n, t_n) \Delta W_n \quad (115)$$

où $\Delta W_n = W_{t_{n+1}} - W_{t_n} \sim \mathcal{N}(0, \Delta t)$.

Résolution numérique d'équations différentielles ordinaires

Numériquement, quand on cherche à résoudre $\frac{dX}{dt} = \mathcal{M}(X)$ en se donnant un pas de temps d'intégration Δt , il existe 2 méthodes d'ordre 1 en temps.

Euler Explicite (Euler forward)

$$X_{n+1} = X_n + \Delta t \mathcal{M}(X_n) + \mathcal{O}(\Delta t) \quad (116)$$

Euler Implicit (Euler backward)

$$X_{n+1} = X_n + \Delta t \mathcal{M}(X_{n+1}) + \mathcal{O}(\Delta t) \quad (117)$$

En quelque sorte, on pourrait dire que la méthode d'Euler-Maruyama est une “méthode d'Euler explicite stochastique”.

0.6.6 Convergence de la Méthode Euler-Maruyama

En considérant l'EDS linéaire simple :

$$dX_t = \mu dt + \sigma dW_t, X_0 = 0 \quad (118)$$

Résolution analytique En intégrant les 2 membres de $t = 0$ à $t = t_n$:

$$X_{t_n} - X_0 = \int_0^{t_n} \mu dt + \int_0^{t_n} \sigma dW_t \quad (119)$$

Avec $X_0 = 0$, et $W_0 = 0$, donc :

$$X_{t_n} = \mu t_n + \sigma W_{t_n} \quad (120)$$

Résolution numérique Par la méthode d'Euler-Maruyama :

En discrétilisant selon un pas de temps Δt ,

$$X_{n+1} = X_n + \mu\Delta t + \sigma\Delta W_n \quad (121)$$

où $\Delta W_n = W_{t_{n+1}} - W_{t_n} \sim \mathcal{N}(0, \Delta t)$.

Remarque : de la manière où nous avons défini le mouvement brownien, la variance du processus croît de Δt à chaque pas de temps.

Calcul de l'erreur En posant :

$$e_n = X_{t_n} - X_n \quad (122)$$

Et en utilisant d'une part l'expression analytique, d'autre part l'expression numérique, on obtient :

$$\begin{aligned} X_{t_n} &= \mu t_n + \sigma W_{t_n} \\ X_n &= X_0 + \sum_{k=0}^{n-1} (\mu\Delta t + \sigma\Delta W_k) = \mu n\Delta t + \sigma \sum_{k=0}^{n-1} \Delta W_k \end{aligned} \quad (123)$$

et

$$e_n = \mu t_n - \mu n\Delta t + \sigma (W_{t_n} - \sum_{k=0}^{n-1} \Delta W_k) \quad (124)$$

Simplification de l'erreur

Comme $t_n = n\Delta t$, le terme déterministe s'annule :

$$\mu t_n - \mu n\Delta t = 0 \quad (125)$$

Il reste :

$$e_n = \sigma (W_{t_n} - \sum_{k=0}^{n-1} \Delta W_k) \quad (126)$$

Donc :

$$\sum_{k=0}^{n-1} \Delta W_k = W_{t_n} - \Delta W_n \quad (127)$$

L'erreur s'écrit :

$$e_n = \sigma (W_{t_n} - (W_{t_n} - \Delta W_n)) = \sigma \Delta W_n \quad (128)$$

Erreur Quadratique moyenne Comme $\Delta W_n \sim \mathcal{N}(0, \Delta t)$, on a :

$$\mathbb{E}[e_n^2] = \sigma^2 [(\Delta W_n)^2] = \sigma^2 \Delta t \quad (129)$$

Conclusion L'erreur quadratique moyenne étant proportionnelle à Δt , cela montre que l'erreur tend vers zéro quand $\Delta t \rightarrow 0$. La méthode d'Euler-Maruyama converge vers la solution exacte de cette EDS.

0.6.7 Applications

Les équations différentielles stochastiques ont de nombreuses applications en ingénierie, météorologie, biologie, finance.

Mouvement brownien géométrique : C'est un processus stochastique à valeurs positives. Il est utilisé en mathématiques financières pour modéliser un cours de bourse. La dynamique de ce processus est la base du modèle de Black-Scholes.

$$dX_t = \mu X_t dt + \sigma X_t dW_t \quad (130)$$

Remarque : l'équation du premier ordre $dX_t = \mu dt + \sigma dW_t$ avait déjà été utilisée par Louis Bachelier (1900) pour établir un modèle de cours de bourse.

Equation de Langevin : Elle modélise le mouvement d'une particule dans un "bain thermique", soumise à une force $\xi(t)$ fluctuante (qui pourrait être un bruit blanc). Dans ce cas, la différentielle de la vitesse associée au principe fondamentale de la dynamique est :

$$dV(t) = -\lambda V(t)dt + dW_t \quad (131)$$

0.6.8 Références

Gratton, S., Sequential Filtering, (2024-2025)

Baehr, C., Modélisations Probabilistes d'Equations de la Mécanique des Fluides, Etat de l'art en modélisation stochastique (2004)

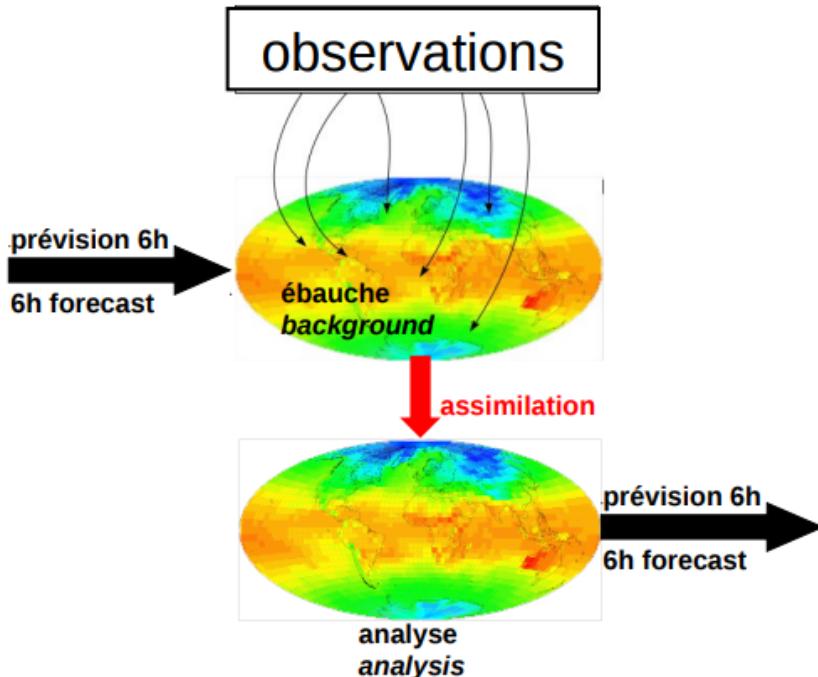
Smith L., Itô and Stratanovitch, A guide for the perplexed

0.7 Filtre Kalman d'Ensemble

Le filtre Kalman d'Ensemble (EnKF) utilise des méthodes de Monte-Carlo pour évaluer la covariance de l'erreur d'analyse, sur un Filtre de Kalman Evensen [1994].

0.7.1 Assimilation de données

Nous évoluons dans le cadre des méthodes d'assimilation de données pour la Météo.



0.7.2 Ensemble Kalman Filter (EnKF)

Statistique de l'erreur

Représentation de la statistique d'erreur

Nous reprenons les notations introduites par Evensen [2009].

Pour le filtre de Kalman (**KF**) usuel, et pour un champ X_t les matrices de covariance sont exprimées par rapport à "l'état vrai" du système :

$$\begin{aligned} P^f &= \overline{(X_t^f - X_t^t)(X_t^f - X_t^t)^T} \\ P^a &= \overline{(X_t^a - X_t^t)(X_t^a - X_t^t)^T} \end{aligned} \quad (132)$$

La limitation apparaît immédiatement, dans la mesure où "l'état vrai" du système n'est jamais connu. Le Filtre Kalman d'Ensemble (**EnKF**) propose donc de représenter la covariance de l'erreur en référence à une moyenne d'ensemble $\overline{X_t}$. On a alors :

$$\begin{aligned} P^f &\approx P_c^f = \overline{(X_t^f - \overline{X}_t)(X_t^f - \overline{X}_t)^T} \\ P^a &\approx P_c^a = \overline{(X_t^a - \overline{X}_t)(X_t^a - \overline{X}_t)^T} \end{aligned} \quad (133)$$

Echantillonage par Méthodes de Monte-Carlo

Ici, dans le cadre du Filtre Kalman d'Ensemble, la covariance de l'erreur est représentée en utilisant un ensemble d'états du systèmes, échantillonné par méthodes de Monte-Carlo.

Ainsi, en se référant au Théorème Central Limite pour les estimateurs de Monte-Carlo, l'erreur liée à un ensemble de taille N décroît proportionnellement à $\frac{1}{\sqrt{N}}$.

Par extension, en passant à la limite, on peut imaginer un échantillonage de l'espace suffisant, de telle manière que le champ X_t soit décrit par une densité de probabilité limite :

$$\phi(X_t) = \frac{dN}{N} \quad (134)$$

avec dN la densité de points par unité de volume, et N le nombre de points échantillonnés.

Remarque : La covariance de l'erreur pourrait être représentée en théorie par la densité de probabilité limite formée par une infinité de points.

Vers une équation différentielle stochastique

En reprenant la dynamique du système associée au filtre de Kalman, le filtre Kalman d'Ensemble propose une résolution "stochastique" de cette dynamique par méthodes de Monte-Carlo. L'objectif des méthodes de Monte-Carlo étant d'éviter les calculs lourds de matrice de covariance d'une part, et d'éviter l'emploi d'une fermeture analytique des équations d'évolution de la covariance de l'erreur (*ou tout simplement d'éviter une modélisation rigide de cette covariance*).

En associant la dynamique du système d'une part, et son échantillonnage d'autre part, nous pouvons introduire une équation différentielle stochastique pour modéliser l'évolution du système :

$$dX_t = f(X_t)dt + g(X_t)dW_t \quad (135)$$

où, comme précédemment, W_t est un mouvement brownien. Cette équation différentielle pouvant intégrée au sens d'**Itô** et résolue par des méthodes numériques telles que la méthode d'**Euler-Maruyama**.

De Markov Chain Monte-Carlo à l'EnKF

Bon ... Encore un peu de théorie ! Mais en observant de plus près, on arrive finalement à associer les éléments principaux du cours, à savoir le filtre de Kalman et les méthodes Markov Chain Monte Carlo.

Equation de Kolmogorov ou de Fokker-Planck

L'équation de Fokker-Planck régit les transitions d'un processus de Markov. Il s'agit de l'équation de diffusion pour une densité de probabilité lorsque l'espace et le temps sont continus.

Son expression est la suivante, en 1 dimension :

$$\frac{\partial}{\partial t}P(x,t) = -c\frac{\partial}{\partial x}P(x,t) + D\frac{\partial^2}{\partial x^2}P(x,t) \quad (136)$$

où c est une vitesse de transport et D un coefficient de diffusion.

En dimensions supérieures :

$$\frac{\partial\phi}{\partial x} + \sum_i \frac{\partial(f_i\phi)}{\partial x_i} = \frac{1}{2} \sum_{i,j} \frac{\partial^2\phi(gQg^T)_{ij}}{\partial x_i \partial x_j} \quad (137)$$

Lorsque nous utilisons un modèle additif d'erreurs gaussiennes, et que celles-ci forment un processus de Markov, c'est l'équation de Fokker-Planck (aussi appelée équation de Kolmogorov) qui régit l'évolution de la densité de probabilité $\phi(X_t)$ pour l'état du modèle.

Formellement, l'**EnKF** utilise un estimateur **Markov Chain Monte-Carlo** pour résoudre l'équation de **Fokker-Planck**.

Remarque : nous avions vu l'équation maîtresse, pour les phénomènes continus en temps mais discrets en espace. De manière analogue, l'équation de Fokker-Planck décrit l'évolution d'une densité de probabilité pour les phénomènes continus en temps et en espace. Elle s'attache à décrire l'évolution de cette densité selon un transport et une diffusion.

Analogie avec l'équation de la chaleur

L'équation de Fokker-Planck nous permet ici de décrire fondamentalement l'évolution de la statistique de l'erreur. En considérant un flux de probabilités (*analogie avec l'équation de la chaleur*), l'équation décrit un changement de densité de probabilité sur un volume local qui dépend d'un flux de probabilité sur ce volume (issu de la dynamique) et d'un terme de diffusion (qui tend à diffuser ou aplatiser la densité de probabilité, dans la mesure où l'erreur statistique croît).

Remarque : l'équation qui régit formellement les transitions d'un processus de Markov sera nommée "équation de Fokker-Planck" par les physiciens, et "équation de Kolmogorov" pour les mathématiciens.

Monte-Carlo ou Markov Chain Monte-Carlo ?

Monte-Carlo ou Markov Chain Monte-Carlo permettent de construire 2 estimateurs pour résoudre les mêmes équations.

Ici, on pourrait résoudre l'équation de Fokker-Planck des 2 manières :

- En utilisant un estimateur de Monte-Carlo : il s'agit des méthodes de filtrage particulaire (**Particle Filter**).
- En faisant appel à un estimateur Markov Chain Monte-Carlo : on obtient alors le filtre Kalman d'ensemble (**Ensemble Kalman Filter**).

Cela permet d'introduire 2 méthodes d'assimilation de données : le **filtrage particulier** et le **filtre kalman d'ensemble**. Nous exposons ici le filtre kalman d'ensemble en pratique.

A retenir

On souhaitait utiliser les méthodes stochastiques (MC et MCMC) pour simuler et étudier l'évolution de l'erreur sur un système chaotique (par exemple : un modèle Météo). En construisant le raisonnement, nous venons d'établir que nous souhaitons résoudre numériquement l'équation de Kolmogorov, ce qui nous permettrait d'avoir accès à une densité de probabilité de l'état du système.

Bon ... Toutes ces maths pour tomber sur un filtre à particule, ça pue un peu comme un vieux diesel ! Passons à une implémentation pratique de ces méthodes.

0.7.3 Formulation de l'EnKF pour l'analyse

Analyse

En météo, on parle d'analyse pour les champs de conditions initiales produits par l'étape d'assimilation de données.

Définissons un ensemble d'observations de départ

$$d_j = d + \epsilon_i, j \in \{1, \dots, N\} \quad (138)$$

Cet ensemble peut être défini ou généré en perturbant des observations., de telle manière que les **erreurs de mesure** $\epsilon_{j_1 \leq j \leq N}$ ont une moyenne nulle. La matrice de covariance de l'erreur de mesure est définie comme :

$$R_e = \overline{\epsilon \epsilon^T} \quad (139)$$

Remarque : On notera que, dans un ensemble de taille infinie, R converge vers la matrice R définie pour le filtre Kalman.

Le filtre Kalman entre alors en jeu avec une étape de prédition et une étape de correction.

Prédiction : $\hat{X}_j = X_j^f$

Correction : $X_j^a = \hat{X}_j + P_e^f H^T (H P_e^f H^T + R_e)^{-1} (d_j + H X_j^f)$

où X^f est un champ issu de la prévision, H est un opérateur d'observation, R_e est la covariance de l'erreur de mesure et P_e^f est la matrice de covariance de prévision du filtre de Kalman.

En utilisant l'expression du gain du filtre,

$$K_e = P_e^f H^T (H P_e^f H^T + R_e)^{-1} \quad (140)$$

Et la covariance issue de la minimisation de la variance pour l'ajustement du filtre Kalman classique s'exprime comme :

$$P_e^a = (I - K_e H) P_e^f \quad (141)$$

Remarque : ici, l'EnKF se comporte comme un filtre Kalman dans la limite d'un ensemble de taille infinie.

Synthèse

Ici la matrice de covariance liée au bruit d'état (et à la dynamique) s'exprime comme

$$Q_e = \overline{dW_k dW_k^T} \quad (142)$$

où dW_k représente le tirage issu du mouvement brownien $dW_t \sim \mathcal{N}(0, dt)$. Ici, Q_e converge également vers Q (matrice de covariance du bruit d'état pour le KF) à la limite d'un ensemble infini.

Les propriétés remarquables sont que l'estimateur de la moyenne évolue selon la dynamique du système :

$$\overline{\bar{X}_{k+1}} = \overline{f(\bar{X}_k)} = f(\overline{\bar{X}_k}) + (\text{non-linear terms}) \quad (143)$$

La matrice de covariance d'erreur évolue selon :

$$P_e^{k+1} = F P_e^k F^T + Q_e + (\text{non-linear terms}) \quad (144)$$

où F est l'opérateur tangent linéaire du modèle.

Le Filtre Kalman d'Ensemble se comporte comme un Filtre de Kalman dans le cas d'un ensemble infini. **Dans le cas d'une dynamique non-linéaire**, les résidus non-linéaires sont pris en compte implicitement par l'EnKF, tandis qu'un Filtre de Kalman Etendu (EKF) néglige ces termes.

En pratique, la formulation du filtre de Kalman d'Ensemble ne fait pas intervenir d'opérateur Tangent-Linéaire, ce qui le rend très facile à implémenter.

0.8 Assimilation de données

0.8.1 Description formelle

On note $x_{1:L} = (x_1, , x_L) \in \mathbb{R}^{LxD}$ une trajectoire d'états (ou L et D sont respectivement la longueur de la trajectoire et la dimension du vecteur d'état). On note $p(x_{i+1}|x_i)$ la dynamique de transition de l'état x_i vers l'état x_{i+1} .

L'observation $y \in \mathbb{R}^M$ de la trajectoire $x_{1:L}$ suit un processus d'observation $p(y|x_{1:L})$, formulé comme $y = \mathcal{A}(x_{1:L}) + \eta$. \mathcal{A} est une fonction (opérateur) d'observation et $\eta \in \mathbb{R}^M$ une erreur d'observation (ou de mesure).

Par des techniques d'Assimilation de Données, nous cherchons à inférer les trajectoires plausibles $x_{1:L}$ étant donné l'observation y , c'est à dire, d'estimer la trajectoire à posteriori :

$$p(x_{1:L}|y) = \frac{p(y|x_{1:L})}{p(y)} p(x_1) \prod_{i=1}^{L-1} p(x_{i+1}|x_i) \quad (145)$$

Liens avec l'estimation bayésienne et les chaînes de Markov

Le problème ci-dessus est effectivement formulé comme un problème d'estimation bayésienne où :

- $p(x_{1:L}|y)$ représente la vraisemblance : (*likelihood*) en anglais,
- $p(y|x_{1:L})$ la distribution empirique des données d'observations,
- $p(y)$ la loi marginale des données,
- $p(x_1) \prod_{i=1}^{L-1} p(x_{i+1}|x_i)$ le modèle,

La trajectoire est effectivement représentée par une chaîne de Markov, avec :

- une distribution initiale, appelée **background** en météo : $p(x_1)$
- une fonction de transition (la dynamique discrète) : $p(x_{i+1}|x_i)$

0.9 Modèles de diffusion pour l'assimilation de données

0.9.1 Diffusion Models for Data Assimilation

L'utilisation de Modèles de Diffusion pour les systèmes d'assimilation de donnée est un champ de recherche émergeant, avec des applications prometteuses. L'assimilation variationnelle et les filtres Kalman d'Ensemble constituent le fondement des méthodes d'assimilation (4DEnVar), développées dans les années 1990 et améliorées continuellement depuis pour un usage dans les systèmes opérationnels.

Fondamentalement, ces systèmes reposent sur la résolution numérique de l'équation de Fokker-Planck, l'équation différentielle stochastique de diffusion.

De manière analogue, les réseaux de neurones génératifs peuvent reconstruire les bruits sur ces modèles, à partir d'un générateur stochastique. Les références suivantes donnent un aperçu de ces méthodes : Rozet and Louppe [2023], Huang et al. [2024], Hodyss and Morzfeld [2025].

Bibliography

- G. Evensen. Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162, 5 1994. ISSN 0148-0227. doi: 10.1029/94jc00572. URL <http://dx.doi.org/10.1029/94JC00572>.
- G. Evensen. The ensemble Kalman filter for combined state and parameter estimation. *IEEE Control Systems*, 29(3):83–104, 6 2009. ISSN 1941-000X. doi: 10.1109/mcs.2009.932223. URL <http://dx.doi.org/10.1109/MCS.2009.932223>.
- D. Hodyss and M. Morzfeld. Using Diffusion Models to do Data Assimilation, 2025. URL <https://arxiv.org/abs/2506.02249>.
- L. Huang, L. Gianinazzi, Y. Yu, P. D. Dueben, and T. Hoefer. Diffda: a Diffusion Model for Weather-scale Data Assimilation, 2024. URL <https://arxiv.org/abs/2401.05932>.
- F. Rozet and G. Louppe. Score-based Data Assimilation, 2023. URL <https://arxiv.org/abs/2306.10574>.
- K. Sedlmeier, S. Mieruch, G. Schädler, and C. Kottmeier. Compound extremes in a changing climate – a Markov chain approach. *Nonlinear Processes in Geophysics*, 23(6):375–390, 11 2016. ISSN 1607-7946. doi: 10.5194/npg-23-375-2016. URL <http://dx.doi.org/10.5194/npg-23-375-2016>.