

Notes sur le livre Feynman-Kac Formulae

Version de travail compilée le 24 février 2006

Table des matières

1	Quelques notations fondamentales	1
1.1	Modèle de Feynman-Kac	1
1.2	Dynamique	2
1.3	Structure de semi-groupe	3
1.4	Interprétations en termes de processus de Markov non linéaire	5
1.5	Le système de particules	6
2	Estimations uniformes en temps	7
2.1	Transformation du terme clé	7
2.2	Lemme clé de type Burkholder	9
2.3	Le résultat de convergence	10
3	Propagation du chaos	10
3.1	Introduction	10
3.2	Propagation faible	11
3.3	Propagation du chaos forte	12
3.3.1	Nouvelle formulation de la variation totale	12
3.3.2	Un peu d'entropie relative	13
3.3.3	Inégalité de Csiszár-Kullback	14
3.3.4	Encore une belle inégalité de Csiszár	14
3.3.5	Encore une hypothèse de régularité	15
3.3.6	Densité de la loi du système de particules	16
3.3.7	Le résultat de convergence	17
4	Stabilité des semi-groupes de Feynman-Kac, mesures invariantes	19
4.1	Mesurer l'écart entre des mesures	19
4.2	Les noyaux markoviens contractent	20
4.3	Une première majoration	21
4.4	Encore un peu de semi-groupes	23
4.5	Majoration des coefficients de contraction	25
4.6	Mesure invariante	26
5	Algorithmes de type Feynman-Kac-Metropolis-Hastings	27
5.1	Le modèle de Metropolis-Hastings	28
5.1.1	Le cas général	28
5.1.2	Le cas réversible et Gibbs	28
5.2	Les modèles de Feynman-Kac-Metropolis	29
5.2.1	Quelques notations	29
5.2.2	Description de la dynamique	29
5.3	Comparaison heuristique	30
5.4	Système de particules associé	30

6	Recuit simulé par Feynman-Kac-Metropolis	31
6.1	Le contexte	31
6.2	Contraction du noyau	32
6.2.1	Renversement du temps	32
6.2.2	Propriété de contraction	33
6.3	Convergence du recuit simulé	34
6.3.1	Propriétés de régularité	35
6.3.2	Comportement asymptotique	36
7	Applications	38
7.1	Filtrage et modèle de Markov caché	38
7.2	Chaîne de Markov contrainte ou conditionnelle	40
7.3	Simulation d'événements rares	40

1 Quelques notations fondamentales

1.1 Modèle de Feynman-Kac

Soit $((E_n, \mathcal{E}_n))_n$ une suite d'espaces mesurables. On note $\mathcal{P}(E_n)$ l'ensemble des mesures de probabilité sur E_n et $\mathcal{B}_b(E_n)$ l'ensemble des applications mesurables bornées de E_n dans \mathbb{R} .

Pour tout n , on dispose d'un noyau de Markov $M_n(x_{n-1}, dx_n)$ de E_{n-1} dans $\mathcal{P}(E_n)$ et on se donne également $\mu \in \mathcal{P}(E_0)$. On dispose ainsi d'une chaîne de Markov inhomogène $(X_n)_n$ telle que $X_n \in E_n$, de loi initiale μ et de transitions $(M_n)_n$. La loi de la chaîne jusqu'au temps n est donnée par

$$\mathbb{P}_{\mu,n}(d(x_0, \dots, x_n)) = \mu(dx_0)M_1(x_0, dx_1) \cdots M_n(x_{n-1}, dx_n).$$

Soit, pour tout n , $G_n : E_n \rightarrow]0, 1]$ une fonction mesurable bornée¹.

Définition 1.1 (Mesure de Feynman-Kac) *Les modèles de prédiction $(\mathbb{Q}_n)_n$ et de correction $(\widehat{\mathbb{Q}}_n)_n$ associés au couple (G, M) sont les suites de mesures à valeurs dans les espaces des trajectoires définies par*

$$\begin{aligned} \forall n, \quad \mathbb{Q}_{\mu,n}(d(x_0, \dots, x_n)) &:= \frac{1}{Z_n} \left(\prod_{p=0}^{n-1} G_p(x_p) \right) \mathbb{P}_{\mu,n}(d(x_0, \dots, x_n)), \\ \forall n, \quad \widehat{\mathbb{Q}}_{\mu,n}(d(x_0, \dots, x_n)) &:= \frac{1}{\widehat{Z}_n} \left(\prod_{p=0}^n G_p(x_p) \right) \mathbb{P}_{\mu,n}(d(x_0, \dots, x_n)). \end{aligned}$$

Les constantes de normalisation

$$Z_n := \mathbb{E}_\mu \left(\prod_{p=0}^{n-1} G_p(X_p) \right) \quad \text{et} \quad \widehat{Z}_n := \mathbb{E}_\mu \left(\prod_{p=0}^n G_p(X_p) \right)$$

sont aussi appelée fonctions de partition.

Définition 1.2 *Les modèles de prédiction non normalisé $(\gamma_n)_n$ et normalisé $(\eta_n)_n$ sont les suites de mesures (positives pour le premier et de probabilité pour le second) telles que, pour toute fonction*

¹On peut autoriser G_n à prendre des valeurs supérieures à 1 sans grandes difficultés et à s'annuler (ce qui demande plus d'attention).

$f_n \in \mathcal{B}_b(E_n)$,

$$\gamma_n(f_n) := \mathbb{E}_\mu \left(f_n(X_n) \prod_{p=0}^{n-1} G_p(X_p) \right) \quad \text{et} \quad \eta_n(f_n) := \frac{\gamma_n(f_n)}{\gamma_n(1)}.$$

On définit de même les modèles de correction non normalisé $(\hat{\gamma}_n)_n$ et normalisé $(\hat{\eta}_n)_n$ par

$$\hat{\gamma}_n(f_n) := \mathbb{E}_\mu \left(f_n(X_n) \prod_{p=0}^n G_p(X_p) \right) \quad \text{et} \quad \hat{\eta}_n(f_n) := \frac{\hat{\gamma}_n(f_n)}{\hat{\gamma}_n(1)}.$$

Remarque 1.3 On peut exprimer la suite $(\gamma_n)_n$ en fonction de $(\eta_n)_n$ en remarquant que $\gamma_n(1)$ est égal à $\gamma_{n-1}(G_{n-1})$:

$$\begin{aligned} \gamma_n(f_n) &= \eta_n(f_n) \gamma_n(1) = \eta_n(f_n) \gamma_{n-1}(G_{n-1}) \\ &= \eta_n(f_n) \eta_{n-1}(G_{n-1}) \gamma_{n-1}(1) = \cdots = \eta_n(f_n) \prod_{p=0}^{n-1} \eta_p(G_p). \end{aligned}$$

On obtient encore les relations

$$\hat{\eta}_n(f_n) = \frac{\eta_n(f_n G_n)}{\eta_n(G_n)} \quad \text{et} \quad \hat{\gamma}_n(f_n) = \hat{\eta}_n(f_n) \prod_{p=0}^n \eta_p(G_p). \quad (1)$$

Dans les applications, on s'intéresse souvent plus à γ_n qu'à η_n et plus aux processus de correction qu'à ceux de prédiction. Cependant la remarque précédente montre que si l'on sait décrire le comportement de la suite $(\eta_n)_n$ alors on sera en mesure de faire de même pour les autres quantités. Tout le travail consiste donc à étudier les propriétés de la suite $(\eta_n)_n \dots$

1.2 Dynamique

L'objet de cette section est de mettre en évidence les relations de récurrence qui permettent d'exprimer en particulier la dynamique des suites $(\gamma_n)_n$ et $(\eta_n)_n$.

La suite $(\gamma_n)_n$ vérifie la relation de récurrence suivante :

$$\gamma_n = \gamma_{n-1} Q_n \quad (2)$$

avec, pour $x_{n-1} \in E_{n-1}$ et $f_n \in \mathcal{B}_b(E_n)$,

$$Q_n(f_n)(x_{n-1}) := G_{n-1}(x_{n-1}) M_n(f_n)(x_{n-1}),$$

ou encore pour $\mu_{n-1} \in \mathcal{P}(E_{n-1})$ et $f_n \in \mathcal{B}_b(E_n)$,

$$\mu_{n-1}(Q_n(f_n)) = \mu_{n-1}(G_{n-1} M_n(f_n)).$$

En effet, grâce à la propriété de Markov, on peut écrire

$$\begin{aligned} \gamma_n(f_n) &= \mathbb{E}_\mu \left(f_n(X_n) \prod_{p=0}^{n-1} G_p(X_p) \right) = \mathbb{E}_\mu \left(\mathbb{E}_\mu \left(f_n(X_n) \middle| X_0, \dots, X_{n-1} \right) \prod_{p=0}^{n-1} G_p(X_p) \right) \\ &= \mathbb{E}_\mu \left(\mathbb{E}_\mu \left(f_n(X_n) \middle| X_{n-1} \right) \prod_{p=0}^{n-1} G_p(X_p) \right) = \mathbb{E}_\mu \left(M_n(f_n)(X_{n-1}) \prod_{p=0}^{n-1} G_p(X_p) \right) \\ &= \mathbb{E}_\mu \left(G_{n-1}(X_{n-1}) M_n(f_n)(X_{n-1}) \prod_{p=0}^{n-2} G_p(X_p) \right) = \gamma_{n-1}(G_{n-1} M_n(f_n)). \end{aligned}$$

La relation (2) est linéaire mais $\gamma_n(E_n) = \gamma_{n-1}(G_{n-1}) \leq 1$: il y a perte de masse. Pour la suite $(\eta_n)_n$ les données sont inversées : la masse est constante mais la récurrence n'est plus linéaire comme on va le voir un petit peu plus loin. Voici l'expression de la récurrence pour $(\eta_n)_n$.

Proposition 1.4 *Les suites de mesures $(\eta_n)_n$ et $(\hat{\eta}_n)_n$ vérifient toutes les deux une relation de récurrence*

$$\eta_n = \Phi_n(\eta_{n-1}) \quad \text{et} \quad \hat{\eta}_n = \hat{\Phi}_n(\hat{\eta}_{n-1})$$

où $\Phi_n, \hat{\Phi}_n : \mathcal{P}(E_{n-1}) \rightarrow \mathcal{P}(E_n)$ sont des fonctionnelles non linéaires qui se décomposent en deux parties :

$$\Phi_n(\eta) := \Psi_{n-1}(\eta)M_n \quad \text{avec} \quad \Psi_n(\eta)(dx) := \frac{1}{\eta(G_n)}G_n(x)\eta(dx)$$

et

$$\hat{\Phi}_n(\eta) := \Psi_n(\eta M_n) = \hat{\Psi}_{n-1}(\eta)\hat{M}_n \quad \text{avec} \quad \hat{\Psi}_n(\eta)(dx) := \frac{1}{\eta(\hat{G}_n)}\hat{G}_n(x)\eta(dx)$$

où le couple (\hat{G}_n, \hat{M}_n) est donné par

$$\hat{G}_n := M_{n+1}(G_{n+1}) \quad \text{et} \quad \hat{M}_n(f_n) := \frac{M_n(f_n G_n)}{M_n(G_n)}.$$

Remarque 1.5 *Parfois, on pourra utiliser la notation bien pratique car compacte suivante :*

$$G_n \cdot \eta_n(dx) = \frac{G_n(x)}{\eta_n(G_n)} \eta_n(dx)$$

désigne la mesure de probabilité absolument continue par rapport à η_n de densité G_n (à la normalisation près). C'est encore la mesure que nous avons appelé $\Psi_n(\eta_n)$.

Que ce soit pour les mesures de prédiction ou de correction, la récurrence se fait donc en deux pas :

$$\eta_n \in \mathcal{P}(E_n) \xrightarrow[G_n]{\text{correction}} \hat{\eta}_n \in \mathcal{P}(E_n) \xrightarrow[M_{n+1}]{\text{prédiction}} \eta_{n+1} \in \mathcal{P}(E_{n+1})$$

et

$$\hat{\eta}_n \in \mathcal{P}(E_n) \xrightarrow[M_{n+1}]{\text{prédiction}} \eta_{n+1} \in \mathcal{P}(E_{n+1}) \xrightarrow[G_{n+1}]{\text{correction}} \hat{\eta}_{n+1} \in \mathcal{P}(E_{n+1})$$

Remarque 1.6 *On peut résumer l'action de Φ_n en remarquant que pour $\mu \in \mathcal{P}(E_{n-1})$ et $f \in \mathcal{B}_b(E_n)$, on a*

$$\Phi_n(\mu)(f) = \frac{\mu(G_{n-1}M_n(f))}{\mu(G_{n-1})} = \frac{\mu(G_{n-1}M_n(f))}{\mu(G_{n-1}M_n(1))} = \frac{\mu(Q_n f)}{\mu(Q_n(1))}. \quad (3)$$

1.3 Structure de semi-groupe

L'idée de cette section est de remarquer que les composées des noyaux $(Q_n)_n$ et $(\Phi_n)_n$ forment des semi-groupes (facile et pas extraordinaire) qui possèdent les mêmes types d'expression (très remarquable).

Rappelons que l'on note Q_n l'opérateur qui à $\mu \in \mathcal{P}(E_{n-1})$ et $f_n \in \mathcal{B}_b(E_n)$ associe

$$\mu(Q_n(f_n)) = \mu(G_{n-1}M_n f_n) = \int_{E_{n-1}} G_{n-1}(x)M_n f_n(x) \mu(dx).$$

Définition 1.7 *Soit $(Q_{q,n})_{0 \leq q \leq n}$ le semi-groupe défini par*

$$Q_{n,n} =: Id \quad \text{et} \quad Q_{q,n} := Q_{q+1} \cdots Q_{n-1}Q_n \quad \text{pour } q < n.$$

Cette définition peut s'exprimer en termes de chaîne de Markov :

$$Q_{q,n}(f_n)(x_q) = \mathbb{E}_{q,\delta_{x_q}} \left(f_n(X_n) \prod_{p=q}^{n-1} G_p(X_p) \right)$$

Remarquons que, par définition, $Q_{n-1,n} = Q_n$. La propriété de semi-groupe, c'est-à-dire le fait que, pour $q \leq p \leq n$, $Q_{q,p}Q_{p,n} = Q_{q,n}$ est évidente. En appliquant plusieurs fois (2), on obtient

$$\gamma_n(f_n) = \gamma_{n-1}(Q_n(f_n)) = \gamma_{n-2}(Q_{n-1}(Q_n(f_n))) = \dots = \gamma_q(Q_{q,n}(f_n)).$$

On a alors immédiatement les relations

$$\eta_n(f_n) = \frac{\gamma_n(f_n)}{\gamma_n(1)} = \frac{\gamma_q(Q_{q,n}(f_n))}{\gamma_q(Q_{q,n}(1))} = \frac{\eta_q(Q_{q,n}(f_n))}{\eta_q(Q_{q,n}(1))}.$$

Ceci peut être reformulé en termes d'espérance :

$$\eta_n(f_n) = \frac{\mathbb{E}_{q,\eta_q} \left(f_n(X_n) \prod_{p=q}^{n-1} G_p(X_p) \right)}{\mathbb{E}_{q,\eta_q} \left(\prod_{p=q}^{n-1} G_p(X_p) \right)},$$

où \mathbb{E}_{q,η_q} désigne l'espérance pour la chaîne de Markov dont la loi au temps q est η_q .

Ainsi, il apparaît que η_n est l'image de η_q par une transformation analogue à celle qui fait passer de η_{n-1} à η_n .

Définition 1.8 Soit $(\Phi_{q,n})_{0 \leq q \leq n}$ le semi-groupe défini par

$$\Phi_{n,n} =: Id \quad \text{et} \quad \Phi_{q,n} := \Phi_n \circ \Phi_{n-1} \circ \dots \circ \Phi_{q+1}.$$

Encore une fois, $\Phi_{n-1,n} = \Phi_n$. Le fait remarquable ici est que $\Phi_{q,n}$ envoie $\mathcal{P}(E_q)$ dans $\mathcal{P}(E_n)$ d'une manière analogue à Φ_n comme le montre le résultat suivant.

Lemme 1.9 L'application $\Phi_{q,n}$ envoie $\mathcal{P}(E_q)$ dans $\mathcal{P}(E_n)$ de la façon suivante : pour $\mu_q \in \mathcal{P}(E_q)$ et $f_n \in \mathcal{B}_b(E_n)$, on a

$$\Phi_{q,n}(\mu_q)(f_n) = \frac{\mu_q(Q_{q,n}(f_n))}{\mu_q(Q_{q,n}(1))} = \frac{\mathbb{E}_{q,\mu_q} \left(f(X_n) \prod_{p=q}^{n-1} G_p(X_p) \right)}{\mathbb{E}_{q,\mu_q} \left(\prod_{p=q}^{n-1} G_p(X_p) \right)},$$

où \mathbb{E}_{q,μ_q} désigne l'espérance pour la chaîne de Markov dont la loi au temps q est μ_q .

Preuve. Montrons ceci par récurrence sur n :

$$\begin{aligned} \Phi_{q,n+1}(\mu_q)(f_{n+1}) &= \Phi_{n+1}(\Phi_{q,n}(\mu_q))(f_{n+1}) \stackrel{(3)}{=} \frac{\Phi_{q,n}(\mu_q)(Q_{n+1}(f_{n+1}))}{\Phi_{q,n}(\mu_q)(Q_{n+1}(1))} \\ &\stackrel{\text{Hyp. Rec.}}{=} \frac{\mu_q(Q_{q,n}(Q_{n+1}(f_{n+1})))}{\mu_q(Q_{q,n}(Q_{n+1}(1)))} = \frac{\mu_q(Q_{q,n+1}(f_{n+1}))}{\mu_q(Q_{q,n+1}(1))}, \end{aligned}$$

et la propriété est clairement vraie pour $n = q$. □

Dans la suite il sera utile d'écrire $\Phi_{q,n}$ sous la forme suivante

$$\Phi_{q,n}(\mu_q)(f_n) = \frac{\mu_q(G_{q,n}P_{q,n}(f_n))}{\mu_q(G_{q,n})} \quad \text{avec} \quad G_{q,n} := Q_{q,n}(1) \quad \text{et} \quad P_{q,n}(f_n) := \frac{Q_{q,n}(f_n)}{Q_{q,n}(1)}. \quad (4)$$

Remarque 1.10 *Encore une fois, $Q_{q,n}$ agit linéairement (à gauche) sur les mesures alors que ce n'est pas le cas de $\Phi_{q,n}$.*

Les hypothèses permettant d'obtenir les résultats clé comme la propriété de contraction, la convergence à l'équilibre (dans le cas homogène) ou la propagation du chaos portent sur les familles $(G_{q,n})_{q \leq n}$ et $(P_{q,n})_{q \leq n}$. On supposera dans la suite que les constantes suivantes sont finies :

$$r_{q,n} = \sup_{x_q, y_q \in E_q} (G_{q,n}(x_q)/G_{q,n}(y_q)) \quad \text{et} \quad \beta(P_{q,n}) = \sup_{x_q, y_q \in E_q} \|P_{q,n}(x_q, \cdot) - P_{q,n}(y_q, \cdot)\|_{VT}. \quad (5)$$

1.4 Interprétations en termes de processus de Markov non linéaire

La suite $(\eta_n)_n$ est une suite de mesures de probabilité définies par récurrence (à un pas). On aimerait donc l'interpréter comme la suite des lois marginales d'un processus de type markovien. Cependant, la récurrence est non linéaire : le processus sera donc un peu biscornu. De manière générale, on cherche à réécrire la relation de récurrence de la manière suivante :

$$\eta_{n+1} = \Phi_{n+1}(\eta_n) = \eta_n K_{n+1, \eta_n} \quad (6)$$

où $K_{n+1, \eta}$ est une application de $\mathcal{P}(E_n)$ dans $\mathcal{P}(E_{n+1})$. Il n'y a pas unicité de choix pour ces noyaux². Imposons tout de même à $K_{n+1, \eta}$ d'être de la forme

$$K_{n+1, \eta} = S_{n, \eta} M_{n+1}$$

où $S_{n, \eta}$ est un noyau markovien de E_n dans lui-même. La condition imposée est que pour toute mesure $\eta \in \mathcal{P}(E_n)$, $\eta S_{n, \eta} = \Psi_n(\eta)$. Il n'y a toujours pas unicité. Par exemple, les noyaux suivants conviennent

$$S_{n, \eta}(x_n, dy_n) = \varepsilon_n G_n(x_n) \delta_{x_n}(dy_n) + (1 - \varepsilon_n G_n(x_n)) \Psi_n(\eta)(dy_n),$$

à condition que le réel ε_n soit choisi pour que $\varepsilon_n G_n$ soit inférieur à 1 (on peut prendre par exemple $\varepsilon_n = 0$)³.

Vérifions que l'on a bien, pour tout $\eta \in \mathcal{P}(E_n)$, $\eta S_{n, \eta} = \Psi_n(\eta)$. Soit $f \in \mathcal{B}_b(E_n)$. Alors

$$S_{n, \eta}(f)(x) = \varepsilon_n G_n(x) f(x) + (1 - \varepsilon_n G_n(x)) \Psi_n(\eta)(f).$$

Intégrons cette relation par rapport à η :

$$\eta(S_{n, \eta}(f)) = \varepsilon_n \eta(G_n f) + \Psi_n(\eta)(f) - \varepsilon_n \eta(G_n) \Psi_n(\eta)(f) = \Psi_n(\eta)(f),$$

puisque $\Psi_n(\eta)(f) = \eta(G_n f) / \eta(G_n)$.

On retrouve le schéma à deux pas

$$\eta_n \xrightarrow{S_{n, \eta_n}} \hat{\eta}_n = \eta_n S_{n, \eta_n} \xrightarrow{M_{n+1}} \eta_{n+1} = \hat{\eta}_n M_{n+1}.$$

Essayons d'interpréter la mesure η_n comme la loi au temps n d'un processus de type markovien et de décrire sa dynamique.

$$X_n \xrightarrow{\text{saut autoguidé}} \hat{X}_n \xrightarrow{\text{exploration}} X_{n+1}.$$

- **Saut autoguidé** : étant donné la position et la loi de la particule X_n au temps n , elle saute sur un nouveau site \hat{X}_n choisi avec la mesure

$$\mathcal{L}(\hat{X}_n | X_n) = S_{n, \eta_n}(X_n, \cdot) = \varepsilon_n G_n(X_n) \delta_{X_n} + (1 - \varepsilon_n G_n(X_n)) \Psi_n(\eta_n).$$

²Déjà dans le cas linéaire, il est clair que si l'on se donne une suite de probabilités, il existe un bon nombre de chaînes de Markov inhomogènes admettant ces marginales.

³Le rôle de ε_n est crucial lorsque G_n est borné mais pas par 1.

En d'autres termes, la particule reste sur place avec probabilité $\varepsilon_n G_n(X_n)$ et l'on pose $\widehat{X}_n = X_n$ (ce qu'elle tendance à faire d'autant plus qu'elle dans une région à fort potentiel). Sinon, elle saute sur un nouveau site selon la loi

$$\Psi_n(\eta_n)(dx_n) = \frac{1}{\eta_n(G_n)} G_n(x_n) \eta_n(dx_n).$$

Elle aura ainsi tendance à rejoindre une zone à fort potentiel parmi les zones où elle était susceptible de se trouver avant de sauter. On voit ici la non linéarité du processus : la façon de sauter dépend de l'endroit où l'on se trouve mais dépend également (et surtout) de toute la loi au temps de saut.

– **Exploration** : la particule \widehat{X}_n rejoint un nouveau site choisi selon la mesure

$$\mathcal{L}(X_{n+1}|\widehat{X}_n) = M_{n+1}(\widehat{X}_n, \cdot).$$

Si l'on choisit $\varepsilon_n = 0$, alors $\mathcal{L}(\widehat{X}_n|X_n)$ ne dépend pas de X_n : \widehat{X}_n est choisi selon la loi $\widehat{\eta}_n$. Si $\varepsilon_n = 0$ pour tout n alors les variables aléatoires $(X_n)_n$ sont indépendantes de lois respectives $(\eta_n)_n$.

Examinons le cas particulier où l'on choisit $G_n = 1$ pour tout n . La suite $(\eta_n)_n$ est tout simplement donnée par $\eta_{n+1} = \eta_n M_{n+1}$: η_n est la loi de la chaîne de Markov inhomogène de loi initiale η_0 et de transitions $(M_n)_n$. Dans le modèle $\varepsilon_n = 0$ pour tout n , on a $S_{n,\eta_n}(x_n, \cdot) = \eta_n$ et $K_{n,\eta_n}(x_n, \cdot) = \eta_{n+1}$. La loi de (X_0, \dots, X_n) est donnée par

$$\mathbb{K}_{\eta_0, n} = \eta_0(dx_0)\eta_1(dx_1) \cdots \eta_n(dx_n).$$

Si à présent $\varepsilon_n = 1$, la partie saut disparaît, on a $S_{n,\eta}(x_n, \cdot) = \delta_{x_n}$ et $K_{n,\eta_n}(x_n, \cdot) = \eta_n M_{n+1}(x_n, \cdot)$. La loi de (X_0, \dots, X_n) est donnée par

$$\mathbb{K}_{\eta_0, n} = \eta_0(dx_0)M_1(x_0, dx_1) \cdots M_n(x_{n-1}, dx_n).$$

C'est la loi de la chaîne de Markov. Cette interprétation contient donc plus de structure et est plus riche en termes d'interprétation probabiliste que le premier.

1.5 Le système de particules

On souhaite manipuler des objets plus simples et plus classiques que ce processus non linéaire dont la dynamique nécessite à chaque instant la connaissance de sa loi. On va linéariser le problème. Le prix à payer sera que le nouvel objet sera de grande dimension.

Le système de particules associé aux transitions $(K_{n,\eta_n})_{\eta_n \in \mathcal{P}(E_n), n}$ est une chaîne de Markov inhomogène

$$\left(\xi_n^{(N)} \right)_n = \left(\left(\xi_n^{(1,N)}, \dots, \xi_n^{(N,N)} \right) \right)_n$$

qui au temps n prends ses valeurs dans l'espace produit E_n^N . On oubliera dans la suite l'indice $(\cdot)^{(N)}$: ξ_n désigne le vecteur des N particules au temps n et ξ_n^p est la position de la particule p parmi N au temps n . La loi initiale est la mesure produit $\eta_0^{\otimes N}$. Les transitions de E_n^N vers E_{n+1}^N sont données par

$$\mathbb{P}_{\eta_0}^N(\xi_{n+1} \in dx_{n+1} | \xi_n) := \prod_{p=1}^N K_{n+1, m(\xi_n)}(\xi_n^p, dx_{n+1}^p),$$

où $dx_{n+1} = dx_{n+1}^1 \times \cdots \times dx_{n+1}^N$ est un voisinage infinitésimal de $x_{n+1} = (x_{n+1}^1, \dots, x_{n+1}^N) \in E_{n+1}^N$ et l'on a utilisé l'abus de notation :

$$m(\xi_n) := \frac{1}{N} \sum_{p=1}^N \delta_{\xi_n^p} = \eta_n^N.$$

La dynamique du système de particules imite celle du processus non linéaire. Chaque transition se décompose en deux étapes élémentaires. Pour le voir, il suffit d'utiliser la décomposition de $K_{n,\eta}$:

$$\mathbb{P}_{\eta_0}^N(\xi_{n+1} \in dx_{n+1} | \xi_n) = \prod_{p=1}^N (S_{n,m(\xi_n)} M_{n+1})(\xi_{n-1}^p, dx_n^p).$$

On peut ainsi mettre en évidence deux étapes, la sélection et la mutation :

$$\xi_n \xrightarrow{\text{sélection}} \widehat{\xi}_n \xrightarrow{\text{mutation}} \xi_{n+1}.$$

- **Sélection** : étant donné la configuration $\xi_n \in E_n$ chacune des N particules choisit sa nouvelle position indépendamment des autres selon la loi

$$\mathcal{L}(\widehat{\xi}_n^p | \xi_n) = S_{n,m(\eta_n)}(\xi_n^p, \cdot) = \varepsilon_n G_n(\xi_n^p) \delta_{\xi_n^p} + (1 - \varepsilon_n G_n(X_n)) \Psi_n(m(\xi_n)).$$

En d'autres termes, la particule reste sur place avec probabilité $\varepsilon_n G_n(\xi_n^p)$ et l'on pose $\widehat{\xi}_n^p = \xi_n^p$. Sinon, elle saute sur un nouveau site selon la loi

$$\Psi_n(m(\xi_n))(dx_n) = \sum_{i=1}^N \frac{G_n(\xi_n^i)}{\sum_{j=1}^N G_n(\xi_n^j)} \delta_{\xi_n^i}(dx_n).$$

La particule ne peut aller que sur un site déjà occupé et elle aura tendance à rejoindre les sites situés dans un fort potentiel.

- **Mutation** : chaque particule $\widehat{\xi}_n^p$ rejoint, indépendamment des autres, un nouveau site choisi selon la mesure

$$\mathcal{L}(\xi_{n+1}^p | \widehat{\xi}_n^p) = M_{n+1}(\widehat{\xi}_n^p, \cdot).$$

Comme on peut le voir d'après la forme produit de la définition de la transition du système de particules, conditionnellement à ξ_n , les positions (ξ_{n+1}^p) sont indépendantes. Si de plus, on a $\varepsilon_n = 0$ alors elles sont également identiquement distribuées de loi $\Phi_n(m(\xi_n))$. Dans ce cas particulier on parle d'*algorithmes génétiques*.

Dans tous les cas, pour toute fonction $f_{n+1} \in \mathcal{B}_b(E_{n+1})$,

$$\begin{aligned} \mathbb{E}(\eta_{n+1}^N(f_{n+1}) | \xi_n) &= \frac{1}{N} \sum_{i=1}^N \mathbb{E}(f_{n+1}(\xi_{n+1}^i) | \xi_n) = \frac{1}{N} \sum_{i=1}^N (K_{n+1, \eta_n^N} f_{n+1})(\xi_n^i) \\ &= \eta_n^N(K_{n+1, \eta_n^N}(f_{n+1})) = \Phi_{n+1}(\eta_n^N)(f_{n+1}). \end{aligned} \quad (7)$$

2 Estimations uniformes en temps

On souhaite montrer que η_n^N converge vers η_n lorsque N le nombre de particules tend vers $+\infty$. Pour cela on utilise la décomposition suivante, avec la convention $\Phi_{-1}(\eta_{-1}^N) = \eta_0$,

$$\eta_n^N - \eta_n = \sum_{q=0}^n [\Phi_{q,n}(\eta_q^N) - \Phi_{q,n}(\Phi_q(\eta_{q-1}^N))]$$

puisque $\Phi_{q-1,n} = \Phi_{q,n} \circ \Phi_q$. Chacun des termes devrait être petit avec N . En effet, d'après la définition du système de particules, on a vu que $\mathbb{E}(\eta_q^N | \xi_{q-1}) = \Phi_q(\eta_{q-1}^N)$ et on espère un résultat de type loi des grands nombres. On espère ensuite contrôler comment ce faible écart au départ se propagera le long du semi-groupe $(\Phi_{q,n})_q$. Sous de bonnes hypothèses de stabilité, un semi-groupe peut oublier exponentiellement vite les conditions initiales. Chaque terme de la somme devrait donc pouvoir être majoré par une quantité de la forme $c(q)d(N)$ avec $c(q)$ le terme général d'une série convergente et $d(N)$ qui tend vers 0 quand N tend vers l'infini. C'est effectivement ce que l'on peut établir en précisant que, comme dans la loi des grands nombres, $d(N) = \sqrt{N}$.

2.1 Transformation du terme clé

Le but de cette section est de montrer que l'on peut écrire le terme à contrôler

$$\Phi_{q,n}(\eta_q^N) - \Phi_{q,n}(\Phi_q(\eta_{q-1}^N))$$

en fonction de la différence des deux mesures η_q^N et $\Phi_q(\eta_{q-1}^N)$ appliquées à une certaine fonction. Ceci n'est pas donné gratuitement dans la mesure où $\Phi_{q,n}$ n'agit pas linéairement sur les mesures (voir (4)).

Définition 2.1 On introduit trois nouvelles quantités $G_{q,n}^N$, $P_{q,n}^N$ et $Q_{q,n}^N$.

– $G_{q,n}^N : E_q \rightarrow \mathbb{R}^+$ défini, pour $x_q \in E_q$, par

$$G_{q,n}^N(x_q) := \frac{G_{q,n}(x_q)}{\Phi_q(\eta_{q-1}^N)(G_{q,n})}. \quad (8)$$

– $P_{q,n}^N : \mathcal{B}_b(E_n) \rightarrow \mathcal{B}_b(E_q)$ défini, pour $f_n \in \mathcal{B}_b(E_n)$ et $x_q \in E_q$, par

$$P_{q,n}^N(f_n)(x_q) := P_{q,n}(f_n)(x_q) - \Phi_{q,n}(\Phi_q(\eta_{q-1}^N))(f_n)(x_q). \quad (9)$$

– $Q_{q,n}^N : \mathcal{B}_b(E_n) \rightarrow \mathcal{B}_b(E_q)$ défini, pour $f_n \in \mathcal{B}_b(E_n)$ et $x_q \in E_q$, par

$$Q_{q,n}^N(f_n)(x_q) := G_{q,n}^N(x_q)P_{q,n}^N(f_n)(x_q). \quad (10)$$

Remarquons que l'on peut encore réécrire $P_{q,n}^N$ de plusieurs façons (utiles dans la suite) :

$$\begin{aligned} P_{q,n}^N(f_n)(x_q) &= P_{q,n}(f_n)(x_q) - \Phi_{q,n}(\Phi_q(\eta_{q-1}^N))(f_n)(x_q) \\ &\stackrel{(4)+(8)}{=} P_{q,n}(f_n)(x_q) - \Phi_q(\eta_{q-1}^N)(G_{q,n}^N P_{q,n}(f_n)) \end{aligned} \quad (11)$$

$$= \int (P_{q,n}f_n(x_q) - P_{q,n}f_n(y_q))G_{q,n}^N(y_q)\Phi_q(\eta_{q-1}^N)(dy_q). \quad (12)$$

Avec les notations ci-dessous, on peut écrire $\Phi_{q,n}(\eta_q^N)(f_n) - \Phi_{q,n}(\Phi_q(\eta_{q-1}^N))(f_n)$ de la façon suivante :

$$\begin{aligned} \Phi_{q,n}(\eta_q^N)[f_n - \Phi_{q,n}(\Phi_q(\eta_{q-1}^N))(f_n)] &\stackrel{(4)}{=} \frac{1}{\eta_q^N(G_{q,n})}\eta_q^N [G_{q,n}P_{q,n}(f_n - \Phi_{q,n}(\Phi_q(\eta_{q-1}^N))(f_n))] \\ &\stackrel{(9)}{=} \frac{1}{\eta_q^N(G_{q,n})}\eta_q^N [G_{q,n}P_{q,n}^N(f_n)] \\ &\stackrel{(8)}{=} \frac{1}{\eta_q^N(G_{q,n}^N)}\eta_q^N [G_{q,n}^N P_{q,n}^N(f_n)] \\ &\stackrel{(10)}{=} \frac{1}{\eta_q^N(G_{q,n}^N)}\eta_q^N [Q_{q,n}^N(f_n)]. \end{aligned}$$

Remarque 2.2 Par la définition (8) de $G_{q,n}^N$, on a

$$\Phi_q(\eta_{q-1}^N)(G_{q,n}^N) = \Phi_q(\eta_{q-1}^N)\left(\frac{G_{q,n}}{\Phi_q(\eta_{q-1}^N)(G_{q,n})}\right) = 1.$$

De plus,

$$\Phi_q(\eta_{q-1}^N)(Q_{q,n}^N f_n) \stackrel{(10)}{=} \Phi_q(\eta_{q-1}^N)(G_{q,n}^N P_{q,n}^N f_n) \stackrel{(11)}{=} 0. \quad (13)$$

On peut ainsi écrire

$$\Phi_{q,n}(\eta_q^N)(f_n) - \Phi_{q,n}(\Phi_q(\eta_{q-1}^N))(f_n) \stackrel{(13)}{=} \frac{1}{\eta_q^N(G_{q,n}^N)} [\eta_q^N - \Phi_q(\eta_{q-1}^N)] (Q_{q,n}^N f_n).$$

Avec (12), pour toute fonction $f_n \in \mathcal{B}_b(E_n)$, on utilise les constantes définies en (5) pour obtenir les deux majorations suivantes :

$$\begin{aligned} |P_{q,n}^N f_n(x_q)| &\leq \int |P_{q,n} f_n(x_q) - P_{q,n} f_n(y_q)| G_{q,n}^N(y_q) \Phi_q(\eta_{q-1}^N)(dy_q) \\ &\leq \text{osc}(P_{q,n} f_n) \leq \beta(P_{q,n}). \end{aligned}$$

Ceci assure alors

$$\frac{\|Q_{q,n}^N f_n\|}{\eta_q^N(G_{q,n}^N)} \leq \frac{\|G_{q,n}^N\|}{\eta_q^N(G_{q,n}^N)} \|P_{q,n}^N f_n\| \leq r_{q,n} \beta(P_{q,n}).$$

On en déduit le contrôle désiré

$$|\Phi_{q,n}(\eta_q^N)(f_n) - \Phi_{q,n}(\Phi_q(\eta_{q-1}^N))(f_n)| \leq r_{q,n} \beta(P_{q,n}) \left| (\eta_q^N - \Phi_q(\eta_{q-1}^N)) \left(\overline{Q}_{q,n}^N f_n \right) \right|,$$

où $\overline{Q}_{q,n}^N f_n := Q_{q,n}^N f_n / \|Q_{q,n}^N f_n\|$ est une fonction bornée par 1.

La première partie du contrat est remplie : on voit ici l'effet du semi-groupe. Sous de bonnes hypothèses sur $r_{q,n}$ et $\beta(P_{q,n})$, on obtiendra des contrôles uniformes en temps. Reste à contrôler l'erreur sur un pas de temps due à l'approximation particulière. Pour cela, il faut un résultat de type loi des grands nombres quantitatif.

2.2 Lemme clé de type Burkholder

Lemme 2.3 (Lemme 7.3.3 p. 223) *On se donne une suite de lois $(\mu_i)_i$, des fonctions $(h_i)_i$ bornées telles que $\mu_i(h_i) = 0$. Soit X_1, \dots, X_N indépendantes de lois respectives μ_1, \dots, μ_N . Alors, pour tout $p \geq 1$,*

$$\sqrt{N} [\mathbb{E}(|m(X)(h)|^p)]^{1/p} \leq d(p)^{1/p} \sigma(h)$$

où

$$m(X)(h) := \frac{1}{N} \sum_{i=1}^N h_i(X_i) \quad \text{et} \quad \sigma^2(h) := \frac{1}{N} \sum_{i=1}^N \text{osc}(h_i)^2.$$

La constante $d(p)$ est explicite :

$$d(2n) = (2n)_n 2^{-n} \quad \text{et} \quad d(2n-1) = \frac{(2n-1)_n}{\sqrt{n-1/2}} 2^{-(n-1/2)} \quad \text{avec} \quad (n)_p = \frac{n!}{(n-p)!}.$$

Remarque 2.4 *Le cas $p = 2$ est facile à établir : soit X' une copie indépendante de X . On a alors*

$$\begin{aligned} \mathbb{E}(|m(X)(h)|^2) &= \frac{1}{2} \mathbb{E}(|m(X)(h) - m(X')(h)|^2) \\ &= \frac{1}{2N^2} \sum_{i=1}^N \mathbb{E}[(h_i(X_i) - h_i(X'_i))^2] \leq \frac{1}{2N} \sigma^2(h). \end{aligned}$$

Plus généralement, pour $p \geq 1$,

$$\mathbb{E}(|m(X)(h)|^p) \leq \mathbb{E}(|m(X)(h) - m(X')(h)|^p)$$

Remarque 2.5 *Le contrôle précis de $d(p)$ permettra d'obtenir des contrôles efficaces de transformations de Laplace (pour faire des grandes déviations) par simple développement en séries entières.*

Remarque 2.6 *On pourrait obtenir le même type de résultat en considérant des fonctions bornées plutôt que des fonctions centrées à oscillation finie. Reste à lancer une grande consultation nationale sur le sujet...*

2.3 Le résultat de convergence

L'idée est d'appliquer le lemme 2.3 après avoir conditionné par \mathcal{F}_{q-1}^N . Pour cela, il faut insister sur le fait que conditionnellement à ξ_{q-1} , les particules ξ_q^1, \dots, ξ_q^N sont indépendantes mais pas tout à fait identiquement distribuées (en général) car elles ne partent pas du même point mais passent ensuite dans la même moulinette. En particulier on a, d'après (7), pour toute fonction $f_q \in \mathcal{B}_b(E_q)$,

$$\mathbb{E}(\eta_q^N(f_q) | \xi_{q-1}^N) = \Phi_q(\eta_{q-1}^N)(f_q).$$

Posons pour tout $i = 1, \dots, N$,

$$X_i = \xi_q^i, \quad \mu_i = \mathcal{L}(\xi_q^i | \xi_{q-1}) \quad \text{et} \quad h_i(x) = \frac{1}{N} \overline{Q}_{q,n}^N f_n(x) - \frac{1}{N} \mathbb{E}(\overline{Q}_{q,n}^N f_n(\xi_q^i) | \xi_{q-1}).$$

On a bien $\mu_i(h_i) = 0$ pour tout $i = 1, \dots, N$ et

$$(\eta_q^N - \Phi_q(\eta_{q-1}^N))(\overline{Q}_{q,n}^N f_n) = \frac{1}{N} \sum_{i=1}^N \overline{Q}_{q,n}^N f_n(\xi_q^i) - \frac{1}{N} \sum_{i=1}^N \mathbb{E}(\overline{Q}_{q,n}^N f_n(\xi_q^i) | \xi_{q-1}) = m(X)(h).$$

Il ne reste plus qu'à écrire

$$\mathbb{E}\left(\left|(\eta_q^N - \Phi_q(\eta_{q-1}^N))(\overline{Q}_{q,n}^N f_n)\right|^p\right) \leq \mathbb{E}\left(\mathbb{E}\left(\left|(\eta_q^N - \Phi_q(\eta_{q-1}^N))(\overline{Q}_{q,n}^N f_n)\right|^p \middle| \xi_{q-1}\right)^p\right).$$

Puisque $\overline{Q}_{q,n}^N f_n$ est bornée par 1, son oscillation est inférieure à 2. D'après le lemme 2.3,

$$\sqrt{N} \left[\mathbb{E}\left(\left|(\eta_q^N - \Phi_q(\eta_{q-1}^N))\overline{Q}_{q,n}^N f_n\right|^p\right) \right]^{1/p} \leq 2d(p)^{1/p} r_{q,n} \beta(P_{q,n}).$$

Théorème 2.7 (Th 7.4.4. p. 246) *Pour tous $n \geq 0$, $p \geq 1$ et $f_n \in \text{Osc}_1(E_n)$,*

$$\sqrt{N} \left[\mathbb{E}\left(\left|(\eta_n^N f_n - \eta_n f_n)^p\right|\right) \right]^{1/p} \leq 2d(p)^{1/p} \sum_{q=0}^n r_{q,n} \beta(P_{q,n}).$$

Sous de bonnes conditions de contraction on obtient un résultat uniforme en temps : il existe une constante C telle que

$$\sup_{n \geq 0} \sup_{f_n \in \text{Osc}_1(E_n)} \sqrt{N} \left[\mathbb{E}\left(\left|(\eta_n^N f_n - \eta_n f_n)^p\right|\right) \right]^{1/p} \leq C.$$

3 Propagation du chaos

3.1 Introduction

On montre ici la chose suivante : la loi des q particules fixées parmi N converge quand N tend vers l'infini vers une loi produit dont toutes les marges sont égales à la loi du processus non linéaire sous-jacent. Il faut comprendre chaos au sens d'indépendance et propagation au sens où l'indépendance des distributions initiales se propage sur à tout temps dans le système infini. Nous allons en effet montrer que cette indépendance asymptotique est vraie non seulement pour la loi à un instant mais aussi au sens des lois sur les trajectoires de longueur n . La démonstration repose essentiellement sur le théorème 2.7 et sur une formulation équivalente de la propagation du chaos. Puisque les lois des systèmes de particules sont échangeables (*i.e.* invariantes par permutation des coordonnées) la propagation du chaos est équivalente à la convergence en loi de la mesure empirique du système de particule vers la loi du processus non linéaire sous-jacent.

On note

– la loi du processus non linéaire sur l'intervalle de temps $[0, n]$:

$$\mathbb{K}_{\eta_0, n}(d(x_0, \dots, x_n)) := \eta_0(dx_0)K_{1, \eta_0}(dx_1) \cdots K_{n, \eta_{n-1}}(dx_n) \in \mathcal{P}(E_{[0, n]}),$$

– la loi du système de particules sur l'intervalle $[0, n]$:

$$\mathbb{P}_{\eta_0, n}^N \in \mathcal{P}(E_{[0, n]}^N),$$

– la loi des q premières particules du système de taille N sur l'intervalle $[0, n]$:

$$\mathbb{P}_{\eta_0, n}^{(N, q)} \in \mathcal{P}(E_{[0, n]}^q),$$

– la mesure empirique⁴ du système de particules sur le même intervalle :

$$\mathbb{K}_{\eta_0, n}^N := \frac{1}{N} \sum_{i=1}^N \delta_{\xi_{[0, n]}^i} \in \mathcal{P}(E_{[0, n]}),$$

– $\Pi_n \subset \mathcal{C}_b(E_{[0, n]})$ le sous-espace des fonctions cylindriques :

$$F_n = f_0 \otimes f_1 \otimes \cdots \otimes f_n \quad \text{avec} \quad f_0 \in \mathcal{C}_b(E_0), \dots, f_n \in \mathcal{C}_b(E_n) \quad \text{et} \quad \forall_{p=0}^n \|f_p\| \leq 1.$$

Remarque 3.1 Par définition, on a, pour $q \leq N$,

$$\mathbb{P}_{\eta_0, n}^{(N, q)}(F_n^1 \otimes \cdots \otimes F_n^q) := \mathbb{E}_{\eta_0}^N \left(\prod_{i=1}^q F_n^i(\xi_{[0, n]}^i) \right) = \mathbb{E}_{\eta_0}^N \left(\prod_{i=1}^q \prod_{j=1}^n f_j^i(\xi_j^i) \right).$$

Théorème 3.2 La suite de mesures $(\mathbb{P}_{\eta_0, n}^N)_N$ est $\mathbb{K}_{\eta_0, n}$ -(faiblement) chaotique⁵, c'est-à-dire que

$$\lim_{N \rightarrow \infty} \mathbb{E}_{\eta_0}^N \left(\prod_{i=1}^q F_n^i(\xi_{[0, n]}^i) \right) = \prod_{i=1}^q \mathbb{K}_{\eta_0, n}(F_n^i),$$

si et seulement si la suite des mesures empiriques $(\mathbb{K}_{\eta_0, n}^N)_N$ converge en loi vers $\mathbb{K}_{\eta_0, n}$.

Remarque 3.3 On parle de propagation du chaos faible car c'est un résultat en espérance. On peut aussi montrer des résultats du type

$$\lim_{N \rightarrow \infty} \left\| \mathbb{P}_{\eta_0, n}^{(N, q)} - (\mathbb{K}_{\eta_0, n})^{\otimes q} \right\|_{vt} = 0.$$

On parlera alors de propagation du chaos forte.

3.2 Propagation faible

Théorème 3.4 Pour tout $n \in \mathbb{N}$ et tout $p \geq 1$, on a

$$\sup_{F_n \in \Pi_n} \mathbb{E}_{\eta_0}^N \left(\left| \mathbb{K}_{\eta_0, n}^N(F_n) - \mathbb{K}_{\eta_0, n}(F_n) \right|^p \right)^{1/p} \leq \frac{a(p)b(n)}{\sqrt{N}}.$$

⁴Attention : c'est une variable aléatoire à valeurs dans $\mathcal{P}(E_{[0, n]})$.

⁵La loi $\mathbb{P}_{\eta_0, n}^{(N, q)}$ des q premières particules parmi N converge vers la mesure produit $(\mathbb{K}_{\eta_0, n})^{\otimes q}$.

Preuve. On procède par récurrence en écrivant⁶ :

$$\mathbb{K}_{\eta_0,n}^N(F_n) - \mathbb{K}_{\eta_0,n}(F_n) = \mathbb{K}_{\eta_0,n-1}^N(F_{n-1,n}) - \mathbb{K}_{\eta_0,n-1}(F_{n-1,n}) + \mathbb{E}_{\eta_0}^N(I_1) + \mathbb{E}_{\eta_0}^N(I_2)$$

où l'on a posé

$$\begin{aligned} F_{n-1,n} &= f_0 \otimes f_1 \otimes \cdots \otimes f_{n-1} K_{n,\eta_{n-1}}(f_n) \in \mathcal{C}_b(E_{[0,n-1]}) \\ I_1 &:= \frac{1}{N} \sum_{i=1}^N \left[\prod_{p=0}^{n-1} f_p(\xi_p^i) \right] \left[f_n(\xi_n^i) - K_{n,\eta_{n-1}}^N(f_n)(\xi_{n-1}^i) \right] \\ I_2 &:= \frac{1}{N} \sum_{i=1}^N \left[\prod_{p=0}^{n-1} f_p(\xi_p^i) \right] \left[K_{n,\eta_{n-1}}^N(f_n)(\xi_{n-1}^i) - K_{n,\eta_{n-1}}(f_n)(\xi_{n-1}^i) \right]. \end{aligned}$$

Conditionnellement à \mathcal{F}_{n-1}^N les variables aléatoires

$$\left[\prod_{p=0}^{n-1} f_p(\xi_p^i) \right] \left[f_n(\xi_n^i) - K_{n,\eta_{n-1}}^N(f_n)(\xi_{n-1}^i) \right], \quad i = 1, \dots, N$$

sont indépendantes centrées comme on l'a déjà remarqué pour établir le théorème 2.7. D'après le lemme 2.3, on a le contrôle $\sqrt{N} \mathbb{E}_{\eta_0}^N(|I_1|^p)^{1/p} \leq a(p)$. Pour estimer I_2 , on revient à la définition de $K_{n,\eta}$ (voir (6)) pour écrire

$$\left(K_{n,\eta_{n-1}}^N - K_{n,\eta_{n-1}} \right)(f_n) = (1 - \varepsilon_{n-1} G_{n-1})(\Phi_n(\eta_{n-1}^N) - \Phi_n(\eta_{n-1}))(f_n).$$

Encore d'après le travail à un temps, on obtient

$$\sqrt{N} \mathbb{E}_{\eta_0}^N \left(\left\| K_{n,\eta_{n-1}}^N(f_n) - K_{n,\eta_{n-1}}(f_n) \right\|^p \right)^{1/p} \leq a(p)b(n),$$

ce qui fournit le contrôle $\sqrt{N} \mathbb{E}_{\eta_0}^N(|I_2|^p)^{1/p} \leq a(p)b(n)$. En posant

$$J_n = \sup_{F_n \in \Pi_n} \mathbb{E}_{\eta_0}^N \left(\left| \mathbb{K}_{\eta_0,n}^N(F_n) - \mathbb{K}_{\eta_0,n}(F_n) \right|^p \right)^{1/p},$$

on a montré que $J_n \leq a(p)C(n) + J_{n-1}$, ce qui suffit à conclure. \square

3.3 Propagation du chaos forte

3.3.1 Nouvelle formulation de la variation totale

Rappelons que l'on définit la distance en variation totale entre μ et ν par

$$\|\mu - \nu\|_{vt} = \frac{1}{2} \sup \left\{ \int g d\nu - \int g d\mu, \quad g \text{ bornée par } 1 \right\}.$$

Proposition 3.5 *Si ν admet f pour densité par rapport à μ alors*

$$\|\nu - \mu\|_{vt} = \frac{1}{2} \int |f - 1| d\mu.$$

Preuve. Pour toute fonction g bornée par 1, on a

$$\int g d\nu - \int g d\mu = \int g(f - 1) d\mu \leq \int |f - 1| d\mu,$$

l'égalité étant atteinte pour $g = \text{sgn}(f - 1)$. \square

⁶On a le schéma suivant : $\square - \circ = (\times - \circ) + (\square - \Delta) + (\Delta - \times)$

3.3.2 Un peu d'entropie relative

Définition 3.6 Soit μ et ν deux mesures de probabilités sur E . On définit l'entropie relative de ν par rapport à μ par

$$\text{Ent}(\nu|\mu) = \begin{cases} \int \frac{d\nu}{d\mu} \ln \frac{d\nu}{d\mu} d\mu = \int \ln \frac{d\nu}{d\mu} d\nu & \text{si } \nu \ll \mu, \\ +\infty & \text{sinon.} \end{cases}$$

Soit f μ -intégrable de E dans \mathbb{R}_+ , on définit son entropie par rapport à μ par

$$\text{Ent}_\mu(f) = \int f \ln f d\mu - \int f d\mu \ln \int f d\mu.$$

Remarque 3.7 Si ν admet f pour densité par rapport à μ , on a

$$\text{Ent}(\nu|\mu) = \text{Ent}_\mu(f).$$

Proposition 3.8 (Formule variationnelle de l'entropie) Soit $\mu \in \mathcal{P}(E)$ et f positive d'intégrale 1 par rapport à μ . Alors

$$\text{Ent}_\mu(f) = \sup \left\{ \int fg d\mu, \int e^g d\mu = 1 \right\} = \sup \{ \mu(fg) - \ln \mu(e^g), g \text{ bornée} \}.$$

Preuve. Démontrons la première égalité. D'après l'inégalité de Jensen appliquée à \ln et la mesure de probabilité de densité f par rapport à ν , on a, pour toute fonction g telle que e^g soit d'intégrale 1 pour μ ,

$$\int fg d\mu - \int f \ln f d\mu = \int \ln \frac{e^g}{f} f d\mu \leq \ln \int e^g d\mu = 0.$$

Réciproquement, on montre que l'on peut choisir g arbitrairement proche de $\ln f$. La seconde égalité se déduit par homothétie. \square

Remarque 3.9 La proposition ci-dessus peut encore s'écrire de la façon suivante : soit ν absolument continue par rapport à μ alors

$$\text{Ent}(\nu|\mu) = \sup \{ \nu(g) - \ln \mu(e^g), g \text{ bornée} \}.$$

Lemme 3.10 Si ν est absolument continue par rapport à μ de densité $f \in L^2(\mu)$ alors

$$\text{Ent}(\nu|\mu) \leq \|f - 1\|_{L^2(\mu)}^2.$$

Preuve. On utilise la majoration $\ln u \leq u - 1$ pour $u \geq 0$:

$$\text{Ent}(\nu|\mu) = \int \ln f d\nu \leq \int (f - 1) d\nu = \int (f - 1)f d\mu = \int (f - 1)^2 d\mu,$$

puisque $\mu(f - 1) = 0$. \square

3.3.3 Inégalité de Csiszár-Kullback

Théorème 3.11 Soit μ et ν dans $\mathcal{P}(E)$ tels que ν admet la densité f par rapport à μ .

$$\|\nu - \mu\|_{vt} \leq \sqrt{\frac{1}{2}\text{Ent}(\nu|\mu)} = \sqrt{\frac{1}{2}\text{Ent}_\mu(f)}.$$

Preuve. La démonstration repose sur une inégalité astucieuse due à Pinsker : pour tout $u \geq 0$,

$$3(u - 1)^2 \leq (2u + 4)(u \log u - u + 1).$$

D'après les inégalités de Cauchy-Schwarz et de Pinsker,

$$\begin{aligned} \int |f - 1| d\mu &\leq \int \sqrt{\frac{2f + 4}{3}} \sqrt{f \log f - f + 1} d\mu \\ &\leq \sqrt{\int \frac{2f + 4}{3} d\mu} \sqrt{\int f \log f d\mu} \\ &\leq \sqrt{2 \int f \log f d\mu} = \sqrt{2\text{Ent}_\mu(f)}. \end{aligned}$$

On conclut en utilisant la proposition 3.5. □

3.3.4 Encore une belle inégalité de Csiszár

Définition 3.12 On dit qu'une mesure de probabilité $\mu^{(N)}$ sur l'espace produit E^N est échangeable si elle est invariante par permutation des coordonnées.

Théorème 3.13 (Csiszár) Soit (E, \mathcal{E}) un espace mesurable et $\mu^{(N)}$ une mesure échangeable sur E^N absolument continue par rapport à une mesure produit $\eta^{\otimes N}$. Si, pour $1 \leq q \leq N$, $\mu^{(q,N)}$ désigne la loi des q -premières coordonnées, on a

$$\text{Ent}\left(\mu^{(q,N)}|\eta^{\otimes q}\right) \leq \frac{q}{N} \left(1 + \frac{\{N/q\}}{[N/q]}\right) \text{Ent}\left(\mu^{(N)}|\eta^{\otimes N}\right),$$

où $[x]$ désigne la partie entière de x et $\{x\} = x - [x]$.

Remarque 3.14 Pour tout $x \geq 1$,

$$x \left(1 + \frac{\{x\}}{[x]}\right) \leq \frac{2}{x}.$$

Preuve. D'après la formule variationnelle de l'entropie, on a

$$\text{Ent}\left(\mu^{(N)}|\eta^{\otimes N}\right) \geq \mu^{(N)}(g^{(q)}) - \ln \eta^{\otimes N}(\exp g^{(q)})$$

avec $g^{(q)}$ de la forme

$$g^{(q)}(x_1, \dots, x_N) = \sum_{p=1}^{[N/q]} \varphi(x_{(p-1)q+1}, \dots, x_{(p-1)q+q})$$

avec φ bornée. Par échangeabilité,

$$\mu^{(N)}(g^{(q)}) = [N/q] \mu^{(q,N)}(\varphi) \quad \text{et} \quad \eta^{\otimes N}(\exp g^{(q)}) = (\eta^{\otimes q}(\varphi))^{[N/q]}.$$

Pour toute fonction φ bornée sur E^q ,

$$[N/q] \left(\mu^{(q,N)}(\varphi) - \ln \eta^{\otimes q}(\varphi) \right) \leq \text{Ent} \left(\mu^{(N)} | \eta^{\otimes N} \right).$$

L'arbitraire sur φ assure que

$$\text{Ent} \left(\mu^{(q,N)} | \eta^{\otimes q} \right) \leq \frac{1}{[N/q]} \text{Ent} \left(\mu^{(N)} | \eta^{\otimes N} \right).$$

Il ne reste plus qu'à remarquer que, pour $x \geq 1$,

$$\frac{1}{[x]} = \frac{1}{x} \frac{[x] - \{x\}}{[x]} = \frac{1}{x} \left(1 + \frac{\{x\}}{[x]} \right)$$

pour achever la preuve. □

3.3.5 Encore une hypothèse de régularité

Pour contrôler l'inégalité de Csiszár ait un intérêt, il faut se placer sous des hypothèses qui assurent au moins que la loi du système de particules admet une densité par rapport à la loi des N particules non linéaires indépendantes (sinon le membre de droite de ladite inégalité sera infini). Pour que ceci soit garanti, il faut que les noyaux $(K_{p+1, \mu_p}(x, dv))_{\mu_p \in \mathcal{P}(E_p)}$ soient absolument continues par rapport à $K_{p+1, \eta_p}(x, dv)$.

Définition 3.15 *Pour tout $p \geq 1$, on dit que la condition $(M)^p$ est satisfaite si, pour tout $n \geq 1$ et tout $x_{n-1} \in E_{n-1}$, $M_n(x_{n-1}, \cdot)$ est absolument continu par rapport à η_n et*

$$M_n(x_{n-1}, dx_n) = k(x_{n-1}, x_n) \eta_n(dx_n) \quad \text{avec, pour tout } p \geq 1, \quad \sup_{x_{n-1} \in E_{n-1}} k(x_{n-1}, \cdot) \in L^p(\eta_n).$$

Cette propriété n'est pas très parlante. Voici une condition suffisante bien plus explicite.

Lemme 3.16 *Supposons que la propriété de contraction $(M)_1$ soit satisfaite, c'est-à-dire que, pour tout $n \in \mathbb{N}$, il existe $\varepsilon_n(M) > 0$ tel que, pour tous $x_n, y_n \in E_n$,*

$$M_n(x_n, \cdot) \geq \varepsilon_n(M) M_n(y_n, \cdot).$$

Alors la condition $(M)^p$ est satisfaite pour tout $p \geq 1$.

Preuve. Soit $a_{n-1} \in E_{n-1}$. Posons

$$p_n(dy) = M_n(a, dy) \quad \text{et} \quad m_n(x, y) = \frac{dM_n(x, \cdot)}{dM_n(a, \cdot)}(y) \in [\varepsilon_n(M), \varepsilon_n(M)^{-1}].$$

Pour tout $x \in E_{n-1}$, il est clair que

$$\varepsilon_n(M) p_n(dy) \leq M_n(x, dy) = m_n(x, y) p_n(dy) \leq \frac{1}{\varepsilon_n(M)} p_n(dy).$$

D'autre part,

$$\int \sup_x \left(\frac{dM_n(x, \cdot)}{d\eta_n} \right)^p d\eta_n = \int \sup_x \left(\frac{dM_n(x, \cdot)}{dp_n} \right)^p \left(\frac{dp_n}{d\eta_n} \right)^{p-1} dp_n.$$

Puisque

$$\begin{aligned} \frac{d\eta_n}{dp_n}(y) &= \frac{1}{\eta_{n-1}(G_{n-1})} \int_{E_{n-1}} G_{n-1}(x_{n-1}) \frac{dM_n(x_{n-1}, \cdot)}{dp_n}(y) \eta_{n-1}(dx_{n-1}) \\ &= \frac{\eta_{n-1}(G_{n-1}m_n(\cdot, y))}{\eta_{n-1}(G_{n-1})}, \end{aligned}$$

on a donc l'encadrement suivant :

$$\varepsilon_n(M) \leq \frac{d\eta_n}{dp_n}(y) \leq \frac{1}{\varepsilon_n(M)}.$$

On en déduit immédiatement que

$$\int \sup_x \left(\frac{dM_n(x, \cdot)}{d\eta_n} \right)^p d\eta_n \leq \frac{1}{\varepsilon_n(M)^{2p-1}} < +\infty,$$

pour tout $p \geq 1$. □

Exemple 1 Soit $E_n = \mathbb{R}$ pour tout $n \in \mathbb{N}$ et M_n défini par

$$M_n(x, dy) := \frac{c(n)}{2} e^{-c(n)|y-A(x)|} dy,$$

avec $c(n) > 0$ et $\text{osc}(A_n) < +\infty$. La condition $(M)_1$ est satisfaite avec $\varepsilon_n(M) = \exp(-c(n)\text{osc}A_n)$. En effet,

$$\ln \frac{M_n(x, \cdot)}{M_n(y, \cdot)}(z) = c(n)(|z - A_n(y)| - |z - A_n(x)|),$$

et donc

$$\left\| \ln \frac{M_n(x, \cdot)}{M_n(y, \cdot)}(z) \right\|_{\infty} \leq c(n)|A_n(x) - A_n(y)| \leq c(n)\text{osc}(A_n).$$

3.3.6 Densité de la loi du système de particules

Rappelons que la loi des trajectoires jusqu'à l'instant n de N particules non linéaires indépendantes est donnée par

$$\mathbb{K}_{\eta_0, n}^{\otimes N}(d(x_0, \dots, x_n)) = \prod_{i=1}^N \eta_0(dx_0^i) K_{1, \eta_0}(dx_1^i) \cdots K_{n, \eta_{n-1}}(dx_n^i) \in \mathcal{P}(E_{[0, n]}^N),$$

avec $x_p = (x_p^i)_{1 \leq i \leq N} \in E_p^N$. Rappelons l'expression de K_{p+1, μ_p} :

$$K_{p+1, \mu_p}(u, dv) = \varepsilon_p G_p(u) M_{p+1}(u, dv) + (1 - \varepsilon_p G_p(u)) \frac{\mu_p(G_p M_{p+1}(\cdot, dv))}{\mu_p(G_p)}. \quad (14)$$

De plus, si l'on note $m(x_p)$ la mesure empirique du vecteur x_p alors la loi de N particules en interaction est donnée par

$$\mathbb{P}_{\eta_0, n}^N(d(x_0, \dots, x_n)) = \prod_{i=1}^N \eta_0(dx_0^i) K_{1, m(x_0)}(dx_1^i) \cdots K_{n, m(x_{n-1})}(dx_n^i) \in \mathcal{P}(E_{[0, n]}^N).$$

Ainsi, si $\xi_{[0,n]}$ désigne les trajectoires du système de particules jusqu'au temps n (la dépendance en N est implicite), pour toute fonction $F \in \mathcal{B}_b(E_{[0,n]}^N)$, on a

$$\mathbb{E}(F(\xi_{[0,n]})) = \int_{E_{[0,n]}^N} F(x_0, \dots, x_n) \left[\prod_{p=0}^{n-1} \prod_{i=1}^N \frac{dK_{p+1,m(x_p)}(x_p^i, \cdot)}{dK_{p+1,\eta_p}(x_p^i, \cdot)}(x_{p+1}^i) \right] \mathbb{K}_{\eta_0,n}^{\otimes N}(d(x_0, \dots, x_n)).$$

La mesure $\mathbb{P}_{\eta_0,n}^N$ admet donc une densité par rapport à la mesure $\mathbb{K}_{\eta_0,n}^{\otimes N}$ qui s'écrit sous la forme $\exp H_n^N(x_0, \dots, x_n)$ avec

$$\begin{aligned} H_n^N(x_0, \dots, x_n) &= \sum_{p=0}^{n-1} \sum_{i=1}^N \ln \frac{dK_{p+1,m(x_p)}(x_p^i, \cdot)}{dK_{p+1,\eta_p}(x_p^i, \cdot)} \\ &= N \sum_{p=0}^{n-1} \int_{E_p \times E_{p+1}} \ln \frac{dK_{p+1,m(x_p)}(u, \cdot)}{dK_{p+1,\eta_p}(u, \cdot)}(v) m(x_p, x_{p+1})(d(u, v)) \end{aligned}$$

avec la notation

$$m(x_p, x_{p+1})(d(u, v)) = \frac{1}{N} \sum_{i=1}^N \delta_{(x_p^i, x_{p+1}^i)}((u, v)).$$

Remarque 3.17 On peut réécrire $H_n^{(N)}(\xi_{[0,n]})$ comme une fonctionnelle de la mesure empirique du système de particules :

$$H_n^{(N)}(\xi_{[0,n]}) = N \mathcal{H}_n(m(\xi_{[0,n]}))$$

avec pour toute mesure μ sur $E_{[0,n]}$,

$$\mathcal{H}_n(\mu) = \sum_{p=0}^{n-1} \int_{E_p \times E_{p+1}} \ln \frac{dK_{p+1,\mu_p}(u, \cdot)}{dK_{p+1,\eta_p}(u, \cdot)}(v) \mu_{p,p+1}(d(u, v))$$

où $\mu_{p,p+1}$ est la marginale de μ sur $E_p \times E_{p+1}$.

Il est clair que la loi du système de particules est échangeable puisque sa densité par rapport à la mesure produit $\mathbb{K}_{\eta_0,n}^{\otimes N}$ ne s'exprime qu'en fonction de la mesure empirique du système.

3.3.7 Le résultat de convergence

Théorème 3.18 Sous les conditions $(M)_1$ et (G) , on obtient

$$\text{Ent}(\mathbb{P}_{\eta_0,n}^{q,N} | \mathbb{K}_{\eta_0,n}^{\otimes q}) \leq c(\varepsilon(M), \varepsilon(G)) \frac{qn}{N}.$$

Preuve. Puisque $\mathbb{P}_{\eta_0,n}^N$ est échangeable, on a, d'après le théorème 3.13,

$$\text{Ent}(\mathbb{P}_{\eta_0,n}^{q,N} | \mathbb{K}_{\eta_0,n}^{\otimes q}) \leq \frac{2q}{N} \text{Ent}(\mathbb{P}_{\eta_0,n}^N | \mathbb{K}_{\eta_0,n}^{\otimes N}) = \frac{2q}{N} \mathbb{E} \left(H_n^{(N)}(\xi_{[0,n]}) \right) = 2q \mathbb{E}(\mathcal{H}_n(m(\xi_{[0,n]}))).$$

D'après l'expression de \mathcal{H}_n ,

$$\begin{aligned} |\mathbb{E}(\mathcal{H}_n(m(\xi_{[0,n]})))| &= \sum_{p=0}^{n-1} \int_{E_{p+1}} \ln \frac{dK_{p+1,m(\xi_p)}(\xi_p^1, \cdot)}{dK_{p+1,\eta_p}(\xi_p^1, \cdot)}(v) K_{p+1,m(\xi_p)}(\xi_p^1, dv) \\ &= \sum_{p=0}^{n-1} \mathbb{E}(\text{Ent}(K_{p+1,m(\xi_p)}(\xi_p^1, \cdot) | K_{p+1,\eta_p}(\xi_p^1, \cdot))) \\ &\leq \sum_{p=0}^{n-1} \mathbb{E} \left(\int_{E_{p+1}} \left| \frac{dK_{p+1,m(\xi_p)}(\xi_p^1, \cdot)}{dK_{p+1,\eta_p}(\xi_p^1, \cdot)}(v) - 1 \right|^2 K_{p+1,\eta_p}(\xi_p^1, dv) \right), \end{aligned}$$

d'après le lemme 3.10. Pour pouvoir utiliser le théorème 2.7, il faut se ramener au contrôle de l'écart entre $\eta_p^N f_p$ et $\eta_p f_p$ pour une certaine fonction f_p bornée bien choisie. Pour ce faire, on va faire apparaître artificiellement la mesure η_{p+1} pour utiliser la condition $(M)^2$. On peut réécrire le membre de droite ci dessus de la façon suivante :

$$\sum_{p=0}^{n-1} \mathbb{E} \left(\int_{E_{p+1}} \left| \frac{dK_{p+1,m(\xi_p)}(\xi_p^1, \cdot)}{d\eta_{p+1}}(v) - \frac{dK_{p+1,\eta_p}(\xi_p^1, \cdot)}{d\eta_{p+1}}(v) \right|^2 \frac{d\eta_{p+1}}{dK_{p+1,\eta_p}(\xi_p^1, \cdot)}(v) \eta_{p+1}(dv) \right).$$

Puisque

$$M_{p+1}(x_p, dx_{p+1}) = k(x_p, x_{p+1})\eta_{p+1}(dx_{p+1}),$$

on a

$$\frac{K_{p+1,\mu_p}(u, \cdot)}{d\eta_{p+1}}(v) = \varepsilon_p G_p(u) k_{p+1}(u, v) + (1 - \varepsilon_p G_p(u)) \frac{\mu_p(G_p k_{p+1}(\cdot, v))}{\mu_p(G_p)}.$$

La mesure $K_{p+1,\eta_p}(u, \cdot)$ n'est pas tout à fait égale à η_{p+1} (elle dépend de $u \in E_p$) mais on a, d'après (14),

$$\eta_p K_{p+1,\eta_p}(dv) = \int K_{p+1,\eta_p}(u, dv) \eta_p(du) = \frac{\eta_p(G_p M_{p+1}(\cdot, dv))}{\eta_p(G_p)} = \Phi_{p+1}(\eta_p) = \eta_{p+1}(dv).$$

On a donc, d'après la définition de k_{p+1} ,

$$\frac{\eta_p(G_p k_{p+1}(\cdot, v))}{\eta_p(G_p)} = \frac{d\Phi_{p+1}(\eta_p)}{d\eta_{p+1}}(v) = 1. \quad (15)$$

Ainsi

$$\frac{K_{p+1,\eta_p}(u, \cdot)}{d\eta_{p+1}}(v) = \varepsilon_p G_p(u) k_{p+1}(u, v) + (1 - \varepsilon_p G_p(u)) \geq (1 - \varepsilon_p G_p(u)). \quad (16)$$

Ceci permet de réécrire la différence des densités :

$$\begin{aligned} \frac{dK_{p+1,\mu_p}(u, \cdot)}{d\eta_{p+1}}(v) - \frac{dK_{p+1,\eta_p}(u, \cdot)}{d\eta_{p+1}}(v) &= (1 - \varepsilon_p G_p(u)) \left(\frac{\mu_p(G_p k_{p+1}(\cdot, v))}{\mu_p(G_p)} - \frac{\eta_p(G_p k_{p+1}(\cdot, v))}{\eta_p(G_p)} \right) \\ &= (1 - \varepsilon_p G_p(u)) \left(\frac{\mu_p(G_p k_{p+1}(\cdot, v))}{\mu_p(G_p)} - 1 \right). \end{aligned}$$

On injecte cette expression pour obtenir

$$|\mathbb{E}(\mathcal{H}_n(m(\xi_{[0,n]})))| \leq \sum_{p=0}^{n-1} \int_{E_{p+1}} \eta_{p+1}(dv) \mathbb{E} \left(\left| \frac{m(\xi_p)(G_p k_{p+1}(\cdot, v))}{m_p(\xi_p)(G_p)} - 1 \right|^2 \right),$$

puisque, d'après (16),

$$\frac{d\eta_{p+1}}{dK_{p+1,\eta_p}(\xi_p^1, \cdot)}(v) (1 - \varepsilon_p G_p(\xi_p^1))^2 \leq 1 - \varepsilon_p G_p(\xi_p^1) \leq 1.$$

Il ne nous reste plus qu'à contrôler l'espérance du membre de droite. Notons

$$\tilde{G}_p := \frac{G_p}{\eta_p(G_p)}, \quad r_p := \frac{\sup G_p}{\inf G_p}, \quad |k_{p+1}|(v) := \sup_u k_{p+1}(u, v)$$

et enfin

$$f_p^v := \frac{1}{r_p(1 + |k_{p+1}|(v))} \tilde{G}_p(k_{p+1}(\cdot, v) - 1).$$

Remarquons que \tilde{G}_p est borné par r_p , $|f_p^v|$ par 1 et que, d'après (15),

$$\eta_p(f_p^v) = \frac{1}{r_p(1 + |k_{p+1}|(v))} \left(\frac{\eta_p(G_p k_{p+1}(\cdot, v))}{\eta_p(G_p)} - 1 \right) = 0.$$

On a alors

$$\begin{aligned} \left| \frac{m(\xi_p)(G_p k_{p+1}(\cdot, v))}{m_p(\xi_p)(G_p)} - 1 \right| &= \left| \frac{m(\xi_p)(\tilde{G}_p k_{p+1}(\cdot, v))}{m_p(\xi_p)(\tilde{G}_p)} - 1 \right| = \frac{1}{m_p(\xi_p)(\tilde{G}_p)} \left| m(\xi_p)(\tilde{G}_p(k_{p+1}(\cdot, v)) - 1) \right| \\ &\leq r_p^2(1 + |k_{p+1}|(v)) |m(\xi_p)(f_p^v) - \eta_p(f_p^v)|. \end{aligned}$$

On en conclut donc que

$$|\mathbb{E}(\mathcal{H}_n(m(\xi_{[0,n]})))| \leq \sum_{p=0}^{n-1} r_p^2 \int_{E_{p+1}} \eta_{p+1}(dv)(1 + |k_{p+1}|(v))^2 \mathbb{E}(|m(\xi_p)(f_p^v) - \eta_p(f_p^v)|^2).$$

Les hypothèses $(M)_1$ et (G) assurent le contrôle de l'espérance d'après le théorème 2.7 indépendamment de v et l'hypothèse $(M)_1$ implique $(M)^2$ d'après le lemme 3.16. \square

4 Stabilité des semi-groupes de Feynman-Kac, mesures invariantes

4.1 Mesurer l'écart entre des mesures

Cette section recense différentes façons de mesurer l'écart entre deux mesures de probabilité. Commençons pas la plus importante en pratique.

Définition 4.1 *La variation totale de μ et ν deux mesures de probabilité sur (E, \mathcal{E}) est donnée par*

$$\|\mu - \nu\|_{vt} := \frac{1}{2} \sup_{A \in \mathcal{E}} (\mu(A) - \nu(A)).$$

Remarque 4.2 *La variation totale est un outil précieux car, en particulier, c'est une distance. En fait c'est plus que cela. À toute mesure μ finie signée sur (E, \mathcal{E}) , on associe sa variation totale :*

$$\|\mu\|_{vt} := \frac{1}{2} \sup_{A, B \in \mathcal{E}} (\mu(A) - \mu(B)).$$

C'est une norme sur l'espace vectoriel des mesures finies. À toute mesure signée finie μ on peut associer μ^+ et μ^- deux mesures positives étrangères et E^+ et E^- deux éléments de E tels que

$$\forall A \in \mathcal{E}, \quad \mu(A) = \mu^+(A) - \mu^-(A) \quad \text{et} \quad \mu^+(E^-) = \mu^-(E^+) = 0.$$

On parle de décomposition de Hahn-Jordan. On a alors une autre expression pour la variation totale de μ :

$$\|\mu\|_{vt} := \frac{\mu^+(E) + \mu^-(E)}{2} = \frac{\mu(E^+) - \mu(E^-)}{2}.$$

Si μ est de masse nulle alors $\mu^+(E) = \mu^-(E)$ et $\mu(E^+) = -\mu(E^-)$.

On a d'autres formulations équivalentes et souvent plus pratiques.

Proposition 4.3 *On a encore*

$$\|\mu - \nu\|_{vt} = \frac{1}{2} \sup_{f, \|f\| \leq 1} (\mu(f) - \nu(f)) = \sup_{f, \text{osc} f \leq 1} (\mu(f) - \nu(f)).$$

Preuve. Il faut utiliser la décomposition de Hahn-Jordan de la mesure signée $m := \mu - \nu$ de masse 0 :

$$m := \mu - \nu = m^+ - m^- \quad \text{avec} \quad \|\mu - \nu\|_{vt} = m^+(E) = m^-(E).$$

Pour toute fonction d'oscillation bornée par 1,

$$\begin{aligned} |\mu(f) - \nu(f)| &= \left| \int f(x) m^+(dx) - \int f(y) m^-(dy) \right| \\ &= \|\mu - \nu\|_{vt} \left| \int (f(x) - f(y)) \frac{m^+(dx)}{m^+(E)} \frac{m^-(dy)}{m^-(E)} \right| \leq \|\mu - \nu\|_{vt}. \end{aligned}$$

L'inégalité inverse s'obtient en remarquant que $\mathbf{1}_A$ est d'oscillation 1. \square

Remarque 4.4 *D'autres fonctionnelles que la variation totale peuvent être utilisé pour quantifier l'écart entre deux mesures de probabilité. Citons par exemple l'entropie relative (ou entropie de Boltzmann, information de Kullback...)* :

$$H(\mu, \nu) = \text{Ent}(\mu|\nu) = \begin{cases} \int \ln \left(\frac{d\mu}{d\nu} \right) d\nu & \text{si } \mu \ll \nu, \\ +\infty & \text{sinon.} \end{cases}$$

On peut adapter tout ce qui suit à une large classe de telles fonctionnelles contentant les principales **distances** entre mesures de probabilité.

4.2 Les noyaux markoviens contractent

Nous donnons ici des justifications mathématiques au célèbre lieu commun

Les noyaux markoviens font décroître l'entropie.

Définition 4.5 *Le coefficient de contraction de Dobrushin d'un noyau markovien M de (E, \mathcal{E}) dans (F, \mathcal{F}) est noté $\beta(M)$ et vaut*

$$\beta(M) := \sup_{x, y \in E} \|M(x, \cdot) - M(y, \cdot)\|_{vt} \in [0, 1].$$

Remarque 4.6 *L'intérêt de ce coefficient est le suivant. Soit M est un noyau markovien de E dans F tel que*

$$\exists \varepsilon > 0, \forall x, y \in E, \quad M(x, \cdot) \geq \varepsilon M(y, \cdot),$$

au sens où $M(f)(x) \geq \varepsilon M(f)(y)$ pour toute fonction positive de $\mathcal{B}_b(F)$. Alors

$$\beta(M) \leq 1 - \varepsilon.$$

En particulier, si $E = F$, alors $\|M^n(x, \cdot) - M^n(y, \cdot)\|_{vt}$ converge exponentiellement vite vers 0. On peut ainsi obtenir une estimation de la vitesse de convergence vers la mesure invariante.

Proposition 4.7 *Soit M un noyau markovien de (E, \mathcal{E}) dans (F, \mathcal{F}) . Pour toute mesure $\mu \in \mathcal{M}(E)$, on a*

$$\|\mu M\|_{vt} \leq \beta(M) \|\mu\|_{vt} + (1 - \beta(M)) \frac{|\mu(E)|}{2}. \quad (17)$$

De plus, $\beta(M)$ est la norme de M en tant qu'endomorphisme sur $\mathcal{M}_0(E)$ (l'ensemble des mesures signées de masse nulle). Il est donc égal à

$$\beta(M) = \sup_{\mu \in \mathcal{M}_0(E)} \frac{\|\mu M\|_{vt}}{\|\mu\|_{vt}}. \quad (18)$$

Preuve. Supposons tout d'abord que $\mu \in \mathcal{M}_0(E)$. Alors, $\|\mu\|_{vt} = \mu(E^+) = -\mu(E^-)$. De plus, μM est aussi de masse nulle et

$$\|\mu M\|_{vt} = \sup_{A \in \mathcal{E}} \mu M(A).$$

Or, pour tout $A \in \mathcal{F}$,

$$\begin{aligned} \mu M(A) &= \mu(\mathbf{1}_{E^+} M(\mathbf{1}_A)) + \mu(\mathbf{1}_{E^-} M(\mathbf{1}_A)) \leq \mu(\mathbf{1}_{E^+} M(\mathbf{1}_A)) + \inf_{y \in E} M(y, A) \mu(E^-) \\ &\leq \int_{E^+} \sup_{y \in E} (M(x, A) - M(y, A)) \mu(dx) \leq \beta(M) \mu(E^+) = \beta(M) \|\mu\|_{vt}, \end{aligned}$$

ce qui démontre (17) (dans le cas μ de masse nulle) et

$$\beta(M) \geq \sup_{\mu \in \mathcal{M}_0(E)} \frac{\|\mu M\|_{vt}}{\|\mu\|_{vt}}.$$

□

Voici le résultat fondamental établissant la propriété de contraction des noyaux markoviens.

Théorème 4.8 *Pour toutes probabilités μ et ν sur (E, \mathcal{E}) et tout noyau markovien M de E sur F , on a*

$$H(\mu M, \nu M) \leq \beta(M) H(\mu, \nu).$$

Ce résultat s'appuie sur le lemme suivant.

Lemme 4.9 *Soit μ et ν deux mesures bornées sur $(\mathbb{R}_+^2, \mathcal{B}(\mathbb{R}_+^2))$ admettant un premier moment et telles que*

– μ et ν agissent de la même manière sur les fonctions affines :

$$\mu(\mathbb{R}_+^2) = \nu(\mathbb{R}_+^2), \quad \int s \mu(ds, dt) = \int s \nu(ds, dt) \quad \text{et} \quad \int t \mu(ds, dt) = \int t \nu(ds, dt),$$

– pour tous $a, b \in \mathbb{R}$,

$$\int |as - bt| \mu(ds, dt) \leq \int |as - bt| \nu(ds, dt).$$

Alors pour toute fonction convexe et homogène h sur \mathbb{R}_+^2 , $\mu(h) \leq \nu(h)$.

4.3 Une première majoration

Rappelons pour commencer l'expression du semi-groupe de Feynman-Kac :

$$\Phi_{p,n}(\mu_p)(f_n) = \frac{\mu_p(Q_{p,n}(f_n))}{\mu_p(Q_{p,n}(1))} = \frac{\mathbb{E}_{p,\mu_p} \left(f(X_n) \prod_{q=p}^{n-1} G_q(X_q) \right)}{\mathbb{E}_{p,\mu_p} \left(\prod_{q=p}^{n-1} G_q(X_q) \right)}.$$

Dans la suite il sera très utile d'écrire $\Phi_{p,n}$ sous la forme suivante, inspirée de (4)

$$\Phi_{p,n}(\mu_p) = \Psi_{p,n}(\mu_p) P_{p,n},$$

où

$$P_{p,n}(f_n) := \frac{Q_{p,n}(f_n)}{Q_{p,n}(1)} \quad \text{et} \quad \Psi_{p,n}(\mu_p)(dx_p) = \frac{G_{p,n}(x_p)}{\mu_p(G_{p,n})} \mu_p(dx_p) \quad \text{avec} \quad G_{p,n} = Q_{p,n}(1)$$

sont des applications de $\mathcal{P}(E_p)$ dans $\mathcal{P}(E_n)$ et lui-même respectivement.

Proposition 4.10 *Pour tous $0 \leq p \leq n$, on a*

$$\begin{aligned}\beta(P_{p,n}) &= \sup_{\mu_p, \nu_p \in \mathcal{P}(E_p)} \frac{\|\Phi_{p,n}(\mu_p) - \Phi_{p,n}(\nu_p)\|_{vt}}{\|\Psi_{p,n}(\mu_p) - \Psi_{p,n}(\nu_p)\|_{vt}} \\ &\leq \sup_{\mu_p, \nu_p \in \mathcal{P}(E_p)} \|\Phi_{p,n}(\mu_p) - \Phi_{p,n}(\nu_p)\|_{vt}.\end{aligned}$$

Preuve. Premièrement, puisque $\Phi_{p,n}(\mu_p) = \Psi_{p,n}(\mu_p)P_{p,n}$, la formulation (18) assure que

$$\|\Phi_{p,n}(\mu_p) - \Phi_{p,n}(\nu_p)\|_{vt} \leq \beta(P_{p,n})\|\Psi_{p,n}(\mu_p) - \Psi_{p,n}(\nu_p)\|_{vt}$$

De plus, pour tous $x_p, y_p \in E_p$ distincts,

$$\Psi_{p,n}(\delta_{x_p}) = \delta_{x_p}, \quad \Phi_{p,n}(\delta_{x_p}) = P_{p,n}(x_p, \cdot) \quad \text{et} \quad \|\delta_{x_p} - \delta_{y_p}\|_{vt} = 1.$$

On a donc

$$\beta(P_{p,n}) = \sup_{x,y \in E_p} \frac{\|\delta_x P_{p,n} - \delta_y P_{p,n}\|_{vt}}{\|\delta_x - \delta_y\|_{vt}} \leq \begin{cases} \sup_{\mu_p, \nu_p \in \mathcal{P}(E_p)} \frac{\|\Phi_{p,n}(\mu_p) - \Phi_{p,n}(\nu_p)\|_{vt}}{\|\Psi_{p,n}(\mu_p) - \Psi_{p,n}(\nu_p)\|_{vt}}, \\ \sup_{\mu_p, \nu_p \in \mathcal{P}(E_p)} \|\Psi_{p,n}(\mu_p)P_{p,n} - \Psi_{p,n}(\nu_p)P_{p,n}\|_{vt}. \end{cases}$$

□

On peut à présent estimer le coefficient de contraction de la transformation $\Phi_{p,n}$.

Théorème 4.11 *Pour tous $0 \leq p \leq n$ et $\mu_p, \nu_p \in \mathcal{P}(E_p)$, on a*

$$\|\Phi_{p,n}(\mu_p) - \Phi_{p,n}(\nu_p)\|_{vt} \leq \beta(P_{p,n}) \frac{\|G_{p,n}\|_{osc}}{\nu_p(G_{p,n}) \vee \nu_p(G_{p,n})} \|\mu_p - \nu_p\|_{vt}.$$

Remarque 4.12 *Avec les définitions (5), on a immédiatement, pour $0 \leq p \leq n$,*

$$\|\Phi_{p,n}(\mu_p) - \Phi_{p,n}(\nu_p)\|_{vt} \leq 2r_{p,n}\beta(P_{p,n})\|\mu_p - \nu_p\|_{vt}.$$

Pour obtenir ce théorème, il faut utiliser la proposition 4.10 et le lemme suivant qui contrôle l'action de $\Psi_{p,n}$.

Lemme 4.13 *Soit G une fonction strictement positive et mesurable sur (E, \mathcal{E}) . On associe à G la transformation de Boltzmann-Gibbs Ψ de $\mathcal{P}(E)$ dans lui-même définie par*

$$\Psi(\mu)(dx) = \frac{1}{\mu(G)} G(x) dx.$$

Pour tous $\mu, \nu \in \mathcal{P}(E)$, on a

$$\|\Psi(\mu) - \Psi(\nu)\|_{vt} \leq \frac{\|G\|_{osc}}{\mu(G) \vee \nu(G)} \|\mu - \nu\|_{vt}.$$

Preuve. Pour établir ce résultat on écrit pour $f \in \mathcal{B}_b(E)$:

$$\Psi(\mu)(f) - \Psi(\nu)(f) = \frac{1}{\mu(G)} \mu[G(f - \Psi(\nu)(f))].$$

On remarque ensuite

$$\begin{aligned}G(x)[f(x) - \Psi(\nu)(f)] - G(y)[f(y) - \Psi(\nu)(f)] \\ = (G(x) - G(y))[f(x) - \Psi(\nu)(f)] + G(y)(f(x) - f(y)).\end{aligned}$$

Ceci permet d'obtenir

$$\text{osc}(G(f - \Psi(\nu)(f))) \leq \|G\|_{\text{osc}} \text{osc}(f)$$

Puisque $G(f - \Psi(\nu)(f))$ est centrée pour ν , on a

$$\begin{aligned} \Psi(\mu)(f) - \Psi(\nu)(f) &\leq \frac{1}{\mu(G)} (\mu[G(f - \Psi(\nu)(f))] - \nu[G(f - \Psi(\nu)(f))]) \\ &\leq \frac{\|G\|_{\text{osc}} \text{osc}(f)}{\mu(G)} \|\mu - \nu\|_{vt}. \end{aligned}$$

Par symétrie entre μ et ν , on obtient le résultat annoncé en prenant le supremum sur les fonctions d'oscillation inférieure à 1. \square

4.4 Encore un peu de semi-groupes

Il nous faut à présent contrôler $\beta(P_{p,q})$ en fonction des paramètres du modèle (chaîne de Markov et potentiels). Pour cela, il est crucial de réinterpréter $P_{p,q}$. Pour n fixé, notons, pour $p \leq q \leq n$,

$$R_{p,q}^{(n)}(f_q) := \frac{Q_{p,q}(f_q G_{q,n})}{Q_{p,q}(G_{q,n})}.$$

Le noyau $P_{p,n}$ peut être vu comme la transition de E_p dans E_n d'une chaîne de Markov inhomogène de transitions $(R^{(n)})_{0 \leq p \leq q \leq n}$. En effet, on a bien $R_{p,n}^{(n)} = P_{p,n}$ et, pour tous $p \leq p' \leq q' \leq r' \leq n$ et tous fonction $f_{r'} \in \mathcal{B}_b(E_{r'})$,

$$\begin{aligned} R_{p',q'}^{(n)} R_{q',r'}^{(n)}(f_{r'}) &= \frac{Q_{p',q'}(R_{q',r'}^{(n)}(f_{r'}) G_{q',n})}{Q_{p',q'}(G_{q',n})} = \frac{1}{Q_{p',q'}(G_{q',n})} Q_{p',q'} \left(\frac{Q_{q',r'}(f_{r'} G_{r',n})}{Q_{q',r'}(G_{r',n})} G_{q',n} \right) \\ &\stackrel{(*)}{=} \frac{Q_{p',q'}(Q_{q',r'}(f_{r'} G_{r',n}))}{Q_{p',q'}(G_{q',n})} \stackrel{(**)}{=} \frac{Q_{p',r'}(f_{r'} G_{r',n})}{Q_{p',q'} Q_{q',n}(1)} = R_{p',r'}^{(n)}. \end{aligned}$$

Pour (*), on a utilisé que

$$Q_{q',r'}(G_{r',n}) = Q_{q',r'} Q_{r',n}(1) = Q_{q',n}(1) = G_{q',n}.$$

Pour (**), on a utilisé le même argument et la propriété de semi-groupe de $(Q_{p,q})$.

Cette écriture de $P_{p,n}$ permettra de le saucissonner en tranches de longueur m pour obtenir des contrôles de contraction...

On dit qu'un semi-groupe $(I_{p,q})_{p \leq q}$ vérifie la condition $(I)_m$ pour $m \in \mathbb{N}^*$ si, pour tout $p \in \mathbb{N}$ il existe $\varepsilon_p(I) \in]0, 1[$ tel que, pour tous $x_p, y_p \in E_p$,

$$I_{p,p+m}(x_p, \cdot) \geq \varepsilon_p(I) I_{p,p+m}(y_p, \cdot),$$

ou encore, pour toute fonction positive $f_{p+m} \in \mathcal{B}_b(E_{p+m})$,

$$I_{p,p+m}(f_{p+m})(x_p) \geq \varepsilon_p(I) I_{p,p+m}(f_{p+m})(y_p).$$

On utilisera cette notation pour les semi-groupes M et Q . on dira également que $R^{(n)}$ vérifie la condition $(R^{(n)})_m$ si, pour tous $0 \leq p + m \leq n$ et tous $x_p, y_p \in E_p$,

$$R_{p,p+m}^{(n)}(x_p, \cdot) \geq \varepsilon_p(R^{(n)}) R_{p,p+m}^{(n)}(y_p, \cdot).$$

Enfin, on supposera que, pour tout n , le potentiel G_n vérifie

$$\forall x_n, y_n \in E_n, \quad G_n(x_n) \geq \varepsilon_n(G) G_n(y_n),$$

et on notera

$$\varepsilon_{p,n}(G) = \varepsilon_p(G) \cdots \varepsilon_{n-1}(G) = \prod_{k=p}^{n-1} \varepsilon_k(G).$$

Proposition 4.14 *Si $(Q)_m$ est vérifiée alors $(R^{(n)})_m$ aussi et, pour tout $0 \leq p+m \leq n$, on a*

$$\varepsilon_p(R^{(n)}) \leq \varepsilon_p^2(Q) \quad \text{et} \quad \beta(R_{p,p+m}^{(n)}) \leq 1 - \varepsilon_p^2(Q).$$

De plus, pour tous $x_p, y_p \in E_p$,

$$G_{p,n}(x_p) \geq \varepsilon_p(Q) G_{p,n}(y_p).$$

Si (G) et $(M)_m$ sont satisfaites, alors $(Q)_m$ l'est aussi et

$$\varepsilon_p(Q) \geq \varepsilon_{p,p+m}(G) \varepsilon_p(M) \quad \text{et} \quad \beta(R_{p,p+m}^{(n)}) \leq 1 - \varepsilon_{p+1,p+m}(G) \varepsilon_p^2(M).$$

Remarque 4.15 *Remarquons que dans le cas (G) et $(M)_m$ sont satisfaites, le coefficient de contraction $\beta(R_{p,p+m}^{(n)})$ ne dépend pas de G_p . Ceci aura une conséquence très intéressante sur le contrôle de $\beta(P_{0,n})$ sous $(M)_1$ (voir remarque 4.18).*

Preuve. Si $(Q)_m$ est satisfaite, alors pour toute fonction positive $f_q \in \mathcal{B}_b(E_q)$, tous $x_p, y_p \in E_p$ et $0 \leq p+m \leq q \leq n$, on a

$$\begin{aligned} R_{p,q}^{(n)}(f_q)(x_p) &= \frac{Q_{p,p+m}[Q_{p+m,q}(f_q G_{q,n})](x_p)}{Q_{p,p+m}[Q_{p+m,q}(G_{q,n})](x_p)} \\ &\geq \varepsilon_p^2(Q) \frac{Q_{p,p+m}[Q_{p+m,q}(f_q G_{q,n})](y_p)}{Q_{p,p+m}[Q_{p+m,q}(G_{q,n})](y_p)} = \varepsilon_p^2(Q) R_{p,q}^{(n)}(f_q)(y_p). \end{aligned}$$

De même, pour $0 \leq p+m \leq n$,

$$\frac{G_{p,n}(x_p)}{G_{p,n}(y_p)} = \frac{Q_{p,n}(1)(x_p)}{Q_{p,n}(1)(y_p)} = \frac{Q_{p,p+m}(G_{p+m,n})(x_p)}{Q_{p,p+m}(G_{p+m,n})(y_p)} \geq \varepsilon_p(Q).$$

Pour la seconde partie de la proposition, on écrit

$$\begin{aligned} \frac{Q_{p,q}(f_q)(x_p)}{Q_{p,q}(f_q)(y_p)} &= \frac{G_p(x_p) M_{p+1}(Q_{p+1,q}(f_q))(x_p)}{G_p(y_p) M_{p+1}(Q_{p+1,q}(f_q))(y_p)} \\ &\geq \varepsilon_p(G) \frac{M_{p+1}(G_{p+1} M_{p+2} Q_{p+2,q}(f_q))(x_p)}{M_{p+1}(G_{p+1} M_{p+2} Q_{p+2,q}(f_q))(y_p)} \\ &\geq \varepsilon_p(G) \varepsilon_{p+1}(G) \frac{M_{p+1} M_{p+2} Q_{p+2,q}(f_q)(x_p)}{M_{p+1} M_{p+2} Q_{p+2,q}(f_q)(y_p)} \\ &\geq \varepsilon_p(G) \varepsilon_{p+1}(G) \cdots \varepsilon_{p+m-1}(G) \frac{M_{p+1} M_{p+2} \cdots M_{p+m} Q_{p+m,q}(f_q)(x_p)}{M_{p+1} M_{p+2} \cdots M_{p+m} Q_{p+m,q}(f_q)(y_p)} \\ &\geq \varepsilon_{p,p+m}(G) \frac{M_{p,p+m} Q_{p+m,q}(f_q)(x_p)}{M_{p,p+m} Q_{p+m,q}(f_q)(y_p)} \geq \varepsilon_{p,p+m}(G) \varepsilon_p(M), \end{aligned}$$

Enfin, comme ci-dessus on écrit

$$\begin{aligned}
R_{p,p+m}^{(n)}(f_{p+m})(x_p) &= \frac{Q_{p,p+m}(f_{p+m}G_{p+m,n})(x_p)}{Q_{p,p+m}(G_{p+m,n})(x_p)} = \frac{G_p(x_p)}{G_p(x_p)} \frac{M_{p+1}Q_{p+1,p+m}(f_{p+m}G_{p+m,n})(x_p)}{M_{p+1}Q_{p+1,p+m}(G_{p+m,n})(x_p)} \\
&= \frac{M_{p+1}(G_{p+1}M_{p+2}Q_{p+2,p+m}(f_{p+m}G_{p+m,n}))(x_p)}{M_{p+1}(G_{p+1}M_{p+2}Q_{p+2,p+m}(G_{p+m,n}))(x_p)} \\
&\geq \varepsilon_{p+1}(G) \frac{M_{p,p+2}Q_{p+2,p+m}(f_{p+m}G_{p+m,n})(x_p)}{M_{p,p+2}Q_{p+2,p+m}(G_{p+m,n})(x_p)} \\
&\geq \cdots \geq \varepsilon_{p+1,p+m} \frac{M_{p,p+m}(f_{p+m}G_{p+m,n})(x_p)}{M_{p,p+m}(G_{p+m,n})(x_p)} \\
&\geq \varepsilon_{p+1,p+m}\varepsilon_p(M)^2 \frac{M_{p,p+m}(f_{p+m}G_{p+m,n})(y_p)}{M_{p,p+m}(G_{p+m,n})(y_p)}.
\end{aligned}$$

Remarque 4.16 *On utilise pour conclure une formulation équivalente du coefficient de Dobrushin : soit M un noyau de Markov de E dans F , alors*

$$\beta(M) = 1 - \inf \sum_{i=1}^n M(x, A_p) \wedge M(y, A_p),$$

où l'infimum est pris sur tous les éléments x, y de E et toutes les partitions mesurables finies de F .

Soit alors une partition $(A_l)_{1 \leq l \leq n}$ de E_{p+m} :

$$\begin{aligned}
\sum_{i=1}^n R_{p,p+m}^{(n)}(\mathbf{1}_{A_i})(x) \wedge R_{p,p+m}^{(n)}(\mathbf{1}_{A_i})(y) &\geq \varepsilon_{p+1,p+m}\varepsilon_p(M)^2 \sum_{i=1}^n \frac{M_{p,p+m}(x, A_i)}{M_{p,p+m}(G_{p+m,n})(x)} \\
&\geq \varepsilon_{p+1,p+m}\varepsilon_p(M)^2,
\end{aligned}$$

en envoyant tout sur x et en recomposant F . □

4.5 Majoration des coefficients de contraction

On utilise les résultats précédents pour contrôler $\beta(P_{p,q})$ dans un premier temps.

Corollaire 4.17 *Si $(Q)_m$ est vérifiée alors,*

$$\beta(P_{p,p+q}) \leq \prod_{k=0}^{[q/m]-1} (1 - \varepsilon_{p+km}(Q)^2).$$

Si (G) et $(M)_m$ sont vérifiées alors,

$$\beta(P_{p,p+q}) \leq \prod_{k=0}^{[q/m]-1} \left(1 - \varepsilon_{p+km}^{(m)}(G, M)\right),$$

avec $\varepsilon_p^{(m)}(G, M) = \varepsilon_p(M)^2 \varepsilon_{p+1,p+m}(G)$. Si $(M)_1$ est vérifiée alors

$$\beta(P_{p,n}) \leq \prod_{k=0}^{n-1} (1 - \varepsilon_k^2(M)).$$

Remarque 4.18 Le jeu de conditions (G) , $(M)_m$ est bien plus précis que $(Q)_m$. Par exemple, dans le cas homogène, on obtient les majorations suivantes :

$$\beta(P_{0,nm}) \leq \begin{cases} (1 - \varepsilon^{2m}(G)\varepsilon^2(M))^n & \text{sous } (Q)_m, \\ (1 - \varepsilon^{m-1}(G)\varepsilon^2(M))^n & \text{sous } (G) \text{ et } (Q)_m. \end{cases}$$

De plus, si $(M)_1$ est valide, l'estimation ne dépend plus du potentiel :

$$\beta(P_{0,n}) \leq (1 - \varepsilon^2(M))^n.$$

Il nous reste à transférer ces résultats aux propriétés de contraction aux semi-groupes non linéaires $\Phi_{p,q}$.

Proposition 4.19 Pour tous $p \in \mathbb{N}$ et $n \geq m$,

$$\|\Phi_{p,p+n}(\mu_p) - \Phi_{p,p+n}(\nu_p)\|_{vt} \leq \begin{cases} \prod_{k=0}^{[q/m]-1} (1 - \varepsilon_{p+km}(Q)^2) & \text{sous } (Q)_m, \\ \frac{2 \prod_{k=0}^{[q/m]-1} (1 - \varepsilon_{p+km}^{(m)}(G, M))}{\varepsilon_p(M)\varepsilon_{p+m}(G)} & \text{sous } (G) \text{ et } (M)_m, \\ \prod_{k=0}^{n-1} (1 - \varepsilon_k^2(M)) & \text{sous } (M)_1. \end{cases}$$

Remarque 4.20 Dans le cas homogène, sous (G) et $(M)_m$ par exemple, on obtient

$$\|\Phi_{0,nm}(\mu_0) - \Phi_{0,nm}(\nu_0)\|_{vt} \leq \frac{2(1 - \varepsilon^2(M)\varepsilon^{m-1}(G))^n}{\varepsilon(M)\varepsilon^m(G)}.$$

Si $(M)_1$ est satisfaite, G n'intervient pas dans le taux de décroissance.

Remarque 4.21 On peut encore retranscrire ces résultats pour les semi-groupes de prédiction. Ces derniers ont l'avantage de rester bien définis lorsque l'on autorise les potentiels à s'annuler. Les estimations pour les semi-groupes de prédiction peuvent être exprimées en fonction des conditions $(M)_m$, (G) etc.

4.6 Mesure invariante

On se place dans le cadre homogène : tous les triplets $(E_n, G_n, M_n)_n$ sont égaux et on oublie donc l'indice n . On dira que μ est Φ -invariante pour (G, M) si $\mu = \Phi(\mu)$.

Proposition 4.22 Pour tout potentiel $G : E \mapsto [0, +\infty[$ et toute mesure $\mu \in \mathcal{P}(E)$ tels que $\mu(G) \wedge \mu M(G) > 0$, on a

$$\mu \text{ est } \Phi\text{-invariante pour } (G, M) \implies \Psi(\mu) \text{ est } \widehat{\Phi}\text{-invariante pour } (G, M),$$

et

$$\mu \text{ est } \widehat{\Phi}\text{-invariante pour } (G, M) \implies \mu M \text{ est } \Phi\text{-invariante pour } (G, M).$$

Preuve. Si μ est Φ -invariante pour (G, M) alors

$$\widehat{\Phi}(\Psi(\mu)) = \Psi(\Psi(\mu)M) = \Psi(\Phi(\mu)) = \Psi(\mu).$$

Dans le second cas, on écrit

$$\Phi(\mu M) = \Psi(\mu M)M = \widehat{\Phi}(\mu)M = \mu M.$$

□

Supposons que $\eta = \Phi(\eta) \in \mathcal{P}(E)$ soit un point fixe de Φ et soit γ_n et η_n les mesures de Feynman-Kac non normalisée et normalisée issues de η . Alors

$$\eta_n = \Phi^n(\eta) = \eta \quad \text{et} \quad \gamma_n(f) = \mathbb{E}_\eta \left(f(X_n) \prod_{p=0}^{n-1} G(X_p) \right) = \eta(f) \eta(G)^n,$$

d'après la remarque 1.3.

Théorème 4.23 *Supposons que les conditions (G) et $(M)_m$ soient vérifiées pour un entier $m \geq 1$ et deux réels $\varepsilon(G) > 0$ et $\varepsilon(M) > 0$. Alors il existe une unique mesure η Φ -invariante pour (G, M) et, pour toute mesure $\mu \in \mathcal{P}(E)$,*

$$\|\Phi^n(\mu) - \eta\|_{VT} \leq c_1(1 - \varepsilon^2(M)\varepsilon^{m-1}(G))^{[n/m]} \|\mu - \eta\|_{VT},$$

où c_1 ne dépend que de $\varepsilon(G)$ et $\varepsilon(M)$.

5 Algorithmes de type Feynman-Kac-Metropolis-Hastings

Un des principaux problèmes de simulation en probabilités est de pour générer des réalisations d'une loi donnée. Lorsque cette mesure π est définie sur un espace S (de grande taille même fini), il est très difficile de la simuler directement. L'une des idées les plus efficaces, même si elle paraît de prime abord assez peu simplificatrice, est de construire une chaîne de Markov $(Y_n)_n$ sur S de mesure invariante π . Pour obtenir une réalisation de π (ou presque), il ne reste plus qu'à utiliser un théorème de convergence de la loi de Y_n vers π pour dire que si n est *assez grand* Y_n peut être pris pour une réalisation de π . Bien sûr, en général, on ne sait pas quantifier ce *assez grand*... Cette approche a été proposée pour la première fois par Metropolis (et ses collègues) dans un article publié dans une revue de chimie...

Très souvent la mesure π s'écrit sous la forme

$$\pi(dx) = \frac{1}{Z} e^{-V(x)} \nu(dx) \tag{19}$$

où ν est une mesure de référence sur S , V est une fonction positive sur S que l'on appellera énergie et $Z = \nu(e^{-V})$ est la constante de normalisation. On dit que π est une mesure de Gibbs.

Remarque 5.1 *Ces mesures sont très naturelles car ce sont les mesures qui maximisent l'entropie à énergie fixée. Plus précisément, on cherche les mesures sur S qui maximisent l'entropie (mathématique)*

$$-\sum_{x \in S} p(x) \ln p(x),$$

sous la contrainte $\sum_{x \in S} V(x)p(x) = \langle V \rangle$ fixé. Il apparaît que la solution est unique et s'écrit sous la forme

$$\pi_\beta(x) = \frac{1}{Z_\beta} e^{-\beta V(x)} \quad \text{avec} \quad Z_\beta = \sum_{x \in S} e^{-\beta V(x)},$$

pour un certain β bien choisi.

En pratique, il est impossible de calculer V en tout point de S et donc Z n'est pas connu. Il est souvent très petit ou très grand ce qui pose également des problèmes numériques.

L'idée est de construire un processus admettant π pour mesure invariante. Celui-ci n'est bien sûr pas unique.

5.1 Le modèle de Metropolis-Hastings

5.1.1 Le cas général

Cet algorithme permet de construire une chaîne de Markov homogène sur $(Z_n)_n = ((Y_n, Y'_n))_n$ à valeurs dans $E = S^2$ dont la première composante est une chaîne de Markov sur S de mesure invariante π . On pose

$$G(y, y') := \frac{d(\pi \times K)_2}{d(\pi \times K)_1}(y, y') \quad \text{avec} \quad \begin{cases} d(\pi \times K)_1(d(y, y')) = \pi(dy)K(y, dy'), \\ d(\pi \times K)_2(d(y, y')) = \pi(dy')K(y', dy). \end{cases}$$

La dynamique de Z se décompose en deux temps de la manière suivante :

$$Y_n \xrightarrow{\text{exploration}} Y'_n \xrightarrow{\text{sélection}} Y_{n+1}.$$

L'exploration se fait selon le noyau $K : Y'_n$ est choisi selon la mesure $K(Y_n, \cdot)$. La transition $(Y_n \rightarrow Y'_n)$ est ensuite acceptée avec probabilité $1 \wedge G(Y_n, Y'_n)$ et alors $Y_{n+1} = Y'_n$; dans le cas contraire Y_{n+1} reste en Y_n : c'est l'étape de sélection.

Le processus $(Y_n)_n$ est alors une chaîne de Markov de transition K_{MH} donnée par

$$K_{MH}(y, dy') = (1 \wedge G(y, y'))K(y, dy') + \left(1 - \int_S (1 \wedge G(y, z))K(y, dz)\right)\delta_y(dy'),$$

et de mesure invariante π .

Remarque 5.2 *Pour implémenter l'algorithme, il n'est pas nécessaire de connaître π . Il suffit de pouvoir évaluer la fonction G ce qui est possible dès que l'on connaît π à une constante multiplicative près.*

5.1.2 Le cas réversible et Gibbs

On suppose que le noyau de transition K est réversible par rapport à ν , c'est-à-dire que

$$\nu(dx)K(x, dy) = \nu(dy)K(y, dx)$$

et que la mesure π est une mesure de Gibbs de la forme (19). L'algorithme de Metropolis devient très simple. À l'étape n , la chaîne Y est en $Y_n = y$. On génère Y' suivant la loi $K(Y_n, dy')$.

- si $V(Y') \leq V(y)$ alors on pose $Y_{n+1} = Y'$.
- si $V(Y') > V(y)$, alors

$$Y_{n+1} = \begin{cases} Y' & \text{avec probabilité } e^{-(V(Y')-V(y))}, \\ y & \text{avec probabilité } 1 - e^{-(V(Y')-V(y))}. \end{cases}$$

La chaîne rejoint toujours un point d'énergie plus basse et peut également remonter à contre-courant de l'énergie avec une probabilité positive.

Proposition 5.3 *Le processus $(Y_n)_n$ ainsi construit est une chaîne de Markov de noyau de transition*

$$Q(y, dy') = e^{-[V(y')-V(y)]_+} K(y, dy').$$

De plus, π est l'unique mesure invariante de X et elle est de plus réversible.

5.2 Les modèles de Feynman-Kac-Metropolis

5.2.1 Quelques notations

On note $E = S \times S$. À un noyau markovien K sur S , on associe le noyau markovien M^K sur E défini de la façon suivante :

$$M^K((y, z), d(y', z')) := \delta_z(dy')K(y', dz').$$

En d'autres termes, si $(Y_n)_n$ est une chaîne de Markov sur S de transition K alors X définie par $X_n = (Y_n, Y_{n+1})$ est une chaîne de Markov sur E de transition M^K .

On associe à toute mesure $\pi \in \mathcal{P}(S)$ et tout noyau markovien K sur S les mesures $(\pi \times K)_1$ et $(\pi \times K)_2$ sur E définies par

$$(\pi \times K)_1 = \pi(dy)K(y, dy') \quad \text{et} \quad (\pi \times K)_2 = \pi(dy')K(\bar{y}', dy).$$

La proposition qui suit est évidente mais cruciale pour la suite.

Proposition 5.4 *Soit M un noyau markovien sur E . Pour toutes mesures μ et ν dans $\mathcal{P}(E)$, on a*

$$\mu \ll \nu \quad \text{et} \quad \mu M = \eta \implies \begin{cases} \eta \text{ est } \Phi\text{-invariante pour } \left(\frac{d\mu}{d\nu}, M \right), \\ \mu \text{ est } \widehat{\Phi}\text{-invariante pour } \left(\frac{d\mu}{d\nu}, M \right). \end{cases}$$

En particulier, pour toute mesure $\pi \in \mathcal{P}(S)$ et tous noyaux markoviens K, L sur S , on a

$$(\pi \times L)_2 \ll (\pi \times K)_1 \implies \begin{cases} (\pi \times K)_1 \text{ est } \Phi\text{-invariante pour } \left(\frac{d(\pi \times L)_2}{d(\pi \times K)_1}, M^K \right), \\ (\pi \times L)_2 \text{ est } \widehat{\Phi}\text{-invariante pour } \left(\frac{d(\pi \times L)_2}{d(\pi \times K)_1}, M^K \right). \end{cases}$$

Soit π une probabilité sur S et K et L deux noyaux markoviens sur S tels que, pour tout $y \in S$,

$$L(y, \cdot) \ll \pi \ll K(y, \cdot).$$

La proposition ci-dessus assure que les mesures $(\pi \times K)_1$ et $(\pi \times L)_2$ sont respectivement Φ - et $\widehat{\Phi}$ -invariantes pour le couple

$$(G, M) = \left(\frac{d(\pi \times L)_2}{d(\pi \times K)_1}, M^K \right).$$

5.2.2 Description de la dynamique

Le processus $(Z_n)_n = ((Y_n, Y'_n))_n$ possède à nouveau une dynamique en deux temps prescrite par les transitions (non linéaires) $K_{\eta_{n-1}} = S_{\eta_{n-1}}M^K$. La suite $(\eta_n)_n$ des lois des v.a. $(Y_n)_n$ est donnée par récurrence par la formule

$$\eta_n = \eta_{n-1}S_{\eta_{n-1}}M^K = \Psi(\eta_{n-1})M^K,$$

avec

$$S_\eta((y, y'), \cdot) = \varepsilon G(y, y')\delta_{(y, y')}(\cdot) + (1 - \varepsilon G(y, y'))\Psi(\eta)(\cdot),$$

et

$$\Psi(\eta)(d(y, y')) = \frac{1}{\eta(G)}G(y, y')\eta(d(y, y')).$$

La dynamique est donc décomposée en deux phases :

$$Y_n \xrightarrow{\text{exploration}} Y'_n \xrightarrow{\text{sélection}} Y_{n+1}.$$

L'exploration se fait toujours suivant le noyau K , comme pour l'algorithme de Metropolis. La sélection est cependant bien différente. La transition $(Y_n \rightarrow Y'_n)$ est acceptée avec probabilité $\varepsilon G(Y_n, Y'_n)$ et alors $Y_{n+1} = Y'_n$. En cas de rejet, Y_{n+1} est tirée selon la loi $\Psi(\eta_{n-1})$. Ainsi Y ne reste pas en place...

Proposition 5.5 *La mesure $(\pi \times K)_1$ définie par*

$$(\pi \times K)_1(d(y, y')) = \pi(dy)K(y, dy')$$

est Φ -invariante pour (G, M^K) .

Remarque 5.6 *En particulier, la loi de la première composante du processus convergera vers la mesure cible π .*

Remarque 5.7 (Le cas réversible et Gibbs) *On suppose à nouveau donné un noyau de transition K réversible par rapport à ν et π de type Gibbs. On peut alors choisir $L = K$. Le potentiel G s'écrit à nouveau sous une forme bien plus simple :*

$$G(y, y') = e^{-(V(y')-V(y))}.$$

5.3 Comparaison heuristique

L'avantage le plus évident de l'algorithme de Metropolis est sa simplicité et notamment sa structure homogène et linéaire. Son principal défaut vient du fait que si la chaîne Y est dans un puits d'énergie à un instant donné, elle risque d'y rester longtemps puisque la probabilité de rester sur place est alors proche de 1. Ceci a pour principale conséquence de ralentir fortement la convergence à l'équilibre vers la mesure invariante π .

Pour l'algorithme de Feynman-Kac, c'est l'inverse. Il possède une structure plus complexe et ne peut donc pas être simulé directement mais le processus ne restera pas coincé dans un minimum local de l'énergie. Pour contourner le problème de simulation, il faut bien entendu faire appel aux systèmes de particules associés à la dynamique non linéaire.

5.4 Système de particules associé

On associe naturellement au processus de Feynman-Kac le système de particules

$$(\xi_n)_n = ((Y_n^1, \dots, Y_n^N, Y_n'^1, \dots, Y_n'^N))_n$$

à valeurs dans S^{2N} dont la dynamique se décompose encore en deux étapes :

$$(Y_n^i)_{1 \leq i \leq N} \xrightarrow{\text{mutation}} (Y_n'^i)_{1 \leq i \leq N} \xrightarrow{\text{sélection}} (Y_{n+1}^i)_{1 \leq i \leq N}.$$

Durant l'étape de mutation chaque particule se déplace de manière indépendante selon le noyau K et à la position Y_n^i de la particule i on associe $Y_n'^i$ selon la loi $K(Y_n^i, \cdot)$. Cette nouvelle position est acceptée avec probabilité $\varepsilon G(Y_n^i, Y_n'^i)$, auquel cas $Y_{n+1}^i = Y_n'^i$. En cas de rejet, Y_{n+1}^i est tirée selon la loi

$$\sum_{j=1}^N \frac{G(Y_n^i, Y_n'^j)}{\sum_j G(Y_n^j, Y_n'^j)} \delta_{Y_n'^j}.$$

La particule i aura donc tendance à rejoindre une particule localisée dans une région à fort potentiel.

Remarque 5.8 *Ce système de particules peut être vu comme des algorithmes de Metropolis en interaction.*

6 Recuit simulé par Feynman-Kac-Metropolis

6.1 Le contexte

On cherche les points où une fonction V positive définie sur un ensemble S fini (mais grand) atteint son minimum (global). On sait qu'une recherche exhaustive est impossible (car S est grand), qu'un algorithme déterministe (de type plus grande pente) restera généralement bloqué dans des minima locaux, donc qu'il est intéressant de développer des algorithmes stochastiques.

À cette fin, on remarque que, si on se donne une mesure ν de référence sur S , les mesures de probabilité de Gibbs

$$\pi_\beta(x) = \frac{1}{Z_\beta} e^{-\beta V(x)} \nu(x)$$

se concentrent sur le lieu du minimum (ν -essentiel pour être précis) de V quand β tend vers l'infini. Or, les algorithmes classiques de Metropolis, ou de Feynman-Kac-Metropolis pour ce qui nous intéresse, permettent d'obtenir ces mesures π_β comme mesures invariantes de chaînes de Markov contractantes, donc de les simuler (sans calculer $Z_\beta = \sum_{x \in S} e^{-\beta V(x)} \nu(x)$, ce qui est essentiel, car ce calcul est aussi dur qu'une recherche exhaustive du minimum!).

Une méthode de recuit simulé consiste à faire tourner un algorithme de Metropolis en faisant croître β (c'est-à-dire en refroidissant le système, d'où le nom, inspiré de la sidérurgie) assez vite pour que les mesures invariantes π_β approchent effectivement le lieu du minimum, mais assez lentement pour que l'algorithme de Metropolis ait le temps d'approcher la mesure invariante. On va voir que les dynamiques non-linéaires de type Feynman-Kac-Metropolis convergent plus vite que les résultats connus avec Metropolis classique. Tout le problème, déjà mis en valeur par Arnaud, est qu'on contrôle beaucoup moins bien l'approximation de la dynamique non-linéaire par le système de particules...

Nous nous placerons dans le cadre suivant (les résultats sont valables dans un cadre plus général) :

- V est une fonction positive définie sur un ensemble S fini. On cherche le lieu du minimum de V . On notera $\omega(V) = \sup(V) - \inf(V)$ l'oscillation de V .
- ν mesure de probabilité de référence sur S et K noyau markovien sur S tel que ν est réversible pour K :

$$\nu(x)K(x, x') = \nu(x')K(x', x) \quad \forall x, x' \in S$$

On supposera aussi que K a de bonnes propriétés de contraction, c'est-à-dire qu'il existe $\varepsilon(K) > 0$ tel que

$$K(y, \cdot) \geq \varepsilon(K)K(y', \cdot) \quad \forall y, y' \in S \quad (\text{K})$$

Ceci implique en particulier que ν est la seule probabilité invariante de K et que

$$\|\mu K^n - \nu\|_{\text{vt}} \leq \beta(K^n) \|\mu - \nu\|_{\text{vt}} \quad (20)$$

avec $\beta(K^n) \leq (1 - \varepsilon(K))^n$.

- Pour tout $\beta \geq 0$

$$\pi_\beta(x) = \frac{1}{Z_\beta} e^{-\beta V(x)} \nu(x)$$

avec $Z_\beta = \sum_{x \in S} e^{-\beta V(x)} \nu(x)$.

- Sur l'espace produit $E = S^2$, on définit le noyau markovien associé à la loi des couples pour K

$$M^K((x, x'), (y, y')) = \delta_{x'}(y) K(y, y')$$

et le potentiel

$$G_\beta(x, x') = e^{-\beta(V(x') - V(x))}$$

On associe (comme d'habitude) au couple (M^K, G) deux dynamiques non-linéaires sur $\mathcal{P}(E)$: $\Phi_\beta(\mu) = \Psi_\beta(\mu)M^K$ et $\widehat{\Phi}_\beta(\mu) = \Psi_\beta(\mu M^K)$, avec $\Psi_\beta(\mu) = \frac{G_\beta}{\mu(G_\beta)}\mu$. On sait (voir proposition 5.4) que

$$\begin{cases} (\pi_\beta \times K)_1(x, x') = \pi_\beta(x)K(x, x') & \text{est } \Phi_\beta\text{-invariante,} \\ (\pi_\beta \times K)_2(x, x') = \pi_\beta(x')K(x', x) & \text{est } \widehat{\Phi}_\beta\text{-invariante.} \end{cases}$$

et on peut espérer que les résultats de contraction obtenus dans un cadre plus général s'appliqueront ici. En fait, on peut même obtenir mieux, ce sera le contenu de la section 6.2, et on l'utilisera section 6.3 pour montrer que l'algorithme de recuit-simulé par Feynman-Kac-Metropolis converge pour tout réchauffement par paliers sous-linéaire. On travaillera avec la dynamique $\widehat{\Phi}_\beta$, car elle a de meilleures propriétés de contraction et donne des mesures dont la répartition est plus directement reliée à l'énergie.

6.2 Contraction du noyau

En utilisant habilement notre cadre spécifique, en particulier son bon comportement par renversement du temps, on obtient

Théorème 6.1 *Pour tout $n \geq 1$*

$$\|\widehat{\Phi}_\beta^{n+2}(\mu) - \widehat{\Phi}_\beta^{n+2}(\nu)\|_{vt} \leq 4\varepsilon^{-3}(K)e^{\beta\omega(V)}\beta(K^n)\|\mu - \nu\|_{vt}$$

Pour simplifier les notations dans cette section, on notera $\pi = \pi_1$ et $G = G_1$, ce qui permettra de travailler sans le coefficient β , la prise en compte de ce coefficient (multiplicatif devant la fonction V) étant immédiate.

6.2.1 Renversement du temps

On note $X_n = (Y_n, Y_{n+1})$ la chaîne de Markov sur $E = S^2$ de matrice de transition M^K . Remarquons que Y_n est alors une chaîne de Markov de matrice de transition K , et que le potentiel G décrit juste le défaut de réversibilité de π pour K :

$$\begin{aligned} \pi(y_0)K(y_0, y_1)G(y_0, y_1) &= \frac{1}{Z}e^{-V(y_0)}\nu(y_0)K(y_0, y_1)e^{V(y_0)-V(y_1)} \\ &= \frac{1}{Z}e^{-V(y_1)}\nu(y_1)K(y_1, y_0) = \pi(y_1)K(y_1, y_0) \end{aligned}$$

où on a utilisé le fait que ν est K -réversible. En itérant cette formule, on obtient pour tout $n \geq 1$

$$\begin{aligned} \pi(y_0)K(y_0, y_1)K(y_1, y_2) \cdots K(y_n, y_{n+1}) \prod_{k=0}^n G(y_k, y_{k+1}) \\ = \pi(y_{n+1})K(y_{n+1}, y_n)K(y_n, y_{n-1}) \cdots K(y_1, y_0) \end{aligned}$$

En sommant/intégrant cette formule, on obtient pour toute fonction φ mesurable bornée sur S^{n+2}

$$\mathbb{E}_{(\pi \times K)_1} \left(\varphi(Y_0, Y_1, \dots, Y_{n+1}) \prod_{k=0}^n G(Y_k, Y_{k+1}) \right) = \mathbb{E}_{(\pi \times K)_1} (\varphi(Y_{n+1}, Y_n, \dots, Y_0))$$

puis

Lemme 6.2 Pour tous $a, b \in S$ et pour toute fonction φ mesurable bornée sur S^{n+2}

$$\begin{aligned} \mathbb{E}_{(a,b)} \left(\varphi(Y_0, Y_1, \dots, Y_{n+1}) \prod_{k=0}^n G(Y_k, Y_{k+1}) \right) \\ = \frac{d\pi K^n}{d\pi}(b) \mathbb{E}_{(\pi \times K)_1} (\varphi(Y_{n+1}, Y_n, \dots, Y_0) \mid (Y_{n+1}, Y_n) = (a, b)) \end{aligned} \quad (21)$$

Preuve. On procède comme pour l'identité précédente :

$$\begin{aligned} \mathbb{E}_{(a,b)} \left(\varphi(Y_0, Y_1, \dots, Y_{n+1}) \prod_{k=0}^n G(Y_k, Y_{k+1}) \right) \\ = \int K(b, dy_2) \prod_{p=2}^n K(y_p, dy_{p+1}) G(b, y_2) \prod_{p=2}^n G(y_p, y_{p+1}) \varphi(a, b, y_2, \dots, y_{n+1}) \\ = \frac{1}{\pi(b)} \int \pi(dy_{n+1}) \prod_{p=n}^2 K(y_{p+1}, dy_p) K(y_2, b) \varphi(a, b, y_2, \dots, y_{n+1}) \\ = \frac{1}{\pi(b)K(b, a)} \mathbb{E}_{(\pi \times K)_1} (\varphi(Y_{n+1}, Y_n, \dots, Y_0) 1_{\{(Y_{n+1}, Y_n) = (a, b)\}}) \\ = \frac{d\pi K^n}{d\pi}(b) \mathbb{E}_{(\pi \times K)_1} (\varphi(Y_{n+1}, Y_n, \dots, Y_0) \mid (Y_{n+1}, Y_n) = (a, b)) \end{aligned}$$

car $\mathbb{E}_{(\pi \times K)_1} ((Y_{n+1}, Y_n) = (a, b)) = \pi K^n(b)K(b, a)$. □

6.2.2 Propriété de contraction

On avait vu (théorème 4.11) que

$$\|\widehat{\Phi}^n(\eta) - \widehat{\Phi}^n(\mu)\|_{\text{vt}} \leq 2\beta(\widehat{P}^n) \sup_{x, x'} \frac{\widehat{G}^n(x)}{\widehat{G}^n(x')} \|\eta - \mu\|_{\text{vt}} \quad (22)$$

avec

$$\begin{aligned} \widehat{Q}^n f(a, b) &= \mathbb{E}_{(a,b)} \left(f(X_n) \prod_{k=0}^n G(Y_k, Y_{k+1}) \right) \\ &= \frac{d\pi K^n}{d\pi}(b) \mathbb{E}_{(\pi \times K)_1} (f(Y_1, Y_0) \mid Y_n = b) \\ \widehat{P}^n f(a, b) &= \frac{\widehat{Q}^n f}{\widehat{Q}^n 1} = \mathbb{E}_{(\pi \times K)_1} (f(Y_1, Y_0) \mid Y_n = b) \\ \widehat{G}^n(a, b) &= \widehat{Q}^n 1 = \frac{d\pi K^n}{d\pi}(b) \end{aligned} \quad (23)$$

Ces expressions nous permettent de démontrer la

Proposition 6.3 Sous l'hypothèse de contraction (K), on a pour tout $n \geq 1$

$$\begin{aligned} \beta(\widehat{P}^{n+2}) &\leq 2\varepsilon^{-1}(K)\beta(K^n) \\ \widehat{G}^{n+2}(x) &\geq \varepsilon^2(K)e^{-\omega(V)}\widehat{G}^{n+2}(x') \end{aligned}$$

Il suffit alors d'injecter ces estimations dans la formule (22) pour démontrer le théorème 6.1. Pour démontrer la proposition 6.3, on utilise le

Lemme 6.4 Soit $\mu \in \mathcal{P}(S)$, M_1 et M_2 deux noyaux markoviens tels que

$$M_2(y, \cdot) \ll \mu M_1 M_2 \ll M_1(y, \cdot)$$

Alors

$$\left| \frac{dM_1 M_2(y, \cdot)}{d\mu M_1 M_2}(y') - 1 \right| \leq 2\beta(M_1) \sup_{z, z' \in S} \frac{dM_2(z, \cdot)}{dM_2(z', \cdot)}(y')$$

Preuve.[Démonstration du lemme 6.4] On peut en effet écrire

$$\begin{aligned} \left| \frac{dM_1 M_2(y, \cdot)}{d\mu M_1 M_2}(z) - 1 \right| &= \left| \int_S \mu(dy') (M_1(y, dy'') - M_1(y', dy'')) \frac{dM_2(y'', \cdot)}{d\mu M_1 M_2}(z) \right| \\ &\leq 2\beta(M_1) \sup_{y'' \in S} \frac{dM_2(y'', \cdot)}{d\mu M_1 M_2}(z) \end{aligned}$$

et

$$\frac{d\mu M_1 M_2}{dM_2(y'', \cdot)}(z) = \int_S \mu M_1(dy) \frac{dM_2(y, \cdot)}{dM_2(y'', \cdot)}(z) \geq \inf_{z, z' \in S} \frac{dM_2(z', \cdot)}{dM_2(z, \cdot)}(z)$$

□

Preuve.[Démonstration de la proposition 6.3] On peut récrire l'expression (23) comme

$$\begin{aligned} \widehat{P}^{n+2} f(a, b) &= \mathbb{E}_{(\pi \times K)_1}(f(Y_1, Y_0) \mid Y_{n+2} = b) \\ &= \frac{1}{\pi K K^{n+1}(b)} \int \pi(dy_0) K(y_0, dy_1) f(y_1, y_0) K^{n+1}(y_1, b) \\ &= \int \pi(dy_0) K(y_0, dy_1) \frac{dK^{n+1}(y_1, \cdot)}{d(\pi K) K^{n+1}}(b) f(y_1, y_0) \end{aligned}$$

On peut donc appliquer le lemme 6.4 avec $M_1 = K^n$, $M_2 = K$ et $\mu = \pi K$ pour obtenir (si $\|f\|_\infty \leq 1$)

$$\begin{aligned} |\widehat{P}^{n+2} f(a, b) - (\pi \times K)_2| &\leq \int \pi K(dy_1) \left| \frac{dK^{n+1}(y_1, \cdot)}{d(\pi K) K^{n+1}}(b) - 1 \right| \\ &\leq 2\beta(K^n) \sup_{z, z' \in S} \frac{dK(z, \cdot)}{dK(z', \cdot)}(b) \leq 2\varepsilon^{-1}(K)\beta(K^n) \end{aligned}$$

La preuve de l'estimation pour \widehat{G} est assez simple, on la trouvera p. 184 de Del Moral. □

6.3 Convergence du recuit simulé

Pour définir l'algorithme de recuit-simulé, il reste à préciser quelle température on impose au système à chaque étape. Ce sera fait par une fonction croissante $\beta : \mathbb{N}^* \rightarrow \mathbb{R}_+$. Une fois cette fonction fixée, on obtient donc la dynamique non-linéaire de recuit-simulé par itération

$$\widehat{\eta}_n = \widehat{\Phi}_{\beta(n)}(\widehat{\eta}_{n-1})$$

et la question cruciale est de savoir à quelle condition sur la fonction β (le schéma de température) on aura

$$\lim_{n \rightarrow \infty} \|\widehat{\eta}_n - (\pi_{\beta(n)} \times K)_2\|_{\text{vt}} = 0.$$

Les estimations pour cette convergence se décomposent en deux parties indépendantes :

- le temps de relaxation de $\widehat{\Phi}_\beta$ vers $(\pi_\beta \times K)_2$, à β fixé ;
- les oscillations de la fonction $\beta \mapsto (\pi_\beta \times K)_2$.

6.3.1 Propriétés de régularité

On utilisera le lemme

Lemme 6.5 *Pour toutes λ mesure positive et U fonction positive mesurable sur (E, \mathcal{E}) , on définit*

$$\mu_U(dx) = \frac{1}{Z_U} e^{-U(x)} \lambda(dx) \quad \text{avec } Z_U = \lambda(e^{-U}).$$

On a alors pour toutes fonctions U_1, U_2 positives mesurables sur (E, \mathcal{E})

$$\|\mu_{U_1} - \mu_{U_2}\|_{vt} \leq \frac{1}{2} \omega(U_1 - U_2) \quad (24)$$

Preuve. On écrit

$$\begin{aligned} \mu_{U_1} &= \frac{1}{Z_{U_1}} e^{-U_1} \lambda = \frac{1}{Z_{U_1}} e^{(U_2 - U_1)/2} e^{-(U_1 + U_2)/2} \lambda \\ &= \frac{Z_{(U_1 + U_2)/2}}{Z_{U_1}} e^{(U_2 - U_1)/2} \mu_{(U_1 + U_2)/2} \end{aligned}$$

puis on remarque que

$$\frac{Z_{(U_1 + U_2)/2}}{Z_{U_1}} = \mu_{U_1} \left(e^{-(U_2 - U_1)/2} \right) \geq e^{-\sup(U_2 - U_1)/2}$$

donc

$$\mu_1(A) \geq \mu_{(U_1 + U_2)/2}(A) e^{-\omega(U_2 - U_1)/2}$$

On peut obtenir la même chose avec μ_2 à la place de μ_1 , soit le fait que la mesure $\mu(A) = \mu_{(U_1 + U_2)/2}(A) e^{-\omega(U_2 - U_1)/2}$ vérifie pour tout $A \in \mathcal{E}$

$$\mu(A) \leq \mu_1(A) \wedge \mu_2(A).$$

On peut alors conclure par une propriété classique de la distance en variation totale (identité (4.5) du lemme 4.2.2 dans Del Moral), dont je reproduis l'argument ici : on sait (décomposition de Hahn-Jordan) qu'il existe deux mesure positives μ^+ et μ^- et deux ensembles disjoints E_+ et E_- de réunion E tels que $\mu_1 - \mu_2 = \mu^+ - \mu^-$ et $\mu^+(E_-) = \mu^-(E_+) = 0$, et qu'on a alors

$$\begin{aligned} \|\mu_1 - \mu_2\|_{vt} &= \mu(E_+) \\ &= \mu_1(E_+) - \mu_2(E_+) \\ &= 1 - \mu_1(E_-) - \mu_2(E_+) \\ &\leq 1 - \mu(E_-) - \mu(E_+) = 1 - \mu(E) \\ &= 1 - e^{-\omega(U_2 - U_1)/2} \leq \frac{\omega(U_2 - U_1)}{2} \end{aligned}$$

□

On obtient alors un contrôle sur chacune des deux parties de la convergence :

Proposition 6.6

1. Si $\beta_1 \leq \beta_2$,

$$\|(\pi_{\beta_1} \times K)_2 - (\pi_{\beta_2} \times K)_2\|_{vt} \leq \frac{1}{2} (\beta_2 - \beta_1) \omega(V). \quad (25)$$

2. Il existe $C_1(K)$ et $C_2(K)$ (ne dépendant que des propriétés spectrales du noyau K) telles pour tout $\beta \in \mathbb{R}_+$ et toutes probabilités μ et ν sur E , on a pour $\Delta = \Delta(\beta) = C_1(K) + C_2(K)\beta\omega(V)$

$$\|\widehat{\Phi}_\beta^\Delta(\mu) - \widehat{\Phi}_\beta^\Delta(\nu)\|_{vt} \leq \frac{1}{e} \|\mu - \nu\|_{vt}. \quad (26)$$

Preuve. Le point 1. se déduit immédiatement du lemme 6.5, une fois remarqué que $(\pi_\beta \times K)_2$ est une mesure de Gibbs sur $E = S^2$,

$$(\pi_\beta \times K)_2 = \frac{1}{\lambda(e^{-\beta U})} e^{-\beta U} \lambda \quad \text{avec } U(y, y') = V(y') \quad \text{et } \lambda = (\nu \times K)_2.$$

Pour le point 2. on applique le théorème 6.1 pour obtenir

$$\|\widehat{\Phi}_\beta^{n+2}(\mu) - \widehat{\Phi}_\beta^{n+2}(\nu)\|_{vt} \leq 4\varepsilon^{-3}(K)e^{\beta\omega(V)}\beta(K)^n \|\mu - \nu\|_{vt}$$

et on remarque que

$$\begin{aligned} 4\varepsilon^{-3}(K)e^{\beta\omega(V)}\beta(K)^n &\leq \frac{1}{e} \\ \Leftrightarrow 1 + \beta\omega(V) + n \ln \beta(K) + \ln(4\varepsilon^{-3}(K)) &\leq 0 \\ \Leftrightarrow n &\geq \frac{-1}{\ln(\beta(K))} (1 - \ln(\varepsilon^3(K)/4) + \beta\omega(V)) \end{aligned}$$

ce qui donne le résultat souhaité avec $C_2(K) = -1/\ln(\beta(K))$ et $C_1(K) = 2 - (1 - \ln(\varepsilon^3(K)/4))/\ln(\beta(K))$. \square

6.3.2 Comportement asymptotique

On peut désormais définir de manière précise le schéma de température à suivre. On choisit une suite croissante $\beta' : \mathbb{N} \rightarrow \mathbb{R}_+$, à laquelle on associe une suite de temps de relaxation définis par

$$\begin{aligned} t(0) &= 0 \\ t(n+1) &= t(n) + \Delta(\beta'(n)) \end{aligned}$$

puis la vraie suite de températures utilisées, qui sera constante par paliers, $\beta : \mathbb{N}^* \rightarrow \mathbb{R}_+$ définie par $\beta(p) = \beta'(n)$ tant que $t(n) < p \leq t(n+1)$.

La suite $\widehat{\eta}_p$ vérifie donc $\widehat{\eta}_p = \widehat{\Phi}_{\beta'(n)}(\widehat{\eta}_{p-1})$ tant que $t(n) < p \leq t(n+1)$, soit aussi, aux temps $t(n)$,

$$\widehat{\eta}_{t(n+1)} = \widehat{\Phi}_{\beta'(n)}^{\Delta(\beta'(n))}(\widehat{\eta}_{t(n)})$$

On a alors

Théorème 6.7 *Pour toute fonction β' vérifiant*

$$\lim_{n \rightarrow \infty} (\beta'(n+1) - \beta'(n)) = 0$$

on a

$$\lim_{n \rightarrow \infty} \|\widehat{\eta}_{t(n)} - (\pi_{\beta'(n)} \times K)_2\|_{vt} = 0.$$

En particulier, pour $\beta'(n) = (n+1)^a$, avec $0 < a < 1$,

$$\begin{aligned} t(n) &= O(n^{1+a}) \\ (n+1)^{1-a} \|\widehat{\eta}_{t(n)} - (\pi_{\beta'(n)} \times K)_2\|_{vt} &\leq \frac{1}{e} + \frac{7a}{2}\omega(V) \end{aligned}$$

Preuve. On décompose

$$\begin{aligned} \widehat{\eta}_{t(n+1)} - (\pi_{\beta'(n+1)} \times K)_2 &= \left[\widehat{\Phi}_{\beta'(n)}^{\Delta(\beta'(n))} (\widehat{\eta}_{t(n)}) - (\pi_{\beta'(n)} \times K)_2 \right] \\ &\quad + \left[(\pi_{\beta'(n)} \times K)_2 - (\pi_{\beta'(n+1)} \times K)_2 \right] \end{aligned}$$

ce qui permet, après avoir rappelé que $(\pi_{\beta'(n)} \times K)_2$ est invariante par $\widehat{\Phi}_{\beta'(n)}$, d'appliquer les deux estimations de la proposition 6.6 pour obtenir que la suite $I_n = \|\widehat{\eta}_{t(n)} - (\pi_{\beta'(n)} \times K)_2\|_{\text{vt}}$ vérifie pour tout $n \geq 1$

$$I_{n+1} \leq \frac{I_n}{e} + \frac{1}{2}(\beta'(n+1) - \beta'(n))\omega(V)$$

ce qui implique

$$e^{n+1}I_{n+1} \leq 1 + \frac{e}{2}\omega(V) \sum_{p=0}^n e^p(\beta'(p+1) - \beta'(p))$$

soit aussi

$$I_{n+1} \leq e^{-n-1} + \frac{\omega(V)}{2} \frac{e - e^n \sum_{p=0}^n e^p(\beta'(p+1) - \beta'(p))}{e - 1 \sum_{p=0}^n e^p}$$

où le majorant tend vers 0 quand n tend vers l'infini, par l'hypothèse et le lemme suivant (qui est un lemme de Césaro à coefficients variables) :

Lemme 6.8 (Toeplitz-Kronecker) *Soit (a_n) une suite réelle positive telle que $\sum_{n \geq 0} a_n = +\infty$ et (b_n) telle que $\lim_{n \rightarrow \infty} b_n = b$. alors*

$$\lim_{n \rightarrow \infty} \frac{\sum_{p=0}^n a_p b_p}{\sum_{p=0}^n a_p} = b$$

Le cas particulier s'étudie par une série de calculs. L'inégalité de concavité $(p+1)^a - p^a \leq ap^{a-1}$ donne

$$\sum_{p=1}^n e^p((p+1)^a - p^a) \leq a \sum_{p=1}^n \frac{e^p}{p^{1-a}} \leq ae \left(1 + \sum_{p=2}^n \frac{e^{p-1}}{p^{1-a}} \right)$$

Or, pour $p \geq 2$, $p/(p-1) \leq 2$ donc

$$\frac{e^{p-2}}{(p-1)^{1-a}} = \frac{1}{e} \frac{e^{p-1}}{p^{1-a}} \left(\frac{p}{p-1} \right)^{1-a} \leq \frac{2^{1-a}}{e} \frac{e^{p-1}}{p^{1-a}} \leq \frac{2}{e} \frac{e^{p-1}}{p^{1-a}}$$

ce qui s'écrit aussi

$$\left(1 - \frac{2}{e} \right) \frac{e^{p-1}}{p^{1-a}} \leq \frac{e^{p-1}}{p^{1-a}} - \frac{e^{p-2}}{(p-1)^{1-a}}$$

soit enfin

$$\sum_{p=2}^n \frac{e^{p-1}}{p^{1-a}} \leq \left(1 - \frac{2}{e} \right)^{-1} \sum_{p=2}^n \left(\frac{e^{p-1}}{p^{1-a}} - \frac{e^{p-2}}{(p-1)^{1-a}} \right) = \left(1 - \frac{2}{e} \right)^{-1} \left(\frac{e^{n-1}}{n^{1-a}} - 1 \right) \leq \frac{2e^n}{n^{1-a}}.$$

Cette majoration donne pour l'estimation de contraction (avec $ne^{-n} \leq 1$)

$$\begin{aligned} e^{n+1}I_{n+1} &\leq 1 + \frac{1}{2}\omega(V) \sum_{p=1}^n e^p((p+1)^a - p^a) \\ &\leq 1 + \frac{ae}{2}\omega(V) + ae\omega(V) \frac{e^{n+1}}{(n+1)^{1-a}} \end{aligned}$$

soit aussi

$$(n+1)^{1-a}I_{n+1} \leq \frac{1}{e} + a\left(e + \frac{1}{2}\right)\omega(V) \leq \frac{1}{e} + \frac{7a}{2}\omega(V).$$

Comme $\Delta(\beta'(n)) = O(n^a)$, on a

$$t(n) = \sum_{p=0}^{n-1} \Delta(\beta'(p)) = O(n^{1+a}).$$

Cette estimation est cruciale, dans la mesure où elle contrôle le temps effectif pendant lequel l'algorithme doit tourner pour donner le résultat annoncé. \square

7 Applications

Cette section présente quelques problèmes concrets (filtrage, estimation d'événements rares, ...) qui se ramènent à l'étude de formules de Feynman-Kac.

7.1 Filtrage et modèle de Markov caché

On modélise

- les états cachés par une suite $(X_n)_{n \geq 0}$ telle que $X_n \in E_n$,
- les observations par une suite $(Y_n)_{n \geq 0}$ telle que $Y_n \in F_n$.

On suppose que

- $(X_n)_{n \geq 0}$ est une chaîne de Markov de transitions $(M_n)_n$ et de loi initiale η_0 ,
- le canal est sans mémoire : conditionnellement aux états cachés, les observations sont indépendantes :

$$\mathbb{P}(Y_{0:n} \in dy_{0:n} | X_{0:n} = x_{0:n}) = \prod_{k=0}^n \mathbb{P}(Y_k \in dy_k | X_{0:n} = x_{0:n}).$$

- l'observation k ne dépend que de l'état de la chaîne caché à l'instant k et il existe une mesure λ_k sur F_k telle que, pour tout $x_k \in E_k$,

$$\mathbb{P}(Y_k \in dy_k | X_{0:n} = x_{0:n}) = \mathbb{P}(Y_k \in dy_k | X_k = x_k) = g_k(x_k, y_k)\lambda_k(dy_k).$$

Les quantités $\mathbb{P}(Y_k \in dy_k | X_k = x_k)$ sont appelées probabilités d'émission.

Cette dernière hypothèse est assez forte : les lois conditionnelles sont toutes absolument continues par rapport à une même mesure.

Remarque 7.1 *Un des cas les plus simples est le filtre de Kalman-Bucy. On se donne des variables aléatoires $(U_n)_n, (V_n)_n$ i.i.d. de loi $\mathcal{N}(0, 1)$. Le processus complet est alors défini par*

$$\begin{cases} X_{n+1} = aX_n + \sigma U_{n+1}, \\ Y_n = bX_n + \alpha V_n. \end{cases}$$

La v.a. X_n modélise l'écart entre la position théorique d'un avion (donnée par le plan de vol) et sa véritable position au temps n . Le réel a traduit le comportement du pilote. Si $0 \leq a < 1$, il corrige les erreurs, si $-1 < a < 0$ il a tendance à sur-corriger (c'est sûrement un apprenti) et si $|a| \geq 1$, on peut se demander s'il y a un pilote dans l'avion... La suite $(X_n)_n$ n'est rien d'autre qu'un processus auto-régressif d'ordre 1. La seconde équation modélise les erreurs de mesure du radar qui observe l'avion. Dans ce cas très particulier, la loi de X_n conditionnellement aux observations $Y_{0:n}$ est encore une gaussienne dont on peut calculer les paramètres. On n'a donc pas besoin de ce qui suit...

Définition 7.2 Les quantités $\mathbb{P}(Y_k \in dy_k | X_k = x_k)$ sont appelées probabilités d'émission. On appelle fonctions de vraisemblance les fonctions $(G_k)_k$ définies par

$$G_k : x \in E_k \mapsto G_k(x) := g_k(x, Y_k).$$

Exemple 2 On suppose que $Y_k = h_k(X_k) + V_k \in \mathbb{R}^d$ où $(V_k)_k$ est une suite de v.a. i.i.d. de lois respectives $q_k(v) \lambda(dv)$ indépendantes de $(X_n)_n$. On a ainsi,

$$\mathbb{P}(Y_k \in dy | X_k = x) = q_k(y - h_k(x)) dy \quad \text{et} \quad G_k(x) = q_k(y - h_k(x)).$$

Dans le cas particulier où $\mathcal{L}(V_k) = \mathcal{N}(0, 1)$, $G_k(x)$ est proportionnel à $\exp(-|Y_k - h_k(x)|^2/2)$ et va donc favoriser les x tels que $h_k(x)$ est proche de Y_k .

Les deux questions auxquelles on doit répondre (et qui se ramènent à des formules de Feynman-Kac) sont les suivantes :

- on souhaite connaître la loi de X_n sachant $Y_{0:n}$,
- on veut calculer la vraisemblance du modèle (M, G) , c'est-à-dire la fonction des observations suivantes :

$$l_n^{(M,G)} := L_n^{(M,G)} \quad \text{où} \quad \mathbb{P}_{(M,G)}(Y_{0:n} \in dy_{0:n}) = L_n^{(M,G)}(y_{0:n}) \lambda_0(dy_0) \dots \lambda_n(dy_n).$$

Proposition 7.3 Avec les notations précédentes,

$$l_n^{(M,G)} = \hat{\gamma}_n(1) \quad \text{et} \quad \mathbb{P}(X_n \in dx | Y_{0:n}) = \hat{\eta}_n.$$

Tout repose sur la formule de Bayes et la structure markovienne du modèle. Calculons tout d'abord la loi du processus complet $(X_0, \dots, X_n, Y_0, \dots, Y_n)$:

$$\begin{aligned} \mathbb{P}(X_{0:n} \in dx_{0:n}, Y_{0:n} \in dy_{0:n}) &= \mathbb{P}(Y_{0:n} \in dy_{0:n} | X_{0:n} \in dx_{0:n}) \mathbb{P}(X_{0:n} \in dx_{0:n}) \\ &= \prod_{k=0}^n g_k(x_k, y_k) \mathbb{P}(X_{0:n} \in dx_{0:n}) \lambda_0(dy_0) \dots \lambda_n(dy_n). \end{aligned}$$

La loi jointe des observations est obtenue en prenant l'espérance par rapport à la chaîne X :

$$\mathbb{P}(Y_{0:n} \in dy_{0:n}) = \mathbb{E} \left(\prod_{k=0}^n g_k(X_k, y_k) \right) \lambda_0(dy_0) \dots \lambda_n(dy_n).$$

Ainsi, nous avons déterminé $L_n^{(M,G)}$:

$$L_n^{(M,G)}(y_{0:n}) = \mathbb{E} \left(\prod_{k=0}^n g_k(X_k, y_k) \right).$$

Il ne reste plus qu'à évaluer cette fonction en $Y_{0:n}$ pour obtenir :

$$l_n^{(M,G)} = \mathbb{E} \left(\prod_{k=0}^n G_k(X_k) \right) = \hat{\gamma}_n(1).$$

Remarque 7.4 Attention, G_k dépend de Y_k mais l'espérance doit se comprendre à $Y_{0:n}$ fixé.

On peut à présent retourner le conditionnement pour obtenir la loi de $X_{0:n}$ sachant $Y_{0:n}$:

$$\mathbb{P}(X_{0:n} \in dx_{0:n} | Y_{0:n} \in dy_{0:n}) = \frac{\prod_{k=0}^n g_k(X_k, y_k) \mathbb{P}(X_{0:n} \in dx_{0:n})}{\mathbb{E}(\prod_{k=0}^n g_k(X_k, y_k))}.$$

Si seule nous intéresse la loi de X_n sachant $Y_{0:n}$, on écrit pour toute fonction $f \in \mathcal{B}_b(E_n)$,

$$\mathbb{E}(f(X_n) | Y_{0:n} = y_{0:n}) = \frac{\mathbb{E}(f(X_n) \prod_{k=0}^n g_k(X_k, y_k))}{\mathbb{E}(\prod_{k=0}^n g_k(X_k, y_k))}.$$

On en déduit donc que $\mathbb{E}(f(X_n) | Y_{0:n}) = \hat{\eta}_n(f)$.

7.2 Chaîne de Markov contrainte ou conditionnelle

On se donne une chaîne de Markov $(X_n)_n$ comme précédemment et une suite $(A_n)_n$ d'ensembles avec $A_n \subset E_n$. On souhaite estimer

$$P_n = \mathbb{P}(X_0 \in A_0, \dots, X_n \in A_n) \quad \text{et} \quad \mathbb{P}(X_n \in dx | X_0 \in A_0, \dots, X_n \in A_n).$$

Ce problème se met sous la forme Feynman-Kac en posant $G_k(x) = \mathbf{1}_{A_k}(x)$ et on a immédiatement

$$P_n = \hat{\gamma}_n(1) \quad \text{et} \quad \mathbb{E}(f(X_n) | X_0 \in A_0, \dots, X_n \in A_n) = \hat{\eta}_n(f).$$

Si la reformulation du problème en termes de formule de Feynman-Kac est évidente, l'étude théorique de ce cadre dégénéré (G_n peu s'annuler) demande une attention soutenue. Il faut par exemple postuler que la chaîne pourra bien atteindre l'ensemble A_{n+1} depuis A_n . Le problème est encore plus délicat pour le système de particules. En effet, il se peut par exemple que toutes les particules meurent (*i.e.* sortent de l'ensemble A_n) en même temps. Il faut donc mettre en place des contrôles de la durée de vie du système de particules ou encore proposer d'autres algorithmes plus malins.

7.3 Simulation d'événements rares

Soit $(X_t)_t$ un processus de Markov fort à valeurs dans S . On considère un ensemble $B \subset S$ dit critique et on note

$$T_B := \inf \{t \geq 0, X_t \in B\},$$

son temps d'atteinte. Le problème est de proposer une méthode d'estimation de

$$P_B := \mathbb{P}(T_B \leq T)$$

où T est le temps de retour de X dans un ensemble récurrent.

Exemple 3 *On considère une file d'attente $M/M/s/k$ (temps d'arrivée et de service exponentiels, s serveurs et k places dans la file) où les clients sont servis plus vite qu'ils n'arrivent. On se demande alors par exemple qu'elle est la probabilité que la file soit pleine (pas souvent) avant qu'elle ne soit vide (0 est un état récurrent).*

On considère une suite finie d'ensembles imbriqués $(B_l)_{0 \leq l \leq L}$ tels que

$$B = B_L \subset B_{L-1} \subset \dots \subset B_1 \subset B_0 = S,$$

et on note

$$T_l = \inf \{t \geq 0, X_t \in B_l\}$$

le temps d'atteinte de l'ensemble B_l . Par définition, la suite $(T_l)_l$ est croissante. L'idée est d'utiliser la formule de Bayes et d'écrire

$$P_B = \prod_{l=0}^L p_l \quad \text{avec} \quad p_l = \mathbb{P}(T_l \leq T | T_{l-1} \leq T) \quad \text{pour } l \geq 1 \text{ et } p_0 = 1.$$

Pour $l = 1, \dots, L$, on introduit la l -ième excursion :

$$\mathcal{X}_l := (X_t, T_{l-1} \wedge T \leq t \leq T_l \wedge T).$$

Si X a atteint B_{l-1} avant R , \mathcal{X}_l est le bout de trajectoire de X entre l'instant où il est sorti de B_{l-1} et l'instant où il est entré dans $B_l \cup R$, c'est un élément de l'ensemble

$$E := \bigcup_{0 \leq s \leq t} \mathcal{D}([s, t], S)$$

où $\mathcal{D}([s, t], S)$ désigne les fonctions càdlàg de $[s, t]$ dans S . Il faut noter que la suite $(\mathcal{X}_l)_{1 \leq l \leq L}$ est une chaîne de Markov inhomogène (sur un espace un peu compliqué) dont on notera $(M_l)_l$ les noyaux définis par

$$M_l(e, de') = \mathbb{P}(\mathcal{X}_l \in de' | \mathcal{X}_{l-1} = e)$$

La propriété de Markov forte, la probabilité conditionnelle ci-dessus ne dépend que du point terminal de l'excursion e (et pas de e tout entier). Pour une excursion $e = (x_t, t' \leq t \leq t'') \in E$, notons $\pi(e) = x_{t''}$ son extrémité terminale et

$$G_l(e) = \mathbf{1}_{\{\pi(e) \in B_l\}}.$$

On a alors par définition

$$G_l(\mathcal{X}_l) = 1 \quad \text{ssi} \quad \pi(\mathcal{X}_l) \in B_l \quad \text{ssi} \quad X_{T_l \wedge T} \in B_l \quad \text{ssi} \quad T_l \leq T$$

On dispose d'un processus de Markov \mathcal{X} et d'une famille de potentiels $(G_l)_l$. On peut leur associer les mesures de prédiction non normalisée :

$$\hat{\gamma}_l(f) = \mathbb{E} \left(f(\mathcal{X}_l) \prod_{k=0}^l G_k(\mathcal{X}_k) \right) = \mathbb{E}(f(\mathcal{X}_l) \mathbf{1}_{\{T_l \leq T\}}),$$

et normalisée :

$$\hat{\eta}_l = \frac{\hat{\gamma}_l(f)}{\hat{\gamma}_l(1)} = \mathbb{E}(f(\mathcal{X}_l) | T_l \leq T).$$

La fonction f dépend *a priori* de toute la l -ième excursion. Dans le cas particulier (intéressant en pratique) où elle ne dépend que de sa valeur terminale, elle s'écrit $f = \varphi \circ \pi$ avec φ définie sur S et

$$\hat{\eta}_l(f) = \hat{\eta}_l(\varphi \circ \pi) = \mathbb{E}(\varphi(X_{T_l}) \mathbf{1}_{\{T_l \leq T\}}).$$

On peut aussi rappeler la définition des mesures de correction non normalisée :

$$\gamma_l(f) = \mathbb{E} \left(f(\mathcal{X}_l) \prod_{k=0}^{l-1} G_k(\mathcal{X}_k) \right) = \mathbb{E}(f(\mathcal{X}_l) \mathbf{1}_{\{T_{l-1} \leq T\}}),$$

et normalisée :

$$\eta_l = \frac{\gamma_l(f)}{\gamma_l(1)} = \mathbb{E}(f(\mathcal{X}_l) | T_{l-1} \leq T).$$

En particulier, $\eta_l(G_l) = \mathbb{P}(T_l \leq T | T_{l-1} \leq T) = p_l$. On retrouve ici la formule (1) :

$$\hat{\gamma}_l(1) = \mathbb{P}(T_l \leq T) = \prod_{k=0}^l p_k = \prod_{k=0}^l \eta_k(G_k).$$

On souhaite à présent non plus seulement estimer P_B mais également se faire une idée des trajectoires typiques qui permettent d'atteindre B avant R . Pour cela, on fait recoller les excursions pour créer des trajectoires entières. On note

$$\mathcal{X}_l^0 = \mathcal{X}_{0:l} := (\mathcal{X}_0, \mathcal{X}_1, \dots, \mathcal{X}_l) = (X_t, 0 \leq t \leq T_l \wedge T)$$

la trajectoire jusqu'au temps $T_l \wedge T$. Il faut remarquer que $(\mathcal{X}_{0:l})_{l \leq L}$ est une chaîne de Markov de noyaux $(M_{0:l})_l$ définis par

$$M_{0:l}((e_0, \dots, e_{l-1}), de'_0, \dots, de'_l) = \delta_0(de'_0) \dots \delta_{l-1}(de'_{l-1}) M_l(e'_{l-1}, de'_l).$$

En effet, on ne peut passer d'une trajectoire jusqu'au temps T_{l-1} à une trajectoire jusqu'au temps T_l que si elles coïncident jusqu'au temps T_{l-1} . La propriété de Markov pour $(\mathcal{X}_l)_l$ assure de plus que le choix de l'excursion à rajouter ne dépend que de celle qui la précède. On définit ensuite les fonctions de sélection :

$$G_{0:l}(e_0, \dots, e_l) := G_l(e_l) = \mathbf{1}_{\{\pi(e_l) \in B_l\}}.$$

On peut alors construire les suites de mesures $\widehat{\gamma}_{0:l}, \widehat{\eta}_{0:n}$, etc.

Pour estimer les formules de Feynman-Kac, il faut introduire un système de particules. Les fonctions de sélection ne prennent que deux valeurs (faciles à évaluer) mais peuvent s'annuler. Ceci pose des problèmes de durée de vie du système et de perte de diversité génétique.

La dynamique du système de particules est construite à partir de celle d'un processus (non linéaire) dont les lois marginales sont les mesures $(\eta_n)_n$. Il existe de multiples dynamiques conduisant à ce résultat (voir les sections 1.4 et 1.5). La plus simple s'obtient de la façon suivante : pour passer du temps $l-1$ au temps l on réalise successivement les étapes de

- sélection : générer un N -échantillon $(\widehat{\xi}_{l-1}^1, \dots, \widehat{\xi}_{l-1}^N)$ selon la loi $\widehat{\eta}_{l-1}^N$,
- mutation : générer N excursions indépendantes issues des points $(\widehat{\xi}_{l-1}^1, \dots, \widehat{\xi}_{l-1}^N)$ jusqu'au temps $T_l \wedge T$,
- pondération : tuer les particules qui ont touché R avant d'atteindre B_l et affecter à chaque particule encore en vie un poids identique (égal à l'inverse du nombre de particules en vie) pour obtenir $\widehat{\eta}_l^N$.

L'un des problèmes posés par cette dynamique est que des particules en pleine santé peuvent être éliminées : si M particules sont encore en vie après l'étape de mutation, chacune des particules est affectée du poids $1/M$. On tire alors N particules selon la loi uniforme sur ces M sites. La loi de la répartition des N particules est donc une loi multinomiale de paramètres N et $(1/M, \dots, 1/M)$.

Une façon de remédier à cette perte de diversité est de modifier la dynamique du système en obligeant les M particules encore en vie après l'étape de mutation à rester sur place et en distribuant les $N - M$ autres sur ces M sites uniformément et indépendamment.

Il est possible de mesurer les performances de l'estimation par le système de particules de multiples façons. Par exemple, on peut montrer que

$$\sqrt{N} \left(\frac{\widehat{\gamma}_L^N(1)}{\mathbb{P}(T_B \leq T)} - 1 \right) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, V_L).$$

L'expression de V_L est en général très compliquée et elle fait intervenir des termes que l'on ne connaît pas comme les $(p_l)_l$. Si $S = \mathbb{R}$ et que X est continu ou si X est une chaîne de Markov sur \mathbb{Z} aux plus proches voisins, l'expression de V_L est très simple :

$$V_L = \sum_{l=1}^L \left(\frac{1}{p_l} - 1 \right). \quad (27)$$

Le fait que cette expression soit relativement simple et que l'analyse du problème soit un peu plus facile vient du fait que le processus X entre dans chaque ensemble B_l par le même point x_l .

L'expression (27) suggère que, pour L donné, les ensembles $(B_l)_{l \leq L}$ doivent être choisis tous égaux à $P_B^{1/L}$. Bien sûr ceci est impossible en pratique puisque l'on ne connaît pas P_B et ne la probabilité d'aller d'un ensemble à un autre.

En pratique on procède dans l'autre sens. On choisit une probabilité p (1/4 par exemple) puis on lance N trajectoires dans B_0 jusqu'à ce qu'elles atteignent R . On choisit alors un ensemble B_1 (en fait un niveau x_1) tel qu'une proportion p de trajectoires aient atteint B_1 avant R . On redistribue ensuite les $N(1-p)$ particules parmi les Np qui sont encore en vie et on les fait repartir et on recommence... Un des avantages de la méthode est que l'on contrôle le nombre de morts à chaque étape : le niveau à atteindre est fixé pour. Il permet aussi de former des trajectoires typiques et

pas seulement d'estimer P_B . Quand une particule meurt puis renaît ailleurs, elle hérite du passé de la particule à laquelle elle rejoint. Il y a donc une perte de la diversité des trajectoires (surtout au niveau des ancêtres lointains). Remarquons enfin que cet algorithme ne s'écrit pas sous la forme Feynman-Kac (personne n'est parfait).

Citons un autre algorithme qui évite la mort du système de particules. On fixe les ensembles $(B_l)_l$. Notons H la taille du système de particules. À l'étape l , plutôt que générer H trajectoires et de compter les morts, on simule autant de particules que nécessaire pour avoir H survivants et l'on note N_l^H ce nombre. C'est le nombre de tentatives nécessaires pour l'apparition du H -ième succès dans un schéma de Bernoulli de paramètre p_l . On estime donc p_l par H/N_l^H plutôt que par le rapport de N par le nombre de survivants.