02/02 : Metropolis Hastings, Doeblin, recuit simulé

09/02 : Recuit simulé (fin), condition de dérive

16/02 : LGN ergodique

~~23/02~~
02/03 : TLC ergodique

09/03 : Méthodes particulaires

16/03
23/03) Metropolis Hastings adaptatif

Ch I Algorithme de Metropolis Hastings et convergence de chaînes de Markov

I/ Noyau markovien et chaîne de Markov.

Espace d'états E muni d'une tribu ξ

def : • On appelle noyau markovien sur (E, ξ) ~~une~~ une application $P : E \times \xi \to [0,1]$ telle que :

1) $\forall x \in E$, $A \in \xi \to P(x, A)$ est une proba,

2) $\forall A \in \xi$ $x \in E \to P(x, A)$ est mesurable.

• On appelle chaîne de Markov de noyau de transition P, un processus $(X_n)_{n \in \mathbb{N}}$ à valeurs dans E adapté à une filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$ tq

$$\forall n \in \mathbb{N}, \ \forall A \in \xi, \ \mathbb{P}(X_{n+1} \in A \mid \mathcal{F}_n) = P(X_n, A)$$

Notations :

- Pour $\varphi : E \to \mathbb{R}$ mesurable positive ou tq $\forall x \in E$, $\int_E |\varphi(y)|\, P(x, dy) < +\infty$, on note $P\varphi$ la fonction de $E \to \mathbb{R}$ définie par $P\varphi(x) = \int_E \varphi(y)\, P(x, dy)$.
- Pour μ dans l'espace $\mathcal{P}(E)$ des probas sur (E, ξ) et $\varphi : E \to \mathbb{R}$ mesurable ≥ 0 ou tq $\int |\varphi(x)|\, \mu(dx) < +\infty$ on note $\mu(\varphi) = \int_E \varphi(x)\, \mu(dx)$.
- On note μP la proba sur (E, ξ) définie par $\forall A \in \xi$, $\mu P(A) = \int_{x \in E} P(x, A)\, \mu(dx)$.
- La proba μ est dite invariante par P si $\mu P = \mu$.
- Si Q est un autre noyau markovien, on note PQ le noyau défini par $\forall (x, A) \in E \times \xi$, $PQ(x, A) = \int_{y \in E} Q(y, A)\, P(x, dy)$

. Itérées du noyau P

$$P^0(x,A) = \delta_x(A) = \mathbb{1}_A(x). \qquad P^1 = P$$

$$P^n = PP^{n-1} = P^{n-1}P \qquad \forall n \geq 2$$

Exercice : Mq que si $(X_n)_{n\geq 0}$ est une chaîne de Markov de noyau P tq $X_0 \sim \mu$, alors $X_n \sim \mu P^n$

II Algorithme de Metropolis Hastings :

Objectif : simuler suivant $\pi \in \mathcal{P}(E)$ qui s'écrit

$$\pi(dx) = \frac{\eta(x)\,\lambda(dx)}{\int_E \eta(y)\,\lambda(dy)} \quad \text{où}$$

- λ est une mesure positive de référence sur E (en : lebesgue sur $\mathbb{R}^d$)
- $\eta : E \to \mathbb{R}_+$ tq $\int_E \eta(y)\,\lambda(dy) \in]0,+\infty[$.

Pas besoin de connaître la constante de normalisation pour M.H.

M.H. permet de construire une chaîne de Markov de probabilité invariante π.

exemples d'application :

- physique statistique : $E = \mathbb{R}^d$ λ : Lebesgue

$\eta(x) = e^{-\frac{V(x)}{k_B T}}$ où $V : \mathbb{R}^d \to \mathbb{R}$ fonction potentiel

densité de Boltzmann-Gibbs

k_B : constante de Boltzmann

T : température

On note $\beta = \frac{1}{k_B T}$

- statistiques bayesiennes : E l'espace où vit un paramètre Θ que l'on souhaite estimer.

On $p_{Y|\Theta}(y|\theta)$ la densité de l'observation Y si le paramètre vaut θ et $p_\Theta(\theta)$ la densité a priori sur Θ. Par Bayes, la densité de Θ sachant que l'on a observé y est

$$p_{\Theta|Y}(\theta|y) = \frac{p_{Y|\Theta}(y|\theta)\, p_\Theta(\theta)}{\int_E p_{Y|\Theta}(y,\tilde\theta)\, p_\Theta(\tilde\theta)\, \lambda(d\tilde\theta)}$$

Construction de l'algorithme :

Soit $Q(x, dy) = q(x,y)\,\lambda(dy)$ noyau markovien à densité

$q : E \times E \to \mathbb{R}_+$ tq $\forall x \in E$, on sait simuler suivant $Q(x,\cdot) = q(x,\cdot)\lambda(\cdot)$

ratio de Metropolis Hastings

rapport $\Rightarrow$ pas besoin constante normalisation

$$\alpha(x,y) = \begin{cases} 1 \wedge \dfrac{\eta(y)\,q(y,x)}{\eta(x)\,q(x,y)} & \text{si } \eta(x)\,q(x,y) > 0 \\ 1 & \text{si } \eta(x)\,q(x,y) = 0 \end{cases}$$

à valeurs dans $[0,1]$

- On part de X_0 va à valeurs dans E
- Sachant $(X_0, \dots, X_n)$ on génère une proposition $Y_{n+1} \sim Q(X_n, \cdot)$ et indep une va $U_{n+1} \sim \mathcal{U}[0,1]$

$$X_{n+1} = Y_{n+1}\,\mathbb{1}_{\{U_{n+1} \leq \alpha(X_n, Y_{n+1})\}} + X_n\,\mathbb{1}_{\{U_{n+1} > \alpha(X_n, Y_{n+1})\}}$$

Pour $f : E \to \mathbb{R}$ mesurable bornée

$$\mathbb{E}(f(X_{n+1}) \mid X_0, \dots, X_n) = \mathbb{E}(\mathbb{E}(f(X_{n+1}) \mid X_0, \dots, X_n, Y_{n+1}) \mid X_0, \dots, X_n)$$

$$= \mathbb{E}\big(\alpha(X_n, Y_{n+1}) f(X_{n+1}) + (1-\alpha(X_n, Y_{n+1})) f(X_n) \mid X_0, \dots, X_n \big)$$

$$= \int_E f(y) \, \alpha(X_n, y) \, q(X_n, y) \, \lambda(dy) + f(X_n) \times \int_E (1-\alpha(X_n,y)) q(X_n,y) \lambda(dy)$$

$(X_n)_{n \geq 0}$ est une chaîne de Markov de noyau

$$P(x, dy) = \alpha(x,y) \, q(x,y) \, \lambda(dy) + R(x) \, \delta_x(dy)$$

où $R(x) = \int_E (1-\alpha(x,y)) \, q(x,y) \, \lambda(dy)$

<u>Montrons que π est reversible et donc invariante pour P :</u>

$$\eta(x) q(x,y) \alpha(x,y) = \begin{cases} \eta(x) q(x,y) \wedge \eta(y) q(y,x) & \text{si } \eta(x) q(x,y) > 0 \\ 0 = \eta(x) q(x,y) \wedge \eta(y) q(y,x) & \text{si } \eta(x) q(x,y) = 0 \end{cases}$$

$$= (\eta(x) q(x,y)) \wedge (\eta(y) q(y,x))$$

fonction symétrique de (x,y).

$\Rightarrow \eta(x)\lambda(dx)\,\alpha(x,y)\,q(x,y)\lambda(dy) = \eta(y)\,q(y,x)\,\alpha(y,x)\,\lambda(dx)\lambda(dy).$

égalités de mesures sur $E\times E$

$$\eta(x)\lambda(dx)\,R(x)\,\delta_x(dy) = \eta(y)\,\lambda(dy)\,R(y)\,\delta_y(dx)$$

$$\boxed{\pi(dx)\,P(x,dy) = \pi(dy)\,P(y,dx)}$$ en sommant et normalisant par $\int_E \eta(z)\lambda(dz)$

réversibilité

réversibilité $\Rightarrow$ invariance.

par intég sur $x\in E$, il vient $\pi P(dy) = \pi(dy)\underbrace{\int_{x\in E} P(y,dx)}_{P(y,E)=1}$

Remarques : . La valeur de $\alpha(x,y)$ lorsque $q(x,y)=0$ est sans importance pour P.

En revanche choisir $\alpha(x,y)=1$ si $\eta(x)=0$, maximise la probabilité d'accepter un mouvement si on se trouve en un point de densité nulle

• Autres choix possibles du ratio qui préservent la réversibilité

$$\alpha(x,y) = \begin{cases} a\left(\dfrac{\eta(y)\,q(y,x)}{\eta(x)\,q(x,y)}\right) & \text{si } \eta(x)\,q(x,y) > 0 \\ 1 & \text{si } \eta(x)\,q(x,y) = 0. \end{cases}$$

avec $a : \mathbb{R}_+ \to [0,1]$ tq $\forall u > 0$, $a(u) = u\, a\left(\frac{1}{u}\right)$

choix MH $\longrightarrow$ $a(u) = 1 \wedge u$

Pour a comme préc, $a(u) \leq 1 \wedge u$

Le choix MH maximise la probabilité d'accepter Y_{n+1} suivant $q(X_n, \cdot)$ ...

• MH par marche aléatoire : $E = \mathbb{R}^d$ et $q(x,y) = \varphi(y-x)$ où φ densité de proba paire sur $\mathbb{R}^d$.

$Y_{n+1} = X_n + Z_{n+1}$ où $Z_{n+1} \sim \varphi$ indep de X_n.

$$\alpha(x,y) = \begin{cases} 1 \wedge \dfrac{\eta(y)\,\cancel{\varphi(x-y)}}{\eta(x)\,\cancel{\varphi(y-x)}} & \text{si } \eta(x)\varphi(y-x)>0 \\ 1 & \text{si } \eta(x)\,\varphi(y-x) = 0. \end{cases}$$

par parité de φ.

Dans la suite, nous allons donner des conditions sur P qui assurent que si $(X_n)_{n\geq 0}$ chaîne de Markov de noyau P alors

- P admet une unique proba invariante π et $\mathcal{L}(X_n) \xrightarrow[n\to+\infty]{} \pi$
- $\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) \xrightarrow{p.s.} \pi(f)$ si $\pi(|f|)<+\infty$ loi des Grands Nombres ergodiques.
- $\sqrt{n}\left(\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) - \pi(f)\right) \xrightarrow{?}$ TLC associé.

III Ergodicité de chaîne de Markov à espace $(E, \mathcal{E})$ général

$$\mu P^n \xrightarrow[n \to +\infty]{?} \pi \text{ invariante}$$

1) distance de variation totale :

déf : On appelle distance en variation totale sur $\mathcal{P}(E)$

$$d_{TV}(\mu, \sigma) = \sup_{A \in \mathcal{E}} |\mu(A) - \sigma(A)|$$

décomposition de Jordan-Hahn de la mesure signée $\mu - \sigma$: $\mu - \sigma = \xi_+ - \xi_-$ où ξ_+ et ξ_- sont des mesures ≥ 0 minimales au sens où $\exists\, B \in \mathcal{E}$ tq

$$\xi_-(B) = \xi_+(B^c) = 0$$ (voir cours de Poly).

Prop : La mesure $\mu \wedge \sigma(A) = \sigma(A \cap B) + \mu(A \cap B^c)$ est la plus grande mesure majorée à la fois par μ et par σ.

$$d_{TV}(\mu, \sigma) = 1 - \mu \wedge \sigma(E) = \sup_{|\varphi| \leq \frac{1}{2}} |\mu(\varphi) - \sigma(\varphi)|$$

où le supremum porte sur toutes les fonctions $\varphi : E \to \mathbb{R}$ mesurables tq $\sup_{x \in E} |\varphi(x)| \leq \frac{1}{2}$

NB : En général lorsque A n'est inclus ni ds B ni ds B^c, $\mu \wedge \sigma(A) \neq \mu(A) \wedge \sigma(A)$

preuve : $\mu(A) - \sigma(A) = \xi_+(A \cap B) - \xi_-(A \cap B^c)$

$\Rightarrow$ si $A \subset B$, $\mu(A) - \sigma(A) \geq 0$

si $A \subset B^c$, $\mu(A) - \sigma(A) \leq 0$

$$\mu \wedge \sigma(A) = \underbrace{\sigma(A \cap B)}_{\leq \mu(A \cap B)} + \underbrace{\mu(A \cap B^c)}_{\leq \sigma(A \cap B^c)} \leq \mu(A) \wedge \sigma(A).$$

Soit maintenant η mesure majorée à la fois par μ et par σ.

$$\eta(A) = \eta(A \cap B) + \eta(A \cap B^c) \leq \underbrace{\sigma(A \cap B) + \mu(A \cap B^c)}_{\mu \wedge \sigma(A)}$$

Ainsi $\mu \wedge \sigma$ est la plus grande mesure majorée à la fois par μ et par σ.

$$0 = \mu(E) - \sigma(E) = \xi_+(B) - \xi_-(B^c). \Rightarrow \xi_+(B) = \xi_-(B^c).$$

$$-\xi_-(A \cap B^c) \leq \underbrace{\mu(A) - \sigma(A)}_{\xi_+(A \cap B) - \xi_-(A \cap B^c)} \leq \xi_+(A \cap B)$$

avec majorant atteint pour $A = B$ et minorant atteint pour $A = B^c$

Ainsi $d_{TV}(\mu,\sigma) = \sup_{A\in\mathcal{E}} |\mu(A)-\sigma(A)| = \xi_+(B) = \mu(B)-\sigma(B)$

$$= 1 - \underbrace{(\mu(B^c)+\sigma(B))}_{\sigma(E\cap B)+\mu(E\cap B^c) = \mu\wedge\sigma(E)}.$$

Pour $|\varphi| \leq \frac{1}{2}$,

$$\mu(\varphi)-\sigma(\varphi) = \underbrace{\xi_+(\varphi)}_{\text{maximum dès que } \varphi=\frac{1}{2} \text{ sur } B} - \underbrace{\xi_-(\varphi)}_{\text{minimum dès que } \varphi=-\frac{1}{2} \text{ sur } B^c}.$$

$$\max_{|\varphi|\leq\frac{1}{2}} \{\mu(\varphi)-\sigma(\varphi)\} = \mu\left(1_B-\tfrac{1}{2}\right)-\sigma\left(1_B-\tfrac{1}{2}\right) = \mu(B)-\cancel{\tfrac{1}{2}}-\left(\sigma(B)-\cancel{\tfrac{1}{2}}\right)$$

$$= \xi_+(B) = d_{TV}(\mu,\sigma)$$

$$\min_{|\varphi|\leq\frac{1}{2}} \{\mu(\varphi)-\sigma(\varphi)\} = \mu\left(1_{B^c}-\tfrac{1}{2}\right)-\sigma\left(1_{B^c}-\tfrac{1}{2}\right) = -\xi_-(B^c)$$

$$= -d_{TV}(\mu,\sigma) \quad \blacksquare$$

Lemme : $$d_{TV}(\mu,\sigma) = \inf_{\substack{(X,Y) \text{ tq } X\sim\mu \\ Y\sim\sigma}} \mathbb{P}(X\neq Y)$$

preuve : Soit (X,Y) tq $X\sim\mu$, $Y\sim\sigma$ et $A\in\mathcal{E}$.

$$\mu(A) = \mathbb{P}(X\in A) = \mathbb{P}(X\in A,\, X=Y) + \mathbb{P}(X\in A,\, X\neq Y)$$
$$\sigma(A) = \mathbb{P}(Y\in A) = \mathbb{P}(Y\in A,\, X=Y) + \mathbb{P}(Y\in A,\, X\neq Y)$$

$$-\mathbb{P}(X\neq Y) \leq -\mathbb{P}(Y\in A,\, X\neq Y) \leq \mu(A)-\sigma(A) = \mathbb{P}(X\in A, X\neq Y) - \mathbb{P}(Y\in A, X\neq Y) \leq \underbrace{\mathbb{P}(X\in A, X\neq Y)}_{\leq \mathbb{P}(X\neq Y)}$$

sup sur A et inf sur couplages

$$\boxed{d_{TV}(\mu,\sigma) \leq \inf_{(X,Y)} \mathbb{P}(X\neq Y)}$$

Exhibons un couplage qui réalise la borne inf :

- si $d_{TV}(\mu,\sigma)=0$, $\mu=\sigma$, $X=Y$ convient
- si $d_{TV}(\mu,\sigma)=1$, $X\sim\mu$, $Y\sim\sigma$ quelconque convient

Supposons $0 < \underset{\underset{1-\mu\wedge\sigma(E)}{\parallel}}{d_{TV}(\mu,\sigma)} < 1$ ie $0 < \mu\wedge\sigma(E) < 1$.

Soit $Z \sim \dfrac{\mu\wedge\sigma}{\mu\wedge\sigma(E)}$ $\quad \tilde{X} \sim \dfrac{\mu-\mu\wedge\sigma}{1-\mu\wedge\sigma(E)}$ et $U \sim \mathcal{U}[0,1]$ indep de $(Z,\tilde{X},\tilde{Y})$

$\tilde{Y} \sim \dfrac{\sigma-\mu\wedge\sigma}{1-\mu\wedge\sigma(E)}$

On pose $\overset{Y}{X} = Z\,\mathbb{1}_{\{U\leq\mu\wedge\sigma(E)\}} + \overset{\tilde{Y}}{\tilde{X}}\,\mathbb{1}_{\{U>\mu\wedge\sigma(E)\}}$

$\mathbb{P}(X\neq Y) \leq \mathbb{P}(U>\mu\wedge\sigma(E)) = 1-\mu\wedge\sigma(E) = d_{TV}(\mu,\sigma)$.

Pour $f : E \to \mathbb{R}$ mes bornée

$$\mathbb{E}(f(X)) = \mathbb{E}\left(f(Z)\,\mathbb{1}_{\{U\leq\mu\wedge\sigma(E)\}}\right) + \mathbb{E}\left(f(\tilde{X})\,\mathbb{1}_{\{U>\mu\wedge\sigma(E)\}}\right)$$

$$\underset{\text{indep}}{=} \cancel{\mu\wedge\sigma(E)} \int_E f(z)\,\frac{\cancel{\mu\wedge\sigma}(dz)}{\cancel{\mu\wedge\sigma(E)}} + \cancel{(1-\mu\wedge\sigma(E))} \int_E f(x)\,\frac{\mu-\cancel{\mu\wedge\sigma}\,(d\tilde{x})}{\cancel{1-\mu\wedge\sigma(E)}}$$

$$= \int_E f(\tilde{x})\,\mu(d\tilde{x})$$

ie $X\sim\mu$. De la m façon $Y\sim\sigma$. ☐

Exo : | Mq pour $\gamma > 0$, $d_{TV}(\mu, \sigma) = \frac{1}{2\gamma} \sup_{|\varphi| \leq \gamma} |\mu(\varphi) - \sigma(\varphi)|$

Prop : L'espace $\mathcal{P}(E)$ muni de d_{TV} est complet.

preuve : Soit $(\mu_n)_{n \in \mathbb{N}}$ suite de Cauchy de $\mathcal{P}(E)$.

$$\mu = \sum_{m \in \mathbb{N}} \frac{\mu_m}{2^{m+1}} \in \mathcal{P}(E) \quad \text{et } \forall n \quad \mu_n \ll \mu.$$

Par Radon Nykodym, $\mu_n(dx) = p_n(x)\, \mu(dx)$ où $p_n : E \to \mathbb{R}_+$ mes tq $\int_E p_n(x) \mu(dx) = 1$

$$d_{TV}(\mu_n, \mu_m) = \sup_{|\varphi| \leq \frac{1}{2}} |\mu_n(\varphi) - \mu_m(\varphi)| = \sup_{|\varphi| \leq \frac{1}{2}} \left| \int_E \varphi(x) (p_n(x) - p_m(x)) \mu(dx) \right|$$

$\varphi = \frac{1}{2} \operatorname{sg}(p_n - p_m)$

$$= \frac{1}{2} \int |p_n(x) - p_m(x)| \, \mu(dx) = \frac{1}{2} \| p_n - p_m \|_{L^1(\mu)}.$$

$L^1(\mu)$ est complet. $(p_n)_n$ est une suite de Cauchy dans cet espace. Elle converge donc vers $p_\infty \in L^1(\mu)$

On peut extraire une sous-suite qui cv μ p.p.

ce qui implique que $p_\infty(x) \geq 0$ $\mu(dx)$ p.p.

Comme $\left|\int p_\infty(x)\mu(dx) - \int p_n(x)\mu(dx)\right| \leq \int |p_\infty - p_n(x)|\mu(dx) = \|p_n - p_\infty\|_{L^1(\mu)}$

$$\int p_\infty(x)\mu(dx) = 1$$

Soit $\mu_\infty(dx) = p_\infty(x)\mu(dx)$. $\mu_\infty \in \mathcal{P}(E)$.

$$d_{TV}(\mu_\infty, \mu_n) = \frac{1}{2}\|p_\infty - p_n\|_{L^1(\mu)} \xrightarrow[n \to +\infty]{} 0$$ ✓.

2) Condition de Doeblin et ergodicité uniforme :

<u>def</u> : On dit que le noyau P satisfait la condition de Doeblin si $\exists \alpha > 0$, $\forall x, y \in E$, $P(x,\cdot) \wedge P(y,\cdot)(E) \geq \alpha$

<u>NB</u> : $\Leftrightarrow d_{TV}(P(x,\cdot), P(y,\cdot)) \leq 1 - \alpha$ Impliqué par Doeblin unif

$\exists \alpha > 0, \exists \sigma \in \mathcal{P}(E), \forall x, y \in E, P(x,\cdot) \wedge P(y,\cdot) \geq \alpha\sigma(\cdot)$

<u>Thm:</u> Sous Doeblin, P admet une unique proba invariante π et $\forall \mu \in \mathcal{P}(E), \forall n \in \mathbb{N}$, $d_{TV}(\mu P^n, \pi) \leq (1-\alpha)^n d_{TV}(\mu, \pi)$

<u>Lemme:</u> Sous Doeblin, $\forall x, y \in E$ $|P\varphi(x) - P\varphi(y)| \leq 2(1-\alpha) \sup_{z \in E} |\varphi(z)|$ et $d_{TV}(\mu P, \sigma P) \leq (1-\alpha) d_{TV}(\mu, \sigma)$.

<u>Preuve du thm:</u>

$\mu \in \mathcal{P}(E) \longrightarrow \mu P \in \mathcal{P}(E)$ est une contraction d'après le lemme pour d_{TV}. Comme $\mathcal{P}(E)$ muni de d_{TV} est complet, il existe un unique point fixe π i.e. une unique proba invariante

$$d_{TV}(\mu P^n, \pi) = d_{TV}((\mu P^{n-1})P, (\pi P^{n-1})P) \overset{\text{lemme}}{\leq} (1-\alpha) d_{TV}(\mu P^{n-1}, \pi P^{n-1})$$
$$\leq (1-\alpha)^n d_{TV}(\mu, \pi) \text{ en itérant}$$

Preuve du lemme :

Soit $\varphi : E \to \mathbb{R}$ mes bornée.

$$|P\varphi(x) - P\varphi(y)| = \left| \int_E \varphi(z) \left(P(x,dz) - \cancel{P(x,\cdot)\wedge P(y,\cdot)}(dz)\right) - \int_E \varphi(z) \left(P(y,dz) - \cancel{P(x,\cdot)\wedge P(y,\cdot)}(dz)\right)\right|$$

$P(x,\cdot) - P(x,\cdot)\wedge P(y,\cdot) \geq 0$

$$\leq \int_E |\varphi(z)| \left(P(x,dz) - P(x,\cdot)\wedge P(y,\cdot)(dz)\right) + \int_E |\varphi(z)| \left(P(y,dz) - P(x,\cdot)\wedge P(y,\cdot)(dz)\right)$$

Doeblin

$$\leq 2 \sup_{z\in E} |\varphi(z)| \left(1 - P(x,\cdot)\wedge P(y,\cdot)(E)\right) \leq 2(1-\alpha) \sup_{z\in E} |\varphi(z)|. \quad (*)$$

On pose $c_\varphi = (1-\alpha) \sup_{z\in E} |\varphi(z)| - \sup_{y\in E} P\varphi(y)$

$$P\varphi(x) + c_\varphi \leq \cancel{P\varphi(x)} + (1-\alpha) \sup_{z\in E} |\varphi(z)| - \cancel{P\varphi(x)}$$

$$P\varphi(x) + c_\varphi \geq (1-\alpha) \sup_{z\in E} |\varphi(z)| - \sup_{y\in E} \left(P\varphi(y) - P\varphi(x)\right) \underset{(*)}{\geq} -(1-\alpha) \sup_{z\in E} |\varphi(z)|$$

$$\sup_{x\in E} |P\varphi(x) + c_\varphi| \underset{(*)}{\leq} (1-\alpha) \sup_{z\in E} |\varphi(z)|.$$

$$d_{TV}(\mu P, \sigma P) = \sup_{|\varphi| \leq \frac{1}{2}} |\mu P(\varphi) - \sigma P(\varphi)|$$

$$= \sup_{|\varphi| \leq \frac{1}{2}} |\mu(P\varphi + c_\varphi) - \sigma(P\varphi + c_\varphi)| \underset{(*)}{\leq} \sup_{|\psi| \leq \frac{1-\alpha}{2}} |\mu(\psi) - \sigma(\psi)|$$

$$\overset{\substack{\gamma = \frac{1-\alpha}{2} \\ \text{exo}}}{=} (1-\alpha)\, d_{TV}(\mu, \sigma) \qquad \blacksquare$$

NB : Supposons $\exists\, m \geq 1$ tq P^m satisfait Doeblin

D'après le thm, P^m admet une unique proba invariante π.

$$\pi P P^m = \pi P^m P = \pi P$$

Ainsi πP est invariante par P^m. Par unicité de la proba invariante $\pi P = \pi$ et π est invariante par P.

Comme toute proba invariante par P est invariante par P^m, π est l'unique proba invariante par P.

$$d_{TV}(\mu P, \sigma P) = \sup_{|\psi| \le \frac{1}{2}} |\mu P(\psi) - \sigma P(\psi)| = \sup_{|\psi| \le \frac{1}{2}} |\mu(P\psi) - \sigma(P\psi)|$$

$$\le \sup_{|\psi| \le \frac{1}{2}} |\mu(\psi) - \sigma(\psi)| \le d_{TV}(\mu, \sigma).$$

en combinant lemme pour P^m et l'inégalité

$$d_{TV}(\mu P^n, \sigma P^n) \le (1-\alpha)^{\lfloor \frac{n}{m} \rfloor} \underbrace{d_{TV}(\mu, \sigma)}_{\le 1}$$

$$d_{TV}(\mu P^n, \pi) \le (1-\alpha)^{\lfloor \frac{n}{m} \rfloor}$$

Pour le choix $\mu = \delta_x$, $\sup_{x \in E} d_{TV}(P^n(x, \cdot), \pi) \le (1-\alpha)^{\lfloor \frac{n}{m} \rfloor}$

ce qui implique $\lim_{n \to +\infty} \sup_{x \in E} d_{TV}(P^n(x, \cdot), \pi) = 0$

propriété appelée ergodicité uniforme.

Inversement sous ergodicité unif on choisit n_α assez grand pour que $\sup_{x \in E} d_{TV}(P^{n_\alpha}(x, \cdot), \pi) \le \frac{1-\alpha}{2}$

$$d_{TV}(P^{m_\alpha}(x,\cdot), P^{m_\alpha}(y,\cdot)) \leq d_{TV}(P^{m_\alpha}(x,\cdot), \pi) + d_{TV}(\pi, P^{m_\alpha}(y,\cdot))$$
$$\leq 1-\alpha.$$

Donc P^{m_α} satisfait Doeblin.

Ainsi ergodicité unif $\Leftrightarrow$ Doeblin pour une itérée de P.

3) Algorithme de recuit simulé :

Objectif : minimiser $V : E \to \mathbb{R}$ mesurable

On suppose que $v_* := \inf\{v : \lambda(\{x : V(x) \leq v\}) > 0\} > -\infty$ ess inf de V pour λ.

$\eta_\beta(x) = e^{-\beta V(x)}$ pour $\beta \geq 0$.

On note $B = \{\beta \geq 0 : Z_\beta := \int_E e^{-\beta V(x)} \lambda(dx) < +\infty\}$.

On suppose $\mathcal{B} \neq \emptyset$.

$$Z_\beta = e^{-\beta v_*} \underbrace{\int e^{-\beta \overbrace{(V(x) - v_*)}^{\geq 0 \ \lambda(dx)\ p.p.}} \lambda(dx)}_{\searrow \text{ avec } \beta}$$

Ainsi $\mathcal{B} = [\underline{\beta}, +\infty[$ ou $\mathcal{B} =]\underline{\beta}, +\infty[$ avec $\underline{\beta} \in \mathbb{R}_+$.

Pour $\beta \in \mathcal{B}$, on pose $\pi_\beta(dx) = \frac{\eta_\beta(x)}{Z_\beta} \lambda(dx)$.

<u>Lemme</u> $\forall \varepsilon > 0, \ \pi_\beta(\{x : V(x) \geq v_* + \varepsilon\}) \xrightarrow[\beta \to +\infty]{} 0$

Lorsque la "température" $\to 0$, π_β se concentre sur les minimas de V

Preuve : Soit $\beta > \underline{\beta}$

$$\Pi_\beta\left(\{x : V(x) \geq v_* + \varepsilon\}\right) = \frac{\int \mathbb{1}_{\{V(x) \geq v_* + \varepsilon\}} e^{-\beta(V(x) - v_*)} \lambda(dx)}{\int e^{-\beta(V(x)-v_*)} \lambda(dx)}$$

$$\leq \frac{\int_E \mathbb{1}_{\{V(x) \geq v_* + \varepsilon\}} e^{-\beta(V(x)-v_*)} \lambda(dx)}{e^{-\beta\frac{\varepsilon}{2}} \underbrace{\lambda\left(\{x : V(x) \leq v_* + \tfrac{\varepsilon}{2}\}\right)}_{>0 \text{ par déf de } v_*}}$$

$$\leq \frac{\int_E \mathbb{1}_{\{V(x) \geq v_* + \varepsilon\}} e^{-\frac{\beta}{2}(V(x)-v_*)} \lambda(dx)}{\lambda\left(\{x : V(x) \leq v_* + \tfrac{\varepsilon}{2}\}\right)}.$$

Pour $\beta \geq 2(\underline{\beta}+1)$, $\underbrace{\mathbb{1}_{\{V(x) \geq v_*+\varepsilon\}} e^{-\frac{\beta}{2}(V(x)-v_*)}}_{\xrightarrow[\beta \to +\infty]{} 0} \leq e^{-(\underline{\beta}+1)(V(x)-v_*)}$

Par CVD le numérateur $\to 0$ pour $\beta \to +\infty$ ☒

Algorithme de recuit simulé :

noyau de transition $Q(x, dy) = q(x,y)\,\lambda(dy)$ avec densité q symétrique $(\forall x,y \in E \quad q(x,y) = q(y,x))$

On note P_β le noyau de Metropolis Hastings associé à Q pour la proba cible π_β $(\beta > 0)$

$$\alpha_\beta(x,y) = \begin{cases} 1 \wedge \dfrac{\pi_\beta(y)\,\cancel{q(y,x)}}{\pi_\beta(x)\,\cancel{q(x,y)}} = 1 \wedge e^{-\beta(V(y)-V(x))} & \cancel{q(x,y) > 0} \\ \cancel{1 \quad \text{sinon}} \end{cases}$$

car la valeur du ratio lorsque $q(x,y) = 0$ est sans importance.

$$\alpha_\beta(x,y) = e^{-\beta(V(y)-V(x))^+}$$

On choisit un schéma de température $(\beta_n)_{n\in\mathbb{N}}$

$\mathbb{P}(X_{n+1} \in A \mid \mathcal{F}_n) = P_{\beta_n}(X_n, A)$ avec $\beta_n \xrightarrow[n\to+\infty]{} +\infty$