Ch III Méthodes de Monte Carlo par chaînes de Markov adaptatives

I L'algorithme SHUS (Self Healing Umbrella Sampling) :

1) le cadre :

$\pi(dx) = \eta(x)\,\lambda(dx)$ avec $\eta : E \to \mathbb{R}_+$ mesurable densité de proba par rapport à λ (i.e. $\int_E \eta(x)\,\lambda(dx) = 1$).

On se donne

- une partition mesurable $(E_i)_{1 \le i \le g}$ de E
- on note
  - $I : E \ni x \longmapsto \sum_{i=1}^{g} i\,\mathbb{1}_{E_i}(x)$.
  - $\theta_*(i) = \pi(E_i) = \int_{E_i} \eta(x)\,\lambda(dx)$. Quitte à retirer des E_i de la partition on suppose que $\forall i$, $\theta_*(i) > 0$
  - $\textcircled{H} = \{ \theta = (\theta(1), \dots, \theta(g)) \in]0,1[^g$ tq $\sum_{i=1}^{g} \theta(i) = 1 \}$.

NB : $\theta_* \in \textcircled{H}$.

Métastabilité : θ_* est très loin de l'uniformité

besoin de passer par des strates E_i de très faible proba cible $\theta_*(i)$ pour relier des strates de plus forte probas.

Suggère de modifier la proba cible pour la rendre plus uniforme

Pour $\theta \in \Theta$, on pose $\pi_\theta(dx) = \frac{1}{Z_\theta} \sum_{i=1}^{g} \frac{\eta(x)}{\theta(i)} \mathbb{1}_{E_i}(x) \lambda(dx)$

$$= \frac{1}{Z_\theta} \frac{\eta(x)}{\theta(I(x))} \lambda(dx).$$

où la constante de normalisation Z_θ vaut

$$Z_\theta = \sum_{i=1}^{g} \frac{\pi(E_i)}{\theta(i)} = \sum_{i=1}^{g} \frac{\theta_*(i)}{\theta(i)} \qquad \pi_\theta(E_i) = \frac{1}{Z_\theta} \frac{\pi(E_i)}{\theta(i)} = \frac{1}{Z_\theta} \frac{\theta_*(i)}{\theta(i)}$$

Pour $\theta = \theta_*$, $Z_{\theta_*} = g$ $\quad \pi_{\theta_*}(E_i) = \frac{1}{g}$ les strates E_i sont équipondérées suivant π_{θ_*}

ens : M_1 $\pi_{(\frac{1}{g}, \dots, \frac{1}{g})} = \pi$

Construction d'un couple (X_n, θ_n) à l'instant n tq θ_n approche θ_* et X_n simulé suivant noyau de Metropolis Hastings $P_{\theta_{n-1}}(X_{n-1}, \cdot)$ qui admet $\pi_{\theta_{n-1}}$ comme proba cible.

On met à jour θ_n à partir de θ_{n-1} en pénalisant la strate $I(X_n)$ $\left(\theta_n(I(X_n)) < \theta_{n-1}(I(X_n)) \right.$ et $\left. \theta_n(i) > \theta_{n-1}(i) \text{ pour } i \neq I(X_n) \right)$

<u>Calcul d'intégrale suivant π alors que l'on a simulé suivant π_{θ^*}:</u>

$$\pi_{\theta^*} = \frac{1}{g} \sum_{i=1}^{g} \frac{\eta(x) \, 1_{E_i}(x) \, \lambda(dx)}{\theta_*(i)} = \sum_{i=1}^{g} \frac{1}{g\,\theta_*(i)} 1_{E_i}(x) \, \pi(dx).$$

$$\frac{d\pi}{d\pi_{\theta^*}}(x) = \sum_{i=1}^{g} g\,\theta_*(i) \, 1_{E_i}(x)$$

Pour $f : E \to \mathbb{R}$ mesurable bornée $\int_E f(x)\,\pi(dx) = \sum_{i=1}^{g} \int_{E_i} g\,\theta_*(i)\, f(x)\, \pi_{\theta^*}(dx)$.

Pour $X \sim \pi_{\theta^*}$ $= \mathbb{E}\left(\sum_{i=1}^{g} g\,\theta_*(i) f(X) 1_{E_i}(X) \right)$

peut de variance liée au chgt de proba

$$\mathrm{Var}\left(\sum_{i=1}^{g} g\theta_*(i)\, \mathbb{1}_{E_i}(X)\right) = \sum_{i=1}^{g} (g\theta_*(i))^2\, \pi_{\theta_*}(E_i) - \Big(\sum_{i=1}^{g} \cancel{g}\theta_*(i)\underbrace{\pi_{\theta_*}(E_i)}_{\frac{1}{\cancel{g}}}\Big)^2$$

$$= g\sum_{i=1}^{g} \theta_*^2(i) - 1$$

peut être grande si θ_* loin de l'uniformité (en cas de métastabilité)

ex : 2 strates de poids proches de $\frac{1}{2}$ toutes les autres de poids proche de $0 \longrightarrow \mathrm{Var} \simeq \frac{g}{2} - 1 \gg 1$ si $g \gg 1$.

Introduction d'un paramètre $\alpha \in [0,1]$ pour tempérer le changement de proba !

$$\eta_\theta^\alpha(x) = \frac{1}{Z_\theta^\alpha} \sum_{i=1}^{g} \frac{\eta(x)}{\theta_*^\alpha(i)} \mathbb{1}_{E_i}(x). \qquad Z_\theta^\alpha = \sum_{i=1}^{g} \frac{\theta_*(i)}{\theta^\alpha(i)}$$

On pose $\pi_\theta^\alpha(dx) = \eta_\theta^\alpha(x)\, \lambda(dx)$

$$\pi_\theta^0 = \pi \qquad \pi_\theta^1 = \pi_\theta$$

Pour $X \sim \Pi^{\alpha}_{\theta^*}$ $\qquad Z^{\alpha}_{\theta^*} = \sum_{i=1}^{g} \theta_*^{1-\alpha}(i)$.

$$\frac{d\Pi}{d\Pi^{\alpha}_{\theta^*}}(x) = \sum_{i=1}^{g} Z^{\alpha}_{\theta^*}\, \theta_*^{\alpha}(i)\, 1_{E_i}(x)$$

$$\mathrm{Var}\left(\frac{d\Pi}{d\Pi^{\alpha}_{\theta_*}}(X)\right) = \sum_{i=1}^{g} \left(Z^{\alpha}_{\theta^*}\theta_*^{\alpha}(i)\right)^2 \Pi^{\alpha}_{\theta^*}(E_i) - \left(\sum_{i=1}^{g} \cancel{Z^{\alpha}_{\theta_*}}\, \theta_*^{\alpha}(i)\, \underbrace{\Pi^{\alpha}_{\theta_*}(E_i)}_{\frac{\theta_*^{1-\alpha}(i)}{\cancel{Z^{\alpha}_{\theta_*}}}}\right)^2$$

$$= Z^{\alpha}_{\theta^*} \sum_{i=1}^{g} \theta_*^{1+\alpha}(i) - 1$$

$$= \sum_{i=1}^{g} \theta_*^{1-\alpha}(i) \sum_{i=1}^{g} \theta_*^{1+\alpha}(i) - 1$$

Lemme : $\alpha \longrightarrow f(\alpha) = \sum_{i=1}^{g} \theta_*^{1-\alpha}(i) \sum_{i=1}^{g} \theta_*^{1+\alpha}(i)$ $\nearrow$ sur $[0,1]$.

NB : pour $\alpha = 0$ on obtient $f(0) = \left(\sum_{i=1}^{g} \theta_*(i)\right)^2 = 1$

Preuve du lemme :

$$\frac{d}{d\alpha}\left(\ln f(\alpha)\right) = \frac{d}{d\alpha} \ln\left(\sum_{i=1}^{g} \theta_*^{1+\alpha}(i)\right) + \frac{d}{d\alpha} \ln\left(\sum_{i=1}^{g} \theta_*^{1-\alpha}(i)\right)$$

$$= \frac{\sum_{i=1}^{g} \theta_*^{1+\alpha}(i) \ln \theta_*(i)}{\sum_{i=1}^{g} \theta_*^{1+\alpha}(i)} - \frac{\sum_{i=1}^{g} \theta_*^{1-\alpha}(i) \ln \theta_*(i)}{\sum_{i=1}^{g} \theta_*^{1-\alpha}(i)}.$$

$$= g(1+\alpha) - g(1-\alpha) \geq 0 \quad \text{pour } \alpha \in [0,1]$$

$$\Rightarrow f \nearrow \text{ sur } [0,1].$$

où $$g(b) = \frac{\sum_{i=1}^{g} \ln \theta_*(i)\, \theta_*^{b}(i)}{\sum_{i=1}^{g} \theta_*^{b}(i)}$$ pour $b \geq 0$

$$g'(b) = \frac{\sum_{i=1}^{g} (\ln \theta_*(i))^2\, \theta_*^{b}(i)}{\sum_{i=1}^{g} \theta_*^{b}(i)} - \left(\frac{\sum_{i=1}^{g} (\ln \theta_*(i))\, \theta_*^{b}(i)}{\sum_{i=1}^{g} \theta_*^{b}(i)}\right)^2 \geq 0.$$

en variance

$$\left(\sum_{i=1}^{g} \ln \theta_*(i)\, \theta_*^{\frac{b}{2}}(i)\, \theta_*^{\frac{b}{2}}(i)\right)^2 \overset{C.S.}{\leq} \left(\sum_{i=1}^{g} (\ln(\theta_*(i)))^2\, \theta_*^{b}(i)\right) \sum_{i=1}^{g} \theta_*^{b}(i).$$

$\boxed{}$

2) L'algorithme SHUS$_\alpha^{\kappa}$ avec $\alpha \in [0,1]$ et $\kappa \in]\frac{1}{2},1]$:

On part de $(\tilde{\theta}_0, X_0)$ à valeurs dans $]0,+\infty[^d \times E$ et on passe par récurrence de $(\tilde{\theta}_n, X_n)$ à $(\tilde{\theta}_{n+1}, X_{n+1})$ en effectuant 3 étapes :

1) normalisation : on pose $S_n = \sum_{i=1}^{d} \tilde{\theta}_n(i)$.

et $\theta_n = \dfrac{\tilde{\theta}_n}{S_n} \in \Theta$.

2) tirage de $X_{n+1} \sim P^\alpha_{\theta_n}(X_n, \cdot)$ noyau de Metropolis-Hastings qui cible $\pi^\alpha_{\theta_n}$ (plus généralement P^α_θ cible π^α_θ)

3) mise à jour des poids

pour $1 \leq i \leq d$, $\tilde{\theta}_{n+1}(i) = \tilde{\theta}_n(i) + \dfrac{\gamma}{g_\alpha(S_n)} S_n\, \theta_n^\alpha(i)\, \mathbb{1}_{E_i}(X_{n+1})$

avec $\gamma > 0$

et pour $s > 0$, $g_\alpha(s) = \begin{cases} s \text{ si } \alpha = 1 \\ \ln(1+s)^{\frac{\alpha}{1-\alpha}} \text{ sinon} \end{cases}$

Ga $\tilde{\theta}_{n+1}(I(X_{n+1})) > \tilde{\theta}_n(I(X_{n+1}))$

et $\tilde{\theta}_{n+1}(i) = \tilde{\theta}_n(i)$ pour $i \neq I(X_{n+1})$.

$$\sum_{i=1}^{g} \frac{\theta_*(i)}{\tilde{\theta}_{n+1}^{\alpha}(i)} < \sum_{i=1}^{g} \frac{\theta_*(i)}{\tilde{\theta}_n^{\alpha}(i)}$$

(on suppose $\alpha > 0$).

Pour $i \neq I(X_{n+1})$

$$\Pi^{\alpha}_{\theta_{n+1}}(E_i) = \frac{1}{\sum_{j=1}^{g} \frac{\theta_*(j)}{\tilde{\theta}^{\alpha}_{n+1}(j)}} \frac{\theta_*(i)}{\underbrace{\tilde{\theta}^{\alpha}_{n+1}(i)}_{=\tilde{\theta}^{\alpha}_n(i)}} > \frac{1}{\sum_{j=1}^{g} \frac{\theta_*(j)}{\tilde{\theta}^{\alpha}_n(j)}} \frac{\theta_*(i)}{\tilde{\theta}^{\alpha}_n(i)} = \Pi^{\alpha}_{\theta_n}(E_i)$$

augmentation de la proba cible des strates non visitées à l'instant $n+1$.

Par passage au complémentaire $\Pi^{\alpha}_{\theta_{n+1}}(E_{I(X_{n+1})}) < \Pi^{\alpha}_{\theta_n}(E_{I(X_{n+1})})$

$$\Pi^{\alpha}_{\theta_{n+1}}(E_{I(X_{n+1})}) = 1 - \sum_{i \neq I(X_{n+1})} \Pi^{\alpha}_{\theta_{n+1}}(E_i) \qquad \Pi^{\alpha}_{\theta_n}(E_{I(X_{n+1})}) = 1 - \sum_{i \neq I(X_{n+1})} \Pi^{\alpha}_{\theta_n}(E_i)$$

Le pas de l'algorithme est $\gamma_{n+1} = \frac{\gamma}{\int \alpha(S_n)}$ dans la mise à jour de $\tilde{\theta}_{n+1}$.

$$\tilde{\theta}_{n+1}(i) = S_n \theta_n(i) \left(1 + \gamma_{n+1} \theta_n^{\alpha-1}(i) \mathbb{1}_{E_i}(X_{n+1})\right)$$
$$= \tilde{\theta}_n(i) \left(1 + \gamma_{n+1} \theta_n^{\alpha-1}(i) \mathbb{1}_{E_i}(X_{n+1})\right).$$

<u>Algorithme de Wang Landau de paramètre α noté WL_α :</u>

Pour une suite de pas $(\gamma_n)_{n \geq 1}$ donnée et $(\tilde{\theta}_0, X_0) \in]0, +\infty[^g \times E$

on itère le passage de $(\tilde{\theta}_n, X_n)$ à $(\tilde{\theta}_{n+1}, X_{n+1})$ ds $]0, +\infty[^g \times E$

en effectuant 3 étapes

1) normalisation $S_n = \sum_{i=1}^{g} \tilde{\theta}_n(i)$ $\theta_n = \frac{\tilde{\theta}_n}{S_n} \in \Theta$.

2) generation de $X_{n+1} \sim P^\alpha_{\theta_n}(X_n, \cdot)$ moyen MH ciblant $\pi^\alpha_{\theta_n}$

3) $\tilde{\theta}_{n+1}(i) = \tilde{\theta}_n(i) \left(1 + \gamma_{n+1} \theta_n^{\alpha-1}(i) \mathbb{1}_{E_i}(X_{n+1})\right).$

Algorithme stochastique pour $(\theta_n)_n$.

Typiquement on va avoir besoin du comportement suivant des pas : $\sum \gamma_n = +\infty$ et $\sum \gamma_n^2 < +\infty$.

$$\theta_{n+1}(i) = \frac{\tilde{\theta}_{n+1}(i)}{S_{n+1}} = \frac{1 + \gamma_{n+1}\, \theta_n^{\alpha-1}(i)\, \mathbb{1}_{E_i}(X_{n+1})}{1 + \gamma_{n+1}\, \theta_n^{\alpha}(I(X_{n+1}))} \times \underbrace{\frac{\tilde{\theta}_n(i)}{S_n}}_{\theta_n(i)}.$$

Le choix $\gamma_{n+1} = \dfrac{\gamma}{g_\alpha(S_n)}$ fait dans $\mathrm{SHUS}^\alpha_\alpha$ va assurer que $\gamma_{n+1} \asymp n^{-\alpha}$ $\alpha \in]\frac{1}{2}, 1]$.

ce qui implique que $\sum_{n\geq 0} \gamma_{n+1} = +\infty$ et $\sum_{n \geq 0} \gamma_{n+1}^2 < +\infty$.

3) intuition sur le choix de g_α :

$$S_{n+1} = S_n + \frac{\gamma}{g_\alpha(S_n)} S_n \,\theta_n^\alpha(I(X_{n+1}))$$

On espère que pour $n \to +\infty$, $\theta_n \to \theta_*$ et la loi de X_{n+1} converge vers $\pi^\alpha_{\theta_*}$

si bien que $\mathbb{E}(\theta_n^\alpha(I_n(X_{n+1}))) \longrightarrow \sum_{i=1}^{g} \theta_*^\alpha(i) \underbrace{\pi^\alpha_{\theta_*}(E_i)}_{\frac{\theta_*^{1-\alpha}(i)}{Z^\alpha_{\theta^*}}}$

$$= \frac{1}{Z^\alpha_{\theta^*}} \underbrace{\sum_{i=1}^{g} \theta_*(i)}_{1}$$

On peut espérer que S_n se comporte asympt pour $n \to +\infty$ comme $s(n)$

où $\frac{ds}{dt}(t) = \frac{\gamma s(t)}{Z^\alpha_{\theta^*} g_\alpha(s(t))}$

Cas $\alpha = 1$: $g_1(s) = s$ $\quad \frac{ds(t)}{dt} = \frac{\gamma}{Z^\alpha_{\theta^*}}$

$s(t) \sim \frac{\gamma t}{Z^\alpha_{\theta^*}}$

$S_n \sim \frac{\gamma n}{Z^\alpha_{\theta^*}} \quad \gamma_n \sim \frac{Z^\alpha_{\theta^*}}{n}$

Cas $\frac{1}{2} < \alpha < 1$: $\qquad g_\alpha(s) = \left(\ln(\cancel{1}+s)\right)^{\frac{\alpha}{1-\alpha}}$

$$\frac{d\,s(t)}{dt} = \frac{\gamma}{Z_{\beta^*}^{\alpha}} \frac{s(t)}{\left(\ln(s(t))\right)^{\frac{\alpha}{1-\alpha}}} \qquad \frac{d \ln s(t)}{dt} = \frac{\gamma}{Z_{\beta^*}^{\alpha}} \frac{1}{\left(\ln s(t)\right)^{\frac{\alpha}{1-\alpha}}}$$

$$\frac{d}{dt}\left(\left(\ln(s(t))\right)^{\frac{1}{1-\alpha}}\right) = \frac{\gamma}{(1-\alpha)\,Z_{\beta^*}^{\alpha}}$$

$$\ln(s(t)) \underset{t\to+\infty}{\sim} \left(\frac{\gamma\, t}{(1-\alpha)\,Z_{\beta^*}^{\alpha}}\right)^{1-\alpha} \qquad \left(\ln S\right)_n \sim \left(\frac{\gamma\, n}{(1-\alpha)\,Z_{\beta^*}^{\alpha}}\right)^{1-\alpha}$$

$$\gamma_n \sim \frac{\gamma}{\left(\frac{\gamma\, n}{(1-\alpha)\,Z_{\beta^*}^{\alpha}}\right)^{\alpha}} = \frac{\gamma^{1-\alpha}\left((1-\alpha)\,Z_{\beta^*}^{\alpha}\right)^{\alpha}}{n^{\alpha}} \qquad \text{pour } n \to +\infty.$$

4) Intuition sur la convergence pour $\alpha = 1$:

Supposons que $(\theta_n)_n$ converge vers $\theta_\infty \in \Theta$ lorsque $n \to +\infty$ et montrons qu'alors $\theta_\infty = \theta_*$

$$\frac{1}{n}\sum_{\ell=1}^{n} \mathbb{1}_{E_i}(X_\ell) \longrightarrow \pi^\alpha_{\theta_\infty}(E_i) = \frac{\theta_*(i)}{Z^\alpha_{\theta_\infty}\,\theta^\alpha_\infty(i)}$$

$$\tilde\theta_n(i) = \tilde\theta_0(i) + \gamma\sum_{\ell=1}^{n} \theta^\alpha_{\ell-1}(i)\,\mathbb{1}_{E_i}(X_\ell) \underset{n\to\infty}{\sim} \gamma\,\cancel{\theta^\alpha_\infty(i)}\; n\,\frac{\theta_*(i)}{Z^\alpha_{\theta_\infty}\,\cancel{\theta^\alpha_\infty(i)}}$$

$$S_n = \sum_{i=1}^{g} \tilde\theta_n(i) \underset{n\to+\infty}{\sim} \frac{\gamma n}{Z^\alpha_{\theta_\infty}} \underbrace{\sum_{i=1}^{g} \theta_*(i)}_{1}$$

$$\theta_n(i) = \frac{\tilde\theta_n(i)}{S_n} \xrightarrow[n\to+\infty]{} \frac{\gamma\,\theta_*(i)/Z^\alpha_{\theta_\infty}}{\gamma/Z^\alpha_{\theta_\infty}} = \theta_*(i)$$

i.e. $\theta_\infty = \theta_*$

II Convergence de l'algorithme $SHUS^{\alpha}_{\alpha}$:

Hypothèses :

A1 | la densité cible η est tq $\sup_{x \in E} \eta(x) < +\infty$

| et $\min_{1 \leq i \leq q} \theta_*(i) > 0$

A2 Pour tout $\theta \in \Theta$, le noyau P^{α}_{θ} est un noyau de Metropolis Hastings qui cible π^{α}_{θ} avec proposition $q(x,y)\,\lambda(dy)$ où q est une densité symétrique (i.e $q(x,y) = q(y,x)$ $\forall (x,y) \in E^2$) et vérifie $\inf_{(x,y) \in E \times E} q(x,y) > 0$

Thm: Soit $\gamma > 0$, $\alpha \in [0,1[$, $\alpha \in]\frac{1}{2},1]$. Supposons A1 et A2.

Alors l'algorithme $SHUS_\alpha^\alpha$ issu de n'importe quelle CI $(\theta_0, X_0) \in]0,+\infty[^g \times E$ converge au sens suivant

1) $\mathbb{P}\left(\theta_n \xrightarrow[n\to+\infty]{} \theta_*\right) = 1$.

2) $\forall f : E \to \mathbb{R}$ mesurable bornée

$$\mathbb{E}[f(X_n)] \xrightarrow[n\to+\infty]{} \int_E f(x)\, \pi^\alpha_{\theta_*}(dx).$$

et $$\mathbb{P}\left(\lim_{n\to+\infty} \frac{1}{n}\sum_{l=1}^{n} f(X_l) = \int_E f(x)\, \pi^\alpha_{\theta^*}(dx)\right) = 1.$$

3) $$\lim_{n\to+\infty} \mathbb{E}\left[\left(\sum_{i=1}^{g} \theta_{n-1}^{1-\alpha}(i)\right)\left(\sum_{i=1} \theta_{n-1}^{\alpha}(I(X_n))\right) f(X_n)\right] = \int_E f(x)\pi(dx)$$

$$\mathbb{P}\left(\lim_{n\to+\infty} \frac{1}{n}\sum_{l=1}^{n}\left(\sum_{i=1}^{g} \theta_{l-1}^{1-\alpha}(i)\right)\left(\sum_{i=g}^{g} \theta_{l-1}^{\alpha}(I(X_l))\right) f(X_l) = \int_E f(x)\pi(dx)\right) = 1$$

NB: $\frac{d\pi}{d\pi^\alpha_{\theta_*}}(x) = Z^\alpha_{\theta_*}\, \theta_*^{\alpha}(I(x)) = \left(\sum_{i=1} \theta_*^{1-\alpha}(i)\right) \theta_*^{\alpha}(I(x))$ approché par $\frac{d\pi}{d\pi^\alpha_{\theta_{l-1}}}(x)$

• Pour le même résultat avec $\alpha = 1$, il faut supposer en outre que $\inf_{x \in E} \gamma(x) > 0$.

permet de montrer la récurrence de l'algorithme i.e. son retour une infinité de fois dans un compact de (H). $\limsup_{n \to +\infty} \min_{1 \le i \le g} \theta_n(i) > 0$ ici

$$\theta_{n+1}(i) = \theta_n(i) \frac{1 + \gamma_{n+1} \theta_n^{\alpha-1}(i) \mathbb{1}_{E_i}(X_{n+1})}{1 + \gamma_{n+1} \theta_n^{\alpha}(I(X_{n+1}))}.$$

$$\frac{1+b}{1+c} = 1 + b - c + \frac{c(c-b)}{1+c}$$ appliqué avec $b = \gamma_{n+1} \theta_n^{\alpha-1}(i) \mathbb{1}_{E_i}(X_{n+1})$

$c = \gamma_{n+1} \theta_n^{\alpha}(I(X_{n+1}))$

$$\theta_{n+1}(i) = \theta_n(i) \Big\{ 1 + \gamma_{n+1} \Big(\theta_n^{\alpha-1}(i) \mathbb{1}_{E_i}(X_{n+1}) - \theta_n^{\alpha}(I(X_{n+1})) \Big) + \gamma_{n+1}^2 \frac{\theta_n^{\alpha}(I(X_{n+1})) \big(\theta_n^{\alpha}(I(X_{n+1})) - \theta_n^{\alpha-1}(i) \mathbb{1}_{E_i}(X_{n+1}) \big)}{1 + \gamma_{n+1} \theta_n^{\alpha}(I(X_{n+1}))} \Big\}$$

On pose $H_i(\theta,x) = \theta^{\alpha}(i)\, 1_{E_i}(x) - \theta(i)\, \theta^{\alpha}(I(x))$

et $h_i(\theta) = \int H_i(\theta,x)\, \pi_\theta^{\alpha}(dx)$

$$= \frac{\cancel{\theta^{\alpha}(i)}\,\theta_*(i)}{Z_\theta^{\alpha}\,\cancel{\theta^{\alpha}(i)}} - \theta(i)\sum_{j=1}^{g} \frac{\cancel{\theta^{\alpha}(j)}\,\theta_*(j)}{Z_\theta^{\alpha}\,\cancel{\theta^{\alpha}(j)}}$$

$$= \frac{\theta_*(i) - \theta(i)}{Z_\theta^{\alpha}}$$

$$\Lambda_{m+1}(i) = \gamma_{m+1}\,\theta_m(i)\left(\frac{\theta_m^{2\alpha}(I(X_{m+1})) - \overbrace{\theta_m^{\alpha}(I(X_{m+1}))\,\theta_m^{\alpha-1}(i)\,1_{E_i}(X_{m+1})}^{\theta_m^{2\alpha-1}(I(X_{m+1}))\,1_{E_i}(X_{m+1})}}{1+\gamma_{m+1}\,\theta_m^{\alpha}(I(X_{m+1}))}\right)$$

$$= \gamma_{m+1}\,\frac{\theta_m^{2\alpha}(I(X_{m+1}))\left(\theta_m(i) - 1_{E_i}(X_{m+1})\right)}{1+\gamma_{m+1}\,\theta_m^{\alpha}(I(X_{m+1}))}.$$

$$\theta_{m+1} = \theta_m + \gamma_{m+1}\, H(\theta_m, X_{m+1}) + \gamma_{m+1}\,\Lambda_{m+1}$$

(i par i)

$$|\Lambda_{n+1}|^2 \leq \gamma_{n+1}^2 \, \theta_n^{4\alpha}(I(X_{n+1})) \sum_{i=1}^{I} \underbrace{\left(\theta(i) - 1_{E_i}(X_{n+1})\right)^2}_{\in [-1,1]}$$

$$\leq g \, \gamma_{n+1}^2$$

$$|\Lambda_{n+1}| \leq \gamma_{n+1} \sqrt{g}$$

$$\theta_{n+1} = \theta_n + \gamma_{n+1} h(\theta_n) + \gamma_{n+1} \left(H(\theta_n, X_{n+1}) - h(\theta_n) \right) + \gamma_{n+1} \Lambda_{n+1}$$

Pour faire apparaître un accroissement de mart, on utilise l'équation de Poisson

$$\hat{H}_\theta(x) - P_\theta^\alpha \hat{H}_\theta(x) = H(\theta, x) - h(\theta)$$

g équations

1 pour chaque valeur de i.

$$H(\theta_n, X_{n+1}) - h(\theta_n) = \hat{H}_{\theta_n}(X_{n+1}) - P_{\theta_n}^\alpha \hat{H}_{\theta_n}(X_{n+1})$$

$$H(\theta_n, X_{n+1}) - h(\theta_n) = \hat{H}_{\theta_n}(X_{n+1}) - P^\alpha_{\theta_n}\hat{H}_{\theta_n}(X_{n+1})$$

$$= \underbrace{\hat{H}_{\theta_n}(X_{n+1}) - P^\alpha_{\theta_n}\hat{H}_{\theta_n}(X_n)}_{\text{accroiss mart}} + \underbrace{P^\alpha_{\theta_n}\hat{H}_{\theta_n}(X_n) - P^\alpha_{\theta_{n+1}}\hat{H}_{\theta_{n+1}}(X_{n+1})}_{\text{telescopique}}$$

$$+ \underbrace{P^\alpha_{\theta_{n+1}}\hat{H}_{\theta_{n+1}}(X_{n+1}) - P^\alpha_{\theta_n}\hat{H}_{\theta_n}(X_{n+1})}_{\text{petit si } \theta \to P^\alpha_\theta \hat{H}_\theta \text{ est régulière}}$$

car $\theta_{n+1} - \theta_n = O(\gamma_{n+1})$.

$$\theta_{n+1} = \theta_n + \gamma_{n+1}\left(\hat{H}_{\theta_n}(X_{n+1}) - P^\alpha_{\theta_n}\hat{H}_{\theta_n}(X_n)\right) + \gamma_{n+1} h(\theta_n) + R_{n+1}$$

où $R_{n+1} = \underbrace{\gamma_{n+1}\Delta_{n+1}}_{R^1_{n+1}} + \underbrace{\gamma_{n+1}\left(P^\alpha_{\theta_n}\hat{H}_{\theta_n}(X_n) - P^\alpha_{\theta_{n+1}}\hat{H}_{\theta_{n+1}}(X_{n+1})\right)}_{R^2_{n+1}}$

$+ \underbrace{\gamma_{n+1}\left(P^\alpha_{\theta_{n+1}}\hat{H}_{\theta_{n+1}}(X_{n+1}) - P^\alpha_{\theta_n}\hat{H}_{\theta_n}(X_{n+1})\right)}_{R^3_{n+1}}$

h s'appelle le champ moyen de l'algorithme.

$$\sum_{k=0}^{n-1} R^2_{\ell+1} = \sum_{\ell=0}^{n-1} \gamma_{\ell+1}\left(P^{\alpha}_{\theta_\ell}\hat{H}_{\theta_\ell}(X_\ell) - P^{\alpha}_{\theta_{\ell+1}}\hat{H}_{\theta_{\ell+1}}(X_{\ell+1})\right)$$

$$= \gamma_1 P^{\alpha}_{\theta_0}\hat{H}_{\theta_0}(X_0) - \gamma_n P^{\alpha}_{\theta_n}\hat{H}_{\theta_n}(X_n)$$

$$+ \sum_{\ell=1}^{n-1}(\gamma_{\ell+1}-\gamma_\ell)\, P^{\alpha}_{\theta_\ell}\hat{H}_{\theta_\ell}(X_\ell)$$

pour $\gamma_\ell = \frac{c}{\ell^\alpha}$ $\qquad \frac{\gamma_{\ell+1}-\gamma_\ell}{c} = \frac{1}{(\ell+1)^\alpha} - \frac{1}{\ell^\alpha}$

$$= \frac{1}{\ell^\alpha}\left(\left(1+\frac{1}{\ell}\right)^{-\alpha} - 1\right) \underset{\ell\to+\infty}{\sim} -\frac{\alpha}{\ell^{1+\alpha}}$$

<u>Fonction de Lyapunov</u> : $\omega(\theta) = \sum_{i=1}^{q} \theta_*(i) \ln\left(\frac{\theta_*(i)}{\theta(i)}\right)$ = entropie relative de θ^*/θ.

<u>Lemme</u> : $\omega \geq 0$ sur Θ ne s'annule qu'en θ_*

et $\nabla\omega \cdot h(\theta) < 0$ pour $\theta \in \Theta \setminus \{\theta_*\}$

⇒ " en moyenne l'algo tend à se rapprocher de θ_* "

Preuve : Par convexité de $x \to x \ln x$ (dérivée $\ln x + 1$, dérivée seconde $\frac{1}{x}$) sur $\mathbb{R}_+$

$$\omega(\theta) = \sum_{i=1}^{g} \frac{\theta_*(i)}{\theta(i)} \ln\left(\frac{\theta_*(i)}{\theta(i)}\right) \times \theta(i) \overset{\text{Jensen}}{\geq} \underbrace{\sum_{i=1}^{g} \frac{\theta_*(i)}{\theta(i)} \times \theta(i) \ln\Big(\underbrace{\sum_{i=1}^{g} \frac{\theta_*(i)}{\theta(i)} \theta(i)}_{1}\Big)}_{0}$$

La fonction étant strictement convexe on a même inégalité stricte si $\frac{\theta_*(i)}{\theta(i)}$ n'est pas constante ie si $\theta \neq \theta_*$

$$\nabla \omega(\theta) = -\left(\frac{\theta_*(1)}{\theta(1)}, \dots, \frac{\theta_*(g)}{\theta(g)}\right) \qquad h(\theta) = \frac{\theta_* - \theta}{Z^{\pm}_{\theta}}$$

$$h \cdot \nabla \omega(\theta) = -\frac{1}{Z^{\pm}_{\theta}} \sum_{i=1}^{g} \left(\theta_*(i) - \theta(i)\right) \frac{\theta_*(i)}{\theta(i)} = \frac{1}{Z^{\pm}_{\theta}} \Big(\underbrace{\sum_{i=1}^{g} \theta_*(i)}_{1} - \sum_{i=1}^{g} \frac{\theta_*^2(i)}{\theta(i)}\Big)$$

$$\sum_{i=1}^{g} \left(\frac{\theta_*(i)}{\theta(i)}\right)^2 \theta(i) \overset{\text{Jensen}}{\geq} \left(\sum_{i=1}^{g} \frac{\theta_*(i)}{\theta(i)} \theta(i)\right)^2 = 1$$

avec inégalité stricte si θ n'est pas proportionnelle à θ_* ie si $\theta \neq \theta_*$