Séance du 21/02

Notations : $f_N : \mathbb{R}^d \to \mathbb{R}$ fonction "cible".

$$\hat{y}^{(N)}(x;\theta) = \frac{1}{N}\sum_{i=1} \sigma_*(x;\theta_i)$$

$$R_N(\theta) = \mathbb{E}\left[\left(f(x) - \hat{y}^{(N)}(x;\theta)\right)^2\right] = \text{risque quadratique moyen.}$$

$$= \mathbb{E}\left[f(x)^2\right] + \frac{2}{N}\sum_i V(\theta_i) + \frac{1}{N^2}\sum_{i,j} U(\theta_i,\theta_j)$$

$$V(\theta) = \mathbb{E}\left[-f(x)\,\sigma_*(x;\theta)\right]$$

$$U(\theta_1,\theta_2) = \mathbb{E}\left[\sigma_*(x;\theta_1)\,\sigma_*(x;\theta_2)\right]$$

$$\Psi(\theta;\rho) = V(\theta) + \int U(\theta,\theta')\,d\rho(\theta')$$

$$G(\theta;\rho) = -\nabla_\theta \Psi = \mathbb{E}\left[f(x)\nabla_\theta \sigma_*(x;\theta)\right] - \int \mathbb{E}\left[\nabla_\theta \sigma_*(x;\theta)\,\sigma_*(x;\theta')\right] d\rho(\theta')$$

$$= \mathbb{E}\left[\left(y - \int \sigma_*(x;\theta')\,d\rho(\theta')\right)\nabla_\theta \sigma_*(x;\theta)\right]$$

<u>Remarque</u> : on définit

$$R(\rho) = \mathbb{E}[f(x)^2] + 2\int V(\theta)\, d\rho(\theta) + \int U(\theta,\theta')\, d\rho(\theta)\, d\rho(\theta')$$

le risque pour une distribution ρ des paramètres.

Alors $\left| \inf_\theta R_N(\theta) - \inf_\rho R(\rho) \right| \leq \frac{C}{N}$.

Et $\inf_\theta R_N(\theta)$ approche le risque minimum à $O(\frac{1}{N})$ près (théo. de Barron). Donc si on trouve un jeu de paramètres tels que $R_N(\theta)$ est proche de $\inf_\rho R(\rho)$, on a gagné.

<u>Dynamique</u> : SGD

$$\theta_i^{k+1} = \theta_i^k + \underbrace{2\varepsilon\big(y - \hat{y}(x_k; \theta^k)\big) \cdot \nabla_\theta \sigma_*(x_k; \theta_i^k)}_{F[\theta](x_k;\, \theta_i^k)}$$

Equation pour $\rho_t(\theta)$:

$$\partial_t \rho = 2\nabla_\theta\left(\nabla_\theta \Psi(\theta, \rho_t)\, \rho_t \right)$$

On introduit la dynamique :

$$\bar{\theta}_i^t = \theta_i(0) - 2\int_0^t \nabla_\theta \Psi(\bar{\theta}_i^s, \rho_s)\, ds$$

alors ρ_s = loi de $\bar{\Theta}_i(s)$, et les $\bar{\Theta}_i(s)$ sont i.i.d.

<u>But</u> : démontrer que $\hat{\rho}_k^{(N)} = \frac{1}{N} \sum_i \delta(\theta - \theta_i^k)$

est proche de $\rho_{k\varepsilon}$.

On va démontrer que $|\theta_i^k - \bar{\Theta}_i(k\varepsilon)|$ petit

$\bar{\Theta}_i^k$.

<u>Hypothèses</u> : On suppose toute la régularité qu'on veut sur $\sigma_{\underline{x}}$, sur $\mathbb{P}_x$.
On admet toutes les propriétés souhaitées sur ρ_t (en particulier : Lipschitz en t).
On appellera C toutes les constantes.

<u>Théo</u> : t fixé $\quad k = t/\varepsilon$

$$\max_i \mathbb{E}\left[|\theta_i^k - \bar{\Theta}_i^k|\right] \leq e^{ct} \times (\dots)$$

petit quand $N \to \infty$, $\varepsilon \to 0$

$$\frac{1}{2}\left(\theta_i^k - \bar{\theta}_i^k\right) = \varepsilon \sum_{\ell=0}^{k-1} F[\theta^\ell](x_\ell;\theta_i^\ell) - \int_0^{k\varepsilon} G(\bar{\theta}_i(s),\rho_s)\,ds$$

$$= \varepsilon \sum_{\ell=0}^{k-1} F[\theta^\ell](x_\ell;\theta_i^\ell) - \sum_{\ell=0}^{k-1} \int_{\ell\varepsilon}^{(\ell+1)\varepsilon} G(\bar{\theta}_i(s),\rho_s)\,ds$$

$$G(\bar{\theta}_i^\ell;\rho_{\ell\varepsilon}) + \underbrace{C\varepsilon}_{\bar{\theta}_i(s),\ \rho_s \text{ Lipch. en } s.}$$

$$\frac{1}{2}\left|\theta_i^k - \bar{\theta}_i^k\right| \leq \varepsilon \left| \sum_{\ell=0}^{k-1} F[\theta^\ell](x_\ell,\theta_i^\ell) - G(\bar{\theta}_i^\ell;\rho_{\ell\varepsilon}) \right| + C\varepsilon^2 k$$

$\mathcal{F}_\ell$ = tribu engendré par les $\theta_i(0)$ et $x_0,\dots,x_\ell$.

Alors $\mathbb{E}\left[F[\theta^\ell](x_\ell;\theta_i^\ell) \,\middle|\, \mathcal{F}_{\ell-1}\right] = G(\theta_i^\ell, \hat{\rho}_\ell^{(N)})$

mesure empirique des θ_i^ℓ

On introduit

$$Z_i^\ell := F[\theta^\ell](x_\ell;\theta_i^\ell) - G(\theta_i^\ell;\hat{\rho}_\ell^{(N)})$$

Alors $\sum_{\ell=0}^{k-1} Z_i^\ell = M_i^{k-1}$: martingale.

Termes à estimer : $\varepsilon \Pi_i^{k-1}$ et $\underbrace{G(\theta_i^\ell, \hat{\rho}_\ell^{(N)}) - G(\bar{\theta}_i^\ell, \rho_{\ell\varepsilon})}_{E}$

$$
\begin{aligned}
|E| \leq\ & \left|G(\theta_i^\ell, \hat{\rho}_\ell^{(N)}) - G(\bar{\theta}_i^\ell, \hat{\rho}_\ell^{(N)})\right| && \leq C\,|\theta_i^\ell - \bar{\theta}_i^\ell| \\
&+ \left|G(\bar{\theta}_i^\ell, \hat{\bar{\rho}}_\ell^{(N)}) - G(\bar{\theta}_i^\ell, \hat{\rho}_\ell^{(N)})\right| && \leq C\,\frac{1}{N}\sum_j |\theta_j^\ell - \bar{\theta}_j^\ell| \\
&+ \left|G(\bar{\theta}_i^\ell, \hat{\bar{\rho}}_\ell^{(N)}) - G(\bar{\theta}_i^\ell, \rho_{\ell\varepsilon})\right| && =: Y_i^\ell
\end{aligned}
$$

$$\mathbb{E}(|Y_i^\ell|) \leq \frac{C}{\sqrt{N}}.$$

$$
\begin{aligned}
\mathbb{E}(|\theta_i^k - \bar{\theta}_i^k|) \leq\ & C\varepsilon \sum_{\ell=0}^{k-1} \mathbb{E}(|\theta_i^\ell - \bar{\theta}_i^\ell|) + C\varepsilon \sum_{\ell=0}^{k-1} \frac{1}{N} \sum_j \mathbb{E}(|\theta_j^\ell - \bar{\theta}_j^\ell|) \\
&+ \underbrace{L\varepsilon^2 k + \frac{C\varepsilon k}{\sqrt{N}} + \varepsilon C\sqrt{k}}
\end{aligned}
$$

$k\varepsilon = t$ fixé

petit quand $\begin{cases} N \to \infty \\ \varepsilon \to 0 \end{cases}$

$$\Delta_k = \max_i \mathbb{E}(|\theta_i^k - \bar{\theta}_i|)$$

$$\Delta_k \leq 2C\varepsilon \sum_{\ell=0}^{k-1} \Delta_\ell + \alpha_k$$

Gronwall discret pour terminer.

Si $u_k \leqslant C \sum_{\ell=0}^{k-1} u_\ell + \varepsilon_k$

Alors $u_k \leqslant \varepsilon_k + C(1+C)^k \sum_{\ell=0}^{k-1} \varepsilon_\ell (1+C)^{-\ell}$

$$\Rightarrow \Delta_k \leqslant \varepsilon_k + C'(1+\varepsilon C)^k \varepsilon_k$$

$$\leqslant C_1 e^{Ct} \varepsilon_k$$

$$\varepsilon_k \leqslant L\varepsilon t + \frac{Ct}{\sqrt{N}} + C\sqrt{\varepsilon t}$$