Optimisation stochastique

Le but: calculer $h(\vartheta) = \mathbb{E}\left[H(\vartheta, Z)\right]$.

On considère la procédure récursive

$$\vartheta_{n+1} = \vartheta_n - \gamma_{n+1} H(\vartheta_n, Z_{n+1}), \quad (Z_n)_{n\geq 1} \ i.i.d. \sim \mathbb{P}_Z.$$

ϑ_0 est indépendant de $(Z_n)_{n\geq 1}$.

Théorème (lemme de Robbins–Siegmund)

1. Il existe une fonction de Lyapounov $V(\vartheta) \in C^1$ telle que
    a. Elle est coercive et $\nabla V(\vartheta)$ est Lipschitz.
    b. $\langle \nabla V, h \rangle \geq 0$. (*Intuition*: $\dot{\vartheta} = -h(\vartheta)$ donc $\dfrac{d}{dt} V(\vartheta(t)) = -\langle \nabla V, h \rangle \leq 0$)
    c. $|\nabla V|^2 \leq C(1 + V)$, donc V est à croissance sous-quadratique.
2. $\|H(\vartheta, Z)\|_2 \leq C(1 + V(\vartheta))^{1/2}$.
3. $\sum_{n\geq 1} \gamma_n = \infty$ et $\sum_{n\geq 1} \gamma_n^2 < \infty$.
4. $V(\vartheta_0) \in L^1(\mathbb{P})$ et $\vartheta_0 \perp\!\!\!\perp (Z_n)$.

Alors

1. $V(\vartheta_n) \overset{p.s.}{\to} V_\infty \in L^1$.
2. $(V(\vartheta_n))_{n\geq 1}$ est L^1-bornée.
3. $\sum_{n\geq 1} \gamma_n \langle \nabla V, h \rangle (\vartheta_n) \in L^1$ et donc finie p.s.
4. $\sum_{n\geq 1} |\Delta \vartheta_n|^2 \in L^1$ et donc finie p.s.

Robbins–Monro

On cherche la racine ϑ^* telle que $h(\vartheta^*) = 0$.

Robbins–Monro = Robbins–Siegmund où $V(\vartheta) = \dfrac{1}{2}|\vartheta - \vartheta^*|^2$.

Théorème

1. Pour tout $\vartheta \neq \vartheta^*$ on a $\langle h(\vartheta), \vartheta - \vartheta^* \rangle > 0$.
2. Pour tout $\vartheta \in \mathbb{R}^d$ on a $\|H(\vartheta, Z)\|_2 \leq C(1 + |\vartheta - \vartheta^*|^2)^{1/2}$.
3. $\sum_{n \geq 1} \gamma_n = \infty$ et $\sum_{n \geq 1} \gamma_n^2 < \infty$.
4. $\vartheta_0 \in L^2(\mathbb{P})$ et $\vartheta_0 \perp\!\!\!\perp (Z_n)$.

Alors $\{h = 0\} = \{\vartheta^*\}$ et $\vartheta_n \to \vartheta^*$ p.s. et dans $L^p(\mathbb{P})$ pour $0 < p < 2$. De plus, $(|\vartheta_n - \vartheta^*|)_{n \geq 0}$ est $L^2(\mathbb{P})$-bornée.

Gradient stochastique

On cherche un minimum de $V(\vartheta)$.

Gradient stochastique = Robbins–Siegmund où $h(\vartheta) = \nabla V(\vartheta)$.

Théorème

1. $V(\vartheta)$ coercive, $\nabla V(\vartheta)$ est Lipschitz, $\{\vartheta\colon\ \nabla V(\vartheta) = 0\} = \{\vartheta^*\}$.
2. $\nabla V(\vartheta) = \mathbb{E}\left[H(\vartheta, Z)\right]$ avec $\|H(\vartheta, Z\|_2 \leq C(1 + V(\vartheta))^{1/2}$.
3. $\sum_{n \geq 1} \gamma_n = \infty$ et $\sum_{n \geq 1} \gamma_n^2 < \infty$.
4. $V(\vartheta_0) \in L^1(\mathbb{P})$ et $\vartheta_0 \perp\!\!\!\perp (Z_n)$.

Alors $\vartheta_n \to \vartheta^*$ p.s. et $|\nabla V(\vartheta_n)| \xrightarrow{L^p} 0$ pour $0 < p < 2$.