

Algorithmes stochastiques

Vitesses de convergence des algorithmes de gradient stochastiques

A. Godichon-Baggioni

Convergence presque sûre

APPROCHE DIRECTE

Théorème

On suppose que la fonction G est strictement convexe, i.e pour tout $h \neq m$,

$$\langle \nabla G(h), h - m \rangle > 0$$

*et que l'hypothèse **(PS0)** est vérifiée, i.e pour tout h*

$$\mathbb{E} \left[\|\nabla_h g(X, h)\|^2 \right] \leq C \left(1 + \|h - m\|^2 \right).$$

Alors

$$m_n \xrightarrow[n \rightarrow +\infty]{p.s} m.$$

APPROCHE VIA LE DÉVELOPPEMENT DE TAYLOR

Théorème

On suppose que m est l'unique zéro du gradient et l'unique minimiseur de G . On suppose également qu'il existe C, C' tels que pour tout h ,

$$\|\nabla^2 G(h)\|_{op} \leq C \quad \text{et} \quad \mathbb{E} \left[\|\nabla_h g(X, h)\|^2 \right] \leq C' (1 + G(h) - G(m)).$$

Alors

$$m_n \xrightarrow[n \rightarrow +\infty]{p.s.} m.$$

EXERCICE : APPROCHE LYAPUNOV

Démontrer le théorème suivant :

Théorème

On suppose qu'il existe une fonction $V : \mathbb{R}^d \rightarrow \mathbb{R}_+$ vérifiant :

- $V(m) = 0$ et $\forall h \neq m, V(h) \neq 0$
- Il existe une constante C telle que pour tout h

$$\mathbb{E} \left[\|\nabla_h g(X, h)\|^2 \right] \leq C(1 + V(h))$$

- Il existe $\alpha > 0$ tel que pour tout h

$$\langle \nabla G(h), \nabla V(h) \rangle \geq \alpha V(h)$$

- Le gradient de V est L -lipschitz : pour tout h, h' ,

$$\|\nabla L(h) - \nabla L(h')\| \leq L\|h - h'\|$$

Alors

$$m_n \xrightarrow[n \rightarrow +\infty]{p.s} m.$$

APPLICATION AU MODÈLE LINÉAIRE

Théorème

Si X admet un moment d'ordre 4, ϵ admet un moment d'ordre 2 et si $\mathbb{E}[XX^T]$ est définie positive, alors

$$\theta_n \xrightarrow[n \rightarrow +\infty]{p.s} \theta.$$

APPLICATION À LA RÉGRESSION LOGISTIQUE

Théorème

On suppose que X admet un moment d'ordre 2 et que la Hessienne de G en θ est positive. Alors

$$\theta_n \xrightarrow[n \rightarrow +\infty]{p.s} \theta.$$

Vitesses de convergence presque sûre

CADRE

On considère une suite de pas de la forme $\gamma_n = c_\gamma n^{-\alpha}$ avec $c_\gamma > 0$ et $\alpha \in (1/2, 1)$. On suppose que les hypothèses suivantes sont vérifiées :

(PS1) Il existe $\eta > \frac{1}{\alpha} - 1$ et C_η tels que pour tout h ,

$$\mathbb{E} \left[\|\nabla_h g(X, h)\|^{2+2\eta} \right] \leq C_\eta \left(1 + \|h - m\|^{2+2\eta} \right).$$

(PS2) La fonction G est deux fois continument différentiable sur une voisinage de m et

$$\lambda_{\min} := \lambda_{\min} (\nabla^2 G(m)) > 0.$$

VITESSE DE CONVERGENCE

Théorème

On suppose que les hypothèses (PS1) et (PS2) sont vérifiées. Alors

$$\|m_n - m\|^2 = O\left(\frac{\ln n}{n^\alpha}\right) \quad p.s.$$

APPLICATION AU MODÈLE LINÉAIRE

Théorème

Soit $\eta > 0$. On suppose que X admet un moment d'ordre $4 + 4\eta$ et que ϵ admet un moment d'ordre $2 + 2\eta$. De plus on suppose que $\mathbb{E}[XX^T]$ est positive. Alors

$$\|\theta_n - \theta\|^2 = O\left(\frac{\ln n}{n^\alpha}\right) \quad p.s.$$

APPLICATION AU MODÈLE LINÉAIRE

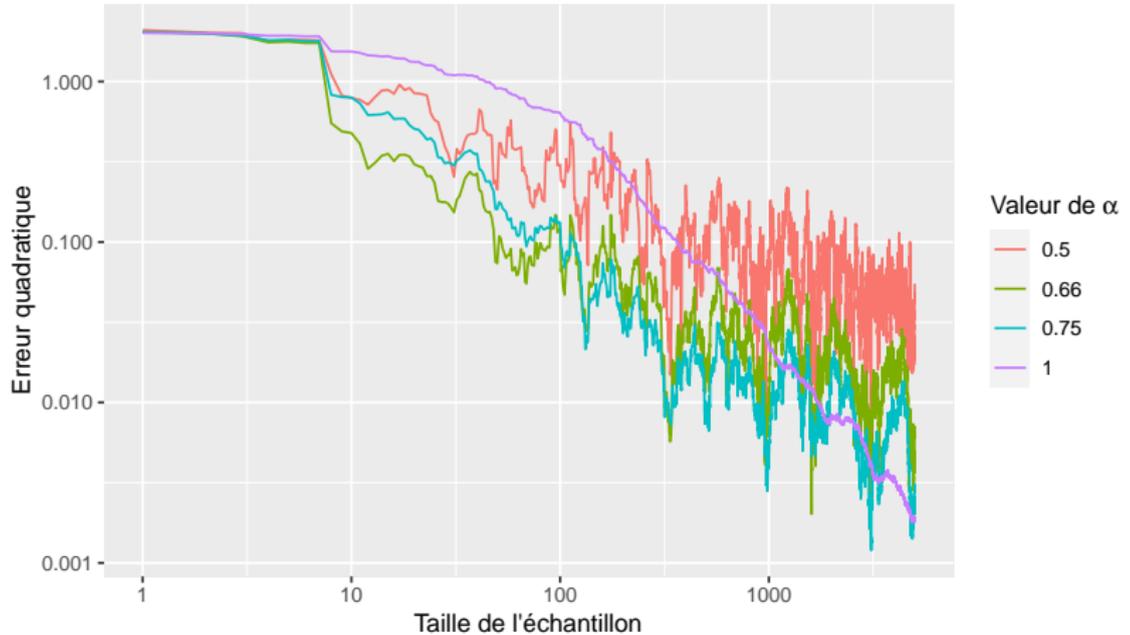


FIGURE – Evolution de l'erreur quadratique de θ_n en fonction de la taille d'échantillon n dans le cadre de la régression linéaire.

APPLICATION AU MODÈLE LINÉAIRE

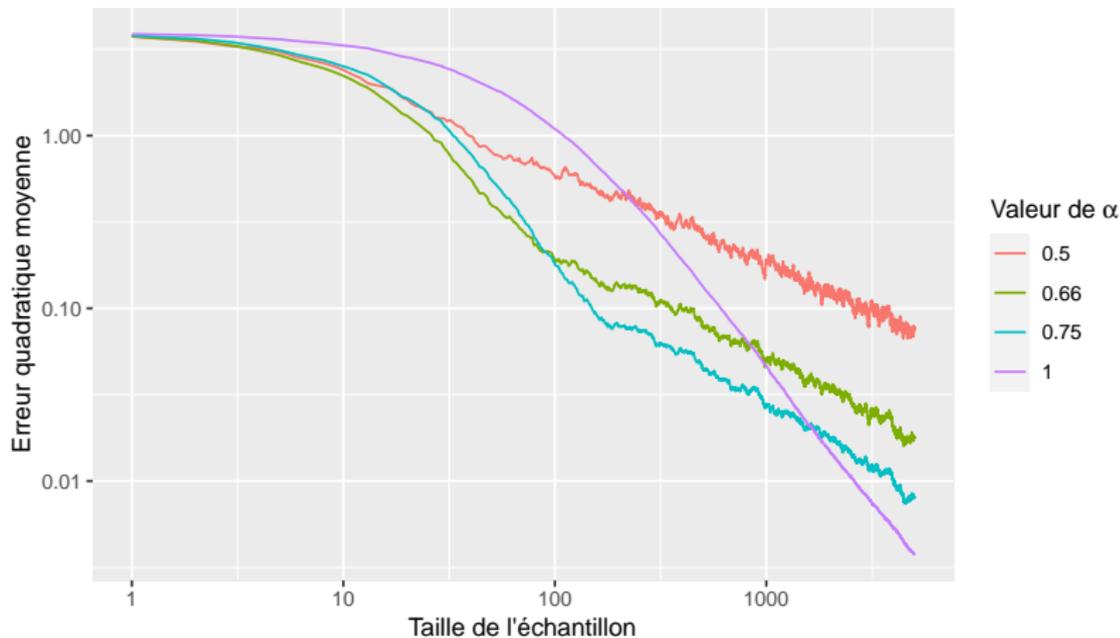


FIGURE – Evolution de l'erreur quadratique moyenne de θ_n en fonction de la taille d'échantillon n et du choix de α dans le cadre de la régression linéaire.

APPLICATION À LA RÉGRESSION LOGISTIQUE

Théorème

On suppose qu'il existe $\eta > 0$ tel que X admette un moment d'ordre $2 + 2\eta$. On suppose également que $\nabla^2 G(\theta)$ est positive. Alors

$$\|\theta_n - \theta\|^2 = O\left(\frac{\ln n}{n^\alpha}\right) \quad p.s.$$

APPLICATION À LA RÉGRESSION LOGISTIQUE

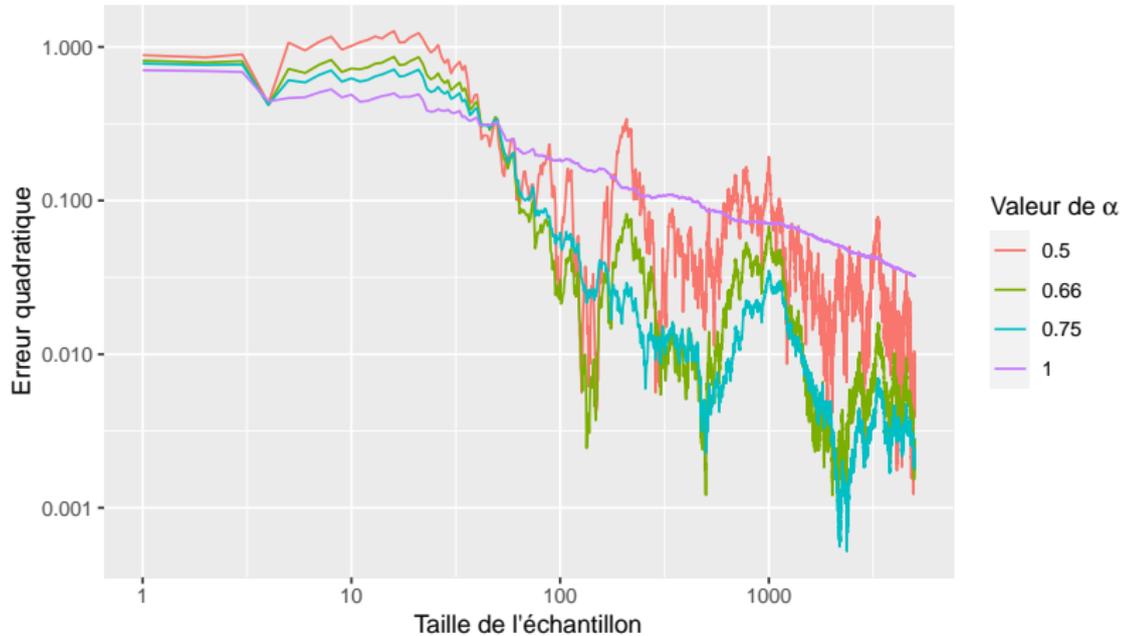


FIGURE – Evolution de l'erreur quadratique de θ_n en fonction de la taille de l'échantillon n dans le cadre de la régression logistique.

APPLICATION À LA RÉGRESSION LOGISTIQUE

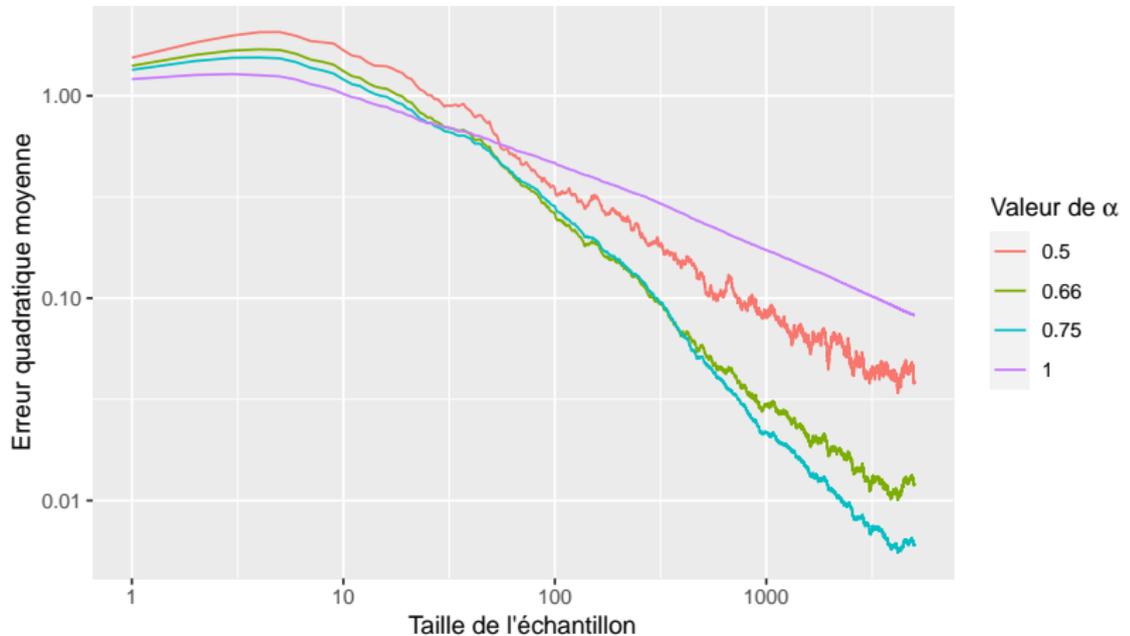


FIGURE – Evolution de l'erreur quadratique moyenne de θ_n en fonction de la taille de l'échantillon n et du choix du paramètre α dans le cadre de la régression logistique.

REMARQUES

En prenant $\gamma_n = c_\gamma n^{-1}$ et $c_\gamma > \frac{1}{2\lambda_{\min}}$, on peut montrer

$$\sqrt{n} (m_n - m) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \Sigma_{RM})$$

avec

$$\Sigma_{RM} = c_\gamma \int_0^{+\infty} e^{-s(H - \frac{1}{2c_\gamma} I_d)} \Sigma e^{-s(H - \frac{1}{2c_\gamma} I_d)} ds$$

avec $\Sigma = \mathbb{E} \left[\nabla_h g(X, m) \nabla_h g(X, m)^T \right]$.

EXERCICE

► Modèle linéaire :

On prend $\theta = (-2, -1, 0, 1, 2)^T$, $X_i \sim \mathcal{N}(0, I_5)$ et $\epsilon_i \sim \mathcal{N}(0, 1)$.

1. Générer un échantillon $(X_i, Y_i)_{i=1, \dots, n}$ avec $n = 10000$.
2. Ecrire une fonction qui ressorte l'ensemble de tous les estimateurs $\theta_i, i = 0, \dots, n$.
3. Tracer la courbe de l'évolution de l'erreur quadratique.
4. Générer maintenant 50 échantillons et tracer l'évolution de l'erreur quadratique moyenne.

► Faire de même pour la régression logistique.