

Méthodes d'accélération

A. Godichon-Baggioni

Algorithme de gradient stochastique moyenné

DÉFINITION

Fonction à minimiser :

$$G(h) = \mathbb{E}[g(X, h)]$$

Algorithme de gradient stochastique : Pour tout $n \geq 0$

$$m_{n+1} = m_n - \gamma_{n+1} \nabla_h g(X_{n+1}, m_n)$$

Algorithme moyenné:

$$\bar{m}_n = \frac{1}{n+1} \sum_{k=0}^n m_k$$

où les m_k sont les estimateurs de gradient stochastique.

Ecriture récursive:

$$\bar{m}_{n+1} = \bar{m}_n + \frac{1}{n+2} (m_{n+1} - \bar{m}_n).$$

LEMME DE TOEPLITZ

Lemme

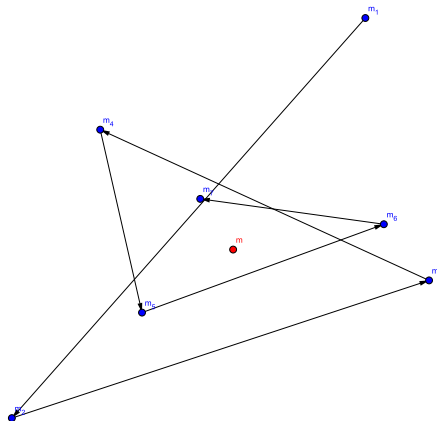
Soit (a_n) positive telle que $\sum_{n \geq 0} a_n = +\infty$ et X_n une suite de variables aléatoires convergeant presque sûrement vers X . Alors

$$\frac{1}{\sum_{k=0}^n a_k} \sum_{k=0}^n a_k X_k \xrightarrow[n \rightarrow +\infty]{p.s.} X.$$

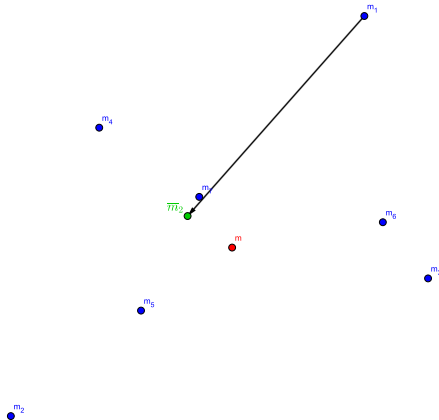
Application :

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

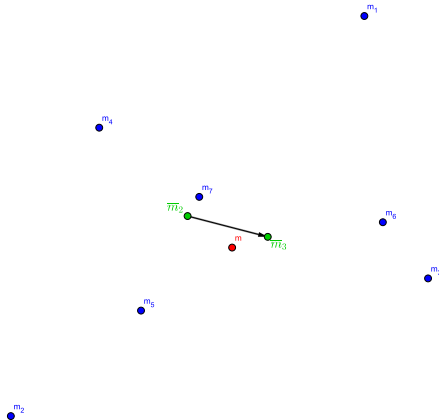
COMMENT ÇA MARCHE?



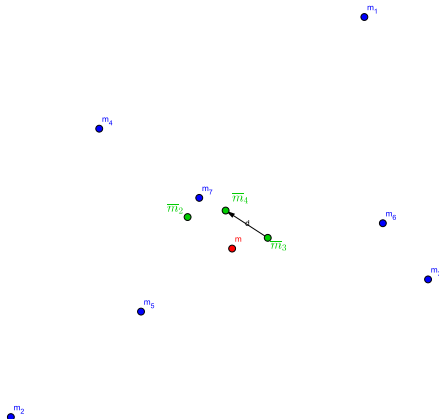
COMMENT ÇA MARCHE?



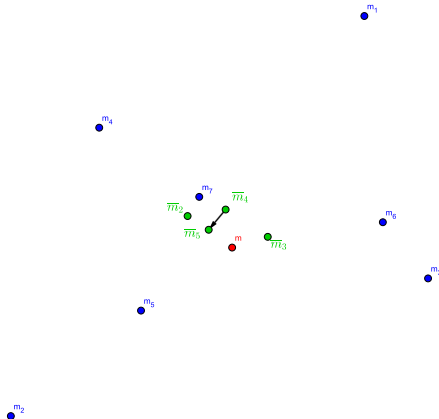
COMMENT ÇA MARCHE?



COMMENT ÇA MARCHE?

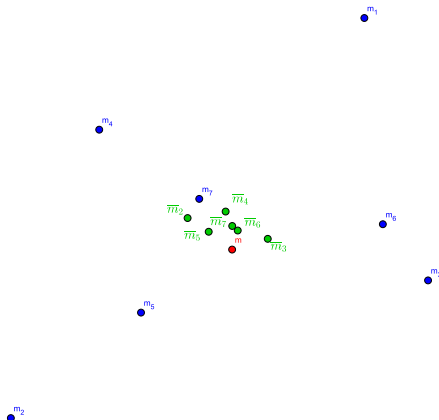


COMMENT ÇA MARCHE?



COMMENT ÇA MARCHE?

COMMENT ÇA MARCHE?



EFFICACITÉ ASYMPTOTIQUE

Théorème

On suppose que les hypothèses (PS1) à (PS4) sont vérifiées. Alors

$$\sqrt{n} (\overline{m}_n - m) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N} \left(0, H^{-1} \Sigma H^{-1} \right)$$

avec $H = \nabla^2 G(m)$ et $\Sigma = \Sigma(m)$.

RÉGRESSION LINÉAIRE

Algorithme moyenné:

$$\theta_{n+1} = \theta_n + \gamma_{n+1} \left(Y_{n+1} - X_{n+1}^T \theta_n \right) X_{n+1}$$

$$\bar{\theta}_{n+1} = \bar{\theta}_n + \frac{1}{n+2} \left(\theta_{n+1} - \bar{\theta}_n \right)$$

avec $\bar{\theta}_0 = \theta_0$.

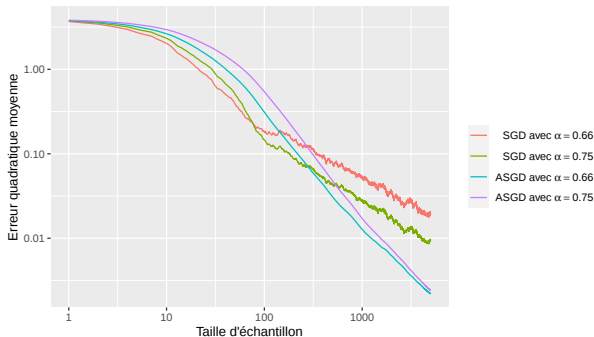


Figure: Evolution de l'erreur quadratique moyenne.

TESTER $H_0: \theta = \theta_0$ "EN LIGNE"

Réécriture du TLC: Sous H_0 ,

$$\sqrt{n} \frac{(\bar{\theta}_n - \theta_0)^T H (\bar{\theta}_n - \theta_0)}{\sigma^2} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \chi_d^2$$

Application: Soit $\bar{H}_n$ et $\hat{\sigma}_n^2$ des estimateurs consistants. Alors

$$C_n := \sqrt{n} \frac{(\bar{\theta}_n - \theta_0)^T \bar{H}_n (\bar{\theta}_n - \theta_0)}{\hat{\sigma}_n^2} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \chi_d^2$$

SIMULATIONS

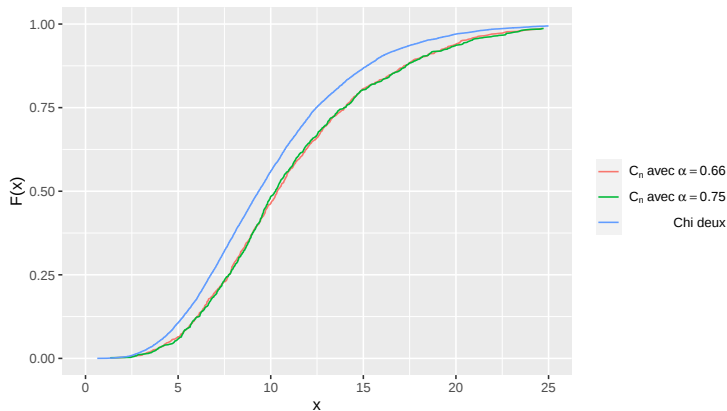


Figure: Comparaison de la fonction de répartition de C_n avec $n = 5000$, pour $\alpha = 0.66$ et $\alpha = 0.75$, et de celle d'une Chi 2 à 10 degrés de liberté dans le cadre du modèle linéaire.

Algorithme de Newton stochastique

Idée

Modèle linéaire:

$$Y = \theta^T X + \epsilon$$

Gradient stochastique: Posons $H = \mathbb{E} [XX^T] = \begin{pmatrix} 10^{-1} & 0 \\ 0 & 10^1 \end{pmatrix}$.

$$\mathbb{E} [\theta_{n+1} - \theta] = \begin{pmatrix} 1 - \gamma_{n+1} 10^{-1} & 0 \\ 0 & 1 - \gamma_{n+1} 10^1 \end{pmatrix} \mathbb{E} [\theta_n - \theta]$$

Choix de γ_n ?

SIMULATIONS

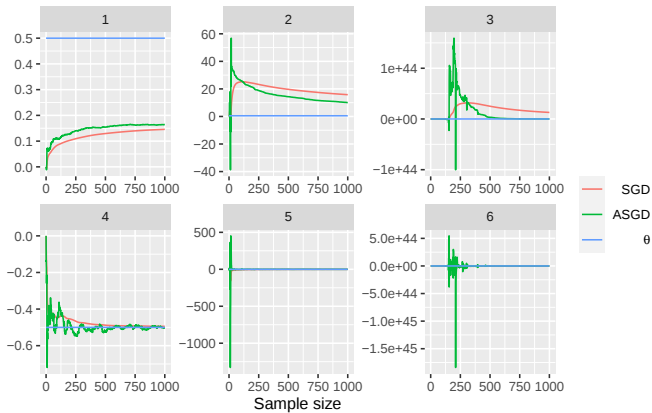


Figure: Evolution des estimateurs de la première coordonnée (en haut) et de la deuxième (en bas) pour, de gauche à droite, $c_\gamma = 0.1$, $c_\gamma = 1$ et $c_\gamma = 10$.

SIMULATIONS (NEWTON STOCHASTIQUE)

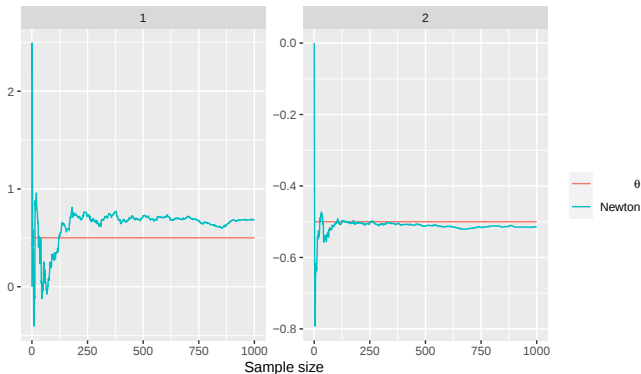


Figure: Evolution des estimateurs de la première coordonnée (à gauche) et de la deuxième coordonnée (à droite) pour l'algorithme de Newton stochastique.

ALGORITHME DE NEWTON STOCHASTIQUE

Algorithme de Newton stochastique:

$$m_{n+1} = m_n - \frac{1}{n+1} \bar{H}_n^{-1} \nabla_h g(X_{n+1}, m_n)$$

Hypothèses sur $\bar{H}_n$:

- ▶ $\bar{H}_n^{-1}$ est symétrique et définie positive.
- ▶ Il existe une filtration $(\mathcal{F}_n)$ telle que
 - ▶ $\bar{H}_n$ et m_n sont $\mathcal{F}_n$ mesurables.
 - ▶ X_{n+1} est indépendant de $\mathcal{F}_n$.
- ▶ On peut contrôler les valeurs propres de $\bar{H}_n$:

$$\lambda_{\max}(\bar{H}_n) = O(1) \quad p.s$$

$$\lambda_{\max}(\bar{H}_n^{-1}) = O(n^\beta) \quad p.s$$

avec $\beta < 1/2$.

- ▶ **(H1)** $\implies \liminf \lambda_{\min}(\bar{H}_n^{-1}) > 0 \text{ p.s.}$

CONVERGENCE

Théorème

On suppose que les hypothèses sur les estimateurs de Hessiennes et les suivantes sont vérifiées :

- *Il existe une constante C telle que*

$$\forall h \in \mathbb{R}^d, \quad \mathbb{E} \left[\|\nabla_h g(X, h)\|^2 \right] \leq C + C (G(h) - G(m))$$

- *La fonction G est deux fois continument différentiable sur un voisinage de m et*

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

- *La Hessienne de G est uniformément bornée: il existe $L_{\nabla G}$ tel que*

$$\forall h \in \mathbb{R}^d, \quad \left\| \nabla^2 G(h) \right\|_{op} \leq L_{\nabla G}$$

Alors

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

EFFICACITÉ ASYMPTOTIQUE

Théorème

Sous certaines hypothèses,

$$\sqrt{n} (m_n - m) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N} \left(0, H^{-1} \Sigma H^{-1} \right)$$

avec $H = \nabla^2 G(m)$ et $\Sigma = \Sigma(m)$.

MODÈLE LINAIRE

Formule de Riccati: Soit A une matrice définie positive, pour tout $u \in \mathbb{R}^d$ et $\lambda \geq 0$, on a $1 + \lambda u^T A^{-1} u \geq 1$ et donc

$$(A + \lambda uu^T)^{-1} = A^{-1} - \lambda (1 + \lambda u^T A^{-1} u)^{-1} A^{-1} uu^T A^{-1}.$$

Algorithme de Newton stochastique:

$$\theta_{n+1} = \theta_n + \frac{1}{n+1} \bar{H}_n^{-1} (Y_{n+1} - X_{n+1}^T \theta_n) X_{n+1}$$

$$H_{n+1}^{-1} = H_n^{-1} - \left(1 + X_{n+1}^T H_n^{-1} X_{n+1}\right)^{-1} H_n^{-1} X_{n+1} X_{n+1}^T H_n^{-1}$$

avec H_0 positive et $\bar{H}_n^{-1} = (n+1)H_n^{-1}$.

Réécriture de $\bar{H}_n$:

$$\bar{H}_n = \frac{1}{n+1} \left(H_0 + \sum_{k=1}^n X_k X_k^T \right).$$

SIMULATIONS

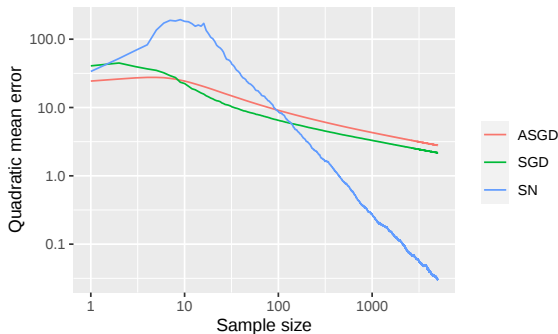


Figure: Evolution de l'erreur quadratique moyenne des estimateurs de gradient θ_n (SGD), de leur version moyennée $\bar{\theta}_n$ (ASGD) et des estimateurs de Newton stochastique $\hat{\theta}_n$ (SN) en fonction de la taille de l'échantillon dans le cadre du modèle linéaire.