Présentation d'article - Convergence of Markovian Stochastic Approximation with Discontinuous Dynamics

G. FORT, E. MOULINES ,A. SCHRECK AND M. VIHOLA

Cours de Statistiques Computationnelles (2020-2021)

A. Stéphanovitch et L. Bleistein

January 4, 2021

Contents

Le problème

Analyse de la convergence de l'algorithme d'approximation stochastique de la forme

$$\theta_{n+1} = \theta_n + \gamma_{n+1} H_{\theta_n}(X_{n+1})$$

sous des hypothèses faibles sur la régularité de $\theta \mapsto H_\theta(x)$.

$\{\gamma_n, \in \mathbb{N}\}$: suite déterministe de pas.
$\{X_n, n \in \mathbb{N}\}$: chaîne de Markov contrôlée.

Cet algorithme a pour but de **trouver les zéros de** $h : \Theta \to \mathbb{R}^d$ **définie par**

$$h(\theta) = \int_X H_\theta(x) \pi_\theta(dx)$$

pour π_θ la probabilité invariante de P_θ, Θ un ouvert de $\mathbb{R}^d$.

Résultats de l’article

- Démontrer la convergence de cet algortihme sous des hypothèses faibles sur la régularité de $\theta \mapsto H_\theta(x)$.
- Application à des exemples avec dynamiques discontinues.
- Nous reproduisons 3 des 4 exemples (code disponible sur Github).

Algorithme

Input: une suite de compacts K_i telle que $\bigcup_{i=1}^{\infty} K_i = \Theta$, une suite de pas $\{\gamma_n, n \in \mathbb{N}\}$, un point d'initialisation (x_*, θ_*).

Tant que $\theta_n \in K_0$**:** on fait tourner l'algortihme d'approximation stochastique ($\theta_{n+1} = \theta_n + \gamma_{n+1} H_{\theta_n}(X_{n+1})$).

Si θ_n **sort du compact** K_0**:** on réinitialise la suite avec (x, θ), et on recommence à faire tourner l'algorithme avec la suite de pas $\{\gamma_{1+k}, k \in \mathbb{N}\}$ et en remplaçant K_0 par K_1.

On répète avec K_i et $\{\gamma_{i+k}, k \in \mathbb{N}\}$ jusqu'à convergence.

Hypothèses

H1 La fonction $(x, \theta) \mapsto H_\theta(x)$ est mesurable de $X \times \theta$ dans $\mathbb{R}^d$. Il existe une fonction mesurable $W : X \to [1, \infty)$ telle que pour tout compact $K \subset \Theta$,

$$\sup_{\theta \in K} |H_\theta|_W < \infty$$

où $|f|_W = sup_X |f|/W$.

H2 (a) Pour tout $\theta \in \Theta$, le noyau P_θ a une unique distribution invariant π_θ .

(b) Pour tout compact $K \subset \Theta$, il existe des constantes $C > 0$ et $\lambda \in (0, 1)$ telles que $\forall x \in X, l \geq 0$,

$$sup_{\theta \in K} ||P_\theta^l(x, \cdot) - \pi_\theta||_W \leq C\lambda^l W(x)$$

et $sup_{\theta \in K} \pi_\theta(W) < \infty$ où $||\mu||_W = sup_{\{f, |f|_W < 1\}} |\mu(f)|$.

(c) Il existe $p > 1$ et pour tout compact $K \subset \Theta$, il existe des constantes $\rho \in (0, 1)$ et $b < \infty$ telles que $sup_{\theta \in K} P_\theta W^p(x) \leq \rho W^p(x) + b$.

Hypothèses

H3 Pour $D_W(\theta, \theta^{'}) = sup_{x\in X} \frac{||P_\theta(x,\cdot) - P_{\theta'}(x,\cdot)||_W}{W(x)}$, il existe $\nu \in (0,1]$ telle que pour tout compact $K \in \Theta$, $\sup\limits_{\theta,\theta' \in K} \frac{D_W(\theta,\theta')}{|\theta-\theta'|^\nu} < \infty$.

H4 Soit $\alpha \in (0,1]$, pour tout compact $K \subset \Theta$, il existe une constant $C > 0$ telle que $\forall\theta > 0$,

$$\sup_{\theta\in K} \int \pi_\theta(dx) \sup_{|\theta'-\theta|\leq\delta} |H_{\theta'}(x) - H_\theta(x)| \leq C\delta^\alpha$$

.

Hypothèses

H5 Il existe une fonction C^1 $w : \Theta \to [0, \infty)$ telle que:
(a) $\forall M > 0$ l'ensemble de niveau $\{\theta \in \Theta, w(\theta) \leq M\}$ est compact.
(b) L'ensemble des points stationnaires $\mathcal{L} = \{\theta \in \Theta, \langle \nabla w(\theta), h(\theta) \rangle = 0\}$ est compact.
(c) Pour tout $\theta \in \Theta \backslash \mathcal{L}$ on a $\langle \nabla w(\theta), h(\theta) \rangle < 0$.

H6 La suite de pas est de la forme $\{\gamma_0/(n+1)^\beta, n \in \mathbb{N}\}$ avec $\gamma_0 > 0$ et
$\beta \in \left(max(\frac{1}{p}, \frac{1+(\alpha \wedge \nu)/p}{1+(\alpha \wedge \nu)}); 1 \right]$.

Théorème principal de l'article

Soit $\{K_n, n \in \mathbb{N}\}$ une suite croissante et recouvrante de compacts et $(x_*, \theta_*) \in X \times K_0$. En supposant H1 à H6, la suite $\{(X_n, \theta_n), n \in \mathbb{N}\}$ donnée par l'algorithme commencé en (x_*, θ_*) est stable :

$$\mathbb{P}\left(\bigcup_{i=0}^{\infty} \bigcap_{k=0}^{\infty} \{\theta_k \in K_i\}\right) = \mathbb{P}\left(\exists\, i \text{ s.t. } \forall k, \theta_k \in K_i\right) = 1.$$

De plus, pour $\mathcal{L} = \{\theta \in \Theta |\ \langle \nabla w(\theta), h(\theta) \rangle = 0\}$, la suite θ_k converge vers une composante connexe de $\mathcal{L}$:

$$\mathbb{P}\left(\lim_{k\to\infty} d(\theta_k, \mathcal{L}) = 0,\ \lim_{k\to\infty} ||\theta_{k+1} - \theta_k|| = 0\right) = 1.$$

Simulations (1)

1. **Calcul de quantile:**

$$\text{Trouver } t \text{ t.q. } F_X(t) = q \in [0,1].$$
$$H_\theta(x) := q - \mathbb{1}_{x \leq \theta}$$

2. **Calcul de médiane multidimensionnelle:**

$$m^* = \text{argmin}_{m \in \mathbb{R}^d} \mathbb{E}\left(||X - m||\right).$$
$$H_\theta(x) := \frac{x - \theta}{||x - \theta||} \mathbb{1}_{x \neq \theta}.$$

3. **Quantilisation (calcul de centres de Voronoï):** trouver $\theta^* = (\theta_1, \ldots, \theta_N) \in (\mathbb{R}^d)^N$ t.q.

$$\theta^* = \text{argmin}_{\theta \in (\mathbb{R}^d)^N} \sum_{i=1}^{N} \mathbb{E}\left(||X - \theta_i|| \mathbb{1}_{C_i(\theta)}(X)\right)$$
$$H_\theta(x) := -2\left((\theta_i - x)\mathbb{1}_{x \in C_i(\theta)}\right)_{1 \leq i \leq N} - \mu \left(\sum_{i \neq j} \frac{\theta_i - \theta_j}{||\theta_i - \theta_j||}\right)_{1 \leq i \leq N}, \mu > 0.$$

Simulations (2)

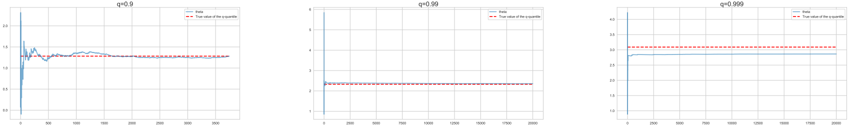


Figure: Estimation de quantiles de gaussienne, pour différents quantiles.

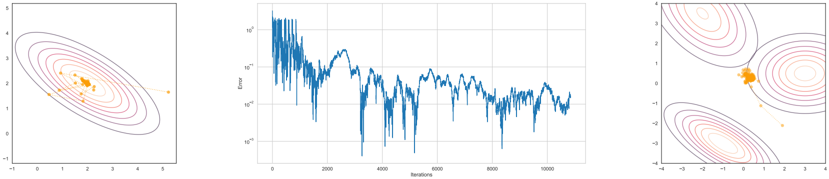
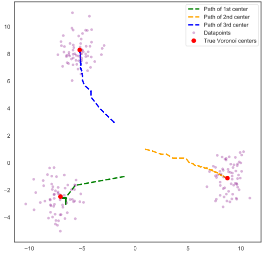


Figure: Estimation de médiane pour une gaussienne et un mélange de gaussiennes en 2D. L'erreur correspond à l'estimation de la gaussienne simple (gauche).

Simulations (3)



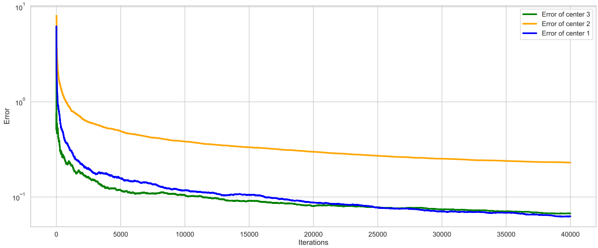


Figure: Estimation de centres de Voronoï, pour 3 Blobs 2D générés avec sklearn.makeblobs.

Conclusion

Questions ouvertes:

1. Vitesse de convergence de l’algorithme ?
2. Critère d’arrêt de l’algorithme ?
3. Comment choisir les (nombreux) hyperparamètres ?