Almost surely convergence of stochastic approximation under general moment assumption

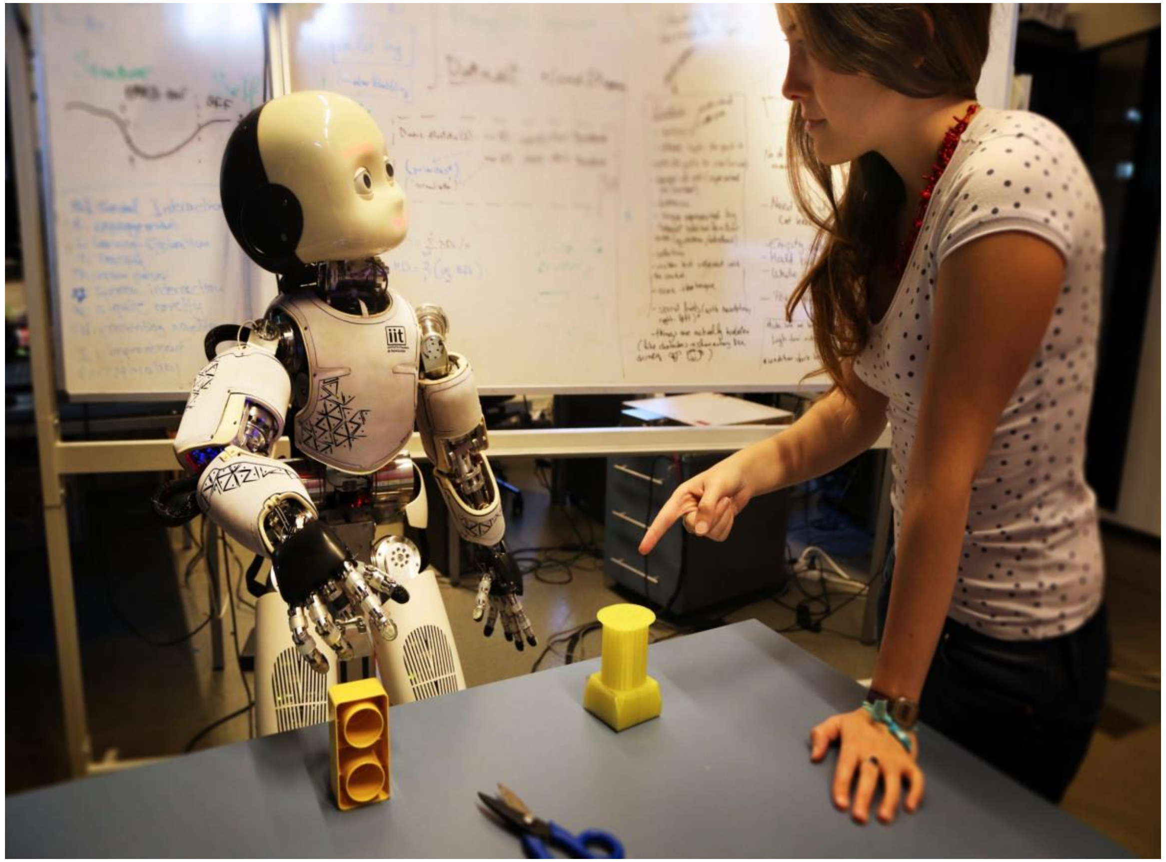
Anh Duc Nguyen, NUS
Quang Nguyen, UIT VNUHCM
Hoang Huy Nguyen, ISyE Georgia Tech
Siva Theja Maguluri, ISyE Georgia Tech

$$\alpha_k = \alpha(k+K)^{-\xi}$$

Stochastic Approximation

Main problem: the fixed point equation with x^* as the solution

$$H(x) = x$$



$\alpha_k = \alpha(k+K)^{-\xi}$

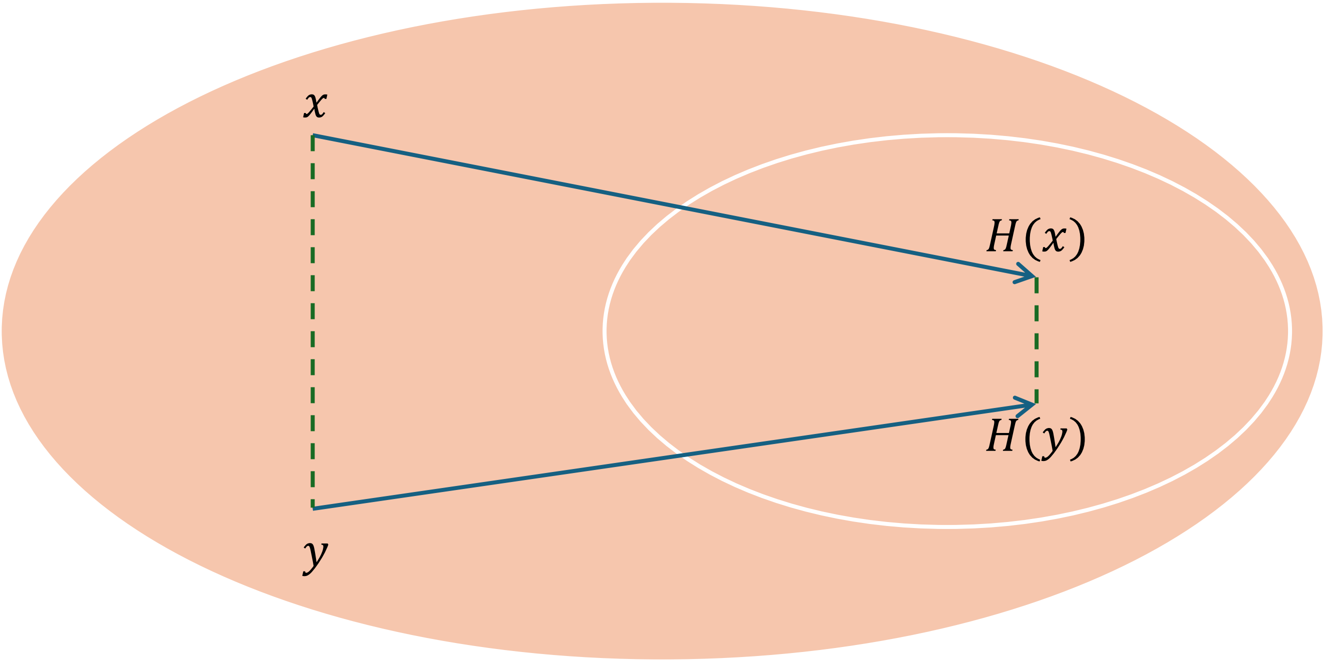
Assumption

Stochastic Approximation Algorithm

$$x_{k+1} = x_k + \alpha_k(H(x_k) - x_k \qquad)$$

Assumption 1: Contractive operator $||H(x) - H(y)||_2 \leq \rho||x-y||_2 \ (\rho < 1)$

Assumption 2: Noise condition $E[w_k] = 0, E\left[||w_k||_2^p\right] < \infty$



$$\boxed{\alpha_k = \alpha(k+K)^{-\xi}}$$

Problem Formulation

Stochastic Approximation Algorithm

$$x_{k+1} = x_k + \alpha_k\big(H(x_k) - x_k + w_k\big)$$

$\boldsymbol{\alpha_k} = (k+1)^{-1}$

$\boldsymbol{\alpha_k} = \alpha(k+K)^{-\xi}$

Law of large number

Stochastic Approximation Algorithm

$$x_{k+1} = x_k + \alpha_k(\qquad\qquad - x_k + w_k)$$

Conjecture:

$$1 \geq \xi > \frac{1}{p} \Rightarrow x_k \xrightarrow{a.s} x^*$$

Generalize LLN !

Law of Large Number

$$x_{k+1} = \frac{\sum_{0\leq i\leq k} w_i}{k+1} \xrightarrow{a.s} 0$$

$$x_{k+1} = (1-\alpha_k)x_k + \alpha_k w_k = \sum c_i w_i \xrightarrow{a.s} 0$$

Method: Truncation

$\alpha_k = \alpha(k+K)^{-\xi}$

Truncation method

Stochastic Approximation Algorithm

$$x_{k+1} = x_k + \alpha_k(H(x_k) - $$

Non-linear SA ✕

Generalized Law of Large Number

$$x_{k+1} = (1-\alpha_k)x_k + \alpha_k w_k = \sum c_i w_i \xrightarrow{a.s} 0$$

Method: Truncation

Truncate

$$\widetilde{w}_k = w_k 1_{|w_k| \le \boxed{B_k}} \Longrightarrow \tilde{x}_{k+1} = (1-\alpha_k)\tilde{x}_k + \alpha_k \widetilde{w}_k$$

Borel Cantelli

$$\tilde{x}_k \to 0 \text{ a.s} \Longrightarrow \sum P(|\tilde{x}_k| > \epsilon) < \infty \ \forall \epsilon$$

$$\tilde{x}_k - x_k \to 0 \text{ a.s} \Longrightarrow \sum P(|\widetilde{w}_k| > \boxed{B_k}) < \infty$$

$\alpha_k = \alpha(k+K)^{-\xi}$

Drift method

Stochastic Approximation Algorithm

$$x_{k+1} = x_k + \alpha_k(H(x_k) - x_k + w_k)$$

Method: Drift approach

Supermartingale Convergence Theorem

$$\begin{cases} E[Y_{k+1}|F_k] \le Y_k - X_k + Z_k \\ \sum \mathrm{Z_k} < \infty, \mathrm{Y_k}, \mathrm{X_k}, \mathrm{Z_k} \ge 0 \end{cases} \Longrightarrow \begin{cases} \mathrm{Y_k} \to \mathrm{Y}^* \ \text{a.s.} \\ \sum \mathrm{X_k} < \infty \ \text{a.s.} \end{cases}$$

Proof recipe

1) Pick a suitable Lyapunov function
2) Establish 1-step Drift
3) Apply Supermartingale Convergence Theorem

$\alpha_k = \alpha(k+K)^{-\xi}$

Challenges

Stochastic Approximation Algorithm

$$x_{k+1} = x_k + \alpha_k(H(x_k) - x_k + w_k)$$

Difficulties

$p = 2$

$$Y_k = V(x_k - x^*) = ||x_k - x^*||_2^2$$

$p < 2$

No second moment

$p > 2$

Second moment is insufficient $1 \geq \xi > \frac{1}{2} > \frac{1}{p}$

$\alpha_k = \alpha(k+K)^{-\xi}$

Main Result

Stochastic Approximation Algorithm

$$x_{k+1} = x_k + \alpha_k(H(x_k) - x_k + w_k)$$

Why Drift + Truncation for 1<p<2

Law of Large Numbers

Similar one-step drift template

Truncation for the moment handling

Stochastic Approximation

Robbins–Siegmund (RS)

$$\begin{cases} \{V_n\}, \{C_n\}, \{D_n\} \geq 0, \text{ filtration } \mathcal{F}_n \\ \sum C_n = \infty \text{ a.s.}, \sum D_n < \infty \text{ a.s.} \\ \mathbb{E}\left[V_{n+1} \mid \mathcal{F}_n\right] \leq \underbrace{(1 - C_n)}_{\text{Contraction}} V_n + \underbrace{D_n}_{\text{Noise}} \end{cases}$$

$$V_n \to 0 \text{ a.s.}$$

$$Y_n = V_n, X_n = C_n V_n, Z_n = D_n$$

Supermartingale Convergence

$$\begin{cases} \mathbb{E}[Y_{n+1} \mid \mathcal{F}_n] \leq Y_n - X_n + Z_n \\ \sum Z_n < \infty, \{Y_n\}, \{X_n\}, \{Z_n\} \geq 0 \end{cases}$$

$$Y_n \to Y^* \text{ a.s.}$$

$$\sum X_n < \infty \text{ a.s.}$$

LLN for $1 < p < 2$

Issue: No 2nd moment → **Solution:** Truncation

Claim: Let Y be a random variable with $\mathbb{E}Y^p < \infty$ for some $p \in (1,2)$. Then, for any threshold $t > 0$: $\mathbb{E}\left[Y^2 1_{\{|Y| \leq t\}}\right] \leq t^{2-p}\mathbb{E}Y^p$.

$$\mathbb{E}[\tilde{Y}^2] \leq t^{2-p}\mathbb{E}Y^p$$

Proof: Apply Hölder's inequality on the indicator of the truncation set.

What is the appropriate truncation threshold t?

LLN: One-step Drift

$$\alpha_n = \alpha(n+K)^{-\xi}$$
$$B_n = (n+1)^{\kappa}$$

Case $p = 2$:

$$\mathbb{E}[V_{n+1}|\mathcal{F}_n] \le (1 - c\alpha_n)V_n + C\alpha_n^2 \mathbb{E}[\varepsilon_{n+1}^2|\mathcal{F}_n]$$

$$(S_{n+1} - \mu)^2 \qquad \sum_n \alpha_n = \infty \qquad \sum_n \alpha_n^2 < \infty$$

Ours:

$$\mathbb{E}[W_{n+1}|\mathcal{F}_n] \le (1 - \alpha_n + \alpha_n^2)W_n + \alpha_n^2\, \mathbb{E}[\tilde{\varepsilon}_{n+1}^2|\mathcal{F}_n] + \alpha_n \mathbb{E}[\tilde{\varepsilon}_{n+1}|\mathcal{F}_n]^2$$

Extra **bias** term

Linearity in α_n

$$\left(\tilde{S}_{n+1} - \mu\right)^2$$

Companion process:

$$\tilde{S}_n = (1 - \alpha_n)\tilde{S}_n + \alpha_n(\mu + \tilde{\varepsilon}_{n+1})$$

$$\tilde{\varepsilon}_{n+1} = \varepsilon_{n+1} 1_{\{|\varepsilon_{n+1}| \le B_n\}}$$
$$\varepsilon_{n+1} = X_{n+1} - \mu$$

Due to truncation, $\tilde{\varepsilon}_{n+1}$ is **biased**!

Convergence of LLN

$$\alpha_n = \alpha(n+K)^{-\xi}$$
$$\tilde{S}_n = (1-\alpha_n)\tilde{S}_n + \alpha_n(\mu + \tilde{\varepsilon}_{n+1})$$

$$\mathbb{E}[W_{n+1}|\mathcal{F}_n] \leq (1-\alpha_n+\alpha_n^2)W_n + \boxed{\alpha_n^2\,\mathbb{E}[\tilde{\varepsilon}_{n+1}^2|\mathcal{F}_n]} + \boxed{\alpha_n\mathbb{E}[\tilde{\varepsilon}_{n+1}]^2}$$

$$B_n = (n+1)^\kappa$$
$$\tilde{\varepsilon}_{n+1} = \varepsilon_{n+1}1_{\{|\varepsilon_{n+1}|\leq B_n\}}$$

$\sum_n$ (constants C_1 & C_2) $\sum_n$

$$\sum_n C_1(n+1)^{-2\xi-\kappa(2-p)} \qquad \sum_n C_2(n+1)^{-\xi-2\kappa(p-1)}$$

$$\xi > \frac{1}{2} \qquad \kappa > \frac{1-\xi}{2(p-1)}$$

$$< \infty \qquad < \infty$$

By Robbins-Siegmund, $\tilde{S}_n \to \mu$ almost surely.

With $\boldsymbol{\kappa} > \mathbf{max}\left\{\frac{1-\xi}{2(p-1)}, \frac{1}{p}\right\}$ and Borel Cantelli lemma, we obtain $S_n \to \mu$ almost surely.

LLN to SA

$\alpha_n = \alpha(n+K)^{-\xi}, B_n = (n+1)^{\kappa}$
$\tilde{\varepsilon}_{n+1} = \varepsilon_{n+1} 1_{\{|\varepsilon_{n+1}|\}} \leq B_n$

LLN 1-step drift $\mathbb{E}[W_{n+1}|\mathcal{F}_n] \leq (1 - \alpha_n + \alpha_n^2)W_n + \alpha_n^2\, \mathbb{E}[\tilde{\varepsilon}_{n+1}^2 \big| \mathcal{F}_n] + \alpha_n \mathbb{E}[\tilde{\varepsilon}_{n+1}]^2$

SA 1-step drift $\mathbb{E}[U_{n+1}|\mathcal{F}_n] \leq \left(1 - \frac{1-\rho}{2}\alpha_n\right) U_n + \alpha_n^2 \mathbb{E}[\|\tilde{\varepsilon}_{n+1}\|^2|\mathcal{F}_n] + C\alpha_n \left\| \mathbb{E}[\tilde{\varepsilon}_{n+1}^2|\mathcal{F}_n] \right\|^2$

extra **bias**

$\|\tilde{x}_n - x^\star\|^2$

ρ-contractive

With $\kappa > \max\left\{\frac{1-\xi}{2(p-1)}, \frac{1}{p}\right\}$, we obtain $x_n \to x^\star$ almost surely.

$$\alpha_k = \alpha(k+K)^{-\xi}$$

Main Result

Stochastic Approximation Algorithm

$$x_{k+1} = x_k + \alpha_k(H(x_k) - x_k + w_k)$$

Result

	LLN	SA
$p < 2$	**Truncation + Drift**	**Truncation + Drift**
$p = 2$	**Drift**	**Drift**
$p > 2$	**Truncation + High moment bound**	*Future work*

$\alpha_k = \alpha(n+K)^{-\xi}$

LLN for $p > 2$

Generalized Law of Large Number

$$x_{k+1} = (1-\alpha_k)x_k + \alpha_k w_k = \sum c_i w_i \xrightarrow{a.s} 0$$

Borel-Cantelli

$$\sum P(|x_k| > \epsilon) < \infty \ \forall \epsilon \Longrightarrow \sum E[|x_k|^p] < \infty$$

Rosenthal

$$E[|x_k|^p] \le C_p\left(p^{\frac{p}{2}} E[|x_k|^2]^{\frac{p}{2}} + p^p \alpha_k^p E[|w_k|^p]\right)$$

Summary

	LLN	SA
$p < 2$	**Truncation + Drift**	**Truncation + Drift**
$p = 2$	**Drift**	**Drift**
$p > 2$	**Truncation + High moment bound**	*Future work*

Challenges solved:

- Picking right truncation threshold
- Handled biased SA
- Establish and analyze 1-step drift

THANK YOU

for your attention

LLN for $p > 2$

$$\mathrm{Var}(S_{b(n+1)})^{\frac{p}{2}} \leq 2^{\frac{p}{2}-1}\left[\left(\prod_{j=b(n)}^{b(n+1)-1}(1-\alpha_j)^2\right)^{\frac{p}{2}} \mathrm{Var}(S_{b(n)})^{\frac{p}{2}} + \left(\sum_{j=b(n)}^{b(n+1)-1}\alpha_j^2 \prod_{k=j+1}^{b(n+1)-1}(1-\alpha_k)^2\,\mathrm{Var}(X_{j+1})\right)^{\frac{p}{2}}\right]$$

$$\sum_{n=1}^{\infty}\prod_{j=b(n)}^{b(n+1)-1}(1-\alpha_j)^p \leq \underbrace{\sum_{n=1}^{n_0}\prod_{j=b(n)}^{b(n+1)-1}(1-\alpha_j)^p}_{\text{First term}} + \underbrace{\sum_{n_0}^{\infty}\frac{1}{(2C)^p(b(n)+1)^{\frac{p}{2}}}}_{\text{Second term}}$$

$$(\sum_{j=b(n)}^{b(n+1)-1}\alpha_j^2\prod_{k=j+1}^{b(n+1)-1}(1-\alpha_k)^2)^{\frac{p}{2}} = \mathcal{O}(\alpha_k^p) + \textbf{Decaying term}$$

Marcinkiewicz-Zygmund

Theorem 5.1 (M-Z). *Assume* $X, X_1, X_2, \ldots$ *are i.i.d. with* $\mathbb{E}X = 0$, *and define*

$$S_n = \sum_{i=1}^{n} a_{i,n} X_i.$$

Suppose for some $1 < q \leq \infty$, *it occurs*

$$\sup_{n \geq 1} \left(\frac{1}{n} \sum_{i=1}^{n} |a_{i,n}|^q \right)^{1/q} < \infty,$$

where for $q = \infty$, *this is interpreted as* $\sup_{n,i} |a_{i,n}| < \infty$. *If* $\mathbb{E}|X|^p < \infty$ *for any* p *such that* $1 \leq 1/p + 1/q < 2$, *then*

$$\frac{S_n}{n} \xrightarrow{\text{a.s.}} 0.$$

Weighted SLLN

Theorem 6.1 (Weighted SLLN). *Let $\{X_k\}_{k\in\mathbb{N}}$ be a sequence of equidistributed random variables. Let also $f:[0,\infty)\to\mathbb{R}$, $g:[0,\infty)\to\mathbb{R}$ be positive functions, and define $\varphi(y)=f(y)g(y)$. Assume φ satisfies:*

1. *φ is strictly increasing and $\varphi([0,\infty))=[0,+\infty)$.*
2. *there exist constants $a,b\in\mathbb{R}$ such that for all $s>0$,*

$$\varphi^2(s)\int_s^\infty \frac{1}{\varphi^2(x)}dx \le as+b.$$

Furthermore, for the truncated sequence $\{Y_k\}_{k\in\mathbb{N}}$ defined by

$$Y_k := X_k\,\mathbf{1}_{\{|X_k|\le\varphi(k)\}} + \varphi(k)\,\mathbf{1}_{\{X_k>\varphi(k)\}} - \varphi(k)\,\mathbf{1}_{\{X_k<-\varphi(k)\}},$$

suppose there is an absolute constant $C>0$ such that for all $1\le m\le n$,

$$\mathbb{E}\left(\max_{m\le\ell\le n}\left|\sum_{k=m}^{\ell}\frac{Y_k-\mathbb{E}Y_k}{\varphi(k)}\right|^2\right)\le C\sum_{k=m}^{n}\frac{\mathrm{Var}(Y_k)}{\varphi^2(k)}.$$

Then the condition $\mathbb{E}\left[\varphi^{-1}(|X_1|)\right]<\infty$ implies that the series

$$\sum_{k=1}^{\infty}\frac{X_k-\mathbb{E}Y_k}{\varphi(k)} \quad \textit{is almost surely convergent.}$$

If, in addition, f is increasing and $\lim_{y\to\infty}f(y)=\infty$, then the following weighted SLLN holds:

$$\frac{1}{f(n)}\sum_{k=1}^{n}\frac{X_k-\mathbb{E}Y_k}{g(k)}\longrightarrow 0 \quad \textit{almost surely as } n\to\infty.$$

Weighted SLLN (cont)

In our context, we want to have

$$\varphi(k) = \frac{1}{\alpha_{k-1}} = k^{\xi}.$$

Recall that we have the recursion

$$S_n = \sum_{k=1}^{n} w_{n,k} X_k,$$

where $w_{n,k} = \alpha_{k-1} \prod_{j=k}^{n-1} (1 - \alpha_j)$. We now consider the factorization

$$w_{n,k}^{-1} = f(n) g(k),$$

where

$$f(n) = \frac{1}{\prod_{j=0}^{n-1} (1 - \alpha_j)}, \quad g(k) = \frac{\prod_{j=0}^{k-1} (1 - \alpha_j)}{\alpha_{k-1}}.$$