Regret bounds for NSA algorithms

Sofiane Saadane
IMT

joint work with Sebastien Gadat and Fabien Panloup

Plan

- What is a bandit algorithm ?
- Stochastic bandit algorithms
- Regret
- NSa and regret
- The PDMP of over-penalized NSa

What is a bandit algorithm ?

An algorithm to determine among many sources which one is the most profitable.

Application

- Clinical trials
- Determine the efficiency of a treatement.
- Advertising, finance,...

Stochastic Bandit Algorithm

- Consider d arms associated to unknown parmaters $p_1, ..., p_d$.
- At step n, you draw an arm $i \in \{1, ..., d\}$ and you receive a reward : $A_n^{(i)}$
- The rewards are i.i.d and $\mathbb{E}(A_n^{(i)}) = p_i$ for all $n \in \mathbb{N}$.

Examples

- MOSS : Bubeck and Audibert 2010
- Draws at each time the arm with the best empirical arm.
- KL-UCB : Garivier and Cappé 2011.
- Builds a confidence interval for the reward.

Regret

How to evaluate the performance of a bandit algorithm ?

Definition of the regret

Denoting by A_k^j the reward associated to arm j at step k and I_k the arm drawn at step k, the regret R_n is the random variable defined as

$$\mathbb{E}R_n := \mathbb{E}\max_{1\leq j\leq d}\sum_{k=1}^{n}\left[A_k^j - A_k^{I_k}\right].$$

where $(A_k^j)_{k,j}$ is a sequence of independent random variables. Difficult to handle...

Pseudo-Regret

Proposition If $\overline{R}_n := \max_{1\leq j\leq d} \mathbb{E}\sum_{k=1}^{n}\left[A_k^j - A_k^{I_k}\right]$. then

$$0 \leq \mathbb{E}R_n - \bar{R}_n \leq \sqrt{\frac{n\log d}{2}}.$$

Furthermore, for every integers n and d and for any (admissible) strategy,

$$\mathbb{E}(R_n) \geq \frac{1}{20}\sqrt{nd} \quad \text{uniformly over}(p_1, ..., p_d).$$

NSa


YES WE SCAN
DEAL WITH IT.
UNITED WE PROGRESS TOWARD A PERFECTLY MONITORED SOCIETY

Narendra Shapiro Bandit (NS)

An historical algorithm introduced by Narendra and Shapiro in 1969. It produces a sequence $X_n = (X_n^1, ..., X_n^d)$ with $\sum_{k=1}^{d} X_n^k = 1$:

$$X_{n+1}^1 = X_n^1 + \begin{cases} \gamma_{n+1}(1 - X_n^1) & \text{if arm 1 is selected and wins} \\ -\gamma_{n+1} X_n^1 & \text{if another arm is selected and wins} \\ 0 & \text{otherwise} \end{cases}$$

with $\gamma_n = (\frac{c}{c+n})^\alpha$ where $\alpha \in (0, 1)$.

NS convergence

- Condition to guarantee the convergence are strong and depend on the probabilities of success of the arms.
- The algorithm is **faillible** and thus we have

$$\mathbb{E}(R_n) \simeq (p_1 - p_2)\mathbb{P}(X_n \to 0) \simeq n$$

NS penalized

NS penalized

2010,D. Lamberton and G.Pagès introduced a penalized version of NS :

$$X_{n+1} = X_n + \begin{cases} \gamma_{n+1}(1 - X_n) & \text{if arm 1 is selected and wins} \\ -\gamma_{n+1} X_n & \text{if arm 2 is selected and wins} \\ -\rho_n \gamma_{n+1} X_n & \text{if arm 1 is selected and loses} \\ \rho_{n+1} \gamma_{n+1}(1 - X_n) & \text{if arm 2 is selected and loses} \end{cases}$$

NS penalized

- Idea : penalize in case of failure.
- Advantage : Infallibility requieres very few hypotesis.
- Weakness : The regret bound we obtain is not uniform in (p_1, p_2).

Over Penalized NS

Over Penalized NS

GPS (2015) present to you the over-penalized NSa :

$$X_{n+1} = X_n + \begin{cases} \gamma_{n+1}(1 - X_n) - \rho_n\gamma_{n+1}X_n & \text{if arm 1 is selected and wins} \\ -\gamma_{n+1}X_n + \rho_{n+1}\gamma_{n+1}(1 - X_n) & \text{if arm 2 is selected and wins} \\ -\rho_n\gamma_{n+1}X_n & \text{if arm 1 is selected and loses} \\ \rho_{n+1}\gamma_{n+1}(1 - X_n) & \text{if arm 2 is selected and loses} \end{cases}$$

- an arm is always penalized to escape from local trap.

Convergence of OP NSa

The algorithm can be written in a recursive way :

$$X_{n+1} = X_n + \gamma_{n+1} h(X_n) + \gamma_{n+1}\rho_{n+1}\kappa(X_n) + \gamma_{n+1}\Delta M_{n+1}$$

- $h(x) = (p_1 - p_2)x(1-x)$
- $\kappa(x) = -(1-p_1)x^2 + (1-p_2)(1-x)^2$
- Equilibria of $\dot{X} = h(X)$: Dirac masses on each arm. Stable one : $(1, 0, ..., 0)$.
- The Kushner-Clark theorem yields the a.s. convergence towards an equilibrium.

Regret for OP NSa

$$\begin{aligned}\bar{R}_n &= \sum_{k=1}^{n} \mathbb{E}(A_k^1) - \mathbb{E}\left(\sum_{k=1}^{n} A_k^{I_k}\right) = p_1 n - \mathbb{E}\left(\sum_{k=1}^{n} X_k p_1 + (1 - X_k) p_2\right) \\ &= \sum_{k=1}^{n} \rho_k \mathbb{E}(\underbrace{\frac{1 - X_k}{\rho_k}}_{Y_k}).\end{aligned}$$

Thus

$$\sup_n \mathbb{E}(Y_n) < +\infty \implies \bar{R}_n \leq C(p_1 - p_2) \sum_{k=1}^{n} \rho_k$$

We have to choose

$$\rho_n = \frac{\rho_1}{\sqrt{n}}, \quad \gamma_n = \frac{\gamma_1}{\sqrt{n}}$$

Comparison of Penalized and Over-Penalized NSa

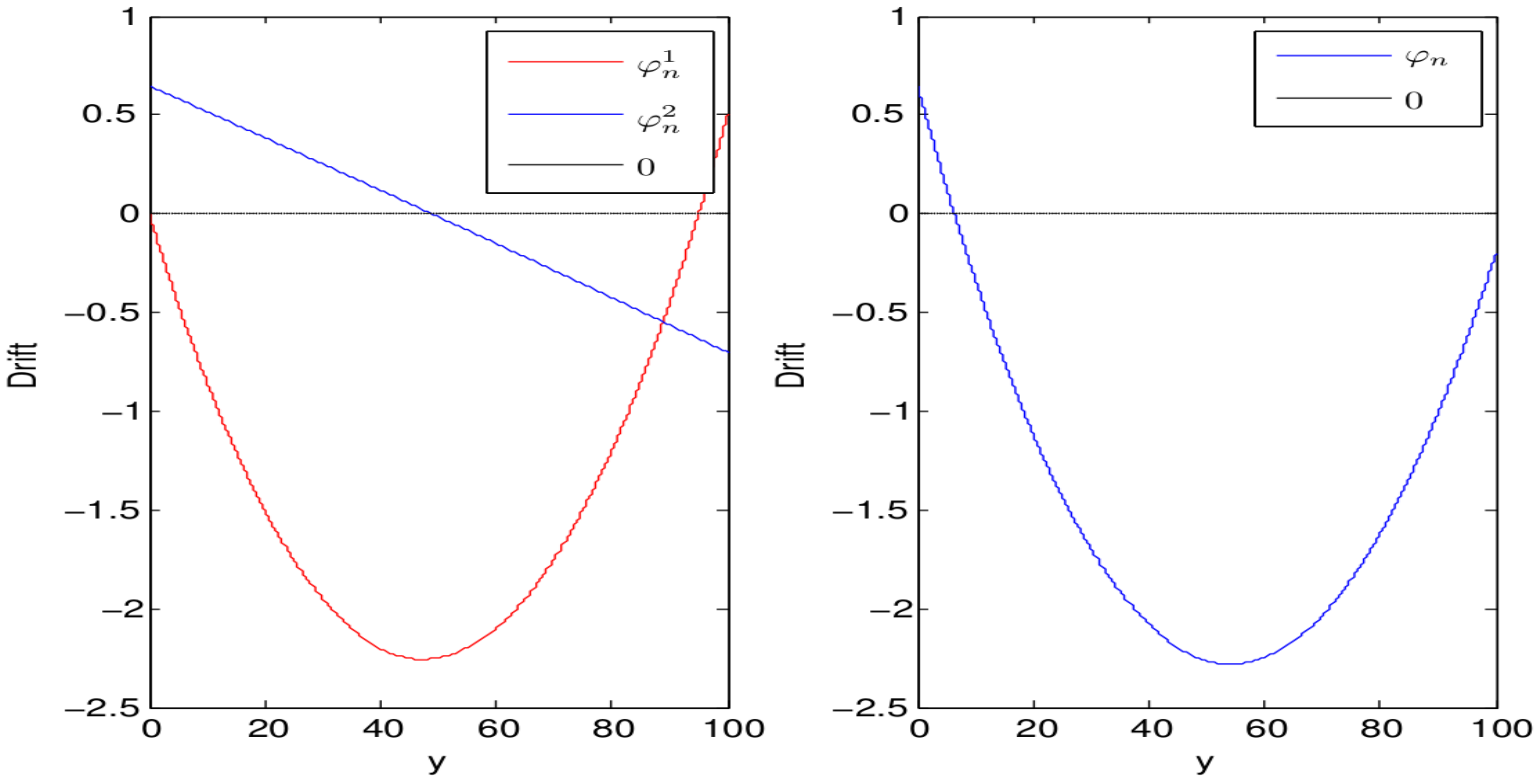


FIGURE : Drift decomposition (left) and global (right) when $y \in [0, \frac{1}{\gamma_n}]$ with $\gamma_1 = \rho_1 = 1$, $p_1 = 0.7$,$p_2 = 0.6$.

Idea of the proof

- Goal : Obtain a uniform bound in (p_1, p_2).
- Arguments : Lyapunov type and awful computations.
- An important quantity : $\pi = p_1 - p_2$.

Theorem

The choice $\gamma_n = 2.63\rho_n = 0.89/\sqrt{n}$ *yields*

$$\forall n, \quad \sup_{(p_1,p_2)\in[0,1],p_2<p_1} \bar{R}_n \leq 31.1\sqrt{2n}.$$

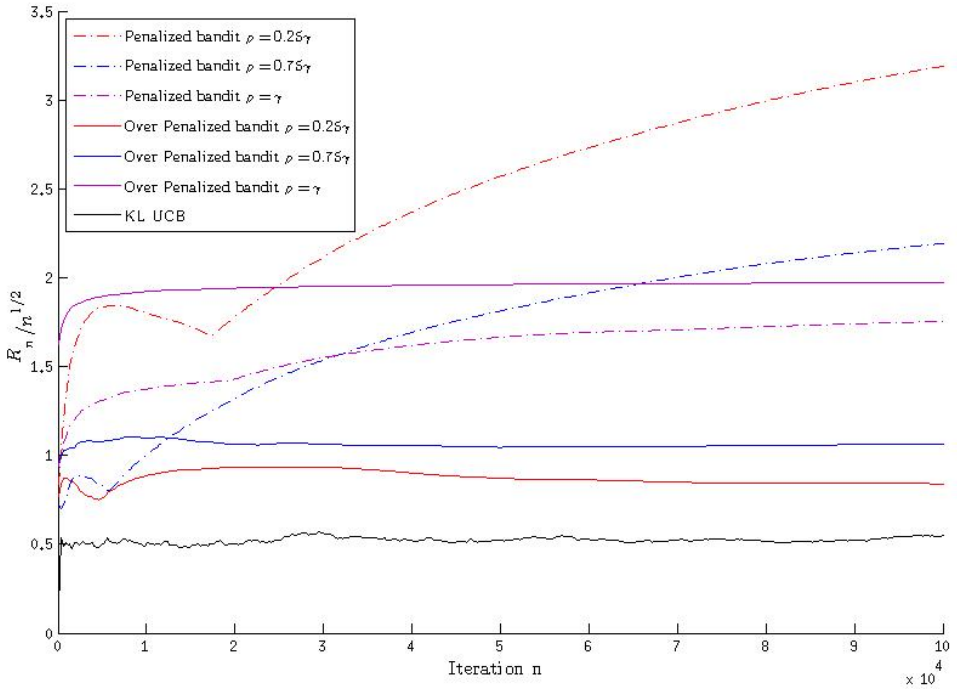
Idea of the proof

- Consider $Z_n^{(r)} = \frac{(1-X_n)^r}{\gamma_n}$ and exhibit a mean reverting property for a proper r (larger than 3) : for all $n \geq 1$,

$$\mathbb{E}(\Delta Z_{n+1}^{(r)} | \mathcal{F}_n) = \gamma_{n+1}(1 - X_n)P_n^r(X_n) + \Delta R_n,$$

  where P_n^r is a polynomial function.
- We study the roots of P_n^r and show that it is negative most of the time.
- A recursion between $Z_n^{(r)}$ and $Z_{n-1}^{(r)}$ completes the proof.

Numerical comparison



- The expected bound is

$$\forall n, \quad \sup_{(p_1,p_2)\in[0,1],p_2<p_1} \bar{R}_n \leq 0.9\sqrt{2n}.$$

And now PDMP

- Consider a multiarmed version of OP NSa :

$$X^1_{n+1} = X^1_n + \begin{cases} \gamma_{n+1}(1 - X^1_n) - \rho_n\gamma_{n+1}X^1_n & \text{if arm 1 is selected and wins} \\ -\gamma_{n+1}X_n + \rho_{n+1}\gamma_{n+1}\frac{(1-X^1_n)}{d-1} & \text{if arm 2 is selected and wins} \\ -\rho_n\gamma_{n+1}X^1_n & \text{if arm 1 is selected and loses} \\ \rho_{n+1}\gamma_{n+1}\frac{(1-X^1_n)}{d-1} & \text{if arm 2 is selected and loses} \end{cases}$$

- The choice of the division of the loose of an arm is arbitrary and other ways of division should be considered.

PDMP

Proposition If $\alpha = \beta \leq 1/2$ and $g = \gamma_1/\rho_1$, then

$$\frac{1}{\rho_n}(X_{n,2}, ..., X_{n,d}) \Rightarrow \mu_{d,g}$$

where μ_d is the (unique) stationary distribution of the Markov process whose generator $\mathcal{L}_d$ acts on compactly supported functions f of $\mathcal{C}^1((\mathbb{R}_+)^{d-1})$ as follows :

$$\begin{aligned}\mathcal{L}_d f(y_2, ..., y_d) &= \sum_{i=2,...,d} \underbrace{\frac{p_i y_i}{g}(f(y_2, ..., y_i + g, ..., y_d) - f(y_2, ..., y_i, ...y_d))}_{\text{Jump part}} \\ &+ \sum_{i=2,...,d} \underbrace{(\frac{1-p_1}{d-1} - p_1 y_i)\partial_i f(y_2, ..., y_d)}_{\text{Deterministic part}}.\end{aligned}$$

About PDMP

Since the trajectory of each coordinate are independent, it is sufficient to consider

$$\mathcal{L}f(x) = (a - bx)f'(x) + cx(f(x + g) - f(x))$$

- Introduced in the 80's by Davis.
- Growing field of interest with various applications (biology).
- Famous example : TCP (Bardet, Guillin, Malrieu and al.)
- Random switching fields : Cloez and Hairer for instance.

Dynamics of the process

- Set $a = \frac{1-p_1}{d-1}$, $b = c = p_1$ and $g = \frac{\rho_1}{\gamma_1}$.
- Exponantial decay between jumps and additive jumps (+g).

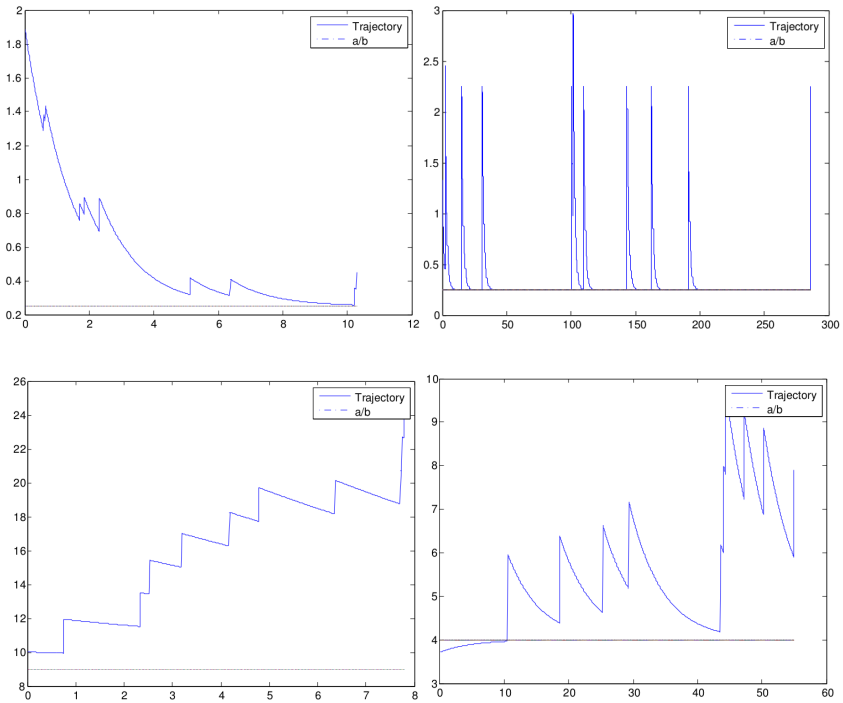


FIGURE : Exact simulation of trajectories of the process.

Ergodicity and rate of convergence

- Few informations are known.
- However speed of of convergence for two measures can be proved.

- Wasserstein distance

$$\mathcal{W}_p(\mu,\nu) = \inf\left\{\mathbb{E}\left((X-Y)^p\right))^{\frac{1}{p}} \mid \mathcal{L}(X)=\mu, \mathcal{L}(Y)=\nu\right\}.$$

- Total variation distance

$$\|\mu P_t - \nu P_t\|_{TV} = \inf \mathbb{P}(X \neq Y)$$

where the infimum is taken over all couple (X, Y) such that $\mathcal{L}(X) = \mu$, $\mathcal{L}(Y) = \nu$.

Results

- Wasserstein Bound

$$\mathcal{W}_p(\mu_t, \mu_\infty) \quad \leq \quad \gamma_p e^{-\frac{t\pi}{p}}.$$

- Total variation bound

$$\|\mu_0 P_t - \mu_\infty P_t\|_{TV} \leq C_\varepsilon e^{-(\alpha\pi-\varepsilon)t} \quad \text{with } \alpha = \frac{1}{2 + \frac{b\pi}{ac}}.$$

Idea of the proof (inspired by Bardet, Malrieu and al.)

- Wasserstein : Use the stochastic monotony of the process.
- Total Variation : Use the Wasserstein coupling and try to stick the trajectories when they are close enough.

Open question

- Are those rate optimal ?
- Total variation distance : not really.
- However, we manage to prove :

$$\left| \int x(\mu_0 - \mu_\infty)(dx) \right| e^{-\pi t} \leq \mathcal{W}_1(\mu_t, \mu_\infty) \leq \mathcal{W}_1(\mu_0, \mu_\infty) e^{-t\pi}$$

- Open question for $\mathcal{W}_p$ with $p \geq 3$.

Futur work

- d-armed regret ?
- More general rewards (continuous maybe).
- Wasserstein lower bound.