Séminaire d'été

Marche aléatoire de l'éléphant

Alice Morinière

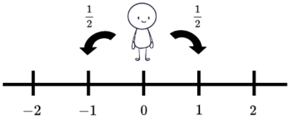
En stage sous la supervision d'Hélène Guérin et Lucile Laulin

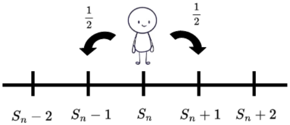
21 Juin 2023

Table des matières

Rappel sur les marches aléatoires

Marche aléatoire symétrique :





$$S_n = \sum_{i=1}^{n} X_i, \quad \text{où} \quad X_i \sim \mathscr{R}\left(\frac{1}{2}\right)$$

Rappel sur les marches aléatoires

Théorème

$$\frac{S_n}{n} \xrightarrow[n\to\infty]{p.s.} 0 \quad et \quad \frac{S_n}{\sqrt{n}} \xrightarrow[n\to\infty]{\mathscr{L}} \mathscr{N}(0,1)$$

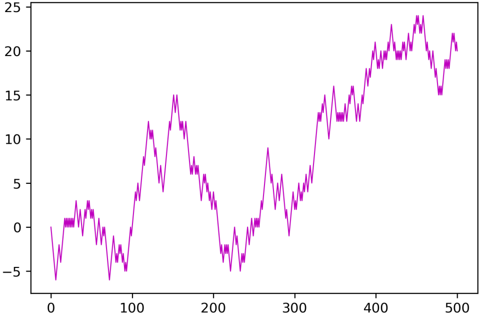
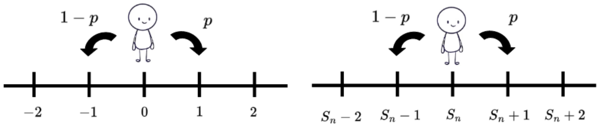


Figure – Marche aléatoire symétrique

Rappel sur les marches aléatoires

Marche aléatoire non symétrique :



$$S_n = \sum_{i=1}^{n} X_i, \quad \text{où} \quad X_i \sim \mathscr{R}(p), \quad \text{avec} \quad p \in [0,1]$$

Rappel sur les marches aléatoires

Théorème

$$\frac{S_n}{n} \xrightarrow[n\to\infty]{p.s.} 2p-1 \quad et \quad \frac{S_n-(2p-1)n}{\sqrt{4p(1-p)n}} \xrightarrow[n\to\infty]{\mathscr{L}} \mathcal{N}(0,1).$$

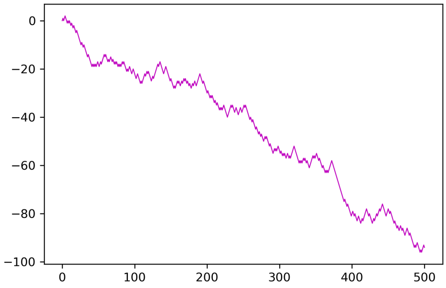


Figure – Marche non symétrique pour p=0.4

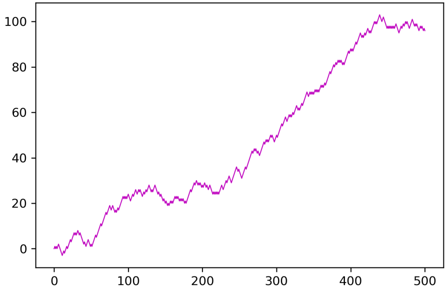


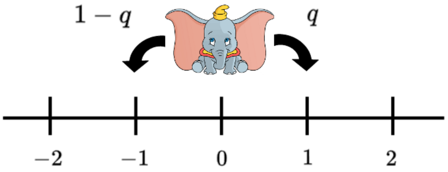
Figure – Marche non symétrique pour p=0.6

Motivations

Domaines d'applications des marches à longue mémoire : physique théorique, informatique, économie, biologie, etc.

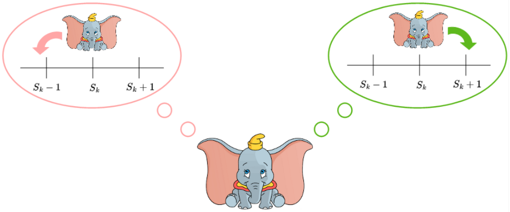
Objectif : Etudier le comportement asymptotique d'un modèle simple de marche aléatoire à longue mémoire

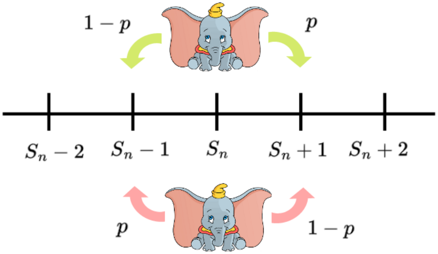
La marche aléatoire de l'éléphant : étape 1



$$S_1 = X_1, \quad \text{où} \quad X_1 \sim \mathscr{R}(q)$$

La marche aléatoire de l'éléphant : étape n





$$k \in [\![1, n-1]\!], \quad X_n = \begin{cases} X_k & \text{avec probabilité } p, \\ -X_k & \text{avec probabilité } 1-p. \end{cases} \quad \text{et} \quad S_n = S_{n-1} + X_n$$

p = paramètre de mémoire de la marche

L'importance du paramètre de mémoire

Trois régimes dépendant de la valeur de p :

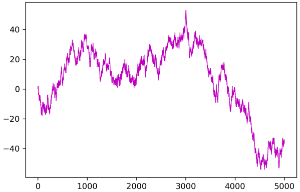


Figure – Régime diffusif pour $0 \leq p < \frac{3}{4}$ (ici $p = \frac{1}{3}$)

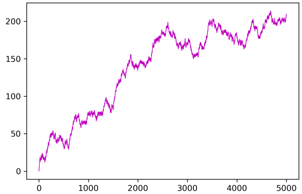


Figure – Régime critique pour $p = \frac{3}{4}$

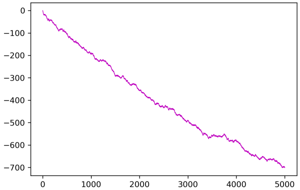


Figure – Régime super diffusif pour $\frac{3}{4} < p \leq 1$ (ici $p = \frac{8}{10}$)

L'approche martingale

Définition

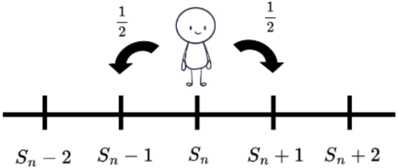
Une suite de variables aléatoires $(M_n)_{n\geq 1}$ est une martingale si

$$\mathbb{E}(M_{n+1}|M_n, \dots, M_1) = M_n.$$

C'est une sous-martingale (respectivement sur-martingale) si

$$\mathbb{E}(M_{n+1}|M_n, \dots, M_1) \geq M_n, \quad (\textit{respectivement} \quad \mathbb{E}(M_{n+1}|M_n, \dots, M_1) \leq M_n).$$

Exemples :



L'approche martingale

Définition

Une suite de variables aléatoires $(M_n)_{n\geq 1}$ est une martingale si

$$\mathbb{E}(M_{n+1}|M_n, ..., M_1) = M_n.$$

C'est une sous-martingale (respectivement sur-martingale) si

$$\mathbb{E}(M_{n+1}|M_n, ..., M_1) \geq M_n, \quad (\textit{respectivement} \quad \mathbb{E}(M_{n+1}|M_n, ..., M_1) \leq M_n).$$

On définit pour $n \geq 0$, $\boldsymbol{M_n := a_n S_n}$, où a_n dépend du paramètre de mémoire p.

Alors $\boldsymbol{(M_n)_{n\geq 0}}$ **est une martingale.**

Résultats obtenus grâce à cette approche

	diffusif $(0 \leq p < \frac{3}{4})$	**critique** $(p = \frac{3}{4})$	**super-diffusif** $(\frac{3}{4} < p \leq 1)$
LGN	$\frac{S_n}{n} \xrightarrow[n\to\infty]{\text{p.s.}} 0$	$\frac{S_n}{\sqrt{n}\log(n)} \xrightarrow[n\to\infty]{\text{p.s.}} 0$	$\frac{S_n}{n^{2p-1}} \xrightarrow[n\to\infty]{\text{p.s.},\mathbb{L}^4} L$
TCL	$\frac{S_n}{\sqrt{n}} \xrightarrow[n\to\infty]{\mathscr{L}} \mathscr{N}\left(0, \frac{1}{3-4p}\right)$	$\frac{S_n}{\sqrt{n\log(n)}} \xrightarrow[n\to\infty]{\mathscr{L}} \mathscr{N}(0,1)$	$\frac{S_n - n^{2p-1}L}{\sqrt{n}} \xrightarrow[n\to\infty]{\mathscr{L}} \mathscr{N}\left(0, \frac{1}{4p-3}\right)$

Illustration des convergences : régime diffusif ($0 \leq p < \frac{3}{4}$)

Théorème ([Bercu(2017)])

$$\frac{S_n}{n} \xrightarrow[n \to \infty]{p.s.} 0.$$

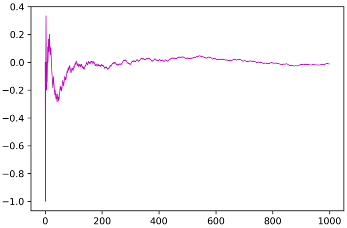


Figure – Evolution de $\frac{S_n}{n}$ en fonction de n, pour $p = \frac{1}{3}$

Illustration des convergences : régime diffusif ($0 \leq p < \frac{3}{4}$)

Théorème ([Bercu(2017)])

$$\frac{S_n}{n} \xrightarrow[n\to\infty]{p.s.} 0.$$

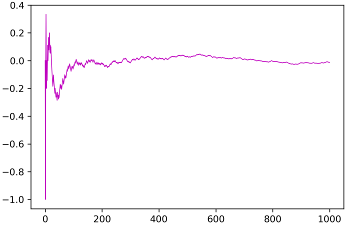


Figure – Evolution de $\frac{S_n}{n}$ en fonction de n, pour $p = \frac{1}{3}$

Théorème ([Bercu(2017)])

$$\frac{S_n}{\sqrt{n}} \xrightarrow[n\to\infty]{\mathscr{L}} \mathscr{N}\left(0, \frac{1}{3-4p}\right).$$

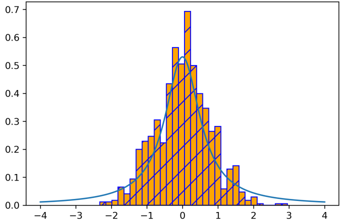


Figure – Comparaison entre l'histogramme d'un échantillon de 1000 valeurs de $\frac{S_n}{\sqrt{n}}$ pour $n = 500$, $p = \frac{1}{3}$ et la représentation de la densité de la loi $\mathscr{N}\left(0, \frac{1}{3-4p}\right)$

Illustration des convergences : régime critique ($p = \frac{3}{4}$)

Théorème ([Bercu(2017)])

$$\frac{S_n}{\sqrt{n}\log(n)} \xrightarrow[n\to\infty]{p.s.} 0.$$

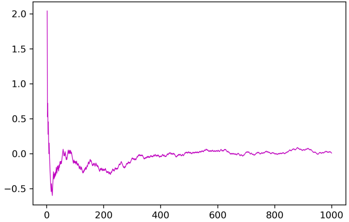


Figure – Evolution de $\frac{S_n}{\sqrt{n}\log(n)}$ en fonction de n, pour $p = \frac{3}{4}$

Illustration des convergences : régime critique ($p = \frac{3}{4}$)

Théorème ([Bercu(2017)])

$$\frac{S_n}{\sqrt{n}\log(n)} \xrightarrow[n\to\infty]{p.s.} 0.$$

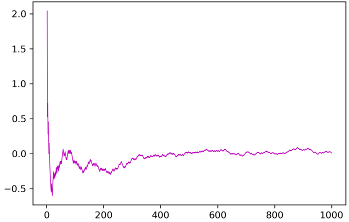


Figure – Evolution de $\frac{S_n}{\sqrt{n}\log(n)}$ en fonction de n, pour $p = \frac{3}{4}$

Théorème ([Bercu(2017)])

$$\frac{S_n}{\sqrt{n\log(n)}} \xrightarrow[n\to\infty]{\mathscr{L}} \mathscr{N}(0,1).$$

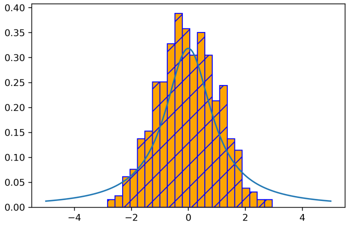
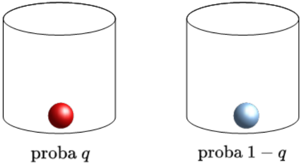


Figure – Comparaison entre l'histogramme d'un échantillon de 1000 valeurs de $\frac{S_n}{\sqrt{n\log(n)}}$ pour $n = 500$, $p = \frac{3}{4}$ et la représentation de la densité de la loi $\mathscr{N}(0,1)$

Modélisation avec des urnes de Polya

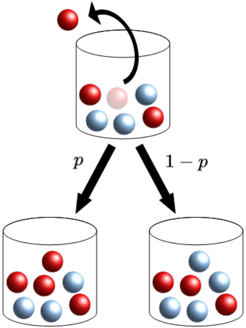
Urne avec des boules rouges et bleues : à l'étape n, $U_n = (R_n, B_n)$ où, R_n est le nombre de boules rouges, et B_n le nombre de boules bleues

Etape 1 :



Modélisation avec des urnes de Polya

Etape n :



Pour une marche alétoire de l'éléphant partant de $S_0 = 0$, et telle que $S_1 = R_1 - B_1$, alors

$$\forall n \in \mathbb{N}, \quad S_n \overset{\mathscr{L}}{=} R_n - B_n.$$

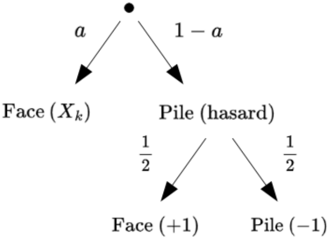
A quoi sert cette approche ?

En partie à obtenir les mêmes résultats qu'avec les martingales, mais pour le reste, nous ne détaillerons pas aujourd'hui...

Pour les intéressés : cette méthode est utilisée entre autres dans [Baur and Bertoin(2016)] et [Laulin(2022)].

Une autre formulation pour notre marche aléatoire

On garde $S_0 = 0$, et $S_1 = X_1$ où $X_1 \sim \mathscr{R}(q)$, puis à l'étape n, on se souvient de nouveau d'une étape k parmi les $n-1$ précédentes, mais cette fois ci Dumbo va choisir son prochain choix à Pile ou Face... On note $a = 2p - 1$.



Les arbres aléatoires récursifs

Définition

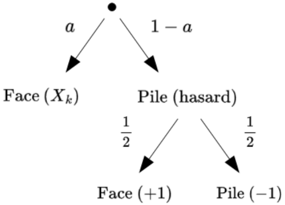
Un arbre aléatoire récursif est un arbre construit de la façon suivante : la racine est numérotée à 1, puis pour chaque nouveau noeud $i \geq 2$, on choisit uniformément au hasard le noeud auquel on le rattache parmi ceux déjà existants, donc uniformément au hasard parmi $[1, i-1]$. On appelle percolation de Bernoulli sur les arêtes le processus suivant :chaque arête de l'arbre est conservée avec une certaine probabilité s, pour $s \in [0, 1]$ et supprimée avec une probabilité 1-s.

Construction d'une forêt de mémoire de la marche

Définition

Un arbre aléatoire récursif est un arbre construit de la façon suivante : la racine est numérotée à 1, puis pour chaque nouveau noeud $i \geq 2$, on choisit uniformément au hasard le noeud auquel on le rattache parmi ceux déjà existants, donc uniformément au hasard parmi $[1, i-1]$. On appelle percolation de Bernoulli sur les arêtes le processus suivant :chaque arête de l'arbre est conservée avec une certaine probabilité s, pour $s \in [0,1]$ et supprimée avec une probabilité 1-s.

Rappel :



Lien entre S_n et la forêt de mémoire

$$S_n = \sum_{i=1}^{n} |c_{i,n}| m_i$$

Des résultats sur la taille des clusters

Théorème ([Baur and Bertoin(2015)])

Pour tout $0 \leq a \leq 1$, *on a*

$$\frac{|c_{1,n}|}{n^a} \xrightarrow[n\to\infty]{p.s.} X_1,$$

où la variable X_1 *a une distribution de Mittag-Leffler de paramètre* a.

Pour $i \geq 2$, *si* $|c_{i,n}| \neq 0$, *on a*

$$\frac{|c_{i,n}|}{n^a} \xrightarrow[n\to\infty]{p.s.} \rho_i,$$

où ρ_i *a la même loi que* $\beta_i^a X_1$, *et* β_i *est une variable de loi Beta de paramètre* $(1, i-1)$ *indépendante de* X_1.

Comment utiliser ces résultats ?

Définition

Pour $k \geq 1$, i_k est l'indice étant à la racine du k-ième cluster, en les numérotant par ordre d'apparition. Donc

$$i_1 = 1 \quad \text{et} \quad \forall k \geq 1, \quad i_{k+1} \overset{\mathscr{L}}{=} i_k + G_k \quad \text{où} \quad G \sim \mathscr{G}(1-a),$$

cela signifie donc que pour $k \geq 2$, $i_k \overset{\mathscr{L}}{=} 1 + B_k$ où B_k suit une loi binomiale négative de paramètres $(1-a, k-1)$.
Et, pour tout $n \geq 1$, on peut écrire

$$S_n = \sum_{k \geq 1} |c_{i_k,n}| m_{i_k}.$$

Un résultat obtenu grâce à cette méthode

Rappel : L'approche martingale nous a donné : Pour $p > \frac{3}{4}$, et donc $a := 2p - 1 > \frac{1}{2}$,

$$\frac{S_n}{n^a} \xrightarrow[n \to \infty]{\text{p.s.}} L.$$

On vient de voir :

$$S_n = \sum_{k \geq 1} |c_{i_k,n}| m_{i_k}.$$

Et les études des arbres nous donnent :

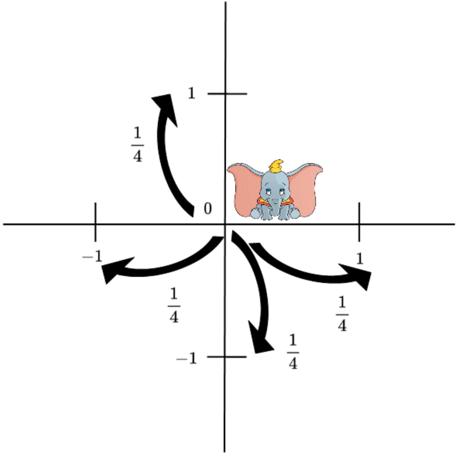
$$\frac{|c_{1,n}|}{n^a} \xrightarrow[n \to \infty]{\text{p.s.}} X_1, \quad \text{et} \quad \frac{|c_{i_k,n}|}{n^a} \xrightarrow[n \to \infty]{\text{p.s.}} \rho_{i_k}.$$

Si on mixe le tout, on obtient :

$$\boxed{L = X_1 m_1 + \sum_{k \geq 2} \rho_{i_k} m_{i_k}.}$$

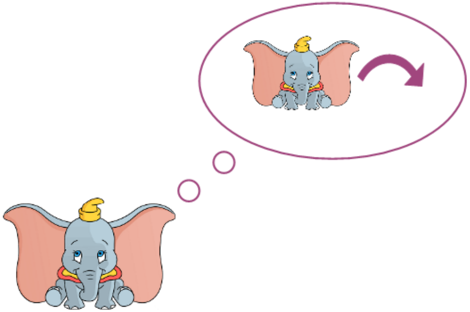
Et en dimension d ?

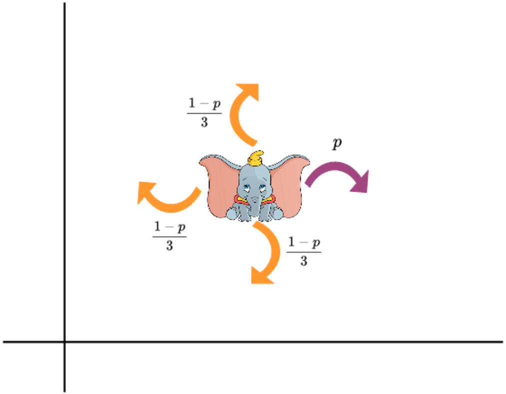
Exemple en dimension 2 : étape 1



Et en dimension d ?

Exemple en dimension 2 : étape n





L'importance du paramètre de mémoire

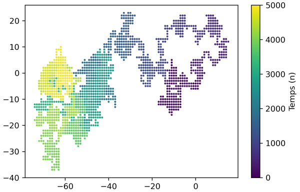


Figure – Régime diffusif pour $0 \leq p < \frac{2d+1}{4d}$ (ici $p = \frac{1}{3}$)

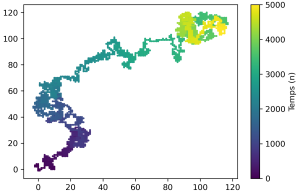


Figure – Régime critique pour $p = \frac{2d+1}{4d}$ (ici $\frac{5}{8}$)

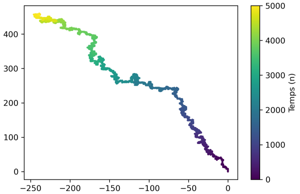
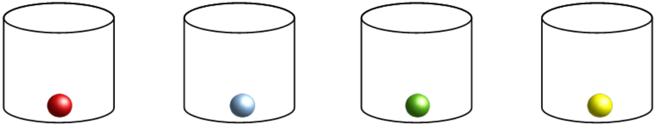


Figure – Régime super diffusif pour $\frac{2d+1}{4d} < p \leq 1$ (ici $p = \frac{8}{10}$)

Les urnes de Polya en dimension d

Urne avec des boules de $2d$ couleurs différentes : à l'étape n, $U_n = (C_n^1, ..., C_n^{2d})$ où C_i est le nombre de boules de la couleur i

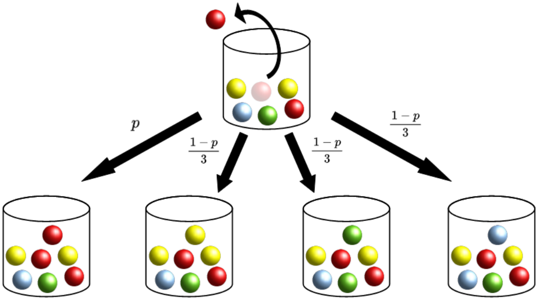
Exemple en dimension 2 : étape 1 :



$\frac{1}{4}$ $\frac{1}{4}$ $\frac{1}{4}$ $\frac{1}{4}$

Les urnes de Polya en dimension d

Exemple en dimension 2 : étape n :

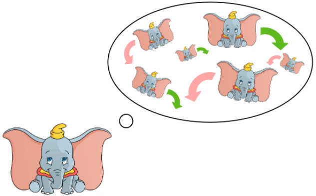


Pour une marche alétoire de l'éléphant partant de $S_0 = 0$, et telle que $S_1 = (C_1^1 - C_1^{d+1})e_1 + ... + (C_1^d - C_1^{2d})e_d$, alors

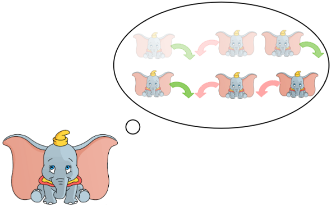
$$\forall n \in \mathbb{N}, \quad S_n \overset{\mathcal{L}}{=} (C_n^1 - C_n^{d+1})e_1 + ... + (C_n^d - C_n^{2d})e_d.$$

Ouvertures : Quelques extensions

- mémoire renforcée (traité par [Laulin(2022)])

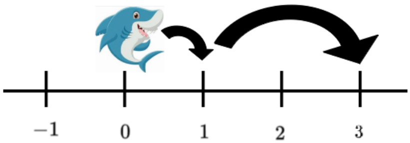


- amnésie (traité par [Laulin(2022)])



Ouvertures : Quelques extensions

- nage aléatoire du requin (traité par [Businger(2018)])



- Autres : changer le paramètre de mémoire au cours du temps, se rappeler d'une séquence plutôt que seulement d'un pas, marche continue, etc

Bibliographie

Erich Baur and Jean Bertoin.
The fragmentation process of an infinite recursive tree and Ornstein-Uhlenbeck type processes.
2015.

Erich Baur and Jean Bertoin.
Elephant random walks and their connection to pólya-type urns.
Physical review E, 94(5) :052134, 2016.

Bernard Bercu.
A martingale approach for the elephant random walk.
Journal of Physics A : Mathematical and Theoretical, 51(1) :015201, nov 2017.
doi : 10.1088/1751-8121/aa95a6.

Silvia Businger.
The Shark Random Swim (Lévy Flight with Memory).
Journal of Statistical Physics, 172 :701–717, 2018.

Lucile Laulin.
Autour de la marche aléatoire de l'éléphant.
PhD thesis, Bordeaux, 2022.

Merci de votre écoute !