

---

# A TEMPLATE FOR THE *arxiv* STYLE

---

A PREPRINT

David S. Hippocampus\*

Department of Computer Science  
Cranberry-Lemon University  
Pittsburgh, PA 15213  
hippo@cs.cranberry-lemon.edu

Elias D. Striatum

Department of Electrical Engineering  
Mount-Sheikh University  
Santa Narimana, Levand  
stariate@ee.mount-sheikh.edu

15 décembre 2020

## Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

**Keywords** First keyword · Second keyword · More

## 1 Introduction : Le problème des bandits stochastiques

Dans le problème des bandits stochastiques classiques, un agent doit choisir à chaque pas de temps  $t$  une action  $k_t \in \{1, 2, \dots, K\}$ . L'agent reçoit alors une récompense  $r_t \sim \nu(\mu_{k_t})$  où  $\nu(\mu_{k_t})$  est une distribution d'espérance  $\mu_{k_t}$ . Le but de l'agent est de maximiser la somme des récompenses sur un horizon de temps  $T$ , plus formellement on souhaite munir l'agent d'une politique d'action maximisant  $\mathbb{E} \left[ \sum_{t=1}^T r_{k_t} \right]$ . Cette espérance sera maximiser si l'action optimale  $k_\star = \underset{k}{\operatorname{argmax}} \mu_k$  de moyenne optimale  $\mu_\star = \max_k \mu_k$  est sélectionnée le plus souvent possible. Le problème de maximisation ainsi défini revient à minimiser le pseudo-regret cumulatif défini par  $\mathcal{R}(T) = T\mu_\star - \mathbb{E} \left[ \sum_{t=1}^T r_t \right]$ . Empiriquement, on peut vérifier la performance en calculant son regret cumulatif empirique

$$R(T) = \sum_{t=1}^T \Delta_{k_t} \quad (1)$$

où  $\Delta_{k_t} = \mu_\star - \mu_{k_t}$  est le regret instantané cumulé par l'agent au temps  $t$ . La mesure de performance définie par (1) est la base de plusieurs expérimentations comparatives effectuées dans la littérature sur les bandits stochastiques. Le lecteur notera que l'architecture de ce problème suggère que les distributions  $\nu$  définissant

---

\*Use footnote for providing further information about author (webpage, alternative address)—*not* for acknowledging funding agencies.

les récompenses ont une espérance. Or, plusieurs distributions de probabilité n'ont pas d'espérance. PARLER D'APPLICATION. Nous souhaitons donc tenter d'étendre le problème des bandits stochastiques sur des bras d'espérance non définie. Nous proposerons ici plusieurs façons de faire en s'inspirant de certaines distributions.

## 2 Extension des bandits stochastiques pour la loi de Cauchy

Une variable aléatoire  $X$  suit une loi de Cauchy ( $X \sim \text{Cauchy}(L, a)$ ) si sa fonction de densité est  $f(x; L, a) = \frac{1}{\pi a \left[1 + \left(\frac{x-L}{a}\right)^2\right]}$  avec  $L \in \mathbb{R}$  et  $a > 0$ . Cette loi de probabilité n'a pas d'espérance, mais est symétrique par rapport  $L$  (la localisation de la loi) et possède donc un centre naturel en  $L$ . Si on se place dans un problème on l'on doit choisir entre deux lois de Cauchy pour maximiser l'observation tirée de ces lois, il semble naturel de choisir la loi de localisation maximale pour augmenter les chances d'avoir une grande observation. En fait, il est facile de démontrer que si  $X_1 \sim \text{Cauchy}(L_1, a_1)$  et  $X_2 \sim \text{Cauchy}(L_2, a_2)$ , alors

$$\mathbb{P}[X_1 > X_2] > 0.5 \Leftrightarrow L_1 > L_2 \quad (2)$$

$$\mathbb{P}[X_1 > t] > \mathbb{P}[X_2 > t] \quad \forall t \in \mathbb{R} \Leftrightarrow L_1 > L_2 \text{ et } a_1 = a_2 \quad (3)$$

Autrement dit, on peut toujours établir une relation de dominance entre deux lois de Cauchy de paramètres de localisations distincts en utilisant la définition (2). Dans le cas où les paramètres d'échelles sont égaux, on peut établir une relation de dominance plus forte en utilisant (3). Conséquemment, si on définit un bandit stochastiques à  $K$  bras de distribution de Cauchy, il semble naturel de définir l'action optimale comme  $k_\star = \operatorname{argmax}_k L_k$  à partir de la localisation optimale  $k_\star = \operatorname{argmax}_k L_k$ . Le regret associé au choix de l'action  $k$  devient donc  $\Delta_k = L_\star - L_k$ . En appliquant cela dans (1), on obtient une mesure de performance empirique permettant d'évaluer la performance d'un agent dans cette configuration. PARLER DES ALGORITHMES CLASSIQUES EN UNE PHRASE OU DEUX EN CITANT DES ARTICLES. Les algorithmes classiques basées sur la moyenne empirique des récompenses échoueront à avoir de bons résultats dans cette configuration étant donné que la moyenne empirique n'estime pas bien le paramètre de localisation de la loi. En fait cet estimateur est divergent dans ce cas. Conséquemment, nous devons d'abord trouver de bons estimateurs de la localisation d'une loi de Cauchy.

### 2.1 Estimateurs de localisation d'une loi de Cauchy

Soit  $\mathcal{X} = \{X_1, X_2, \dots, X_T\}$  une séquence d'observations provenant d'une loi Cauchy( $L, a$ ) où  $L$  est inconnu et  $a$  connu. Soit  $X_{(i)}$  la  $i^{\text{ième}}$  statistique d'ordre de  $\mathcal{X}$ . On définit les estimateurs empiriques suivant pour  $L$ .

- La médiane :  $\text{MED}(\mathcal{X})$
- La moyenne  $\alpha$ -tronquée :  $TM_\alpha(\mathcal{X}) = \frac{1}{T-2r} \sum_{i=r+1}^{T-r} X_{(i)}$ , où  $r = \lfloor T\alpha \rfloor$  et  $0 < \alpha < 0.5$
- L'estimateur de maximum de vraisemblance :  $\text{MLE}(\mathcal{X})$
- L-estimator :  $\text{LE}(\mathcal{X}) = \frac{1}{T} \sum_{i=1}^T J\left(\frac{i}{T+1}\right) X_{(i)}$ , où  $J(u) = \frac{\sin(4\pi(u-0.5))}{\tan(\pi(u-0.5))}$ ,

Remarquons que MLE ne possède pas de forme close, conséquemment il doit être estimé numériquement en programmant un algorithme d'optimisation. DÉTAILLÉ (RÉFÉRENCE ALGO). PARLER DE L'ARTICLE DE RÉFÉRENCE pour L-ESTIMATOR. Aussi, on a noté que LE n'était pas défini lorsque  $T$  est pair en raison du fait que  $u$  pourra alors prendre la valeur 0.5, causant une indétermination de la forme 0/0 de la fonction  $J(u)$ . Or c'est une singularité effaçable et donc  $J(u)$  peut être prolongé analytiquement en  $u = 0.5$  en définissant  $J(0.5) := \lim_{u \rightarrow 0.5} J(u) = 4$ .

PARLER DES RÉSULTATS DE CES ESTIMATEURS

### 2.2 Adaptation d'algorithmes sur des bandits Cauchy

Les algorithmes ETC,  $\epsilon$ -greedy et Boltzmann/Softmax (CITÉS DES ARTICLES) peuvent être adaptés pour jouer sur des bandits Cauchy en remplaçant l'estimation empirique  $\hat{\mu}_k(t-1)$  permettant d'estimer le bras

optimal par un estimateur  $\widehat{L}_k(t-1)$  de la localisation  $L_k$  de la distribution du bras no  $k$ . Chaque estimateurs présentés à la section 2.1 nous donne une adaptation différente de ces algorithmes.

PRÉSENTER DES RÉSULTATS D'ALGORITHMES SUR CES ESTIMATEURS (GRAPHE DE PSEUDO REGRET)

MIXTURE D'EXPERT ?

### 3 Extension des bandits stochastiques pour les lois de Pareto à espérance infinie

Une variable aléatoire  $X$  suit une loi de Pareto ( $X \sim \text{Pareto}(L, a)$ ) si sa fonction de densité est

$$f(x; L, a) = \begin{cases} \frac{aL^a}{x^{a+1}} & \text{si } x \geq L \\ 0 & \text{sinon} \end{cases} \quad (4)$$

où  $L > 0$  et  $a > 0$

VOICI PLUSIEURS LOIS DE PARETO (Montrer deux graphiques, un avec L variables et a fixes, et un autre L fixe avec a variables)

La fonction de répartition est  $F_X(x) = \mathbb{P}[X \leq x] = \begin{cases} 1 - (L/x)^a & \text{si } x \geq L \\ 0 & \text{sinon} \end{cases}$ ,

et son espérance est  $\mu = \begin{cases} \frac{aL}{a-1} & \text{pour } a > 1 \\ \infty & \text{sinon} \end{cases}$

Le cas où  $a \leq 1$  est le cas qui nous intéresse puisque l'espérance n'existe pas. Dans ce dernier cas, nous souhaiterons définir le regret de façon la plus cohérente possible avec le cas où l'espérance existe. Pour la suite, il importe donc d'abord d'observer pour deux variables  $X_1 \sim \text{Pareto}(L_1, a_1)$  et  $X_2 \sim \text{Pareto}(L_2, a_2)$ , alors dans le cas où  $L_1 = L_2 = L$  (localisation fixe), on a

$$\mu_1 > \mu_2 \Leftrightarrow \frac{a_1 L}{a_1 - 1} > \frac{a_2 L}{a_2 - 1} \Leftrightarrow a_1 < a_2 \quad (5)$$

et que dans le cas  $a_1 = a_2 = a$  on a que

$$\mu_1 > \mu_2 \Leftrightarrow \frac{aL_1}{a-1} > \frac{aL_2}{a-1} \Leftrightarrow L_1 > L_2 \quad (6)$$

Ainsi, si deux distributions de Pareto sont comparées sur la base de leur espérance, cette dernière grandit lorsque  $L$  grandit et grandit lorsque  $a$  rapetisse.

Dans le cadre général où l'on peut avoir  $a \leq 1$  (espérance infinie), on ne tentera de comparer deux lois de Pareto à l'aide d'une relation de dominance de première ordre. Soit  $\mathcal{D}_1$  et  $\mathcal{D}_2$  deux distributions de probabilité, on dira que  $\mathcal{D}_1$  domine  $\mathcal{D}_2$  et on écrira  $\mathcal{D}_1 \succeq \mathcal{D}_2$  si  $\mathbb{P}[X_1 > t] \geq \mathbb{P}[X_2 > t] \forall t \in \mathbb{R}$ . Notons que cela est équivalent à  $F_{X_1}(t) < F_{X_2}(t) \forall t \in \mathbb{R}$ . CITER OUVRAGE DÉFINISSANT LA DOMINANCE DE PREMIER ORDRE.

On note que si  $\mathcal{D}_1 = \text{Pareto}(L_1, a)$  et  $\mathcal{D}_2 = \text{Pareto}(L_2, a)$  avec  $a > 0$ , alors  $\mathcal{D}_1 \succeq \mathcal{D}_2 \Leftrightarrow L_1 \geq L_2$ . Dans le cas où  $\mathcal{D}_1 = \text{Pareto}(L, a_1)$  et  $\mathcal{D}_2 = \text{Pareto}(L, a_2)$  avec  $a_1 > 0$  et  $a_2 > 0$ , alors  $\mathcal{D}_1 \succeq \mathcal{D}_2 \Leftrightarrow L_1 \leq L_2$ . Les démonstrations sont simples en utilisant l'expression de la fonctions de répartition de la loi de Pareto. Notons que si  $\mathcal{D}_1 = \text{Pareto}(L_1, a_1)$  et  $\mathcal{D}_2 = \text{Pareto}(L_2, a_1)$  avec  $L_1 \neq L_2$  et  $a_1 \neq a_2$ , il n'est pas possible de définir un ordre de dominance entre  $\mathcal{D}_1$  et  $\mathcal{D}_2$  puisque la fonction de répartition associée à ces distributions s'intersecteront pour un certain  $t \in ]\max(L_1, L_2); \infty[$

On note que les relations de dominance ainsi définies sont compatibles avec la «dominance» établie sur la comparaison des espérances, relations (5) et (6), dans le cas d'espérance définie.

Avec ces observations, nous pouvons définir intuitivement et naturellement des notions de regrets pour certaines configurations de bandits de Pareto.

### 3.1 Bandits de Pareto avec paramètre $a$ constant

On suppose ici un bandit stochastiques à  $K$  bras dont le numéro  $k$  est de distribution Pareto( $L_k, a$ ) Ici,  $a > 0$  est fixe mais quelconque, le cas où  $a \leq 1$  est particulièrement intéressant dans le contexte de notre étude. On définit dans cette situation le bras optimal et la notion de regret en se basant sur les localisations  $L_k$  des bras.

$$L_\star = \max_k L_k, \quad k^\star = \operatorname{argmax}_k L_k \quad \text{et} \quad \Delta_k = L_\star - L_k$$

ALGO DÉFINIT AVEC DES ESTIMATEURS DE L

### 3.2 Bandits de Pareto avec paramètre $a$ constant

On suppose ici un bandit stochastiques à  $K$  bras dont le numéro  $k$  est de distribution Pareto( $L, a_k$ ) où  $L$  est fixe. On remarque que les  $a_k$  étant variables, certains peuvent être inférieure à 1 et d'autres non. Cette configuration peut donc contenir des bras à distribution d'espérance non-définie et d'autres à distribution à espérance définie. On définit dans cette situation le bras optimal et la notion de regret en se basant sur les paramètres  $a_k$  des bras.

$$a_\star = \min_k a_k, \quad k^\star = \operatorname{argmin}_k a_k \quad \text{et} \quad \Delta_k = a_k - a_\star$$

JOUER THOMPSON SAMPLING (conjugué existe pour le paramètre  $k$ ).

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula. See Section 3.2.

### 3.3 Headings : second level

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

$$\xi_{ij}(t) = P(x_t = i, x_{t+1} = j | y, v, w; \theta) = \frac{\alpha_i(t) a_{ij}^{w_t} \beta_j(t+1) b_j^{v_{t+1}}(y_{t+1})}{\sum_{i=1}^N \sum_{j=1}^N \alpha_i(t) a_{ij}^{w_t} \beta_j(t+1) b_j^{v_{t+1}}(y_{t+1})} \quad (7)$$

#### 3.3.1 Headings : third level

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

**Paragraph** Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

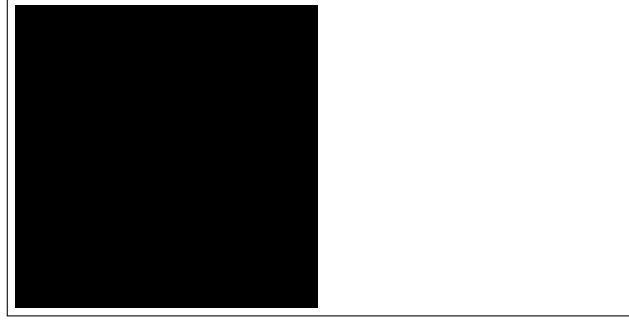


FIGURE 1 – Sample figure caption.

TABLE 1 – Sample table title

Part		
Name	Description	Size ( $\mu\text{m}$ )
Dendrite	Input terminal	$\sim 100$
Axon	Output terminal	$\sim 10$
Soma	Cell body	up to $10^6$

## 4 Examples of citations, figures, tables, references

### 4.1 Citations

Citations use `natbib`. The documentation may be found at

<http://mirrors.ctan.org/macros/latex/contrib/natbib/natnotes.pdf>

Here is an example usage of the two main commands (`citet` and `citep`) : Some people thought a thing [Kour and Saabne, 2014a, Hadash et al., 2018] but other people thought something else [Kour and Saabne, 2014b]. Many people have speculated that if we knew exactly why Kour and Saabne [2014b] thought this...

### 4.2 Figures

Suspendisse vitae elit. Aliquam arcu neque, ornare in, ullamcorper quis, commodo eu, libero. Fusce sagittis erat at erat tristique mollis. Maecenas sapien libero, molestie et, lobortis in, sodales eget, dui. Morbi ultrices rutrum lorem. Nam elementum ullamcorper leo. Morbi dui. Aliquam sagittis. Nunc placerat. Pellentesque tristique sodales est. Maecenas imperdiet lacinia velit. Cras non urna. Morbi eros pede, suscipit ac, varius vel, egestas non, eros. Praesent malesuada, diam id pretium elementum, eros sem dictum tortor, vel consectetur odio sem sed wisi. See Figure 1. Here is how you add footnotes.<sup>2</sup> Sed feugiat. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Ut pellentesque augue sed urna. Vestibulum diam eros, fringilla et, consectetur eu, nonummy id, sapien. Nullam at lectus. In sagittis ultrices mauris. Curabitur malesuada erat sit amet massa. Fusce blandit. Aliquam erat volutpat. Aliquam euismod. Aenean vel lectus. Nunc imperdiet justo nec dolor.

### 4.3 Tables

See awesome Table 1.

The documentation for `booktabs` ('Publication quality tables in LaTeX') is available from :

<https://www.ctan.org/pkg/booktabs>

### 4.4 Lists

— Lorem ipsum dolor sit amet

---

2. Sample of the first footnote.

- consectetur adipiscing elit.
- Aliquam dignissim blandit est, in dictum tortor gravida eget. In ac rutrum magna.

## Références

- George Kour and Raid Saabne. Real-time segmentation of on-line handwritten arabic script. In *Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on*, pages 417–422. IEEE, 2014a.
- Guy Hadash, Einat Kermany, Boaz Carmeli, Ofer Lavi, George Kour, and Alon Jacovi. Estimate and replace : A novel approach to integrating deep neural networks with existing applications. *arXiv preprint arXiv :1804.09028*, 2018.
- George Kour and Raid Saabne. Fast classification of handwritten on-line arabic characters. In *Soft Computing and Pattern Recognition (SoCPaR), 2014 6th International Conference of*, pages 312–318. IEEE, 2014b. doi:10.1109/SOCPAR.2014.7008025.