

# 适应性的重尾分布多臂老虎机

## Adaptive Heavy-Tailed Multi-Armed Bandits

黄佳泰 \* 戴言\* 黄隆波

清华大学 交叉信息研究院

2022年5月22日



- ① 研究对象
- ② 研究背景
- ③ 研究成果
- ④ 研究意义

## 多臂老虎机问题

- 学习者 (agent) 与环境 (environment) 将进行  $T$  轮交互。\*



- 第  $t$  轮, 学习者选择决策 (action)  $i_t \in \{1, 2, \dots, K\}$ , 环境同时决定损失 (loss)  $l_t \in \mathbb{R}^K$ 。学习者将看到并承受  $l_t(i_t)$ 。

---

\* 图片来自 Reinforcement Learning – Multi-Arm Bandit Implementation by Jeremy Zhang.

# 多臂老虎机问题

- 学习者 (agent) 与环境 (environment) 将进行  $T$  轮交互。\*



- 第  $t$  轮, 学习者选择决策 (action)  $i_t \in \{1, 2, \dots, K\}$ , 环境同时决定损失 (loss)  $l_t \in \mathbb{R}^K$ 。学习者将看到并承受  $l_t(i_t)$ 。
- 学习者的目标是最小化它的遗憾值 (regret):

$$\mathcal{R}_T \triangleq \mathbb{E} \left[ \sum_{t=1}^T l_t(i_t) \right] - \min_{i^* \in [K]} \mathbb{E} \left[ \sum_{t=1}^T l_t(i^*) \right],$$

其中, 期望对学习者和环境中的随机性求得。

\* 图片来自 Reinforcement Learning – Multi-Arm Bandit Implementation by Jeremy Zhang.

## 环境的分类

- 通常假设环境是随机 (stochastic) 的, 即存在  $K$  个概率分布  $\nu_1, \nu_2, \dots, \nu_K$ , 使所有  $l_t(k)$  都是  $\nu_k$  中的独立随机样本。

## 环境的分类

- 通常假设环境是随机 (stochastic) 的, 即存在  $K$  个概率分布  $\nu_1, \nu_2, \dots, \nu_K$ , 使所有  $l_t(k)$  都是  $\nu_k$  中的独立随机样本。
- 通常还假设二阶矩有界, 即  $\mathbb{E}_{x \sim \nu_k}[x^2] \leq 1, \forall 1 \leq k \leq K$ , 例如 [Seldin and Slivkins, 2014]、[Zimmert and Seldin, 2019]。

## 环境的分类

- 通常假设环境是随机 (stochastic) 的, 即存在  $K$  个概率分布  $\nu_1, \nu_2, \dots, \nu_K$ , 使所有  $l_t(k)$  都是  $\nu_k$  中的独立随机样本。
- 通常还假设二阶矩有界, 即  $\mathbb{E}_{x \sim \nu_k}[x^2] \leq 1, \forall 1 \leq k \leq K$ , 例如 [Seldin and Slivkins, 2014]、[Zimmert and Seldin, 2019]。
- 本文则允许重尾分布 (heavy-tailed distribution), 即只存在一个重尾参数  $(\alpha, \sigma)$  ( $1 < \alpha \leq 2$ ) 使得  $\mathbb{E}_{x \sim \nu_k}[x^\alpha] \leq \sigma^\alpha$ 。这就是 [Bubeck et al., 2013] 中提出的重尾分布多臂老虎机。

## 环境的分类

- 通常假设环境是随机 (stochastic) 的, 即存在  $K$  个概率分布  $\nu_1, \nu_2, \dots, \nu_K$ , 使所有  $l_t(k)$  都是  $\nu_k$  中的独立随机样本。
- 通常还假设二阶矩有界, 即  $\mathbb{E}_{x \sim \nu_k}[x^2] \leq 1, \forall 1 \leq k \leq K$ , 例如 [Seldin and Slivkins, 2014]、[Zimmert and Seldin, 2019]。
- 本文则允许重尾分布 (heavy-tailed distribution), 即只存在一个重尾参数  $(\alpha, \sigma)$  ( $1 < \alpha \leq 2$ ) 使得  $\mathbb{E}_{x \sim \nu_k}[x^\alpha] \leq \sigma^\alpha$ 。这就是 [Bubeck et al., 2013] 中提出的重尾分布多臂老虎机。
  - 我们还允许  $\alpha, \sigma$  是未知的, 这就要求算法有高适应性。



## 环境的分类

- 通常假设环境是随机 (stochastic) 的, 即存在  $K$  个概率分布  $\nu_1, \nu_2, \dots, \nu_K$ , 使所有  $l_t(k)$  都是  $\nu_k$  中的独立随机样本。
- 通常还假设二阶矩有界, 即  $\mathbb{E}_{x \sim \nu_k}[x^2] \leq 1, \forall 1 \leq k \leq K$ , 例如 [Seldin and Slivkins, 2014]、[Zimmert and Seldin, 2019]。
- 本文则允许重尾分布 (heavy-tailed distribution), 即只存在一个重尾参数  $(\alpha, \sigma)$  ( $1 < \alpha \leq 2$ ) 使得  $\mathbb{E}_{x \sim \nu_k}[x^\alpha] \leq \sigma^\alpha$ 。这就是 [Bubeck et al., 2013] 中提出的重尾分布多臂老虎机。
  - 我们还允许  $\alpha, \sigma$  是未知的, 这就要求算法有高适应性。
- 本文还提出了对抗 (adversarial) 环境: 可能存在  $T$  个不同的分布  $\{\nu_i^t\}_{t,i}$  使  $l_t(i) \sim \nu_i^t$ 。这些分布间不必有任何规律, 只要求它们各自满足重尾分布的条件即可。
  - 这对算法适应性进一步提出了更高的要求。

- ① 研究对象
- ② 研究背景
- ③ 研究成果
- ④ 研究意义

## 相关工作

- [Bubeck et al., 2013] 引入了重尾分布多臂老虎机模型，但他们只允许环境是随机的，还要求  $\alpha$  与  $\sigma$  均已知。此外，他们的结果与下界还有  $\log T$  的因子差，其中  $T$  为游戏长度。

## 相关工作

- [Bubeck et al., 2013] 引入了重尾分布多臂老虎机模型，但他们只允许环境是随机的，还要求  $\alpha$  与  $\sigma$  均已知。此外，他们的结果与下界还有  $\log T$  的因子差，其中  $T$  为游戏长度。
- [Lee et al., 2020] 考虑了学习者不知道  $\sigma$  的情况，但仍要求  $\alpha$  已知，并且仍只考虑了随机环境。此外，他们的算法与理论下界也还有  $\log K$  的因子差，其中  $K$  为手臂个数。

## 相关工作

- [Bubeck et al., 2013] 引入了重尾分布多臂老虎机模型，但他们只允许环境是随机的，还要求  $\alpha$  与  $\sigma$  均已知。此外，他们的结果与下界还有  $\log T$  的因子差，其中  $T$  为游戏长度。
- [Lee et al., 2020] 考虑了学习者不知道  $\sigma$  的情况，但仍要求  $\alpha$  已知，并且仍只考虑了随机环境。此外，他们的算法与理论下界也还有  $\log K$  的因子差，其中  $K$  为手臂个数。
- [Zimmert and Seldin, 2019] 在非重尾分布的环境（即  $\alpha = 2$ ,  $\sigma = 1$ ）中，可以同时随机与对抗环境下都达到理论下界。然而，他们无法处理重尾分布，更不用说未知  $\alpha, \sigma$  的情况。

- ① 研究对象
- ② 研究背景
- ③ 研究成果
- ④ 研究意义

## 研究成果

一共提出了如下三个算法：

- 算法一（**对抗环境**）：已知  $\alpha, \sigma$  的情况下，可以同时在随机和对抗环境下都达到理论下界。这顺便证明了从统计学意义上说，即使是重尾分布，对抗环境也不比随机环境难。

## 研究成果

一共提出了如下三个算法：

- 算法一（**对抗环境**）：已知  $\alpha, \sigma$  的情况下，可以同时在随机和对抗环境下都达到理论下界。这顺便证明了从统计学意义上说，即使是重尾分布，对抗环境也不比随机环境难。
- 算法二（**重尾参数**）：可以在  $\alpha, \sigma$  都未知的情况下工作，在随机和对抗环境下都达到接近下界的遗憾值保证。



## 研究成果

一共提出了如下三个算法：

- 算法一（**对抗环境**）：已知  $\alpha, \sigma$  的情况下，可以同时在随机和对抗环境下都达到理论下界。这顺便证明了从统计学意义上说，即使是重尾分布，对抗环境也不比随机环境难。
- 算法二（**重尾参数**）：可以在  $\alpha, \sigma$  都未知的情况下工作，在随机和对抗环境下都达到接近下界的遗憾值保证。
- 算法三（**最坏情况**）：同样在  $\alpha, \sigma$  都未知的情况下工作。即使在对抗环境下，它的最坏情况理论保证也完全达到下界。这标志着重尾参数未知在统计学意义上也不是额外难题。

① 研究对象

② 研究背景

③ 研究成果

④ 研究意义

# 研究意义

- [Zimmert and Seldin, 2019] 指出，在实际应用中，所有损失向量不容易是独立同分布的。
  - [Putta and Agrawal, 2022] 指出，事先知道  $\sigma$  是难以实现的。
- [Zhang et al., 2020] 指出，在许多真实数据集中，损失服从一个重尾分布，而方差（二阶矩）不一定存在。
  - 由于这是后验求得的，事先知道  $\alpha$  也是非常困难的。

## 研究意义

- [Zimmert and Seldin, 2019] 指出，在实际应用中，所有损失向量不容易是独立同分布的。
  - [Putta and Agrawal, 2022] 指出，事先知道  $\sigma$  是难以实现的。
- [Zhang et al., 2020] 指出，在许多真实数据集中，损失服从一个重尾分布，而方差（二阶矩）不一定存在。
  - 由于这是后验求得的，事先知道  $\alpha$  也是非常困难的。

我们的算法具相比前作有较强适应性优势，故应用前景显著。

# 谢谢大家!

欢迎提问~

- [Bubeck et al., 2013] Bubeck, S., Cesa-Bianchi, N., and Lugosi, G. (2013).  
Bandits with heavy tail.  
*IEEE Transactions on Information Theory*, 59(11):7711–7717.
- [Lee et al., 2020] Lee, K., Yang, H., Lim, S., and Oh, S. (2020).  
Optimal algorithms for stochastic multi-armed bandits with heavy tailed rewards.  
*Advances in Neural Information Processing Systems*, 33:8452–8462.
- [Putta and Agrawal, 2022] Putta, S. R. and Agrawal, S. (2022).  
Scale-free adversarial multi armed bandits.  
In *International Conference on Algorithmic Learning Theory*, pages 910–930. PMLR.

- [Seldin and Slivkins, 2014] Seldin, Y. and Slivkins, A. (2014).  
One practical algorithm for both stochastic and adversarial bandits.  
In *International Conference on Machine Learning*, pages 1287–1295.  
PMLR.
- [Zhang et al., 2020] Zhang, J., Karimireddy, S. P., Veit, A., Kim, S.,  
Reddi, S., Kumar, S., and Sra, S. (2020).  
Why are adaptive methods good for attention models?  
*Advances in Neural Information Processing Systems*, 33:15383–15393.
- [Zimmert and Seldin, 2019] Zimmert, J. and Seldin, Y. (2019).  
An optimal algorithm for stochastic and adversarial bandits.  
In *The 22nd International Conference on Artificial Intelligence and  
Statistics*, pages 467–475. PMLR.