Open 问题：固定预算下的最优最佳臂识别

Chao Qin CQ2199@COLUMBIA.EDU
哥伦比亚大学

编辑：Po-Ling Loh 和 Maxim Raginsky

arXiv:2303.00950v1 [cs.LG] 2 Mar 2023

摘要

最佳臂识别或纯探索问题自Bubeck 等人（2009）和Audibert 等人（2010）以来，一直受到学习理论会议社区的广泛关注。对于任何具有唯一最佳臂的老虎机实例，其在所谓的固定置信度设定下的渐近复杂度已由Garivier 和 Kaufmann（2016）和Chernoff （1959）完全刻画；然而，对于其被称为固定预算设定的“对偶”设定，人们对其渐近复杂度知之甚少。本文讨论固定预算设定下与实例相关的渐近复杂度所涉及的开放问题和猜想。

关键词：多臂老虎机、最佳臂识别、纯探索、渐近复杂度

1. 引言与问题表述

我们考虑所谓的最佳臂识别（BAI）或纯探索问题，其中臂的数量是有限的。实验者可以按顺序选择臂进行测量，并观测其质量的独立噪声观测。实验者的目标是通过以自适应且智能的方式分配测量资源，有把握地识别出最佳臂。BAI 问题数十年来也以不同名称被研究，例如在统计学和运筹学文献中称为排序与选择或序数优化。机器学习文献主要在两种设定下研究 BAI 问题。一种称为固定置信度设定，其目标是在保证停止时间后的错误决策概率低于预先指定水平的同时，使收集样本的期望数量最小化；另一种称为固定预算设定，其目标是在给定样本预算用尽后，使错误决策概率最小化。对于任何具有唯一最佳臂的老虎机实例，其在固定置信度设定下的渐近复杂度已经得到完整刻画。例如，参见Garivier 和 Kaufmann（2016）和Chernoff （1959）。尽管这两种设定看似彼此“对偶”，但固定预算设定下与实例相关的渐近复杂度长期以来一直不明确。本文简要介绍固定置信度设定下已有的结果，并讨论固定预算设定下与实例相关的渐近复杂度所涉及的开放问题和猜想。

我们用粗体字母表示向量。一个老虎机实例 $\boldsymbol{\mu}$ 由 k 个未知分布或臂 $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)$ 组成，其对应的期望为 $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$。为便于说明，我们假设老虎机实例 $\boldsymbol{\mu}$ 有唯一的最佳臂。记该臂为 $I^*(\boldsymbol{\mu}) \triangleq \arg\max_{i \in [k]} \theta_i$，其中$[k]$，$\{1, \ldots, k\}$。实验者并不知道老虎机实例 $\boldsymbol{\mu}$，但希望在实验结束时以高置信度识别最佳臂 $I^*(\boldsymbol{\mu})$。在每个时间点 $t = 1$、2、.....，她可以根据迄今收集到的信息选择一个臂 $I_t \in [k]$ 进行测量，随后观测到一个独立噪声观测值 $Y_{t,\, I_t}$，该观测值从分布 μ_{I_t}中抽取。

© 2022年 C. Qin。

2. 固定置信度设定及其已知结果

在固定置信度设定下，实验者可以随时停止收集样本，并在停止后返回对最佳臂身份的估计。因此，实验者的算法由三条规则构成：决定每个时间点采样哪个臂的采样规则、决定每个时间点是否停止的停止规则，以及在停止时间 τ返回基于前 τ 个观测值对最佳臂身份所作估计 I_τ 的决策规则。

令 $\mathcal{S}$ 为具有唯一最佳臂的老虎机实例类。 Garivier 和 Kaufmann (2016) 研究了能够保证错误决策概率（在停止时间处）一致地低于预先指定水平 $\delta > 0$的算法，即

$$\forall \boldsymbol{\mu} \in \mathcal{S}, \quad \mathbb{P}_{\boldsymbol{\mu}} \left(\hat{I}_{\tau_\delta} \neq I^*(\boldsymbol{\mu}) \right) \leq \delta \tag{1}$$

其中 τ_δ 是一个 几乎必然有限的 停止时间。符号 $\mathbb{P}_{\boldsymbol{\mu}}(\cdot)$ 表示我们评估事件发生的概率时，所选臂的观测值是根据老虎机实例 $\boldsymbol{\mu}$抽取的。在学习理论文献中，这类算法称为 δ-概率近似正确 或 δ-*PAC*算法。在这类算法中，我们希望最小化收集样本的期望数量，记为 $\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]$。Garivier 和 Kaufmann (2016) 证明，对于任意 δ-PAC 算法，

$$\forall \boldsymbol{\mu} \in \mathcal{S}, \quad \liminf_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]}{\log(1/\delta)} \geq \Gamma^*_{\mathrm{fc}}(\boldsymbol{\mu}) \tag{2}$$

其中

$$\Gamma^*_{\mathrm{fc}}(\boldsymbol{\mu}) = \left(\sup_{\boldsymbol{w} \in \Sigma_k} \inf_{\boldsymbol{\nu} \in \mathrm{Alt}(\boldsymbol{\mu})} \sum_{i=1}^{k} w_i \mathrm{KL}(\mu_i \| \nu_i) \right)^{-1}. \tag{3}$$

这里， Σ_k 是维度为 $k-1$的概率单纯形；$\mathrm{Alt}(\boldsymbol{\mu})$ $\{\boldsymbol{\nu} \in \mathcal{S} : I^*(\boldsymbol{\nu}) \neq I^*(\boldsymbol{\mu})\}$ 是其唯一最佳臂不同于 $\boldsymbol{\mu}$的唯一最佳臂的老虎机实例集合；$\mathrm{KL}(p\|q)$表示分布 p 与 q之间的 Kullback-Leibler（KL）散度。 Γ^*_{fc}中的下标 fc 是“固定置信度”（fixed-confidence）的缩写。除了方程（2）中的信息论下界外，Garivier 和 Kaufmann（2016）还提出了所谓的 Track-and-Stop 算法，该算法为δ-PAC，并且能够保证

$$\forall \boldsymbol{\mu} \in \mathcal{S}, \quad \limsup_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]}{\log(1/\delta)} \leq \Gamma^*_{\mathrm{fc}}(\boldsymbol{\mu}). \tag{4}$$

由于公式（2）和（4）中的下界与上界相同，函数 $\Gamma^*_{\mathrm{fc}} : \mathcal{S} \to \mathbb{R}$刻画了固定置信度设定下的渐近复杂度。

3. 固定预算设定及其未解决的问题

在固定预算设定中，样本预算 n 是预先确定的。收集 n 个样本后，实验者需要估计最佳臂的身份，记作 I_n。因此，算法仅由采样规则和决策规则组成。在固定预算设定中，实验者的目标是最小化错误决策概率，其定义为

$$p_{\boldsymbol{\mu},n} \triangleq \mathbb{P}_{\boldsymbol{\mu}} \left(\hat{I}_n \neq I^*(\boldsymbol{\mu}) \right).$$

从某种意义上说，这一设定与固定置信度设定似乎是“对偶”的：固定置信度设定是在错误决策概率始终较小的约束下最小化样本数，而这里则是在样本预算固定的约束下最小化错误决策概率。然而，对于固定预算设定下的渐近复杂度，目前所知甚少。

开放问题 1。首要且最重要的开放问题是，是否存在一个理想的算法类别 $\mathcal{A}$ 和一个定义良好的函数 $\Gamma^*_{\mathrm{fb}}:\ \mathcal{S}\to\mathbb{R}$，使得对于 $\mathcal{A}$ 中的任意算法，

$$\forall \boldsymbol{\mu}\in\mathcal{S},\quad \liminf_{n\to\infty}\frac{n}{\log(1/p_{\boldsymbol{\mu},n})}\geq \Gamma^*_{\mathrm{fb}}(\boldsymbol{\mu})$$

并且存在一个属于 $\mathcal{A}$ 的算法，能够保证

$$\forall \boldsymbol{\mu}\in\mathcal{S},\quad \limsup_{n\to\infty}\frac{n}{\log(1/p_{\boldsymbol{\mu},n})}\leq \Gamma^*_{\mathrm{fb}}(\boldsymbol{\mu}).$$

这里，Γ^*_{fb} 中的下标 fb 是“固定预算”（fixed-budget）的缩写。

关于潜在算法类别的讨论。Kaufmann 等人（2016）研究了所谓的一致算法：对于任意 $\boldsymbol{\mu}\in\mathcal{S}$，错误决策的概率 $p_{\boldsymbol{\mu},\ n}$ 会在 n 趋于无穷时趋于零。一致算法的类别相对较大，我们认为它可能并不是用于刻画固定预算设定下渐近复杂度的合适算法类别。请注意，在固定置信度设定中，公式（1）中定义的 δ-PAC 算法类别具有较强限制性，因为它要求对于任意老虎机实例 $\boldsymbol{\mu}\in\mathcal{S}$，在停止时间（stopping time）处错误决策的概率始终保持较小。这一限制有助于分析固定置信度设定下的渐近复杂度。我们认为，有必要在固定预算设定中提出一个自然但限制性更强的算法类别。例如，除了要求错误决策的概率收敛至零之外，我们可能还需要控制该类别中算法的收敛速率。一种潜在的算法类别包含所有表现不劣于均匀采样的算法，即对于该类别中的任意算法，对于任意老虎机实例 $\boldsymbol{\mu}\in\mathcal{S}$，其 $\limsup_{n\to\infty}\frac{n}{\log(1/p_{\boldsymbol{\mu},\ n})}$ 的值都不高于均匀采样。这引出了以下开放问题。

开放问题2。这个开放问题是：在固定预算设定下，是否存在一种不同于均匀采样本身的算法，其表现始终不逊于均匀采样。

事实上，在固定置信度设定下，可以证明，对于任意老虎机实例 $\boldsymbol{\mu}$, $\Gamma^*_{\mathrm{fc}}(\boldsymbol{\mu})$，其值小于或等于均匀采样下 $\limsup_{n\to\infty}\frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]}{\log(1/\delta)}$ 的值。这意味着，这些固定置信度设定下的渐近最优算法，其表现始终不逊于均匀采样。我们倾向于认为，这些算法在固定预算设定下也优于均匀采样，但这个开放问题的答案尚不明确。

4. 猜想

本节陈述文献中已有的两个猜想。遗憾的是，一般而言，这两个猜想都不成立。

猜想1。由于固定预算设定与固定置信度设定彼此“对偶”，一个猜想是 $\Gamma^*_{\mathrm{fb}}=\Gamma^*_{\mathrm{fc}}$。

猜想 2。另一个猜想是 $\Gamma^*_{\mathrm{fb}} = \Gamma^*_{\mathrm{na}}$ 其中 Γ^*_{na} 后文定义的量，是 非自适应 版本的固定预算设定中的渐近复杂度，该设定由 Glynn 和 Juneja (2004) 研究（下标 na在 Γ^*_{na} 中是“非自适应”的缩写）。他们考虑的采样规则会固定概率向量 $\boldsymbol{w}$，用于确定每个时间步选择 k 个臂的概率，因此不会根据依次选取的臂所产生的观测值进行调整。他们证明，对于任意老虎机实例 $\boldsymbol{\mu} \in \mathcal{S}$,

$$\forall \boldsymbol{w} \in \Sigma_k, \quad \liminf_{n\to\infty} \frac{n}{\log(1/p_{\boldsymbol{\mu},n})} \geq \Gamma^*_{\mathrm{na}}(\boldsymbol{\mu})$$

and

$$\exists \boldsymbol{w}^*(\boldsymbol{\mu}) \in \Sigma_k, \quad \limsup_{n\to\infty} \frac{n}{\log(1/p_{\boldsymbol{\mu},n})} \leq \Gamma^*_{\mathrm{na}}(\boldsymbol{\mu})$$

其中

$$\Gamma^*_{\mathrm{na}}(\boldsymbol{\mu}) = \left(\sup_{\boldsymbol{w}\in\Sigma_k} \inf_{\boldsymbol{\nu}\in \mathrm{Alt}(\boldsymbol{\mu})} \sum_{i=1}^{k} w_i \mathrm{KL}(\nu_i \| \mu_i) \right)^{-1}. \tag{5}$$

乍看之下，方程（$\Gamma^*_{\mathrm{na}}(\boldsymbol{\mu})$）中的复杂度项 与方程（5）中的 $\Gamma^*_{\mathrm{fc}}(\boldsymbol{\mu})$）看起来相同。事实上，两者并不相同，因为 KL 散度通常不具对称性；但对于高斯分布，$\Gamma^*_{\mathrm{na}}(\boldsymbol{\mu}) = \Gamma^*_{\mathrm{fc}}(\boldsymbol{\mu})$。请注意，最优采样向量 $\boldsymbol{w}^*(\boldsymbol{\mu})$）取决于对未知老虎机实例 $\boldsymbol{\mu}$的了解，因此事先并不知道。因此，对于我们关注的自适应固定预算设定，针对每个老虎机实例 $\boldsymbol{w}^*(\boldsymbol{\mu})$）始终固定最优采样向量 $\boldsymbol{\mu}$ 的采样规则并非有效选择。

这两个猜想都不成立。Ariu 等人 (2021) 的结果表明，对于伯努利老虎机，这两个猜想通常都不成立。受 Carpentier 和 Locatelli (2016) 的构造启发，Ariu 等人(2021) 构造了一组臂数较多的老虎机实例，并证明这两个猜想不可能对所有实例都成立。我们认为，对于高斯老虎机也可以证明类似的否定结果。

5. 双臂老虎机的已知结果

虽然这两个猜想一般而言都不成立，但 Kaufmann 等人（2016）证明，对于已知方差的双臂高斯老虎机，即对于任意此类老虎机实例 $\boldsymbol{\mu}$，$\Gamma^*_{\mathrm{fb}}(\boldsymbol{\mu}) = \Gamma^*_{\mathrm{fc}}(\boldsymbol{\mu}) = \Gamma^*_{\mathrm{na}}(\boldsymbol{\mu})$ 均成立。该文还证明，最优采样规则是非自适应的，即采样向量固定为 $(\frac{\sigma_1}{\sigma_1+\sigma_2}, \frac{\sigma_2}{\sigma_1+\sigma_2})$，其中 σ_1 和 σ_2 是两条臂的已知方差。最近，Kato 等（2022）证明，当两条臂的未知均值之差趋于零时，即使方差也未知，所提算法的上界仍与 Kaufmann 等人（2016）给出的依赖于实例的下界相匹配。Adusumilli（2022）研究了双臂高斯老虎机的扩散渐近域，并证明相同的采样向量 $(\frac{\sigma_1}{\sigma_1+\sigma_2}, \frac{\sigma_2}{\sigma_1+\sigma_2})$ 也是极小极大最优的。然而，对于双臂伯努利老虎机，Kaufmann 等人（2016）证明，尽管最优采样向量存在，但它需要知道两条臂的未知均值，而这些均值事先并不可知。目前尚不清楚是否存在一种算法，无需满足这一要求便能达到渐近最优性。

致谢

我们感谢 Kaito Ariu、R´emy Degenne、Sandeep Juneja、Masahiro Kato、Junpei Komiyama、Wouter M.Koolen、Pierre M´enard、Daniel Russo 和 Assaf Zeevi 进行的富有成效的讨论。

参考文献

Karun Adusumilli. 两臂最佳臂识别的极小极大策略，2022年。URLhttps://arxiv.org/abs/2204.05527。Kaito Ariu、Masahiro Kato、Junpei Komiyama、Kenichiro McAlinn 和 Chao Qin。策略选择与最佳臂识别：探索采样的渐近分析，2021年。URLhttps://arxiv.org/abs/2109.08229。Jean-Yves Audibert、S´ebastien Bubeck 和 R´emi Munos。多臂老虎机中的最佳臂识别。载于第23届年度学习理论会议，第41–53页，2010年。S´ebastien Bubeck、R´emi Munos 和 Gilles Stoltz。多臂老虎机问题中的纯探索。载于算法学习理论国际会议，第23–37页。施普林格，2009年。Alexandra Carpentier 和 Andrea Locatelli。固定预算最佳臂识别老虎机问题的紧致（下）界。载于 第29届年度学习理论会议，第590–604页，2016年。Herman Chernoff。序贯实验设计。《数学统计年鉴》，30(3)：755–770，1959年。Aur´elien Garivier 和 Emilie Kaufmann。固定置信度下的最优最佳臂识别。载于第29届年度学习理论会议，第998–1027页，2016年。Peter Glynn 和 Sandeep Juneja。关于序数优化的大偏差视角。载于2004年冬季仿真会议论文集，第1卷。IEEE，2004年。Masahiro Kato、Kaito Ariu、Masaaki Imaizumi、Masahiro Nomura 和 Chao Qin。小间隙下固定预算的最佳臂识别，2022年。URLhttps://arxiv.org/abs/2201.04469。Emilie Kaufmann、Olivier Capp´e 和 Aur´elien Garivier。关于多臂老虎机模型中最佳臂识别的复杂性。《机器学习研究杂志》，17(1)：1–42，2016年。