分组多臂老虎机中的固定预算约束最佳臂识别

Raunak Mukherjee[1] **Sharayu Moharir**[2]

[1]印度理工学院孟买分校电气工程系，[2]印度理工学院孟买分校电气工程系

arXiv:2603.04007v1 [cs.LG] 4 Mar 2026

摘要

我们研究分组多臂老虎机中的固定预算约束最佳臂识别问题，其中每个臂由多个相互独立、具有随机奖励的属性组成。只有当某个臂的所有属性均值都高于给定阈值时，该臂才被视为可行臂。我们的目标是找到总体均值最大的可行臂。首先，我们推导出该设定下任意算法的错误概率下界。随后，我们提出可行性约束逐次拒绝（FCSR），这是一种在确保可行性的同时识别最佳臂的新算法。我们证明，在指数方面，该算法对问题参数的依赖达到最优，误差仅在常数因子范围内。实验结果表明，FCSR 优于自然基线方法，同时保留了可行性保证。

1 引言

纯探索随机多臂老虎机问题旨在从给定的臂集合中识别最优臂。最佳臂识别（BAI）最常见的两种范式包括固定置信度设置 [Kauf- mann 等人， 2016]以及固定预算设置 [Audibert 等人， 2010]。固定预算最佳臂识别（FBBAI）在在线广告、推荐系统等领域有诸多应用。一般而言，它适用于包含有限测试阶段、随后进入商业化阶段的情形 [Audibert 等人， 2010]。

许多服务更适合建模为向客户提供的一组相互独立的服务的集合。通常，人们希望其中每项服务的质量都达到某一标准 [Dharod et al., 2024]。例如，一家典型的汽车维修店可能提供洗车服务、空调维修服务、轮胎和车轮养护服务、汽车检查等。为了评估此类服务，让客户分别为每项服务评分，并为每项服务维护一个评分是合理的。评估此类服务整体表现的一项合理指标，是各项服务评分的（加权）平均值。此外，若要认定一项服务设施可接受，可能还需要每项服务的评分都超过某一阈值

类似地，在在线广告中，人们可能希望找到能够面向多个人口统计细分群体投放的最佳广告创意（例如图片或视频广告）。每种广告创意都可以视为一个臂，其在各个人口统计细分群体中的（随机）表现可建模为该广告创意的一个属性。我们希望找到一种最佳广告创意，使其在任何人口统计细分群体中的表现都不会差到令人无法接受。

我们将此问题建模为分组赌博机设定下的纯探索多臂赌博机问题，其中每个臂是一组属性，这些属性被建模为相互独立采样的随机变量。如果某个臂的所有属性的平均奖励都超过给定阈值，则称该臂为可行。

我们的贡献。 我们提出了可行性约束连续拒绝算法（FCSR），这是一种新颖的混合采样算法，其对问题参数的依赖在指数上达到最优，误差仅为一个常数因子。FCSR 完全无需参数 [1]，也就是说，它不需要了解问题实例（而这类信息在实践中通常无法获得）。FCSR 是一种混合采样策略，它综合了现有的固定预算算法，并引入了一种新颖的采样启发式方法（SAMPLEUNTILFEASIBLE）。

我们首先为该设定定义一个新的复杂度参数 H_{FC}来说明这一点；若约束被放松，该参数可恢复固定预算最佳臂识别设定中众所周知的难度参数。我们以依赖于实例的参数 H_{FC}推导了错误概率的非渐近下界。随后，我们证明了FCSR错误概率的匹配上界，从而确立了指数中至多相差常数因子的最优性。 我们将FCSR与所考虑的基线方法进行测试，这些基线方法见于

[1]关于使用“无参数”和“最优”来描述算法的类似用法，参见 Yang and Tan [2021] 。

若干合成数据集以及真实数据集上的类似设定中。我们的结果表明，FCSR总体上优于基线方法。

相关工作。 在固定预算设定下，无约束单属性纯探索多臂赌博机的形式化问题已经得到了充分研究。Successive Rejects [Au- dibert et al.、 2010] 以及 Sequential Halving [Karnin et al. 等算法将探索预算划分为多个阶段，并在每个阶段结束时使用淘汰规则淘汰/拒绝一个或多个臂。随后，最后存活的臂被标记为最优臂。它们是在该设定下已知的最佳算法之一，同时相对简单。需要注意的是，在无约束情形下，决定拒绝哪个臂相对直接（按照每个臂的经验均值进行排序）。相比之下，Faizal and Nair [2022] 指出，在约束设定下，对具有最大经验均值的臂的追求必须与可能违反约束的风险相平衡。

Wang et al. [2021]研究了最佳臂识别问题中的线性约束以及单调安全约束。Hou et al.[2022]研究了一个在臂的方差上施加可行性约束的固定置信度最佳臂识别问题。这些工作与我们的工作不同，因为它们没有考虑识别由多个属性构成的臂。Katz-Samuels and Scott [2019]研究了识别满足一般可行性约束的最佳多维臂这一问题；然而，该工作聚焦于固定置信度设定，其中反馈是多维的，而本文则关注固定预算的单维反馈设定。

Locatelli 等人 [2016] 提出了阈值化多臂老虎机问题（TBP），并为该问题设计算法 APT。在 TBP 中，有 K 个标量值分布和一个阈值 τ。目标是在固定预算 T内，识别所有均值大于 τ 的分布。Katz and Samuels [2018] 通过考虑多维臂，以及识别均值属于给定多面体的臂这一问题，将 TBP 设置推广到上述情形。上述工作中考虑的约束与本文考虑的可行性约束非常相似。不过，上述论文关注的是识别满足这些约束的所有臂，而不是寻找满足约束的最佳臂。

Kagrecha et al. [2023]研究了一个类似的多维约束赌博机设定，但目标是遗憾最小化。Dharod et al. [2024] 和 Faizal and Nair [2022]研究了最佳臂识别目标。不过，前者考虑的是在对另一属性施加约束的情况下最大化单一属性，后者则采用与本文相同的多臂老虎机设定，但处于固定置信度机制下。据我们所知，固定预算设定仍是一个开放问题，本文对此给出了相关结果。

2 问题表述

约束分组多臂老虎机设定。我们有 K 个臂，每个臂有 M 个属性。令 $[K] = \{1, 2, \ldots, K\}$ 和$[M] = \{1, 2, \ldots, M\}$。臂 i的属性 j表示为元组（i, j），并与未知的奖励分布$\nu_{i,j}$相关联。所有属性的随机奖励均为独立同分布，并假设服从 R次高斯分布。我们将由乘积分布 $\mathcal{B} = \bigotimes_{(i,\ j)\ \in \mathcal{A}} \nu_{ij}$定义的对象称为多臂老虎机实例。

定义（R-次高斯分布）。令 $R > 0$。若对于所有 $t \in \mathbb{R}$，分布 ν满足 $\mathbb{E}_{X\sim\nu}[\exp(t(X - \mathbb{E}[X])] \leq \exp\left(\frac{R^2t^2}{2}\right)$，则称其为 R-次高斯分布。

属性 (i, j) 的平均奖励 定义为 $\mu_{i,j} := \mathbb{E}_{X\sim\nu_{i,j}}[X]$。臂 i 的平均奖励 是其属性均值的平均值，定义为

$$\mu_i := \frac{1}{M}\sum_{j=1}^{M} \mu_{i,j}. \tag{1}$$

本文考虑所有属性的简单平均值，但相关结果很容易推广到加权平均的情况。若臂 i 的所有属性 (i, j) 的平均奖励高于给定阈值 $\tau \in \mathbb{R}$，则称该臂可行 。我们将所有可行臂构成的集合定义为 $\mathcal{F} := \{i \in [K] : \min_{j\in[M]} \mu_{i,\ j} > \tau\}$。令 $\mathcal{F}^c := [K]\backslash\mathcal{F}$ 表示所有不可行 臂构成的集合。

如果存在，最佳可行臂为

$$i^\star := \arg\max_{i\in\mathcal{F}} \mu_i.$$

我们假设唯一最佳臂存在。推广到多个最优臂的情形是直接的。如果不存在可行臂，则将给定实例判定为不可行。我们通过定义 $i^\star = 0$ 为表示该实例不可行性的标志 来表示这种情况。

目标。 给定一个未知的多臂老虎机实例 $\mathcal{B}$，在任意时刻 t，学习器会选择一个属性（i_t, j_t），并获得从分布 $X_{i,\ j}(t) \sim \nu_{i,\ j}$ 中抽取的随机奖励。自适应学习器根据过去观察到的样本，在时刻 t 做出决策，并在耗尽预算 T 后输出决策$I_t \in \{0, 1, \ldots, K\}$。学习器的目标是尽量降低其偏离真实 $i^\star$的概率。形式上，给定学习器已知的预算 T 和阈值τ，以及学习器未知的多臂老虎机实例 $\mathcal{B}$，学习器旨在最小化

$$\mathbb{P}_{\mathcal{B}}(I_t \neq i^\star). \tag{2}$$

3 下界

本节给出此设定下任何策略所导致的最小错误概率的一个基本下界。

令 $\mathcal{R}=\{i:\ i\in\mathcal{F}^c,\ \mu_i\geq\mu_{i^\star}\}$。注意，若 $i^\star=0$，则$\mathcal{R}=\mathcal{F}^c$。下文将高风险臂称为 $\mathcal{R}$，它包含所有平均均值奖励高于最优臂的不可行臂。当自适应学习器错误地报告最佳臂时，即 $I_T\neq i^\star$，存在三种不同的可能情况：(i) 最佳臂 $i^\star$ 存在，但被错误地判定为不可行；(ii) 某个可行次优臂 $i'\in\mathcal{F}_S$被认为其均值奖励高于 $i^\star$；或者 (iii) 某个高风险 臂 $r\in\mathcal{R}$ 被判定为可行（并且根据定义，r 的均值奖励高于 $i^\star$）。不存在最佳臂且 $i^\star=0$的情况属于情形 (iii) 的一种子情况。

复杂度参数定义。我们定义一些与此设置下问题实例难度相关的参数。定义属性 $(i,\ j)$ 的阈值间隙为 $\Delta_{i,\ j}:=|\mu_{i,\ j}-\tau|$。此外，令 $\Delta_{0,\ j}=\infty\,\forall\,j\in[M]$。定义臂 $i\neq i^\star$ 的次优间隙为 $\Delta_i:=|\mu_{i^\star}-\mu_i|\,\forall\,i\in\{0,\ldots,K\}$。令 $\mu_0:=\infty$，因此若最佳臂不存在 ($i^\star=0$)，则 $\Delta_i=\infty\,\forall\,i$。定义 $\Delta_{0,\ j}$ 和 μ_0，是为了确保即使最佳臂不存在，以下参数仍然定义良好。

假设所有臂均按其平均均值奖励排序编号，即 $\mu_1\geq\mu_2\geq\cdots\geq\mu_K$。我们分别在某个老虎机问题 $\mathcal{B}$ 上定义参数均值难度、风险难度和可行性难度，如下所示：

$$H_2^R(\mathcal{B}):=\max_{i\in[|\mathcal{R}|+2:K]}i\Delta_i^{-2},\tag{3}$$

$$H_f(\mathcal{B}):=\frac{K}{\log(K)}\max_{j\in[M]}\bar{\Delta}_{i^\star j}^{-2},\tag{4}$$

$$H_{\text{tbp}}(i):=\sum_{j\in[M]}\bar{\Delta}_{i,j}^{-2}\ ,\ H_{\text{tbp}}(\mathcal{B}):=K\max_{i\in\mathcal{F}^C}H_{\text{tbp}}(i).\tag{5}$$

令 $H_{\text{tbp}}:=0$，其中 $|\mathcal{F}^C|=0$，即不存在不可行臂。对 $\mathcal{B}$ 的依赖通常是隐含的。因此，将多臂老虎机实例 $\mathcal{B}$ 的总体可行性约束难度参数定义为

$$H_{FC}(\mathcal{B}):=\max\left\{H_2^R,\,H_{\text{tbp}},\,H_f\right\}.\tag{6}$$

注 1。请注意，在允许 $\tau=-\infty$ 的情况下，所有臂几乎处处可行，且该问题实例等同于带分组臂的无约束最佳臂识别。我们有

$$H_f=0\qquad H_{tbp}=0\qquad H_{FC}=\text{和。}$$

$H_2=\sum_{i\in[K]}i\Delta_i^{-2}$ 因此，max，并恢复了固定预算无约束设定中广泛使用的难度指数（见 *Audibert* 等人 *[2010]*）。如果将每个分组臂视为具有相同臂均值的单维臂，那么这正是我们预期在标准最佳臂识别中看到的难度指数。

定理 1（下界）。设 $\mathcal{C}_{FC}(a;\ K)$ 表示满足 $K=M\geq2$ 且其难度H_{FC} 被某个常数 a上界约束的老虎机实例集合。若

$$T\geq\frac{4}{60^2}(a\log(K))^2\log(6T(K+M+1)),\tag{7}$$

存在一个多臂老虎机实例 $\mathcal{G}\in\mathcal{C}_{FC}(a)$，使得任意学习器错误报告最佳臂的概率至少为

$$\mathbb{P}_{\mathcal{G}\in\mathcal{C}_{FC}(a)}\geq\frac{1}{6}\exp\left(\frac{-1200T}{\log(K)H_{FC}(\mathcal{G})}\right).\tag{8}$$

证明概要与讨论。证明通过构造两个对抗性老虎机问题族来进行，这两个问题族是 $\mathcal{C}_{FC}(a)$ 的子集，并刻画了两种主要失效模式：(i) 一个可行性族 $\mathcal{C}_F$，其中只有一个臂越过阈值 τ；以及 (ii) 一个风险族 $\mathcal{C}_R$，该族提高竞争臂的均值，或降低最佳臂的单个属性。我们将 Carpentier 和 Locatelli[2016] 所采用的技术推广到这一更一般的设定。对于每个问题族，我们建立一个集中事件 ξ，并对某个采样不足的臂应用测度变换论证，从而得到所需结果。对两个问题族取最大值，便得到带有组合难度 $H_{FC}=\max\{H_{tbp},H_2^R,H_f\}$ 的上述界。主要的新贡献是构造了一类丰富的多维老虎机实例，用于刻画违反可行性约束与均值判别错误之间的多种失效模式。完整证明见附录C。

4 可行性约束连续拒绝算法

4.1 符号说明

定义 $[i:\ j]=\{i,i+1,\ldots,j\}$ 和 $[j]=[1:\ j]$，其中i 和 j 为整数。令 $X_{i,j}(t)\sim\nu_{i,j}$ 表示在时间$t>0$ 拉取属性 (i,j) 时观察到的随机奖励（除非另有说明，$i\in[K]$，$j\in[M]$）。令 $C_{i,j}(t)$ 表示在时间 t 之前拉取 (i,j) 的次数，$S_{i,j}$ 表示在 t之前观察到的所有奖励之和。此外，定义属性经验均值 $\hat{}\mu_{i,j}(t):=S_{i,j}(t)/\max\{C_{i,j}(t)1\}$。因此，在时间 t 拉取属性 (i,j) 时，我们进行如下更新

$$\begin{aligned}&S_{i,j}(t+1)\leftarrow S_{i,j}(t)+X_{i,j}(t)\,,\\&C_{i,j}(t+1)\leftarrow C_{i,j}(t)+1\,,\\&\hat{\mu}_{i,j}(t+1)=\frac{S_{i,j}(t+1)}{\max\{C_{i,j}(t+1),1\}}\,,\end{aligned}\tag{9}$$

并且 $S_{k,\ m}(t+1)=S_{k,\ m}(t)\quad C_{k,\ m}(t+1)=C_{k,\ m}(t)$，$\hat{}\mu_{k,\ m}(t+1)=\hat{}\mu_{k,\ m}(t)$ 对所有其他属性 $(k,\ m)\neq(i,\ j)$ 均保持不变。

令 $S(t)\in\mathbb{R}^{K\times M}$ 为一个矩阵，其每个元素由 $S_{i,j}(t)$ 给出，臂 i 的属性和向量 $S_i(t)$ 为 $S(t)$ 的第 i 行。类似地，定义 $C(t)\in\mathbb{N}^{K\times M}$、$C_i(t)\in\mathbb{N}^M$，以及属性经验均值矩阵 $\hat{}\mu(t)\in\mathbb{R}^{K\times M}$。

FCSR算法维护由内部采样子程序更新的全局属性统计矩阵 S、C $\hat{}\mu$

因此，为简便起见，在提及全局属性统计量 t 时，我们省略时间步 S、C $\hat{}$$\mu$以及奖励 $X_{i,\ j}$。但是，我们会明确标记仅在子程序作用域内使用的统计量。

4.2 算法描述

我们简要回顾 FCSR 所基于的两种固定预算老虎机算法。逐次拒绝 （SR） [Audibert 等人，2010]是一种固定预算最佳臂识别策略，分 $K-1$ 个轮次进行。在每一轮次中，所有存活臂均按照预定调度被均匀采样，并计算每条臂均值的经验估计。估计最差的臂将被淘汰。

针对阈值化多臂老虎机问题的*APT* 程序[Locatelli 等人， 2016] 处理的是一个不同的问题：给定均值未知的臂和已知阈值，它自适应地分配样本，以判定每个均值位于阈值之上还是之下，并将样本集中于经验均值接近阈值的臂。APT 达到最优性能。算法 1 描述了臂i属性集合的 APT 采样程序，并对全局臂统计向量 S_i, C_i 进行任意初始化。SR 与 APT 均能在指数中仅相差常数因子的意义下达到最优性能 [Carpentier 和 Locatelli，2016，Locatelli 等人， 2016]。

算法 1 APT$(i, T;\ S_i, C_i)$

要求： 臂 i、预算 T、阈值 $\tau \in \mathbb{R}$ 和全局统计量 S_i, C_i
初始化： 使用已有的全局 $S_i, C_i, \hat{}\mu_{i,j} \quad \forall\ j \in [M]$。
对于 $t = 0, 1, \ldots, T-1$ **执行**
  对每个属性 $j = 1, \ldots, M$ **执行**
    计算经验差距: $\hat{\Delta}_j(t) = \lceil \mu_{i,\ j} - \tau |$
    计算: $B_j(t+1) = \sqrt{C_{i,\ j}} \cdot \bar{\Delta}_{ij}(t)$
  结束循环
  选择属性 $J_{t+1} = \arg\min_{1 \le j \le M} B_j(t+1)$。
  观察奖励 $X_{i,J_{t+1}} \sim \nu_{i,J_{t+1}}$.
  使用 (9) 更新全局 $S_{i,\ J_{t+1}}$, $C_{i,\ J_{t+1}}$, $\hat{}\mu_{i,\ J_{t+1}}$。
结束循环
返回更新后的统计量 $S_i,\ C_i$。

FCSR 结合了这两种思想：SR 根据估计臂均值提供跨臂的全局淘汰计划，而 APT 则在局部用于将样本集中分配给接近可行性阈值的属性。SR 淘汰次优臂的同时，APT 采样则针对淘汰风险臂。此外，一种新颖的采样策略 SAMPLEUNTILFEASIBLE（SUF）会将总体预算中的专门部分用于确保最佳可行臂不会被错误淘汰。SUF 有选择地对不可行属性进行采样，以防止最佳臂过早被淘汰。这种组合自然形成了一个三阶段结构，下面将对此进行更详细的介绍。

算法2 UNIFORM$(i,\ T;\ S_i,\ C_i)$

要求： 臂 i、预算 T和全局统计量 $S_i,\ C_i$
  对于臂 i的属性 $j = 1,\ 2,\ \ldots M$**执行**
    对于样本 $s = 1,\ 2,\ \ldots \lfloor T/M \rfloor$**执行**
      观察奖励 $X_{i,j} \sim \nu_{i,j}$
      更新 $S_{i\ ,J_{t+1}}$ $C_{i,J_{t+1}}, \hat{}\mu_{i,J_{t+1}}$ 使用 (9)。
    结束循环
  结束循环
  返回 更新后的统计量 $C_i,\ S_i$

给定总样本预算 T 以及分数超参数$f,\ g \in$（0，1）。我们预留预算的总比例 fT，并将其均分：对每个臂 i 设置可行性预算P_i

$$P_i \leftarrow \lfloor fT/K \rfloor.$$

初始化共享的剩余可行性样本池 $T_{\text{extra}} \leftarrow 0$。当臂 i 被淘汰且其可行性预算为 $P_i > 0$时，我们将这些未使用样本重新利用，将其加入“额外”样本池 $T_{\text{extra}} \leftarrow T_{\text{extra}} + P_i$，随后用于进一步对属性进行均匀采样（见下文的均匀阶段）。剩余的（$1-f$）T个样本按照SR调度在各轮次间分配。

在原始SR算法中， n_r 是轮次 r结束时所有臂收到的样本总数。因此，$\Delta n_r := n_r - n_{r-1}$ 是每个存活臂的“每轮”预算。形式上，

$$n_r := \left\lceil \frac{(1-f)T}{\bar{n}(K+1-r)} \right\rceil, \quad \Delta n_r := n_r - n_{r-1} \qquad (10)$$

其中 $n_0 := 0$。 $\bar{n} = \frac{1}{2} + \sum_{k=2}^{K} \frac{1}{k}$ 是归一化常数。在轮次 $r = 1,$ ....., $K-1$ 中，令 S_r 表示存活臂集合。每个存活臂 $i \in S_r$ 接受三个顺序采样阶段（所有整数数量均按下文所述进行舍入）。令 $|S_r|$ 表示存活臂的数量，并令 $\mathcal{A} = [K]$为所有臂的集合。

1. **均匀阶段**。 向每个臂分配 $\lfloor (1-g)\Delta n_r \rfloor$ 个样本，并在该臂的各属性间均匀分配（算法2）。此外，将剩余池 T_{extra} 中的 $\lfloor T_{\text{extra}}/|S_r| \rfloor$ 个样本重新分配给每个存活臂；相应地减少 T_{extra}。 2. **风险（APT）阶段**。 根据 APT 阈值子程序（算法1），每个臂分配 $\lfloor g\Delta n_r \rfloor$ 个样本。臂 i 的 APT 分配在第 r轮的均匀阶段观测之后进行。 3. **可行性阶段**。 对于臂 i 中仍经验不可行 的任何属性 j，依次在该属性上抽取额外样本，最多不超过该臂剩余的可行性预算 P_i。当属性（$i,\ j$）达到经验可行时，停止对其抽样。从中扣除每个可行性

样本。如果臂 i 之后被弃用，其剩余的P_i 将转入池中：$T_{extra} \leftarrow T_{\text{extra}} + P_i$。参见 SAMPLEUNTILFEASIBLE（算法3）。

在第 r 轮的所有阶段结束后，为每个仍存活的臂 i 计算淘汰分数。臂 i 在某个时刻 t 的分数定义为

$$s(i;t) := \begin{cases} \hat{\mu}_i(t) & \text{if } \min_{j\in[M]} \hat{\mu}_{ij}(t) > \tau, \\ \min_{j\in[M]} \hat{\mu}_{ij}(t) & \text{otherwise.} \end{cases} \tag{11}$$

重复此过程，直到只剩下一条臂；若该臂可行，则返回该臂，否则返回 -1。完整伪代码请参见算法4。

算法3 SAMPLEUNTILFEASIBLE(i, P_i; S_i, C_i, τ)

要求： 臂 i，可行性预算 P_i，全局 $S_i \in \mathbb{R}^M$，$C_i \in \mathbb{N}^M$，阈值 τ
初始化： 使用已有的 S_i, C_i, $\hat{\mu}_{i,j}\ \forall j \in [M]$.
当 $P_i \geq 1$ **时**
  $\mathcal{F}^c_{,(i)} \leftarrow \{j : \hat{\mu}_{i,j} \leq \tau\}$ ▷ 不可行属性集合
  若 $\mathcal{F}^c_{,(i)} = \emptyset$，则中断
  $j^* \leftarrow \min_{j\in\mathcal{F}^c_{,(i)}} j$ ▷ 选择最低索引
  当 $\hat{\mu}_{i,j^\star} \leq \tau$ **时**
    观察奖励 $X_{i,j^\star} \sim \nu_{i,j^\star}$
    使用 (9) 更新 $S_{i,j^\star}, C_{i,j^\star}, \hat{\mu}_{i,j^\star}$。
    $P_i \leftarrow P_i - 1$
    若 $P_i = 0$ ▷ 则跳出 当可行性预算耗尽时
  结束 while 循环
结束 while
返回 更新后的统计量 S_i, C_i。

算法4 FCSR(T, τ, f, g; S, C)

要求： K 个臂、每个臂有 M 个属性、预算 T、阈值τ、比例 $f, g \in (0, 1)$ 以及矩阵 S, C。确保：最优可行臂或 -1。初始化：全局统计量 $S_{i,j}, C_{i,j}, \hat{\mu}_{i,j} \leftarrow 0$. $P_i \leftarrow \lfloor \frac{fT}{K} \rfloor \forall\, i, T_{\text{extra}} \leftarrow 0$。对于 $r = 1$ 到 $K-1$，执行：按照 (10) 定义 $\Delta n_r \leftarrow n_r - n_{r-1}$。$T^{\text{arm}}_{\text{extra}} \leftarrow T_{\text{extra}}/|\mathcal{A}|$; $T_{\text{extra}} \leftarrow 0$。对于每个臂 $i \in \mathcal{A}$，执行：运行 UNIFORM($i, \lfloor (1-g)\Delta n_r \rfloor + T^{\text{arm}}_{\text{extra}}$; S_i, C_i)。运行 APT($i, \lfloor g \cdot \Delta n_r \rfloor, \tau$; S_i, C_i)。运行 SAMPLEUNTILFEASIBLE(i, P_i, τ; S_i, C_i)。结束循环。使用 (11) 计算 $s(i)\ \forall\, i$。淘汰 $\mathcal{A} \leftarrow \mathcal{A} \setminus \{\arg\min_{i\in\mathcal{A}} s(i)\}$. $T_{\text{extra}} \leftarrow T_{\text{extra}} + P_{\arg\min_{i\in\mathcal{A}} s(i)}$。结束循环。如果最后剩下的臂 $i_{\text{final}} \in \mathcal{A}$ 可行，则返回i_{final}.；否则返回 0。

4.3 理论分析

定理2 (FCSR的性能). 设 K、$M \geq 1$。给定参数 f、$g \in (0, 1)$ 以及 τ。设 $c = 1/32R^2$ FCSR(T, τ, f, g) 的错误概率满足

$$\mathbb{P}(e) \leq 3K^2 \exp\left(\frac{-cT}{\log(K) H_{FC}(\mathcal{B})}\right), \tag{12}$$

对于所有

$$T \geq \max\Big\{256 H_{tbp} R^2 \log((\log(T)+1)M), \frac{4K^2M}{f}, \frac{K \max\{H_{tbp}{}^{-1}, (H_2^R)^{-1}\}}{\log(K)}\Big\}.$$

证明概要。 主要创新在于采样过程 SUF，而整体分析建立在 Audibert 等人对逐次淘汰的分析之上 [2010]。我们根据真实均值对臂排序，并将最优臂 $i^\star$ 被淘汰这一事件分解为三种情况：(i) $i^\star$ 被错误地被判定为不可行（f_i），(ii) 次优可行臂因估计误差而胜过$i^\star$（s_i），或（iii）风险臂（不可行臂）被错误地判定为可行，并淘汰 $i^\star$（r_i）。对各轮次应用并集界可得 $\sum_i \mathbb{P}(e_i) \leq \sum_i \mathbb{P}(f_i) + \sum_i \mathbb{P}(s_i) + \sum_i \mathbb{P}(r_i)$。我们分别对每一项进行界定，即可得到所述结果。完整证明见附录D。

比较 FCSR 下错误概率的上界（12）与下界（8），可以发现，在这两种情况下，错误概率都由基本难度参数 H_{FC} 刻画（带有一个 log(K) 因子），这表明 FCSR 在指数上达到最优性，[2] 误差仅相差常数因子。

与 APT 采样这一自然替代方案相比，新颖的采样过程 SUF 对 $\mathbb{P}(f_i)$ 给出了强得多的上界。直观地说，假设由于某个属性的经验均值低于 τ，最佳臂被判定为不可行，那么 APT 可能会把样本集中在其他不那么关键、但被判定为可行的属性上，仅仅因为这些属性更接近阈值。相比之下，SUF 只针对这些关键的经验不可行属性集中采样。

引理3。 假设 $i^\star$ 存在且 $T \geq$ 最大值$\{\frac{4K^2M}{f}\}$，在 *SUF* 下，$i^\star$ 在第$r \in [K-1]$ 轮结束时被判定为不可行的概率上界为

$$\mathbb{P}(i^\star \in \mathcal{F}^c_r) \leq \exp\Big(-\frac{fT}{16R^2 \log K\, H_f}\Big). \tag{13}$$

证明概要与讨论。证明利用了以下包含关系

[2] See Appendix B for a more detailed discussion on *optimality*

最佳臂在第 i 轮结束时被判定为不可行这一事件，是第 i轮结束时可行性预算耗尽这一事件的子集。我们构造随机变量Z_j，用于统计直到最佳臂的属性 j 被判定为可行所需的样本数，并通过停时分析得到一个尾界。随后，我们对所有属性的这一数量进行汇总，并通过Chernoff界得到结果。完整证明见附录9。

为了从形式上看出SUF的优越性能，考虑如下情形：将每轮预算 Δn_r （定义见(10)）的比例 f 通过APT均匀分配到全部 K 条臂上，用于“可行性”采样；此时错误概率的上界为

$$\mathbb{P}_{APT}(f_i) \leq \exp\left(\frac{-f(T-K)}{\overline{\log}(K)K(K+1-i)H_{tbp}(i^\star)}\right). \tag{14}$$

上述结果源于对APT采样错误概率上界的应用（见附录D中的引理10）。此情形下分母中的因子按 $\mathcal{O}(K^3)$ 缩放，而SUF对应的缩放为 $\mathcal{O}(K)$。这是由于分母中额外含有 K 因子，并且 $H_{tbp}(i^\star)$是对所有属性求和，而 H_f 只是一个最大值。

5 数值分析

我们在合成数据与真实世界数据上评估FCSR。我们构造了四个合成老虎机实例，用以改变难度指数，并对照基线算法检验FCSR在各类困难问题实例中的稳健性。我们还在源自MovieLens数据集 [Harper and Konstan, 2015b]的随机生成老虎机实例上测试FCSR，以评估其在低预算实际场景中的表现。

5.1 实验设置

FCSR 的超参数设为 $g = 0.3$ 和 $f = 0.2$，通过经验选择，使样本分配在均值区分和可行性检验之间大致均衡。分配给 UNIFORM 的样本比例为 $= 0.8 \times 0.7 = 0.56$。分配给 SUF 的比例为 $= 0.2$，分配给 APT 的比例为 $= 0.8 \times 0.3 =$0.24。

合成实验使用具有 $K = 10$臂和 $M = 5$ 属性的老虎机实例，运行 $N = 2000$次迭代。属性服从 $\nu_{ij} \sim \mathcal{N}(\mu_{ij},$ 0.3)，其中 $\mathcal{N}(0, 1)$表示均值为0、方差为1的正态分布。算法使用错误概率 $\mathbb{P}$(error)进行评估，其定义为算法未能识别真正的最佳可行臂 $i^\star$的运行所占比例。我们绘制错误概率的自然对数log($\mathbb{P}$(error))相对于总预算 T的曲线。报告值为$N = 2000$次试验的平均值。

5.2 基线方法

均匀采样（US）。 总预算在所有臂-属性对之间平均分配。计算经验均值以及经验可行臂集合。预算耗尽后，选择经验总体均值最高的臂；若没有可行臂，则算法按约定返回0。

逐次拒绝（SR）。我们采用一种 SR 的变体，类似于 Faizal 和 Nair[2022] 的先检验不可行性方案。样本按照逐次拒绝调度进行分配。与 FCSR 类似，使用式 (11) 中的评分规则淘汰臂。若最终剩下的臂可行，则将其标记为最优；否则返回 0。

探索后承诺（ETC）。这是经典探索后承诺算法的两阶段变体。在第一阶段，将总预算中的固定比例用于对所有属性进行均匀采样。按照 SR 对臂进行排序，并将排名前 M 个臂组成候选集。随后，将剩余预算均匀分配给这些候选臂的属性。最后返回排名最高的臂。

5.3 合成实例

我们设计了四种不同的问题实例，用于对算法的不同方面进行压力测试。这些实例由一个较小的常数 $a \in$ (0.001,0.1) 参数化，以控制难度。为方便记号，定义$[i:$ $j] = \{i, i+1, \ldots, j\}$ 其中 i和 j为整数。令$\mu_{[a:b],}$ $_{[c:d]} = x$ 表示所有 $m \in [a:\ b]$ 以及所有$k \in [c:d]$的 $\mu_{mk} = x$ 。

实验1：风险实例。 我们设定 $\tau = 0.5$ 和 a =0.01。只有可行臂 $i^\star = 10$，其 $\mu_{i^\star,[1:M]}$ =0.7。其余 $K - 1 = 9$ 个臂均不可行，但总体均值较高。$\mu_{[1:9],5} = 0.5 - a \leq \tau$。$\mu_{[1:9],[1:4]}$ $=0.8 + \frac{a}{4}$。因此， $\mu_{[1:9]} = 0.74$。

结果与讨论。 实验1旨在考验那些追求总体均值较高的臂、却未充分验证其可行性的算法。我们观察到，在实验1中，FCSR显著优于对风险臂鲁棒性较低的基线算法，SR排名第二。所有4项实验的结果见图 1。

实验2：可行性实例。 我们设置 $\tau = 0.5$ 和$a = 0.01$。最优臂 $i^\star = 10$ 的均值较高，且$\mu_{i^\star,\ [1:\ 4]} = 0.8$，但其第五个属性仅勉强可行，$\mu_{i^\star,\ 5} = 0.5 + a$。其余 $K - 1 = 9$ 个臂均轻松满足可行性要求，$\mu_{[1:9],}$ $_{[1:5]} = 0.6$但明显次优。

结果与讨论。 实验2的难点在于收集足够的证据，以确认

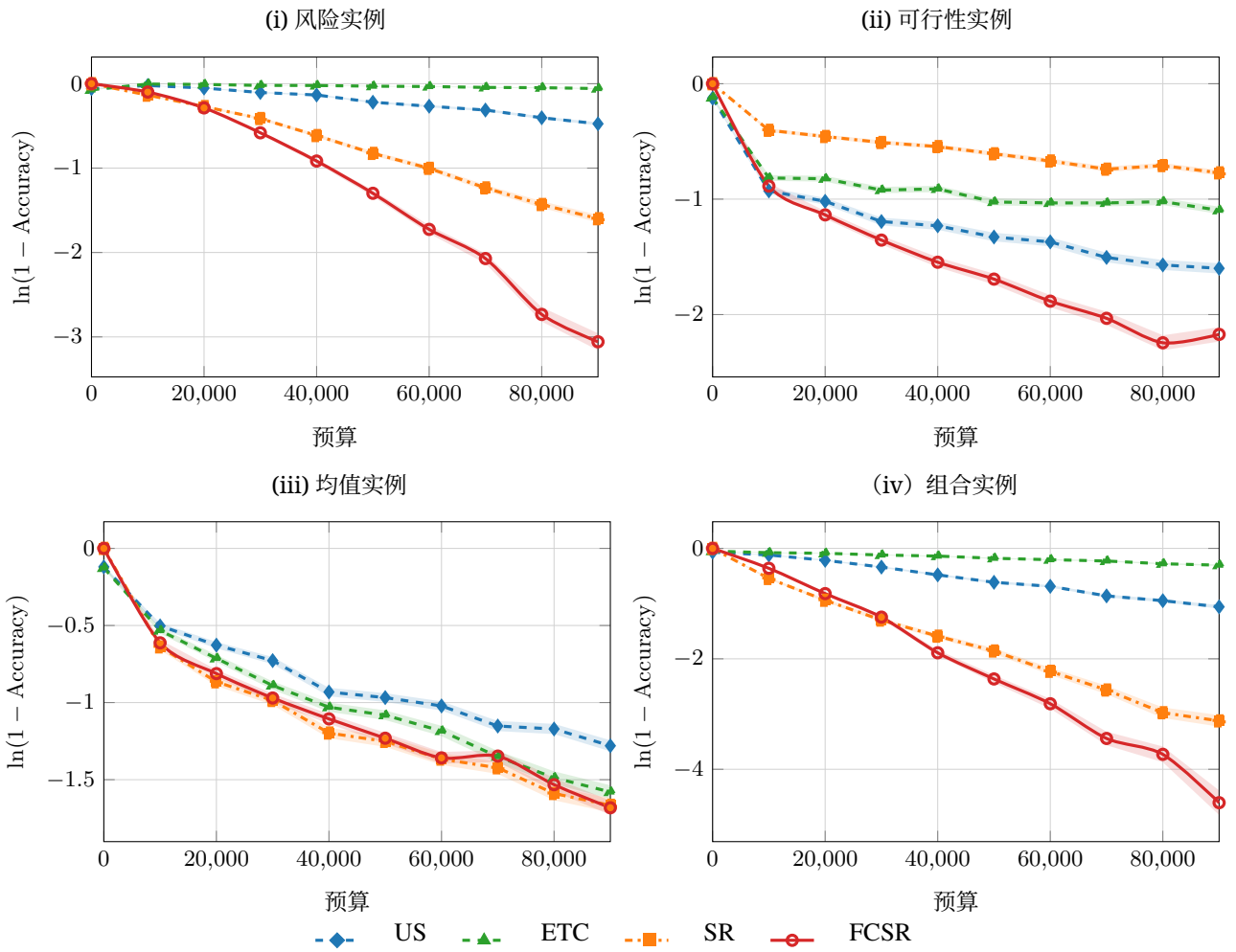


图 1：合成实例上的性能比较。每个子图展示四种算法（US、ETC、SR、FCSR）的 ln(1 −准确率) 与预算的关系。阴影带使用德尔塔方法近似 $\mathrm{Var}(\ln(1-\hat{A})) \approx \frac{1}{N}\frac{\hat{A}}{(1-\hat{A})}$ 计算， $N = 2000$。

最佳臂的可行性，同时避免误将某个次优但“更安全”的臂识别为最佳臂。在实验2的可行性实例和实验4的组合实例中，FCSR仍在很大程度上优于基线方法，但差距不像实验1中那么明显。

实验3：均值识别实例。我们设定$\tau = 0.3$ 和 $a = 0.003$。所有臂均明显可行。最佳臂 $i^\star = 1$ 具有$\mu_{i^\star,\,[1:\,5]} = 0.7$。每个后续臂的均值 $k \in \{2, \ldots, 10\}$ 逐步降低，按照参数为 a的算术平均值递减，即$\mu_k = \mu_{mk} = 0.7 - (k-1)a$ 对于所有 $k \in [2:\ 10]$ 和$m \in [1:\ 5]$。

结果与讨论。 实验3作为对照，与标准最佳臂识别等价。它测试 FCSR 在将预算分配到各种采样策略时所经历的性能下降。在实验3中，我们观察到 SR 是最佳算法，FCSR 紧随其后。这是意料之中的，因为实验3本质上是一个最佳臂识别问题。

实验4：组合实例。 我们设置 $\tau = 0.5$ 和$a = 0.01$。该实例由前面三个实例组合而成。最佳臂 $i^\star = 10$ 的臂均值为 $\mu_{i^\star}$ =0.7，其中一个属性接近阈值 $\mu_{i^\star,5} = 0.5 + a$，其余属性则为 $\mu_{i^\star,[1:4]} = 0.75$。前5个臂具有风险，且 $\mu_{[1:5],5} = 0.5 - a$，以及 $\mu_{[1:5],[1:4]} = 0.9 + \frac{a}{4}\sigma$ 接下来的4个臂在均值识别方面较难，也就是说，所有属性都相等，且臂均值以参数 a 按等差数列递减，从 $\mu_k = 0.7$到 $\mu_k = 0.7 - 4a$，适用于 $k \in [6:\ 9]$。

结果与讨论。实验4结合了上述3个困难实例。我们观察到，FCSR的对数错误概率仍然呈现相对更陡的下降趋势，并且优于所有其他基线方法。由于该问题实例同时包含均值区分和可行性检查，我们观察到，与实验1相比，其他算法（尤其是SR）在本实验中并没有那么不利。

表1：包含多种类型电影的电影组合。真实均值为 < 0.73 的行已着色 d.

臂编号	真实臂均值	属性编号	类型	电影名称	真实属性均值
0	0.826	1	喜剧	《公主新娘》（1987）	0.826
		2	动作	《星球大战：第四集——新希望》（1977）	0.824
		3	剧情	美国丽人（1999年）	0.821
		4	惊悚	移魂都市（1998年）	0.761
		5	科幻	鬼玩人3：魔界英豪（1993年）	0.747
1	0.740	1	喜剧	灼热的马鞍（1974年）	0.772
		2	动作	星球大战：第六集——绝地归来（1983年）	0.799
		3	剧情	桂河大桥（1957年）	0.819
		4	惊悚	空中监狱（1997年）	0.640
		5	科幻	X档案：征服未来（1998年）	0.668
2	0.710	1	喜剧	我的堂兄维尼（1992年）	0.721
		2	动作	碟中谍（1996年）	0.680
		3	剧情	离开拉斯维加斯（1995年）	0.735
		4	惊悚	魔鬼代言人（1997年）	0.709
		5	科幻	疯狂的麦克斯（1979年）	0.706

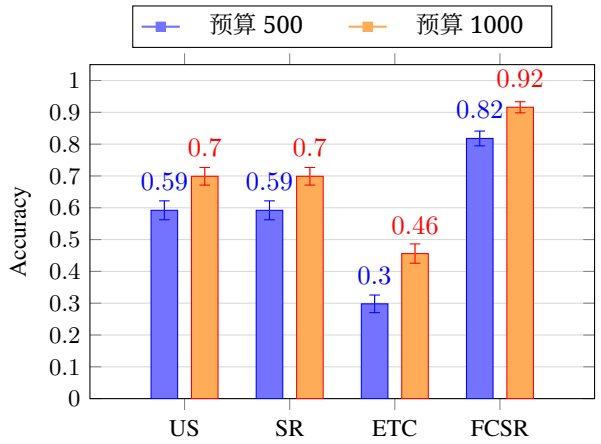


图2：两种预算下不同算法的准确率。误差线表示伯努利分布均值的95%置信区间，使用经验方差 $\mathrm{Var}(X) = \hat{p}(1-\hat{p})$ 计算，即 $\hat{p} \pm 1.96\sqrt{\hat{p}(1-\hat{p})/N}$，其中进行了 $N = 1000$ 次独立运行。

5.4 MOVIELENS数据集

我们使用MovieLens-25M数据集构建了一个分组老虎机实例 [Harper and Konstan, 2015a]。在此设置中，每个臂代表一个由不同类型（属性）的电影组成的“电影组合”。目标是找出每种类型都具有较高观众吸引力的组合，这反映了内容提供商需要策划一个能够持续满足不同人口群体需求的组合。

若某个臂中每部电影的平均归一化评分都超过阈值（我们设定阈值为 $\tau = 0.73$，相当于 3.65/5），则将其定义为可行臂，以确保组合中的所有类型都达到基本质量标准。最佳臂是整体平均评分最高的可行组合。我们筛选电影集，仅保留用户评分不少于 800 条的电影，以确保评分稳定性。在剩余电影中，我们找出出现频率最高的前 5 种类型，作为属性 $M = 5$，并定义 $K = 3$ 个臂，其中每个臂代表一个电影组合。

臂 i 的属性 j 对应的真实平均奖励定义为该电影所有用户评分的平均值，再除以 5.0 将其归一化到 $[0, 1]$ 区间。当老虎机算法选择该属性时，它会从该电影的实际历史评分中，以均匀有放回抽样的方式获得随机奖励。

讨论。 我们观察到，对于表1中的实例，无论是 $T = 500$ 还是 $T = 1000$（在图2中分别按此顺序绘制），FCSR在MovieLens数据集上的表现都优于基线方法。因此，即使在实际应用和预算有限的情况下，FCSR也切实可行。

6 结论

我们提出了可行性约束连续拒绝算法（FCSR），这是一种用于分组多臂老虎机中固定预算最佳臂识别的无参数算法。我们定义复杂度参数 H_{FC}，并据此推导该设定的下界。我们证明了FCSR误差概率的上界，并表明其在常数因子意义下与下界相匹配，从而证明了FCSR的最优性。为此，我们提出SAMPLEUNTILFEASIBLE，这是实现最优性能所必需的一种新颖子程序。实验上，FCSR在若干合成压力测试中优于自然基线方法，并且在低预算场景下适用于实际应用。

参考文献

Jean-Yves Audibert、Sébastien Bubeck与Rémi Munos。最佳臂识别：面向固定预算与固定置信度的统一方法。载于神经信息处理系统进展（*NIPS*），2010。URL http://papers.nips.cc/paper/4640-best-arm-identification-a-unified-approach-to-fixed-budget-and-fixed-confidence.pdf。

Alexandra Carpentier与Andrea Locatelli。固定预算最佳臂识别多臂老虎机问题的紧（下）界，2016。URL https://arxiv.org/abs/ 1605.09004。Sahil Dharod、Malyala Preethi Sravani、Sakshi Heda与Sharayu Moharir。分组多臂老虎机中的约束最佳臂识别，2024年。URL https://arxiv.org/abs/2412.08031。Fathima Zarin Faizal与Jayakrishnan Nair。固定预算下的约束纯探索多臂赌博机。*arXiv*，abs/2211.14768，2022年。URL https://api.semanticscholar.org/CorpusID:254044503。F. Maxwell Harper与Joseph A. Konstan。MovieLens 数据集：历史与背景。*ACM*交互式智能系统汇刊，5(4)，2015年12月a。ISSN 2160-6455。doi: 10.1145/2827872。URL https://doi.org/10.1145/2827872。F. Maxwell Harper与Joseph A. Konstan。MovieLens 数据集：历史与背景。*ACM*交互式智能系统汇刊（*TiiS*），5(4):19:1–19:19，2015b。doi: 10.1145/2827872。Yunlong Hou、Vincent Y. F. Tan与Zixin Zhong。近乎最优的方差约束最佳臂识别，2022年。URL https://arxiv.org/abs/2201.1 0142。Anmol Kagrecha、Jayakrishnan Nair与Krishna Jagannathan。多准则多臂赌博机的约束遗憾最小化，2023年。URL https: //arxiv.org/abs/2006.09649。Zohar Karnin、Tomer Koren与Oren Somekh。多臂赌博机中的近乎最优探索。载于Sanjoy Dasgupta与David McAllester编，第*30*届国际机器学习会议论文集，《机器学习研究论文集》第28卷，第1238–1246页，美国佐治亚州亚特兰大，2013年6月17日至19日。PMLR。URL https://proceedings.mlr.press/v28/karnin13.html。Gábor Katz与Yonatan Samuels。可行臂识别。载于《机器学习研究论文集》，编，机器学习研究论文集，《机器学习研究论文集》第80卷，第 1–17 页。PMLR，2018年 8 月 8–10 日。网址：http://proc eedings.mlr.press/v80/katz18a.html。

Julian Katz-Samuels 和 Clayton Scott。最佳可行臂识别。载于 Kamalika Chaudhuri 和 Masashi Sugiyama 编，第二十二届国际人工智能与统计会议论文集，第 89 卷，机器学习研究论文集，第 1593–1601 页。PMLR，2019年 4 月 16–18 日。网址：https://proceedings.mlr.press/v89/katz-samuels19a.html。

Emilie Kaufmann、OlivierCappé 和 AurélienGarivier。多臂老虎机模型中最佳臂识别的复杂度，2016。网址：https://arxiv.org/ abs/1407.4443。

Andrea Locatelli、Maurilio Gutzeit 和 Alexandra Carpen- tier。阈值化多臂老虎机问题的最优算法，2016。网址：https://arxiv.org/abs/ 1605.08671。

Chao Qin。开放问题：固定预算下的最佳臂识别最优方法，2023年。网址：https://arxiv.or g/abs/2303.00950。

ZhenlinWang、Andrew Wagenmaker 和 Kevin Ja mieson。带安全约束的最佳臂识别，2021。网址：https://arxiv.org/abs/2111.12151。

Junwen Yang 和 Vincent Y. F. Tan. 线性老虎机中极小极大最优最佳臂识别的探索。*CoRR*，abs/2105.13017，2021。网址：https://arxiv.org/abs/2105.13017。

分组赌博机中的固定预算约束最佳臂识别（补充材料）

Raunak Mukherjee[1] **Sharayu Moharir**[2]

[1]印度理工学院孟买分校电气工程系，[2]印度理工学院孟买分校电气工程系

A 伪代码

令 $X_{i,j}(t) \sim \nu_{i,j}$ 表示在时刻 $t > 1$ 拉动属性 (i, j) $i \in [K], j \in [M]$ 时所观测到的随机奖励。令 $C_{i,j}(t)$ 表示 (i, j) 在时刻 t 之前被拉动的次数，$S_{i,j}$ 表示 t 之前所有观测奖励之和，以及 $\hat{\mu}_{i,j}(t) = S_{i,m}(t)/\max\{C_{i,m}(t)\ 1\}$，即属性经验均值。因此，在时刻 t 拉动属性 (i, j) 时，我们更新

$$S_{i,j}(t+1) \leftarrow S_{i,j}(t) + X_{i,j}(t)\,, \quad C_{i,j}(t+1) \leftarrow C_{i,j}(t) + 1, \quad \hat{\mu}_{i,j}(t+1) = \frac{S_{i,m}(t+1)}{\max\{C_{i,m}(t+1), 1\}}. \tag{15}$$

令 $S(t) \in \mathbb{R}^{K\times M}$ 为一个矩阵，其中每个元素由 $S_{i,j}(t)$ 给出，且 $S_{(i)}(t)$ 为 $S(t)$ 的第 i 行（即臂 i 的和向量）。类似地，定义 $C(t) \in \mathbb{N}^{K\times M}, C_{(i)}(t) \in \mathbb{N}^{M}\ \forall i$。评分规则针对每个臂 i 在某一时刻 t 定义为

$$s(i;t) := \begin{cases} \hat{\mu}_i(t) & \text{if } \min_{j\in[M]} \hat{\mu}_{ij}(t) > \tau, \\ \min_{j\in[M]} \hat{\mu}_{ij}(t) & \text{otherwise.} \end{cases} \tag{16}$$

为简洁起见，提及 FCSR 使用的全局属性统计量 S，C，$\hat{\mu}$ 和奖励 $X_{i,\ j}$ 时，我们省略时间步 t。但是，对于仅在子程序范围内使用的任何统计量，我们都会显式标注其索引。此外，回顾臂 $\Delta n_r = n_r - n_{r-1}$（满足 $n_0 = 0$）的每轮预算定义；对于所有 $r > 0$，n_r 定义为 $n_r \leftarrow \lceil \lfloor (1-f)T \rfloor / \bar{n}(K+1-r) \rceil$。归一化常数的显式形式为 $\bar{n} \leftarrow \frac{1}{2} + \sum_{k=2}^{K} \frac{1}{k}$。

算法5 FCSR(T，τ，f，g；S，C)

输入： K 条臂，每条臂有 M 个属性，预算 T，阈值 τ，比例 f，$g \in (0, 1)$ 及矩阵 S，C。**输出：** 最优可行臂或-1。

初始化： 全局统计量 $S_{i,\ j}$，$C_{i,\ j}$，$\hat{\mu}_{i,\ j} \leftarrow 0\ \forall\, i \in [K]$，$j \in [M]$.2: $P_i \leftarrow \lfloor \frac{fT}{K} \rfloor \forall\, i$。 $T_{\text{extra}} \leftarrow 0$.3: **对于** $r = 1$ 至 $K-1$ **执行**4: $n_r \leftarrow \lceil \lfloor (1-f)T \rfloor / \bar{n}(K+1-r) \rceil$，以及 $\Delta n_r \leftarrow n_r - n_{r-1}$（其中 $n_0 = 0$）。 5: $T_{\text{extra}}^{\text{arm}} \leftarrow T_{\text{extra}}/K$； $T_{\text{extra}} \leftarrow 0$。 6: **对于** 每个臂 $i \in \mathcal{A}$ **执行**

▷对每个活跃臂采样

7: 运行 UNIFORM($i, \lfloor (1-g)\ \cdot \Delta n_r \rfloor + T_{\text{extra}}^{\text{arm}}$； S_i, C_i)。8: 运行 APT($i, \lfloor g \cdot \Delta n_r \rfloor, \tau$； S_i, C_i)。9: 运行 SAMPLEUNTILFEASIBLE(i, P_i, τ； S_i, C_i)。10: for 循环结束 11: 根据公式 16 计算 $s(i)\ \forall\, i$，并剔除 $\mathcal{A} \leftarrow \mathcal{A} \setminus \{\arg\min_{i\in\mathcal{A}} s(i)\}$。12: $T_{\text{extra}} \leftarrow T_{\text{extra}} + P_{\arg\min_{i\in\mathcal{A}} s(i)}$。▷重新利用被剔除臂的可行性预算 13: for 循环结束 14: 如果最后剩余的臂 $i_{\text{final}} \in \mathcal{A}$ 可行，则返回 i_{final}.15: 否则返回 0。

算法6 UNIFORM(i, T; S_i, C_i)

要求： 臂 i，预算 T 以及全局统计量 S_i, C_i

对于臂 i 的属性 $j = 1, 2, \ldots M$ **执行**

  对于样本 $s = 1, 2, \ldots \lfloor T/M \rfloor$ **执行** ▷ 对每个属性均匀采样

    观测奖励 $X_{i,j} \sim \nu_{i,j}$

    更新全局 $S_i, C_{i,J_{t+1}}$ 以及 $\hat{\mu}_{i,J_{t+1}}$ 使用公式15。

  结束 for

结束 for

返回 更新后的统计量 C_i, S_i

算法7 APT(i, T; S_i, C_i)

要求： 臂 i、预算 T、阈值 $\tau \in \mathbb{R}$ 和全局统计量 S_i, C_i

初始化： 使用现有的全局 S_i, C_i, $\hat{\mu}_{i,j}\ \forall j \in [M]$。

对 $t = 0, 1, \ldots, T-1$ **执行**

  对每个属性 $j = 1, \ldots, M$ **执行**

    计算经验差距：$\hat{\Delta}_j(t) = |\hat{\mu}_{i,j} - \tau|$

    计算 APT 评分：$B_j(t+1) = \sqrt{C_{i,j} \cdot \dfrac{\hat{\Delta}_{ij}}{(t)}}$

  结束循环

  选择属性 $J_{t+1} = \arg\min_{1 \le j \le M} B_j(t+1)$ 并拉取它。

  观察奖励 $X_{i,J_{t+1}} \sim \nu_{i,J_{t+1}}$。

  更新全局 $S_i, C_{i,J_{t+1}}$ 以及 $\hat{\mu}_{i,J_{t+1}}$ 使用公式15。

结束循环

返回 更新后的统计量 S_i, C_i。

算法8 SAMPLEUNTILFEASIBLE(i, P_i; S_i, C_i, τ)

要求： 臂 i，可行性预算 P_i，全局 $S_i \in \mathbb{R}^M, C_i \in \mathbb{N}^M$，阈值 τ**1: 初始化：** 使用已有的 S_i、C_i、$\hat{\mu}_{i,j}\ \forall j \in [M]$.**2: 当** $P_i \geq 1$**时执行3:** $\mathcal{F}^c$、$_{(i)} \leftarrow \{j : \hat{\mu}_{i,j} \leq \tau\}$ ▷ 不可行属性集合 **4:** 若 $\mathcal{F}^c, _{(i)} = \emptyset$**则中断5:** $j^* \leftarrow$ 最小值$_{j \in \mathcal{F}^c, _{(i)}} j$ ▷选择索引最低的不可行属性 **6: 当** $\hat{\mu}_{i,j^\star} \leq \tau$ **执行** ▷持续采样，直到经验均值超过阈值 **7:** 观测奖励$X_{i,j^\star} \sim \nu_{i,j^\star}$**8:** 更新全局 $S_{i,j^\star}$, $C_{i,j^\star}$ 和 $\hat{\mu}_{i,j^\star}$，使用公式15。 **9:** $P_i \leftarrow P_i - 1$**10:** 若 $P_i = 0$ ▷ 则中断可行性预算耗尽时停止 **11: 结束循环12: 结束循环13: 返回** 更新后的统计量 S_i, C_i。

下面的引理证明了 FCSR 的有效性，即它不会超出给定的预算 T。

引理4（预算合规性）。令 $T \in \mathbb{N}$和 $f, g \in (0, 1)$。在 FCSR 中，设定 $P_i = \lfloor fT/K \rfloor$ 和 $T_{\mathrm{SR}} = T - \sum_{i=1}^{K} P_i$。令$n_0 = 0 < n_1 < \cdots < n_{K-1}$ 定义为

$$n_r = \left\lceil \frac{\lfloor (1-f)T \rfloor}{\bar{n}\,(K+1-r)} \right\rceil, \qquad \Delta n_r = n_r - n_{r-1}.$$

在第 r 轮，每个活跃臂进行 $\lfloor g \Delta n_r \rfloor$ 次 *APT* 抽取和 $\lfloor (1-g)\Delta n_r \rfloor$ 次均匀抽取。因此，算法进行的抽取总次数至多为 T。

证明。 将抽取分为可行性抽取和 SR 抽取。对于可行性抽取：每个臂最多可从其可行性预算中使用 B_i 次抽取

；任何重新分配的可行性抽取均来自同一个池。因此

$$P_{\text{feas}} \le \sum_{i=1}^{K} P_i \le fT$$

SR：第 r 轮有 $K+1-r$ 个活跃臂。对于每个这样的臂，

$$\lfloor g\Delta n_r \rfloor + \lfloor (1-g)\Delta n_r \rfloor \le \Delta n_r,$$

因此，第 r 轮的SR抽取次数至多为（$K+1-r$）Δn_r。对各轮求和可得

$$P_{\text{SR}} \le \sum_{r=1}^{K-1} (K+1-r)\Delta n_r.$$

根据 n_r的构造以及向下取整/向上取整的单调性，将理想的实数值调度 $\tilde{n}_r = \lfloor (\frac{1-f}{\bar{n}} \frac{T}{K+1-r}) \rfloor$ 替换为$n_r = \lceil \tilde{n}_r \rceil$，只会因取整效应使加权和 $\sum_r (K+1-r)\Delta n_r$ 至多减少，因此

$$\sum_{r=1}^{K-1} (K+1-r)\Delta n_r \le \lfloor (1-f)T \rfloor \le T_{\text{SR}}.$$

因此 $P_{\text{SR}} \le T_{\text{SR}}$。总之，综合可得，

$$P_{\text{total}} = P_{\text{feas}} + P_{\text{SR}} \le \sum_{i=1}^{K} P_i + T_{\text{SR}} = T.$$

□

B 术语讨论

在固定预算文献中，如果上界指数中的各项与下界中的各项在常数因子意义下相匹配，通常就称该算法为最优。关于使用“最优”来描述 APT 的情况，参见 Carpentier 和 Locatelli [2016]。下界的推导方法是考虑一类多臂老虎机实例，并对该类实例中任意算法的最坏情况误差给出下界。

Carpentier 和 Locatelli [2016] 还将逐次拒绝称为最优，因为与他们证明的下界相比，其指数仅相差常数因子。本例中的下界同样考虑了一个难度有界的对抗类老虎机实例，这与本文中的设定一致。即使在普通的单维无约束最佳臂识别问题中，也没有已知算法能够与下界完全匹配 [Qin, 2023]。由于我们对下界采用了类似的方法，而 FCSR 在指数上仅差常数因子地达到了该下界，因此 FCSR 是这一设定下的最优算法。

C 下界证明

C.1 证明策略

我们定义两类老虎机实例，$\mathcal{C}_F$“可行性类别”和 $\mathcal{C}_R$“风险类”，以对任意策略 Alg 由于以下原因导致的错误概率给出下界：(i) 将最优可行臂判定为不可行；以及 (ii) 将可行次优臂或均值较高的不可行臂误报为最佳臂。随后我们论证，任意算法 Alg的错误概率下界至少是这些单独类别上的最坏情况错误概率的最大值，即

$$\max_{\mathcal{G} \in \mathcal{C}_F \cup \mathcal{C}_R} \mathbb{P}_{\mathcal{G}}(e) \ge \max \left\{ \max_{\mathcal{G} \in \mathcal{C}_F} \mathbb{P}_{\mathcal{G}}(e), \max_{\mathcal{G} \in \mathcal{C}_R} \mathbb{P}_{\mathcal{G}}(e) \right\}$$

其中错误事件定义为 $e = 1\{I_T \neq i^\star\}$ 尽管我们在构造中假设 $i^\star$ 存在，并对 K 和 M作出进一步假设，但我们指出，下界仍然适用于任意 Alg，因为我们能够找出一个特定问题实例，其中 Alg 表现不佳。

我们使用引理 5 来建立可行性类别上的下界，并基于 Carpentier 和 Locatelli [2016] 以及 Locatelli 等人 [2016] 给出一种新颖的多维构造，以刻画 Alg 在风险类上必须做出的权衡。

C.2 可行性类别

我们首先针对分组臂，给出阈值化多臂老虎机问题（TBP） [Locatelli 等人， 2016] 的下界。主要区别在于，Locatelli 等人 [2016] 中的证明使用的是臂分布为高斯分布的一维多臂老虎机实例，而为了整体证明的需要，我们需要臂分布为伯努利分布的分组老虎机实例。若满足 $M = 1$，并在预算 T的取值范围上附加一个条件，我们的证明即可恢复 Locatelli 等人 [2016] 中的结果。为此，我们改写了 Carpentier 和 Locatelli [2016]中针对标准最佳臂识别多臂老虎机问题的紧下界证明。

C.2.1 多维 TBP。

该设定与分组多臂老虎机问题非常相似。给定 K 个臂，每个臂有 M 个属性；在每一轮 $t > 0$开始时，学习器选择对一个属性进行采样（k_t， m_t），其中 $k \in [K]$， $m \in [M]$，并观察到一个随机奖励 $X(t) \sim \nu_{k,\ m_\circ}$ 经过 T轮后，学习器必须输出一个臂集合 $\widehat{S}_\tau \subseteq [K]_\circ$ 令 $\widehat{S}_\tau^C = [K]/\widehat{S}_{\tau\circ}$ 与之前一样，令$\mu_{k,\ m}$ 表示（k， m）的属性均值， μ_k 表示臂 k的臂总体均值。

对于某个 $\tau \in \mathbb{R}$，将可行臂集合定义为 $\mathcal{S}_\tau := \{k\colon\ \mu_k > \tau\}_\circ$ 不可行臂集合记为$S_\tau^C := [K]/S_{\tau\circ}$ 学习器对任意选择的可行性进行错误分类的事件为

$$\mathcal{L}(T) = 1\{\widehat{S}_\tau \cap S_\tau^C \neq \emptyset \ \vee\ \widehat{S}_\tau^C \cap S_\tau \neq \emptyset\}.$$

学习器旨在最小化 $\mathbb{E}[\mathcal{L}(T)]_\circ$ 这是阈值化多臂老虎机问题的一个特殊情形： $\epsilon = 0$，且均值为 $\mu_k = \tau$ 的臂被归类为不可行。

定义可行性类别 $\mathcal{C}_F(d;\ K) := \{\mathcal{G}^k\colon\ k \in \{0\} \cup [K]\}$，这是由间隙常数 $d \in (0, \frac{1}{4}]$ 和 $K = M$（臂数及属性数）参数化的一组 $K+1$ 老虎机问题。令所有这些问题实例均满足 $\tau = 0.5$，并令 $d \in (0, \frac{1}{4}]$ 为一个用于参数化各实例难度的常数。令 $\nu' = \mathcal{B}(\frac{1}{2} - d)$ 和 $\nu = \mathcal{B}(\frac{1}{2} + d)$，其中 $\mathcal{B}(p)$ 表示均值参数为 p 的伯努利分布。对于 $0 \leq k \leq K$，$\mathcal{G}^k$ 是一个问题实例，其中臂 $k \in [K]$ 的分布为 $\nu_{k,\ m} = \nu\ \forall\ m \in [M]_\circ$ 对于所有臂 $j \neq k$，臂分布均为 $\nu_{j,\ m} = \nu'\ \forall\ m \in [M]_\circ$

引理 5 (伯努利 TBP 下界)。 对于任意 $K = M \geq 2\ d \in (0, \frac{1}{4}]$, 若 $T \geq 4(\log(K)H_f)^2\log(6TK)/(60)^2$,，对于任意老虎机算法，均有

$$\max_{\mathcal{G}^i \in \mathcal{C}_F(d;K)} \mathbb{E}_{\mathcal{G}^i}[\mathcal{L}(T)] \geq \frac{1}{6}\exp\left(\frac{-120T}{\log(K)H_f}\right) \tag{17}$$

其中， $\mathbb{E}_{\mathcal{G}^i}$ 是根据问题 $\mathcal{G}^i$ 的样本计算的期望。

证明。 我们采用 Carpentier 和 Locatelli [2016] 的证明策略，并做了少量调整。考虑这样一个事件：在某些适当条件下，某个任意学习器 Alg 报告不存在不可行的选项/属性，即在多臂老虎机实例 $\mathcal{G}^0$ 中 $\widehat{S}_\tau^C = \emptyset_\circ$ 然后，我们通过测度变换论证，给出 Alg 错误地未改变其输出的概率下界，即使在实例 $\mathcal{G}^i$ 中， i 个选项是按构造不可行的。

由于对于 $\mathcal{G}^k$，臂 k 可行，且所有其他臂都不可行，因此 $k = i^\star_\circ$ 于是， $\Delta_{i^\star j}^{-2} = d^{-2}\ \forall\ j \in [M]_\circ$ 我们注意到

$$Md^{-2} = \log K\frac{K}{\log K}\max_j \bar{\Delta}_{i^\star j}^{-2} = \log(K)H_f. \tag{18}$$

步骤 1：经验 KL 散度集中的高概率事件 对于定义在 $\mathbb{R}$ 上的两个分布 ν, ν'，且满足 ν 关于 ν' 绝对连续的情况，我们用 $\mathrm{KL}(\nu, \nu') = \int_{\mathbb{R}} \log\left(\frac{d\nu}{d\nu'(x)}(x)\right) d\nu(x)$, 表示分布 ν 与 ν' 之间的 Kullback-Leibler 散度。令 $k \in \{1, ..., K\}_\circ$ 记

$$\mathrm{KL} := \mathrm{KL}(\nu', \nu) = \mathrm{KL}(\nu, \nu') = (1-2p)\log\big(\frac{1-p}{p}\big),$$

两个参数分别为 $p = 0.5 + d$ 和 $1 - p = 0.5 - d$ 的伯努利分布 ν 和 ν' 之间的 Kullback-Leibler 散度。由于 $p \in [1/4, 1/2)$，以下不等式成立：

$$\mathrm{KL} \leq 10d^2. \tag{19}$$

令 $1 \le t \le T$。我们将经验臂 KL 散度定义为：

$$\widehat{\mathrm{KL}}_k(t) = \frac{1}{t}\sum_{s=1}^{t}\sum_{m=1}^{M} \mathbf{1}\{X_{k,m}(s)=1\}\log(\frac{p}{1-p}) + \mathbf{1}\{X_{k,m}(s)=0\}\log(\frac{1-p}{p}),$$

其中，根据定义，对任意 $s \le t$, $X_{k,\,m}(s) \sim_{i.i.d} \nu^i_{k,\,m}$。定义臂 KL 散度具有集中性的事件如下：

$$\xi = \Big\{\forall 1 \le k \le K, \forall 1 \le t \le T, |\widehat{\mathrm{KL}}_k(t)| - \mathrm{KL}_k \le 2\sqrt{\frac{\log(6TK)}{t}}\Big\}.$$

现在我们陈述以下　　断言；这是一个适用于所有满足 $0 \le i \le K$ 的老虎机问题 $\mathcal{G}^i$ 的 $|\widehat{\ }\mathrm{KL}_{k,\,t}|$ 集中性界　　.

断言。 </style}有 $\mathbb{P}_{\mathcal{G}^i}(\xi) \ge 5/6.$

我们验证上述断言。若 $k \ne i$（因此 $\nu^i_{k,\,m} = \nu\ \forall\, m$），则 $\mathbb{E}_{\mathcal{G}^k}[\widehat{\ }\mathrm{KL}_k(t)] = \mathrm{KL}$。若 $k = i$（因此 $\nu^i_{k,\,m} = \nu'$），则 $\backsim \mathbb{E}_{\mathcal{G}^i}[\mathrm{KL}_k(t)] = -\mathrm{KL}$（因为臂参数在 $k = i$ 中互换了）。此外注意，由于 $p \in [1/4, 1/2)$

$$|\log(\frac{d\nu_{k,m}}{d\nu'_{k,m}}(X_{k,m}(s)))| = |\mathbf{1}\{X_{k,m}(s)=1\}\log(\frac{p}{1-p}) + \mathbf{1}\{X_{k,m}(s)=0\}\log(\frac{1-p}{p})| \le \log(3).$$

$\widehat{\ }$因此，$\mathrm{KL}_k(t)$ 是一组独立同分布样本的和，这些样本均以 $\log(3)$ 为界，其均值为 $\pm\mathrm{KL}$，具体取决于i的取值。我们可以对该量应用 Hoeffding 不等式，并且以大于 $1-(6KT)^{-1}$ 的概率有

$$|\widehat{\mathrm{KL}}_k(t)| - \mathrm{KL}_k \le \sqrt{2}\log(3)\sqrt{\frac{\log(6TK)}{t}}.$$

此断言以及对所有 $1 \le k \le K$ 和 $1 \le t \le T$取并集界，意味着 $\mathbb{P}_{\mathcal{G}^i}(\xi) \ge 5/6$，因为我们有 $\sqrt{2}\log(3) < 2$，因此该断言得以验证。

$\widehat{\ }$**步骤 2：测度变换** 任取一个策略 $\mathcal{A}lg$，使其在预算 T结束时返回 S_τ。令 $T_{k,m}$ 表示 $\mathcal{A}lg$ 在每个选择（k, m）上收集的样本数，并令 $T_k = \sum_m T_{k,m}$ 表示针对臂 k收集的样本数。根据固定预算设定的定义，这些量是随机的，但有 $\sum_{k,m} T_{k,m} = \sum_k T_k = T$。对于任意 $t_{k,m} = \mathbb{E}_{\mathcal{G}^0}[T_{k,m}].$，记为 $t_k = \mathbb{E}_{\mathcal{G}^0}[T_k].$。此外，有 $\sum_{k,m} t_{k,m} = \sum_k t_k = T$。

我们回顾测度变换恒等式（例如参见 Audibert 等人 [2010]），该恒等式指出，对于任意可测事件 $\mathcal{E}$ 以及任意 $1 \le i \le K$：

$$\mathbb{P}_{\mathcal{G}^i}(\mathcal{E}) = \mathbb{E}_{\mathcal{G}^0}\Big[\mathbf{1}\{\mathcal{E}\}\exp\big(-T_i\widehat{\mathrm{KL}}_{i,T_i}\big)\Big], \tag{20}$$

由于乘积分布 $\mathcal{G}^i$ 和 $\mathcal{G}^0$ 在臂 i 的所有属性 $j \in [M]$ 上均不同，$T_i\,\widehat{\ }\mathrm{KL}_{i,\,T_i}$ 即为在这两个不同的老虎机问题中，臂 i 下观测到的所有样本的对数似然比。令 $1 \le i \le K$。现在考虑事件

$$\mathcal{E}_i = \{\widehat{S}_\tau = \emptyset\} \cap \{\xi\} \cap \{T_i \le 6t_i\},$$

即算法在结束时报告所有选择均可行的事件，其中 ξ 成立，并且选择 i 被拉取的次数小于 $6t_i$。根据公式 (20)，我们有

$$\begin{aligned}
\mathbb{P}_{\mathcal{G}^i}(\mathcal{E}_i) &= \mathbb{E}_{\mathcal{G}^0}\Big[\mathbf{1}\{\mathcal{E}_i\}\exp\big(-T_i\widehat{\mathrm{KL}}_{i,T_i}\big)\Big] \\
&\ge \mathbb{E}_{\mathcal{G}^0}\Big[\mathbf{1}\{\mathcal{E}_i\}\exp\Big(-T_i\mathrm{KL} - 2\sqrt{T_i\log(6TK)}\Big)\Big] \\
&\ge \mathbb{E}_{\mathcal{G}^0}\Big[\mathbf{1}\{\mathcal{E}_i\}\exp\Big(-6t_i\mathrm{KL} - 2\sqrt{T\log(6TK)}\Big)\Big] \\
&\ge \exp\Big(-6t_i\mathrm{KL} - 2\sqrt{T\log(6TK)}\Big)\mathbb{P}_{\mathcal{G}^0}(\mathcal{E}_i),
\end{aligned} \tag{21}$$

$\widehat{\ }$由于在 $\mathcal{E}_i$ 上，我们有 ξ 成立，且 $T_i \le 6t_i$；又因为对于任意 $t \le T$，$\mathbb{E}_1\mathrm{KL}_{i,\,t} = \mathrm{KL}_i$。

步骤 3：下界 o n $\mathbb{P}_{\mathcal{G}^0}(\mathcal{E}_i)$ 对于任意合理算法 假设对于我们考虑的算法 $\mathcal{A}lg$，er

$$\mathbb{E}_{\mathcal{G}^0}[\widehat{S}_\tau = \emptyset] \leq 1/2, \tag{22}$$

也就是说，$\mathcal{A}lg$ 在问题 $\mathcal{G}^0$ 上出错的概率小于 $1/2$。注意，如果 $\mathcal{A}lg$ 不满足这一条件，它在问题 $\mathcal{G}^0$ 上的表现就会很差，并且在我们定义的 $K+1$ 老虎机问题上，其成功概率始终不超过 $1/2$。

根据马尔可夫不等式和 $\mathbb{E}_{\mathcal{G}^0}[T_k] = t_k$，对于任意 $1 \leq k \leq K$，有：

$$\mathbb{P}_{\mathcal{G}^0}(T_k \geq 6t_k) \leq \frac{\mathbb{E}_{\mathcal{G}^0}[T_k]}{6t_k} = 1/6. \tag{23}$$

因此，结合公式 (22)、(23) 和命题 C.2.1，根据并集界，对于任意 $2 \leq i \leq K$，有

$$\mathbb{P}_{\mathcal{G}^0}(\mathcal{E}_i) \geq 1 - (1/6 + 1/2 + 1/6) = 1/6.$$

将这一事实与公式 (21) 以及对任意 $1 \leq i \leq K$、$\mathbb{P}_{\mathcal{G}^i}$，$(\widehat{S}_\tau = \emptyset) \geq \mathbb{P}_{\mathcal{G}^i}(\mathcal{E}_i)$ 这一事实结合起来，可得对任意$1 \leq i \leq K$

$$\begin{aligned}\mathbb{P}_{\mathcal{G}^i}(\widehat{S}_\tau = \emptyset) &\geq \frac{1}{6}\exp\Big(-6t_i\mathrm{KL}_i - 2\sqrt{T\log(6TK)}\Big) \\ &\geq \frac{1}{6}\exp\Big(-60t_i d_i^2 - 2\sqrt{T\log(6TK)}\Big),\end{aligned} \tag{24}$$

最后一步用到了公式 (19)。

步骤 4：结论。 由于 $\sum_{1\leq k\leq K} d^{-2} = \log(K)H_f$，并且由于 $\sum_{1\leq k\leq K} t_k = T$，因此存在 $1 \leq i \leq K$ 使得

$$t_i \leq \frac{T}{\log(K)H_f d^2},$$

因为逆否命题会立即得出矛盾。对于该 i，由等式 (24) 可知

$$\mathbb{P}_{\mathcal{G}^i}(\widehat{S}_\tau \neq \emptyset) \geq \frac{1}{6}\exp\Big(-60\frac{T}{\log(K)H_f} - 2\sqrt{T\log(6TK)}\Big).$$

注意，$\forall\, k \in [K]$ 我们有 $H_{tbp}(\mathcal{G}^0) = H_{tbp}(\mathcal{G}^k) = \sum_k d_k^{-2}$。由于对任意问题实例 $\mathcal{G}^i$，$i \in [K]$，事件 $\{\widehat{S}_\tau = \emptyset\} \subseteq \mathcal{L}(T)$ 均成立，并且根据假设 $T \geq 4(\log(K)H_f)^2\log(6TK)/(60)^2$，我们可以吸收第二项并得到所需结果

$$\max_{i\in\{1,\ldots,K\}} \mathbb{E}_{\mathcal{G}^i}(\mathcal{L}(T)) \geq \mathbb{P}_{\mathcal{G}^i}(\widehat{S}_\tau \neq \emptyset) \geq \frac{1}{6}\exp\Big(\frac{-120T}{\log(K)H_f}\Big).$$

□

C.3 风险类别

C.4 老虎机问题的定义

设 $\tau = 3/8, K,\ M \geq 2$。设臂1的属性分布分别记为 $\nu_{1j}^R = \mathcal{B}(\tau + d_R)$ 和 $\nu_{1j}^{R'} = \mathcal{B}(\tau - d_R)$，其中所有 $j \in [M]$ 和 $d_R = 1/8$ 均满足 $\tau + d_R = 1/2$。设臂1的臂分布为 $\nu_1^R = \bigotimes_{j\in[M]} \nu_{1j}^R$。设$\beta \in (0,\ 1)$为一个常数，用于刻画这类问题的难度。对于 $2 \leq i \leq K$，定义分布 $\nu_{ij}^R := \mathcal{B}(\frac{1}{2} - d_i)$ 和 $\nu_{ij}^{R'} := \mathcal{B}(\frac{1}{2} + d_i)$，其中

$$d_i = \frac{\beta i}{16K}\sqrt{\frac{K-1}{MK}}.$$

设臂 $2 \le i \le K$ 的臂分布定义为 $\nu_i^R = \bigotimes_{j\in[M]} \nu_{ij}^R$。现在定义 $K+M$ 个多臂老虎机实例。定义"基准"多臂老虎机实例，其中臂 $1 = i^\star$ 可行且均值最高：

$$\mathcal{G}_R^1 = \bigotimes_{i\in[K]} \nu_i^R.$$

注意，在这种情况下，臂1具有最高的均值 $\mu_1 = 1/2$，所有其他臂 $\mu_k = 1/2 - d_i < 1/2$。此外，所有其他臂都是可行的，因为 $MK \ge K-1 \implies d_K \le 1/16 \implies \mu_K \ge 7/16 > \tau$。
前 $K-1$ 个问题实例的定义方式与 Carpentier 和 Locatelli [2016] 类似，其中问题实例 j 中的 $i^\star = j$。这是因为 $\mu_j\left(\mathcal{G}_R^j\right) = 1/2 + d_i > \mu_1$。形式上，对 $2 \le j \le K$ 令并定义

$$\mathcal{G}_R^j = \bigotimes_{i\in[K]/\{j\}} \nu_i^R \bigotimes_j \nu_j^{R'}.$$

接下来的 M 个问题实例在臂1的某个属性上不可行，因此 $i^\star = 2$ 和 $1 \in \mathcal{F}^C$。对于 $K+1 \le j \le M+K+1$（即 $j-K \in [M]$），定义

$$\mathcal{G}_R^j = \bigotimes_{i\in[K]-\{1\}} \nu_i^R \bigotimes_{i\in[M]-\{j-K\}} \nu_{1i}^R \bigotimes_{j-K} \nu_{1,j-K}^{R'}.$$

定义风险类为 $\mathcal{C}_\mathcal{R}(\beta; K, M) =: \{\mathcal{G}_R^i\colon 1 \le i \le K+M+1\}$ 其中，β 是控制每个多臂老虎机问题 $\mathcal{G}_R^i$ 中间隙的参数，而 K, M 则影响臂和属性的数量。与仅识别均值最高的臂相比，这类多臂老虎机实例更加困难，因为学习器不仅必须识别最佳臂，还必须学会区分最佳臂是否可行。

C.5 风险类下界证明

引理 6 (风险类下界)。对于任意 $K, M \ge 2$，若 $T \ge 4(\log(K)\max\{H_{tbp}(\mathcal{G}_R^i)\ H_2^R(\mathcal{G}_R^i)\})^2 \log(6T(K+M+1))/60^2$，则对于任意老虎机算法，我们有以下下界

$$\max_{\mathcal{G}_R^i \in \mathcal{C}_R(\beta;K,M)} \mathbb{P}_{\mathcal{G}_R^i}(e) \ge \frac{1}{6}\exp\left(\frac{-1200T}{\log(K)\max\{H_2^R(\mathcal{G}_R^i), H_{tbp}(\mathcal{G}_R^i)\}}\right). \tag{25}$$

证明。与引理 5 中一样，我们将 Carpentier 和 Locatelli [2016]的论证改编到我们新构造的多臂老虎机问题类别中。此外，我们证明，尽管学习器事先知道最佳臂可能不可行（因为在我们的构造中只有臂1不可行），但 FCSR 给出的界在指数上与常数因子相匹配。这之所以可行，是因为我们精心构造了风险类 $\mathcal{C}_R$，也得到了难度参数前乘的那个难以获得的 $\log(K)$ 因子。下界中缺失的 $\log(K)$ 因子，正是标准最佳臂识别问题中上界与下界之间存在差距的原因，直到 Carpentier 和 Locatelli [2016]解决了这一问题。

步骤 1：经验 KL 散度集中的高概率事件令 $\nu_{i,\ j}^w\ \forall\ w \in [K+M+1]$ 表示问题实例 $\mathcal{G}_R^w$ 中属性 $(i,\ j)$ 的分布。定义问题 $w \in [K]$ 相对于问题 1 的经验 KL 散度为

$$\widehat{\mathrm{KL}}_w(t) = \begin{cases} \frac{1}{t}\sum_{s=1}^t \sum_{j=1}^M \log\frac{d\nu_{i,j}^1}{d\nu_{i,j}^w}(X(t)) & : w \in [K], i = w \\ \frac{1}{t}\sum_{s=1}^t \log\frac{d\nu_{1,j}^1}{d\nu_{1,j}^w}(X(t)) & : w \in [K+1 : K+M+1], j = w-K \end{cases}$$

其中 $X(t) \sim_{\text{i.i.d}} \nu_{i,\ j}^1$。令 $\mathrm{KL}_w = \mathbb{E}_{\mathcal{G}_R^1}[\widehat{\mathrm{KL}}_w(t)]$。问题 w 的经验 KL 散度：$\widehat{\mathrm{KL}}_w(t)$ 给出了在基准实例 $\mathcal{G}_R^1$ 下观测到的所有样本相对于问题实例 $\mathcal{G}_R^w$ 的对数似然比。上述定义仅包含由扰动臂/属性产生的非零项。

定义 $d_w = d_i$，其中 $w,\ i \in [K]$ 以及 $d_w = d_R = 1/8$，其中 $w \in [K+1 :\ K+M+1]$。对于 $w \in [K]$，扰动后的臂分布为伯努利分布，其参数以 $1/2$ 为中心对称，因此 $\mathrm{KL}_w = \mathrm{KL}(0.5-d_w,\ 0.5+d_w) =$

KL$(0.5+d_w,\ 0.5-d_w)$。在这种情况下，对所有 $i,\ j$，以下不等式均成立（Carpentier 和 Locatelli 的公式 (2) [2016]）

$$\mathrm{KL}_w \le 10 d_w^2.$$

对于 $w \in [K+1:\ K+M+1]$，我们有如下不等式

$$\mathrm{KL}_w = \mathrm{KL}(0.5, 0.25) \le 10 d_R^2 = 10 d_w^2.$$

定义事件

$$\xi' = \left\{ \forall\, w \in [K+M+1],\, \forall\, 1 \le t \le T,\ \left| \widehat{\mathrm{KL}}_w(t) - \mathrm{KL}_w \right| \le 2\sqrt{\frac{\log(6T(K+M+1))}{t}} \right\}. \tag{26}$$

断言。有 $\mathbb{P}_{\mathcal{G}_R^1}(\xi') \ge 5/6.$

我们验证上述断言。对于 $w \in [K]$，臂 w 的均值参数在 0。5 附近扰动。更明确地说，$0.5 - d_w = \mu_{k,\ m}$ 对于 $k = w$ 以及所有满足 $\mathcal{G}_R^1$ 和 $0.5 + d_w = 1 - \mu_{k,\ m}$ 的 m，对于 $k = w$ 以及所有满足 $\mathcal{G}_R^w$ 的 m。$\mu_w \in [1/4, 1/2)$ 因此，

$$|\log(\frac{d\nu_{k,m}^1}{d\nu_{k,m}^w}(X(s)))| = |\mathbf{1}\{X(s)=1\}\log(\frac{\mu_{k,m}}{1-\mu_{k,m}}) + \mathbf{1}\{X(s)=0\}\log(\frac{1-\mu_{k,m}}{\mu_{k,m}})| \le \log(3),$$

其中 $X(s) \sim \nu_{k,\ m}^1$。现在对于 $w \in [K+M+1]$，只有臂1的属性 $j = w - K$ 受到扰动，因此我们有

$$|\log(\frac{d\nu_{1,j}^1}{d\nu_{1,j}^w}(X(s)))| = |\mathbf{1}\{X(s)=1\}\log(\frac{0.5}{0.25}) + \mathbf{1}\{X(s)=0\}\log(\frac{0.5}{0.75})| \le \log(3).$$

因此，对于任意 w，$\mathrm{KL}_w(t)$ 是一组独立同分布样本之和，每个样本都以 log(3) 为界，其均值为 KL_w。我们可以对所有 $K,\ M,\ t$ 应用 Hoeffding 不等式，并且以大于 $1 - (6T(K+M+1))^{-1}$ 的概率，有

$$|\widehat{\mathrm{KL}}_w(t) - \mathrm{KL}_w| \le \sqrt{2}\log(3)\sqrt{\frac{\log(6T(K+M+1))}{t}}.$$

该断言与对所有 $1 \le w \le K+M+1$ 和 $1 \le t \le T$ 取并集界可推出 $\mathbb{P}_{\mathcal{G}_R^1}(\xi) \ge 5/6$，因为我们有$\sqrt{2}\log(3) < 2$，因此该断言得证。

步骤 2：测度变换 设任意策略 $\mathcal{A}lg$ 在预算 T用完时返回臂 I_T。令 $T_{k,m}$表示 $\mathcal{A}lg$ 在每种选择 (k,m) 上收集的样本数，$T_k = \sum_m T_{k,m}$ 表示为臂 k收集的样本数。令广义索引 w 在 $2 \le w \le K$时指代臂 $i \in [K]$，在 $K+1 \le w \le K+M+1$时指代属性 $(1, w-K)$。因此，w 对扰动臂/属性集合进行索引。定义 T_w 为

$$T_w = \begin{cases} \sum_{j=1}^M T_{w,j} & : w \in [K], \\ T_{1,w-K} & : w \in [K : K+M+1]. \end{cases}$$

根据固定预算设定的定义，有 $\sum_w T_w = \sum_{k=2}^K T_k + \sum_{m=1}^M T_{1,m} = \sum_{k,m} T_{k,m} = T$。对任意 $t_{k,m} = \mathbb{E}_{\mathcal{G}_R^1}[T_{k,m}]$，令 $t_k = \mathbb{E}_{\mathcal{G}_R^1}[T_k]$ 和 $t_w = \mathbb{E}_{\mathcal{G}_R^1}[T_w]$。因此，也有 $\sum_w t_w = T$。

我们回顾一下测度变换恒等式（例如参见 Audibert 等人 [2010]），该恒等式指出，对于任意可测事件 $\mathcal{E}$ 以及任意 $1 \le w \le K+M+1$：

$$\mathbb{P}_{\mathcal{G}_R^w}(\mathcal{E}) = \mathbb{E}_{\mathcal{G}_R^1}\Big[\mathbf{1}\{\mathcal{E}\}\exp\big(-T_w\widehat{\mathrm{KL}}_w(T_w)\big)\Big], \tag{27}$$

因为乘积分布 $\mathcal{G}_R^w$ 和 $\mathcal{G}_R^1$ 在问题实例 w中所有受到扰动的属性上都不同。$T_w\ \widehat{\mathrm{KL}}_w(T_w)$ 充当对数似然比。令 $1 \le w \le K+M+1$。现在考虑事件

$$\mathcal{E}_w = \{I_T = 1\} \cap \{\xi'\} \cap \{T_w \le 6t_w\},$$

即算法报告臂1为最佳臂、 ξ' 成立，且在 w 中拉取受扰动属性集合的次数小于 $6t_w$的事件。根据公式（27），我们有

$$\begin{aligned}\mathbb{P}_{\mathcal{G}_R^w}(\mathcal{E}_w) &= \mathbb{E}_{\mathcal{G}_R^1}\Big[\mathbf{1}\{\mathcal{E}_w\}\exp\big(-T_w\widehat{\mathrm{KL}}_w(T_w)\big)\Big] \\ &\geq \mathbb{E}_{\mathcal{G}_R^1}\Big[\mathbf{1}\{\mathcal{E}_w\}\exp\Big(-T_w\mathrm{KL}_w-2\sqrt{T_i\log(6T(K+M+1)}\Big)\Big] \\ &\geq \mathbb{E}_{\mathcal{G}_R^1}\Big[\mathbf{1}\{\mathcal{E}_w\}\exp\Big(-6t_w\mathrm{KL}_w-2\sqrt{T\log(6T(K+M+1))}\Big)\Big] \\ &\geq \exp\Big(-6t_w\mathrm{KL}_w-2\sqrt{T\log(6T(K+M+1))}\Big)\mathbb{P}_{\mathcal{G}_R^1}(\mathcal{E}_w),\end{aligned} \tag{28}$$

由于 $\widehat{\ }$在 $\mathcal{E}_w$上，我们有 ξ'成立，且有 $T_w \leq 6t_w$；又因为对于任意 $t \leq T$，$\mathbb{E}_{\mathcal{G}_R^1}\mathrm{KL}_w(t) = \mathrm{KL}_w$。

步骤 3：关于 的下界 $\mathbb{P}_{\mathcal{G}_R^1}(\mathcal{E}_w)$ **对于任意合理算法，**假设对于我们考虑的算法 $\mathcal{A}lg$

$$\mathbb{P}_{\mathcal{G}_R^1}(I_T \neq 1) \leq 1/2, \tag{29}$$

也就是说，算法 $\mathcal{A}lg$ 在问题 $\mathcal{G}_R^1$ 上出错的概率小于 $1/2$。注意，如果 $\mathcal{A}lg$ 不满足这一点，那么它在问题 $\mathcal{G}_R^1$ 上的表现就很差，并且在我们定义的 $K+M+1$ 个老虎机问题上，其成功概率不会大于 $1/2$。

对于任意 $1 \leq w \leq K+M+1$，由马尔可夫不等式和 $\mathbb{E}_{\mathcal{G}_R^1}[T_w] = t_w$ 可得，

$$\mathbb{P}_{\mathcal{G}_R^1}(T_k \geq 6t_k) \leq \frac{\mathbb{E}_{\mathcal{G}_R^1}[T_w]}{6t_w} = 1/6. \tag{30}$$

因此，结合公式 （29）、（30）和断言 C.5，由并集界可知，对于任意 $2 \leq w \leq K+M+1$

$$\mathbb{P}_{\mathcal{G}_R^1}(\mathcal{E}_i) \geq 1-(1/6+1/2+1/6) = 1/6.$$

将这一事实与公式（28）以及对于任意 $1 \leq w \leq K+M+1, \mathbb{P}(e) \geq \mathbb{P}_{\mathcal{G}_R^w}(I_T=1) \geq \mathbb{P}_{\mathcal{G}_R^w}(\mathcal{E}_i)$ 成立这一事实结合起来，可知对于任意 $1 \leq w \leq K+M+1$

$$\begin{aligned}\mathbb{P}_{\mathcal{G}_R^w}(e) &\geq \frac{1}{6}\exp\Big(-6t_w\mathrm{KL}_w-2\sqrt{T\log(6T(K+M+1))}\Big) \\ &\geq \frac{1}{6}\exp\Big(-60t_wd_w^2-2\sqrt{T\log(6T(K+M+1))}\Big).\end{aligned} \tag{31}$$

步骤 4：结论。我们首先给出构造中阈值化难度的上界。记 $H_{tbp}^{(i)} := H_{tbp}(\mathcal{G}_R^i)$。由于在我们的构造中，唯一不可行的臂是臂 1，因此有

$$H_{tbp}^{(i)} = K\max_{i\in\mathcal{F}^C}\sum_j\bar{\Delta}_{ij}^{-2} \leq K\cdot Md_R^{-2} = 64MK, \tag{32}$$

其中，在 H_{tbp} 的定义中，臂按其真实均值编索引。上述不等式是通过选取最大索引 K 以及潜在的唯一不可行臂对应的 $\sum_j\bar{\Delta}_{ij}^{-2} = Md_R^{-2}$ 得到的。定义经典难度 $H^{(i)} = H(\mathcal{G}_R^i) := \sum_{i\neq 1}\Delta_i^{-2}$。

我们可以通过考察最简单的实例 $\min_{2\leq i\leq K+M+1}H^{(i)}$，对任意老虎机问题 $\mathcal{G}_R^i$ 的经典难度参数给出下界。这个实例 K 中，臂 K 被翻转。注意，对于任意其他臂 $i\neq K$，此处 $\Delta_i \geq (2d_K)^{-2}$。因此，对于任意 $i\in[K+M+1]$

$$(K-1)(2d_K)^{-2} \leq \min_{2\leq i\leq K+M+1}H^{(i)} \leq H^{(i)}. \tag{33}$$

由 $\beta\in(0,1)$ 和 d_K 的定义

$$d_K = \frac{\beta}{16}\sqrt{\frac{K-1}{MK}} \implies (K-1)(2d_K)^{-2} = (K-1)\frac{256MK}{4\beta^2(K-1)} \geq 64MK.$$

结合上述结果与 (32) 和 (33)，对于任意 i，我们有

$$H^{(i)} \geq H_{tbp}^{(i)}. \tag{34}$$

现在定义 $H_R(i) := H_{tbp}^{(i)} + H^{(i)}$，它是一个组合难度参数，可对所需的 H_{FC} 给出上界，并令 $h^* = \sum_{2 \leq i \leq K+M+1} \frac{1}{d_i^2 H_R(i)}$。存在 $2 \leq i \leq K+M+1$ 使得

$$t_i \leq \frac{T}{h^* d^2 H_R(i)}.$$

将其代入 (31)，我们得到：

$$\mathbb{P}_{\mathcal{G}_R^i}(e) \geq \mathbb{P}_{\mathcal{G}_R^i}(I_T = 1) \geq \exp(-\frac{60T}{h^* H_R(i)} - 2\sqrt{T \log(6T(K+M+1))}). \tag{35}$$

现在我们给出 $h^\star$的下界。对于任意 $1 \leq i \leq K$，有

$$d_i^2 H(i) = d_i^2 \sum_{k \neq i} \frac{1}{(d_i + d_k)^2} \leq d_i^2 \Big(\frac{i}{d_i^2} + \sum_{k > i} \frac{1}{d_k^2} \Big) \leq i + i^2 \sum_{K \geq k \geq i} \frac{1}{k^2} \leq i + i^2 (\frac{1}{i} - \frac{1}{K}) \leq 2i.$$

这意味着

$$d_i^2 H_R(i) = d_i^2 H_{tbp}(i) + d_i^2 H(i) \leq 2 d_i^2 H(i) \leq 4i. \tag{36}$$

因此，利用积分对黎曼和进行下界估计，我们得到 $h^\star$的如下下界：

$$\begin{aligned} h^* &\geq \sum_{2 \leq i \leq K+M+1} \frac{1}{d_i^2 H_R(i)} \geq \sum_{2 \leq i \leq K} \frac{1}{d_i^2 H_R(i)} \\ &\geq \frac{1}{4} \sum_{i=2}^{K} \frac{1}{i} \geq \frac{1}{4} (\log(K+1) - \log(2)) \geq \frac{1}{10} \log(K), \end{aligned}$$

其中，最后一个不等式可通过经验验证。对于所有问题实例，风险臂集合 $\mathcal{R} = \emptyset$，因此$H_2(\mathcal{G}_R^i) = \max_{i \neq i^\star} i\Delta_i^2 \leq H^{(i)}$。（这是固定预算最佳臂识别文献中的一个众所周知的结果；参见 Carpentier 和 Locatelli [2016]）。因此，$H_R(i) \geq \max\{H_{tbp}(\mathcal{G}_R^i), H_2(\mathcal{G}_R^i)\}$。于是，将我们对 $h^\star$ 的下界代入 (35)，可得存在某个多臂老虎机实例 $\mathcal{G}_R^i$，它属于 $\mathcal{C}_\mathcal{R}$，使得

$$\mathbb{P}_{\mathcal{G}_R^i}(e) \geq \exp\left(\frac{-600T}{\log(K) \max\{H_{tbp}(\mathcal{G}_R^i), H_2^R(\mathcal{G}_R^i)\}} - 2\sqrt{T \log(6T(K+M+1))} \right). \tag{37}$$

选取 $T \geq 4(\max\{H_{tbp}(\mathcal{G}_R^i), H_2^R(\mathcal{G}_R^i)\})^2 \log(6T(K+M+1))/600^2$，第二项可由第一项给出下界，从而得到所需结果。

注： 由 (33) 可得 $H^{(i)} \geq \max\{H_{tbp}(\mathcal{G}_R^i), H_2^R(\mathcal{G}_R^i)\}$。可以验证，基准实例具有最高的经典难度 $H^{(i)}$，因此对所有 i，我们有

$$\max\{H_{tbp}(\mathcal{G}_R^i), H_2^R(\mathcal{G}_R^i)\} \leq H^{(1)} \leq \sum_{i=2}^{K} (\frac{\beta i}{16K})^2 \frac{K-1}{MK} \leq \frac{\beta^2 K}{256M}. \tag{38}$$

□

C.6 完整下界

定理 7（下界）。令 $\mathcal{C}_{FC}(a;\ K)$ 表示具有 $K = M \geq 2$ 且其难度 H_{FC}的上界为某个常数 a的老虎机实例集合。若

$$T \geq \frac{4}{60^2} (a \log(K))^2 \log(6T(K+M+1)) \tag{39}$$

存在一个多臂老虎机实例 $\mathcal{G} \in \mathcal{C}_{FC}(a)$，使得任何学习器错误报告最佳臂的概率至少为

$$\mathbb{P}_{\mathcal{G} \in \mathcal{C}_{FC}(a)} \geq \frac{1}{6} \exp\left(\frac{-1200T}{\log(K) H_{FC}(\mathcal{G})} \right).$$

证明。取 $K = M \geq 2$ 和 $a = \max\{\beta^2/256,\ {}^{M}\!/_{\log(K)}\, d^{-2}\}$。由引理 6 中的式（38）可得 th at

$$\max\{H_2^R, H_{tbp}\} \leq \beta^2/256.$$

此外，由引理 5 中的式 (18)，我们有 $H_f \leq \frac{Md^{-2}}{K}\log()$。因此，$\mathcal{C}_R(\beta;\ K, M) \subset \mathcal{C}_{FC}(a)$，且 $\mathcal{C}_F(\beta;\ K) \subset \mathcal{C}_{FC}(a)$。由此，类 $\mathcal{C}_{FC}, H_{FC} \leq a$中的任意问题的难度以及问题类 $\mathcal{C}_F$都足够丰富，可以应用引理 6 和引理 5。此外，若 T 满足式 (39) 给出的界，则它自动满足使用上述引理所需的界，因为 $\log(6TK) < \log(6T(K + M + 1))$。

对于 $\mathcal{C}_{FC}$ 上任意算法 Alg 的错误概率下界，至少等于各个子类 $\mathcal{C}_R, \mathcal{C}_F$上的最坏情况错误概率的最大值，即

$$\max_{\mathcal{G}\in\mathcal{C}_{FC}} \mathbb{P}_{\mathcal{G}}(e) \geq \max\left\{\max_{\mathcal{G}\in\mathcal{C}_F} \mathbb{P}_{\mathcal{G}}(e), \max_{\mathcal{G}\in\mathcal{C}_R} \mathbb{P}_{\mathcal{G}}(e)\right\}. \tag{40}$$

我们使用引理 5 为式 (40) 中的第一项建立下界，并使用引理 6 为第二项建立下界。选择使分子中的整体表达式最小的常数，在指数的分母中将最大值移至内部，并根据 $H_{FC}(\mathcal{G}) = \max\{H_{tbp}(\mathcal{G})\ H_2^R(\mathcal{G})\ H_f(\mathcal{G})\}$ 的定义，我们得到

$$\mathbb{P}_{\mathcal{G}\in\mathcal{C}_{FC}(a;K)}(e) \geq \frac{1}{6}\exp\left(\frac{-1200T}{\log(K)H_{FC}(\mathcal{G})}\right).$$

□

D FCSR 误差上界证明

引理 8。（*Ville* 不等式的） 设 $(M_t, \mathcal{F}_t)_{t\geq 0}$ 是满足 $M_0 \leq 1$ 的非负超鞅。那么对于任意 $a >$ 0,

$$\mathbb{P}\big(\sup_{t\geq 0} M_t \geq a\big) \leq \frac{1}{a}.$$

令 $s_r(i)$ 表示 FCSR 在第 $r \in \{1, 2, \ldots, K-1\}$ 轮分配完所有样本后，臂 i 的得分。定义 $\mathcal{F}_r := \{i : s_r(i) > \tau\}$ 为在第 r 轮结束时被判定为可行的所有臂的集合，且 $\mathcal{F}_r^c := [K]/\mathcal{F}_r$ 为被判定为不可行的所有臂的集合。

引理 9。假设存在最佳臂 $i^\star$，则臂 $i^\star$ 在第 $r \in$ 轮结束时被判定为不可行的概率 $\{1, 2, \ldots, K-1\}$ 的上界为

$$\mathbb{P}(i^\star \in \mathcal{F}_r^c) \leq \exp\Big(-\frac{fT}{16R^2 \log K\, H_f}\Big), \tag{41}$$

对于所有 $T \geq$ 最大值$\{{}^{4K^2M}\!/_{f}\}$，其中 $f \in (0,\ 1)$ 是一个静态超参数。回顾 $H_f = {}^{K}\!/_{\text{对数}K\text{最}}$大值 $\bar{\Delta}_{i^\star j}^{-2}$。

证明。回顾一下，FCSR算法（算法5）在每轮 $r \in [K-1]$ 中，以如下方式为所有存活臂分配相同数量的样本：首先对所有 M 属性进行均匀采样（算法6），然后按照 APT例程（算法7）采样，最后按照 SUF例程（算法8）采样。

定义。 固定臂 $i^\star$，令 $A_j^{(s)}$ 表示通过 Uniform 和 APT 例程对 $i^\star$ 的属性 $j \in [M]$ 进行采样的次数，即在调用 SUF 之前，j 在第 s 轮中获得的样本数。
令 $\mathcal{H}_s$ 为截至轮次 $s-1$结束时由所有观测值生成的σ代数。定义 $T_j^{(s)}$ 为FCSR在轮次 s 之前对属性 j 进行采样的次数，并令 $S_j^{(s)}$ 为所观测到的 $T_j^{(s)}$ 个奖励的相应总和。定义偏移量

$$C_j^{(s)} := S_j^{(s)} - T_j^{(s)}\,\tau.$$

注意，若经验均值 $S_j^{(s)}/T_j^{(s)} > \tau$ 与 j 在轮次 $s-1$中被判定为可行，则 $C_j^{(s)} > 0$。因此，若 $C_j^{(s)} \le 0$，则 j 的经验均值在 $s-1$ 结束时小于或等于 τ（其中 $s-1>1$）。后一种情形仅在可行性预算 P_1 于轮次 $s-1$耗尽时出现；否则，根据定义，SUF会继续对该属性进行采样，直到经验均值越过 τ。

令 $X_j(t)$ 表示属性 j 的第 t 个样本（假设其为均值为 μ_j 的 R-次高斯分布）。我们可以将 SUF 在第 s 轮被调用处理属性 j 时将看到的样本流定义为

$$Y_j^{(s)}(u) := X_j\big(T_j^{(s)} + u\big), \qquad u \ge 1.$$

令 $Z^{(s),\,\mathrm{act}}_{\cdot,\,j}$ 表示第 s 轮分配给属性 j 的样本数。我们将量 $Z_j^{(s)}$ 定义为对 $Z^{(s),\,\mathrm{act}}_{\cdot,\,j}$ 的紧密近似：

$$Z_j^{(s)} := \begin{cases} \inf\Big\{ t \ge 0 : \dfrac{S_j^{(s)} + \sum_{u=1}^{t} Y_j^{(s)}(u)}{T_j^{(s)} + t} > \tau \Big\}, & \text{if } C_j^{(s)} > 0, \\ 0, & \text{if } C_j^{(s)} \le 0, \end{cases}$$

约定 $\inf \emptyset = +\infty$。注意，$Z_j = Z^{(s),\,\mathrm{act}}_{\cdot,\,j}$，除非在对属性 j 进行采样的过程中，P_1 耗尽（此时 $Z_j > Z^{(s),\,\mathrm{act}}_{\cdot,\,j}$）。这是因为，如果 $C_j^{(s)} \le 0$，那么 j 必定已在第 $s-1$ 轮被判定为不可行，这意味着可行性预算 P_1 必定已耗尽。因此，$Z_j^{(s)} = 0 = Z^{(s),\,\mathrm{act}}_{\cdot,\,j}$。如果 $C_j^{(s)} > 0$ 且 P_1 未在采样过程中耗尽，SUF 会继续采样，直到经验均值恰好越过 τ，与 Z_j 的定义完全相同。因此，逐路径有

$$Z_j^{(s)} \ge Z_j^{(s),\mathrm{act}}.$$

定义总量

$$S^{\mathrm{act}} := \sum_{s=1}^{r}\sum_{j=1}^{M} Z_j^{(s),\mathrm{act}}, \qquad S := \sum_{s=1}^{r}\sum_{j=1}^{M} Z_j^{(s)}.$$

$S^{\mathrm{act}} \le P_1$ 和 $S \ge S^{\mathrm{act}}$ 始终成立。如果算法在第 r轮结束时宣布臂 $i^\star$ 不可行，那么截至第 r轮执行的 SUF 调用中，可行性预算 P_1 必然已经耗尽；并且由于 $\{S \ge P_1\} \Longrightarrow \{S^{\mathrm{act}} = P_1\}$，我们有

$$\{i^\star \in \mathcal{F}_r^c\} \subseteq \{S \ge P_1\}. \tag{I}$$

因此，

$$\mathbb{P}(i^\star \in \mathcal{F}_r^c) \le \mathbb{P}(S \ge P_1),$$

只需对 $\mathbb{P}(S \ge P_1)$进行界定即可。由包含关系 (I) 和切尔诺夫

$$\mathbb{P}(S \ge P_1) \le e^{-\theta P_1}\, \mathbb{E}[e^{\theta S}]. \tag{42}$$

现在，我们首先对 $E[e^{\theta Z_j^{(s)}}]$进行界定，进而界定 $E[e^{\theta S}]$。

逐属性矩母函数。固定任意轮次 $s \in \{1,\dots,r\}$ 和属性 j，并以 H_s为条件。我们首先在 (i) $C_j^{(s)} \le 0$和 (ii) $C_\ell > 0$ 的情况下，对 $P(Z_j^{(s)} \ge t)$进行界定。

(i) 如果 $C_j^{(s)} \le 0$，那么 $Z_j^{(s)} = 0$ 且

$$\mathbb{P}(Z_j^{(s)} \ge t) = 0. \tag{43}$$

(ii) 若 $C_j^{(s)}$，定义中心化未来和 $\widetilde{S}_j^{(s)}(t)$ 以及指数过程 $M_\ell(t;\lambda)$ 为

$$\widetilde{S}_j^{(s)}(t) := \sum_{u=1}^{t} \big(Y_j^{(s)}(u) - \mu_{i^\star j}\big), \qquad M_\ell(t;\lambda) := \exp\Big(-\lambda \widetilde{S}_\ell(t) - \tfrac{\lambda^2 R^2}{2} t \Big).$$

对任意 $\lambda > 0$，过程 $M_\ell(t;\,\lambda)$ 是一个非负超鞅（关于 $\sigma(\mathcal{H}_s, Y_{(s)\,j}(1)\ldots, Y_{(s)\,j}(t))$），且 $M_\ell(0;\,\lambda) = 1$。在事件 $\{Z_j^{(s)} \ge t\}(t \ge 1)$ 上，利用 $Z_j^{(s)}$ 的定义并从两边减去 $(T_j^{(s)} + t)\mu_{i^\star j}$，可得如下不等式：

$$\frac{S_j^{(s)} + \sum_{u=1}^{t-1} Y_j^{(s)}(u)}{T_j^{(s)} + t - 1} \le \tau \implies \widetilde{S}_j^{(s)}(t-1) \le -(t-1)\bar{\Delta}_{i^\star j} - C_j^{(s)}.$$

因此，在 $\{Z_j^{(s)} \ge t\}$上，

$$M(t-1;\lambda) \ge \exp\Big(\lambda\big((t-1)\bar{\Delta}_{i^\star j} + C_j^{(s)}\big) - \tfrac{\lambda^2 R^2}{2}(t-1)\Big).$$

由于 $\sup_{t\ge 0} M(t;\,\lambda) \ge M(t-1;\,\lambda)$，我们在给定 $\mathcal{H}_s$（即第 s 轮之前所有样本的历史）条件下应用 Ville 不等式（引理 8），其中 $a = \exp\Big(\lambda\big((t-1)\bar{\Delta}_{i^\star j} + C_j^{(s)}\big) - \tfrac{\lambda^2 R^2}{2}(t-1)\Big)$，从而得到

$$\mathbb{P}\big(Z_j^{(s)} \ge t \mid \mathcal{H}_s\big) \le \mathbb{P}\big(\sup_{t\ge 0} M_t \ge a\big) \le \exp\Big(-\lambda\big((t-1)\bar{\Delta}_{i^\star j} + C_j^{(s)}\big) + \tfrac{\lambda^2 R^2}{2}(t-1)\Big).$$

项 $-\lambda C_j^{(s)}$ 非正（因为 $C_j^{(s)} > 0$），因此可以舍去。对 λ 优化（取 $\lambda = \bar{\Delta}_{i^\star j}/R^2$）可得，对于$t \ge 1$,

$$\mathbb{P}\big(Z_j^{(s)} \ge t \mid \mathcal{H}_s\big) \le \exp\Big(-\frac{\bar{\Delta}_{i^\star j}^2}{2R^2}\,(t-1)\Big).$$

定义 $\alpha_j := \bar{\Delta}_{i^\star j}^2/(2R^2)$ 和 $\alpha := \min_j \alpha_j$。因此，对于情形 (i) 和 (ii)，一致地，对所有 $Z_j^{(s)}$，我们有

$$\mathbb{P}\big(Z_j^{(s)} \ge t \mid \mathcal{H}_s\big) \le e^{-\alpha(t-1)} \qquad (t \ge 1).$$

对于 $0 < \theta < \alpha$,

$$\begin{aligned}
\mathbb{E}\big[e^{\theta Z_j^{(s)}} \mid \mathcal{H}_{\ell-1}\big] &= 1 + \sum_{t\ge 1}(e^{\theta t} - e^{\theta(t-1)})\mathbb{P}(Z_j^{(s)} \ge t \mid \mathcal{H}_s) \\
&\le 1 + \sum_{t\ge 1}(e^{\theta t} - e^{\theta(t-1)})e^{-\alpha(t-1)} \\
&\le 1 + \frac{e^\theta - 1}{1 - e^{-(\alpha-\theta)}} = \frac{e^\theta - e^{-(\alpha-\theta)}}{1 - e^{-(\alpha-\theta)}} =: G(\theta).
\end{aligned}$$

总体矩母函数。由各属性的独立性，我们有,

$$\mathbb{E}\big[e^{\theta S}\big] = \mathbb{E}\big[e^{\theta \sum_{s=1}^r \sum_{j=1}^M Z_j^{(s)}}\big] = \prod_{j=1}^M E\Big[e^{\theta \sum_{s=1}^r Z_j^{(s)}}\Big].$$

利用期望的迭代法则，以及属性 j 在轮次 s 中的未来样本相对于先前各轮次的条件独立性，我们得到如下结果

$$E\Big[e^{\theta \sum_{s=1}^r Z_j^{(s)}}\Big] \le \mathbb{E}\Big[\mathbb{E}\big[e^{\theta Z_j^{(r)}} e^{\theta \sum_{s=1}^{r-1} Z_j^{(r)}} \,\big|\, \mathcal{H}_r\big]\Big] \le E\Big[e^{\theta \sum_{s=1}^{r-1} Z_j^{(r)}}\Big]\mathbb{E}\big[e^{\theta Z_j^{(r)}} \,\big|\, \mathcal{H}_r\big] \le E\Big[e^{\theta \sum_{s=1}^{r-1} Z_j^{(r)}}\Big]G(\theta)$$

其中最后一个不等式由每次调用的界 $\mathbb{E}\big[e^{\theta Z_j^{(s)}} \,\big|\, \mathcal{H}_s\big] \le G(\theta)$得到。将这一条件化论证对$s = r-1,\ r-2,$ ....., 1进行迭代，可得

$$E\Big[e^{\theta \sum_{s=1}^r Z_j^{(s)}}\Big] \le G(\theta)^r.$$

最后,

$$\mathbb{E}\big[e^{\theta S}\big] = \prod_{j=1}^M E\Big[e^{\theta \sum_{s=1}^r Z_j^{(s)}}\Big] \le G(\theta)^{Mr}.$$

因此，根据包含关系 (I) 和切尔诺夫不等式,

$$\mathbb{P}(i^\star \in \mathcal{F}_r^c) \le \exp\big(-\theta B_1 + Mr \log G(\theta)\big), \qquad 0 < \theta < \alpha.$$

选择 $\theta = \alpha/2$。若 $B_1 \geq \frac{4}{\alpha} Mr\log G(\frac{\alpha}{2})$，则 $Mr\log G(\alpha/2) \leq \frac{\alpha}{4} B_1$，因此

$$-\theta B_1 + Mr \log G(\theta) = -\tfrac{\alpha}{2} B_1 + Mr \log G(\tfrac{\alpha}{2}) \leq -\tfrac{\alpha}{4} B_1.$$

因此

$$\mathbb{P}(i^\star \in \mathcal{F}_r^c) \leq \exp\Big(-\frac{\alpha}{4} B_1\Big).$$

回顾 $B_1 = \lfloor fT/K \rfloor$。对于任意 $x \geq 1$，我们有 $\lfloor x \rfloor \geq x/2$。因此，若 $\frac{fT}{K} \geq 2$，则 $B_1 = \Big\lfloor \frac{fT}{K} \Big\rfloor \geq \frac{1}{2} \frac{fT}{K}$。在此条件下，该界为

$$\mathbb{P}(i^\star \in \mathcal{F}_r^c) \leq \exp\Big(-\frac{\alpha}{8} \frac{fT}{K}\Big). \tag{44}$$

Eq对于$B_1 \geq \frac{4}{\alpha} Mr\log G(\frac{\alpha}{2})$等价地，利用 $fT/K \geq 2$和 $B_1 \geq fT/2K$，只需要求

$$T \geq \frac{8K}{f\alpha} Mr \log G(\tfrac{\alpha}{2}).$$

由于$H_f = \frac{K}{\log K} \max_j \bar{\Delta}_{i^\star j}^{-2}$，我们得到$\alpha = \frac{K}{2R^2 \log K H_f}$。将其代入方程44，即得所需的界

$$\mathbb{P}(i^\star \in \mathcal{F}_r^c) \leq \exp\Big(-\frac{fT}{16R^2 \log K H_f}\Big),$$

对所有 $T \geq Mr\log \frac{K^{16R^2 H_f}}{f} \log G(\frac{K}{4R^2 \log K H_f})$成立。可以验证，粗略上界 $G(\theta) < \theta$成立。因此我们需要 $T \geq \max\{\frac{4KMr}{f}\}$，而这在 $T \geq \max\{\frac{4K^2 M}{f}\}$时成立。

□

引理10 (带有 P随机插入的 APT)。设 K 臂服从均值为 μ_i 的 R-次高斯分布。固定阈值 τ，并设定$\varepsilon = 0$。定义

$$\Delta_i := |\mu_i - \tau|, \qquad H := \sum_{i=1}^{K} \Delta_i^{-2}.$$

设总抽样预算为 $T_{\mathrm{tot}} = T + P$，并假设在这 T_{tot} 次抽样中，存在一个恰好包含 T 个时间步的子序列，在这些时间步中算法按照 *APT* 规则选择臂（将这些步骤称为APT 抽样步骤）。在这些T 个 *APT* 抽样步骤上运行 *APT*，对于复杂度至多为 H 的问题类 $\mathcal{B}_{H,\,R}$，可得如下保证：若

$$T \geq 256\, H\, R^2 \log\big((\log T + 1)K\big),$$

then

$$\sup_{B \in \mathcal{B}_{H,R}} \mathbb{E}\big[\mathcal{L}(T)\big] \leq \exp\Big(-\frac{T}{128\, R^2\, H}\Big).$$

证明。Locatelli 等人 [2016] 的证明通过定义一个有利事件 ξ 展开，在该事件中经验均值集中，并证明 APT 在 ξ 上不会出错。在我们的设定中，P额外的样本不会影响集中事件 ξ 的概率。因此，该证明与 Locatelli 等人 [2016] 在附录A.2中的证明完全相同，只需修改为：下文所有经验均值、计数和索引都仅根据总共 $T + P$ 次抽取中的 APT 抽样抽取来计算。形式上，令 $t_1 < t_2, \ldots, < t_T$ 为按照 APT规则抽取的样本索引。将这些样本依次编号为 1、2、…，T，其余非 APT 样本则依照真实时间步递增顺序编号为 $T+1$、…，$T+P$。

设 $\mathbb{A}$ 为一组 K R-次高斯臂，且 $\delta = (4\sqrt{2})^{-1}$。将集中事件 ξ 定义如下：

$$\xi = \left\{ \forall i \in \mathbb{A}, \forall s \in \{1, ..., T\} : \left| \frac{1}{s} \sum_{t=1}^{s} X_{i,t} - \mu_i \right| \leq \sqrt{\frac{T\delta^2}{Hs}} \right\}.$$

证明的其余部分可直接得出，因为即使新增样本并非按照 APT 规则对 K 臂进行采样，集中事件 $\mathbb{P}(\xi)$ 的概率也只会随样本增加而增大。

□

定理 11 (FCSR的性能)。 设 $K, M \geq 1$。给定参数 $f, g \in (0,1)$ 以及 τ。设 $c = 1/32R^2$ 为一个常数， R 为次高斯参数。*FCSR*(τ, *f*, *g*) 的错误概率满足

$$\mathbb{P}(e) \leq 3K^2 \exp\left(-\frac{cT}{\log(K) H_{FC}(\mathcal{B})}\right) \tag{45}$$

对于所有

$$T \geq \max\left\{256 H_{tbp} R^2 \log((\log(T)+1)M), \frac{4K^2 M}{f}, \frac{K \max\{{H_{tbp}}^{-1}, (H_2^R)^{-1}\}}{\log(K)}\right\}.$$

证明。 我们的证明与 Audibert 等人 [2010] 的方法类似。定义 $n_i = \lceil \quad - {}^{n-K}{}_{\log(K)(K+1-i)}\rceil$。~~按照~~真实臂均值的顺序对所有臂进行编号。在第 i 轮，至少有一条最差的 i 臂仍然存留。若最佳臂 $i^\star$ 在第 $1 \leq i \leq K-1-|\mathcal{R}|$ 轮被淘汰，则至少有一条索引为 $|\mathcal{R}|+2 \leq j \leq K$ 的次优臂存留，因此必须胜过 $i^\star$；或者最佳臂 $i^\star$ 被判定为不可行。将前一种事件记为 s_i，后一种事件记为 $f_i = \{i^\star \in \mathcal{F}_i^C\}$。

现在，若最佳臂 $i^\star$ 在第 $K-|\mathcal{R}| \leq i \leq K-1$ 轮被淘汰，则至少有一条索引为 $1 \leq j \leq |\mathcal{R}|$ 的风险臂必须被判定为可行，才能胜过最佳臂 $i^\star$。此事件是这样一个事件的子集：存在某条 不可行臂 $r \in \mathcal{F}^C$ 在第 i 轮结束时被判定为可行。将此事件记为 r_i。因此，在这种情况下，我们可以用最佳臂被淘汰的概率的上界，即它被判定为不可行的概率，或事件 r_i 发生的概率。

根据上述两种情况以及所有轮次 $i \in [K-1]$ 的并集界，我们对 $i^\star$ 未被标记为最佳臂的概率给出上界：

$$\sum_{i=1}^{K-1} \mathbb{P}(e_i) \leq \sum_{i=1}^{K-1} \mathbb{P}(f_i) + \sum_{i=1}^{K-1-|\mathcal{R}|} \mathbb{P}(s_i) + \sum_{K-|\mathcal{R}|}^{K-1} \mathbb{P}(r_i). \tag{46}$$

我们利用引理 8 对 $\mathbb{P}(f_i) = \mathbb{P}(i^\star \in \mathcal{F}_r^C)$ 进行上界估计，因为根据假设，我们有 $T \geq \max\{{}^{4KMr}_{f}\}$。~~因此~~，对于任意 $i \in [K-1]$,

$$\sum_{i=1}^{K-1} \mathbb{P}(f_i) \leq K \exp\Big(-\frac{fT}{16R^2 \log K\, H_f}\Big). \tag{47}$$

为了对 $\mathbb{P}(r_i)$ 进行上界估计，注意到，若要将某个不可行臂判定为可行，即 $r \in \mathcal{F}_i$，则必须在经过 $\lfloor g(1-f)n_i \rfloor$ 个 APT 样本后将其判定为可行。根据假设，我们有 $T \geq 256\, H_{tbp}\, R^2 \log\big((\log\, T+1)M\big)$，因此，由引理10和并集界可得

$$\sum_{K-|\mathcal{R}|}^{K-1} \mathbb{P}(r_i) \leq \sum_{K-|\mathcal{R}|}^{K-1} \sum_{r \in \mathcal{F}^C} \mathbb{P}(r \in \mathcal{F}_i) \leq K^2 \exp\left(\frac{-g(1-f)(T-K)-1}{\overline{\log}(K) \max_{r \in \mathcal{F}^C} K H_{tbp}(r)}\right), \tag{48}$$

其中，最后一个不等式是将最大值移入指数、利用 $(K+1-i) \leq K$ 并对取整下界进行下界估计后得到的。最后， $\mathbb{P}(s_i)$ 的上界由以下事实得到：对于轮次 $1 \leq i \leq K-1-|\mathcal{R}|$，某个索引为 $|\mathcal{R}|+2 \leq j \leq K$ 的臂胜过 $i^\star$ 的概率为

$$\begin{aligned}\sum_{K-|\mathcal{R}|}^{K-1} \mathbb{P}(s_i) \leq \sum_{K-|\mathcal{R}|}^{K-1} \sum_{k=K+1-i}^{K} \mathbb{P}(\hat{\mu}_{i^\star} \leq \hat{\mu}_k) &\leq K^2 \exp\left(-\lfloor (1-f)(1-g)n_k \rfloor \Delta^2_{(K+1-i)}\right) \\ &\leq K^2 \exp\left(\frac{-(1-f)(1-g)(T-K)-1}{\overline{\log}(K) H_2^R}\right).\end{aligned} \tag{49}$$

上述不等式源于这样一个事实：截至第 i 轮，已根据 APT规则分配了 $\lfloor (1-g)(1-f)n_i \rfloor$ 个样本，并且利用了关于下取整的下界。我们可以将 $\log(K)$ 上界为 $\leq 2\log(K)$。由于 $f,\ g \in (0,\ 1)$，且根据我们对 T 的假设

$$T \geq \frac{K}{\log(K)} \max\{{H_{tbp}}^{-1}, (H_2^R)^{-1}\} \implies T \geq 2\max\left\{\frac{g(1-f)K}{2\log(K) H_{tbp}}, \frac{(1-g)(1-f)K}{2\log(K) H_2}\right\}.$$

因此，我们可以在 (49) 和 (48) 中用 $T/2$ 替换 $(T-K)$，并忽略分子中的 -1。最后，通过取最大值并利用 (47)、(48)、(49)，再结合 $f,\ g \in (0,\ 1)$，将 (46) 上界化，即可得到所需结果。因此，

$$\mathbb{P}(e) \leq \sum_{i=1}^{K-1} \mathbb{P}(e_i) \leq 3K^2 \exp\left(\frac{-T}{32R^2 \log(K) \max\{H_f, H_{tbp}, H_2^R\}}\right). \tag{50}$$

□