arXiv:2602.01400v1 [stat.ML] 1 Feb 2026

基于在线社会福利函数的资源分配

Kanad Pardeshi [1] **Samsara Foubert** [2] **Aarti Singh** [1]

摘要

在许多现实场景中，集中式决策者必须在多个时间步内，反复向一群人口分配有限资源。获得资源的个体会产生某种随机效用；为了刻画分配对整体人口的影响，随后使用社会福利函数（SWF）汇总个体效用的期望值。我们对这一场景进行形式化，并提出一个通用的置信序列框架，用于基于SWF的在线学习与推断。该框架适用于任何单调、凹且利普希茨连续的SWF。我们的关键洞见是，仅凭单调性，就足以将个体效用的置信序列提升为最优福利的任意时刻有效界。基于这一基础，我们提出SWF-UCB，这是一种与SWF无关的在线学习算法，可实现近乎最优的$\tilde{\mathcal{O}}\left(n+\sqrt{nkT}\right)$遗憾（每个$T$时间步为$n$个体分配$k$资源）。我们将该框架应用于三个规范立场各异的社会福利函数族：加权幂平均、Kolm和基尼，并分别为其提供量身定制的预言机算法。实验验证了$\sqrt{T}$的缩放特性，并揭示了k与社会福利函数参数之间丰富的相互作用。该框架可自然地支持多种推断应用，例如序贯假设检验、最优停止和策略评估。

1. 引言

决策者经常面临这样的问题：如何随着时间推移，向一定人口分配有限资源，例如公园管理部门向偏远地区安排护林员巡逻，慢性病护理项目在患者之间分配社区卫生工作者探访，或学区在学生之间分配辅导名额。评估此类分配需要将个体结果聚合为一个连贯的集体福利衡量指标。社会福利函数（SWF）为这种聚合提供了一种有原则的方法，其公理基础源自福利经济学，并将公平性与效率之间的权衡以适于优化的形式加以编码。近期的机器学习研究已在学习理论框架内对社会福利函数进行形式化，并为从批量数据中估计社会福利提供了样本复杂度界限（Cousins，2021年；Pardeshi 等人，2024）。

现实世界中的资源分配本质上是序贯的：个体效用未知，必须根据观测到的结果进行学习，但数据积累期间决策又不能延后。每次资源分配都会影响人口的福利，而次优决策造成的损失会随时间累积。这里有两个相互交织的挑战。首先，即使效用估计存在不确定性，决策者也必须有效地分配资源，并以某种方式平衡减少不确定性的策略与次优分配的实际代价。其次，他们需要随时获得有效的统计评估和福利保证：当前策略是否已达到预期的福利阈值？不确定性是否足够低，可以确定采用当前策略并大规模部署？

这些问题会因预算周期、利益相关方审查或优先事项变化等外部因素而适应性地出现。因此，两项挑战都要求统计保证无论何时调用都持续有效。近期研究开始探讨多臂老虎机中的在线福利优化，通过汇总决策时间序列中的福利来衡量随时间变化的公平性（详见第2节）。与此不同，我们考虑的是固定人口反复获得资源、且福利在每个决策点衡量这一持久人口内部公平性的情形。这种固定人口的表述具有独特结构：分配策略是针对个体的连续概率向量，而非离散的臂选择；每轮分配多种资源；并且决策受到约束，即分配概率之和必须等于可用资源数量。因此，最优策略取决于所有个体之间的相对效用，从而产生标准多臂老虎机问题中不存在的相互依赖关系。我们通过统一的基于置信序列的框架应对这两项挑战。

置信序列是在任意停止时刻均保持有效的区间估计，这一特性有助于指导策略决策或认证统计结论——

[1]卡内基梅隆大学机器学习系 [2]卡内基梅隆大学软件与社会系统系。通信请联系：Kanad Pardeshi< kpardesh@andrew.cmu.edu>。

预印本。*2026*年*2*月*3*日。

。我们的框架对社会福利函数（SWF）提出了三个自然假设——单调性、凹性和 Lipschitz 连续性——每个假设都发挥着独特且不可或缺的作用。我们的核心洞见是，仅凭社会福利函数的单调性，就足以将个体效用的逐坐标置信序列提升为当前分配总体福利的任意时刻有效界（定理 4.1）。这一提升原理将一种灵活的统计基础工具与决策者实际需要的信息联系起来：对期望效用上的最优福利值提供随时间始终成立的保证。同样的机制不仅支持在线学习中的乐观分配策略，还自然地产生推断应用——序贯假设检验、最优停止和策略评估。

我们做出以下贡献。首先，我们形式化了跨多个时间步、针对人口进行资源分配的问题，其目标函数由基于社会福利函数的事前效用聚合给出。其次，我们建立了一个置信序列提升定理，仅依赖单调性即可提供任意时刻有效的福利界。第三，我们将该框架具体应用于三个公理上各不相同的社会福利函数族——加权幂平均、Kolm 和基尼，涵盖从功利主义目标到平等主义目标的整个谱系，并针对每种情况开发了定制的高效策略优化预言机。第四，我们提出了 SWF-UCB，这是一种用于在线福利最大化的通用算法，并证明其能够针对 $\tilde{\mathcal{O}}\left(n+\sqrt{nkT}\right)$ 资源在人口 k 中的分配，在 n 个时间步内实现近最优的 T 遗憾。第五，我们通过实验验证了所预测的 $\sqrt{T}$ 扩展规律，并发现 k 的中间值会产生最高的遗憾——这种非单调依赖关系表明，资源稀缺性与学习难度之间的相互作用具有丰富的结构。最后，我们讨论了同一置信序列框架如何自然地应用于有意义的推断任务，例如序贯假设检验、最优停止和策略评估（附录 A.1）。

2. 相关工作

我们的框架通过在线福利最大化和时间一致统计推断这一共同视角，统一了多臂老虎机、公平分配和福利经济学等多个研究方向。然而，与以往关注福利或公平性的多臂老虎机研究不同，后者要么只关注特定目标函数，要么只关注学习保证；我们的方法支持一般的单调社会福利函数（SWF），提供近乎最优的遗憾保证，并能在部分反馈下对优化后的福利进行任意时刻有效的推断。

多臂老虎机变体。多选择老虎机（Anan-tharam等人，1987；Gai等人，2012）在多个臂均可被拉动时，考虑臂奖励之和，这对应于我们设定中的功利主义福利。组合老虎机（Cesa-Bianchi 与 Lugosi，2012；Kveton 等人，2015）研究如何找出一组最佳臂进行拉动，以获得最高奖励。相比之下，在我们的设定中，所有个体都必须以非零概率获得资源，因此我们要找出能带来最大福利的分配概率。（Chen等人，2013）研究了具有非线性超臂奖励的组合老虎机，并假设其满足单调性和有界光滑性。我们利用社会福利函数的单调性，建立了一般性的置信序列提升结果（定理4.1），并为我们的算法（SWF-UCB）开发了高效且精确的最优分配预言机。

公平多臂老虎机。多臂老虎机中的公平性通常要求每个臂至少被拉动一定次数（Chen等人，2020年；Wang 等人，2021年），或要求任人唯贤的公平性（Joseph 等人，2016年；2018年）。（Lim 等人，2024）研究了多重播放多臂老虎机，其中每次播放对应一位用户，并以处境最差用户的表现（即平等主义福利）衡量遗憾。我们的工作涵盖平等主义福利，并为在人口中分配资源时的学习与推断提供了一个框架。

多臂老虎机中的基于福利的遗憾。(Barman 等人，2023年) 和 (Sawarni 等人，2023年) 针对老虎机设定引入了纳什遗憾，通过每轮遗憾的几何平均数来衡量累积损失。(Sarkar 等人，2025) 和 (Krishna 等人，2025) 将这一思路扩展到了 p均值福利族。这一系列工作会跨时间步聚合效用，而我们的工作则在每个时间步聚合人口中的效用。我们还为不同的社会福利函数提供了一个通用框架；在不同的聚合维度下，p均值福利族也是我们所考虑的福利族中的一个特例。

社会福利函数。我们社会福利函数族的公理基础源自福利经济学。Atkin-son (1970年) 提出了厌恶不平等的福利函数，并用单一参数刻画公平与效率之间的权衡：即我们的 WPM 族。Kolm (1976年) 通过平移不变性刻画了 Kolm-Pollak 族，而Weymark (1981年) 则通过依赖排名的权重对广义基尼福利进行了形式化。在机器学习领域，Cousins(2021年；2023年) 提出了 Hölder 连续性界，从而为 PAC 型学习保证提供了支持。Pardeshi 等人(2024) 研究了互补问题，即从偏好数据中学习社会福利函数。据我们所知，我们的工作首次为分配问题提供了基于这些福利族、具有可证明遗憾保证的在线学习和时间一致推断框架。

3. 问题设定

我们考虑一个由 n 个体组成的人口，其中集中式决策者在离散时间步分配相同的不可分资源。在每个时间步 t，$k \leq n$

资源到达并在人口中进行分配。每个个体最多获得一个资源，个体 i 在时间 t 获得资源时的效用为$u_{t,i} \overset{iid}{\sim} \mathcal{D}_i$，其中 $\mathcal{D}_i$ 是均值为 $\mu_i = \mathbb{E}_{U\sim\mathcal{D}_i}[U]$的未知分布。我们假设 $\mu_i > 0$，未获得资源的个体的效用为零。

分配根据一个（随机化）策略进行，由向量 $\mathbf{p} \in [0,1]^n$ 给出，其中 $\sum_i p_i = k$。每个 p_i 是个体 i 获得其中一个 k资源的边际概率。因此，策略 $\mathbf{p}$ 对个体 i的事前期望效用为 $\mu_i p_i$。我们用 $\boldsymbol{\mu} \odot \mathbf{p}$表示期望效用向量。

我们的模型假设，$\{\mathcal{D}_i\}_{i\in[n]}$ 个体效用的分布保持静态，不随时间变化。相反，我们允许分配**策略p**随时间变化，并对分配策略进行在线学习和推断。分配的有效性由使用社会福利函数（SWF）聚合事前期望效用来确定，记为 $M(\boldsymbol{\mu} \odot \mathbf{p})$。

我们出于两个原因考虑事前效用。首先，事前效用表示分配之前的期望效用。相比之下，分配后实现效用会在给定时间步将零效用赋予所有未获得资源的个体，这可能导致若干常见的社会福利函数在部分分配下退化或失去信息。因此，在我们的设定中，事后福利不是一个有意义的优化或比较对象。其次，事前效用具有自然的长期解释：对于固定的分配**策略p**，向量 $\boldsymbol{\mu}\odot \mathbf{p}$对应于按照**p**反复分配资源时获得的时间平均效用。这一解释对于面向推断的场景尤为重要——例如策略评估、序贯检验和最优停止——在这些场景中，需要反复部署单个策略，并通过长期平均值评估其表现。

我们考虑满足三个自然假设的社会福利函数族：

(A1) 单调性：设$\mathbf{v}_1$,$\mathbf{v}_2 \in \mathbb{R}^n_+$ 为两个效用向量。若$v_{1,i} \geq v_{2,i}$对所有 $i \in [n]$都成立，则 $M(\mathbf{v}_1) \geq M(\mathbf{v}_2)$。(A2) 凹性：$M(\mathbf{v})$关于$\mathbf{v}$是凹的。(A3) 利普希茨连续性：$M(\mathbf{v})$关于$\mathbf{v}$ 在 ℓ_∞ 范数下是利普希茨连续的。

每个假设都以模块化方式用于我们的框架中。单调性用于根据观测效用构建 $M(\boldsymbol{\mu} \odot \mathbf{p})$ 的置信序列（第4节），这是我们的核心统计洞见。凹性使策略优化和最优解的高效计算变得易于处理（第5.1节）。最后，Lipschitz连续性支持我们的理论分析和遗憾保证（第5.2节），涉及我们的算法在在线SWF最大化中的最优性。

3.1. 社会福利函数族

我们考虑三种常见的社会福利函数族。

1. 加权幂平均（*WPM*）的参数为$\mathbf{w} \in \Delta_{n-1}$ 和 $q \in (-\infty, 1] \cup \{-\infty\}$。其定义如下

$$M_{\text{WPM}}(\boldsymbol{\mu} \odot \mathbf{p}; \mathbf{w}, q) = \begin{cases} \min_{i\in[n]} \mu_i p_i & \text{如果 } q = -\infty \\ \prod_i (\mu_i p_i)^{w_i} & \text{如果 } q = 0 \\ \left(\sum_i w_i (\mu_i p_i)^q\right)^{1/q} & \text{否则} \end{cases}.$$

直观地说，$\mathbf{w}$ 编码了赋予各个体的相对权重。q 编码了所采用的公平性概念：$q = 0$ 对应于纳什社会福利，而 $q = 1$ 对应于功利主义社会福利（$M_{\text{WPM}} = \sum_i w_i \mu_i p_i$）。WPM满足相对不平等厌恶（Cousins，2023年）：对于任意 $a > 0$，则 $M_{\text{WPM}}(a\mathbf{u}) = aM_{\text{WPM}}(\mathbf{u})$。

2. （加权）科尔姆社会福利具有参数$\mathbf{w} \in \Delta_{n-1}$ 和 $q \in (-\infty, 0] \cup \{-\infty\}$。其定义为

$$M_{\text{Kolm}}(\boldsymbol{\mu} \odot \mathbf{p}; \mathbf{w}, q) = \begin{cases} \min_{i\in[n]} \mu_i p_i & \text{如果 } q = -\infty \\ \sum_i w_i \mu_i p_i & \text{如果 } q = 0 \\ \frac{1}{q} \cdot \log\left(\sum_i w_i \exp\left(q\mu_i p_i\right)\right) & \text{否则} \end{cases}.$$

$\mathbf{w}$ 和 q 编码个体的相对权重和公平性概念（与 WPM 相同）。Kolm 社会福利函数满足绝对不平等厌恶公理（Kolm，1976年）：对于任意 $a > 0$ $M_{\text{Kolm}}(\mathbf{u} + a\mathbf{1}_n) = M_{\text{Kolm}}(\mathbf{u}) + a$。

3. 基尼社会福利具有参数 $\mathbf{w} \in [0,1]^n$，使得 $w_1 \geq w_2 \geq \ldots \geq w_n \geq 0$，且

$$M_{\text{Gini}}(\boldsymbol{\mu} \odot \mathbf{p}; \mathbf{w}) = \sum_i w_i (\boldsymbol{\mu} \odot \mathbf{p})_{(i)},$$

其中 $(\boldsymbol{\mu} \odot \mathbf{p})_{(i)}$ 表示 i向量 $\boldsymbol{\mu}\odot\mathbf{p}$中第 小的元素。我们注意到，适当选择 $\mathbf{w}$，M_{Gini} 与功利主义福利相同（$\mathbf{w} = \mathbf{1}_n$），也与平等主义福利相同（$w_1 = 1$，$w_i = 0$ 对于 $i \geq 2$，并采用适当的$\mathbf{w}$。基尼社会福利函数满足基于排序的不平等敏感性（Weymark，1981年）：效用的排序很重要，但个体身份并不重要。

这三个社会福利函数族以不同方式涵盖了从平等主义福利到功利主义福利的社会福利表述空间。至关重要的是，这三个社会福利函数都满足我们的假设（A1）至（A3）。

命题3.1。$M_{WPM}(\mathbf{v})$、$M_{Kolm}(\mathbf{v})$和 $M_{Gini}(\mathbf{v})$ 关于$\mathbf{v}$均具有单调性且为凹函数。此外，

1. 设$u_{t,i} \in [u_{\min}, u_{\max}]$，其中 $u_{\max} \geq u_{\min} > 0$。则 M_{WPM} 关于 ℓ_∞ 范数是利普希茨连续的，其常数上界为 $u_{\max}/u_{\min}$。

2. M_{Kolm} 关于 ℓ_∞ 范数是利普希茨连续的，常数为1。

3. M_{Gini} 关于 ℓ_∞ 范数是利普希茨连续的，常数为 $\sum_i w_i$。

我们在附录B.1中证明该结果。

注记3.2。虽然我们给出的上界 $u_{\max}/u_{\min}$ 作为WPM族利普希茨常数的上界，对所有$q \in (-\infty, 1] \cup \{-\infty\}$，（Cousins，2021年）给出了更紧的利普希茨连续性界，针对 $q < 0$ 以及H¨older连续性界，针对 $q \in [0,1]$。这些界可用于获得更紧的理论保证（例如，代入定理5.2）。

我们将最优分配$\mathbf{p}^*$（关于社会福利函数 $M(\cdot)$）定义为

$$\mathbf{p}^* = \arg\max_{\mathbf{p}\in\mathcal{P}_k} M(\boldsymbol{\mu}\odot\mathbf{p}),$$

其中 $\mathcal{P}_k = \{\mathbf{p} \in [0,1]^n \mid \sum_i p_i = k\}$。

4. 利用单调性的置信序列框架

我们首先为最优福利值 $M(\boldsymbol{\mu}\odot\mathbf{p}^*)$ 构造一个置信序列（CS），作为学习、检验和停止程序的统计基础。给定观测数据 $\{x_t\}_{t\geq1}$ 和某个 $\delta \in (0,1)$，针对目标量 θ 的有效（$1-\delta$）置信序列由一系列区间 $\left\{[x_t^\downarrow, x_t^\uparrow]\right\}_{t\geq1}$ 组成，使得 $x_t^\downarrow$，且 $x_t^\uparrow$ 取决于过去的数据 $\{x_s\}_{s<t}$ 和 δ。

这一系列区间满足以下性质

$$\mathbb{P}(\exists t \in \mathbb{N} : \theta \notin [x_t^\downarrow, x_t^\uparrow]) \leq \delta,$$

其中 $\delta \in (0,1)$。这一保证具有时间统一性，即对所有时间步 t 同时成立。置信序列是序贯统计推断的核心工具，使决策者能够设计可灵活停止的自适应实验。在我们的设定中，它们为静态或通过自适应学习得到的分配策略所产生的福利值提供任意时刻有效保证，从而无需牺牲统计有效性即可进行假设检验并作出停止决策。

例如，决策者在检验假设 $H_0: \theta = x_0$ 与 $H_1: \theta \neq x_0$ 时，一旦满足 $x_0 \notin [x_t^\downarrow, x_t^\uparrow]$，即可拒绝原假设。时间一致的保证确保这种自适应停止仍能得到有效检验。

在我们的设定中，我们观察效用 $u_{t,\,i}$（对应于被分配资源的个体），而我们的目标量是$M(\boldsymbol{\mu}\odot\mathbf{p}^*)$。置信序列构造对于真实均值效用 μ_i，在给定观测 $\{u_{t,i}\}_{t\geq1}$（Howard 等人，2020年；2021年）时已相当成熟。然而，构造一个 $M(\boldsymbol{\mu}\odot\mathbf{p}^*)$ 的置信序列，因为该量依赖于 $\boldsymbol{\mu}$ 本身以及 $\mathbf{p}^*$，而后者是在给定 $\boldsymbol{\mu}$ 时优化得到的。因此，我们需要 $\boldsymbol{\mu}$ 的一个估计值，并可对其优化以获得分配 $\mathbf{p}_t$，从而得到既有效又具信息量的置信序列。

我们的关键观察是，仅凭单调性（A1）就足以将个体效用的逐坐标置信序列提升为最优社会福利的有效置信序列。这一提升原理构成了后续所有算法和统计保证的基础。

定理 4.1。*(置信序列提升)* 设 $\left\{\boldsymbol{\mu}_t^\downarrow\right\}_{t\geq1}$ 和 $\left\{\boldsymbol{\mu}_t^\uparrow\right\}_{t\geq1}$ 是两个序列，且 $[\mu_{t,i}^\downarrow, \mu_{t,i}^\uparrow]$ 是关于 μ_i 的有效$(1-\delta/n)$置信序列。此外，设

$$\mathbf{p}_t^\downarrow = \arg\max_{\mathbf{p}\in\mathcal{P}_k} M(\boldsymbol{\mu}_t^\downarrow\odot\mathbf{p}), \quad \textit{and}$$
$$\mathbf{p}_t^\uparrow = \arg\max_{\mathbf{p}\in\mathcal{P}_k} M(\boldsymbol{\mu}_t^\uparrow\odot\mathbf{p}).$$

那么，我们以$(1-\delta)$ 的概率一致地有，

$$M(\boldsymbol{\mu}\odot\mathbf{p}^*) \geq M(\boldsymbol{\mu}\odot\mathbf{p}_t^\downarrow) \geq M(\boldsymbol{\mu}_t^\downarrow\odot\mathbf{p}_t^\downarrow), \quad \textit{and}$$
$$M(\boldsymbol{\mu}\odot\mathbf{p}^*) \leq M(\boldsymbol{\mu}_t^\uparrow\odot\mathbf{p}^*) \leq M(\boldsymbol{\mu}_t^\uparrow\odot\mathbf{p}_t^\uparrow).$$

Thus, $\left\{[M(\boldsymbol{\mu}_t^\downarrow\odot\mathbf{p}_t^\downarrow), M(\boldsymbol{\mu}_t^\uparrow\odot\mathbf{p}_t^\uparrow)]\right\}_{t\geq1}$ *is a valid* $(1-\delta)$ *CS for* $M(\boldsymbol{\mu}\odot\mathbf{p}^*)$.

我们在附录 B.2 中仅使用单调性假设证明这一结果。这是一个一般性结果，使我们能够将任意有效的个体平均效用置信集 μ_i 提升为最优 SWF 的置信集。此外，如果 $\mathbf{p}$ 是某个固定且已知的分配，我们注意到 $\left\{[M(\boldsymbol{\mu}_t^\downarrow\odot\mathbf{p}) \; M(\boldsymbol{\mu}_t^\uparrow\odot\mathbf{p})]\right\}_{t\geq1}$ 是一个有效的$(1-\delta)$ 置信集，适用于 $M(\boldsymbol{\mu}\odot\mathbf{p})$。这个提升后的置信集可通过优化上界，立即支持 UCB 风格策略，并支持假设检验和最优停止等双侧推断任务。正文主要关注 SWF 最大化的在线学习；推断应用则见附录 A.1。

5. 在线 SWF 最大化

我们考虑在线寻找最优分配$\mathbf{p}^*$这一任务。令$\mathbf{p}_t$ 表示时刻 t的分配，其事前社会福利为 $M(\boldsymbol{\mu}\odot\mathbf{p}_t)$。我们将遗憾定义为

$$R(T) = \sum_{t=1}^{T} \left(M(\boldsymbol{\mu}\odot\mathbf{p}^*) - M(\boldsymbol{\mu}\odot\mathbf{p}_t)\right),$$

它衡量的是：若已知真实的平均效用 $\boldsymbol{\mu}$，相对于最佳固定随机化分配所造成的福利损失。本任务通过考虑功利主义社会福利（由我们选取的全部 3 个社会福利函数实现），重新得到每轮进行 k 次拉取的随机多重选择多臂老虎机设定。

因此，我们的设定也可以解释为将多选择老虎机推广至资源分配，其中奖励由社会福利函数定义为一种聚合指标。

5.1. 利用凹性的算法

在这项任务中，我们面临经典的探索-利用权衡：提高在时间 t 将资源分配给个体 i 的概率 $p_{t,\,i}$，能让我们更准确地估计 μ_i。然而，这会以牺牲其他本可获得该资源的个体为代价。更复杂的是，分配概率 $p_{t,\,i}$ 既取决于对个体 i 的估计效用 μ_i，也取决于对其他个体平均效用的估计。尽管如此，我们开发了 SWF-UCB 算法，该算法以 UCB 为灵感并针对我们的场景进行了调整，结果表明其性能达到最优。

该算法根据个体平均效用的上置信序列 $\boldsymbol{\mu}_t^{\uparrow}$ 构造策略向量 $\mathbf{p}_t$。通过使用$\mathbf{p}_t$抽取实际分配 $S_t \subseteq [n]$，最后利用观测效用更新获配个体的估计值。我们的通用、与社会福利函数无关的算法见算法1。

算法 1广义 SWF-UCB

要求： 时间范围 T，个体数量 n，每轮资源数 k

要求： 初始置信上界向量 $\boldsymbol{\mu}_1^{\uparrow}$

1: **对于** $t = 1, 2, \ldots, T$ **do**
2: **如果** $t \le \lceil n/k \rceil$**则**
3: $S_t = [tk, ((t+1)k,\ n) \bmod n+1]$
4: **else**
5: 策略优化：
6: $\mathbf{p}_t \leftarrow$ 取最大值$_{\mathbf{p}\in\mathcal{P}_k} M(\boldsymbol{\mu}_t^{\uparrow} \odot \mathbf{p})$
7: 分配抽样：
8: 样本 $S_t \subseteq [n]$ 使得 $|S_t| = k$ 且 $\mathbb{P}(i \in S_t) = p_{t,\,i}$ 对所有 i
9: **结束 if**
10: 反馈：
11: 观察 $u_{t,\,i}$ 对于 $i \in S_t$
12: 置信序列更新：
13: **对于** $i = 1, 2, \ldots, n$ **执行**
14: **如果** $i \in S_t$ **那么** $\mu_{t+1,i}^{\uparrow} \leftarrow$ 更新$(\mu_{t,\,i}^{\uparrow}, u_{t,i})$
15:
16: **else**
17: $\mu_{t+1,i}^{\uparrow} \leftarrow \mu_{t,i}^{\uparrow}$
18: **结束如果**
19: **结束循环**
20: **结束循环**

我们将在下文介绍这些步骤的技术细节。

策略优化（第 6 行）。在每个 t，我们求解一个带约束的优化问题 $\mathbf{p}_t^{\uparrow} = \arg\max_{\mathbf{p}} M(\boldsymbol{\mu}_t^{\uparrow} \odot \mathbf{p})$。我们对 $M(\mathbf{v})$ 关于 $\mathbf{v}$ (A2) 的凹性假设确保该目标函数具有可处理性。然而，由于每个时间步都要解决这个问题，为确保实际可行性，我们需要高效的预言机。我们为所考虑的三个社会福利函数族提供精确的预言机：

定理 5.1。对于以下每类社会福利函数——*WPM*、*Kolm* 和基尼——该优化问题

$$\mathbf{p}_t^{\uparrow} = \arg\max_{\mathbf{p}\in\mathcal{P}} M(\boldsymbol{\mu}_t^{\uparrow} \odot \mathbf{p})$$

都存在一个可精确求解最优解的预言机。

- 对于 *WPM* 和 *Kolm*，最优解具有由标量 λ 参数化的注水形式，该标量可在 $\mathcal{O}(n\log\, n)$ 时间内求得。
- 对于基尼社会福利函数，可以通过基于块的贪心算法在 $\mathcal{O}(kn)$ 时间内计算出最优解。

由于 WPM 和 Kolm 社会福利函数均可微，我们采用 KKT 条件来求解最优解。这些解具有以下闭式解，其参数为 λ；对于 WPM，该参数以乘法形式出现，而对于 Kolm，则以加法形式出现。我们分别在附录 C.1 和附录 C.2中介绍 WPM 和 Kolm 社会福利函数的基于注水的算法。

基尼社会福利函数并非处处可微，因此标准的 KKT分析较为困难。然而，我们证明，对于最优解，使个体 σ 满足 w_i 的排列，可以很容易地确定其与个体 $\sigma(i)$ 配对（命题 C.1）。我们为这一社会福利函数族设计了一种基于块的注水算法（附录 C.3），用于求解一个类似于保序回归中 PAVA 的参数化线性规划问题（Best & Chakravarti，1990）。

为了直观说明这些解的性质，我们考察平等主义社会福利函数和功利主义社会福利函数，这两种设定在三个社会福利函数族中都很常见。在功利主义设定下，最优解是将资源分配给具有最高 $w_i\mu_i^{\uparrow}$ 的 k 个体。在平等主义设定下，最优解为$p_{t,\,i} \propto \left[\lambda/\mu_{t,i}^{\uparrow}\right]_{[0,1]}$，其中 $[x]_{[0,1]}$表示将 x 截断在 0 和 1 之间，而 λ 的取值满足 $\sum_i p_i = k$。该解确保所有个体的 $\mu_i^{\uparrow} p_i$ 值都相同，直到具有最小 $\mu_i^{\uparrow}$ 的个体获得一个 $p_i = 1$资源。

分配抽样（第 8 行）。抽取的集合 $S_t \subseteq [n]$应满足基数约束（$|S_t| = k$）和边际概率约束（$\mathbb{P}(i \in S_t) = p_{t,\,i}$ 对所有 i）。我们采用依赖舍入（Gandhi 等人，2006年），这是一种调查设计中常用的抽样技术，以确保满足这些约束。我们在附录D中给出了依赖舍入算法。

置信度更新（第 15 行）。我们选择序列$\left\{\boldsymbol{\mu}_{t,\,i}^{\uparrow}\right\}_{t\ge 1}$，使其成为 μ_i的有效 CS。我们的更新规则

基于一个广为人知的 CS（Howard 等人，2021年），其表达式为

$$\textsc{Update}(\mu_{t,i}^{\uparrow}, u_{t,i}) = \frac{N_{t,i}\hat{\mu}_{t,i} + u_{t,i}}{N_{t,i}+1} + 1.7\sqrt{\frac{\log(5.2\, n/\delta) + \log\log(2N_{t,i}+1)}{N_{t,i}+1}},$$

在时间 t 之前，资源已分配给个体 i 的次数为 $N_{t,i}$，而 $\hat{\mu}_{t,i}$ 是经验均值。

5.2. 利用利普希茨连续性的结果

我们通过以下上界证明，算法 1能够实现次线性遗憾。

定理 5.2. 设对于所有 i，效用分布 $\mathcal{D}_i$ 均为 *1*-次高斯分布。设 $M(\mathbf{v})$ 关于 ℓ_∞ 范数是 L-利普希茨连续的。设 $a = 3\log(n/\delta)/2$。那么，以概率 $(1-\delta)$，对于所有 $T \in \mathbb{N}$，

$$R(T) \lesssim L\sqrt{\log\log T + \log(2n/\delta)} \cdot \left(n\log\left(\frac{2n}{\delta}\right) + \sqrt{nkT - n\log\left(\frac{2n}{\delta}\right)}\right)$$

因此，选择 $\delta \asymp 1/\sqrt{nkT}$ 后，我们可以得到 $\mathbb{E}[R(T)]$ 的界，其阶为 $\tilde{\mathcal{O}}(L(n+\sqrt{nkT})$。我们利用 $M(\cdot)$ 的利普希茨连续性（A3），用 $\boldsymbol{\mu}_t^{\uparrow} \odot \mathbf{p}_t$ 和 $\boldsymbol{\mu} \odot \mathbf{p}_t$ 给遗憾设定上界。随后，通过证明 $N_{t,i}$ 的任意时刻有效上界，进而得到遗憾的上界。

功利主义福利设定对应于多重选择随机多臂老虎机设定，其中 k 个臂在每个时间步被同时拉动，总奖励为每臂奖励之和。因此，我们立即得到遗憾 $R(T)$ 的如下下界，该下界来自先前文献（Kveton 等人，2015；Cesa-Bianchi 与 Lugosi，2012）：

命题 5.3. 在线 *SWF* 最大化任务的遗憾下界为 $\Omega(\sqrt{nkT})$。

因此，我们的算法在 k 和 T 上是阶数最优的，在 n 上接近最优，与已知界相匹配，至多多对数因子。

在某些情况下，可以将上界收紧许多。例如，当 $k = n$ 时，每个个体在每个时间步都会获得一个资源，因此 $\mathbf{p}_t = \mathbf{p}^* = \mathbf{1}_n$，从而遗憾为零。我们推测，最坏情况遗憾实际上是在某个中间 k 处达到的，而不是在 $k = 1$ 或$k = n$处。直观而言，增加 k 会提高给定时间步中错误分配的代价。然而，对于 $k \gtrsim n/2$，可能的分配空间的直径也会减小，从而导致相对于最优分配的最大可能偏差减小。这导致了一个关于遗憾关于 k的非单调依赖关系。我们在第6节中改变所进行的实验从经验上验证了这一行为。

另一种此类情形是带有 $q = 0$ 的 WPM SWF 设置。最优解的形式为 $\mathbf{p}_t = [\lambda w_i]_{[0,1]}$，且$\lambda$ 的选取使得 $\sum_i p_{t,i} = k$。该最优解与估计值 $\boldsymbol{\mu}_t^{\uparrow}$ 无关，因此在这种情况下遗憾也接近于零。这表明，尽管在最坏情况下遗憾上界是次线性且接近最优的，但仍存在一些非平凡设定可使该上界更紧。我们将对特殊情况以及遗憾上界对 k 的精确依赖关系的进一步分析留待未来工作（第 7 节）。

6. 实验

我们的理论为满足假设（A1）-（A3）的任意社会福利函数提供了适用于 SWF-UCB 的通用近最优界。虽然定理5.2 给出了最坏情况遗憾保证，但并未充分刻画遗憾如何取决于关键问题参数。在实验中，我们研究遗憾对三个问题参数的经验依赖关系：时间范围 T、公平性（幂）参数 q以及分配资源 k。

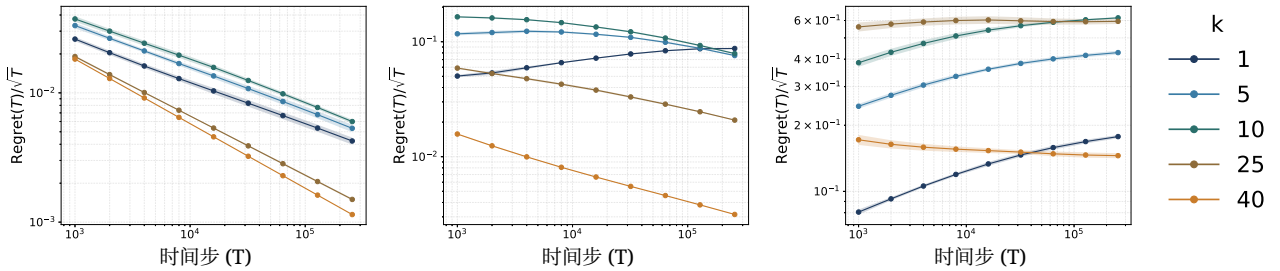
我们针对三类社会福利函数族——WPM、Kolm 和基尼——进行模拟，人口规模为 $n = 50$。个体获得资源后的效用分布为$U_i \sim 0.1 + 0.9X_i$，其中 X_i服从 Beta 分布，其参数为 (α_i, β_i)，并随机选取。每项实验均重复运行 5 次，每次采用随机种子，同时保持臂分布不变并改变随机抽样过程。

我们设置权重向量 $\mathbf{w}$，使得 $w_i \propto 0.9^{i-1}$ 且 $\sum_i w_i = 1$。这相当于在 WPM 和 Kolm 中赋予某些个体更高的权重，而在基尼中则更接近平等主义分配。

变化的时域 T。我们首先验证，我们的算法在社会福利函数和分配机制上的遗憾均实现了预期的 $\sqrt{T}$ 次线性增长。对于所有三种社会福利函数，我们考虑 T在以下范围内 $[10^3, 2.56\cdot 10^5]$ 按对数尺度变化。对于 WPM 和 Kolm，我们将默认幂参数设为 $q = -2$。图 1 绘制了遗憾随时间步数量的变化曲线，其中 k。

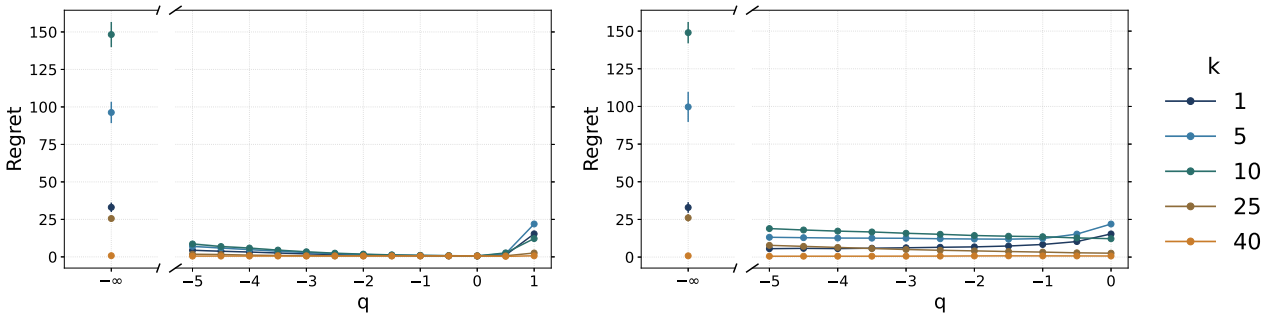
对于所有三种社会福利函数，我们观察到 $R(T)/\sqrt{T}$ 有上界，这表明 $R(T)$ 是 $\mathcal{O}(\sqrt{T})$。遗憾先随着 k的增加而增大，随后减小。然而，不同 k 对应的遗憾相对大小会随着 T的增加而变化。

变化的幂值 q。接下来，我们研究公平性参数如何影响学习难度。我们考察幂值 q 对 WPM 和 Kolm 社会福利函数的影响趋势。所有实验均运行 $T = 10^4$ 个时间步。我们绘制遗憾随 q 变化的曲线，并采用不同的 k 取值，如图2所示。对于



(a) WPM 社会福利函数 *(b)* Kolm 社会福利函数 *(c)* 基尼社会福利函数

图 *1.* 归一化遗憾 $R(T)/\sqrt{T}$ 与时间范围 T 的关系，比较 WPM、Kolm 和基尼社会福利函数。在两个数量级的 T 范围内，归一化遗憾始终有界，这与我们的理论 $\tilde{O}(\sqrt{T})$ 保证一致。



*(a)*WPM 社会福利函数 *(b)* Kolm 社会福利函数

图 *2.* WPM 和 Kolm 社会福利函数随幂值 q 变化的趋势。我们考虑从平等主义（WPM 和 Kolm 的 $q=-\infty$）到功利主义（WPM 的 $q=1$ 和 Kolm 的 $q=0$）的范围。尽管观察到的遗憾平滑变化，但随着 k变化，仍存在显著差异。

对于 WPM 社会福利函数，遗憾先降至 $q=0$，随后升至 $q=1$。当$q=0$时，遗憾变化不大，因为分配概率仅取决于$\mathbf{w}$。对于 Kolm 社会福利函数，较小的 k 值（$k=1$，5）表现为遗憾随 q 增大而略有上升。随着 k 增大，遗憾则转为随 q 增大而略有下降。对于 $q=-\infty$，遗憾先随k增大而上升，随后下降。

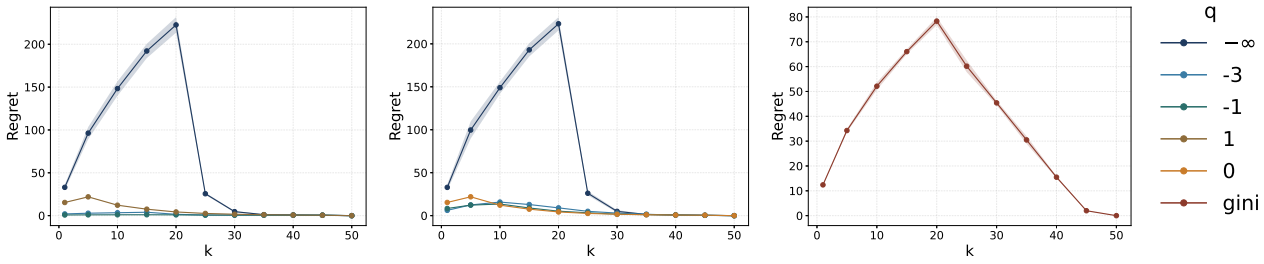
资源数量变化 k。 最后，我们考察资源数量如何影响遗憾变化。我们对所有实验均运行 $T=10^4$ 个时间步。图 3 展示了遗憾随 k 变化的情况，其中 q取不同值。

对于 $q=-\infty$ in WPM 和 Kolm（平等主义情形），遗憾会持续增加，直到 $k=20$，之后在更高的 k 下急剧下降。随着平等主义情形中的资源数量增加，将资源分配给预测效用最低的个体的概率也会增加。然而，一旦将资源以概率 1 分配给实际 μ_i 最低的个体，我们便有 $M(\boldsymbol{\mu}\odot\mathbf{p}_t) = \min_i \mu_i = M(\boldsymbol{\mu}\odot\mathbf{p}_t^{\uparrow})$，因此遗憾不会增加。

对于 WPM 和 Kolm，随着 q 增加，遗憾曲线会趋于平缓。对于基尼，曲线与平等主义情形相似，遗憾先增加至 $k=20$，随后逐渐下降。随着 k 增加，遗憾通常会先增加，然后在 $k=n$ 处降至零。这体现了两者之间的权衡：一方面，随着 k 增加，次优分配的代价会上升；另一方面，对于 $k\gtrsim n/2$，分配空间的直径会减小。前一因素在 k 较小时占主导，而后一因素则在 k 较大时开始起作用。

附录 E 报告了这些采用线性权重衰减的效用分布的实验结果。我们通过实验证实了理论上的 $\tilde{O}(\sqrt{T})$ 保证，如图 4所示。我们还观察到，随着 k（图 6）和 q（图 5）的变化，会出现不同趋势，这表明权重向量也使问题空间更加丰富。

摘要。 综合来看，我们的实验展示了三个关键发现。首先，$\sqrt{T}$缩放在我们三个规范性上彼此不同的社会福利函数族中均得到经验验证。其次，遗憾对 k 存在非单调依赖，表明最高遗憾出现在中等 k。第三，在线福利学习展现出多选择老虎机中不存在的丰富、结构化行为，其相互作用存在于



(a) WPM 社会福利函数　*(b)* Kolm 社会福利函数　*(c)* 基尼社会福利函数

图 *3.* 三种社会福利函数在分配资源数量 k 变化时的趋势。我们观察到，对于平等主义福利，在 $k = 20$ 之后数值急剧下降。随着 q 增加，WPM 和 Kolm 社会福利函数的变化都变得不那么渐进。几何权重与平等主义福利有些相似；对于基尼社会福利函数，我们也观察到在 k 变化时呈现出类似的模式，不过曲线平滑得多。

社会福利函数的选择、社会福利函数的参数以及 k。

7. 讨论

社会福利函数族的用法。本文中的三个社会福利函数族体现了不同的规范性价值和公平观念，我们将在下文讨论它们的潜在用途。

我们将 WPM 族和 Kolm 族放在一起讨论，因为它们的理论形式相似。WPM 族遵循相对不平等厌恶，适用于效用的相对差异很重要的情形。幂 $q \in (-\infty,0)$ 在平等主义福利与纳什福利之间进行插值，而 $q \in (0,1)$ 则在纳什福利与功利主义福利之间进行插值。相比之下，Kolm 族遵循绝对不平等厌恶，因此适用于效用的加性差异很重要的情形。幂 $q \in (-\infty,0)$ 在平等主义福利与功利主义福利之间进行插值。对于 WPM 族和 Kolm 族，权重向量$\mathbf{w}$ 可理解为编码个体在社会福利中的相对重要性。因此，$\mathbf{w}$ 可以编码脆弱性、优先级类别或社会重要性。

当个体之间的位置不平等或基于排名的不平等很重要时，基尼族很有用：福利取决于谁相较于他人处于较差境况，而非效用的绝对差距或相对差距。权重向量 $\mathbf{w}$ 对不同排名赋予社会优先级，使政策制定者能够优先改善排名靠后的个体的状况，而不受其绝对效用的影响。

直接扩展。 我们的在线社会福利函数最大化 5 框架可无缝用于推断应用。定理4.1 提供了有效的福利置信序列，而第5.1 节中的预言机算法则提供了一种高效学习动态策略 $\mathbf{p}_t^{\downarrow}$ 或 $\mathbf{p}_t^{\uparrow}$ 的方法。我们在附录A.1中探讨了三个此类应用——序贯检验、最优停止和策略评估。

虽然我们考虑每个时间步 t 到达固定数量的资源 k，但该框架和分析可以很容易地扩展以适应可变数量的资源k_t 在每个 t。定理4.1 仍然成立，神谕算法也可以在每个时间步使用不同的 k_t 运行。在这种情况下，我们猜想遗憾保证的形式为 $\tilde{O}(n + \sqrt{n \sum_{t=1}^{T} k_t})$。

最后，尽管为简化分析，我们假设未分配资源时的效用为零，但该框架也应能够扩展到未分配效用 $u_{t\,i}^{(0)}$，随机分布的情形。我们将在A.2中对此作进一步说明。

未来工作。在定理4.1中，我们使用$(1-\delta/n)$ 个体平均效用的置信序列来构造$(1-\delta)$ 社会福利的置信序列。通过考虑$(1-\delta_i)$ 平均效用的置信序列（其中 $\sum_i \delta_i = \delta$），并对个体间的置信度进行更精细的分配，可以得到更紧的上界。选择 δ_i 取决于社会福利函数参数，并有可能提高统计和计算效率。

定理5.2 给出了适用于任何满足（A1）-（A3）的社会福利函数的遗憾上界，表明 SWF-UCB 在极小极大意义下近乎最优。然而，对于特定的社会福利函数，可以得到更紧的上界，并明确体现对公平性参数（例如 q 或$\mathbf{w}$）的依赖，从而有可能改进常数和自适应性（注3.2）。

从实证结果来看，遗憾随 k 呈现非单调相变：随着分配的资源增多，遗憾先增大后减小。这种转变取决于社会福利函数参数，表明公平性、不确定性与资源可用性之间存在复杂的相互作用。

虽然我们假设效用独立同分布，但定理 4.1并不限定不确定性的来源，只要求个体效用具有有效的时间一致置信序列。这表明，当此类置信序列适用于具有状态的奖励过程时，无需改变福利层级推断机制，便可能将方法扩展到 restless（Whittle，1988；Wang 等人，2020年）或马尔可夫（Neu 等人，2010年；Ortner 等人，2012）多臂老虎机。

影响声明

本文提出了一个理论框架，利用置信序列和社会福利函数进行在线资源分配与推断。鉴于本文工作的抽象性质，我们预计其不会带来显著的直接社会风险。然而，自动化资源分配系统的建模与开发会引发重要的社会层面考量。我们的建模假设个体效用可以被观测或估计，但在实践中，效用属于潜在构念，可能难以引出、并不稳定或存在争议。根据定义，社会福利函数参数（w， q）体现了关于公平与效率权衡的规范性判断，因此有必要通过价值敏感设计等成熟方法让利益相关者参与其中。此外，保障受保护群体之间的公平性，可能还需要社会福利函数最大化以外的约束。尽管我们的研究以现实世界的分配问题为动机（第1节），但针对如何在具体应用中落实该框架提供详细指导，不在本文当前讨论范围之内。

参考文献

Anantharam, V., Varaiya, P. 和 Walrand, J. 多次选择多臂老虎机问题的渐近高效分配规则——第一部分：独立同分布奖励。*IEEE* 自动控制汇刊，32(11):968–976，1987。doi: 10.1109/TAC.1987.1104491。Atkinson, A. B. 不平等的测量。《经济理论杂志》，2(3):244–263，1970。Barman, S., Khan, A., Maiti, A. 和 Sawarni, A. 多臂老虎机遗憾的公平性与福利量化。载于*AAAI* 人工智能大会论文集，第 37 卷，第 6762–6769 页，2023 年。Best, M. J. 和 Chakravarti, N. 保序回归的活动集算法：一个统一框架。数学规划，47:425–439，1990。网址 https://api.semanticscholar.org/CorpusID:31879613。Cesa-Bianchi, N. 和 Lugosi, G. 组合老虎机。计算机与系统科学杂志，78(5):1404–1422，2012。ISSN 0022-0000。doi: https://doi.org/10 . 1016/j.jcss.2012.01.001。网址 https://www.sciencedirect.com/science/article/pii/S0022000012000219。JCSS 云计算特刊，2011年。Chen, W., Wang, Y. 和 Yuan, Y. 组合多臂老虎机：通用框架与应用。载于 Dasgupta, S. 和 McAllester, D.（编），第 *30* 届国际机器学习会议论文集，第 28 卷，选自机器学习研究论文集，第 151–159 页，美国佐治亚州亚特兰大，2013年6月17–19日。PMLR. 网址 https://proceedings.mlr.press/v28/chen13a.html。Chen, Y., Cuellar, A., Luo, H., Modi, J., Nemlekar, H. 和 Nikolaidis, S. 公平情境多臂老虎机。载于第*19*届自主智能体与多智能体系统国际会议论文集，2020年。Cousins, C. 可证明公平、以福利为中心的机器学习公理化理论。载于 Ranzato, M.、Beygelzimer, A.、Dauphin, Y.、Liang, P. 和 Vaughan, J. W.（编），神经信息处理系统进展，第34卷，第16610–16621页。Curran Associates, Inc.，2021年。网址 https://proceedings.neurips.cc/paper_files/paper/2021/file/8b0bb3eff8c1e5bf7f206125959921d7-Paper.pdf。Cousins, C. 重审公平PAC学习与基数福利公理。载于 Ruiz, F.、Dy, J. 和 van de Meent, J.-W.（编），第*26*届人工智能与统计国际会议论文集，第206卷，机器学习研究论文集，第6422–6442页。PMLR，2023年4月25日至27日。网址 https://proceedings.mlr.press/v206/cousins23a.html。Gai, Y.、Krishnamachari, B. 和 Jain, R. 含未知变量的组合网络优化：具有线性奖励和个体观测的多臂老虎机。*IEEE/ACM*网络汇刊，20(5)：1466–1478，2012。Gandhi, R.、Khuller, S.、Parthasarathy, S. 和 Srinivasan, A. 依赖舍入及其在近似算法中的应用。*ACM*期刊，53(3)：324–360，2006年5月。ISSN 0004-5411。doi：10.1145/1147954.1147956。网址 https://doi.org/10.1145/1147954.1147956。Grafstr¨om, A. 不等概率抽样设计的熵。统计方法学，7(2)：84–97，2010年。Howard, S. R.、Ramdas, A.、McAuliffe, J. 和 Sekhon, J. 通过非负超鞅构造时间一致的切尔诺夫界。概率调查，17(none)：257–317，2020年。doi：10.1214/18-PS321。网址 https://doi.org/10.1214/18-PS321。Howard, S. R.、Ramdas, A.、McAuliffe, J. 和 Sekhon, J. 时间一致的非参数非渐近置信序列。统计年鉴，49(2)：第1055–1080页，2021年。ISSN 00905364, 21688966。网址 https://www.jstor.org/stable/27169410。Joseph, M.、Kearns, M.、Morgenstern, J. H. 和 Roth, A. 学习中的公平性：经典与情境老虎机。神经信息处理系统进展，29，2016年。

Joseph, M.、Kearns, M.、Morgenstern, J.、Neel, S. 和 Roth, A. 面向无限老虎机和情境老虎机的任人唯贤的公平性。载于*2018*年*AAAI/ACM* 人工智能、伦理与社会大会论文集，AIES '18，第158–163页，美国纽约州纽约市，2018年。计算机协会。ISBN 9781450360128。doi: 10.1145/3278721.3278764。网址 https://doi.org/10.1145/3278721.3278764。

Kolm, S.-C. 不平等不等式 I.《经济理论杂志》，12(3):416–442，1976年。

Krishna, A., John, P. G., Barik, A., 和 Tan, V. Y. 随机多臂老虎机的 p-均值遗憾。载于 *AAAI* 人工智能大会论文集，第 39 卷，第 17966–17973 页，2025年。Kveton, B., Wen, Z., Ashkan, A., 和 Szepesvari, C. 随机组合半老虎机的紧遗憾界。载于 人工智能与统计，第 535–543 页。PMLR，2015。Lim, E., Tan, V. Y., 和 Soh, H. 平等主义分配的随机多臂老虎机。 *arXiv* 预印本 *arXiv:2410.05856*，2024。Neu, G., Antos, A., Gy¨orgy, A., 和 Szepesv´ari, C. 具有老虎机反馈的在线马尔可夫决策过程。23，2010年。网址: https://proceedings.neurips.cc/paper_files/paper/2010/file/7bb060764a818184ebb1cc0d43d382aa-Paper.pdf。

Ortner, R., Ryabko, D., Auer, P., 和 Munos, R. 非静止马尔可夫老虎机的遗憾界。载于 国际算法学习理论会议，第 214–228 页。施普林格，2012。

Pardeshi, K. S., Shapira, I., Procaccia, A. D., 和 Singh, A. 社会福利函数的学习。载于 Globerson, A., Mackey, L., Belgrave, D., Fan, A., Paquet, U., Tomczak, J., 和 Zhang, C.（编），神经信息处理系统进展，第 37 卷，第 41733–41766 页。Cur- ran Associates, Inc.，2024。doi: 10.52202/079017-1320。网址: https://proceedings.neurips.cc/paper_files/paper/2024/file/499b12df1531fe8ee0febcf08381f3a4-Paper-Conference.pdf。

Ramdas, A., Ruf, J., Larsson, M., 和 Koolen, W. 可接受的任意时刻有效序贯推断必须依赖于非负鞅。*arXiv* 预印本 *arXiv:2009.03167*，2020年。

Sarkar, D., Pandey, N., 和 Chowdhury, S. R. 重新审视多臂老虎机中的社会福利：UCB（几乎）就是你所需的一切，2025年。网址 https://arxiv.org/abs/2510.21312。

Sawarni, A., Pal, S., 和 Barman, S. 线性多臂老虎机的纳什遗憾保证。载于 Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., 和 Levine, S.（编），神经信息处理系统进展，第 36 卷，第 33288–33318 页。Curran Associates, Inc.，2023年。网址: https://proceedings.neurips.cc/paper_files/paper/2023/file/69bf9fd8d3b7b792b6c8c19149024d22-Paper-Conference.pdf。

Wang, L., Bai, Y., Sun, W. 和 Joachims, T. 随机多臂老虎机中的曝光公平性。载于国际机器学习会议，第 10686–10696 页。PMLR，2021年。

Wang, S., Huang, L. 和 Lui, J. C. S.《Restless-ucb：一种用于在线无休止多臂老虎机的高效低复杂度算法》。33：11878–11889，2020年。网址 https://proceedings.neurips.cc/paper_files/paper/2020/file/89ae0fe22c47d374bc9350ef99e01685-Paper.pdf。

Weymark, J. A. 广义基尼不平等指数。数学社会科学，1(4)：409–430，1981年。

Whittle, P. 无休止多臂老虎机：变化世界中的活动分配。应用概率学杂志，25(A)：287–298，1988年。

A. 进一步讨论

A.1. 推断应用

现在，我们讨论在线推断应用。这是 CS 提升定理（定理4.1）的自然副产品。我们列举基于社会福利函数的分配场景中的一些应用及其解读。这些应用说明，福利层级置信序列如何在自适应数据收集中实现有效推断。

我们注意到，由于定理 4.1 的一般性，其他序贯推断任务（Howard 等人，2020年）也可以使用这一框架加以处理。此外，在 $\mathbf{p}_t$ 动态学习的情况下，我们在第5.1节提出的预言机算法可用于高效地学习它们。

我们从序贯推断的角度重新表述定理 4.1 。

推论 A.1。 设 $[\boldsymbol{\mu}_t^{\downarrow}, \boldsymbol{\mu}_t^{\uparrow}]$ 是一个 $(1-\delta)$ 关于 $\boldsymbol{\mu}$ 的 *CS*。

1. （固定分配策略 $\mathbf{p}$）：设 $\mathbf{p}$ 为已知的固定分配策略。设 $W^{\downarrow}{}_{\mathbf{p}\,t} = M(\boldsymbol{\mu}_t^{\downarrow} \odot \mathbf{p})$ 和 $W^{\uparrow}{}_{\mathbf{p}\,t} = M(\boldsymbol{\mu}_t^{\uparrow} \odot \mathbf{p})$。那么，$\left\{[W^{\downarrow}{}_{\mathbf{p}\,t}, W^{\uparrow}{}_{\mathbf{p}\,t}]\right\}_{t\geq 1}$ 是 $W_{\mathbf{p}} = M(\boldsymbol{\mu} \odot \mathbf{p})$ 的一个 $(1-\delta)$ 置信序列。

2. （动态分配策略 $\mathbf{p}_t$）：设

$$\mathbf{p}_t^{\downarrow} = \arg\max_{\mathbf{p}\in\mathcal{P}_k} M(\boldsymbol{\mu}_t^{\downarrow} \odot \mathbf{p}) \quad \textit{and} \quad \mathbf{p}_t^{\uparrow} = \arg\max_{\mathbf{p}\in\mathcal{P}_k} M(\boldsymbol{\mu}_t^{\uparrow} \odot \mathbf{p}).$$

Let $W_t^{\downarrow} = M(\boldsymbol{\mu}_t^{\downarrow} \odot \mathbf{p}_t^{\downarrow})$ *and* $W_t^{\uparrow} = M(\boldsymbol{\mu}_t^{\uparrow} \odot \mathbf{p}_t^{\uparrow})$. *Then* $\left\{[W_t^{\downarrow}, W_t^{\uparrow}]\right\}_{t\geq 1}$ *is a* $(1-\delta)$ *CS for* $W^* = M(\boldsymbol{\mu} \odot \mathbf{p}^*)$.

第 5.1 节中提出的预言机算法可用于自适应地学习 $\mathbf{p}_t^{\downarrow}$ 或 $\mathbf{p}_t^{\uparrow}$，且效率很高。现在我们讨论如何将这些置信序列用于推断。

序贯检验。 此应用包含两个组成部分的用例：

1. 固定策略$\mathbf{p}$：在这种情况下，决策者希望检验分配策略$\mathbf{p}$ 是否超过某个福利值 W^*。这里的随机性仅来自效用样本。待检验的假设为

$$H_0 : M(\boldsymbol{\mu} \odot \mathbf{p}) < W_0, \quad \text{versus} \quad H_1 : M(\boldsymbol{\mu} \odot \mathbf{p}) > W_0.$$

根据推论A.1的第一部分，我们得到检验：在 $\tau = \inf\left\{t \text{ 处拒绝原假设：} \ W^{\downarrow}{}_{\mathbf{p}\,t} > W_0\right\}$。由于置信序列具有时间一致性，我们有

$$\mathbb{P}_{H_0}(\exists t : W_{\mathbf{p},t}^{\downarrow} > W_0) \leq \delta,$$

这意味着该检验是有效的。

2. 动态策略$\mathbf{p}_t$：在这种情况下，决策者希望了解最优分配策略$\mathbf{p}^*$ 是否实现了超过某一数值 W^*的福利。此处，策略与福利均依赖于数据。待检验的假设为

$$H_0 : M(\boldsymbol{\mu} \odot \mathbf{p}^*) < W_0, \quad \text{versus} \quad H_1 : M(\boldsymbol{\mu} \odot \mathbf{p}^*) > W_0.$$

根据推论的第二部分A.1，我们得到检验：在 $\tau = \inf\left\{t \text{ 处拒绝原假设：} \ W_t^{\downarrow} > W_0\right\}$。由于置信序列具有时间一致性，我们有

$$\mathbb{P}_{H_0}(\exists t : W_t^{\downarrow} > W_0) \leq \delta,$$

这意味着该检验是有效的。

由于该界限在所有时刻都成立，因此这种检验可以自适应地进行：当收集到足够证据以拒绝原假设时，即可停止实验（Ramdas 等人，2020年）。这一思路也可用于检验分配所达到的值是否低于某个阈值 W_*，方法是考虑上置信序列 $\left\{M\left(\boldsymbol{\mu}_t^{\uparrow} \odot \mathbf{p}_t^{\uparrow}\right)\right\}_{t\geq 1}$。

最优停止。考虑这样一种决策场景：在测试阶段动态学习策略。当当前分配所实现的社会福利超过某个值 W_0 时，测试阶段即告结束，随后部署该策略。与序贯检验不同，这里的目标并非拒绝假设，而是决定何时从探索转向部署。

令 $\mathcal{E}=\left\{\forall t\geq 1\colon\ \boldsymbol{\mu}_t^{\downarrow}\preceq\boldsymbol{\mu}\right\}$，其成立的概率至少为 $(1-\delta)$。在 $\mathcal{E}$ 上，由于 M 是单调函数，对所有 t，

$$M(\boldsymbol{\mu}\odot\mathbf{p}_t^{\downarrow})\geq M(\boldsymbol{\mu}_t^{\downarrow}\odot\mathbf{p}_t^{\downarrow})=W_t^{\downarrow}$$

因此，我们在此情况下的停止保证为：在 $\tau=\inf\left\{t\colon\ W_t^{\downarrow}>W_0\right\}$，

$$\mathbb{P}(M(\boldsymbol{\mu}\odot\mathbf{p}_\tau^{\downarrow})>W_0)\geq 1-\delta.$$

停止实验后，我们部署 **策略**$_\tau^{\downarrow}$。

策略评估。 我们的方法也可用于比较两个策略 **策略**$^{(1)}$ 和**策略**$^{(2)}$的社会福利。这使我们能够比较这些策略，并以自适应方式从中选择较优者进行部署。

有趣的是，只要所产生的采样过程确保覆盖所有个体（例如，对所有 i，$p_{t,i}\geq\gamma>0$），就可以使用这两个策略中的任意一个（或它们的组合）来收集数据，因为我们只需要个体效用$\mu_{t,i}$的置信序列。根据推论 A.1 的第一部分，我们为 $\mathbf{p}^{(1)}$ 和 $\mathbf{p}^{(2)}$构造 $(1-\delta/2)$ 福利置信序列，分别记为$\left\{[W^{\downarrow}_1(\,)_t, W^{\uparrow}(1)_t]\right\}_{t\geq1}$和 $\left\{[W^{\downarrow}_2(\,)_t, W^{\uparrow}(2)_t]\right\}_{t\geq1}$。当 $W^{\downarrow}(i)_t > W^{\uparrow}(j)_t$时，以置信度 $(1-\delta)$，可以选择策略 i而非策略 j进行部署。

A.2. 扩展：未分配时的随机效用

在第 3–6 节中，我们假设个体未获得资源时的效用为零。现在讨论如何将基于置信序列的框架扩展到未分配时随机观测到的效用。

令 $u_t^{(1)},_i \overset{iid}{\sim}\mathcal{D}_i^{(1)}$ 和 $u_t^{(0)},_i \overset{iid}{\sim}\mathcal{D}_i^{(0)}$ 分别表示时刻 t 个体 i 获得资源和未获得资源时的随机效用。令 $\mu_i^{(1)}$ 和 $\mu_i^{(0)}$ 分别表示为个体 i 分配资源和不分配资源时的平均效用。因此，分配策略$\mathbf{p}$的事前效用为

$$\overline{\boldsymbol{\mu}}(\mathbf{p})=\boldsymbol{\mu}^{(1)}\odot\mathbf{p}+\boldsymbol{\mu}^{(0)}\odot(\mathbf{1}_n-\mathbf{p}),$$

因此，社会福利为 $M(\boldsymbol{\mu}(\mathbf{p}))$。由此可见，相较于不分配资源，为个体 i 分配资源可带来 $\mu_i^{(1)}-\mu_i^{(0)}$ 的有效效用增益。

扩展置信序列。 利用观测值，我们可以构造 $(1-\delta/2n)$ 关于 $\mu_i^{(1)}$ 和 $\mu_i^{(0)}$ 的双侧置信序列，分别记为 $\left\{[\mu_{t,i}^{(1)\uparrow},\mu_{t,i}^{(1)\downarrow}]\right\}_{t\geq1}$ 和 $\left\{[\mu_{t,i}^{(0)\downarrow},\mu_{t,i}^{(0)\uparrow}]\right\}_{t\geq1}$。根据并集界，可得以下关于 $\boldsymbol{\mu}(\mathbf{p})$ 的 $(1-\delta)$ 置信序列：

—

$$\left\{\left[\left(\boldsymbol{\mu}_t^{(1)\downarrow}-\boldsymbol{\mu}_t^{(0)\uparrow}\right)\mathbf{p}+\boldsymbol{\mu}_t^{(0)\downarrow},\left(\boldsymbol{\mu}_t^{(1)\uparrow}-\boldsymbol{\mu}_t^{(0)\downarrow}\right)\mathbf{p}+\boldsymbol{\mu}_t^{(0)\uparrow}\right]\right\}_{t\geq1}.$$

如果令

$$\mathbf{p}_t^{\downarrow}=\arg\max_{\mathbf{p}\in\mathcal{P}_{\leq k}}\left(\boldsymbol{\mu}_t^{(1)\downarrow}-\boldsymbol{\mu}_t^{(0)\uparrow}\right)\mathbf{p}+\boldsymbol{\mu}_t^{(0)\downarrow},\quad\text{and}$$
$$\mathbf{p}_t^{\uparrow}=\arg\max_{\mathbf{p}\in\mathcal{P}_{\leq k}}\left(\boldsymbol{\mu}_t^{(1)\uparrow}-\boldsymbol{\mu}_t^{(0)\downarrow}\right)\mathbf{p}+\boldsymbol{\mu}_t^{(0)\uparrow},$$

于是我们有如下（$1-\delta$）CS：

$$\left\{\left[\left(\boldsymbol{\mu}_t^{(1)\downarrow}-\boldsymbol{\mu}_t^{(0)\uparrow}\right)\mathbf{p}_t^{\downarrow}+\boldsymbol{\mu}_t^{(0)\downarrow},\left(\boldsymbol{\mu}_t^{(1)\uparrow}-\boldsymbol{\mu}_t^{(0)\downarrow}\right)\mathbf{p}_t^{\uparrow}+\boldsymbol{\mu}_t^{(0)\uparrow}\right]\right\}_{t\geq1}.$$

由于 $M(\cdot)$ 具有单调性，并按照定理 4.1 中类似的一系列不等式推导过程，我们得到 $(1-\delta)$ 关于 $M(\boldsymbol{\mu}(\mathbf{p}))$ CS

$$\left\{\left[M\left(\left(\boldsymbol{\mu}_t^{(1)\downarrow}-\boldsymbol{\mu}_t^{(0)\uparrow}\right)\mathbf{p}_t^{\downarrow}+\boldsymbol{\mu}_t^{(0)\downarrow}\right), M\left(\left(\boldsymbol{\mu}_t^{(1)\uparrow}-\boldsymbol{\mu}_t^{(0)\downarrow}\right)\mathbf{p}_t^{\uparrow}+\boldsymbol{\mu}_t^{(0)\uparrow}\right)\right]\right\}_{t\geq 1}.$$

扩展预言机算法。我们的预言机算法必须求解如下形式的问题

$$\mathbf{p}_t^{\uparrow}=\arg\max_{\mathbf{p}\in\mathcal{P}_{\leq k}} M\left(\left(\boldsymbol{\mu}_t^{(1)\uparrow}-\boldsymbol{\mu}_t^{(0)\downarrow}\right)\mathbf{p}+\boldsymbol{\mu}_t^{(0)\uparrow}\right).$$

对于某个个体 i，如果 $\boldsymbol{\mu}_{t,i}^{(0)\downarrow} > \boldsymbol{\mu}_{t,i}^{(1)\uparrow}$，我们可以有很高的置信度推断，不分配资源会给他们带来更高的效用。在这种情况下，以很高的概率，向个体 i 分配资源只会降低福利目标，而单调性意味着设置 $p_i=0$ 是最优的。因此，可以将这类个体排除在分配之外，只针对其余个体求解优化问题。进而，这使得分配给人口的资源可能少于 k，从而使我们针对$\mathbf{p}$ 的可行集变为

$$\mathcal{P}_{\leq k}=\left\{\mathbf{p}\in[0,1]^n : \sum_i p_i = s, s\in\mathbb{N}, s\leq k\right\}$$

B. 证明

B.1. 命题3.1的证明

证明。单调性与凹性：对于 M_{WPM}、 M_{Kolm} 和M_{Gini}，社会福利公理保证了单调性。现在我们证明这三个社会福利函数族的凹性。

- *WPM*：我们首先注意到，对于 $q=-\infty$, $M_{\text{WPM}}(\mathbf{v}) = \min_i v_i$ 关于$\mathbf{v}$是凹函数，因为对于两个有效向量$\mathbf{v}_1$ 和$\mathbf{v}_2$,

$$\min_i \lambda v_{1,i}+(1-\lambda)v_{2,i} \geq \lambda \min_i v_{1,i}+(1-\lambda)\min_j v_{2,j}$$

对于 $q\in(-\infty,1)$，我们将 WPM 社会福利函数表示为

$$M_{\text{WPM}}(\mathbf{v};\mathbf{w};q)=\left(\sum_i\left(w_i^{1/q}v_i\right)^q\right)^{1/q}$$

设$\mathbf{v}_1$ 和$\mathbf{v}_2$ 是两个有效向量。应用反向闵可夫斯基不等式，可得

$$\begin{aligned}M_{\text{WPM}}(\lambda\mathbf{v}_1+(1-\lambda)\mathbf{v}_2;\mathbf{w},q) &= \left(\sum_i\left(\lambda w_i^{1/q}v_{1,i}+(1-\lambda)w_i^{1/q}v_{2,i}\right)^q\right)^{1/q}\\ &\geq \left(\sum_i\left(\lambda w_i^{1/q}v_{1,i}\right)^q\right)^{1/q}+\left(\sum_i\left((1-\lambda)w_i^{1/q}v_{w,i}\right)^q\right)^{1/q}\\ &= \lambda M_{\text{WPM}}(\mathbf{v}_1;\mathbf{w},q)+(1-\lambda)M_{\text{WPM}}(\mathbf{v}_2;\mathbf{w},q)\end{aligned}$$

- *Kolm*：由于 $q=-\infty$是平等主义情形，而 $q=0$ 是功利主义情形，已通过 WPM 社会福利函数证明它们是凹函数。对于 $q\in(-\infty,0)$， M_{Kolm} 的凹性源于 log-sum-exp函数是凸函数，且其前面乘有 $1/q$ （一个负值）。

- 基尼：基尼社会福利函数也可以表示为

$$M_{\text{Gini}}(\mathbf{v};\mathbf{w})=\min_\sigma\sum_i w_{\sigma(i)}v_i,$$

其中， σ 是集合 $\{1,\ldots,n\}$的一个排列。由于我们取的是一组仿射函数的最小值，因此所得函数是凹函数。

利普希茨连续性：我们现在证明社会福利函数族的利普希茨连续性。根据中值定理，对于两个有效向量$\mathbf{v}_1$和$\mathbf{v}_2$,，存在一个向量$\mathbf{x}$，使得

$$M(\mathbf{v}_1) - M(\mathbf{v}_2) = \nabla M(\mathbf{x}) \cdot (\mathbf{v}_1 - \mathbf{v}_2)$$
$$|M(\mathbf{v}_1) - M(\mathbf{v}_2)| = |\nabla M(\mathbf{x}) \cdot (\mathbf{v}_1 - \mathbf{v}_2)| \overset{(i)}{\leq} \|\nabla M(\mathbf{x})\|_1 \cdot \|\mathbf{v}_1 - \mathbf{v}_2\|_\infty ,$$

其中 (i) 使用 H¨older 的不等 式。因此，为了证明其关于 ℓ_∞ 范数的利普希茨连续性，我们给出 $\|\nabla M(\mathbf{x})\|$ 的界 1

- *WPM*：令 $v_i \in [v_{\min}\ v_{\max}]$，其中 $v_{\max} \geq v_{\min} > 0$。于是，对于 $q \in (-\infty, 0) \cup (0, 1]$,

$$M_{\text{WPM}}(\mathbf{v}; \mathbf{w}, q) = \left(\sum_i w_i v_i^q \right)^{1/q}$$
$$\frac{\partial}{\partial v_i} M_{\text{WPM}}(\mathbf{v}; \mathbf{w}, q) = \frac{1}{q} \cdot \left(\sum_i w_i v_i^q \right)^{1/q-1} \cdot q w_i v_i^{q-1}$$
$$= M_{\text{WPM}}(\mathbf{v}; \mathbf{w}, q) \cdot \frac{w_i v_i^q}{\sum_i w_i v_i^q} \cdot \frac{1}{v_i}$$
$$\implies \sum_i \left| \frac{\partial}{\partial v_i} M_{\text{WPM}}(\mathbf{v}; \mathbf{w}, q) \right| \leq \frac{M}{v_{\min}} \leq \frac{v_{\max}}{v_{\min}}.$$

对于 $q = 0$，我们有

$$M_{\text{WPM}}(\mathbf{v}; \mathbf{w}, 0) = \prod_i v_i^{w_i}$$
$$\frac{\partial}{\partial v_i} = \prod_i v_i^{w_i} \cdot \frac{w_i}{v_i} = M(\mathbf{v}; \mathbf{w}, 0) \cdot \frac{w_i}{v_i}$$
$$\implies \sum_i \left| \frac{\partial}{\partial v_i} M_{\text{WPM}}(\mathbf{v}; \mathbf{w}, q) \right| \leq \frac{M}{v_{\min}} \leq \frac{v_{\max}}{v_{\min}}.$$

对于 $q = -\infty$，我们注意到所有次梯度之和的上界为 1，显然小于 $v_{\max}/v_{\min}$。因此， $v_{\max}/v_{\min}$ 是利普希茨常数。

此外　另外注意，对于 $q = 1, M_{\text{WPM}}(\mathbf{v}; \mathbf{w}, 1), = \sum_i w_i v_i$ 有 $\|\nabla M_{\text{WPM}}(\mathbf{v}; \mathbf{w}, 1)\|_1 = \sum_i w_i = 1.$

- *Kolm*：平等主义和功利主义的情况已在 WPM 情况中讨论。对于$q \in (-\infty, 0)$，我们有

$$M_{\text{Kolm}}(\mathbf{v}; \mathbf{w}, q) = \frac{1}{q} \log \left(\sum_i \exp(q v_i) \right)$$
$$\frac{\partial}{\partial v_i} M_{\text{Kolm}}(\mathbf{v}; \mathbf{w}, q) = \frac{1}{q} \cdot \frac{1}{\sum_i w_i \exp(q u_i)} \cdot w_i q \exp(q u_i) = \frac{w_i \exp(q u_i)}{\sum_i w_i \exp(q u_i)}$$
$$\implies \sum_i \left| \frac{\partial}{\partial v_i} M_{\text{Kolm}}(\mathbf{v}; \mathbf{w}, q \right| = 1$$

- 基尼：我们考虑基尼社会福利函数的另一种表达式：

$$M_{\text{Gini}}(\mathbf{v}; \mathbf{w}) = \min_\sigma \sum_i w_{\sigma(i)} v_i.$$

我们注意到，这是仿射函数 $f_\sigma(\mathbf{v}) = \sum_i w_{\sigma_i} v_i$ 的最小值，其中每个函数都具有 $\|\nabla f_\sigma\|_1 = \sum_i w_i = L$。因此， M_{Gini} 的次梯度满足 $\|\partial M_{\text{Gini}}(\mathbf{v};\mathbf{w})\|_1 \leq L$。

□

B.2. 定理4.1的证明

证明。 对于所有 $i \in [n]$，令 $\left\{[\mu_{t,i}^{\downarrow}, \mu_{t,i}^{\uparrow}]\right\}_{t\geq 1}$ 为 i 的一个（$1-\delta/n$）置信序列。也就是说，

$$\mathbb{P}\left(\exists t \in \mathbb{N} : \mu_i \notin [\mu_{t,i}^{\downarrow}, \mu_{t,i}^{\uparrow}]\right) \leq \frac{\delta}{n}$$

由并集界可知，对于序列 $\left\{[\boldsymbol{\mu}_t^{\downarrow}, \boldsymbol{\mu}_t^{\uparrow}]\right\}_{t\geq 1}$

$$\mathbb{P}\left(\exists t \in \mathbb{N} : \boldsymbol{\mu} \notin [\boldsymbol{\mu}_t^{\downarrow}, \boldsymbol{\mu}_t^{\uparrow}]\right) \leq \delta,$$

其中 $\mathbf{x} \in [\mathbf{x}_t^{\downarrow}, \mathbf{x}_t^{\uparrow}]$ 表示 $x_i \in [x_{t,i}^{\downarrow}, x_{t,i}^{\uparrow}]$ 对所有 $i \in [n]$均成立。

因此，以概率 $(1-\delta)$，我们有

$$M(\boldsymbol{\mu} \odot \mathbf{p}^*) \overset{(i)}{\geq} M(\boldsymbol{\mu} \odot \mathbf{p}_t^{\downarrow}) \overset{(ii)}{\geq} M(\boldsymbol{\mu}_t^{\downarrow} \odot \mathbf{p}_t^{\downarrow}), \quad \text{and}$$
$$M(\boldsymbol{\mu} \odot \mathbf{p}^*) \overset{(iii)}{\leq} M(\boldsymbol{\mu}_t^{\uparrow} \odot \mathbf{p}^*) \overset{(iv)}{\leq} M(\boldsymbol{\mu}_t^{\uparrow} \odot \mathbf{p}_t^{\uparrow}).$$

(i) 是因为 $\mathbf{p}^*$ 对于 $\boldsymbol{\mu}$ 是最优的，而 (ii) 则源于置信序列界以及 $M(\cdot)$ 的单调性。(iii) 源于置信序列界以及 $M(\cdot)$ 的单调性，而 (iv) 是因为 $\mathbf{p}_t^{\uparrow}$ 对于 $\boldsymbol{\mu}_t^{\uparrow}$ 是最优的。

这意味着

$$\mathbb{P}(\exists t \in \mathbb{N} : M(\boldsymbol{\mu} \odot \mathbf{p}^*) \notin [M(\boldsymbol{\mu}_t^{\downarrow} \odot \mathbf{p}_t^{\downarrow}), M(\boldsymbol{\mu}_t^{\uparrow} \odot \mathbf{p}_t^{\uparrow})]) \leq \delta$$

□

B.3. 定理5.2 的证明

证明。 令 $\left\{\boldsymbol{\mu}_t^{\uparrow}\right\}_{t\geq 1}$ 为 $1-\delta/2)$ 关于 $\boldsymbol{\mu}$的置信序列。根据定理 4.1，我们知道，以概率 $(1-\delta/2)$

$$M(\boldsymbol{\mu} \odot \mathbf{p}^*) \overset{(iii)}{\leq} M(\boldsymbol{\mu}_t^{\uparrow} \odot \mathbf{p}^*) \overset{(iv)}{\leq} M(\boldsymbol{\mu}_t^{\uparrow} \odot \mathbf{p}_t^{\uparrow}).$$

因此，我们以概率（$1-\delta/2$）有，

$$\begin{aligned} R(T) &= \sum_{t=1}^{T} M(\boldsymbol{\mu} \odot \mathbf{p}^*) - M(\boldsymbol{\mu} \odot \mathbf{p}_t^{\uparrow}) \\ &\leq \sum_{t=1}^{T} M(\boldsymbol{\mu}_t^{\uparrow} \odot \mathbf{p}^*) - M(\boldsymbol{\mu} \odot \mathbf{p}_t^{\uparrow}) \\ &\leq \sum_{t=1}^{T} M(\boldsymbol{\mu}_t^{\uparrow} \odot \mathbf{p}_t^{\uparrow}) - M(\boldsymbol{\mu} \odot \mathbf{p}_t^{\uparrow}) \end{aligned}$$

如果现在假设 M 关于 $\|\cdot\|_\infty$（最大值）是 L-Lipschitz 的，我们就得到

$$\begin{aligned} R(T) &\leq \sum_{t=1}^{T} M(\boldsymbol{\mu}_t^{\uparrow} \odot \mathbf{p}_t^{\uparrow}) - M(\boldsymbol{\mu} \odot \mathbf{p}_t^{\uparrow}) \\ &\leq \sum_{t=1}^{T} L \left\| \boldsymbol{\mu}_t^{\uparrow} \odot \mathbf{p}_t^{\uparrow} - \boldsymbol{\mu} \odot \mathbf{p}_t^{\uparrow} \right\|_\infty \end{aligned}$$

因此，我们有界　以 $\boldsymbol{\mu}$ 的置信序列表示遗憾。由 H¨older 不等式可得

$$\begin{aligned}\sum_{t=1}^{T} L\|\boldsymbol{\mu}_t^{\uparrow}\odot\mathbf{p}_t^{\uparrow}-\boldsymbol{\mu}\odot\mathbf{p}_t^{\uparrow}\|_\infty &\le \sum_{t=1}^{T} L\|\boldsymbol{\mu}_t^{\uparrow}\odot\mathbf{p}_t^{\uparrow}-\boldsymbol{\mu}\odot\mathbf{p}_t^{\uparrow}\|_1\\ &= L\sum_{t=1}^{T}\sum_{i=1}^{n}|\mu_{t,i}^{\uparrow}-\mu_i|p_{t,i}^{\uparrow}\\ &\lesssim L\sum_{t=1}^{T}\sum_{i=1}^{n}p_{t,i}\sqrt{\frac{\log\log(N_{t,i}+1)+\log(n/\delta)}{N_{t,i}+1}}\\ &\lesssim L\sqrt{\log\log T+\log(n/\delta)}\underbrace{\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{p_{t,i}}{\sqrt{N_{t,i}+1}}}_{S}\end{aligned}$$

现在我们来界定项 S。注意到 $N_{t,\,i}$ 是伯努利随机变量 $N_{t,\,i}=\sum_{s=1}^{t}X_{s,\,i}$之和，其中$X_{s,\,i}\sim$ 伯$(p_{s,\,i})$。因此，我们得到以下任意时刻有效保证（由（Howard 等人，2020年）中的主定理推出）：

$$\mathbb{P}\left(\exists t\in\mathbb{N}:\sum_{s=1}^{t}(p_{s,i}-X_{s,i})\ge a+b\sum_{s=1}^{t}p_{s,i}\right)\le\exp\left(-\frac{2ab}{1+b}\right)$$

令 $b=0.5$，且 $a=(3/2)\cdot\log(2n/\delta)$，使得上述上界为 $\delta/2n$。因此，我们有：以概率 $(1-\delta/2n)$，对所有 $t\in\mathbb{N}$，

$$\begin{aligned}&\sum_{s=1}^{t}(p_{s,i}-X_{s,i})\le a+b\sum_{s=1}^{t}p_{s,i}\\ \implies\ &N_{t,i}=\sum_{s=1}^{t}X_{s,i}\ge 0.5\sum_{s=1}^{t}p_{s,i}-a\end{aligned}$$

现在我们可以考虑两种情况：

1. $\sum_{s=1}^{t}p_{s,\,i}\le 2a$：这里，我们可以说

$$\sum_{s=1}^{t}\frac{p_{s,i}}{\sqrt{N_{s,i}+1}}\le\sum_{s=1}^{t}p_{s,i}\le 2a,$$

　这是一个常数。因此，在这种情况下，遗憾有一个常数上界。

2. $\sum_{s=1}^{t}p_{s,\,i}>2a$：设此结论对 $t\ge t_0$ 成立。于是，以概率 $(1-2\delta)$，

$$\begin{aligned}\sum_{s=t_0}^{t}\frac{p_{s,i}}{\sqrt{N_{s,i}+1}}&\le\sum_{s=t_0}^{t}\frac{1}{\sqrt{0.5(\sum_{k=1}^{s}p_{k,i})-a}}\\ &\le 2\sqrt{2}\cdot\sqrt{\left(\sum_{s=1}^{t}p_{s,i}\right)-2a}\end{aligned}$$

因此，我们有

$$\begin{aligned}\sum_{t=1}^{T}\frac{p_{s,i}}{\sqrt{N_{s,i}+1}}&\lesssim a+\sqrt{\left(\sum_{t=1}^{T}p_{t,i}\right)-a}\\ \implies\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{p_{t,i}}{\sqrt{N_{t,i}+1}}&\lesssim na+\sum_{i=1}^{n}\sqrt{\left(\sum_{t=1}^{T}p_{t,i}\right)-a}\\ &\le na+\sqrt{nkT-na}\end{aligned}$$

因此，我们有

$$R(T) \lesssim L\sqrt{\log\log T + \log\left(\frac{2n}{\delta}\right)}\left[n\log\left(\frac{2n}{\delta}\right) + \sqrt{nkT - n\log\left(\frac{2n}{\delta}\right)}\right]$$

因此，我们得到如下形式的速率： $\mathcal{O}(\sqrt{\log\log T + \log n}(n\log n + \sqrt{nkT}))$。 □

C. 定理的详细内容5.1

C.1. WPM 社会福利函数

C.1.1. 推导所提解

我们的目标函数可表述为

$$\begin{aligned}\text{maximize } M_{\text{WPM}}(\mathbf{u}\odot\mathbf{p};\mathbf{w},q) &= \left(\sum_i w_i(u_ip_i)^q\right)^{1/q}\\ \text{s.t. } \sum_i p_i &= k\\ p_i - 1 &\le 0\\ -p_i &\le 0\end{aligned}$$

首先，考虑 p_i 不受限制的情况。对 p_i 求导后，在最优点处，我们希望

$$\begin{aligned}\frac{\partial}{\partial p_i}M_{\text{WPM}} &= \lambda'\\ \lambda' &= \frac{1}{q}\cdot\left(\sum_i w_i(u_ip_i)^q\right)^{(1/q-1)}\cdot w_iu_i^q\cdot qp_i^{q-1}\\ &= M_{\text{WPM}}(\mathbf{u}\odot\mathbf{p};\mathbf{w},q)\cdot\frac{w_iu_i^q}{\sum_i w_i(u_ip_i)^q}p_i^{q-1}\\ p_i &= \lambda(w_iu_i^q)^{1/(1-q)},\end{aligned}$$

其中

$$\lambda = \left(\frac{\lambda'\cdot\sum_i w_i(u_ip_i^q)}{M_{\text{WPM}}(\mathbf{u}\odot\mathbf{p};\mathbf{w},q}\right)^{1/(q-1)}$$

因此，我们提出的解为

$$p_i = \left[\lambda(w_iu_i^q)^{1/(1-q)}\right]_{[0,1]},$$

其中 p_i 被限制在 0 和 1 之间，并选择 λ 使得 $\sum_i p_i = k$。

C.1.2. 通过 KKT 条件证明最优性

目标函数的拉格朗日函数为

$$\mathcal{L}(\mathbf{p},\boldsymbol{\alpha},\boldsymbol{\beta},\gamma) = -M_{\text{WPM}}(\mathbf{u}\odot\mathbf{p};\mathbf{w},q) + \sum_i\alpha_i(p_i-1) + \sum_i\beta_i(-p_i) + \gamma(\sum_i p_i - k)$$

回顾一下，KKT 条件要求如下：

1. 平稳性：对于所有 i,

$$\frac{\partial}{\partial p_i}\mathcal{L} = -M_{\text{WPM}}(\mathbf{u}\odot\mathbf{p};\mathbf{w},q)\cdot\frac{w_i(u_ip_i)^q}{\sum_i w_i(u_ip_i)^q}\cdot\frac{1}{p_i} + \alpha_i - \beta_i + \gamma = 0$$

2. 原始可行性： $\sum_i p_i = k$，且对于所有 i， $p_i \in [0,1]$。

3. 对偶可行性： $\alpha_i \geq 0, \beta_i \geq 0$ 对所有 i均成立。

4. 互补松弛： $\alpha_i(1-p_i) = 0$ 且 $\beta_i p_i = 0$ 对所有 i。

我们在方程 **??** 中选择所提出的解 $\mathbf{p}^*$，使得 $0 < p_i \leq 1$，且 $\sum_i p_i = k$。这立即表明该解是原始可行的；此外，根据互补松弛，我们知道对所有 i， $\beta_i = 0$。

现在我们考虑关于 α_i的两种情况：

(i) 若 $p_i < 1$，互补松弛要求 $\alpha_i = 0$。在这种情况下，我们有

$$\begin{aligned} p_i^* &= \lambda(w_i u_i^q)^{1/(1-q)} \\ u_i p_i^* &= \lambda(w_i u_i)^{1/(1-q)} \end{aligned}$$

平稳性要求

$$\begin{aligned} \frac{\partial}{\partial p_i}\mathcal{L} = -M_{\text{WPM}}(\mathbf{u}\odot\mathbf{p};\mathbf{w},q)\cdot\frac{w_i(u_ip_i)^q}{\sum_i w_i(u_ip_i)^q}\cdot\frac{1}{p_i} + \alpha_i - \beta_i + \gamma &= 0 \\ M_{\text{WPM}}(\mathbf{u}\odot\mathbf{p};\mathbf{w},q)\cdot\frac{w_i(u_ip_i)^q}{\sum_i w_i(u_ip_i)^q}\cdot\frac{1}{p_i} &= \gamma \\ \frac{M_{\text{WPM}}(\mathbf{u}\odot\mathbf{p};\mathbf{w},q)}{\sum_i w_i(u_ip_i)^q}u_iw_i(u_ip_i)^{q-1} &= \gamma \\ \frac{M_{\text{WPM}}(\mathbf{u}\odot\mathbf{p};\mathbf{w},q)}{\sum_i w_i(u_ip_i)^q}u_iw_i(\lambda(w_iu_i)^{1/(1-q)})^{q-1} &= \gamma \\ \frac{M_{\text{WPM}}(\mathbf{u}\odot\mathbf{p};\mathbf{w},q)}{\sum_i w_i(u_ip_i)^q}\frac{u_iw_i}{u_iw_i}\lambda^{q-1} &= \gamma \\ \frac{M_{\text{WPM}}(\mathbf{u}\odot\mathbf{p};\mathbf{w},q)}{\sum_i w_i(u_ip_i)^q}\lambda^{q-1} &= \gamma \end{aligned}$$

由于左端不依赖于 i，我们可以将 γ 设为左端以满足平稳性。

(ii) 若 $p_i = 1$，则有 $1 = p_i < \lambda(w_iu_i^q)^{1/(1-q)}$。因此，

$$\begin{aligned} 1 &\leq \lambda(w_iu_i^q)^{1/(1-q)} \\ 1 &\overset{(i)}{\leq} \lambda^{1-q}w_iu_i^q \\ \lambda^{q-1} &\overset{(i)}{\leq} w_iu_i^q \\ \lambda^{q-1} &\leq \frac{w_i(u_ip_i)^q}{p_i} \\ \frac{M_{\text{WPM}}(\mathbf{u}\odot\mathbf{p};\mathbf{w},q)}{\sum_i w_i(u_ip_i)^q}\lambda^{q-1} &\overset{(i)}{\leq} M_{\text{WPM}}(\mathbf{u}\odot\mathbf{p};\mathbf{w},q)\cdot\frac{w_i(u_ip_i)^q}{\sum_i w_i(u_ip_i)^q}\cdot\frac{1}{p_i} \\ \gamma &\leq M_{\text{WPM}}(\mathbf{u}\odot\mathbf{p};\mathbf{w},q)\cdot\frac{w_i(u_ip_i)^q}{\sum_i w_i(u_ip_i)^q}\cdot\frac{1}{p_i} \end{aligned}$$

其中我们在(*i*)中令 $p_i = 1$，并根据(*ii*)中的第一种情形令 γ。平稳性要求

$$\frac{\partial}{\partial p_i}\mathcal{L} = -M_{\text{WPM}}(\mathbf{u}\odot\mathbf{p};\mathbf{w},q)\cdot\frac{w_i(u_ip_i)^q}{\sum_i w_i(u_ip_i)^q}\cdot\frac{1}{p_i} + \alpha_i + \gamma = 0$$

这意味着我们有 $\alpha_i \geq 0$，从而确立了对偶可行性。

因此，$\mathbf{p}^*$满足所有必要条件，表明它是一个局部最大值。由于我们是在凸多胞体上最大化一个凹函数，因此它也将是全局最大值。

极限情形。 我们分别讨论此情境下 q 的三个极限值:

1. $q = 1$ （加权功利主义）： 在此，极限解会为 $w_i\mu_i$ 最高的 k 个体分配概率 $p_i = 1$ ，而其他个体的分配概率为 $p_i = 0$ 。
2. $q = 0$ （加权纳什）： 在此，极限解为 $p_i = [\lambda w_i]_{[0,1]}$，其中 λ 的选择满足 $\sum_i p_i = 1$。重要的是，该解不依赖于效用向量 $\boldsymbol{\mu}$；因此，无需任何观测即可推导出最优解。
3. $q = -\infty$ （平等主义）： 在此，极限解为 $p_i = [\lambda/\mu_i]_{[0,1]}$，其中 λ 的选择满足 $\sum_i p_i = 1$。

这些极限情形是目标函数的逐点极限。由于可行集是紧集，且对于所有有限的 q，目标函数均为凹函数，根据标准的连续性论证，闭式解收敛至极限问题的最优解。

C.1.3. 算法

我们在算法2中描述了 WPM 分配预言机的算法。直观而言，该算法通过计算各个体的分配概率填充速率来运行。当填充速率最快的个体 i 达到 $p_i = 1$时，该个体将不再参与后续分配。此过程持续进行，直到 $\sum_i p_i = k$。注水速率的排序（第 15 行）决定了时间复杂度，因此总复杂度为 $\mathcal{O}(n\log\ n)$ 。

算法 2 加权幂平均（WPM）分配求解器

1：输入：效用 $u \in \mathbb{R}^n$**、权重** w**、幂参数** $q \leq 1$**、资源** k**2：输出：最优分配策略** $p \in [0,1]^n$**3：// 步骤 1：计算个体分配率 4：if** $q = -\infty$ **（平等主义）then 5：** $r_i \leftarrow 1/u_i$ **对所有** i**6：else if** $q = 0$**（纳什）then 7：** $r_i \leftarrow w_i$ **对所有** i**8：else if** $q = 1$**（功利主义）then 9：将最大（**$w_i \cdot u_i$**）对应的** k **索引设为 1.0，其余设为 0.0；返回** p**10：else 11：对所有** i**，对速率取对数**$_i \leftarrow$ **ln(**w_i**)+**q**·ln(**u_i**)** $^{1-q}$ **12：** $r \leftarrow$ **Softmax（对数速率）** $\{$**确保** $\sum r_i = 1\}$**13：end if 14：// 步骤 2：在排序容器上进行注水 15：按降序排列** r **：** $r_{(1)} \geq r_{(2)} \geq \cdots \geq r_{(n)}$**，并跟踪索引** π**16：初始化** $p \leftarrow 0$**、** $rem \leftarrow k$**17：预计算后缀和** $R_i = \sum_{j=i}^{n} r_{(j)}$**18：~~for~~** $i = 1$ **到** n **do 19：** $t \leftarrow 1/r_{(i)}$ $\{$**将当前容器填满至 1 所需的时间**。$0\}$**20：if** $t \cdot R_i > rem$ **then 21：** $t_{final} \leftarrow rem/R_i$ $\{$**剩余质量不足以再将任何容器完全填满**$\}$**22：** $p_{\pi(j)} \leftarrow t_{final} \cdot r_{(j)}$ **for** $j \in \{i, \ldots, n\}$**23：** $rem \leftarrow 0$**；break 24：else 25：** $p_{\pi(i)} \leftarrow 1.0$**26：** $rem \leftarrow rem - 1$**27：end if 28：end for 29：返回** p

C.2. Kolm 社会福利函数

C.2.1. 推导所提解

我们希望在约束条件下优化科尔姆社会福利。对于 $q < 0$，问题定义如下：

$$\begin{aligned} \text{maximize} \qquad & M_{\text{Kolm}}(\mathbf{u}; \mathbf{w}, q) = \frac{1}{q} \log \left(\sum_i w_i \exp(q u_i p_i) \right) \\ \text{s.t.} \qquad & p_i \in [0, 1] \qquad \forall i \\ & \sum_i p_i = k, \end{aligned}$$

其中 $k \in [n]$。

首先，考虑 p_i 不受限制的情况。对 p_i 求导后，在最优点处，我们希望

$$\begin{aligned} \frac{\partial}{\partial p_i} M_{\text{Kolm}} &= \lambda \\ \lambda &= \frac{1}{q} \cdot \frac{1}{\sum_i w_i \exp(q u_i p_i)} \cdot w_i q u_i \exp(q u_i p_i) \\ p_i &= \log \left(\frac{\lambda'}{w_i u_i} \right) \cdot \frac{1}{q u_i}, \end{aligned}$$

其中 $\lambda' = \lambda \cdot (\sum_i w_i \exp(q u_i p_i))$。

因此，Kolm 社会福利函数的一般解为

$$\begin{aligned} p_i &= \left[\log \left(\frac{\lambda'}{w_i u_i} \right) \cdot \frac{1}{q u_i} \right]_{[0,1]} \\ &= \left[\frac{\eta + \log(w_i u_i)}{|q| u_i} \right]_{[0,1]}, \end{aligned}$$

其中 $\eta = -\log \lambda'$，并且将 p_i 限制在 $[0, 1]$，使得 $\sum_i p_i = k$。这比 WPM 情况更难，因为在这里 $\eta = \log \lambda'$ 是加性的，而不是乘性的。因此，两个约束 $p_i \geq 0$ 和 $p_i \leq 1$ 都可能处于活跃状态。我们从 η 的角度来理解注水。

C.2.2. 通过 KKT 条件证明最优性

我们注意到

$$M_{\text{Kolm}}(\mathbf{u} \odot \mathbf{p}; \mathbf{w}, q) = \frac{1}{q} \log \left(\sum_i w_i \exp(q u_i p_i) \right),$$

and

$$\begin{aligned} \frac{\partial}{\partial p_i} M_{\text{Kolm}}(\mathbf{u} \odot \mathbf{p}; \mathbf{w}, q) &= \frac{1}{q} \cdot \frac{w_i \exp(q u_i p_i)}{\sum_i w_i \exp(q u_i p_i)} \cdot q u_i \\ &= \frac{w_i \exp(q u_i p_i)}{\sum_i w_i \exp(q u_i p_i)} \cdot u_i \end{aligned}$$

由于 $q \leq 0$，我们注意到 $\partial M_{\text{Kolm}} / \partial p_i$ 是 p_i 的非递增函数。

我们的优化目标函数为：

$$\begin{aligned} \text{maximize} \qquad & \frac{1}{q} \log \left(\sum_i w_i \exp(q u_i p_i) \right) \\ \text{s.t.} \qquad & -p_i \leq 0 \qquad \forall i \\ & p_i - 1 \leq 0 \qquad \forall i \\ & \sum_i p_i = k \end{aligned}$$

该表达式在最优解下的拉格朗日函数为$\mathbf{p}^*$

$$\mathcal{L} = -M_K(\mathbf{u} \odot \mathbf{p}^*; \mathbf{w}, q) - \sum_i \alpha_i p_i^* + \sum_i \beta_i (p_i^* - 1) + \gamma(\sum_i p_i^* - k)$$

关于 p_i 的偏导数为

$$\frac{\partial}{\partial p_i}\mathcal{L}\bigg|_{\mathbf{p}=\mathbf{p}^*} = -\frac{w_i \exp(q u_i p_i^*)}{\sum_i w_i \exp(q u_i p_i^*)} \cdot u_i - \alpha_i + \beta_i + \gamma$$

KKT 条件要求：

1. 平稳性： $0 \in \partial\mathcal{L}/\partial p_i$，这意味着

$$0 = -\frac{w_i \exp(q u_i p_i^*)}{\sum_i w_i \exp(q u_i p_i^*)} \cdot u_i - \alpha_i + \beta_i + \gamma$$

2. 原始可行性：

$$\begin{aligned} -p_i^* &\le 0 \qquad \forall i \\ p_i^* - 1 &\le 0 \qquad \forall i \\ \sum_i p_i^* &= k \end{aligned}$$

通过我们对 $\mathbf{p}^*$ 的构造，所有原始可行性条件均得到满足。

3. 对偶可行性：

$$\begin{aligned} \alpha_i &\ge 0 \qquad \forall i \\ \beta_i &\ge 0 \qquad \forall i \\ \gamma &\ge 0 \end{aligned}$$

4. 互补松弛：

$$\begin{aligned} \alpha_i p_i^* &= 0 \qquad \forall i \\ \beta_i (p_i^* - 1) &= 0 \qquad \forall i \end{aligned}$$

我们考虑三种情况：

- $p_i^* \in (0,1)$：这意味着

$$p_i^* = \frac{\eta^* + \log(w_i u_i)}{|q| u_i},$$

这进而意味着

$$\begin{aligned} \frac{\partial}{\partial p_i} M_{\text{Kolm}}(\mathbf{u} \odot \mathbf{p}; \mathbf{w}, q)\bigg|_{\mathbf{p}=\mathbf{p}^*} &= \frac{w_i \exp(-\eta^*)/(w_i u_i)}{\sum_i w_i \exp(q u_i p_i^*)} \cdot u_i \\ &= \frac{\exp(-\eta^*)}{\sum_i w_i \exp(q u_i p_i^*)} \end{aligned}$$

此外，根据互补松弛，我们有 $\alpha_i = 0$ 和 $\beta_i = 0$。因此，为了满足平稳性，

$$\gamma = \frac{\exp(-\eta^*)}{\sum_i w_i \exp(q u_i p_i^*)},$$

我们注意到，该式与索引 i无关。

- $p_i^* = 0$：这意味着

$$p_i^* = 0 \geq \frac{\eta^* + \log(w_i u_i)}{|q| u_i}.$$

由于 $\partial M_{\text{Kolm}}/\partial p_i$ 是 p_i 的递减函数，因此我们有

$$\begin{aligned} -\frac{\partial}{\partial p_i} M_{\text{Kolm}}(\mathbf{u} \odot \mathbf{p}; \mathbf{w}, q)\Big|_{\mathbf{p}=\mathbf{p}^*} &\geq -\frac{\exp(-\eta^*)}{\sum_i w_i \exp(q u_i p_i^*)} \\ -\frac{\partial}{\partial p_i} M_{\text{Kolm}}(\mathbf{u} \odot \mathbf{p}; \mathbf{w}, q)\Big|_{\mathbf{p}=\mathbf{p}^*} + \gamma &\geq 0 \end{aligned}.$$

互补松弛意味着 $\beta_i = 0$。因此，要满足平稳性，我们应有

$$0 = -\frac{\partial}{\partial p_i} M_{\text{Kolm}}(\mathbf{u} \odot \mathbf{p}; \mathbf{w}, q) - \alpha_i + \gamma,$$

这可以通过某个 $\alpha_i > 0$ 实现。因此，对偶可行性也得到满足。

- $p_i^* = 1$：这意味着

$$p_i^* = 1 \leq \frac{\eta^* + \log(w_i u_i)}{|q| u_i}.$$

由于 $\partial M_{\text{Kolm}}/\partial p_i$ 是 p_i 的递减函数，因此我们有

$$\begin{aligned} -\frac{\partial}{\partial p_i} M_{\text{Kolm}}(\mathbf{u} \odot \mathbf{p}; \mathbf{w}, q)\Big|_{\mathbf{p}=\mathbf{p}^*} &\leq -\frac{\exp(-\eta^*)}{\sum_i w_i \exp(q u_i p_i^*)} \\ -\frac{\partial}{\partial p_i} M_{\text{Kolm}}(\mathbf{u} \odot \mathbf{p}; \mathbf{w}, q)\Big|_{\mathbf{p}=\mathbf{p}^*} + \gamma &\leq 0 \end{aligned}.$$

互补松弛意味着 $\alpha_i = 0$。因此，要满足平稳性，我们应当有

$$0 = -\frac{\partial}{\partial p_i} M_{\text{Kolm}}(\mathbf{u} \odot \mathbf{p}; \mathbf{w}, q) + \beta_i + \gamma,$$

这可以通过某个 $\beta_i > 0$ 实现。因此，对偶可行性也得到满足。

因此，KKT 条件的必要条件均得到满足。由于目标函数中的约束都是仿射函数，这些必要条件也足以保证最优性。这意味着 $\mathbf{p}^*$ 是最优解。

C.2.3. 算法

我们在算法 3中介绍 Kolm 分配预言机的算法。直观地说，该算法通过计算各个体的分配概率的填充速率来推进。与 WPM 算法的关键区别在于，由于 λ在此处是一个加性因子，我们会跟踪两个事件：1）个体 i的概率首次变为非零，此时将其纳入活动集；2）个体 i的概率变为 $p_i = 1$，此时将其从活动集中移除。

当填充最快的个体 i 达到 $p_i = 1$时，该个体将不再获得分配。此过程持续进行，直到$\sum_i p_i = k$。对注水速率进行排序（第 17 行）是时间复杂度的主要部分，因此总复杂度为 $\mathcal{O}(n\log\ n)$。不过，为了简化起见，我们在此实现了一个 $\mathcal{O}(n^2)$ 算法。

C.3. 基尼社会福利函数

对于基尼社会福利函数，优化目标函数为：

$$\begin{aligned} \text{maximize} \quad & M_{\text{Gini}}(\boldsymbol{\mu}^{\uparrow} \odot \mathbf{p}; \mathbf{w}) = \sum_i w_i (\boldsymbol{\mu}^{\uparrow} \odot \mathbf{p})_{(i)} \\ \text{s.t.} \quad & p_i \in [0,1] \qquad \forall i \\ & \sum_i p_i = k \end{aligned}$$

算法 3 Kolm 社会福利分配

输入： 效用 $u \in \mathbb{R}^n$、权重 w、幂参数 $q \leq 0$、资源 k
输出： 分配概率 $p \in [0,1]^n$
// **步骤 1：计算个体分配率**
如果 $q = -\infty$ **则**
  $r_i \leftarrow 1/u_i$ 对所有 i
否则如果 $q = 0$**则**
  $p \leftarrow$ 将具有最大（$w_i \cdot u_i$）的 k 索引设为1.0，其余设为0.0；**返回** p
否则
  $r_i \leftarrow 1/$（$-q \cdot u_i$）对所有 i
结束如果
// **步骤 2：识别关键事件时间（达到 0 或 1）**
如果 $q = -\infty$ **则**
  $T_i^{start} \leftarrow 0, T_i^{end} \leftarrow u_i$ 对所有 i
否则
  $T_i^{start} \leftarrow -\ln(w_i u_i),\ T_i^{end} \leftarrow -q u_i - \ln(w_i u_i)$ 对所有 i
结束如果
将所有 $2n$ 时刻$\{T_i^{start}, T_i^{end}\}$排序为一个序列 $\tau_1, \tau_2, \ldots, \tau_{2n}$
// **步骤 3：在时间区间上注水**
初始化 $p \leftarrow \mathbf{0}$、活跃臂 $\leftarrow \emptyset, t_{prev} \leftarrow \tau_1$
对于 $j = 2$到 $2n$**执行**
  $\Delta t \leftarrow \tau_j - t_{prev}$
  $m_{interval} \leftarrow \sum_{i \in \text{活跃臂}} r_i \cdot \Delta t$
  如果求和$(p) + m_{interval} > k$**则**
    $\Delta t_{final} \leftarrow (k - \text{求和}(p)) / \sum_{i \in \text{活跃臂}} r_i$
    $p_i \leftarrow p_i + r_i \cdot \Delta t_{final}$ 对于所有 $i \in$ 活跃臂；**中断**
  else
    $p_i \leftarrow p_i + r_i \cdot \Delta t$ 对于所有 $i \in$ 活跃臂
    根据 τ_j 是 T^{start}（添加）还是 T^{end}（移除）来更新活跃臂
    $t_{prev} \leftarrow \tau_j$
  结束条件判断
结束循环
返回 p

其中，权重向量$\mathbf{w}$ 满足 $w_1 \geq w_2 \geq \ldots \geq w_n \geq 0$。此外，$(\boldsymbol{\mu}^{\uparrow} \odot \mathbf{p})_{(i)}$ 表示向量$\mathbf{u} \odot \mathbf{p}$的第 i 个顺序统计量。

C.3.1. 个体的最优顺序

命题 C.1. 存在一个最优解 $\mathbf{p}^*$，使得若 π 是满足以下条件的排列

$$\mu^{\uparrow}_{\pi(1)} \leq \mu^{\uparrow}_{\pi(2)} \leq \cdots \leq \mu^{\uparrow}_{\pi(n)}, \tag{1}$$

then

$$\mu^{\uparrow}_{\pi(1)} p^*_{\pi(1)} \leq \mu^{\uparrow}_{\pi(2)} p^*_{\pi(2)} \leq \cdots \mu^{\uparrow}_{\pi(n)} p^*_{\pi(n)}. \tag{2}$$

也就是说，$\mu^{\uparrow}_{\pi\ (i)}\ p_{\pi(i)}$ 的最优顺序是按 $\mu^{\uparrow}_{\pi\ (i)}$ 的非递减顺序排列。

解释：我们可以按非递减的 μ_i 为个体编序，并只考虑满足 $\mu_i p_i$ 非递减的分配；然后只需对 p进行优化。

证明。为简化记号，我们重新标记索引，使得 $\pi(i) = i$。因此，$\mu^{\uparrow}_1 \leq \mu^{\uparrow}_2 \leq \ldots \mu^{\uparrow}_n$。

令 $\mathbf{p}$ 为最优解。令 $u_i = \mu_i p_i$，并令 σ 为一个排列，使得 $\sigma(i)$ 是 $\{u_k\}$ 中第 i 小的元素。我们选择 σ，使得在出现平局时逆序对最少，即 $i < j$ 使得 $u_i > u_j$。

按照 σ 对 u_i's 进行排序后，我们得到若干个 u_i 值相同的"块"。由于逆序对数量最少，每个块中的 μ_i's 都按非递减顺序排列。

如果这些块之间的 μ_i's 按非递减顺序排列，那么问题就解决了。否则，设有两个块 B_1 和 B_2（其中 B_1 包含的 u_i 值小于 B_2），使得这些块之间的 μ_i's 并未按非递减顺序排列。

设 r 为 B_1 中最右侧的元素，l 为 B_2 中最左侧的元素。因此，我们有 $u_r < u_l$（因为它们属于不同的块），且有 $\mu_r > \mu_l$（因为 μ_i's 未按非递减顺序排列）。

设 r' 为按照 σ 排序后位于 r 右侧的元素，设 l' 为位于 l 左侧的元素。由于 l 和 r 分别处于各自块的边界，因此必有 $u_r < u_{r'}$，且 $u_{l'} < u_l$。

因此，存在某个足够小的 $\epsilon > 0$，使得 $p'_r = p_r + \epsilon \in [0,1]$、$p'_l = p_l - \epsilon \in [0,1]$、$u'_r := \mu_r p'_r \le u_{r'}$，且 $u_{l'} \le u'_l := \mu_l p'_l$。

因此，将这部分小概率 ϵ 从 l 调换到 r，不会违反顺序约束。然而，这种调换会使基尼社会福利函数发生如下变化：

$$\Delta M_{\text{Gini}} = \epsilon(w_{\sigma(r)}\mu_r - w_{\sigma(l)}\mu_l) \ge 0,$$

因为 $w_{\sigma(r)} \ge w_{\sigma(l)}$ 且 $\mu_r > \mu_l$。若 $w_{\sigma(r)} > 0$，则该不等式为严格不等式，这意味着$\mathbf{p}$ 不是最优解。若 $w_{\sigma(r)} = 0$，则有 $\Delta M_{\text{Gini}} = 0$，这意味着调换概率不会降低目标函数。重复进行这种调换（目标函数始终不会降低），直到不存在这样的块对为止，即可得到满足 $\mu_1 p_1 \le \ldots \le \mu_n p_n$ 的最优解。 □

C.3.2. 神谕算法的理论性质

我们将基尼社会福利函数的目标函数表示为一个参数化线性规划问题。令 $\tau \in [0,k]$ 为控制概率之和的参数。不失一般性，我们假设各个 $\mu_i^{\uparrow}$ 按非递减顺序排列。我们的目标函数为

$$\begin{aligned} \text{maximize } & \sum_i w_i \mu_i^{\uparrow} p_i \\ \text{such that } & \sum_i p_i = \tau \\ & p_i \in [0,1] \quad \forall\, i \\ & \mu_{i-1}^{\uparrow} p_{i-1} \le \mu_i^{\uparrow} p_i \quad \forall i \in \{2,\ldots,n\} \end{aligned}$$

令 $\mathbf{p}^{\tau}$ 为算法在参数 τ下得到的解。我们的算法从 $\tau = 0$开始，其中 $\mathbf{p}_0 = \mathbf{0}_n$ 为最优解。对于每个 τ，算法都会选择一个索引子集 $S(\tau)$，并随着 τ的增加而提高这些索引对应的概率，同时确保不违反任何约束。选择集合 $S(\tau)$和概率增长速率 $r_p(i)$，以使基尼目标函数的增长率最大化。

命题 C.2. 该算法具有以下性质：

1. 当 $S(\tau)$ 中的某个元素 i 满足 $p_{\tau,i} = 1$时，所选下标集合及其增长率会发生变化。
2. 每个 τ 所选下标对应于一组连续索引 $B(\tau)$。
3. 所选索引 $B(\tau)$ 的概率以如下速率增加

$$r_p(\tau, i) = \frac{1/\mu_i^{\uparrow}}{\sum_{j \in B(\tau)} 1/\mu_j^{\uparrow}},$$

而这会使基尼目标函数的增长率达到

$$r_o(B(\tau)) = \frac{\sum_{i \in B(\tau)} w_i}{\sum_{i \in B(\tau)} 1/\mu_i^{\uparrow}}$$

4. 在任意 τ 处的所有可行无穷小方向中，算法选取的块 $B(\tau)$ 及上述概率增长率，使目标函数的瞬时增长率最大化。

5. 每当集合 $B(\tau)$ 发生变化时，我们就得到一个块序列，每个块内的 $\mu_i p_i$ 值都相同。这些块由满足 $p_i = 1$ 的索引 i 彼此分隔。

证明。 我们通过对 $S(\tau)$ 发生变化的点集进行归纳来证明这一结果。由于目标函数是线性的，且约束集是一个凸多胞体，这种变化只会发生有限次。

基例：我们从 $\tau = 0$ 开始，此时 $\mathbf{p}_\tau = \mathbf{0}_n$ 是显然的最优解。由于这是我们首次选择索引和速率，因此性质 (1) 显然成立。由于此时只有一个块，性质 (5) 在 $\tau = 0$ 处也显然成立。随着 τ 增大，任何允许的增量都将涉及一个由元素组成的后缀块，否则就会违反顺序约束$\mu_{i-1}p_{i-1} \le \mu_i p_i$。因此，性质 (2) 也成立。

设 B 是某个后缀块。我们对该块内概率的速率选择为

$$r_p(i) = \frac{1/\mu_i^{\uparrow}}{\sum_{j\in B} 1/\mu_j^{\uparrow}}.$$

这一选择使效用的增长速率为

$$r_u(i) = \mu_i^{\uparrow} \cdot r_p(i) = \frac{1}{\sum_{j\in B} 1/\mu_j^{\uparrow}}.$$

该增长速率对块内所有元素都相同，因此不会违反顺序约束。此外，这会使基尼社会福利函数以如下速率增加

$$r_o(B) = \sum_{i\in B} w_i r_u(i) = \frac{\sum_{i\in B} w_i}{\sum_{i\in B} 1/\mu_i^{\uparrow}}.$$

因此，命题的第（3）部分是必须满足顺序约束的结果。

设 B' 是 B 的一个子集，且为一个后缀块，使得 $r_o(B') > r_o(B)$；也就是说，如果将所有效用都分配给该子块，基尼社会福利函数的增长速率会更高。那么， B 对该块而言是次优选择，我们应选择 B'。请注意，切换到 B 的子块不会违反顺序约束。我们可以从 $B = [n]$ 开始重复这一论证，直到找到 $r_o(B)$ 值最大的后缀块。这将是我们对 $B(\tau)$ 的选择。由此，我们刻画了第（4）部分中的这个块，说明它在满足约束的同时提供了最大的增长速率。

随着我们提高该块的概率，最左侧的索引 $l(B)$ 填充得最快，因为它的 μ_i 值最小，因此具有最高的填充速率。因此，当我们有 $p_{l(B)} = 1$ 时，这种块和填充速率的选择不再有效，
这满足第 (1) 部分。在此实例中，我们有两个 $\mu_i^{\uparrow} p_i$ 值相等的块（左侧块为 0，右侧块为 $\mu_{l(B)}$），两者之间由 $p_{l(B)} = 1$ 分隔，这满足第 (5) 部分。

归纳情形：令 τ 为某个中间值，此时必须改变所选下标集合及其增长率。我们假设第 (5) 部分成立，即我们有一系列块，这些块由下标 i 分隔，并且满足 $p_i = 1$。

我们注意到，该序列中的一个块与 $B = 0$ 下 $[n]$ 的更大的块相似。因此，如果我们对 $S(\tau)$ 的选择位于某个块内，则第 (1) 至 (5) 部分都会自然得到满足。现在我们证明确实如此。

假设情况相反，即我们选择的 $S(\tau)$ 跨越了多个块。在这种情况下，最优解将包含所有块的后缀。任何可行无穷小方向都可以分解为各个块内后缀方向的凸组合。由于目标函数斜率是线性的，最优方向会将全部权重集中在最大斜率对应的单个块上（若有并列则任意选择）。因此，第 (1) 至 (5) 部分均成立。 □

由于目标函数是线性的，且可行域是一个凸多胞体，因此沿着使瞬时目标增量最大的方向移动时，只要约束尚未变为紧约束，解就会一直保持在该线性规划的最优面上。

C.3.3. 基尼社会福利函数预言机算法

我们在算法4中介绍基于区块的注水算法。该算法通过按块填充概率来执行，并确保不违反顺序约束。所选的块是在遵循顺序约束的同时，能够使基尼社会福利函数值增长速率最大的块。每次块计算需要 $\mathcal{O}(n)$ 时间，且最多需要进行 k 次块计算，因此时间复杂度为 $\mathcal{O}(kn)$。

算法 4 基尼注水求解器

1: 输入：平均效用向量 $u \in \mathbb{R}^n$**、权重向量** w**（其中** $w_1 \geq w_2 \geq \cdots \geq w_n \geq 0$**）、资源数量** k**。2: 输出：使** $M_{Gini}(\mu \odot p; w)$ **最大化的最优分配策略** $p \in [0,1]^n$**。3: 按非递减顺序对** u **排序：** $u_{(1)} \leq u_{(2)} \leq \cdots \leq u_{(n)}$**。4: 令** π **为满足**$u_{(i)} = u_{\pi(i)}$**的排列。5: 对所有** $i \in [n]$**，** $rem = k$**初始化** $p_i = 0$**，并初始化块集合** $\mathcal{B} = \{[1,n]\}$**。6: 预计算后缀和** $W_i = \sum_{j=i}^{n} w_j$ **及效用倒数和** $I_i = \sum_{j=i}^{n} 1/u_{(j)}$**。7: 当** $rem > 0$**时，执行：8: 找出使速率最大的块** $B \in \mathcal{B}$**：** $r(B) = \frac{\sum_{i \in B} w_i}{\sum_{j \in B} 1/u_{(j)}}$**。社会福利函数的增长速率 9: 令** $l(B)$ **为块** B **的最左索引，** R **为最右索引。10:** $\Delta t \leftarrow (1 - p_{\pi(l(B))}) \cdot u_{(l(B))}$ **填满** $p_{l(B)} = 1$**所需的时间 11: 填满块** B**所需的总概率质量 12: 如果** $m \leq rem$**，则 13: 对** $j \in B$ **执行 14:** $p_{\pi(j)} \leftarrow p_{\pi(j)} + \Delta t / u_{(j)}$**15: 结束循环 16:** $rem \leftarrow rem - m$**17:** $\mathcal{B} \leftarrow (\mathcal{B} \setminus \{B\}) \cup \{[l(B)+1, R]\}$ **最左索引现已完全分配 18: 更新剩余子块的预计算后缀和。19: 否则 20: 速率总和** $\leftarrow \sum_{j \in B} 1/u_{(j)}$**21: 对** $j \in B$ **执行 22:** $p_{\pi(j)} \leftarrow p_{\pi(j)} + \frac{rem}{\text{速率总和} \cdot u_{(j)}}$ **23: 结束循环 24:** $rem \leftarrow 0$**25: 结束条件判断 26: 结束循环 27: 返回** p

D. 依赖舍入算法

我们使用依赖舍入算法（Gandhi 等人，2006年；Grafstr¨om，2010年）来抽取集合 S_t 在给定分配策略向量 $\mathbf{p}_t$ 使得 $\mathbb{P}(i \in S_t) = p_{t,\,i}$。这一思想也被称为 π-PS 抽样，并常用于调查设计，即需要以个体/群体间不相等的概率从人口中抽取样本。算法 5给出了依赖舍入的伪代码。

E. 更多实验：线性权重衰减

我们在第 6 节中考虑了权重向量的指数衰减，这会使少数个体获得更大的权重。在本节中，我们考虑一种更平缓的权重线性衰减，即 $w_i = (1 + (i-1)/(n-1))/S$，其中 $S = \sum_{i=1}^{n}(1 + (i-1)/(n-1))$。我们使用与第 6 节完全相同的实验设置，其中 $n = 50$、$U_i \sim 0.1 + 0.9X_i$，且每个 X_i 服从贝塔分布，参数 (α_i, β_i) 与第 6 节中的参数相同。

变化的时域 T。我们令 $q = -2$ 并改变 T范围 $[10^3, 1.28 \cdot 10^5]$。我们在图 T中改变 k，绘制遗憾随 4变化的曲线。对于三种社会福利函数，我们观察到，与参考虚线相比，遗憾大致按 $\mathcal{O}(\sqrt{T})$ 的比例变化，这与我们的理论保证一致。总体而言，我们发现遗憾随着 T的增加而增加，随后随着

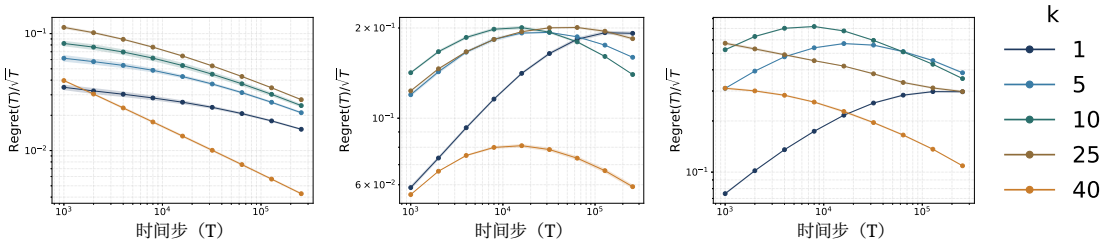
算法 5 π-ps 采样（成对依赖舍入）

要求： $\pi \in [0,1]^N$ 使用 $\sum_i \pi_i = n$

确保： 样本 $S \subseteq [N]$, $|S| = n$

当 π 中至少存在两个分数项**时**
  选取不同的分数项索引 $i \neq j$
  $s \leftarrow \pi_i + \pi_j$, 抽取 $r \sim \mathrm{Unif}[0,1]$
  如果 $s \leq 1$, **则**
    $(\pi_i, \pi_j) \leftarrow \begin{cases} (0, s) & \text{以概率 } \pi_j/s \\ (s, 0) & \text{以概率 } \pi_i/s \end{cases}$
  else
    $(\pi_i, \pi_j) \leftarrow \begin{cases} (1, s-1) & \text{以概率 } (1-\pi_j)/(2-s) \\ (s-1, 1) & \text{以概率 } (1-\pi_i)/(2-s) \end{cases}$
  结束 if
结束 while

返回 $S \leftarrow \{i : \pi_i = 1\}$



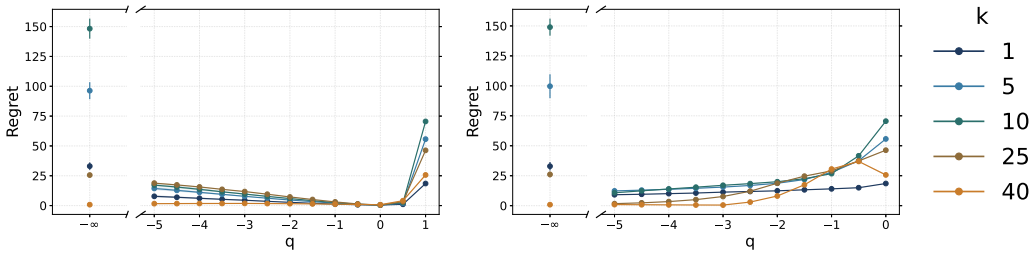
(a) WPM 社会福利函数　*(b)* Kolm 社会福利函数　*(c)* 基尼社会福利函数

图 *4.* 采用权重线性衰减时，WPM、Kolm 和基尼社会福利函数的归一化遗憾 $R(T)/\sqrt{T}$ 与时间范围 T 的关系。在跨越两个数量级的 T范围内，归一化遗憾始终有界，这与我们的理论 $\tilde{\mathcal{O}}(\sqrt{T})$ 保证一致。

k 进一步增加。然而，随着 k 增加，遗憾的相对排序会随 T而变化。我们还观察到，这些趋势与图 1中的趋势大不相同；总体而言，随着 T增加，归一化遗憾下降得更快，这表明最优分配向量更容易学习。

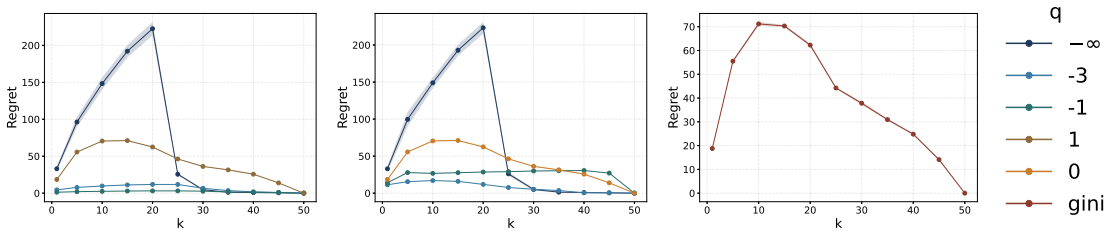
变化的幂值 q。现在，我们绘制不同 q 取值下，分配资源数量 k 变化时观察到的遗憾，如图 5所示。对于 WPM 和 Kolm，当 $q = -\infty$ 时，遗憾值相同；这符合预期，因为权重对平等主义福利没有影响。然而，对于有限的 q，趋势则显著不同。我们观察到，随着 q增加，WPM 社会福利函数的遗憾呈现出明显更强的规律性。对于 Kolm 社会福利函数，不同 k 取值下的趋势则明显更无规律；在指数权重衰减的情况下，并未观察到这一现象。

分配资源数量的变化 k。我们绘制了每个时间步分配资源数量不同时观测到的遗憾 k，见图6。对于 WPM 社会福利函数，当 k 有限时 q，遗憾的增长更为显著，这与图3中观察到的非单调行为不同。Kolm 社会福利函数也呈现出类似行为。基尼社会福利函数与功利主义福利相似，对应于 $q = 1$ 以及 $q = 0$，分别适用于 WPM 社会福利函数和 Kolm 社会福利函数。



*(a)*WPM 社会福利函数　　*(b)* Kolm 社会福利函数

图 *5.* WPM 和 Kolm 社会福利函数在权重线性衰减时随幂值 q 变化的趋势。 $q = -\infty$ 对两者均对应平等主义福利，而社会福利函数的 $q = 1$ 和 Kolm 的 $q = 0$ 则对应功利主义福利。我们观察到，尽管观测到的遗憾平滑变化，但随着 k 变化，结果存在显著差异。



(a) WPM 社会福利函数　　*(b)* Kolm 社会福利函数　　*(c)* 基尼社会福利函数

图 *6*。 三种社会福利函数在分配资源数量 k 变化时的趋势。我们观察到，对于平等主义福利，在 $k = 20$ 之后会急剧下降。随着 q 增加，WPM 和 Kolm 社会福利函数的变化变得不再那么平缓。线性权重与功利主义福利有些相似；对于基尼社会福利函数，随着 k 变化，我们也观察到了类似的模式。