arXiv:2303.02504v2 [cs.LG] 2 Jun 2023

非平稳环境中的 MNL 多臂老虎机

Ayoub Foussoul

美国纽约哥伦比亚大学工业工程与运筹学系，af3209@columbia.edu

Vineet Goyal

美国纽约哥伦比亚大学工业工程与运筹学系，vg2277@columbia.edu

Varun Gupta

布斯商学院，芝加哥大学，美国芝加哥，varun.gupta@chicagobooth.edu

摘要

本文研究非平稳环境中的 MNL-老虎机问题，并提出一种算法，其最坏情况下的期望遗憾为 $\tilde{O}\left(\text{最小值}\left\{\sqrt{NTL},\ N^{\frac{1}{3}}(\Delta_\infty^K)^{\frac{1}{3}}T^{\frac{2}{3}}+\sqrt{NT}\right\}\right)$。其中 N 是臂的数量，L 是变化次数，而 Δ_∞^K 是未知参数的变化度量。此外，我们给出了相匹配的期望遗憾下界（相差对数因子），这表明我们的算法是最优的。我们的方法基于 Agrawal 等人于 2016 年提出的平稳 MNL-老虎机的基于轮次的算法。然而，非平稳性带来了若干挑战，因此我们引入了新的技术和思路来应对这些挑战。具体而言，我们精确刻画了非平稳性导致的估计量偏差，并推导出了新的集中界。

1 引言

MNL 多臂老虎机问题是传统随机多臂老虎机（MAB）问题的一种组合变体。在该问题中，给定一组 N 臂，其收益已知，r_1，……，r_N而参数 ω_1，……，ω_N未知。在每个时间步，决策者选择一个子集 S，其中最多包含 K 个臂。随后，自然按照多项逻辑模型（MNL）随机选择臂 $i \in S$，其概率为 $\omega_i/(1+\sum_{j\in S}\omega_j)$。决策者获得与臂 i对应的收益 r_i。自然也可能不选择任何臂，概率为 $1/(1+\sum_{j\in S}\omega_j)$；此时决策者获得的收益为 0。参数 ω_1，……，ω_N称为 MNL 模型的吸引力参数。

MNL-老虎机问题常见于在线零售场景，其中 N 个臂对应于具有已知价格的 N 种可替代产品。顾客依次到达，并按照相同的、吸引力参数未知的 MNL 模型进行选择

。对于每位顾客，卖家提供一个至多包含 K 种产品的子集，并根据 MNL 模型，针对顾客从所提供集合中作出的随机选择收取相应的收益（如果未发生购买，也可能没有收益）。卖家的目标是在 T 位顾客中最大化总期望收益。

MNL 多臂老虎机是一个研究充分的问题。Agrawal 等人，2016 研究了吸引参数平稳情形下的 MNL-老虎机问题，并提出了一种置信上界算法，其遗憾为 $\tilde{O}(\sqrt{NT})$。这一结果在对数因子范围内是最优的（X. Chen 和 Wang，2017）。具体而言，他们提出了一种基于时期的算法，通过采样来估计吸引参数，从而克服期望收益函数非线性的难题。

然而，在许多应用中，MNL 参数 ω_1、.....、ω_N 会随关注的时间范围而变化。例如，在在线零售中，产品的吸引参数通常会在销售期间发生变化。受这些情形启发，本文研究非平稳参数下的 MNL 多臂老虎机问题。具体而言，我们考虑一种非平稳性的对抗模型，其中吸引参数由对手选择，并开发一种在线学习算法，以最小化决策者的最坏情况动态遗憾（即获得的总收益与每个时间步都采取最优动作所能获得的总收益之间的差距）。我们给出了遗憾界，其取决于吸引参数的变化次数（或切换次数），以及刻画吸引参数总变化量的变化量参数。这项工作延续了一系列关于非平稳环境中强化学习问题的研究；在这些工作中，针对多种学习问题考虑了类似的非平稳性模型，包括多臂老虎机问题（Auer、Cesa-Bianchi 等人，2002；Garivier 和 Moulines，2011；Besbes、Gur 和 Zeevi，2014；Auer、Gajane 和 Ortner，2019；Wei 和 Luo，2021）、上下文多臂老虎机（Luo 等人，2018；Y. Chen 等人，2019）、线性老虎机（Cheung、Simchi-Levi 和 Zhu，2019；Russac、Vernade 和 Capp´e，2019；Zhao 等人，2020）以及其他问题（Hazan 和 Seshadhri，2009；Yang 等人，2016；Russac、Capp´e 和 Garivier，2020；Mao 等人，2020；Touati 和 Vincent，2020）。

处理老虎机问题中非平稳性的一种广泛使用的方法，是运行一个在平稳环境中表现良好的算法（我们称之为基础算法），并持续检测非平稳性。一旦检测到非平稳性，就重新启动基础算法。该方法产生的动态遗憾取决于：在检测到非平稳性之前，基础算法在“近似平稳”环境中累积的遗憾。在经典的 n臂多臂老虎机问题中，基础算法在每个时期计算未知参数的无偏估计量，而评估基础算法在近似平稳环境中累积的遗憾相对容易。对于期望奖励函数为非线性的 MNL-老虎机问题，即使在平稳环境中，我们也不知道有哪种算法能够在每个时期计算吸引力参数的无偏估计量并提供良好的遗憾保证。Agrawal 等人，2016 针对平稳

环境下的 MNL-老虎机问题给出了一种最优算法。他们考虑了一种基于时期的算法，该算法使用多个时间步（一个时期）来获得吸引力参数的单个无偏估计量。然而，在吸引力参数随时间变化的近似平稳环境中，评估基于时期的方法所累积的遗憾会变得更加困难，并且需要新的分析技术。

1.1 我们的贡献

我们为非平稳环境下的 MNL-老虎机问题提出了一种算法，其最坏情况动态遗憾为 $\tilde{O}\left(\min\left\{\sqrt{NTL},\, N^{\frac{1}{3}}(\Delta_\infty^K)^{\frac{1}{3}}T^{\frac{2}{3}}+\sqrt{NT}\right\}\right)$。这里，$L$ 是切换次数，形式化定义为 $L := 1+\sum_{t=1}^{T-1}\mathbf{1}\{\boldsymbol{\omega}(t) \neq \boldsymbol{\omega}(t+1)\}$，其中 $\boldsymbol{\omega}(t)$ 表示第 t 轮的吸引参数向量。此外，Δ_∞^K 是按范数 L_∞^{2K} 定义的变化量，定义为 $\Delta_\infty^K := \sum_{t=1}^{T-1}\|\boldsymbol{\omega}(t) - \boldsymbol{\omega}(t+1)\|_{L_\infty^{2K}}$，其中 $\|x\|_{L_\infty^{2K}} = \sup_{S\subset[N]:|S|\le 2K}\sum_{r\in S}|x_r|$，对 $x\in\mathbb{R}^N$ 成立。我们将该算法称为非平稳环境下 MNL-老虎机的探索-利用算法（EEMNSE）。EEMNSE 由 Agrawal 等人，2016 年提出的基于 UCB 的算法针对近似平稳环境所作的改编（见算法 1），以及 Wei 和 Luo，2021 年提出的 MASTER 算法组成。MASTER 是一种元算法，它以一个基础算法（此处为算法 1）作为输入，以多尺度方式安排该算法的运行，并在检测到环境的非平稳性时重新启动。

挑战与关键技术贡献。即使在平稳环境中，MNL-老虎机问题的一大难点也源于期望收益函数的非线性。Agrawal 等人，2016 通过采用基于时期的方法克服了这一难题：在连续多个时期内提供相同的商品组合（称为一个时期），直到观察到未购买为止。Agrawal 等人，2016 证明，在一个时期内某个臂被选择的次数是对应臂吸引参数的无偏估计量，且服从几何分布 [1]。然而，在非平稳环境中，情况并非如此。特别是，由于吸引力参数可能在一个时期内发生变化，该估计量会产生偏差，其分布也取决于参数的对抗性变化。主要挑战在于紧密刻画非平稳性造成的偏差。

我们引入一种新的随机过程，称为不间断购买过程（UPP），并利用它对偏差进行精确刻画。具体而言，我们将 UPP 夹在两个均值差有界的几何随机变量之间（按照通常随机序）。为分析遗憾，我们利用随机序，为 UPP 推导出新的集中界。具体来说，我们针对 UPP 给出了非对称集中界：其中一个界相对于随机上界的均值，另一个界则相对于随机下界的均值。我们利用这些集中界

[1] 我们想指出，由于有限时间范围造成的截断，估计量的分布并非严格服从几何分布。我们在分析中处理了这一技术问题。

来评估吸引力参数与其估计量之间的差距。最后，我们利用收益函数的一种特殊利普希茨性质，将这一差距转化为收益函数中的差距。我们的方法或许也能为处理基于轮次的算法中的非平稳性提供独立参考。

匹配的下界。 我们还针对切换次数 L 和一种新颖的变化度量Δ_∞^K，分别证明了非平稳环境下 MNL-老虎机问题的下界。具体而言，我们证明，在有 L次切换的非平稳环境中，任何针对 MNL-老虎机问题的多项式时间算法都必须承担至少为 $\tilde{\Omega}\left(\sqrt{NLT}\right)$的遗憾。此外，在变化量为 Δ_∞^K的非平稳环境中，任何算法都必须承担$\tilde{\Omega}\left((N)^{\frac{1}{3}}(\Delta_\infty^K)^{\frac{1}{3}}T^{\frac{2}{3}}\right)$的遗憾。这两个界与我们的在线算法的上界相匹配，至多相差对数因子。因此，对于非平稳环境下的 MNL 多臂老虎机问题，我们的算法是近乎最优的（至多相差对数因子）。

我们还表明，我们提出的变化度量 Δ_∞^K 与其他多臂老虎机场景中常用的无穷范数变化量 L_∞（记为 Δ_∞）不同，且它对遗憾的刻画严格更优。具体而言，我们证明，在最坏情况下，任何多项式时间算法都必须承担 $\tilde{\Omega}\left((KN)^{\frac{1}{3}}(\Delta_\infty)^{\frac{1}{3}}T^{\frac{2}{3}}\right)$的遗憾（注意，我们的算法达到了这一界）。由于 $\Delta_\infty^K \leq 2K\cdot\Delta_\infty$，上界 $\tilde{O}\left(N^{\frac{1}{3}}(\Delta_\infty^K)^{\frac{1}{3}}T^{\frac{2}{3}}\right)$ 对我们算法性能的刻画严格优于 $\tilde{O}\left((NK)^{\frac{1}{3}}(\Delta_\infty)^{\frac{1}{3}}T^{\frac{2}{3}}\right)$。

1.2 其他相关文献

MNL 多臂老虎机最早由 Rusmevichientong、Shen 和 Shmoys 于 2010年 研究。在该工作中，作者开发了一种“探索后提交”算法，其最坏情况下的遗憾为 $O(N^2\log^2 T)$，前提是假设已知最优商品组合与次优商品组合之间的“差距”。随后，Saur´e 和 Zeevi 于 2013年 在类似假设下将这一界改进为 $O(N\log\ T)$。Agrawal 等人，2016 首次为 MNL-老虎机开发了无参数算法。他们的算法实现了最坏情况下为 $\tilde{O}(\sqrt{NT})$ 的遗憾。他们还为该问题证明了 $\Omega(\sqrt{NT/K})$ 的下界，随后 X. Chen 和 Wang，2017 将其改进为 $\Omega(\sqrt{NT})$。最近，Agrawal 等人，2017 开发了一种基于汤普森采样的 MNL-老虎机问题算法，其最坏情况下的遗憾为 $\tilde{O}(\sqrt{NT})$，并表明与以往方法相比，该算法在实践中表现良好。Agrawal 等人，2016；Agrawal 等人，2017 的算法都依赖于一种基于轮次的估计方法，即使用多个时间步（时期）从未知参数的每个估计量中获得一个样本。

我们的问题与多臂老虎机（MAB）范式密切相关（Robbins，1952）。大量文献研究了非平稳环境中的多臂老虎机问题（及其扩展）。其中一些研究考虑了非平稳性的结构化模型

例如休闲多臂老虎机（Whittle，1988；Tekin 和 M. Liu，2012）、休眠多臂老虎机（Gittins，1979；Tekin 和 M. Liu，2012）等（Slivkins 和 Upfal，2008；Levine、Crammer 和 Mannor，2017；Kleinberg 和 Immorlica，2018）。另一些研究则考虑与我们类似的非平稳性对抗模型。其中，Auer、Cesa-Bianchi 等人，2002；Bubeck 和 Slivkins，2012；Seldin 和 Slivkins，2014；Auer 和 Chiang，2016 等研究将性能与单一最佳动作进行比较；还有一些研究考虑动态遗憾的概念。在这些后续研究中，最坏情况动态遗憾的上界由表征环境非平稳程度的参数决定。广泛研究的两个参数是切换次数 L （即未知参数发生变化的次数）和变化量 Δ（即未知参数在整个时间范围内变化总量的度量）（Garivier 和 Moulines，2011；Besbes、Gur 和 Zeevi，2014；Besbes、Gur 和 Zeevi，2015；F. Liu、Lee 和 Shroff，2018；Luo 等人，2018；Cheung、Simchi-Levi 和 Zhu，2019；Y. Chen 等人，2019；Auer、Gajane 和 Ortner，2019；W. Chen 等，2021；Wei 和 Luo，2021）。文献中还考虑了其他参数（Wei、Hong 和 Lu，2016；Suk 和 Kpotufe，2022；Abbasi-Yadkori、Gyorgy 和 Lazic，2022）。

2 问题描述

在 MNL 多臂老虎机问题中，决策者拥有一个包含 N 个商品的选择空间，这些商品的收益（或收入）为 r_1，……，r_N，并且给定一个整数 $K \leq N$。自然选择决策轮次的数量 T，并为每个轮次 t 和商品 j 选择一个吸引参数 $\omega_j(t)$。令 $\boldsymbol{\omega}(t) := (\omega_1(t)……, \omega_N(t))$ 表示轮次 t 时的吸引参数向量。在每个轮次 t，决策者选择一个至多包含 K 个商品的子集 $S(t) \subset [N]$ （称为商品组合）。随后，自然根据具有吸引参数 $\boldsymbol{\omega}(t)$ 的 MNL 选择模型选择（或购买）一个商品 $j(t) \in S(t)$（也可能完全不选择商品）。最后，决策者获得收益 $r(t) = r_{j(t)}$。

具有吸引参数 $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_N)$ 的 MNL选择模型满足：对于每个$S \subset [N]$，当提供的商品组合为 S时，自然以概率 $p^{\boldsymbol{\omega}}(j, S) := \omega_j/(1 + \sum_{r \in S} \omega_r)$ 选择商品 $j \in S$，或以概率 $p^{\boldsymbol{\omega}}(0, S) := 1/(1 + \sum_{r \in S} \omega_r)$ 完全不选择商品。可以将“无选择”或“未购买”选项视为一个额外的商品（商品 0），该商品始终可供自然选择，其吸引参数为 $\omega_0 = 1$。遵循 Agrawal 等人，2016，我们假设对于每个时间步 t，$\omega_j(t) \leq \omega_0 = 1$。这一假设通常在实践中得到验证。例如，在在线零售中，顾客最常见的选择是什么都不买。我们还假设对于每个商品 j，$r_j \leq 1$ 。这一假设是为了简化分析；只要奖励由某个正常数界定，它就不会在数量上改变我们的结果。

决策者的目标是在每一轮 t 最大化其期望收益，表示为，$R(S(t)\ \boldsymbol{\omega}(t)) := \sum_{j \in S(t)} r_j p^{\boldsymbol{\omega}(t)}(j, S(t))$. 然而，由于吸引力参数 $(\boldsymbol{\omega}(t))^{t \in [1,T]}$ 是

未知的，决策者转而寻求最小化其动态遗憾，定义如下：

$$\text{D-Reg} = \sum_{t=1}^{T} R(S^*(t), \boldsymbol{\omega}(t)) - r(t),$$

其中 $S^*(t)$ 是在参数 $\boldsymbol{\omega}(t)$ 下使期望收益最大化的商品组合 .

我们为决策者设计了一种在线算法，并将该算法的最坏情况动态遗憾表示为时间范围 T、切换次数 L、变化量Δ_∞^K 以及该问题的其他已知参数的函数。

3 算法

本节将介绍我们针对非平稳环境下 MNL-老虎机问题设计的算法，我们称之为非平稳环境下 MNL-老虎机的探索-利用算法（EEMNSE）。我们的算法对 Agrawal 等人提出的 UCB 算法进行了大幅调整，2016 （该算法仅适用于平稳环境），使其能够应用于近似平稳环境。我们首先正式定义“近似平稳环境”。随后，我们将介绍算法及其实现的动态遗憾上界。最后，我们将讨论分析中的关键思想和技术。

定义 *1.* 考虑函数 $\Delta: [1,T] \to [0,\infty)$，使得对于每个 $t \in [1, T-1]$，都有$\Delta(t) \geq \max_{S \subset [N]:|S| \leq K} |R(S, \boldsymbol{\omega}(t)) - R(S, \boldsymbol{\omega}(t+1))|$。我们将此函数称为非平稳性度量。设 $\rho: [1,T] \to [0,\infty)$ 是一个非递增函数，满足 $\rho(t) \geq \frac{1}{\sqrt{t}}$，且映射$t \to t \cdot \rho(t)$ 是非递减的。则环境的 (Δ, ρ)-近似平稳部分定义为满足 $\sum_{\tau=1}^{t-1} \Delta(\tau) \leq \rho(t)$ 的时间轮次 $t \in [1,T]$。

如果一个算法在(Δ, ρ)-近似平稳环境中表现良好，就意味着该算法无需了解 Δ，且对于问题的每个实例，该算法都会在每轮 t 开始时输出一个奖励上界 $\hat{} \; R_t \in [0,1]$，并且以至少 $1 - O(\frac{1}{T})$ 的概率，对属于环境 (Δ, ρ)- 近似平稳部分的每一轮 t，以下两个条件均成立：

$$\hat{R}_t \geq \min_{\tau \in [1,t]} R(S^*(\tau), \boldsymbol{\omega}(\tau)) - \sum_{\tau=1}^{t-1} \Delta(\tau) \qquad (1a) \qquad \frac{1}{t}\sum_{\tau=1}^{t} \hat{R}_\tau - r(\tau) \leq \rho(t) + \sum_{\tau=1}^{t-1} \Delta(\tau) \qquad (1b)$$

直观来说，环境的（Δ，ρ）近似平稳部分，是指吸引力参数的变化幅度（通过 Δ衡量）不太大的部分（与 ρ相比）。现在我们来介绍用于 MNL-老虎机问题的算法：

非平稳环境下 MNL-老虎机的探索-利用算法（EEMNSE）。Wei 和 Luo，2021 年提出了一种元算法（称为 MASTER），使得给定

一个在（Δ，ρ）-近似平稳环境中表现良好的基础算法，其中 Δ 和 ρ，MASTER 以多尺度方式调度基础算法，并利用奖励上界 $\hat{}R_t$ 在每个时间步执行非平稳性检验，以检测未知参数是否发生变化。一旦检测到变化，MASTER 就会重新启动。详细描述请参阅 Wei 和 Luo，2021 年。Wei 和 Luo，2021 年将 MASTER 与基础算法结合后的最坏情况动态遗憾界定为 Δ、ρ 以及问题的其他参数的函数。我们的算法由 MASTER 与基础算法的组合构成，该基础算法是对 Agrawal 等人 2016 年提出的 UCB 算法针对近似平稳环境的改编版本。具体而言，由于在近似平稳环境中取得良好性能需要新的集中界，因此需要使用不同的置信上界。我们在算法 1中介绍了这一改编版本。算法 1 按时期运行。在每个时期，我们根据吸引参数当前的置信上界（UCBs），提供使期望收益最大化的商品组合。当发生“未购买”时，该时期结束，并更新 UCBs。

算法 1： 近似平稳环境下 MNL 多臂老虎机的探索与利用

初始化： 对于每个商品 $j \in [N]$，令 $n_j = 0, \tilde{\omega}_j = 0, \bar{\omega}_j = 0, \hat{}\omega_j = 192\log\ NT$。
对于 $t = 1, \ldots, T$ **do**
  提供商品组合 $S = \text{argmax}_{C \subset [N]:|C| \le K} R(C, \hat{}\boldsymbol{\omega})$；购买商品 $j(t)$。
  如果 $j(t) = 0$ **则** // a no-purchase happened and epoch ends here
    对于 $j \in S$，**执行**
      $n_j \leftarrow n_j + 1$； // Update the number of epochs where j was proposed
      $\bar{\omega}_j \leftarrow \frac{n_j - 1}{n_j}\bar{\omega}_j + \frac{1}{n_j}\tilde{\omega}_j$； // Update the mean of the number of purchases
      $\hat{}\omega_j \leftarrow \bar{\omega}_j + \sqrt{\frac{192\bar{\omega}_j \text{对数} NT}{n_j}} + \frac{192 \text{对数} NT}{n_j}$对所有 $j \in [N]$； // Update the UCBs
      $\tilde{\omega}_j \leftarrow 0$； // Reset the number of purchases within an epoch
    end
  else
    $\tilde{\omega}_{j(t)} \leftarrow \tilde{\omega}_{j(t)} + 1$； // Update the number of purchases within current epoch
  end
end

遗憾界。 为了界定 EEMNSE 的动态遗憾，我们评估环境近似平稳时基础算法（算法 1）累积的遗憾。特别是，我们证明存在函数 Δ 和 ρ，使得算法 1 在（Δ，ρ）-近似平稳环境中表现良好，并且 EEMNSE 为 MNL-老虎机问题实现了最优动态遗憾（忽略对数因子）。更确切地说，我们证明以下定理：

定理 1。 对于这些选择，

$$\Delta(t) = 26\|\boldsymbol{\omega}(t) - \boldsymbol{\omega}(t+1)\|_{L_\infty^{2K}}, \quad \forall t \in [1, T-1],$$

and

$$\rho(t) = (149 \log NT)^{\frac{3}{2}} \sqrt{\frac{N}{t}} + (55 \log NT)^3 \frac{N}{t} + \sqrt{\frac{2 \log T}{t}}, \quad \forall t \in [1, T],$$

算法 *1* 在 (Δ, ρ)-近似平稳环境中表现良好，其 $\hat{} R_t = R(S(t), \hat{}\boldsymbol{\omega}(t))$，其中 $\hat{}\boldsymbol{\omega}(t)$ 是在第 t 轮次的上置信界向量。根据 *Wei* 和 *Luo* 的定理 *2*，*2021*，这进一步意味着，即使不知道 Δ_∞^K 和 L，*EEMNSE* 仍能以高概率保证动态遗憾，

$$\text{D-Reg} = \tilde{O}\left(\min\left\{\sqrt{NTL}\,,\ N^{\frac{1}{3}}(\Delta_\infty^K)^{\frac{1}{3}}T^{\frac{2}{3}} + \sqrt{NT}\right\}\right).$$

记号。 我们采用统一记号 $X_j^{(t)}(i)$ 来表示变量 X （例如购买次数）在第 i 步的取值（一步可以是一个时间步、一个时期，或仅仅是一个索引；如果 X不依赖于任何特定的步的概念，我们就省略索引 i）；对于商品 $j \in [N]$ （同样地，如果 X不依赖于商品，我们就省略下标）；以及在时间轮次 1 至 t 之间的取值（当时间范围重要时保留上标，否则省略）。对于每个满足 $i \le k$ 的 i，$k \ge 1$，我们用$X_j^{(t)}(i\colon k)$ 表示随机过程（或随机变量序列） $X_j^{(t)}(i), \ldots, X_j^{(t)}(k)$。

按照这一约定，我们定义以下与算法 1 相关的随机变量和过程：定义 $e_j^{(t)}$ 为截至时间 t （含）商品组合中曾提出 j 的时期数。请注意，这还包括在 t之前未结束的未完成时期。定义 $l_j^{(t)}\left(1\colon e_j^{(t)}\right)$ 为截至时间 t 曾提出 j 的时期（注意， $l_j^{(t)}\left(e_j^{(t)}\right)$可能是一个未完成时期）。定义 $\tilde{\omega}_j^{(t)}\left(1 : e_j^{(t)}\right)$ 分别为在时期 $l_j^{(t)}\left(1 : e_j^{(t)}\right)$ 中商品 j 的购买次数。我们将 $\tilde{\omega}_j^{(t)}\left(1\colon e_j^{(t)}\right)$ 称为购买过程。我们还定义 $\hat{}\boldsymbol{\omega}(t) = (\hat{}\omega_1(t) \ldots, \hat{}\omega_N(t))$ 和 $\bar{\boldsymbol{\omega}}(t) = (\bar{\omega}_1(t) \ldots, \bar{\omega}_N(t))$ 分别为轮次 t 时上置信界向量 $(\hat{}\omega_1, \ldots, \hat{}\omega_N)$ 和平均值向量 $(\bar{\omega}_1, \ldots, \bar{\omega}_N)$，并定义 $n_j(t)$ 为在 $t-1$ 之前曾提出 j 的完整时期数（仅计入在 $t-1$ 之前或恰于 $t-1$ 结束的时期）。令 $\tilde{n}_j(t) = \max\{n_j(t), 1\}$。

对于每个 $t \in [1, T]$ 和商品 $j \in [N]$，定义 $\delta^{(t)} := \sum_{\tau=1}^{t-1} \|\boldsymbol{\omega}(\tau) - \boldsymbol{\omega}(\tau+1)\|_{L_\infty^{2K}}$ 和 $\delta_j^{(t)} := \sum_{\tau=1}^{t-1} |\omega_j(\tau) - \omega_j(\tau+1)|$。最后，对于每个 $\tau \in [1, t]$，令 $\mu_j^{+(t)}(\tau) = \left(\omega_j(\tau) + \delta_j^{(t)}\right)\frac{1+\delta^{(t)}}{1-\delta^{(t)}}$ 和 $\mu_j^{-(t)}(\tau) = \left(\omega_j(\tau) - \delta_j^{(t)}\right)^+ \frac{1-\delta^{(t)}}{1+\delta^{(t)}}$，其中 $(a)^+$ 表示 $a \in \mathbb{R}$ 的正部。注意，对于整数 $m \le n$，我们用 $[m, n]$ 表示集合 $\{m, \ldots\ldots, n\}$，并用 $[m]$ 表示集合 $[1, m]$。

下一节将专门证明定理 1。但在进入证明之前，我们先简要讨论分析中遇到的挑战，以及我们用来克服这些挑战的关键思想。

关键思想。 考虑 Δ 和 ρ 在定理1中给出。我们证明，条件（1a）和（1b）对于 $\hat{} R_t = R$（S（t）、 $\hat{}\boldsymbol{\omega}$（$t$）成立。固定环境中处于（ t 的（Δ， ρ）近似平稳部分的一个时间步。条件（1a）和（1b）要求界定在时间范围 $[1, t]$内使用上置信界而非真实吸引参数时，期望收益函数的变化。

我们分两步进行：第一步，对于任意 $\tau \in [1,t]$，评估上置信界 $\hat{\boldsymbol{\omega}}(t)$ 与真实吸引参数 $\boldsymbol{\omega}(\tau)$ 之间的差距；然后在第二步中，利用期望收益函数的特殊利普希茨性质，将这一差距转化为期望收益方面的差距。第一步的起点是为每个 $j \in [N]$ 推导平均值 $\bar{\omega}_j(t) = \frac{1}{n_j(t)} \sum_{l=1}^{n_j(t)} \tilde{\omega}_j^{(t)}(l)$ 的集中界。回顾一下，$\tilde{\omega}_j^{(t)}(l)$（即在轮次 l 中对 j 的购买次数）被用作真实吸引参数 ω_j 的估计量。理想情况下，我们希望该估计量是以 ω_j 为中心的次高斯变量。在平稳环境中，Agrawal 等人，2016 表明确实如此，并且 $\tilde{\omega}_j^{(t)}(l)$ 的分布（以该轮次中提供的商品组合为条件）服从以 ω_j 为中心的几何分布（正如我们之前提到的，由于有限时间范围造成的截断，$\tilde{\omega}_j^{(t)}(l)$ 的分布并不完全是以 ω_j 为中心的几何分布）。然而，在非平稳环境中，吸引参数会在一个轮次内发生变化，因此得到的估计量如今存在偏差，并且其分布取决于吸引参数的对抗性变化。此外，不同轮次之间的分布不一定相互独立。这违背了平稳环境分析中所使用的估计量的关键性质。

为克服这些问题，我们引入不间断购买过程（UPP）。该过程模拟商品 j 在其被提供的每个时期内的购买次数，假设算法在 t 之后继续处于平稳环境中，其中 $\boldsymbol{\omega}(t') = \boldsymbol{\omega}(t)$，持续 $t' \geq t$，并在从 t之后开始的各个时期提供相同的商品组合 $S = [N]$。（UPP）的优势在于，其分布比购买过程的分布“更好”，而且我们将看到，推导（UPP）的集中界足以得到购买次数均值的集中界。在平稳环境中，（UPP）由一列几何随机变量构成，因此能够解决截断问题。在非平稳环境中，我们利用通常随机序，将（UPP）的变量夹在两个均值相近的几何随机变量之间，从而通过两个“良好”的分布控制非平稳性造成的偏差。具体而言，我们将（UPP）的变量夹在均值为 $\mu_j^{+(t)}(\tau)$ 和 $\mu_j^{-(t)}(\tau)$ 的两个几何随机变量之间。我们的随机界是紧的，因为它们给出了最优的遗憾上界。为得到这些随机界，我们先用独立同分布的 Gumbel(0,1) 随机变量描述（UPP），再用 Gumbel(0,1) 随机变量类似地描述几何随机变量，并通过耦合论证证明这些随机界。接下来是推导集中界。我们推导的是非对称集中界，也就是说，一个界针对 $\mu_j^{+(t)}(\tau)$，另一个针对 $\mu_j^{-(t)}(\tau)$。我们证明这足以满足我们的需求。最后，我们利用收益函数的一个特殊利普希茨性质，将上置信界与真实参数之间的差距转化为期望收益函数中的差距。详细证明见附录。

4 分析

接下来，我们固定定理 1 中定义的 Δ 和 ρ，并对每个 $t \in [1,T]$令 $\hat{} R_t = R(S(t), \hat{}\omega(t))$。我们的目标是证明 Δ 确实是一个非平稳性度量，并证明对于属于环境的（Δ, ρ）-近似平稳部分的每个 t，条件（1a）和（1b）以至少 $1 - O(\frac{1}{T})$ 的概率成立。注意，对于满足 $\delta^{(t)} \geq \frac{1}{2}$的每个时间步 $t \in [1,T]$，我们有$\sum_{\tau=1}^{t-1} \Delta(\tau) = 26\delta^{(t)} \geq 13 \geq 1$，且条件（1a）和（1b）对这样的 t成立。因此，我们只需关注满足 $\delta^{(t)} \leq \frac{1}{2}$ 的时间步 t。

4.1 不间断购买过程（UPP）

令 $t \in [1,T]$ 满足 $\delta^{(t)} \leq \frac{1}{2}$，并令 $j \in [N]$。我们将不间断购买过程 $\zeta_j^{(t)}$ $(1:\infty)$ 定义为：若在时间 t 之后 (i) 算法在平稳环境中继续运行，且对于每个 $t' \geq t$， $\omega(t') = \omega(t)$，并且 (ii) 算法在此后各时期提供相同的商品组合 $S = [N]$，那么该过程模拟商品 j 在其被提出的各时期中的购买次数。

形式上，考虑模拟 $t+1$ 之后各时间轮次的无限过程 $X^{(t)}(t+1:\infty)$，并按如下方式递归定义：对于每个 $t' \geq t+1$，如果变量 $j(t)$ 和 $X^{(t)}(t+1:t'-1)$ 均不为 0，则从以 $S(t) \cup \{0\}$ 为支持集的类别分布中抽取 $X^{(t)}(t')$；给定 $S(t)$，该抽样与问题中的任何其他随机性相互独立，且每个商品 $r \in S(t)$ 被抽中的概率为 $\omega_r(t)/(1+\sum_{k\in S(t)} \omega_k(t))$。否则，从以 $[N] \cup \{0\}$ 为支持集的类别分布中抽取 $X^{(t)}(t')$；该抽样与问题中的任何其他随机性相互独立，且每个商品 $r \in [N]$ 被抽中的概率为 $\omega_r(t)/(1+\sum_{k\in[N]} \omega_k(t))$。该过程模拟 t 之后的虚拟轮次。它先在轮次 t 完成当前时期（提供商品组合 $S(t)$），然后在该时期结束后切换至商品组合 $[N]$。令 i_k 表示发生 $X^{(t)}(i) = 0$ 的第 k 个步骤 $i \geq t+1$ （约定 $i_0 = t+1$）。(UPP) 定义如下：

定义 2. 对每个 $t \in [1,T]$，使得 $\delta^{(t)} \leq \frac{1}{2}$ 且 $j \in [N]$。不间断购买过程 $\zeta_j^{(t)}$ $(1:\infty)$ 定义如下：对于每个时期 $i \leq e_j^{(t)} - 1$， $\zeta_j^{(t)}(i) = \tilde{\omega}_j^{(t)}(i)$。对于$i \geq e_j^{(t)}$，我们区分两种情况。第一种情况是第 t 轮发生了未购买，此时令 $\zeta_j^{(t)}(e_j^{(t)}) = \tilde{\omega}_j^{(t)}(e_j^{(t)})$，并且对于 $k \geq e_j^{(t)} + 1$， $\zeta_j^{(t)}(k)$ 等于满足 $X^{(t)}(i) = j$ 的步数 $i \in \left[i_{k-e_j^{(t)}-1}, i_{k-e_j^{(t)}}\right]$。第二种情况是第 t 轮发生了购买，此时令 $\zeta_j^{(t)}(e_j^{(t)})$ 等于 $\tilde{\omega}_j^{(t)}(e_j^{(t)})$ 与满足$X^{(t)}(i) = j$ 的步数 $i \in [i_0, i_1]$ 之和；并且对于 $k \geq e_j^{(t)} + 1$， $\zeta_j^{(t)}(k)$ 等于满足$X^{(t)}(i) = j$ 的步数 $i \in \left[i_{k-e_j^{(t)}}, i_{k-e_j^{(t)}+1}\right]$。

我们将时期开始时间、轮次 τ 的提供的商品组合 $S(\tau)$ 和购买元素 $j(\tau)$ 的定义扩展到新定义的 (UPP) 虚拟轮次，如下所示，

定义 *3.* 令 $t \in [1,T]$，使得 $\delta^{(t)} \leq \frac{1}{2}$，并令 $j \in [N]$。令 $k \geq 1$。定义 $\zeta_j^{(t)}$ *(k)* 的时期（真实或虚拟）开始时间记为 $s_j^{(t)}$ *(k)*，如下所示：若 $k \leq e_j^{(t)}$ ，则 $s_j^{(t)}$ *(k)* 是时期 $l_j^{(t)}$ *(k)* 的开始时间。若在 t 发生未购买，则对每个 $k \geq e_j^{(t)}+1$，$s_j^{(t)}$ *(k)* $= i_{k-e_j^{(t)}-1}$。否则，对每个 $k \geq e_j^{(t)}+1$，$s_j^{(t)}$ *(k)* $= i_{k-e_j^{(t)}}$。虚拟轮次 $\tau \geq t+1$ 中的提供的商品组合定义如下：对于 $t \leq \tau < s_j^{(t)}\left(e_j^{(t)}+1\right)$，令 *S(τ)* $= S(t)$；对于 $\tau \geq s_j^{(t)}\left(e_j^{(t)}+1\right)$，令 *S(τ)* $= [N]$。最后，每个轮次 $\tau \geq t+1$ 的购买元素定义为 *j(τ)* $= X^{(t)}(\tau)$。

注 1. 令 $t \in [1,T]$ 满足 $\delta^{(t)} \leq \frac{1}{2}$，并令 $j \in [N]$。令 $k \geq 1$。随机变量$\zeta_j^{(t)}(k)$ 几乎必然定义良好。事实上，只有在存在一个持续无限多个时间步的时期 $k' < k$ 时，$\zeta_j^{(t)}(k)$ 才未定义良好。由于每一轮未购买的概率至少为 $\frac{1}{1+N}$，时期 k' 永远持续的概率为0。对所有可能的 k' 应用并集界可得，$\zeta_j^{(t)}(k)$ 几乎必然定义良好。再对所有 t、j 和 k 应用并集界可得，每个 $\zeta_j^{(t)}(k)$ 都几乎必然定义良好。在下文中，我们将不失一般性地仅考虑算法中满足这一条件的样本路径。

在通常随机序下夹逼（UPP）： 对于每个时间步 $s \geq 1$，令 $\Gamma(s)$ 表示 $[N]$ 的子集，使得 $S \in \Gamma(s)$ 当且仅当存在算法的至少一条样本路径，使得 $S(s) = S$。特别要注意，对于每个 $s \leq t$，每个 $S \in \Gamma(s)$ 都满足$|S| \leq K$。以下引理在随机序下用两个几何随机变量夹逼（UPP）的变量：

引理1。令 $t \in [1,T]$ 满足$\delta^{(t)} \leq \frac{1}{2}$，并令 $j \in [N]$。令 $k \geq 1$、$\tau \in [1,t]$、$s \geq 1$和$S \in \Gamma(s)$。令 X 表示在给定 $s_j^{(t)}(k) = s$ 和 $S(s) = S$的条件下的随机变量 $\zeta_j^{(t)}(k)$。令X^+ 为均值为 $\mu_j^{+(t)}(\tau)$ 的几何随机变量，令 X^- 为均值为 $\mu_j^{-(t)}(\tau)$的几何随机变量。那么，

$$X^- \leq_{st} X \leq_{st} X^+,$$

其中 $\leq_{st}$ 表示通常随机序。

4.2 集中界

令$t \in [1,T]$满足$\delta^{(t)} \leq \frac{1}{2}$并令$j \in [N]$。利用引理1，我们可以推导出随机变量 $\zeta_j^{(t)}(k)$的集中界。我们先对给定 $s_j^{(t)}(k) = s$ 和 $S(s) = S$ 时的矩母函数进行界定，其中 $s \geq 1$ 以及 $S \in \Gamma(s)$。

引理 2. 令 $t \in [1,T]$ 满足 $\delta^{(t)} \leq \frac{1}{2}$，并令 $j \in [N]$。令 $k \geq 1$、$\tau \in [1,t]$、$s \geq 1$，且

$S \in \Gamma(s)$，则对于每个 $0 \leq \lambda <$ 对数 $\left(1 + \frac{1}{\mu_j^{+(t)}(\tau)}\right)$，有，

$$\mathbb{E}\left(e^{\lambda \zeta_j^{(t)}(k)} \middle| s_j^{(t)}(k) = s \ , \ S(s) = S\right) \leq \frac{1}{1 - \mu_j^{+(t)}(\tau) \cdot (e^{\lambda} - 1)}.$$

并且对于每个 $\lambda \leq 0$，我们有，

$$\mathbb{E}\left(e^{\lambda \zeta_j^{(t)}(k)} \middle| s_j^{(t)}(k) = s \ , \ S(s) = S\right) \leq \frac{1}{1 - \mu_j^{-(t)}(\tau) \cdot (e^{\lambda} - 1)}$$

现在我们可以推导变量 $\zeta_j^{(t)}(k)$ 的集中界。具体而言，我们有以下引理，

引理 3. 设 $t \in [1, T]$ 满足 $\delta^{(t)} \leq \frac{1}{2}$，且令 $j \in [N]$。令 $k \geq 1$ 和 $\tau \in [1, t]$。对于每个$\eta > 0$，我们有，

$$\mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k} \zeta_j^{(t)}(i) < (1 - \eta)\mu_j^{-(t)}(\tau)\right) \leq e^{-\frac{k\eta^2}{24}\mu_j^{-(t)}(\tau)}$$

and,

$$\mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k} \zeta_j^{(t)}(i) > (1 + \eta)\mu_j^{+(t)}(\tau)\right) \leq e^{-\frac{k \min\{\eta, \eta^2\}}{196}\mu_j^{+(t)}(\tau)}.$$

回顾一下，$\bar{\omega}_j(t) := \frac{1}{\tilde{n}_j(t)} \sum_{i=1}^{n_j(t)} \tilde{\omega}_j^{(t)}(i)$ 表示在 $[1, t-1]$ 中提出 j 时，对 j 在 $n_j(t)$ 个完整时期内的购买次数取均值。由于 $\tilde{n}_j(t)$ 仅包含在 $t-1$ 之前或当时结束的时期，因此对于每个 $i \leq n_j(t)$，我们有 $\tilde{\omega}_j^{(t)}(i) = \zeta_j^{(t)}(i)$，这意味着 $\bar{\omega}_j(t) := \frac{1}{\tilde{n}_j(t)} \sum_{i=1}^{n_j(t)} \zeta_j^{(t)}(i)$。利用这一事实和上述引理 3，我们证明了以下关于 $\bar{\omega}_j(t)$ 的集中界：

引理 4。令 $t \in [1, T]$ 使得 $\delta^{(t)} \leq \frac{1}{2}$，并令 $j \in [N]$。令 $\tau \in [1, t]$。以下集中界成立，

$$\mathbb{P}\left(\hat{\omega}_j(t) < \mu_j^{-(t)}(\tau)\right) \leq \frac{2}{NT^3}$$

and

$$\mathbb{P}\left(\hat{\omega}_j(t) > \mu_j^{+(t)}(\tau) + \sqrt{\frac{2266\mu_j^{+(t)}(\tau)\log NT}{\tilde{n}_j(t)}} + \frac{1364 \log NT}{\tilde{n}_j(t)}\right) \leq \frac{2}{NT^3}.$$

4.3 界定各时期的长度

分析中我们需要的最后一个要素，是证明一个时期的长度以其均值为中心集中。对于 Agrawal 等人 2016 针对平稳环境的分析，他们利用条件期望定律，以均值替代轮次的长度（即 Agrawal 等人 2016 中从 (A.13) 到 (A.14) 的转换）。然而，这一结论并不成立，因为时期总数 L 并不是 $\mathcal{H}_{l-1}$-可测的（其中 $\mathcal{H}_{l-1}$ 是截至时期 $l-1$可获得的信息）。我们采用一种不同的方法，以高概率界定（实际）时期的长度。这样做只会在遗憾中额外损失一个 O（log NT）。更准确地说，我们证明以下引理:

引理 5. 设 $t \in [1,T]$ 满足 $\delta^{(t)} \le \frac{1}{2}$，且设 $j \in [N]$。设 $u \in [1,t]$。我们有:

$$\mathbb{P}\left(\bigcup_{1\le i\le e_j^{(t)}} \left\{ \left| l_j^{(t)}(i) \right| > 5\log NT \left(1 + \sum_{k\in S\left(s_j^{(t)}(i)\right)} \omega_k(u) + \delta_k^{(t)} \right) \right\}\right) \le \frac{1}{NT^3}$$

4.4 综合起来

结合上述各项即可证明定理 1。在证明中，我们首先证明 Δ 确实是一个非平稳性度量，然后证明，对于每个满足 $\sum_{\tau=1}^{t-1}\Delta(\tau) \le \rho(t)$ 且 $\delta^{(t)} \le \frac{1}{2}$ 的 $t \in [1,T]$，条件 (1a) 和 (1b) 以至少 $1 - O(\frac{1}{T})$ 的概率成立，其中 $\hat{}R_\tau = R(S(\tau), \hat{}\omega(\tau))$。后一个结论利用了前述引理，以及期望收益函数的一种特殊利普希茨性质（Agrawal 等人，2016 年的引理 A.3）。证明见附录 A。

5 下界

本节将给出在非平稳环境中，任何用于 MNL 多臂老虎机的多项式时间算法所能达到的遗憾下界，该下界以 L 和 Δ_∞^K为函数。我们还将讨论为何选择变化量 Δ_∞^K，而非 Δ_∞。

对 L 的依赖。 下述定理表明，我们的算法所达到的遗憾上界 $\tilde{}O(\sqrt{NTL})$ 是最优的（差距至多为对数因子）：

定理 2. 固定 T, L, N和 K，并假设 $K \le \frac{N}{4}$。那么，对于每个用于非平稳环境下 *MNL* 多臂老虎机问题的多项式时间算法 $\mathcal{A}$，都存在一个参数为 N 和 K 、且至多有 L 次切换的实例，使得$\mathcal{A}$在 T轮中累积的遗憾至少为$C\cdot\min(\sqrt{NLT}, T)$，其中 $C > 0$是一个与 T、L、N和K无关的绝对常数。

我们对定理 2 的证明采用了 X. Chen 和 Wang 在 2017 年关于平稳环境的研究结果。粗略地说，给定一个算法 $\mathcal{A}$，我们将时间范围划分为 $\left\lceil \frac{T}{L} \right\rceil$ 个窗口，然后按以下方式递归地构造针对 $\mathcal{A}$ 的对抗性实例：假设我们已固定前 $w-1$ 个窗口中的对抗性实例。对于窗口 w，我们采用 X. Chen 和 Wang 在 2017 年给出的、针对 $\mathcal{A}$ 的平稳对抗实例（给定此前已固定的窗口 $w' \leq w-1$）。由此得到一个至多有 L次切换的非平稳实例，并证明 $\mathcal{A}$ 针对该实例累积的期望遗憾至少为 $C \cdot \min(\sqrt{NLT}, T)$。定理 2 的证明见附录B。

对 Δ_∞^K的依赖。下面的定理表明，我们的算法所实现的遗憾上界 $\tilde{O}(N^{\frac{1}{3}}(\Delta_\infty^K)^{\frac{1}{3}}T^{\frac{2}{3}})$ 是最优的（相差至多对数因子）：

定理 3。固定 T、Δ、N和 K，并假设 $K \leq \frac{N}{4}$且 $\Delta \in [\frac{1}{N}, \frac{T}{N}]$。那么，对于每个求解非平稳 *MNL* 多臂老虎机问题的多项式时间算法 $\mathcal{A}$，都存在一个参数为 N 和 K 、变化量为 $\Delta_\infty^K \leq \Delta$ 的实例，使得 $\mathcal{A}$ 在时间范围 T内累积的遗憾至少为$C \cdot N^{\frac{1}{3}}\Delta^{\frac{1}{3}}T^{\frac{2}{3}}$，其中 $C > 0$是一个与 T、Δ、N和 K无关的绝对常数。此外，该实例满足 $\Delta_\infty^K \geq \frac{K}{2} \cdot \Delta_\infty$，其中 Δ_∞ 表示该实例吸引参数在范数 L_∞ 下的变化量。

为了证明定理 3，我们将时间范围划分为长度相等的窗口 M，并递归地构造一个针对算法 $\mathcal{A}$ 的对抗性实例：与之前一样，假设我们已固定前 $w-1$ 个窗口中的实例，在窗口 w中，我们使用一个平稳实例，如 X. Chen 和 Wang 在2017 年提出的实例，来对抗 $\mathcal{A}$（给定先前固定的窗口 $w' \leq w-1$）。通过恰当选择 M，我们的实例的变化量至多为 Δ，并且使得 $\mathcal{A}$累积的遗憾至少为 $C \cdot N^{\frac{1}{3}}\Delta^{\frac{1}{3}}T^{\frac{2}{3}}$。我们还为偶数和奇数窗口选择略有不同的实例，使得 $\Delta_\infty^K \geq \frac{K}{2} \cdot \Delta_\infty$。证明见附录B。

变化量 Δ_∞^K 的选择。在其他相关的多臂老虎机场景中，包括经典的多臂老虎机（MAB）（Besbes、Gur 和 Zeevi，2014；Wei 和 Luo，2021）以及组合半多臂老虎机（W. Chen 等，2021），动态遗憾界都是以范数变化量 L_∞ 的函数给出的。例如，在 MAB 中，对于满足以下条件的非平稳环境，可以达到与参数无关的上界 $\tilde{O}(N^{\frac{1}{3}}(\Delta_\infty)^{\frac{1}{3}}T^{\frac{2}{3}})$：$N$ 是臂的数量，Δ_∞ 是各臂平均奖励的范数变化量 L_∞。由此引出一个问题：在我们的设定中，能否达到类似的上界？（请注意，下界 $\tilde{\Omega}(N^{\frac{1}{3}}(\Delta_\infty^K)^{\frac{1}{3}}T^{\frac{2}{3}})$ 并不排除这种可能性，因为最坏情况实例可能满足两种范数下的变化量 L_∞^{2K} 和 L_∞ 相同。）然而，在我们的设定中，我们在定理 3 中构造的对抗实例满足 $\Delta_\infty^K \geq \frac{K}{2} \cdot \Delta_\infty$。这意味着，就 Δ_∞ 而言，没有任何多项式时间算法能达到优于 $\tilde{\Omega}((NK)^{\frac{1}{3}}\Delta_\infty^{\frac{1}{3}}T^{\frac{2}{3}})$ 的遗憾。由于 $\Delta_\infty^K \leq 2K \cdot \Delta_\infty$，上界 $\tilde{O}(N^{\frac{1}{3}}(\Delta_\infty^K)^{\frac{1}{3}}T^{\frac{2}{3}})$

是对遗憾的严格更优刻画，优于 $\tilde{O}((KN)^{\frac{1}{3}}(\Delta_\infty)^{\frac{1}{3}}T^{\frac{2}{3}})$，这也证明了在 MNL 多臂老虎机设定中，选择变化量 Δ_∞^K 作为更方便的变化量是合理的。

结论与未来方向

受现实场景的启发，我们研究了非平稳环境中的 MNL-老虎机问题。我们设计了一种算法，其动态遗憾界达到最优（相差对数因子）。我们的分析需要运用新的思路和技术，包括不间断购买过程、偏差的随机界、新的集中界等。我们的工作留下了一些值得探索的未来方向：在许多现实场景中，吸引参数的非平稳性遵循某些结构化模式（有些产品刚进入市场时更具“吸引力”，另一些产品则呈现季节性吸引力模式等）。这引出了一个问题：能否利用这些结构来改进 MNL 多臂老虎机的学习算法？另一个有趣的问题是，在真实（或人工生成的）数据上对我们的算法进行实证评估。

参考文献

Abbasi-Yadkori, Yasin, Andras Gyorgy, 和 Nevena Lazic（2022）。《重新审视非平稳随机多臂老虎机的动态遗憾》。载于：*arXiv*预印本 *arXiv:2201.06532*。Agrawal, Shipra 等（2016）。《一种近乎最优的品类选择探索—利用方法》。载于：*2016*年*ACM*经济学与计算会议论文集。——（2017）。《MNL 多臂老虎机的汤普森采样》。载于：学习理论会议（*COLT*）。Auer, Peter, Nicolo Cesa-Bianchi 等（2002）。《非随机多臂老虎机问题》。载于：*SIAM*计算杂志。Auer, Peter 和 Chao-Kai Chiang（2016）。《一种同时适用于随机和对抗性老虎机、伪遗憾近乎最优的算法》。载于：学习理论会议（*COLT*）。Auer, Peter, Pratik Gajane, 和 Ronald Ortner（2019）。《在分布变化次数未知时自适应追踪最佳老虎机臂》。载于：学习理论会议（*COLT*）。Belzunce, Felix, Carolina Martinez Riquelme, 和 Julio Mulero（2015）。随机序导论。Academic Press。Besbes, Omar, Yonatan Gur, 和 Assaf Zeevi（2014）。《奖励非平稳的随机多臂老虎机问题》。载于：神经信息处理系统进展。——（2015）。《非平稳随机优化》。载于：《运筹学》。Bubeck, S´ebastien 和 Aleksandrs Slivkins（2012）。《两全其美：随机与对抗性老虎机》。载于：学习理论会议（*COLT*）。Chen, Wei 等（2021）。《非平稳环境中的组合半老虎机》。载于：人工智能中的不确定性。

Chen, Xi 和 Yining Wang（2017）。《关于 MNL 多臂老虎机商品组合选择模型紧下界的一则说明》。载于：*arXiv*预印本 *arXiv:1709.06109*。Chen, Yifang 等（2019）。《一种非平稳上下文多臂老虎机新算法：高效、最优且无参数》。载于：学习理论会议（*COLT*）。Cheung, Wang Chi, David Simchi-Levi, 和 Ruihao Zhu（2019）。《学习在非平稳条件下进行优化》。载于：第*22*届人工智能与统计国际会议。Garivier, Aur´elien 和 Eric Moulines（2011）。《关于切换多臂老虎机问题的上置信界策略》。载于：算法学习理论：第*22*届国际会议，*ALT 2011*，芬兰埃斯波，*2011*年*10*月*5*日至*7*日。论文集 *22*。Gittins, John C（1979）。《多臂老虎机过程与动态分配指数》。载于：英国皇家统计学会期刊：*B*辑（方法论）。Hazan, Elad 和 Comandur Seshadhri（2009）。《适用于变化环境的高效学习算法》。载于：第*26*届机器学习年度国际会议论文集。Kleinberg, Robert 和 Nicole Immorlica（2018）。《充能多臂老虎机》。载于：*2018*年*IEEE*第*59*届计算机科学基础年度研讨会（*FOCS*）。Levine, Nir, Koby Crammer, 和 Shie Mannor（2017）。《衰退多臂老虎机》。载于：神经信息处理系统进展。Liu, Fang, Joohyun Lee, 和 Ness Shroff（2018）。《用于分段平稳多臂老虎机问题的基于变化检测的框架》。载于：*AAAI*人工智能会议论文集。Luo, Haipeng 等（2018）。《非平稳环境中的高效上下文多臂老虎机》。载于：学习理论会议（*COLT*）。Maddison, Chris J, Daniel Tarlow, 和 Tom Minka（2014）。《A*采样》。载于：神经信息处理系统进展。Mao, Weichao 等（2020）。《无模型非平稳强化学习：近乎最优的遗憾及其在多智能体强化学习和库存控制中的应用》。载于：*arXiv*预印本 *arXiv:2010.03161*。Robbins, Herbert（1952）。《序贯试验设计的若干方面》。载于。Rusmevichientong, Paat, Zuo-Jun Max Shen, 和 David B Shmoys（2010）。《采用多项式逻辑斯谛选择模型并受容量约束的动态商品组合优化》。载于：《运筹学》。Russac, Yoan, Olivier Capp´e, 和 Aur´elien Garivier（2020）。《非平稳广义线性老虎机算法》。载于：*arXiv*预印本 *arXiv:2003.10113*。Russac, Yoan, Claire Vernade, 和 Olivier Capp´e（2019）。《用于非平稳环境的加权线性多臂老虎机》。载于：神经信息处理系统进展。Saur´e, Denis 和 Assaf Zeevi（2013）。《需求学习下的最优动态商品组合规划》。载于：《制造与服务运营管理》。

Seldin, Yevgeny 和 Aleksandrs Slivkins（2014）。“随机与对抗多臂老虎机通用的实用算法”。载于：国际机器学习会议。Slivkins, Aleksandrs 和 Eli Upfal（2008）。“适应变化环境：布朗休闲多臂老虎机”。载于：学习理论会议（*COLT*）。Suk, Joe 和 Samory Kpotufe（2022）。“多臂老虎机中最显著臂切换的跟踪”。载于：学习理论会议（*COLT*）。Tekin, Cem 和 Mingyan Liu（2012）。“休息型与休闲型多臂老虎机的在线学习”。载于：*IEEE*信息论汇刊。Touati, Ahmed 和 Pascal Vincent（2020）。“非平稳线性马尔可夫决策过程中的高效学习”。载于：*arXiv* 预印本 *arXiv:2010.12870*。Wei, Chen-Yu、Yi-Te Hong 和 Chi-Jen Lu（2016）。“非平稳随机环境中最佳专家的跟踪”。载于：神经信息处理系统进展。Wei, Chen-Yu 和 Haipeng Luo（2021）。“无先验知识的非平稳强化学习：一种最优黑盒方法”。载于：学习理论会议（*COLT*）。Whittle, Peter（1988）。“休闲多臂老虎机：变化世界中的活动分配”。载于：应用概率学杂志。Yang, Tianbao 等（2016）。“跟踪缓慢移动的先知：使用真实梯度与噪声梯度的在线学习的最优动态遗憾”。载于：国际机器学习会议。Zhao, Peng 等（2020）。“非平稳线性多臂老虎机的简单方法”。载于：人工智能与统计国际会议。

A 分析：省略的证明

引理1。 设 $t \in [1,T]$ 满足 $\delta^{(t)} \leq \frac{1}{2}$，且设 $j \in [N]$。设 $k \geq 1$、 $\tau \in [1,t]$、 $s \geq 1$和$S \in \Gamma(s)$。令 X 表示在给定 $s_j^{(t)}(k) = s$ 和 $S(s) = S$的条件下的随机变量 $\zeta_j^{(t)}(k)$。令X^+ 为均值为 $\mu_j^{+(t)}(\tau)$的几何随机变量，并令 X^- 为均值为 $\mu_j^{-(t)}(\tau)$的几何随机变量。那么，

$$X^- \leq_{st} X \leq_{st} X^+,$$

其中 $\leq_{st}$ 表示通常随机序。

证明。 我们先回顾一下类别分布的效用表示。

类别分布的效用表示：设 ϵ_1，……， ϵ_m 为 m 服从 Gumbel(0,1) 分布的随机变量。设 $\bar{u}_1$，……，$\bar{u}_m$ 为 m 常数。定义随机变量（u_i）$_{i\in[m]}$使得 $u_i = \bar{u}_i + \epsilon_i$ 对每个 $i \in [m]$均成立。我们将这些变量称为效用。众所周知（例如参见 Maddison、Tarlow 和 Minka，2014），效用的最大值位置

服从类别分布，使得：

$$\mathbb{P}\left(\operatorname*{argmax}_k u_k = j\right) = \frac{e^{\bar{u}_j}}{\sum_{k=1}^{m} e^{\bar{u}_k}}.$$

在此以及论文的其余部分中，由于我们处理的是连续随机变量，最大值位置几乎必然只包含一个元素。因此，在不失一般性的情况下，我们将使用“argmax $u_k = j$”这一记号来指代“argmax $u_k = \{j\}$”。

用效用刻画 X 的分布：设 $(\epsilon_u(m))_{m\geq s,\ u\in S\cup\{0\}}$ 为独立同分布的 Gum- bel(0,1) 变量，并且对每个 $m \geq s$，令

$$Z_m := \operatorname*{argmax}_{u\in S\cup\{0\}} \left\{ \log\left(\omega_u(\min\{m,t\})\right) + \epsilon_u(m) \right\},$$

其中，对每个 $m \geq s$，有 $\omega_0(k) = 1$；并且在论文的其余部分中，我们采用以下约定：$\log(0) = -\infty$，且对任意 $a \in \mathbb{R}$，有 $\mathbb{P}(-\infty \geq a) = \mathbb{P}(-\infty \geq -\infty) = 0$。则：

命题1。 X 与以下随机变量具有相同的分布，

$$\bar{X} = \sum_{m=s}^{\operatorname{argmin}_{r\geq s}\{Z_r=0\}-1} \mathbf{1}\{Z_m = j\}$$

证明。首先，请注意，X 等于该随机和

$$\sum_{m=s}^{\operatorname{argmin}_{r\geq s}\{j(r)=0\}-1} \mathbf{1}\{j(m) = j\}$$

在给定 $s_j^{(t)}(k) = s$ 和 $S(s) = S$ 的条件下。我们分两步证明这一结论。首先，我们证明 $\operatorname{argmin}_{r\geq s}\{Z_r = 0\}$ 与在给定 $s_j^{(t)}(k) = s$和 $S(s) = S$ 的条件下的 $\operatorname{argmin}_{r\geq s}\{j(r) = 0\}$ 具有相同的分布。然后，我们证明，对于每个整数 $b \geq s$，在给定 $\operatorname{argmin}_{r\geq s}\{Z_r = 0\} = b$ 的条件下，随机变量 $\sum_{m=s}^{b-1} \mathbf{1}\{Z_m = j\}$ 与在给定 $\operatorname{argmin}_{r\geq s}\{j(r) = 0\} = b$ 和 $s_j^{(t)}(k) = s$ 和 $S(s) = S$ 的条件下的 $\sum_{m=s}^{b-1} \mathbf{1}\{j(m) = j\}$ 具有相同的分布。由这两个事实即可轻易得出结论。

步骤 1。 对于每个整数 $b \geq s$，我们有，

$$
\begin{aligned}
&\mathbb{P}\left(\operatorname*{argmin}_{r\geq s}\{j(r)=0\}=b \,\middle|\, s_j^{(t)}(k)=s, S(s)=S\right)\\
&=\mathbb{P}\left(j(b)=0, j(b-1)\neq 0,\ldots,j(s)\neq 0 \,\middle|\, s_j^{(t)}(k)=s, S(s)=S\right)\\
&=\mathbb{P}\left(j(b)=0 \,\middle|\, j(b-1)\neq 0,\ \ldots\ldots,\ j(s)\neq 0,\ s_j^{(t)}(k)=s,\ S(s)=S\right)\times\\
&\quad\mathbb{P}\left(j(b-1)\neq 0 \,\middle|\, j(b-2)\neq 0,\ \ldots\ldots,\ j(s)\neq 0,\ s_j^{(t)}(k)=s,\ S(s)=S\right)\times\\
&\quad\cdots\\
&\quad\mathbb{P}\left(j(s)\neq 0 \,\middle|\, s_j^{(t)}(k)=s,\ S(s)=S\right)\\
&=\mathbb{P}\left(j(b)=0\,|S(b)=S\right)\mathbb{P}\left(j(b-1)\neq 0\,|S(b-1)=S\right)\ldots\mathbb{P}\left(j(s)\neq 0\,|S(s)=S\right)\\
&=\frac{1}{1+\sum_{r\in S}\omega_r(\min\{b,t\})}\times\left(1-\frac{1}{1+\sum_{r\in S}\omega_r(\min\{b-1,t\})}\right)\ldots\left(1-\frac{1}{1+\sum_{r\in S}\omega_r(\min\{s,t\})}\right)\\
&=\mathbb{P}\left(Z_b=0\right)\mathbb{P}\left(Z_{b-1}\neq 0\right)\ldots\mathbb{P}\left(Z_s\neq 0\right)\\
&=\mathbb{P}\left(Z_b=0, Z_{b-1}\neq 0,\ldots,Z_s\neq 0\right)\\
&=\mathbb{P}\left(\operatorname*{argmin}_{r\geq s}\{Z_r=0\}=b\right).
\end{aligned}
$$

第三个等式源于以下事实：对于每个 $r \geq s$，$\{j(r) \neq 0,\ \ldots\ldots,\ j(s) \neq 0,\ s_j^{(t)}(k)=s,\ S(s)=S\}$ 都意味着 $\{S(r+1)=S\}$，并且在给定 $\{S(r+1)=S\}$ 的条件下，随机变量 $j(r+1)$ 与 r 之前的所有过去时间步相互独立。第五个等式由类别分布的效用刻画得出，最后，第六个等式源于 $(Z_r)_{r\geq s}$ 的独立性（因为它们是使用相互独立的 Gumbel变量构造的）。

步骤 2. 固定一个整数 $b \geq s$，并令 $m \neq m' \in [s, b-1]$ 和 $j' \in S \cup \{0\}$，我们有，

$$
\begin{aligned}
&\mathbb{P}\left(j(m)=j \,\middle|\, j(m')=j', \operatorname*{argmin}_{r\geq s}\{j(r)=0\}=b, s_j^{(t)}(k)=s, S(s)=S\right)\\
&=\mathbb{P}\left(j(m)=j \,|\, j(m)\neq 0, S(m)=S\right)\\
&=\frac{\mathbb{P}\left(j(m)=j, j(m)\neq 0 \,|\, S(m)=S\right)}{\mathbb{P}\left(j(m)\neq 0 \,|\, S(m)=S\right)}\\
&=\frac{\mathbb{P}\left(j(m)=j \,|\, S(m)=S\right)}{\mathbb{P}\left(j(m)\neq 0 \,|\, S(m)=S\right)}\\
&=\frac{\omega_j(\min\{m,t\})}{\sum_{r\in S}\omega_j(\min\{m,t\})}
\end{aligned}
$$

第一个等号源于以下事实：$\{j(m')=j'$，最小值位置$r\geq s\{j(r)=0\}=b, s_j^{(t)}(k)=s, S(s)=S\}$ 蕴含 $\{j(m)\neq 0, S(m)=S\}$；接着，在给定 $\{j(m)\neq 0, S(m)=S\}$ 的条件下，所选商品 $j(m)$ 与未来轮次相互独立（未来轮次仅取决于$j(m)=0$ 是否成立；一旦固定 $j(m)\neq 0$，未来轮次便与实际抽取的商品 $j(m)$ 相互独立），且与过去轮次相互独立（一旦 $S(m)=S$ 的商品组合固定，无论哪条样本路径导向该商品组合， $j(m)$ 都具有相同的分布）。特别地，这意味着在给定事件 $\{\operatorname{argmin}_{r\geq s}\{j(r)=0\}=b, s_j^{(t)}(k)=s, S(s)=S\}$ 的条件下，变量 $(\mathbf{1}\{j(m)=j\})_{m\in[s,b-1]}$ 独立同分布，且服从均值为 $p=\frac{\omega_j(\min\{m,t\})}{\sum_{r\in S}\omega_j(\min\{m,t\})}$ 的伯努利分布。接下来，我们有，

$$
\begin{aligned}
&\mathbb{P}\left(Z_m=j \,\middle|\, Z_{m'}=j', \operatorname*{argmin}_{r\geq s}\{Z_r=0\}=b\right)\\
&=\mathbb{P}\left(Z_m=j \,|\, Z_{m'}=j', Z_s\neq 0,\ldots,Z_{b-1}\neq 0, Z_b=0\right)\\
&=\mathbb{P}\left(Z_m=j \,|\, Z_m\neq 0\right)\\
&=\frac{\mathbb{P}\left(Z_m=j, Z_m\neq 0\right)}{\mathbb{P}\left(Z_m\neq 0\right)}\\
&=\frac{\mathbb{P}\left(Z_m=j\right)}{\mathbb{P}\left(Z_m\neq 0\right)}\\
&=\frac{\omega_j(\min\{m,t\})}{\sum_{r\in S}\omega_j(\min\{m,t\})},
\end{aligned}
$$

其中第二个等式源于变量 $(Z_m)_{m\geq s}$ 相互独立这一事实。这尤其意味着，在给定 $\{\operatorname{argmin}_{r\geq s}\{Z_r=0\}=b\}$ 的条件下，变量 $(\mathbf{1}\{Z_m=j\})_{m\in[s,b-1]}$ 是均值为 $p=\frac{\omega_j(\min\{m,t\})}{\sum_{r\in S}\omega_j(\min\{m,t\})}$ 的独立同分布伯努利变量。因此，前一组变量和后一组变量具有相同的联合分布，这意味着它们各自的和具有相同的

分布。最后请注意，对于每个 $a \in \mathbb{N}$，我们有：

$$\mathbb{P}(X=a) = \mathbb{P}\left(\sum_{m=s}^{\operatorname{argmin}_{r\geq s}\{j(r)=0\}-1} \mathbf{1}\{j(m)=j\} = a \,\middle|\, s_j^{(t)}(k)=s, S(s)=S\right)$$

$$= \sum_{b=s}^{\infty} \mathbb{P}\left(\sum_{m=s}^{b-1} \mathbf{1}\{j(m)=j\} = a \,\middle|\, s_j^{(t)}(k)=s, S(s)=S, \operatorname*{argmin}_{r\geq s}\{j(r)=0\}=b\right)$$
$$\times \mathbb{P}\left(\operatorname*{argmin}_{r\geq s}\{j(r)=0\}=b \,\middle|\, s_j^{(t)}(k)=s, S(s)=S\right)$$

$$= \sum_{b=s}^{\infty} \mathbb{P}\left(\sum_{m=s}^{b-1} \mathbf{1}\{Z_m=j\}=a \,\middle|\, \operatorname*{argmin}_{r\geq s}\{Z_r=0\}=b\right) \times \mathbb{P}\left(\operatorname*{argmin}_{r\geq s}\{Z_r=0\}=b\right)$$
$$= \mathbb{P}\left(\bar{X}=a\right)$$

□

耦合与随机界的证明：我们区分两种情况：

情况 $s \geq t+1$： 在这种情况下，随机变量 X服从几何分布，其均值为 $\omega_j(t)$。事实上，对于每个 $a \in \mathbb{N}$，我们都有：

$$\mathbb{P}\left(\bar{X}=a\right) = \sum_{b=s+a}^{\infty} \mathbb{P}\left(\sum_{m=s}^{b-1} \mathbf{1}\{Z_m=j\}=a \,\middle|\, \operatorname*{argmin}_{r\geq s}\{Z_r=0\}=b\right) \mathbb{P}\left(\operatorname*{argmin}_{r\geq s}\{Z_r=0\}=b\right)$$
$$= \sum_{b=s+a}^{\infty} \binom{b-s}{a} p^a (1-p)^{b-s-a} \cdot (1-q)^{b-s} q$$
$$= \sum_{b=0}^{\infty} \binom{b+a}{a} p^a (1-p)^b \cdot (1-q)^{b+a} q$$
$$= qp^a(1-q)^a \sum_{b=0}^{\infty} \binom{b+a}{a} ((1-p)(1-q))^b$$
$$= \frac{q \cdot p^a (1-q)^a}{(1-(1-q)(1-p))^{a+1}}$$
$$= \frac{q}{q+s}\left(1-\frac{q}{q+s}\right)^a,$$

其中 $p = \frac{\omega_j(t)}{\sum_{r\in S}\omega_r(t)}$、$q = \frac{1}{1+\sum_{r\in S}\omega_r(t)}$ 和 $s = p(1-q)$。第二个等号由以下事实得出：$(\mathbf{1}\{Z_m=j\})_{m\in[s,b-1]}$ 是均值为 p 的独立同分布伯努利变量，并且随机变量最小值位置 $\operatorname*{argmin}_{r\geq s}\{Z_r=0\}$ 根据独立性服从成功概率为 q 的几何分布

$^{-}(Z_m)_{m\geq s}$。因此， X 服从均值为

$$\frac{1-\left(\frac{q}{q+s}\right)}{\left(\frac{q}{q+s}\right)}=\omega_j(t).$$

现在，由于几何随机变量的均值增大时，其随机序也随之增大，并且

$$\mu_j^{-(t)}(\tau)\leq\omega_j(t)\leq\mu_j^{+(t)}(\tau)$$

成立，因此 $X^-\leq_{st}\bar{X}\leq_{st}X^+$ ，进而 $X^-\leq_{st}X\leq_{st}X^+$。

情形 $s\leq t$**:** 对于这种情形，我们使用定义 Z_m 时所用的相同 Gumbel变量，定义两个随机变量 $\bar{X}^+$ 和 $\bar{X}^-$ ，并证明，

$$X^-\leq_{st}\bar{X}^-\leq_{st}X\leq_{st}\bar{X}^+\leq_{st}X^+.$$

特别地，对于每个 $m\geq s$，定义

$$A_m^+:=\text{argmax}\left\{\log\left(\omega_j(\tau)+\delta_j^{(t)}\right)+\epsilon_j(m)\,,\left\{\log\left(\omega_u(\tau)-\delta_j^{(t)}\right)^+\!+\epsilon_u(m)\right\}_{u\in S\setminus\{j\}}\,,\,\epsilon_0(m)\right\},$$

and,

$$B_m^+:=\text{argmax}\left\{\left\{\log\left(\omega_u(\tau)+\delta_j^{(t)}\right)+\epsilon_u(m)\right\}_{u\in S}\,,\,\epsilon_0(m)\right\},$$

类似地，定义，

$$A_m^-:=\text{argmax}\left\{\log\left(\omega_j(\tau)-\delta_j^{(t)}\right)^+\!+\epsilon_j(m)\,,\left\{\log\left(\omega_u(\tau)+\delta_j^{(t)}\right)+\epsilon_u(m)\right\}_{u\in S\setminus\{j\}}\,,\,\epsilon_0(m)\right\},$$

and,

$$B_m^-:=\text{argmax}\left\{\left\{\log\left(\omega_u(\tau)-\delta_j^{(t)}\right)^+\!+\epsilon_u(m)\right\}_{u\in S}\,,\,\epsilon_0(m)\right\},$$

然后令

$$\bar{X}^+=\sum_{m=s}^{\text{argmin}_{r\geq s}\{B_r^+=0\}-1}\mathbf{1}\{A_m^+=j\},$$

and,

$$\bar{X}^-=\sum_{m=s}^{\text{argmin}_{r\geq s}\{B_r^-=0\}-1}\mathbf{1}\{A_m^-=j\}.$$

我们证明， $\bar{X}^-\leq\bar{X}\leq\bar{X}^+$ 几乎必然成立，这意味着$\bar{X}^-\leq_{st}X\leq_{st}\bar{X}^+$。我们有，

$$
\begin{aligned}
B_r^+ = 0 &\implies \epsilon_0(r) \geq \log\left(\omega_u(\tau) + \delta_j^{(t)}\right) + \epsilon_u(r) \quad \forall u \in S \\
&\implies \epsilon_0(r) \geq \log\left(\omega_u(\min\{r,t\})\right) + \epsilon_u(r) \quad \forall u \in S \\
&\implies Z_r = 0.
\end{aligned}
$$

由于 $\omega_u(\tau) + \delta_j^{(t)} \geq \omega_u(\min\{r,t\})$。因此，

$$
\operatorname*{argmin}_{r \geq s}\{B_r^+ = 0\} \geq \operatorname*{argmin}_{r \geq s}\{Z_r = 0\}
$$

接下来，对于每个 $m \leq$ 最小值位置$_{r \geq s}\{Z_r = 0\}$，我们有，

$$
\begin{aligned}
Z_m = j &\implies \log\left(\omega_j(\min\{m,t\})\right) + \epsilon_j(m) \geq \log\left(\omega_u(\min\{m,t\})\right) + \epsilon_u(m) \quad \forall u \in S \cup \{0\} \\
&\implies \begin{cases} \log\left(\omega_j(\tau) + \delta_j^{(t)}\right) + \epsilon_j(m) \geq \log\left(\omega_u(\tau) - \delta_j^{(t)}\right)^+ + \epsilon_u(m) & \forall u \in S \backslash \{j\} \\ \log\left(\omega_j(\tau) + \delta_j^{(t)}\right) + \epsilon_j(m) \geq \epsilon_0(m) \end{cases} \\
&\implies A_m^+ = j.
\end{aligned}
$$

由于，$(\omega_u(\tau) - \delta_j^{(t)})^+ \leq \omega_u(\min\{m,t\}) \leq \omega_u(\tau) + \delta_j^{(t)}$。因此，

$$
\mathbf{1}\{Z_m = j\} \leq \mathbf{1}\{A_m^+ = j\}
$$

由上述内容可得，几乎必然，

$$
\begin{aligned}
\bar{X}^+ &= \sum_{m=s}^{\operatorname{argmin}_{r \geq s}\{B_r^+ = 0\} - 1} \mathbf{1}\{A_m^+ = j\} \\
&\geq \sum_{m=s}^{\operatorname{argmin}_{r \geq s}\{Z_r = 0\} - 1} \mathbf{1}\{A_m^+ = j\} \\
&\geq \sum_{m=s}^{\operatorname{argmin}_{r \geq s}\{Z_r = 0\} - 1} \mathbf{1}\{Z_m = j\} \\
&= \bar{X}.
\end{aligned}
$$

通过类似的论证，我们得到：

$$
\operatorname*{argmin}_{r \geq s}\{B_r^- = 0\} \leq \operatorname*{argmin}_{r \geq s}\{Z_r = 0\}
$$

并且,

$$\mathbf{1}\{Z_m = j\} \geq \mathbf{1}\{A_m^- = j\}$$

这意味着

$$\bar{X}^- \leq \sum_{m=s}^{\text{argmin}_{r\geq s}\{Z_r=0\}-1} \mathbf{1}\{Z_m = j\} = \bar{X}.$$

现在我们证明 $\bar{X}^+ \leq_{st} X^+$ 和 $\bar{X}^- \geq_{st} X^-$。首先证明 $\bar{X}^+$ 和 $\bar{X}^-$是几何随机变量，其均值为

$$\bar{\mu}^+ = (\omega_j(\tau) + \delta_j^{(t)}) \frac{1 + \sum_{r\in S} \omega_r(\tau) + \delta_r^{(t)}}{1 + \omega_j(\tau) + \delta_j^{(t)} + \sum_{r\in S\setminus\{j\}} (\omega_r(\tau) - \delta_r^{(t)})^+}$$

and

$$\bar{\mu}^- = (\omega_j(\tau) - \delta_j^{(t)})^+ \frac{1 + \sum_{r\in S} (\omega_r(\tau) - \delta_r^{(t)})^+}{1 + (\omega_j(\tau) - \delta_j^{(t)})^+ + \sum_{r\in S\setminus\{j\}} (\omega_r(\tau) + \delta_r^{(t)})}$$

r分别。通过与第一种情况类似的论证，对于每个 $a \in \mathbb{N}$，我们有 ,

$$\begin{aligned}
\mathbb{P}\left(\bar{X}^+ = a\right) &= \mathbb{P}\left(\sum_{m=s}^{\text{argmin}_{r\geq s}\{B_r^+=0\}-1} \mathbf{1}\{A_m^+ = j\} = a\right) \\
&= \sum_{b=s+a}^{\infty} \mathbb{P}\left(\sum_{m=s}^{b-1} \mathbf{1}\{A_m^+ = j\} = a \,\middle|\, \underset{r\geq s}{\text{argmin}}\{B_r^+ = 0\} = b\right) \mathbb{P}\left(\underset{r\geq s}{\text{argmin}}\{B_r^+ = 0\} = b\right) \\
&= \sum_{b=s+a}^{\infty} \binom{b-s}{a} p^a (1-p)^{b-s-a} \cdot (1-q)^{b-s} q \\
&= \frac{q}{q+s} \left(1 - \frac{q}{q+s}\right)^a,
\end{aligned}$$

其中,

$$\begin{aligned}
p &:= \mathbb{P}\left(A_1^+ = j \,\middle|\, B_1^+ \neq 0\right) \\
&= \frac{\mathbb{P}\left(A_1^+ = j \ , \ B_1^+ \neq 0\right)}{\mathbb{P}\left(B_1^+ \neq 0\right)} \\
&= \frac{\mathbb{P}\left(A_1^+ = j\right)}{\mathbb{P}\left(B_1^+ \neq 0\right)} \\
&= \left(\frac{\omega_j(\tau) + \delta_j^{(t)}}{1 + \omega_j(\tau) + \delta_j^{(t)} + \sum_{r\in S\setminus\{j\}} (\omega_r(\tau) - \delta_r^{(t)})^+}\right) \cdot \left(\frac{1 + \sum_{r\in S} (\omega_r(\tau) + \delta_r^{(t)})}{\sum_{r\in S} (\omega_r(\tau) + \delta_r^{(t)})}\right),
\end{aligned}$$

并且， $q := \mathbb{P}\left(B_1^+ = 0\right) = \frac{1}{1+\sum_{r\in S}(\omega_r(\tau)+\delta_r^{(t)})}$，并且， $s = p(1-q)$。因此， $\bar{X}^+$ 是一个成功概率为.....的几何随机变量，

$$\frac{q}{q+s} = \frac{1}{1+(\omega_j(\tau)+\delta_j^{(t)})\frac{1+\sum_{r\in S}\omega_r(\tau)+\delta_r^{(t)}}{1+\omega_j(\tau)+\delta_j^{(t)}+\sum_{r\in S\setminus\{j\}}(\omega_r(\tau)-\delta_r^{(t)})^+}},$$

因此均值

$$\bar{\mu}^+ = (\omega_j(\tau)+\delta_j^{(t)})\frac{1+\sum_{r\in S}\omega_r(\tau)+\delta_r^{(t)}}{1+\omega_j(\tau)+\delta_j^{(t)}+\sum_{r\in S\setminus\{j\}}(\omega_r(\tau)-\delta_r^{(t)})^+}.$$

$^-$类似地，对于 X^-，我们对每个 $a\in\mathbb{N}$ 都有，

$$\begin{aligned}
\mathbb{P}\left(\bar{X}^- = a\right) &= \mathbb{P}\left(\sum_{m=s}^{\operatorname{argmin}_{r\geq s}\{B_r^-=0\}-1} \mathbf{1}\{A_m^- = j\} = a\right) \\
&= \sum_{b=s+a}^{\infty} \mathbb{P}\left(\sum_{m=s}^{b-1}\mathbf{1}\{A_m^- = j\} = a \,\middle|\, \operatorname*{argmin}_{r\geq s}\{B_r^- = 0\} = b\right)\mathbb{P}\left(\operatorname*{argmin}_{r\geq s}\{B_r^- = 0\} = b\right) \\
&= \sum_{b=s+a}^{\infty}\binom{b-s}{a}p^a(1-p)^{b-s-a}\cdot(1-q)^{b-s}q \\
&= \frac{q}{q+s}\left(1-\frac{q}{q+s}\right)^a,
\end{aligned}$$

其中，

$$\begin{aligned}
p &:= \mathbb{P}\left(A_1^- = j \,\middle|\, B_1^- \neq 0\right) \\
&= \frac{\mathbb{P}\left(A_1^- = j\ ,\ B_1^- \neq 0\right)}{\mathbb{P}\left(B_1^- \neq 0\right)} \\
&= \frac{\mathbb{P}\left(A_1^- = j\right)}{\mathbb{P}\left(B_1^- \neq 0\right)} \\
&= \left(\frac{(\omega_j(\tau)-\delta_j^{(t)})^+}{1+(\omega_j(\tau)-\delta_j^{(t)})^+ + \sum_{r\in S\setminus\{j\}}(\omega_r(\tau)+\delta_r^{(t)})}\right)\cdot\left(\frac{1+\sum_{r\in S}(\omega_r(\tau)-\delta_r^{(t)})^+}{\sum_{r\in S}(\omega_r(\tau)-\delta_r^{(t)})^+}\right),
\end{aligned}$$

并且， $q := \mathbb{P}\left(B_1^- = 0\right) = \frac{1}{1+\sum_{r\in S}\omega_r((\tau)-\delta_r^{(t)})^+}$，并且， $s = p(1-q)$。因此， $\bar{X}^-$ 是一个成功概率为

$$\frac{q}{q+s} = \frac{1}{1+(\omega_j(\tau)-\delta_j^{(t)})^+\frac{1+\sum_{r\in S}(\omega_r(\tau)-\delta_r^{(t)})^+}{1+(\omega_j(\tau)-\delta_j^{(t)})^+ + \sum_{r\in S\setminus\{j\}}(\omega_r(\tau)+\delta_r^{(t)})}},$$

因此均值

$$\bar{\mu}^- = (\omega_j(\tau) - \delta_j^{(t)})^+ \frac{1 + \sum_{r\in S}(\omega_r(\tau) - \delta_r^{(t)})^+}{1 + (\omega_j(\tau) - \delta_j^{(t)})^+ + \sum_{r\in S\setminus\{j\}}(\omega_r(\tau) + \delta_r^{(t)})}$$

为完成证明，我们对 $\bar{X}^+$ 和 $\bar{X}^-$ 的均值作如下界定：

$$\frac{1+\delta^{(t)}}{1-\delta^{(t)}} - \frac{1 + \sum_{r\in S}\omega_r(\tau) + \delta_r^{(t)}}{1 + \omega_j(\tau) + \delta_j^{(t)} + \sum_{r\in S\setminus\{j\}}(\omega_r(\tau) - \delta_r^{(t)})^+}$$

$$\geq \frac{1+\delta^{(t)}}{1-\delta^{(t)}} - \frac{1 + \sum_{r\in S}\omega_r(\tau) + \delta_r^{(t)}}{1 + \omega_j(\tau) + \delta_j^{(t)} + \sum_{r\in S\setminus\{j\}}(\omega_r(\tau) - \delta_r^{(t)})}$$

$$= \frac{1}{(1-\delta^{(t)})(1 + \omega_j(\tau) + \delta_j^{(t)} + \sum_{r\in S\setminus\{j\}}(\omega_r(\tau) - \delta_r^{(t)}))} \left[\left(1 + \omega_j(\tau) + \delta_j^{(t)} + \sum_{r\in S\setminus\{j\}} \omega_r(\tau) - \sum_{r\in S\setminus\{j\}} \delta_r^{(t)} \right) \right.$$
$$+ \left(\delta^{(t)} + \delta^{(t)}\omega_j(\tau) + \delta^{(t)}\delta_j^{(t)} + \delta^{(t)} \sum_{r\in S\setminus\{j\}} \omega_r(\tau) - \delta^{(t)} \sum_{r\in S\setminus\{j\}} \delta_r^{(t)} \right)$$
$$\left. + \left(-1 - \sum_{r\in S} \omega_r(\tau) - \sum_{r\in S} \delta_r^{(t)} \right) + \left(\delta^{(t)} + \delta^{(t)} \sum_{r\in S} \omega_r(\tau) + \delta^{(t)} \sum_{r\in S} \delta_r^{(t)} \right) \right]$$

$$= \frac{-2\sum_{r\in S\setminus\{j\}} \delta_r^{(t)} + 2\delta^{(t)} + 2\delta^{(t)} \sum_{r\in S} \omega_r(\tau) + 2\delta^{(t)}\delta_j^{(t)}}{(1-\delta^{(t)})(1 + \omega_j(\tau) + \delta_j^{(t)} + \sum_{r\in S\setminus\{j\}}(\omega_r(\tau) - \delta_r^{(t)}))} \geq 0,$$

其中，最后一个不等式是由以下事实得出的：当 $s \leq t$ 时，$\delta^{(t)} \geq \sum_{r\in S\setminus\{j\}}\delta_r^{(t)}$如 $|S| \leq K$ 。因此，

$$\bar{\mu}^+ \leq \mu_j^{+(t)}(\tau).$$

由于几何随机变量的均值增加时，其随机序也随之增大，因此我们有$_X^+ \leq_{st} X^+$。同样，我们有，

$$\frac{1-\delta^{(t)}}{1+\delta^{(t)}} - \frac{1+\sum_{r\in S}(\omega_r(\tau)-\delta_r^{(t)})^+}{1+(\omega_j(\tau)-\delta_j^{(t)})^+ + \sum_{r\in S\setminus\{j\}}(\omega_r(\tau)+\delta_r^{(t)})}$$

$$\leq \frac{1-\delta^{(t)}}{1+\delta^{(t)}} - \frac{1+\sum_{r\in S}(\omega_r(\tau)-\delta_r^{(t)})}{1+\omega_j(\tau)+\sum_{r\in S\setminus\{j\}}(\omega_r(\tau)+\delta_r^{(t)})}$$

$$= \frac{1}{(1+\delta^{(t)})(1+\omega_j(\tau)+\sum_{r\in S\setminus\{j\}}(\omega_r(\tau)+\delta_r^{(t)}))}\left[\left(1+\omega_j(\tau)+\sum_{r\in S\setminus\{j\}}\omega_r(\tau)+\sum_{r\in S\setminus\{j\}}\delta_r^{(t)}\right)\right.$$
$$+\left(-\delta^{(t)}-\delta^{(t)}\omega_j(\tau)-\delta^{(t)}\sum_{r\in S\setminus\{j\}}\omega_r(\tau)-\delta^{(t)}\sum_{r\in S\setminus\{j\}}\delta_r^{(t)}\right)$$
$$\left.+\left(-1-\sum_{r\in S}\omega_r(\tau)+\sum_{r\in S}\delta_r^{(t)}\right)+\left(-\delta^{(t)}-\delta^{(t)}\sum_{r\in S}\omega_r(\tau)+\delta^{(t)}\sum_{r\in S}\delta_r^{(t)}\right)\right]$$

$$= \frac{\sum_{r\in S\setminus\{j\}}\delta_r^{(t)}+\sum_{r\in S}\delta_r^{(t)}-2\delta^{(t)}-2\delta^{(t)}\sum_{r\in S}\omega_r(\tau)+\delta^{(t)}\delta_j^{(t)}}{(1+\delta^{(t)})(1+\omega_j(\tau)+\sum_{r\in S\setminus\{j\}}(\omega_r(\tau)+\delta_r^{(t)}))}$$
$$\leq \frac{2(\sum_{r\in S}\delta_r^{(t)}-\delta^{(t)})+(\delta^{(t)}\delta_j^{(t)}-\delta_j^{(t)})}{(1+\delta^{(t)})(1+\omega_j(\tau)+\sum_{r\in S\setminus\{j\}}(\omega_r(\tau)+\delta_r^{(t)}))} \leq 0.$$

最后一个不等式由以下事实得出：当 $s \leq t$ 时，$\delta^{(t)} \geq \sum_{r\in S}\delta_r^{(t)}$趋于 $|S| \leq K$，且$\delta^{(t)} \leq \frac{1}{2} < 1$。因此，

$$\mu_j^{-(t)}(\tau) \leq \bar{\mu}^-.$$

由于几何随机变量的均值增大时，其随机序也随之增大，因此有$X^- \leq_{st} X^-$。

□

引理 2. 令 $t \in [1,T]$ 满足$\delta^{(t)} \leq \frac{1}{12}$，并令 $j \in [N]$。令 $k \geq 1$、$\tau \in [1,t]$、$s \geq 1$和$S \in \Gamma(s)$，则对于每个 $0 \leq \lambda < \log\left(1+\mu_j^{+(t)}(\tau)\right)$，我们有，

$$\mathbb{E}\left(e^{\lambda\zeta_j^{(t)}(k)}\middle| s_j^{(t)}(k)=s\ ,\ S(s)=S\right) \leq \frac{1}{1-\mu_j^{+(t)}(\tau)\cdot(e^\lambda-1)}.$$

对于每个 $\lambda \leq 0$，我们有，

$$\mathbb{E}\left(e^{\lambda\zeta_j^{(t)}(k)}\middle| s_j^{(t)}(k)=s\ ,\ S(s)=S\right) \leq \frac{1}{1-\mu_j^{-(t)}(\tau)\cdot(e^{\lambda}-1)}$$

证明。令 $t \in [1,T]$ 使得 $\delta^{(t)} \leq \frac{1}{2}$ 且 $j \in [N]$。令 $s \geq 1$、$S \in \Gamma(s)$、$k \geq 1$ 和 $\tau \in [1,t]$。令 X 表示在 $s_j^{(t)}(k)=s$ 且 $S(s)=S$ 条件下的随机变量 $\zeta_j^{(t)}(k)$。对于每个 $0 \leq \lambda < \log\left(1+\frac{1}{\mu_j^{+(t)}(\tau)}\right)$，根据引理1，$X \leq_{st} X^+$ 其中 X^+ 是均值为 $\mu_j^{+(t)}(\tau)$ 的几何随机变量。这尤其意味着，对于每个 $M>0$，$X\cdot\mathbf{1}\{X\leq M\} \leq_{st} X^+$。随机序的一个经典结果表明，对于非递减函数 ϕ，当 $Z \leq_{st} Y$ 且 $\mathbb{E}(\phi(Z))$ 和 $\mathbb{E}(\phi(Y))$ 均存在时，有 $\mathbb{E}(\phi(Z)) \leq \mathbb{E}(\phi(Y))$（例如参见 Belzunce、Riquelme 和 Mulero（2015）的定理2.2.5）。因此，

$$\mathbb{E}\left(e^{\lambda X\cdot\mathbf{1}\{X\leq M\}}\right) \leq \mathbb{E}\left(e^{\lambda X^+}\right)$$

这意味着 $\mathbb{E}\left(e^{\lambda X}\right)$ 存在，且

$$\mathbb{E}\left(e^{\lambda X}\right) \leq \mathbb{E}\left(e^{\lambda X^+}\right) = \frac{1}{1-\mu_j^{+(t)}(\tau)\cdot(e^{\lambda}-1)}$$

根据单调收敛定理。注意到，得出第一个不等式：

$$\mathbb{E}\left(e^{\lambda\zeta_j^{(t)}(k)}\middle| s_j^{(t)}(k)=s\ ,\ S(s)=S\right) = \mathbb{E}\left(e^{\lambda X}\right).$$

类似地，对任意 $\lambda \leq 0$，由引理 1，$X \geq_{st} X^-$ 其中 X^- 是一个均值为 $\mu_j^{-(t)}(\tau)$ 的几何随机变量，这意味着 $\lambda X\cdot\mathbf{1}\{X\leq M\} \leq_{st} \lambda X^-$ 对任意 $M>0$ 成立，且

$$\mathbb{E}\left(e^{\lambda X\cdot\mathbf{1}\{X\leq M\}}\right) \leq \mathbb{E}\left(e^{\lambda X^-}\right)$$

最后，

$$\mathbb{E}\left(e^{\lambda X}\right) \leq \mathbb{E}\left(e^{\lambda X^-}\right) = \frac{1}{1-\mu_j^{-(t)}(\tau)\cdot(e^{\lambda}-1)}.$$

注意到，得出第二个不等式：

$$\mathbb{E}\left(e^{\lambda\zeta_j^{(t)}(k)}\middle| s_j^{(t)}(k)=s\ ,\ S(s)=S\right) = \mathbb{E}\left(e^{\lambda X}\right).$$

□

引理 3。令 $t \in [1,T]$ 使得 $\delta^{(t)} \leq \frac{1}{2}$，并令 $j \in [N]$。令 $k \geq 1$ 和 $\tau \in [1,t]$。对于每个 ry

$\eta > 0$，我们有：

$$\mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i) < (1-\eta)\mu_j^{-(t)}(\tau)\right) \le e^{-\frac{k\eta^2}{24}\mu_j^{-(t)}(\tau)}$$

and,

$$\mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i) > (1+\eta)\mu_j^{+(t)}(\tau)\right) \le e^{-\frac{k\min\{\eta,\eta^2\}}{196}\mu_j^{+(t)}(\tau)}.$$

证明。对于 $\eta \ge 1$，第一个不等式显然成立，因此我们只关注 $0 < \eta < 1$ 的情况。对于每个 $\lambda \le 0$，我们有：

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i) < (1-\eta)\mu_j^{-(t)}(\tau)\right)\\
&= \mathbb{P}\left(e^{\lambda\sum_{i=1}^{k}\zeta_j^{(t)}(i)} > e^{\lambda k(1-\eta)\mu_j^{-(t)}(\tau)}\right)\\
&\le e^{-\lambda k(1-\eta)\mu_j^{-(t)}(\tau)}\mathbb{E}\left(\prod_{i=1}^{k}e^{\lambda\zeta_j^{(t)}(i)}\right)\\
&= e^{-\lambda k(1-\eta)\mu_j^{-(t)}(\tau)}\mathbb{E}\left(\mathbb{E}\left(e^{\lambda\zeta_j^{(t)}(k)}\middle| s_j^{(t)}(k)\ ,\ S\left(s_j^{(t)}(k)\right)\ ,\ \zeta_j^{(t)}(1),\ldots,\zeta_j^{(t)}(k-1)\right)\prod_{i=1}^{k-1}e^{\lambda\zeta_j^{(t)}(i)}\right)\\
&= e^{-\lambda k(1-\eta)\mu_j^{-(t)}(\tau)}\mathbb{E}\left(\mathbb{E}\left(e^{\lambda\zeta_j^{(t)}(k)}\middle| s_j^{(t)}(k)\ ,\ S\left(s_j^{(t)}(k)\right)\right)\prod_{i=1}^{k-1}e^{\lambda\zeta_j^{(t)}(i)}\right)\\
&\le e^{-\lambda k(1-\eta)\mu_j^{-(t)}(\tau)}\cdot\mathbb{E}\left(\left(\frac{1}{1-\mu_j^{-(t)}(\tau)\cdot(e^{\lambda}-1)}\right)\prod_{i=1}^{k-1}e^{\lambda\zeta_j^{(t)}(i)}\right)\\
&\cdots\\
&\le e^{-\lambda k(1-\eta)\mu_j^{-(t)}(\tau)}\cdot\left(\frac{1}{1-\mu_j^{-(t)}(\tau)\cdot(e^{\lambda}-1)}\right)^{k},
\end{aligned}$$

其中，第一个不等式由马尔可夫不等式得出；第三个等式成立，是因为给定 k第 $s_j^{(t)}(k)$时期的开始时间以及商品组合 $S\left(s_j^{(t)}(k)\right)$后，随机变量 $\zeta_j^{(t)}(k)$与所有过去轮次独立，特别是与 $\zeta_j^{(t)}(1)\ldots,$ $\zeta_j^{(t)}(k-1)$独立。取$\lambda = -\log\left(1+\frac{\eta}{(1-\eta)(1+\mu_j^{-(t)}(\tau))}\right)$，可得，

$$
\begin{aligned}
e^{-\lambda k(1-\eta)\mu_j^{-(t)}(\tau)} &= \left(1+\frac{\eta}{(1-\eta)(1+\mu_j^{-(t)}(\tau))}\right)^{k(1-\eta)\mu_j^{-(t)}(\tau)} \\
&\le \left(1+\frac{\eta}{1+\mu_j^{-(t)}(\tau)}\right)^{k\mu_j^{-(t)}(\tau)} \\
&= e^{k\mu_j^{-(t)}(\tau)\log\left(1+\frac{\eta}{1+\mu_j^{-(t)}(\tau)}\right)} \\
&\le e^{k\mu_j^{-(t)}(\tau)\left(\frac{\eta}{1+\mu_j^{-(t)}(\tau)}-\frac{1}{2}\left(\frac{\eta}{1+\mu_j^{-(t)}(\tau)}\right)^2+\frac{1}{3}\left(\frac{\eta}{1+\mu_j^{-(t)}(\tau)}\right)^3\right)} \\
&\le e^{k\mu_j^{-(t)}(\tau)\left(\frac{\eta}{1+\mu_j^{-(t)}(\tau)}-\frac{1}{6}\left(\frac{\eta}{1+\mu_j^{-(t)}(\tau)}\right)^2\right)} \\
&\le e^{\frac{k\mu_j^{-(t)}(\tau)\eta}{1+\mu_j^{-(t)}(\tau)}}\cdot e^{-\frac{k\mu_j^{-(t)}(\tau)\eta^2}{24}}.
\end{aligned}
$$

第一个不等式来自这样一个事实：对于每个 $x \ge -1$ 和 $r \in (0,1)$，$(1+x)^r \le 1+rx$。第三个不等式由不等式 $\log(1+x) \le x - \frac{x^2}{2} + \frac{x^3}{3}$ 得出，该不等式对每个 $x > -1$均成立。第四个不等式成立，因为$\frac{\eta}{1+\mu_j^{-(t)}(\tau)} \le 1$。最后一个不等式成立，因为 $\mu_j^{-(t)}(\tau) \le 1$。接下来有，

$$
\begin{aligned}
\left(\frac{1}{1-\mu_j^{-(t)}(\tau)\cdot(e^\lambda-1)}\right)^k &= \left(1-\frac{\eta\mu_j^{-(t)}(\tau)}{1+\mu_j^{-(t)}(\tau)}\right)^k \\
&= e^{k\log\left(1-\frac{\eta\mu_j^{-(t)}(\tau)}{1+\mu_j^{-(t)}(\tau)}\right)} \\
&\le e^{-\frac{k\mu_j^{-(t)}(\tau)\eta}{1+\mu_j^{-(t)}(\tau)}},
\end{aligned}
$$

其中，最后一个不等式由以下事实得出：对于每个 $x > -1$，$\log(1+x) \le x$。因此，

$$
\mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i) < (1-\eta)\mu_j^{-(t)}(\tau)\right) \le e^{-\frac{k\mu_j^{-(t)}(\tau)\eta^2}{24}}
$$

现在来看第二个不等式，对于每个 $0 \le \lambda < \log\left(1 + \frac{1}{\mu_j^{+(t)}(\tau)}\right)$，我们有：

$$
\begin{aligned}
&\mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i) > (1+\eta)\mu_j^{+(t)}(\tau)\right) \\
&= \mathbb{P}\left(e^{\lambda\sum_{i=1}^{k}\zeta_j^{(t)}(i)} > e^{\lambda k(1+\eta)\mu_j^{+(t)}(\tau)}\right) \\
&\le e^{-\lambda k(1+\eta)\mu_j^{+(t)}(\tau)}\mathbb{E}\left(\prod_{i=1}^{k} e^{\lambda\zeta_j^{(t)}(i)}\right) \\
&= e^{-\lambda k(1+\eta)\mu_j^{+(t)}(\tau)}\mathbb{E}\left(\mathbb{E}\left(e^{\lambda\zeta_j^{(t)}(k)}\middle| s_j^{(t)}(k)\ ,\ S\left(s_j^{(t)}(k)\right)\ ,\ \zeta_j^{(t)}(1),\ldots,\zeta_j^{(t)}(k-1)\right)\prod_{i=1}^{k-1} e^{\lambda\zeta_j^{(t)}(i)}\right) \\
&= e^{-\lambda k(1+\eta)\mu_j^{+(t)}(\tau)}\mathbb{E}\left(\mathbb{E}\left(e^{\lambda\zeta_j^{(t)}(k)}\middle| s_j^{(t)}(k)\ ,\ S\left(s_j^{(t)}(k)\right)\right)\prod_{i=1}^{k-1} e^{\lambda\zeta_j^{(t)}(i)}\right) \\
&\le e^{-\lambda k(1+\eta)\mu_j^{+(t)}(\tau)}\cdot\mathbb{E}\left(\left(\frac{1}{1-\mu_j^{+(t)}(\tau)\cdot(e^{\lambda}-1)}\right)\prod_{i=1}^{k-1} e^{\lambda\zeta_j^{(t)}(i)}\right) \\
&\cdots \\
&\le e^{-\lambda k(1+\eta)\mu_j^{+(t)}(\tau)}\cdot\left(\frac{1}{1-\mu_j^{+(t)}(\tau)\cdot(e^{\lambda}-1)}\right)^{k}
\end{aligned}
$$

对于 $\lambda = \log\left(1 + \frac{\eta}{1+\mu_j^{+(t)}(\tau)(1+\eta)}\right)$，我们有，

$$
\begin{aligned}
e^{-\lambda k(1+\eta)\mu_j^{+(t)}(\tau)} &= e^{k(1+\eta)\mu_j^{+(t)}(\tau)\log\left(1-\frac{\eta}{(1+\eta)(1+\mu_j^{+(t)}(\tau))}\right)} \\
&\le e^{k(1+\eta)\mu_j^{+(t)}(\tau)\left(-\frac{\eta}{(1+\eta)(1+\mu_j^{+(t)}(\tau))}-\frac{1}{2}\left(\frac{\eta}{(1+\eta)(1+\mu_j^{+(t)}(\tau))}\right)^2\right)} \\
&\le e^{-\frac{k\mu_j^{+(t)}(\tau)\eta}{1+\mu_j^{+(t)}(\tau)}}\cdot e^{-\frac{k\mu_j^{+(t)}(\tau)\eta^2}{98(1+\eta)}},
\end{aligned}
$$

其中，第一个不等式是由 $\log(1-x) \le -x-\frac{x^2}{2}$ 对每个 $x \in (0, 1)$ 都成立这一事实得出的。对于最后一个不等式，回顾 $\mu_j^{+(t)}(\tau) = \left(\omega_j(\tau) + \delta_j^{(t)}\right)\frac{1+\delta^{(t)}}{1-\delta^{(t)}}$，注意到 $\delta^{(t)} \le \frac{1}{2}$，这意味着 $\frac{1+\delta^{(t)}}{1-\delta^{(t)}} \le 3$；又因为 $\delta_j^{(t)} \le \delta^{(t)} \le \frac{1}{2}$，且 $\omega_j(\tau) \le 1$，因此 $\mu_j^{+(t)}(\tau) \le 6$。接下来，我们有，

$$\left(\frac{1}{1-\mu_j^{+(t)}(\tau)\cdot(e^\lambda-1)}\right)^k = \left(1+\frac{\eta\mu_j^{+(t)}(\tau)}{1+\mu_j^{+(t)}(\tau)}\right)^k$$
$$= e^{k\log\left(1+\frac{\eta\mu_j^{+(t)}(\tau)}{1+\mu_j^{+(t)}(\tau)}\right)}$$
$$\le e^{\frac{k\mu_j^{+(t)}(\tau)\eta}{1+\mu_j^{+(t)}(\tau)}}$$

根据 $\log(1+x) \le x$ 对每个 $x > -1$ 都成立这一事实。因此,

$$\mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i) > (1+\eta)\mu_j^{+(t)}(\tau)\right) \le e^{-\frac{k\mu_j^{+(t)}(\tau)\eta^2}{98(1+\eta)}}$$

为完成证明，我们证明对于每个 $\eta > 0$，都有 $\min\{\eta,\eta^2\} \le \frac{2\eta^2}{1+\eta}$。事实上，若 $0 < \eta < 1$，则 $\frac{2\eta^2}{1+\eta} \ge \frac{2\eta^2}{2} = \eta^2 = \min\{\eta,\eta^2\}$。若 $\eta \ge 1$，则 $\frac{2\eta^2}{1+\eta} = 2\eta\frac{\eta}{1+\eta} \ge 2\eta\cdot\frac{1}{2} = \eta = \min\{\eta,\eta^2\}$。 □

引理 4. 设 $t \in [1,T]$ 满足 $\delta^{(t)} \le \frac{1}{2}$，并设 $j \in [N]$。设 $\tau \in [1,t]$。以下集中界成立,

$$\mathbb{P}\left(\hat{\omega}_j(t) < \mu_j^{-(t)}(\tau)\right) \le \frac{2}{NT^3}$$

and

$$\mathbb{P}\left(\hat{\omega}_j(t) > \mu_j^{+(t)}(\tau) + \sqrt{\frac{2266\mu_j^{+(t)}(\tau)\log NT}{\tilde{n}_j(t)}} + \frac{1364\log NT}{\tilde{n}_j(t)}\right) \le \frac{2}{NT^3}.$$

证明。令 $\bar{\zeta}_j(t) := \frac{1}{\tilde{n}_j(t)}\sum_{i=1}^{n_j(t)}\zeta_j^{(t)}(i)$。则 $\bar{\zeta}_j(t) = \bar{\omega}_j(t)$，这是因为前 $n_j(t)$ 轮次都在 $t-1$ 之前结束，因此对每个 $i \le n_j(t)$，都有 $\zeta_j^{(t)}(i) = \tilde{\omega}_j^{(t)}(i)$。接下来，我们有,

$$
\begin{aligned}
&\mathbb{P}\left(\hat{\omega}_j(t) < \mu_j^{-(t)}(\tau)\right) \\
&= \mathbb{P}\left(\bar{\omega}_j(t) + \sqrt{\frac{192\bar{\omega}_j(t)\log NT}{\tilde{n}_j(t)}} + \frac{192\log NT}{\tilde{n}_j(t)} < \mu_j^{-(t)}(\tau)\right) \\
&= \mathbb{P}\left(\bar{\zeta}_j(t) + \sqrt{\frac{192\bar{\zeta}_j(t)\log NT}{\tilde{n}_j(t)}} + \frac{192\log NT}{\tilde{n}_j(t)} < \mu_j^{-(t)}(\tau)\right) \\
&= \sum_{k=0}^{t}\mathbb{P}\left(\left\{\bar{\zeta}_j(t) + \sqrt{\frac{192\bar{\zeta}_j(t)\log NT}{\tilde{n}_j(t)}} + \frac{192\log NT}{\tilde{n}_j(t)} < \mu_j^{-(t)}(\tau)\right\} \cap \{n_j(t)=k\}\right)
\end{aligned}
$$

$$
\begin{aligned}
&\le \mathbb{P}\left(192\log NT < \mu_j^{-(t)}(\tau)\right) \\
&\quad + \sum_{k=1}^{t}\mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i) + \sqrt{\frac{192\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i)\right)\log NT}{k}} + \frac{192\log NT}{k} < \mu_j^{-(t)}(\tau)\right)
\end{aligned}
$$

$$
= \sum_{k=1}^{t}\mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i) + \sqrt{\frac{192\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i)\right)\log NT}{k}} + \frac{192\log NT}{k} < \mu_j^{-(t)}(\tau)\right).
$$

在该不等式中，我们分别讨论 $k=0$ 和 $k\ge 1$ 两种情况；随后在下一个等式中利用以下事实：$\mathbb{P}\left(192\log NT < \mu_j^{-(t)}(\tau)\right) = 0$，因为 $\mu_j^{-(t)}(\tau) \le 1$。现在，令 $k\in[1,t]$，并对该项进行界定

$$
\mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i) + \sqrt{\frac{192\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i)\right)\log NT}{k}} + \frac{192\log NT}{k} < \mu_j^{-(t)}(\tau)\right).
$$

首先，在引理 $\eta = \sqrt{96}$ 的第一个集中不等式中取对数 $k\mu_j^{-(t)}\tau$ ( )，可得，3

$$
\mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i) + \sqrt{\frac{96\mu_j^{-(t)}(\tau)\log NT}{k}} < \mu_j^{-(t)}(\tau)\right) \le \frac{1}{(NT)^4}
$$

接下来，在相同的集中不等式中取 $\eta = \frac{1}{2} + 192$ 对数 $k\mu_j^{-(t)}\tau$ () 可得，

$$\mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i) < \frac{1}{2}\mu_j^{-(t)}(\tau) - \frac{192\log NT}{k}\right) \le e^{-\frac{k\mu_j^{-(t)}(\tau)}{24}\left(\frac{1}{2}+\frac{192\log NT}{k\mu_j^{-(t)}(\tau)}\right)^2}$$
$$\le e^{-\frac{k\mu_j^{-(t)}(\tau)}{24}\left(\frac{1}{4}+\frac{192^2\log^2 NT}{(k\mu_j^{-(t)}(\tau))^2}\right)}$$
$$= e^{-\frac{1}{96}\left(k\mu_j^{-(t)}(\tau)+\frac{4\times 192^2\log^2 NT}{k\mu_j^{-(t)}(\tau)}\right)}$$
$$\le e^{-\frac{2\times 192\log NT}{96}} = \frac{1}{(NT)^4}$$

最后一个不等式源于以下事实：对于每个 $x, a > 0$，$(x + \frac{a}{x}) \ge \sqrt{a}$。最后，对于每个 $k \ge 1$,

$$\mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i) + \sqrt{\frac{192\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i)\right)\log NT}{k}} + \frac{192\log NT}{k} < \mu_j^{-(t)}(\tau)\right)$$
$$\le \mathbb{P}\left(\left\{\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i) + \sqrt{\frac{192\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i)\right)\log NT}{k}} + \frac{192\log NT}{k} < \mu_j^{-(t)}(\tau)\right\}\right.$$
$$\left.\cap\left\{\sum_{i=1}^{k}\zeta_j^{(t)}(i) \ge \left(\frac{1}{2}\mu_j^{-(t)}(\tau) - \frac{192\log NT}{k}\right)^+\right\}\right) + \frac{1}{(NT)^4}$$

$$\le \mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i) + \sqrt{\frac{192\left(\frac{1}{2}\mu_j^{-(t)}(\tau) - \frac{192\log NT}{k}\right)^+\log NT}{k}} + \frac{192\log NT}{k} < \mu_j^{-(t)}(\tau)\right) + \frac{1}{(NT)^4}$$
$$= \mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i) + \sqrt{\left(\frac{96\mu_j^{-(t)}(\tau)\log NT}{k} - \frac{192^2\log^2(NT)}{k^2}\right)^+} + \frac{192\log NT}{k} < \mu_j^{-(t)}(\tau)\right) + \frac{1}{(NT)^4}$$
$$\le \mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i) + \sqrt{\left(\frac{96\mu_j^{-(t)}(\tau)\log NT}{k} - \frac{192^2\log^2(NT)}{k^2}\right)^+ + \frac{192^2\log^2 NT}{k^2}} < \mu_j^{-(t)}(\tau)\right) + \frac{1}{(NT)^4}$$
$$\le \mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i) + \sqrt{\frac{96\mu_j^{-(t)}(\tau)\log NT}{k}} < \mu_j^{-(t)}(\tau)\right) + \frac{1}{(NT)^4}$$
$$\le \frac{2}{(NT)^4},$$

其中，第三个不等式是由 $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ 对每个 a， $b \geq 0$ 得出的。因此，

$$\mathbb{P}\left(\hat{\omega}_j(t) < \mu_j^{-(t)}(\tau)\right) \leq \frac{2}{NT^3}$$

现在，对于第二个集中界，我们有：

$$\begin{aligned}
&\mathbb{P}\left(\hat{\omega}_j(t) > \mu_j^{+(t)}(\tau) + \sqrt{\frac{2266\mu_j^{+(t)}(\tau)\log NT}{\tilde{n}_j(t)}} + \frac{1364\log NT}{\tilde{n}_j(t)}\right) \\
&= \mathbb{P}\left(\bar{\omega}_j(t) + \sqrt{\frac{192\bar{\omega}_j(t)\log NT}{\tilde{n}_j(t)}} + \frac{192\log NT}{\tilde{n}_j(t)} > \mu_j^{+(t)}(\tau) + \sqrt{\frac{2266\mu_j^{+(t)}(\tau)\log NT}{\tilde{n}_j(t)}} + \frac{1364\log NT}{\tilde{n}_j(t)}\right) \\
&= \mathbb{P}\left(\bar{\zeta}_j(t) + \sqrt{\frac{192\bar{\zeta}_j(t)\log NT}{\tilde{n}_j(t)}} > \mu_j^{+(t)}(\tau) + \sqrt{\frac{2266\mu_j^{+(t)}(\tau)\log NT}{\tilde{n}_j(t)}} + \frac{(1364-192)\log NT}{\tilde{n}_j(t)}\right) \\
&= \sum_{k=0}^{t} \mathbb{P}\left(\left\{\bar{\zeta}_j(t) + \sqrt{\frac{192\bar{\zeta}_j(t)\log NT}{\tilde{n}_j(t)}} > \mu_j^{+(t)}(\tau) + \sqrt{\frac{2266\mu_j^{+(t)}(\tau)\log NT}{\tilde{n}_j(t)}} + \frac{(1364-192)\log NT}{\tilde{n}_j(t)}\right\} \cap \{\tilde{n}_j(t) = k\}\right) \\
&\leq \mathbb{P}\left(\mu_j^{+(t)}(\tau) + \sqrt{2266\mu_j^{+(t)}(\tau)\log NT} + (1364-192)\log NT < 0\right) \\
&\quad + \sum_{k=1}^{t} \mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i) + \sqrt{\frac{192\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i)\right)\log NT}{k}} > \mu_j^{+(t)}(\tau) + \sqrt{\frac{2266\mu_j^{+(t)}(\tau)\log NT}{k}} + \frac{(1364-192)\log NT}{k}\right) \\
&= \sum_{k=1}^{t} \mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i) + \sqrt{\frac{192\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i)\right)\log NT}{k}} > \mu_j^{+(t)}(\tau) + \sqrt{\frac{2266\mu_j^{+(t)}(\tau)\log NT}{k}} + \frac{(1364-192)\log NT}{k}\right).
\end{aligned}$$

令 $k \in [1,t]$，并对这一项进行界定

$$\mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i) + \sqrt{\frac{192\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i)\right)\log NT}{k}} > \mu_j^{+(t)}(\tau) + \sqrt{\frac{2266\mu_j^{+(t)}(\tau)\log NT}{k}} + \frac{(1364-192)\log NT}{k}\right).$$

令 $\eta = \sqrt{784\frac{\log NT}{\log k\mu_j^{+(t)}\ \tau\ ()}} + 784\frac{\log NT}{k\mu_j^{+(t)}\ \tau\ ()}$。则

$$\eta \cdot k\mu_j^{+(t)}(\tau) \geq 784\log NT$$

and

$$\eta^2 \cdot k\mu_j^{+(t)}(\tau) \geq 784\log NT.$$

因此，将 η 代入引理 3中的第二个集中不等式可得，

$$\mathbb{P}\left(\sum_{i=1}^{k}\zeta_j^{(t)}(i) > \sqrt{\frac{784\mu_j^{+(t)}(\tau)\log NT}{k}} + \frac{784\log NT}{k} + \mu_j^{+(t)}(\tau)\right) \le \frac{1}{(NT)^4}.$$

接下来，将 $\eta = 1 + \frac{784\,\text{对数}\overline{k\mu_j^{+(t)}\,\tau}}{NT}$（）代入同一个集中不等式，可得：

$$\begin{aligned}\mathbb{P}\left(\sum_{i=1}^{k}\zeta_j^{(t)}(i) > 2\mu_j^{+(t)}(\tau) + \frac{784\log NT}{k}\right) &\le e^{-\frac{k\mu_j^{+(t)}(\tau)}{196}\left(1+\frac{784\log NT}{k\mu_j^{+(t)}(\tau)}\right)} \\ &\le e^{-\frac{784\log NT}{196}} = \frac{1}{(NT)^4}\end{aligned}$$

最后，对于每个 $k \ge 1$，我们有：

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i) + \sqrt{\frac{192\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i)\right)\log NT}{k}} > \mu_j^{+(t)}(\tau) + \sqrt{\frac{2266\mu_j^{+(t)}(\tau)\log NT}{k}} + \frac{(1364-192)\log NT}{k}\right) \\
&\le \mathbb{P}\left(\left\{\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i) + \sqrt{\frac{192\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i)\right)\log NT}{k}} > \mu_j^{+(t)}(\tau) + \sqrt{\frac{2266\mu_j^{+(t)}(\tau)\log NT}{k}} + \frac{(1364-192)\log NT}{k}\right\}\right. \\
&\qquad \left.\cap\left\{\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i) \le 2\mu_j^{+(t)}(\tau) + \frac{784\log NT}{k}\right\}\right) + \frac{1}{(NT)^4} \\
&\le \mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i) + \sqrt{\frac{192\left(2\mu_j^{+(t)}(\tau) + \frac{784\log NT}{k}\right)\log NT}{k}} > \mu_j^{+(t)}(\tau) + \sqrt{\frac{2266\mu_j^{+(t)}(\tau)\log NT}{k}} + \frac{(1364-192)\log NT}{k}\right) + \frac{1}{(NT)^4} \\
&= \mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i) + \sqrt{\frac{384\mu_j^{+(t)}(\tau)\log NT}{k} + \frac{150528\log^2 NT}{k^2}} > \mu_j^{+(t)}(\tau) + \sqrt{\frac{2266\mu_j^{+(t)}(\tau)\log NT}{k}} + \frac{(1364-192)\log NT}{k}\right) + \frac{1}{(NT)^4} \\
&\le \mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i) + \sqrt{\frac{384\mu_j^{+(t)}(\tau)\log NT}{k}} + \sqrt{\frac{150528\log^2 NT}{k^2}} > \mu_j^{+(t)}(\tau) + \sqrt{\frac{2266\mu_j^{+(t)}(\tau)\log NT}{k}} + \frac{(1364-192)\log NT}{k}\right) + \frac{1}{(NT)^4} \\
&= \mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i) > \frac{1364-192-\sqrt{150528}}{784}\frac{784\log NT}{k} + \frac{\sqrt{2266}-\sqrt{384}}{\sqrt{784}}\sqrt{\frac{784\mu_j^{+(t)}(\tau)\log NT}{k}} + \mu_j^{+(t)}(\tau)\right) + \frac{1}{(NT)^4} \\
&\le \mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k}\zeta_j^{(t)}(i) > \frac{784\log NT}{k} + \sqrt{\frac{784\mu_j^{+(t)}(\tau)\log NT}{k}} + \mu_j^{+(t)}(\tau)\right) + \frac{1}{(NT)^4} \\
&\le \frac{2}{(NT)^4},
\end{aligned}$$

其中，第三个不等式由对每个 $a, b \ge 0$ 都成立的 $\sqrt{a+b} \le \sqrt{a} + \sqrt{b}$ 得出，而第四个不等式

成立是因为 $\frac{1364-192-\sqrt{150528}}{784} \geq 1$ 和 $\frac{\sqrt{2266}-\sqrt{384}}{\sqrt{784}} \geq 1$。因此，

$$\mathbb{P}\left(\hat{\omega}_j(t) > \mu_j^{+(t)}(\tau) + \sqrt{\frac{2266\mu_j^{+(t)}(\tau)\log NT}{\tilde{n}_j(t)}} + \frac{1364\log NT}{\tilde{n}_j(t)}\right) \leq \frac{2}{NT^3}$$

□

引理 5。令 $t \in [1,T]$，使得 $\delta^{(t)} \leq \frac{1}{2}$，并令 $j \in [N]$。令 $u \in [1,t]$。我们有，

$$\mathbb{P}\left(\bigcup_{1\leq i\leq e_j^{(t)}}\left\{\left|l_j^{(t)}(i)\right| > 5\log NT\left(1+\sum_{k\in S\left(s_j^{(t)}(i)\right)}\omega_k(u)+\delta_k^{(t)}\right)\right\}\right) \leq \frac{1}{NT^3}$$

证明。对于每个 $i \in [1, e_j^{(t)}]$，时期 $l_j^{(t)}(i)$ 的长度满足

$$\left|l_j^{(t)}(i)\right| \leq s_j^{(t)}(i+1) - s_j^{(t)}(i).$$

注意，如果最后一个时期 $l_j^{(t)}\left(e_j^{(t)}\right)$ 不完整，不等式可能是严格的。因此，

$$\mathbb{P}\left(\bigcup_{1\leq i\leq e_j^{(t)}}\left\{\left|l_j^{(t)}(i)\right| > 5\log NT\left(1+\sum_{k\in S\left(s_j^{(t)}(i)\right)}\omega_k(u)+\delta_k^{(t)}\right)\right\}\right)$$

$$\leq \mathbb{P}\left(\bigcup_{1\leq i\leq e_j^{(t)}}\left\{\left(s_j^{(t)}(i+1)-s_j^{(t)}(i)\right) > 5\log NT\left(1+\sum_{k\in S\left(s_j^{(t)}(i)\right)}\omega_k(u)+\delta_k^{(t)}\right)\right\}\right)$$

$$\leq \mathbb{P}\left(\bigcup_{1\leq i\leq T}\left\{\left(s_j^{(t)}(i+1)-s_j^{(t)}(i)\right) > 5\log NT\left(1+\sum_{k\in S\left(s_j^{(t)}(i)\right)}\omega_k(u)+\delta_k^{(t)}\right)\right\}\right)$$

$$\leq \sum_{i=1}^{T}\mathbb{P}\left(\left(s_j^{(t)}(i+1)-s_j^{(t)}(i)\right) > 5\log NT\left(1+\sum_{k\in S\left(s_j^{(t)}(i)\right)}\omega_k(u)+\delta_k^{(t)}\right)\right)$$

令 $\alpha = \left(1+\sum_{k\in S\left(s_j^{(t)}(i)\right)}\omega_k(u)+\delta_k^{(t)}\right)$. 为得出结论，我们对上述表达式中的各项进行界定。

具体而言，对于每个 $i\in[1,T]$ 和 $j\in[N]$，都有：

$$
\begin{aligned}
&\mathbb{P}\left(s_j^{(t)}(i+1)-s_j^{(t)}(i)>5\log NT\cdot\alpha\right)\\
&=\mathbb{E}\left(\mathbb{P}\left[s_j^{(t)}(i+1)-s_j^{(t)}(i)>5\log NT\cdot\alpha\,\middle|\, s_j^{(t)}(i),S\left(s_j^{(t)}(i)\right)\right]\right)\\
&\le\mathbb{E}\left(\mathbb{P}\left[j\left(s_j^{(t)}(i)\right)\neq 0\,,\,\ldots\,,\,j\left(s_j^{(t)}(i)+\lfloor 5\log NT\cdot\alpha\rfloor-1\right)\neq 0\,\middle|\, s_j^{(t)}(i),S\left(s_j^{(t)}(i)\right)\right]\right)\\
&=\mathbb{E}\left(\prod_{s=s_j^{(t)}(i)}^{s_j^{(t)}(i)+\lfloor 5\log NT\cdot\alpha\rfloor-1}\left(1-\frac{1}{1+\sum_{k\in S\left(s_j^{(t)}(i)\right)}\omega_k(\min\{s,t\})}\right)\right)\\
&\le\mathbb{E}\left(\prod_{s=s_j^{(t)}(i)}^{s_j^{(t)}(i)+\lfloor 5\log NT\cdot\alpha\rfloor-1}\left(1-\frac{1}{1+\sum_{k\in S\left(s_j^{(t)}(i)\right)}\omega_k(u)+\delta_k^{(t)}}\right)\right)\\
&=\mathbb{E}\left(\left(1-\frac{1}{1+\sum_{k\in S\left(s_j^{(t)}(i)\right)}\omega_k(u)+\delta_k^{(t)}}\right)^{\lfloor 5\log NT\cdot\alpha\rfloor}\right)\\
&\le\mathbb{E}\left(\left(1-\frac{1}{1+\sum_{k\in S\left(s_j^{(t)}(i)\right)}\omega_k(u)+\delta_k^{(t)}}\right)^{4\log NT\left(1+\sum_{k\in S\left(s_j^{(t)}(i)\right)}\omega_k(u)+\delta_k^{(t)}\right)}\right)\\
&=\mathbb{E}\left(\exp\left[4\log NT\left(1+\sum_{k\in S\left(s_j^{(t)}(i)\right)}\omega_k(u)+\delta_k^{(t)}\right)\log\left(1-\frac{1}{1+\sum_{k\in S\left(s_j^{(t)}(i)\right)}\omega_k(u)+\delta_k^{(t)}}\right)\right]\right)\\
&\le\mathbb{E}\left(\exp\left[-4\log NT\left(1+\sum_{k\in S\left(s_j^{(t)}(i)\right)}\omega_k(u)+\delta_k^{(t)}\right)\left(\frac{1}{1+\sum_{k\in S\left(s_j^{(t)}(i)\right)}\omega_k(u)+\delta_k^{(t)}}\right)\right]\right)\\
&\le\frac{1}{NT^4}.
\end{aligned}
$$

第二个等式由条件概率链式法则得到。第二个不等式成立，因为对于每个 s 和 u，$\omega_k(\min\{s,t\})\le\omega_k(u)+\delta_k^{(t)}$。第三个不等式成立

对于 $NT \geq e$（因此 $\alpha \cdot \log\ NT \geq 1$）。最后，第四个不等式由以下事实得出：对于每个 $x > -1$，$\log(1-x) \leq -x$。证明可由对 T 个加数应用并集界得到。

□

定理 1. 对于这些选择，

$$\Delta(t) = 26\|\boldsymbol{\omega}(t) - \boldsymbol{\omega}(t+1)\|_{L_\infty^{2K}}, \quad \forall t \in [1, T-1],$$

and

$$\rho(t) = (149 \log NT)^{\frac{3}{2}} \sqrt{\frac{N}{t}} + (55 \log NT)^3 \frac{N}{t} + \sqrt{\frac{2 \log T}{t}}, \quad \forall t \in [1, T],$$

算法 *1* 在 (Δ, ρ)-近似平稳环境中表现良好，其 $\hat{}\ R_t = R(S(t), \hat{}\boldsymbol{\omega}(t))$，其中 $\hat{}\boldsymbol{\omega}(t)$ 是第 t 轮次的上置信界向量。根据 *Wei* 和 *Luo* 的定理 *2*，*2021*，这进一步意味着，即使不知道 Δ_∞^K 和 L，*EEMNSE* 仍能以高概率保证动态遗憾满足，

$$\text{D-Reg} = \tilde{O}\left(\min\left\{\sqrt{NTL}\,,\ N^{\frac{1}{3}}(\Delta_\infty^K)^{\frac{1}{3}}T^{\frac{2}{3}} + \sqrt{NT}\right\}\right).$$

证明. 现在我们给出定理 1 的证明。首先证明 Δ 确实是一个非平稳性度量（见定义 1）。令 $t \in [1, T-1]$ 并考虑某个

$$S^* \in \underset{S \subset [N]: |S| \leq K}{\operatorname{argmax}} |R(S, \boldsymbol{\omega}(t)) - R(S, \boldsymbol{\omega}(t+1))|.$$

我们有，

$$\max_{S\subset[N]:|S|\le K} |R(S,\boldsymbol{\omega}(t)) - R(S,\boldsymbol{\omega}(t+1))|$$

$$= |R(S^*,\boldsymbol{\omega}(t)) - R(S^*,\boldsymbol{\omega}(t+1))|$$

$$= \left| \frac{\sum_{j\in S^*} r_j\omega_j(t)}{1+\sum_{k\in S^*}\omega_k(t)} - \frac{\sum_{j\in S^*} r_j\omega_j(t+1)}{1+\sum_{k\in S^*}\omega_k(t+1)} \right|$$

$$= \left| \frac{\sum_{j\in S^*} r_j(\omega_j(t)-\omega_j(t+1)) + \sum_{j,k\in S^*} r_j(\omega_j(t)\omega_k(t+1) - \omega_k(t)\omega_j(t+1))}{(1+\sum_{k\in S^*}\omega_k(t))(1+\sum_{k\in S^*}\omega_k(t+1))} \right|$$

$$= \left| \frac{\sum_{j\in S^*} r_j(\omega_j(t)-\omega_j(t+1)) + \sum_{j,k\in S^*} r_j(\omega_j(t)\omega_k(t+1) - \omega_j(t)\omega_k(t) + \omega_j(t)\omega_k(t) - \omega_k(t)\omega_j(t+1))}{(1+\sum_{k\in S^*}\omega_k(t))(1+\sum_{k\in S^*}\omega_k(t+1))} \right|$$

$$\le \sum_{j\in S^*} |\omega_j(t)-\omega_j(t+1)| + \frac{\sum_{j\in S^*} r_j\omega_j(t)}{1+\sum_{k\in S^*}\omega_k(t)} \sum_{k\in S^*} |\omega_k(t+1)-\omega_k(t)|$$
$$+ \frac{\sum_{k\in S^*}\omega_k(t)}{1+\sum_{k\in S^*}\omega_k(t)} \sum_{j\in S^*} |\omega_j(t+1)-\omega_j(t)|$$

$$\le 3 \sup_{S\subset[N]:|S|\le 2K} \sum_{r\in S} |\omega_r(t)-\omega_r(t+1)| \le \Delta(t)$$

现在固定一个 $t\in[1,T]$，使得 $\sum_{\tau=1}^{t-1}\Delta(\tau) \le \rho(t)$ 且 $\delta^{(t)} \le \frac{1}{2}$，并证明对于每个 $\tau\in[1,T]$，条件 (1a) 和 (1b) 以至少 $1-O(\frac{1}{T})$ 的概率成立，其中 $\hat{} R_\tau = R(S(\tau), \hat{}\boldsymbol{\omega}(\tau))$。不过，在讨论条件 (1a) 和 (1b) 之前，我们先回顾 Agrawal 等人，2016 提出的以下引理，该引理刻画了在最优解处评估时，期望收益函数相对于吸引参数的变化。

引理 6。 *[Agrawal 等人的引理 A.3, 2016]* 考虑吸引力参数 $\boldsymbol{\omega}$，并令 S^* 为关于 $\boldsymbol{\omega}$ 的最优产品组合。则对于每个 $\boldsymbol{\omega}' \ge \boldsymbol{\omega}$ （逐元素），都有$R(S^*, \boldsymbol{\omega}') \ge R(S^*, \boldsymbol{\omega})$。

现在证明这些条件。首先注意到，$(M_t = \sum_{\tau=1}^{t} R(S(\tau)\ \boldsymbol{\omega}(\tau)) - r(\tau))_{1\le t\le T}$是关于该滤过的鞅

$$\mathcal{F}_0 = \{\Omega,\emptyset\} \subset \mathcal{F}_1 = \sigma(M_1, S(1)) \subset \cdots \subset \mathcal{F}_T = \sigma(M_T, S(T)),$$

其中， $\sigma(M_t, S(t))$表示由随机变量 M_t 和 $S(t)$生成的 σ-代数，而 Ω 表示概率空间。由于对于所有 τ， $R(S(\tau)\ \boldsymbol{\omega}(\tau)) = \mathbb{E}(r(\tau)|S(\tau))$，因此 $(M_t)^{t\in[1,T]}$ 的共同均值为 0。于是，根据 Azuma 不等式，有

$$\mathbb{P}\left(\sum_{\tau=1}^{t} R(S(\tau), \boldsymbol{\omega}(\tau)) - r(\tau) > \epsilon\right) \leq e^{-\frac{\epsilon^2}{2t}},$$

这意味着，以至少 $1 - \frac{1}{T}$ 的概率，有 $\sum_{\tau=1}^{t} R(S(\tau)\ \boldsymbol{\omega}(\tau)) - r(\tau) \leq \sqrt{2t \log(T)}$ 成立。接下来，令

$$\omega_j^{\mathsf{up}}(t) := \frac{1+\delta^{(t)}}{1-\delta^{(t)}} \cdot \hat{\omega}_j(t) + \delta_j^{(t)}.$$

上述事实结合引理 4 和 5 可得，以至少 $1 - \frac{6}{T}$ 的概率，有下式成立：

$$\sum_{\tau=1}^{t} R(S(\tau), \boldsymbol{\omega}(\tau)) - r(\tau) \leq \sqrt{2t \log(T)},$$

$$\left\{\omega_j(\tau) \leq \omega_j^{\mathsf{up}}(t) \quad \forall \tau \leq t \in [1,T], j \in [N]\right\},$$

$$\left\{\hat{\omega}_j(t) \leq \mu_j^{+(t)}(\tau) + \sqrt{\frac{2266\mu_j^{+(t)}(\tau)\log NT}{\tilde{n}_j(t)}} + \frac{1364 \log NT}{\tilde{n}_j(t)} \quad \forall \tau \leq t \in [1,T], j \in [N]\right\},$$

and

$$\left\{\left|l_j^{(t)}(i)\right| \leq 5 \log NT \left(1 + \sum_{k \in S\left(s_j^{(t)}(i)\right)} \omega_k(u) + \delta_k^{(t)}\right) \quad \forall i \in [1, e_j^{(t)}], u \leq t \in [1,T], j \in [N]\right\}.$$

因此，以至少 $1 - \frac{6}{T}$ 的概率，我们有：

条件 (1a)。 令 $\tau \in$ 最小值位置 $R(S^*(u)\ \boldsymbol{\omega}(u))$。我们有，$u\in[1,t]$

$$
\begin{aligned}
\hat{R}_t - \min_{u\in[1,t]} R(S^*(u),\boldsymbol{\omega}(u)) &= R(S(t),\hat{\boldsymbol{\omega}}(t)) - R(S^*(\tau),\boldsymbol{\omega}(\tau)) \\
&\geq R(S^*(\tau),\hat{\boldsymbol{\omega}}(t)) - R(S^*(\tau),\boldsymbol{\omega}(\tau)) \\
&\geq R(S^*(\tau),\hat{\boldsymbol{\omega}}(t)) - R(S^*(\tau),\boldsymbol{\omega}^{\mathsf{up}}(t)) \\
&= \frac{\sum_{k\in S^*(\tau)} r_k\hat{\omega}_k(t)}{1+\sum_{k\in S^*(\tau)}\hat{\omega}_k(t)} - \frac{\sum_{k\in S^*(\tau)} r_k\omega_k^{\mathsf{up}}(t)}{1+\sum_{k\in S^*(\tau)}\omega_k^{\mathsf{up}}(t)} \\
&\geq \frac{\sum_{k\in S^*(\tau)} r_k(\hat{\omega}_k(t)-\omega_k^{\mathsf{up}}(t))}{1+\sum_{k\in S^*(\tau)}\hat{\omega}_k(t)} \\
&= \frac{\frac{-2\delta^{(t)}}{1-\delta^{(t)}}\sum_{k\in S^*(\tau)} r_k\hat{\omega}_k(t) - \sum_{k\in S^*(\tau)} r_k\delta_k^{(t)}}{1+\sum_{k\in S^*(\tau)}\hat{\omega}_k(t)} \\
&\geq -2\frac{\delta^{(t)}}{1-\delta^{(t)}} - \delta^{(t)} \\
&\geq -5\delta^{(t)} \\
&\geq -\sum_{u=1}^{t-1}\Delta(u).
\end{aligned}
$$

第一个不等式由以下事实得出： $S(t)$ 是置信上界 $\hat{\boldsymbol{\omega}}(t)$ 的最优解。第二个不等式由以下事实得出：根据引理6，在 $S^*(t)$ 下是最优商品组合 $\boldsymbol{\omega}(t)$。第三个不等式成立，因为 $\hat{\boldsymbol{\omega}}(t)$ 按坐标成立。第四个不等式成立，因为 $\delta_k^{(t)} \leq \delta^{(t)}$ 如 $|S^*(\tau)| \leq K$ 且对所有 k 都有 $r_k \leq 1$。下一个不等式则由以下事实得出：$\delta^{(t)} \leq \frac{1}{2}$。

条件(1b)。我们有，

$$
\frac{1}{t}\sum_{\tau=1}^{t}\hat{R}_\tau - r(\tau) = \frac{1}{t}\left(\sum_{\tau=1}^{t}\hat{R}_\tau - R(S(\tau),\boldsymbol{\omega}(\tau))\right) + \frac{1}{t}\left(\sum_{\tau=1}^{t} R(S(\tau),\boldsymbol{\omega}(\tau)) - r(\tau)\right).
$$

第二项的上界为 $\sqrt{2\log_t T}$。对于第一项，~~对每~~个 $\tau \in [1,t]$，我们都有：

$$\hat{R}_\tau - R(S(\tau)\ \boldsymbol{\omega}(\tau))$$

$$= R(S(\tau),\ \hat{}\boldsymbol{\omega}(\tau))\ - R(S(\tau)\ \boldsymbol{\omega}(\tau))$$

$$\leq R(S(\tau), \boldsymbol{\omega}^{\mathsf{up}}(\tau)) - R(S(\tau), \boldsymbol{\omega}(\tau))$$

$$= \frac{\sum_{k\in S\ (\tau)}\ r_k\omega_k^{\mathsf{up}}(\tau)}{1+\sum_{k\in S\ (\tau)}\ \omega_k^{\mathsf{up}}(\tau)} - \frac{\sum_{k\in S\ (\tau)}\ r_k\omega_k}{1+\sum_{k\in S(\tau)}\ \omega_k(\tau)}$$

$$\leq \frac{\sum_{k\in S(\tau)}(\omega_k^{\mathsf{up}}(\tau) - \omega_k(\tau))}{1+\sum_{k\in S(\tau)}\omega_k(\tau)}$$

$$= \frac{\sum_{k\in S(\tau)}\ \left(\frac{1+\delta^{(\tau)}}{1-\delta^{(\tau)}}\hat{}\omega_k(\tau)\ - \omega_k(\tau)\right)}{1+\sum_{k\in S(\tau)}\ \omega_k(\tau)} + \frac{\sum_{k\in S(\tau)}\ \delta_k^{(\tau)}}{1+\sum_{k\in S(\tau)}\ \omega_k(\tau)}$$

$$\leq \frac{\sum_{k\in S(\tau)}\left(\frac{1+\delta^{(\tau)}}{1-\delta^{(\tau)}}\hat{\omega}_k(\tau) - \omega_k(\tau)\right)}{1+\sum_{k\in S(\tau)}\omega_k(\tau)} + \delta^{(\tau)}$$

$$\leq \frac{\sum_{k\in S(\tau)}\left(\frac{1+\delta^{(\tau)}}{1-\delta^{(\tau)}}\mu_k^{+(\tau)}(\tau) - \omega_k(\tau)\right)}{1+\sum_{k\in S(\tau)}\omega_k(\tau)} + \frac{1+\delta^{(\tau)}}{1-\delta^{(\tau)}}\cdot\frac{\sum_{k\in S(\tau)}\sqrt{\frac{2266\mu_k^{+(\tau)}(\tau)\log NT}{\tilde{n}_k(\tau)}} + \frac{1364\log NT}{\tilde{n}_k(\tau)}}{1+\sum_{k\in S(\tau)}\omega_k(\tau)} + \delta^{(\tau)}$$

$$\leq \frac{\sum_{k\in S(\tau)}\left(\frac{1+\delta^{(\tau)}}{1-\delta^{(\tau)}}\mu_k^{+(\tau)}(\tau) - \omega_k(\tau)\right)}{1+\sum_{k\in S(\tau)}\omega_k(\tau)} + 3\frac{\sum_{k\in S(\tau)}\sqrt{\frac{2266\mu_k^{+(\tau)}(\tau)\log NT}{\tilde{n}_k(\tau)}} + \frac{1364\log NT}{\tilde{n}_k(\tau)}}{1+\sum_{k\in S(\tau)}\omega_k(\tau)} + \delta^{(\tau)}.$$

第一个不等式根据以下事实得出： $S(\tau)$ 是 $\hat{}\boldsymbol{\omega}(\tau)$ 的最优产品组合，并结合引理6，可得 $\hat{}\boldsymbol{\omega}(\tau)$ $\leq \boldsymbol{\omega}^{\mathsf{up}}(\tau)$。第二个不等式是因为 $\boldsymbol{\omega}(\tau)\ \leq \boldsymbol{\omega}^{\mathsf{up}}(\tau)$。第三个不等式是因为 and $\sum_{k\in S(\tau)}\delta_k^{(\tau)} \leq \delta^{(\tau)}$。第四个不等式根据我们在开头假设的集中界得出。最后一个不等式成立是因为 $\delta^{(\tau)} \leq \delta^{(t)} \leq \frac{1}{2}$，这意味着 $\frac{1+\delta^{(\tau)}}{1-\delta^{(\tau)}} \leq 3$。我们分别对上述各项进行界定。

对于第一项，我们有：

$$\frac{\sum_{k\in S(\tau)}\left(\frac{1+\delta^{(\tau)}}{1-\delta^{(\tau)}}\mu_k^{+(\tau)}(\tau)-\omega_k(\tau)\right)}{1+\sum_{k\in S(\tau)}\omega_k(\tau)}$$

$$=\frac{\sum_{k\in S(\tau)}\left(\frac{(1+\delta^{(\tau)})^2}{(1-\delta^{(\tau)})^2}-1\right)\omega_k(\tau)}{1+\sum_{k\in S(\tau)}\omega_k(\tau)}+\left(\frac{1+\delta^{(\tau)}}{1-\delta^{(\tau)}}\right)^2\cdot\frac{\sum_{k\in S(\tau)}\delta_k^{(\tau)}\omega_k(\tau)}{1+\sum_{k\in S(\tau)}\omega_k(\tau)}$$

$$\leq\frac{4\delta^{(\tau)}}{(1-\delta^{(\tau)})^2}\cdot\frac{\sum_{k\in S(\tau)}\omega_k(\tau)}{1+\sum_{k\in S(\tau)}\omega_k(\tau)}+9\cdot\frac{\sum_{k\in S(\tau)}\delta_k^{(\tau)}\omega_k(\tau)}{1+\sum_{k\in S(\tau)}\omega_k(\tau)}$$

$$\leq\frac{4\delta^{(\tau)}}{(\frac{1}{2})^2}\cdot\frac{\sum_{k\in S(\tau)}\omega_k(\tau)}{1+\sum_{k\in S(\tau)}\omega_k(\tau)}+9\cdot\frac{\sum_{k\in S(\tau)}\delta_k^{(\tau)}\omega_k(\tau)}{1+\sum_{k\in S(\tau)}\omega_k(\tau)}$$

$$\leq 16\delta^{(\tau)}+9\delta^{(\tau)}$$
$$=25\delta^{(\tau)}$$

第一个等式只是将 $\mu_k^{+(\tau)}(\tau)$ 替换为其表达式所得。第一个不等式基于以下事实:

$$\frac{(1+x)^2}{(1-x)^2}-1=\frac{4x}{(1-x)^2}$$

且 $\frac{1+\delta^{(\tau)}}{1-\delta^{(\tau)}}\leq 3$。第二个不等式依据 $\delta^{(\tau)}\leq\frac{1}{2}$，下一个不等式是因为$\sum_{k\in S(\tau)}\delta_k^{(\tau)}\leq\delta^{(\tau)}$，并且对于每个 k 和 τ，$\omega_k(\tau)\leq 1$。

现在来看第二项,

$$
\begin{aligned}
&3\frac{\sum_{k\in S(\tau)}\sqrt{\frac{2266\mu_k^{+(\tau)}(\tau)\log NT}{\tilde{n}_k(\tau)}}+\frac{1364\log NT}{\tilde{n}_k(\tau)}}{1+\sum_{k\in S(\tau)}\omega_k(\tau)}\\
&=3\frac{\sum_{k\in S(\tau)}\sqrt{\frac{2266\frac{1+\delta^{(\tau)}}{1-\delta^{(\tau)}}\left(\omega_k(\tau)+\delta_k^{(\tau)}\right)\log NT}{\tilde{n}_k(\tau)}}+\frac{1364\log NT}{\tilde{n}_k(\tau)}}{1+\sum_{k\in S(\tau)}\omega_k(\tau)}\\
&\le 3\frac{2\sum_{k\in S(\tau)}\sqrt{\frac{2266\left(\omega_k(\tau)+\delta_k^{(\tau)}\right)\log NT}{\tilde{n}_k(\tau)}}+\frac{1364\log NT}{\tilde{n}_k(\tau)}}{1+\sum_{k\in S(\tau)}\omega_k(\tau)}\\
&=6\sqrt{2266\log NT}\cdot\frac{\sum_{k\in S(\tau)}\sqrt{\frac{\omega_k(\tau)+\delta_k^{(\tau)}}{\tilde{n}_k(\tau)}}}{1+\sum_{k\in S(\tau)}\omega_k(\tau)}+4092\log NT\frac{\sum_{k\in S(\tau)}\frac{1}{\tilde{n}_k(\tau)}}{1+\sum_{k\in S(\tau)}\omega_k(\tau)}\\
&\le 6\sqrt{2266\log NT}\cdot\sqrt{\frac{\sum_{k\in S(\tau)}\omega_k(\tau)+\delta_k^{(\tau)}}{1+\sum_{k\in S(\tau)}\omega_k(\tau)}\cdot\frac{\sum_{k\in S(\tau)}\frac{1}{\tilde{n}_k(\tau)}}{1+\sum_{k\in S(\tau)}\omega_k(\tau)}}\\
&\quad+4092\log NT\frac{\sum_{k\in S(\tau)}\frac{1}{\tilde{n}_k(\tau)}}{1+\sum_{k\in S(\tau)}\omega_k(\tau)}\\
&\le 6\sqrt{2266\log NT}\cdot\sqrt{\frac{\delta^{(\tau)}+\sum_{k\in S(\tau)}\omega_k(\tau)}{1+\sum_{k\in S(\tau)}\omega_k(\tau)}\cdot\frac{\sum_{k\in S(\tau)}\frac{1}{\tilde{n}_k(\tau)}}{1+\sum_{k\in S(\tau)}\omega_k(\tau)}}\\
&\quad+4092\log NT\frac{\sum_{k\in S(\tau)}\frac{1}{\tilde{n}_k(\tau)}}{1+\sum_{k\in S(\tau)}\omega_k(\tau)}\\
&\le 6\sqrt{2266\log NT}\cdot\sqrt{\frac{\sum_{k\in S(\tau)}\frac{1}{\tilde{n}_k(\tau)}}{1+\sum_{k\in S(\tau)}\omega_k(\tau)}}+4092\log NT\frac{\sum_{k\in S(\tau)}\frac{1}{\tilde{n}_k(\tau)}}{1+\sum_{k\in S(\tau)}\omega_k(\tau)}.
\end{aligned}
$$

第一个不等式成立是因为 $\frac{1+\delta^{(\tau)}}{1-\delta^{(\tau)}}\le 4$。第二个不等式依据柯西-施瓦茨不等式。第三个不等式是因为 $\sum_{k\in S(\tau)}\delta_k^{(\tau)}\le\delta^{(\tau)}$，最后一个不等式是因为 $\delta^{(\tau)}\le\delta^{(t)}\le 1$。因此，

$$
\begin{aligned}
&\sum_{\tau=1}^{t} \hat{R}_\tau - R(S(\tau), \boldsymbol{\omega}(\tau)) \\
&\leq 26t\delta^{(t)} + 6\sqrt{2266 \log NT} \cdot \sum_{\tau=1}^{t} \sqrt{\frac{\sum_{k \in S(\tau)} \frac{1}{\tilde{n}_k(\tau)}}{1 + \sum_{k \in S(\tau)} \omega_k(\tau)}} + 4092 \log NT \sum_{\tau=1}^{t} \frac{\sum_{k \in S(\tau)} \frac{1}{\tilde{n}_k(\tau)}}{1 + \sum_{k \in S(\tau)} \omega_k(\tau)} \\
&\leq 26t\delta^{(t)} + 6\sqrt{2266 \log NT} \cdot \sqrt{t \cdot \sum_{\tau=1}^{t} \frac{\sum_{k \in S(\tau)} \frac{1}{\tilde{n}_k(\tau)}}{1 + \sum_{k \in S(\tau)} \omega_k(\tau)}} + 4092 \log NT \sum_{\tau=1}^{t} \frac{\sum_{k \in S(\tau)} \frac{1}{\tilde{n}_k(\tau)}}{1 + \sum_{k \in S(\tau)} \omega_k(\tau)}.
\end{aligned}
$$

第一个不等式由以下事实得出： $\delta^{(\tau)} \leq \delta^{(t)}$ 对于 $\tau \leq t$；第二个不等式则应用了柯西-施瓦茨不等式。接下来，固定某个时间步 $u \in [1, t]$，我们有：

$$
\begin{aligned}
\sum_{\tau=1}^{t} \frac{\sum_{k \in S(\tau)} \frac{1}{\tilde{n}_k(\tau)}}{1 + \sum_{k \in S(\tau)} \omega_k(\tau)} &\leq \sum_{\tau=1}^{t} \frac{\sum_{k \in S(\tau)} \frac{1}{\tilde{n}_k(\tau)}}{1 + \sum_{k \in S(\tau)} \omega_k(u) - \delta_k^{(t)}} \\
&= \sum_{j=1}^{N} \sum_{i=1}^{e_j^{(t)}} \frac{\left| l_j^{(t)}(i) \right| \cdot \frac{1}{\max\{1, i-1\}}}{1 + \sum_{k \in S\left(s_j^{(t)}(i)\right)} \omega_k(u) - \delta_k^{(t)}} \\
&\leq \sum_{j=1}^{N} \sum_{i=1}^{e_j^{(t)}} \frac{5 \log NT \left(1 + \sum_{k \in S\left(s_j^{(t)}(i)\right)} \omega_k(u) + \delta_k^{(t)}\right) \frac{1}{\max\{1, i-1\}}}{1 + \sum_{k \in S\left(s_j^{(t)}(i)\right)} \omega_k(u) - \delta_k^{(t)}}
\end{aligned}
$$

第一个不等式由以下事实得出：对于每个 $\tau, u \in [1, t]$, $1 + \sum_{k \in S(\tau)} \omega_k(\tau) \geq 1 + \sum_{k \in S(\tau)} \omega_k(u) - \delta_k^{(t)} \geq 0$。等号成立的原因如下：对于每个时间步 $\tau \in [1, t]$，以及在 τ 提供的每个商品 k，求和中都包含一项 $\mathbf{1}_{\tilde{n}_k(\tau)} \frac{1 + \sum_{k \in S(\tau)} \omega_k(u) - \delta_k^{(t)}}{}$。重新排列这些项，对于每个商品 j 及其被提议的每个时期 $i \in [1, e_j^{(t)}]$，都有一项 $\mathbf{1}_{\tilde{n}_j(\tau)} \frac{1 + \sum_{k \in S(\tau)} \omega_k(u) - \delta_k^{(t)}}{}$，对应于属于时期 $l_j^{(t)}(i)$ 的每个 τ。等号成立，这是因为在时期 $l_j^{(t)}(i)$ 的所有时间轮次中，提供的商品组合固定为 $S(s_j^{(t)}(i))$，且 $\tilde{n}_k(\tau)$ 在该时期的所有时间轮次中都固定为 $\max\{1, i-1\}$。最后一个不等式由时期长度的集中界得出。接下来，注意到，

$$
\begin{aligned}
4-\frac{1+\sum_{k\in S\left(s_j^{(t)}(i)\right)}\omega_k(u)+\delta_k^{(t)}}{1+\sum_{k\in S\left(s_j^{(t)}(i)\right)}\omega_k(u)-\delta_k^{(t)}} &= \frac{3+3\sum_{k\in S\left(s_j^{(t)}(i)\right)}\omega_k(u)-5\sum_{k\in S\left(s_j^{(t)}(i)\right)}\delta_k^{(t)}}{1+\sum_{k\in S\left(s_j^{(t)}(i)\right)}\omega_k(u)-\delta_k^{(t)}} \\
&\geq \frac{3-5\delta^{(t)}}{1+\sum_{k\in S\left(s_j^{(t)}(i)\right)}\omega_k(u)-\delta_k^{(t)}} \\
&\geq \frac{3-\frac{5}{2}}{1+\sum_{k\in S\left(s_j^{(t)}(i)\right)}\omega_k(u)-\delta_k^{(t)}} \geq 0
\end{aligned}
$$

因此,

$$
\begin{aligned}
\sum_{\tau=1}^{t}\frac{\sum_{k\in S(\tau)}\frac{1}{\tilde{n}_k(\tau)}}{1+\sum_{k\in S(\tau)}\omega_k(\tau)} &\leq 20\log NT\cdot\sum_{j=1}^{N}\sum_{i=1}^{e_j^{(t)}}\frac{1}{\max\{1,i-1\}} \\
&\leq 20\log NT\cdot(N\log T+2) \\
&\leq 40N\log^2 NT
\end{aligned}
$$

对于 $N\log\ T\geq 2$，注意，轮次数量为 $e_j^{(t)}\leq T$。因此,

$$
\frac{1}{t}\sum_{\tau=1}^{t}\hat{R}(\tau)-R(\tau)\leq 26\delta^{(t)}+1807(\log NT)^{\frac{3}{2}}\sqrt{\frac{N}{t}}+163680(\log NT)^3\frac{N}{t}.
$$

最后,

$$
\begin{aligned}
\frac{1}{t}\sum_{\tau=1}^{t}\hat{R}(\tau)-r(\tau) &\leq 26\delta^{(t)}+1807(\log NT)^{\frac{3}{2}}\sqrt{\frac{N}{t}}+163680(\log NT)^3\frac{N}{t}+\sqrt{\frac{2\log T}{t}} \\
&\leq \sum_{\tau=1}^{t-1}\Delta(\tau)+\rho(t)
\end{aligned}
$$

□

B 下界：省略的证明

定理 2. 固定 T、L、N和 K，并假设 $K\leq\frac{N}{4}$。那么，对于每个用于非平稳环境中 *MNL*-老虎机问题的多项式时间算法 $\mathcal{A}$，都存在一个参数为 N 和 K 且至多发生 L 次切换的实例，使得 $\mathcal{A}$累积的遗憾至少为

$C \cdot \min(\sqrt{NLT}、T)$，经过 T轮次，其中 $C > 0$是一个与 T、L、N和K无关的绝对常数。

我们对定理 2 的证明使用了 X. Chen 和 Wang，2017 关于平稳环境的研究结果。在这篇论文中，Chen 和 Wang 给出了平稳环境下 MNL 多臂老虎机的最优下界 $\tilde{\Omega}(\sqrt{NT})$。遵循 X. Chen 和 Wang，2017，我们假设 $K \leq \frac{N}{4}$。这一假设在实践中很容易验证。例如，在在线零售中，每轮我们能向顾客展示的商品数量通常远小于商品总数。

证明。固定 T、L、N 和 $K \leq \frac{N}{4}$，并令 A 为一个用于非平稳环境下 MNL 多臂老虎机的多项式时间算法。若 $L = 1$，则可以直接选取 X. Chen 和 Wang，2017 的定理 1 给出的一个（平稳）对抗性实例；由于平稳环境中的切换次数为 $L = 1$，因此该情形下结论成立。否则，若 $L \geq 2$，我们将决策时域划分为 L 个区间，每个区间包含 $\left\lceil \frac{T}{L} \right\rceil$ 个时间步（最后一个区间可能少于此数）。我们递归地构造一个对抗性实例，使得对于某个通用常数 $C > 0$，A 在前 $L-1$ 个区间中的每个区间内都遭受至少 $C \cdot \min(\sqrt{N\left\lceil \frac{T}{L} \right\rceil}, \left\lceil \frac{T}{L} \right\rceil)$ 的期望遗憾。具体而言，假设我们已固定前 l（$\leq L-2$）个区间的对抗性实例。对于第 $l+1$ 个区间接下来的 $\left\lceil \frac{T}{L} \right\rceil$ 个时间步，X. Chen 和 Wang，2017 的定理 1 表明，存在一个（平稳）对抗性实例，使得算法 A（更准确地说，是 A 在第 $l+1$ 个区间内、以我们已固定的对抗性实例部分为条件所采取的随机策略）在第 $l+1$ 个区间内遭受至少 $C \cdot \min(\sqrt{N\left\lceil \frac{T}{L} \right\rceil}, \left\lceil \frac{T}{L} \right\rceil)$ 的遗憾，其中 $C > 0$ 是某个通用常数。我们为第 $l+1$ 个区间选取该实例，并继续构造。最后一个区间（第 L 个区间）中的实例则任意补全。我们的构造至多包含 L 次切换，算法 A 在整个时间范围T内遭受的遗憾至少为 $(L-1) \cdot C\min(\sqrt{N\left\lceil \frac{T}{L} \right\rceil}, \left\lceil \frac{T}{L} \right\rceil)$ $\geq \frac{C}{2} \cdot \min(\sqrt{NLT}, T)$。

□

定理 3。固定 T, Δ, N和 K，并假设 $K \leq \frac{N}{4}$且 $\Delta \in [\frac{1}{N}, \frac{T}{N}]$。那么，对于任何用于非平稳 *MNL* 多臂老虎机问题的多项式时间算法 $\mathcal{A}$，都存在一个参数为 N 和 K、变化量为 $\Delta_\infty^K \leq \Delta$ 的实例，使得$\mathcal{A}$ 在时间范围 T内累积的遗憾至少为$C \cdot N^{\frac{1}{3}}\Delta^{\frac{1}{3}}T^{\frac{2}{3}}$，其中 $C > 0$是一个与 T、Δ、N和 K无关的绝对常数。此外，该实例满足 $\Delta_\infty^K \geq \frac{K}{2} \cdot \Delta_\infty$，其中 Δ_∞ 表示该实例吸引参数在范数 L_∞ 下的变化量。

为证明定理 3，我们将时间范围划分为长度适当的窗口，并利用 X. Chen和 Wang 在 2017 年的结果，为每个窗口构造一个平稳对抗实例。通过恰当地选择这些对抗实例，我们得到的最终实例满足 $\Delta_\infty^K \geq \frac{K}{2} \cdot \Delta_\infty$。

证明。固定 T、N、$K \leq \frac{N}{4}$ 和 $\Delta \in [\frac{1}{N}, \frac{T}{N}]$，并令 A 为用于非平稳环境中 MNL-Bandit 的多项式时间算法。我们将决策时域划分为 $\left\lceil \frac{T}{M} \right\rceil$ 个区间，每个区间的

长度为 M （最后一个区间可能除外，其时间步可能较少），其中 M 将在后文确定。令 M_l 表示第 l 个区间的长度，并令 $\epsilon \in (0, 1)$。我们递归地构造一个针对算法 A 的对抗性实例。具体而言，假设我们已经固定了前 l （$\leq \left\lceil \frac{T}{M} \right\rceil - 1$）个区间的对抗性实例。对于第（$l+1$）个区间，我们进行如下操作：对于每个 $\eta \in (0, 1\ 2)$，X. Chen 和 Wang，2017 证明了存在一种针对 A 的对抗策略 $\pi(\eta)$，其中恰有 K 件商品的吸引参数为 $\frac{1+\eta}{K}$，而其他每件商品的吸引参数均为 $\frac{1}{K}$，且在该策略下，A 至少会产生如下遗憾

$$\frac{\eta}{9}\left(\frac{2M_{l+1}}{3} - M_{l+1}\sqrt{\frac{126 \cdot M_{l+1} \cdot \eta^2}{N}}\right)$$

在第 $l+1$ 个区间的 M_{l+1} 轮中。如果 l 为偶数，我们为第 $l+1$ 个区间选择 $\pi(\epsilon)$，并继续构造。如果 l 为奇数，我们选择 $\pi(\frac{\epsilon}{2})$。那么，根据这些选择，

$$\epsilon = \min\left(\frac{1}{17}\sqrt{\frac{N}{M}}, \frac{1}{2}\Delta \cdot \frac{M}{T}\right) \quad (\leq \frac{1}{2})$$

and

$$M = \left\lceil N^{\frac{1}{3}}\left(\frac{T}{\Delta}\right)^{\frac{2}{3}}\right\rceil,$$

所构造实例中的吸引力参数最多变化 $\left\lceil \frac{T}{M} \right\rceil - 1$ 次（发生在区间之间）；每次变化时，最多有 $2K$ 个元素的吸引参数发生变化，且变化量（绝对值）不超过 $\frac{\epsilon}{K}$。因此，所构造实例的变化量 Δ_∞^K 至多为

$$\frac{T}{M} \cdot 2K \cdot \frac{\epsilon}{K} = \frac{2T}{M}\epsilon \leq \Delta.$$

此外，根据构造，在任意两个区间（一个奇数区间和一个偶数区间）之间，至少有 K 个商品的吸引参数发生了至少 $\frac{\epsilon}{2}$ （绝对值）的变化。由于两个区间之间的范数变化 L_∞ （即单个参数发生的最大变化）至多为 ϵ，因此可得

$$\Delta_\infty^K \geq \frac{K}{2}\Delta_\infty.$$

现在，算法 A 在整个时间范围内遭受的期望遗憾至少为，

$$\begin{aligned}
\sum_{l=1}^{\lceil \frac{T}{M} \rceil} \frac{1}{9} \cdot \frac{\epsilon}{2} \left(\frac{2M_l}{3} - M_l \sqrt{\frac{126 M_l \epsilon^2}{N}} \right) &\geq \sum_{l=1}^{\lceil \frac{T}{M} \rceil} \frac{1}{9} \cdot \frac{\epsilon}{2} \left(\frac{2M_l}{3} - M_l \sqrt{\frac{126 M \epsilon^2}{N}} \right) \\
&= \frac{1}{9} \cdot \frac{\epsilon}{2} \left(\frac{2T}{3} - T \sqrt{\frac{126 M \epsilon^2}{N}} \right) \\
&= \frac{\epsilon T}{18} \left(\frac{2}{3} - \sqrt{126} \left(\epsilon \sqrt{\frac{M}{N}} \right) \right) \\
&\geq \frac{\epsilon T}{18} \left(\frac{2}{3} - \frac{\sqrt{126}}{17} \right) \\
&\geq C \cdot \epsilon T \\
&\geq C \cdot T \cdot \min \left(\frac{1}{17} \sqrt{\frac{N}{M}}, \frac{1}{2} \Delta \cdot \frac{M}{T} \right) \\
&\geq C \cdot T \cdot \min \left(\frac{1}{17} N^{\frac{1}{3}} T^{-\frac{1}{3}} \Delta^{\frac{1}{3}}, \frac{1}{2} N^{\frac{1}{3}} \Delta^{\frac{1}{3}} T^{-\frac{1}{3}} \right) \\
&\geq \frac{C}{17} N^{\frac{1}{3}} T^{\frac{2}{3}} \Delta^{\frac{1}{3}}
\end{aligned}$$

其中， $C > 0.0003$ 是一个与问题参数无关的绝对常数。第一个不等式成立是因为，对于每个 l，都有 $M_l \leq M$；接下来的等式成立是因为 $\sum_{l=1}^{\lceil \frac{T}{M} \rceil} M_l = T$。

□