arXiv:2603.05619v1 [stat.AP] 5 Mar 2026

先检验后惩罚：重复博弈的统计方法

Aymeric Capitaine[1, *]，Antoine Scheid[1, *]，Etienne Boursier[2]，Alain Durmus[1]，和 Michael I. Jordan[3]

[1]CMAP、CNRS、巴黎综合理工学院，法国[2]Inria Saclay、Université Paris Saclay、LMO，法国[3]Inria Paris、巴黎高等师范学院、PSL大学，法国；加州大学伯克利分校，美国[*]共同贡献

2026年3月9日

摘要

我们研究折扣无限重复博弈。在这类博弈中，参与者同意采用一个合作混合行动组合，但每一步只能观察到实现的纯行动。这种不完美监控会破坏经典的触发策略，因为无法确定地识别偏离。为解决这一问题，我们研究如何利用假设检验来维持合作。首先，我们提出一个将统计推断直接融入策略行为的框架。我们引入宽松均衡概念，允许参与者忽略由监控过程中的罕见但极端实现所引发的概率趋于零的历史。在这一框架内，我们形式化提出一种通用先检验后惩罚策略：参与者事前承诺采用合作混合行动组合，持续检验观察到的行动是否符合这一约定，并在累积了足够的偏离统计证据后永久转而采取惩罚。在检验程序满足温和条件时，对于足够有耐心的参与者，这一构造能够维持任何可行且个人理性的收益，从而在不完美监控下得到类似民间定理的结果。随后，我们提出该策略的两种具体实现。第一种依赖任意时刻有效序贯检验，可在无限期内统一控制第一类错误，并能在有限期望时间内检测出与收益相关的偏离。不过，这种策略只考虑平稳偏离，并导出纳什均衡。第二种采用固定大小的批次检验，能够应对任意偏离并实现子博弈完美纳什均衡，但代价是无法对错误惩罚提供全局任意时刻保证。总之，这些结果阐明了统计可靠性与博弈论稳健性之间的根本权衡，为重复策略环境中的数据驱动合作提供了原则性基础。

1 引言

在许多经济领域，维持策略主体之间的合作至关重要，涉及产业组织、金融市场、环境监管和国际贸易等。基于惩罚的策略是一种常见的解决方案：参与者商定采取一项合作行动，并

只要观察到对手也采取相同行动，就继续遵循这一行动。一旦发现偏离，他们便转入惩罚阶段。如果惩罚威胁可信，且相对于偏离带来的短期收益足够严厉，那么任何参与者都没有动机偏离合作策略。

在重复博弈的文献中占据核心地位的是民间定理（Friedman，1971年），它形式化了这类基于惩罚的策略为何有效（Fudenberg and Tirole，1991年；Mailath and Samuelson，2006年）。该定理断言，给定阶段博弈中任何可行且个人理性的收益组合，在博弈无限重复且参与者足够有耐心时，都可以作为子博弈完美纳什均衡（Moore and Repullo，1988年）得以维持。这一基础性结果在完美监控下已得到证明，即假设参与者能够准确观察对手的策略；在某些不完美监控情形下也已得到证明，在这些情形中，参与者只能观察到与对手策略不完全相关的信号。

后一种情形具有相当大的现实意义，因为经济主体往往不得不在不完全信息结构下运作。然而，这种策略向来难以分析，因此催生了多种方法；这些方法主要给出存在性结果，而非提供可实施的策略。因此，这些方法通常无法产生能够直接利用现代学习技术以及许多现实环境中丰富数据的程序。鉴于现实中的经济主体经常运用统计分析来实施基于惩罚的合作策略，这一点尤为遗憾。以下两个例子便说明了这一点。

财务审计提供了一个具体的例子。企业与审计师会在多个时期内反复互动，企业理应如实报告财务报表，而偏离则表现为虚假报告或操纵。由于这类行为通常无法在单个时期内被确定地发现，因此合规情况需通过统计程序进行评估。正如美国上市公司会计监督委员会（Public Company Accounting Oversight Board，PCAOB）所确立的那样，审计师会根据趋势或回归模型形成定量期望，并调查与这些基准在统计上不一致的结果。当积累了足够的证据后，偏离行为会招致制裁或法律责任。尽管监控存在噪声，但延迟而严厉的惩罚前景能够遏制短期偏离，并维持如实报告。

职业体育中的反兴奋剂执法，是通过明确的统计检验在重复互动中维持合作的又一例子。运动员反复选择是否遵守规则参赛，但个人行动无法被观察到，生物测量结果也存在噪声。在世界反兴奋剂机构实施的运动员生物护照计划下，合规情况通过纵向统计模型进行评估，以检验运动员的生物标志物谱在统计上是否符合自然变异。当累积的

证据表明清白行为极不可能时，就会触发制裁（Agency，2023）。在这里，惩罚的威胁再次促成了合作。

不完美监控下重复博弈的相关文献与现实经济互动之间存在差距：前者大多抽象掉了以统计学为基础的合作策略，而后者则普遍涉及数据和学习技术。这引出了以下问题：我们如何才能将假设检验的运用形式化，以维持重复博弈中的合作？

此外，如果统计监控能够利用预测和推断方法，它就会更具吸引力。这引出了第二个更注重实践的问题：是否存在一种简单直接的检验程序，能够切实应用于这一问题的现实情境？

我们通过为不完美公共监控下的重复博弈建立一个统计框架来回答第一个问题，该框架基于两个关键概念。首先，由于统计推断具有概率性，我们引入了一些均衡概念，放宽了纳什均衡和子博弈精炼纳什均衡的要求，允许参与者忽略概率极低的历史，以及由监控过程的极端实现所产生的尾部事件。其次，我们将一种通用的先检验后惩罚 策略形式化：在博弈开始之前，参与者事前承诺采用一个能够实现目标可行收益的合作混合行动组合。随后，他们只观察实现的纯行动，并持续检验原假设，即每个对手都遵循规定的混合策略。只要序贯检验未能拒绝原假设，合作就会持续。一旦检验发现了足够的偏离证据，参与者就会永久转向报复性惩罚组合。在检验程序满足温和条件的情况下——我们之后设计的检验恰好满足这些条件——对于足够耐心的参与者而言，这一范式能够维持任何可行且个人理性的收益，从而在不完美公共监控下得到一种民间定理式结果。我们强调，我们的统计方法使我们能够绕开分析不完美监控下均衡时通常采用的标准可分解性和自生成技术（Abreu 等，1990）。相反，我们借鉴完美监控文献中常见的证明技术，并利用概率工具处理随机信号带来的随机性。这种方法的价值在于，它能够推导出一种简单的策略，依靠一条惩罚规则，并明确控制惩罚的时机。

我们通过引入两种显式检验策略来回答第二个问题。第一种策略利用任意时刻有效检验和 e 过程（Ramdas和Wang，2025）。该框架能够同时对所有游戏轮次中的假设进行检验，并且无论选择何种停止时间，都能保持有效性。假设参与者的策略是平稳的，我们证明，可以构造一种策略，使其作为纳什均衡几乎实现任何可行的理性收益组合。在这一背景下，两类统计错误至关重要。一方面，当原假设为真却被拒绝时，就会发生第一类错误，这对应于在参与者仍进行合作时进入惩罚阶段。任意时刻有效检验确保该事件的概率保持在预先设定的阈值以下。我们强调，与文献中通常采用的基于信念的方法或无信念方法不同，我们的方法首次为博弈提供了这样的

	均衡 type	非平稳策略	第一类错误保证	第二类错误保证
任意时间（第3节）	Nash	✗	✓	✓
批次（第4节）	子博弈完美纳什均衡	✓	✗	✓

表1：各方法属性比较。

保证；后两类方法分别见（Sekiguchi，1997年）和（Ely等人，2005年），但它们通常忽视了这一风险。另一方面，当参与者未能检测到偏离且未触发惩罚时，就会发生第二类错误。我们建立了一个关于参与者发生偏离时合作期望停止时间的有限上界，该上界取决于其偏离的幅度。我们关于期望停止时间上界的结果在检验文献中尚属新颖，也可能具有独立的研究价值。

第二种策略旨在克服前一种方法的两个局限：它仅支持平稳偏离，并且只能提供纳什均衡（而非子博弈精炼纳什均衡）。我们不再在所有时期持续检验同一个原假设，而是提出一种批次先检验后惩罚策略，将博弈划分为连续的轮次区块。在每个批次结束时，参与者根据该区块中实际行动的经验分布进行统计检验，以评估其是否与规定的合作混合策略一致。如果某个批次中的原假设被拒绝，参与者就永久切换至惩罚策略组合。我们证明，这一构造可以逼近任意可行且个人理性的收益，并产生一个子博弈精炼纳什均衡。然而，该批次程序无法对整个博弈期限内的第一类错误和第二类错误提供统一且明确的控制。

总体而言，我们提出的两种检验程序之间存在权衡。一方面，随时检验方法具有明确的第一类错误和第二类错误保证，因此当需要此类要求时，它是一个颇具吸引力的选择，通常原因在于风险厌恶或公平性。然而，先检验后惩罚的随时检验方法从博弈论角度来看较弱，因为它只能阻止平稳偏离，并且不具备子博弈完美性。相反，批量先检验后惩罚可以处理任意偏离，并且构成子博弈精炼纳什均衡，但无法针对假阳性提供统计保证。我们在表1中总结了这一讨论。归根结底，在随时检验程序和批次程序之间进行选择，取决于参与者更看重统计可靠性，还是更强的博弈论保证。

总之，我们的主要贡献包括：

1. **统计监测框架**。我们将玩家规定混合行动、但只能观察到实现的纯行动时的重复战略互动形式化。为考虑统计监测中的罕见事件，我们扩展纳什均衡/子博弈完美均衡，以纳入那些

发生概率很小的历史。

2. **通用先检验后惩罚策略**。我们提出一种通用策略：参与者事前承诺采用以某个可行收益为目标的合作混合策略组合，检验对手的行动是否符合这一约定，并在发现偏离时转而实施惩罚。在序贯检验满足温和条件时，该策略可在不完美监控下恢复民间定理的结论。

3. **显式检验策略**。最后，我们用两种序贯检验实例化这一框架：第一种检验能控制无限期内的误惩罚，同时仍可在有限时间内检测到偏离。第二种基于区块的变体可处理任何形式的偏离，并提供更强的高概率序贯保证，但代价是全局任意时间第一类错误控制较弱。

相关工作

我们的研究处于多个不同领域的交汇处：重复博弈、算法博弈论和假设检验。

首先，民间定理一直是大量文献研究的主题。阶段博弈的重复进行会扩大纳什均衡下可维持的收益组合，这一思想至少可追溯至 Friedman（1971年）；他假设存在完美公共监控，并考察时间平均收益。在相同条件下，这一结果也适用于子博弈精炼纳什均衡（Rubinstein，1980），并进一步推广到贴现收益（Fudenberg 和 Maskin，1986年）。民间定理也在不完美监控 下得到确立（Aumann 等人，1995年）：参与者无法观察对手实际选择的混合策略，只能观察到与之相关的公共信号。借助可分解性和自生成技术（Abreu 等人，1990），可以证明，只要满足两两可识别性，任何可行的理性收益都能在子博弈精炼纳什均衡下得以维持，尽管这一证明并非构造性的（Fudenberg 等人，1994）。私人信号带来了进一步的细化（Mailath 和 Morris，2002）：参与者不再观察共同信号，而是接收不同且可能相互冲突的信号，这阻碍了即时协调。为解决这一困难，人们提出了多种方法，包括允许参与者之间进行通信（Compte，1998）、假设私人信号近乎完美（Ely 和 Välimäki，2002；Hörner 和 Olszewski，2006），或采用无信念策略（Ely等人，2005）。此外，Gossner（1995）研究了以混合策略进行的惩罚阶段能否得以维持；Sugaya（2022）则通过较长的审查阶段以及借助行动进行隐式通信来协调惩罚，在私人监控下确立了一个存在性民间定理。近期进展从多个方向拓展了重复博弈分析的范围，包括参与者可能随时间面对不同阶段博弈的不确定重复博弈（Laclau，2012）、有限博弈中的不完美监控（Horner 和 Renault，2025），以及存在转换成本的环境（Tsodikovich 等人，2024）。民间定理类型的结果也已

在更复杂的策略环境中得到确立，例如竞争机制（Attar 等人，2021）、匹配市场（Deb 等人，2016）、随机博弈（Aiba，2014）和网络博弈（Laclau，2012）。最后，近期研究将民间定理与算法共谋联系起来（Álvaro Cartea 等人，2026）；算法共谋是一个迅速发展的研究领域，探讨合谋行为能否从算法之间的互动中产生（Calvano 等人，2020）。

本文为经济学与机器学习交叉领域日益丰富的文献（Jordan，2025）作出贡献，旨在超越传统的不确定性下均衡分析，以可实施的学习算法取代隐式的信念更新规则（Erev and Roth，1998；Cesa-Bianchi and Lugosi，2006；Roughgarden，2010）。在这一研究方向中，参与者依据观察到的结果，利用数据驱动的程序调整行动。这种方法已成功应用于经典微观经济学问题的研究，包括动态定价（Bertsimas and Perakis，2006；Mueller et al.，2019）、匹配（Liu et al.，2020；Johari et al.，2021）、外部性的管理（Scheid et al.，2024）、逆向选择与道德风险（Bates et al.，2022；Capitaine et al.，2024）、拍卖竞价（Morgenstern and Roughgarden，2016；Feng et al.，2018）、社会选择（Brandt et al.，2012）等。委托—代理问题是一个典型例子，近期的研究运用学习技术来执行合同（Fallah and Jordan，2024；Collina et al.，2024）、推断参与者的偏好（Kolumbus et al.，2024；Scheid et al.，2025），或与未知对手博弈（Mansour et al.，2020；Kolumbus and Nisan，2022；Arunachaleswaran et al.，2025）。该领域的一个重要目标是提供有限时间和有限样本保证，例如给出收敛至均衡或社会最优结果的显式收敛速率（Syrgkanis et al.，2015）。本文聚焦于无限期重复博弈，通过引入明确的基于检验的策略来维持合作，推进了这一研究计划。

最后，我们提出的偏离检测机制主要建立在统计假设检验的经典理论之上（Fisher，1925年；Neyman 和 Pearson，1933；Wald，1945；Lehmann 和 Romano，2005年）。从证据的贝叶斯表述（Andraszewicz 等，2015）到以最优性为导向的检验构造通论（Lehmann 和 Romano，2005年），再到现代多重假设检验理论（Benjamini 和 Hochberg，1995年；Shaffer，1995），大量研究塑造了当代假设检验的发展。我们的技术方法更具体地依赖于任意时刻有效序贯检验的近期理论，尤其是 e过程的概念（Shafer 和 Vovk，2005；Ramdas 等，2022年；Howard 等，2021年）。e过程提供了基于鞅的统计证据度量，可以在线更新，无需预先固定样本量，同时在任意停止规则下保持有效。特别是，e过程（均值上界为一的非负过程）可通过 Ville 不等式（Ville，1939）直接实现无限期内统一的第一类错误控制。尽管任意时刻有效检验在现代序贯决策问题中日益得到应用，但据我们所知，我们的工作是首个将任意时刻有效推断作为核心执行基元，应用于重复博弈论

测试与惩罚构造。

组织结构

本文结构如下。第2节介绍我们的框架，并回顾完美监控下的民间定理。第3节介绍随时测试后惩罚策略，并说明只要能够控制第一类和第二类错误，该策略便能在不完美监控下导出民间定理。随后，我们利用 e 过程明确构造满足这些条件的序贯检验。最后，在第4节中，我们介绍批次先检验后惩罚策略，并在较弱的条件下建立不完美监控下的另一版本民间定理。这里，我们同样给出满足上述条件的简单序贯检验。

2 设定与民间定理

记号 对于任意有限可测空间 A，我们用 $\Delta(A)$ 表示 A 上概率测度的集合。特别地，若 $A = \{a_1, \ldots, a_K\}$，对于任意 $w \in \Delta(A)$，我们用 $w[a_\ell] \in [0,1]$ 表示 w 在 $a_\ell \in A$ 上的质量。若 $A = A^1 \times \ldots \times A^N$ 是若干个空间的乘积，对于任意 $i \in [N] A^{-i} = A^1 \times \ldots \times A^{i-1} \times A^{i+1} \times \ldots \times A^N$，我们记作 $A = A^i \times A^{-i}$.。类似地，若 $w = w^1 \otimes \ldots \otimes w^N$ 是 A 上的乘积测度，对于任意 $i \in [N]$，我们用 $w^{-i} = w^1 \otimes \ldots \otimes w^{i-1} \otimes w^{i+1} \otimes \ldots \otimes w^N$ 表示，因此 $w = w^i \otimes w^{-i}$。最后，对于任意有限集合 X，我们用 conv(X) 表示 X 的凸包。

2.1 框架

我们考虑一个具有无限期和 $N > 0$ 名参与者的重复博弈。每位玩家 $i \in [N]$ 都可以使用一个纯行动集 $A^i = \{a_1^i, \ldots, a_K^i\}$，满足 $\mathrm{Card}(A^i) = K > 0$。他们的偏好由效用函数 $u^i : A \to [0,1]$ 表示，其中 $A = A^1 \times \ldots \times A^N$。在任意轮次 $t \geqslant 0$，每位玩家 i 选择一种混合策略 $w_t^i \in \Delta(A^i)$，因此其期望收益为 $u^i(w_t^i, w_t^{-i})$。这里稍微滥用一下记号，将 $u^i(w_t^i, w_t^{-i}) = \sum_{a^i} 写作 {}_{a^{-i}}\, u^i(a^i, a^{-i})$ $w_t^i[a^i]\, w_t^{-i}[a^{-i}]$。在任意时刻 $t \geqslant 0$，令 H_t 为 A 的 t 元笛卡尔积。玩家可以查看公共历史 $h_t \in H_t$，其中包含过去实现的纯行动，记作 $h_t = \{(A_0^i, A_0^{-i}) \ldots, (A_{t-1}^i, A_{t-1}^{-i})\}$，并可据此制定策略。注意，我们的信号集就是纯行动集 A，对应于典范信号空间（Hörner和Olszewski，2006年）。形式上，给定 $H = \cup_{t \geqslant 0} H_t$，玩家 i 的行动序列 $(w_0^i, w_1^i, \ldots)$ 按照一个（行为）策略生成

$$s^i : \mathcal{H} \to \Delta(\mathcal{A}^i),$$

它将历史映射为纯行动上的分布。我们用 $\mathcal{S}^i$ 表示玩家 i的此类策略集合，并记作 $\mathcal{S} = \mathcal{S}^1 \times \ldots \times \mathcal{S}^N$。请注意，我们关注的是公开策略，它们只取决于

公共历史 $\mathrm{h}_t \in \mathcal{H}$，而非私人历史[1]。接下来，重复博弈按如下方式展开。从历史 $\mathrm{h}_0 = \emptyset$开始，每位玩家 $i \in [N]$ 选择一个策略 $s^i \in \mathcal{S}^i$，满足 $s^i(\mathrm{h}_0) = w_0^i$，并且对任意 $t \geqslant 0$

1. 他们采用混合策略 $w_t^i = s^i(\mathrm{h}_t)$,

2. 对每位参与者 $i \in N$，从中抽取一个纯行动 $A_t^i \sim w_t^i$，

3. 参与者获得收益 $u^i(A_t^i,\ A_t^{-i})$，并观察到 $A_t^{-i} \in \mathcal{A}^{-i}$。随后将下一轮的历史更新为 $\mathrm{h}_{t+1} = \mathrm{h}_t \cup \sigma(\{A_t^i,\ A_t^{-i}\})$。

请注意，由于第 3 步，监控并不完善，因为公共历史只包含纯行动，而不包含混合策略。对于任意策略组合 $\mathbf{s} = (s^1,\ \ldots,\ s^N) \in \mathcal{S}$，策略和参与者各自独立的随机化共同在无限行动历史 $\mathcal{A}^{\mathbb{N}}$的空间上诱导出唯一的概率测度 $\mathbb{P}^{\mathbf{s}}$。我们用 $\mathbb{E}^{\mathbf{s}}$ 表示关于 $\mathbb{P}^{\mathbf{s}}$的期望。参与者 i在策略组合 $\mathbf{s} = (s^1,\ \ldots,\ s^N) \in \mathcal{S}$下的折扣收益为

$$U^i(\mathbf{s}) = (1-\beta) \sum_{t=0}^{\infty} \beta^t \, \mathbb{E}^{\mathbf{s}} \left[u^i(A_t^1, \ldots, A_t^N) \right], \tag{1}$$

其中，$0 < \beta < 1$ 是一个共同折扣因子。最后，对于任意时刻 $t \geqslant 0$ 和历史 $\mathrm{h}_t \in \mathcal{H}$，我们还将玩家 i 采用策略$\mathbf{s}$时的延续收益 定义为

$$U^i(\mathbf{s}\,;\, \mathrm{h}_t) = (1-\beta) \sum_{\ell=t}^{\infty} \beta^{\ell-t} \, \mathbb{E}^{\mathbf{s}} \left[u^i(A_\ell^1, \ldots, A_\ell^N) \,|\, \mathrm{h}_t \right]. \tag{2}$$

现在我们介绍本文使用的两个公共均衡概念。

定义 1 (($\varepsilon, \mathfrak{S}$)–纳什均衡)。 对于任意 $\mathfrak{S} = \mathfrak{S}^1 \times \ldots \times \mathfrak{S}^N \subseteq \mathcal{S}$ 和 $\varepsilon \geqslant 0$，若策略组合$\mathbf{s}_\star = (s_\star^1, \ldots,\ s_\star^N) \in \mathcal{S}$满足以下条件，则称其为($\varepsilon, \mathfrak{S}$)-纳什均衡

$$U^i(s_\star^i, s_\star^{-i}) \geqslant U^i(s^i, s_\star^{-i}) - \varepsilon \qquad \text{for any} \quad i \in [N] \quad \text{and} \quad s^i \in \mathfrak{S}^i \,.$$

定义1对应于经典的 ε-纳什均衡概念，其中参与者可用的偏离集合可以限制为策略的一个子集 $\mathfrak{S} \subset \mathcal{S}$（自然地，设定 $\mathfrak{S} = \mathcal{S}$即可恢复 ε-纳什均衡）。在下文中，我们将($\varepsilon, \mathfrak{S}$)-纳什均衡称为($\varepsilon$, $\mathfrak{S}$)–纳什均衡，并将(ε,$\mathcal{S}$)-纳什均衡称为 ε-纳什均衡。从较高层面来看，这一概念的动机在于：将所有可能的偏离纳入带有统计检验的均衡分析，在信息论上是不可行的——详见第3节的详尽讨论。我们还引入如下子博弈精炼均衡概念。

[1]对于任意玩家 $i \in [N]$，我们确实可以定义私人历史 $\tilde{\mathrm{h}}_t = \{(w_0^i, A_0^i, A_0^{-i}) \ldots, (w_{t-1}^i, A_{t-1}^i, A_{t-1}^{-i})\} \in \tilde{\mathcal{H}}_t$，并考虑私人策略 $\tilde{s}^i : \cup_{t \geqslant 0} \tilde{\mathcal{H}} \to \Delta(\mathcal{A}^i)$。然而，这会破坏重复博弈的递归结构，并使分析变得复杂得多（Mailath and Samuelson，2006年）。将我们的研究扩展到这一情形，是一个值得未来探索的方向。

定义2 ((ε,δ)-HP-SPNE)。 对于任意 $\delta \in (0,1)$和 $\varepsilon \geqslant 0$，若对于任意 $t \geqslant 0$，我们将策略组合$\mathbf{s}_\star = (s_\star^1, \ldots, s_\star^N) \in \mathcal{S}$称为 δ-高概率 ε-子博弈精炼纳什均衡

$$\text{there exists} \quad \mathsf{H_t} \subseteq \mathcal{H}_t \quad \text{such that} \quad \mathbb{P}^{\mathbf{s}_\star}(\mathrm{h}_t \in \mathsf{H}_t) \geqslant 1-\delta\,,$$

且对于任意历史 $\mathrm{h}_t \in \mathrm{H}_t$,

$$U^i(s_\star^i, s_\star^{-i}\,;\mathrm{h}_t) \geqslant U^i(s^i, s_\star^{-i}\,;\mathrm{h}_t) - \varepsilon \quad \text{for any } i \in [N] \quad \text{and} \quad s^i \in \mathcal{S}^i\,.$$

据我们所知，重复博弈文献尚未对定义 2 进行形式化表述。它可以被视为子博弈（公开）精炼的一种概率放宽形式，其中只要求在均衡策略下以高概率出现的历史上满足近似序贯理性。注意，如果 $\delta = 0$，我们便得到经典的 ε-子博弈完美纳什均衡概念。同样，由于我们的方法具有统计性质，必须在均衡分析中排除退化子博弈。在本文余下部分，我们将这一均衡概念称为（ε，δ）-HP-SPNE。我们将（0，0）-HP-SPNE称为SPNE。

2.2 完美监控下的民间定理

我们现在回顾混合策略下的经典民间定理，最终目标是将其扩展到不完美监控的情形。为便于阐述，本节仅假设参与者享有完美监控。具体而言，本节假定重复博弈按如下方式展开：从历史 $\mathrm{h}_0 = \emptyset$开始，玩家 i 选择一个策略 $s^i \in \mathcal{S}^i$，使得 $s^i(\mathrm{h}_0) = w_0^i$，且对于任意 $t \geqslant 0$

1. 他们采用混合策略 $w_t^i = s^i(\mathrm{h}_t)$,

2. 他们获得收益 $u^i(w_t^i,\ w_t^{-i})$，并观察到 $w_t^{-i} \in \Delta(\mathcal{A}^{-i})$。公共历史在下一轮更新为 $\mathrm{h}_{t+1} = \mathrm{h}_t \cup \sigma(\{w_t^i,\ w_t^{-i}\})$。

请注意，对于任意 $t \geqslant 0$，历史 $\mathrm{h}_t = \{(w_0^i,\ w_0^{-i})\ldots\ldots,\ (w_{t-1}^i,\ w_{t-1}^{-i})\}$ 包含迄今为止实际采用的混合策略，而非这些策略的实现结果。这种情况对应于完美监控。

民间定理指出以下事实：与单阶段博弈不同，在单阶段博弈中，理性智能体只有在某个收益组合对应于纳什均衡时才能实现该收益组合；而在重复博弈中，任何高于某个阈值的可行收益组合 $\boldsymbol{v} = (v^1,\ \ldots\ldots,\ v^N) \in [0,1]^N$都可以维持为（子博弈完美）纳什均衡（SPNE）。

构造实现 $\boldsymbol{v}$ 的子博弈完美纳什均衡（SPNE）的标准方法，是采用 冷酷触发策略（Fudenberg 和 Maskin，1986年）。总的来说，参与者同意坚持策略组合 $\mathbf{w}_{\boldsymbol{v}} = (w_{\boldsymbol{v}}^1,\ \ldots\ldots,\ w_{\boldsymbol{v}}^N) \in \Delta(\mathcal{A})^N$，以实现 $\boldsymbol{v}$，即 $(u^1(\mathbf{w}_{\boldsymbol{v}})\ldots\ldots,\ u^N(\mathbf{w}_{\boldsymbol{v}})) = \boldsymbol{v}$。如果其中一人偏离策略，其对手就会永远改用惩罚行动（通常是纳什均衡行动）。如果参与者足够重视

未来，也就是说，β 足够接近 1，那么惩罚的威胁就能阻止参与者偏离 $\mathbf{w}_{\boldsymbol{v}}$。这确保 $\mathbf{w}_{\boldsymbol{v}}$ 是一个能诱导出收益 $\boldsymbol{v}$ 的子博弈完美纳什均衡（SPNE）。

我们在下文正式引入冷酷触发策略和民间定理。固定某个参考混合策略纳什均衡$\mathbf{b} = (b^1, \ldots b^N) \in \Delta(\mathcal{A})$（由纳什定理可知其存在），并记 $u^i = u^i(\mathbf{b})$为玩家 i的相应收益。对任意历史$\mathrm{h}_t \in \mathcal{H}$，冷酷触发策略组合$\mathbf{s}_{\boldsymbol{v}} = (s^1_{\boldsymbol{v}}, \ldots, s^N_{\boldsymbol{v}}) \in \mathcal{S}$对任意 $i \in [N]$由下式给出

$$s^i_{\boldsymbol{v}}(\mathrm{h}_t) = \begin{cases} w^i_{\boldsymbol{v}} & \text{if } \mathrm{h}_t = (\mathbf{w}_{\boldsymbol{v}}, \ldots, \mathbf{w}_{\boldsymbol{v}})\,, \\ b^i & \text{otherwise}\,. \end{cases} \tag{3}$$

下文中，我们用 $\bar{u}_i = \max_{\mathbf{w}} u^i(\mathbf{w})$表示玩家 i的最大收益，并用$\mathcal{U}^i = \mathrm{conv}(u^i(\mathcal{A}))$表示玩家 i在混合策略下的可行即时收益集合。

定理1. 假设完美监控。固定一个收益组合 $\boldsymbol{v} = (v^1, \ldots, v^N) \in U^1 \times \ldots \times U^N$，使得 $v^i \geqslant u^i$ 对任意 $i \in [N]$成立。若对任意玩家 i有 $\beta \geqslant (\bar{u}^i - v^i)(\bar{u}^i - u^i)^{-1}$，则按$(3)$定义的$\mathbf{s}_{\boldsymbol{v}}$ 满足：*(i)* 对任意 $i \in [N]$有 $\bar{U}^i(\mathbf{s}_{\boldsymbol{v}}) = v^i$；*(ii)* $\mathbf{s}_{\boldsymbol{v}}$是一个子博弈完美纳什均衡。

定理1的证明是众所周知的（Fudenberg 和 Maskin，1986年），为完整起见，见附录B。该定理指出，只要参与者足够有耐心（由折扣因子 β 的下界体现），冷酷触发策略（3）就能使约定的效用剖面 $\boldsymbol{v}$ 达成子博弈完美纳什均衡。

请注意，在定理1中，惩罚对应于纳什均衡，这与文献中的标准做法一致（Fudenberg 和 Maskin，1986年）。通过考虑能够带来极小极大收益的惩罚，可以加强这一结果，但这需要更复杂的分析——尤其是必须激励惩罚者自身继续实施惩罚（Fudenberg 和 Maskin，1986年）。为使表述清晰简洁，本文因此仅讨论纳什均衡惩罚。

3 通过任意时间检验实现不完美监控下的均衡

遗憾的是，上一节介绍的完美监控情形在现实生活中很少出现。因此，我们转而探讨：在参与者无法观察实际的混合策略、只能观察从这些策略中抽取的纯行动的不完美信息情形下，能否得到类似的结果。在这种情形下，无法采用严厉触发策略（3），因为参与者无法直接观察对手采用的混合策略。具体而言，我们不再能确定某个玩家 j 是否偏离了 $w^j_{\boldsymbol{v}}$（除非抽取到一个不属于$w^j_{\boldsymbol{v}}$ 支撑集的纯行动）。不过，我们将证明，可以通过统计检验以很高的概率判断是否发生了这种偏离。如果某个玩家收集到足够的统计证据来支持这一假设，就可以像（3）中那样触发惩罚。这就是我们将在下文正式介绍的先检验后惩罚策略的基本思想。

假设参与者就可行收益 $\boldsymbol{v}=(v^1,\dots,v^N)\in\mathcal{U}^1\times\dots\times\mathcal{U}^N$以及混合策略组合$\mathbf{w}_{\boldsymbol{v}}=(w^1_{\boldsymbol{v}},\dots,w^N_{\boldsymbol{v}})\in\Delta(\mathcal{A})$达成一致，使得对于任意 $i\in[N]$，$v^i=u^i(\mathbf{w}_{\boldsymbol{v}})$。接着，我们定义$\boldsymbol{\sigma}_{\boldsymbol{v}}=(\sigma^1_{\boldsymbol{v}},\dots,\sigma^N_{\boldsymbol{v}})\in\mathcal{S}$，其中对于任意 $j\in[N]$，$\sigma^j_{\boldsymbol{v}}(\mathbf{h}_t)=w^j_{\boldsymbol{v}}$对所有 $t\geqslant 0$ 以及 $\mathbf{h}_t\in\mathcal{H}_t$均成立。然后，每个玩家$j$ 都可以按如下方式对 $i\neq j$进行检验

$$\mathrm{H}^i_0\colon\ s^i=\sigma^i_{\boldsymbol{v}}\qquad\text{versus}\qquad\mathrm{H}^i_1\colon\ s^i\neq\sigma^i_{\boldsymbol{v}}\,. \tag{4}$$

H^i_0 为真意味着玩家 i 坚持采用合作策略 $w^i_{\boldsymbol{v}}$。相反，拒绝 H^i_0 意味着得出结论：玩家 i 没有采用约定策略。

为了判断 H^i_0 是否为真，玩家 $j\neq i$ 可以依赖检验序列 $\boldsymbol{\psi}^i=\{\psi^i_t\}_{t\geqslant 0}$，其中$\psi^i_0=0$，且对于任意 $t\geqslant 1,\psi^i_t\colon\ \mathcal{H}_t\to\{0,1\}$。如果 $\psi^i_t=1$，则 H^i_0 在时间 t被拒绝。这里，我们假设所有参与者 $j\neq i$ 都对 H^i_0 使用相同的检验（因此 ψ^i_t中没有 j 索引），因为他们在时间 t都能访问共同的公共历史 $\mathbf{h}_t\in\mathcal{H}_t$。于是，我们可以定义先检验后惩罚 策略 $\mathbf{s}_{\boldsymbol{v}}\in\mathcal{S}$，其中对于 $j\in[N]$，$s^j_{\boldsymbol{v}}$ 由下式给出

$$s^j_{\boldsymbol{v}}(\mathrm{h}_t)=\begin{cases}w^j_{\boldsymbol{v}} & \text{if }\ \prod_{s=0}^{t}\prod_{i=1}^{N}(1-\psi^i_s(\mathrm{h}_s))=1\,,\\ b^j & \text{otherwise}\,.\end{cases} \tag{5}$$

非正式地说，在时间 $t\geqslant 0$，如果参与者尚未检测到任何偏离，即对于所有 $i\in[N]$，$\psi^i_t=0$，他们就继续采用合作混合策略组合 $w_{\boldsymbol{v}}$。一旦检测到偏离，也就是说，对于某个 i，$\psi^i_t=1$，所有参与者都会在接下来的轮次中永久切换到惩罚策略。

为简化记号，对于任意 $i\in[N]$，我们将与检验序列 $\boldsymbol{\psi}^i$ 相关联的（随机）拒绝时间定义为

$$\tau_{\boldsymbol{\psi}^i}=\inf\{t\geqslant 0\colon\ \psi^i_t=1\}\in\mathbb{N}\cup\{\infty\}\,, \tag{6}$$

约定 $\inf\varnothing=\infty$。换言之，$\tau_{\boldsymbol{\psi}^i}$ 是 H^i_0 被拒绝的第一个时刻；也就是说，这是检测到玩家 i 发生偏离的第一个时刻。类似地，我们将$\tau_{\boldsymbol{\psi}}$ 定义为

$$\tau_{\boldsymbol{\psi}}=\min_{i\in[N]}\tau_{\boldsymbol{\psi}^i}\,,$$

这是任意原假设 $\mathrm{H}_{10},\dots,\mathrm{H}^N_0$ 被拒绝的第一个时刻，也就是说，这是检测到任意玩家发生偏离的时刻。从 (5) 可以清楚看出，对于任意 $t\leqslant\tau_{\boldsymbol{\psi}}$，参与者都采用$\mathbf{w}_{\boldsymbol{v}}$，而对于任意 $t>\tau_{\boldsymbol{\psi}}$，则转而采用惩罚$\mathbf{b}$。

请注意，先检验后惩罚策略 (5) 模仿了完美监控下的冷酷触发策略 (3)，但有一个显著区别：后者一旦观察到偏离就会触发惩罚（完美监控允许如此），而前者一旦某个原假设 H^i_0 被拒绝就会触发惩罚（因为监控并不完美）。如果检验是完美的——也就是说，当且仅当玩家 i 确实发生偏离时，H^i_0 才会被拒绝——那么这两种方法就等同了。然而，统计检验永远不可能完美。它们可能会出现第一类错误（原假设 H^i_0 为真时却拒绝它）和第二类错误（原假设 H^i_0 为假时却未能拒绝它），因为它们依赖随机观测。

在我们的情形中，这意味着先检验后惩罚策略 (5) 可能会在参与者实际上仍遵循$\mathbf{w}_{\boldsymbol{v}}$时错误地触发惩罚，这对应于第一类错误；也可能未能惩罚发生偏离的玩家，这对应于第二类错误。这两类错误都可能使先检验后惩罚程序无法实现其目标，即在（子博弈精炼）纳什均衡下确保目标收益组合 $\boldsymbol{v}$ 。

因此，自然会要求序贯检验族 $\boldsymbol{\psi}=(\boldsymbol{\psi}^1,\ \ldots\ldots,\ \boldsymbol{\psi}^N)$ 满足某些条件，以确保第一类和第二类错误保持较低水平，从而使 $\mathbf{s}_{\boldsymbol{v}}$ 有效。特别是针对第一类错误，我们提出以下要求。

条件1 （第一类错误概率较低）。 存在 $\gamma\in(0,\ 1)$，使得对任意 $i\in[N$]

$$\mathbb{E}^{\mathbf{s}_{\boldsymbol{v}}}[\mathbb{1}\{\tau_{\boldsymbol{\psi}^i}<\infty\}]=\mathbb{P}^{\mathbf{s}_{\boldsymbol{v}}}\left(\tau_{\boldsymbol{\psi}^i}<\infty\right)\leqslant\gamma/N\,. \tag{7}$$

如果条件1成立，那么对任意 $i\in[N]$，检验序列 $\boldsymbol{\psi}^i$ 都满足：在合作策略 $\mathbf{w}_{\boldsymbol{v}}$ 下，错误地拒绝 H_0^i 的概率保持较低。等价地说，在玩家 i 合作时，错误地检测到其偏离的概率较低。请注意，条件1立即限制了错误进入惩罚阶段的概率，因为根据并集界可得

$$\mathbb{P}^{\mathbf{s}_{\boldsymbol{v}}}(\tau_{\boldsymbol{\psi}}<\infty)=\mathbb{P}^{\mathbf{s}_{\boldsymbol{v}}}(\cup_{i\in[N]}\{\tau_{\boldsymbol{\psi}^i}<\infty\})\leqslant\sum_{i\in[N]}\mathbb{P}^{\mathbf{s}_{\boldsymbol{v}}}(\tau_{\boldsymbol{\psi}^i}<\infty)\leqslant\gamma\,.$$

这确保了合作能够以较高概率长期维持。请注意，条件1会随着参与者数量的增加而变得更加严格。事实上，参与者越多，假设检验就越多，这会增加误报的概率，因此需要更严格地控制每项检验。控制假发现率（Benjamini 和 Hochberg，1995年），尤其是通过任意时刻扩展（Wang 和 Ramdas，2022年），或许能以一种不那么保守的方式处理玩家面对多个对手时不断增加的假设数量，同时仍能限制误报。

关于第二类错误，序贯检验族 $\boldsymbol{\psi}=(\boldsymbol{\psi}^1,\ \ldots\ldots,\ \boldsymbol{\psi}^N)$ 应满足：若玩家发生偏离，应能以较高概率在有限时间内检测到。请注意，要检测与协作策略 $\mathbf{w}_{\boldsymbol{v}}$ 任意接近的偏离，极其困难，甚至可能毫无希望。因此，我们将关注与 $\mathbf{w}_{\boldsymbol{v}}$“相距足够远”的偏离。为将这一想法形式化，对于任意 $\varepsilon>0$ 和 $i\in[N]$，我们定义以 $w^i_{\boldsymbol{v}}$ 为中心、半径为 ε 的全变差球

$$\mathrm{B}^i(\mathbf{w}_{\boldsymbol{v}}\,,\varepsilon)=\{s\in\mathcal{S}^i\colon\ \|s(\mathrm{h}_t)-w^i_{\boldsymbol{v}}\|_1\leqslant 2\varepsilon\text{ for any }t\geqslant 0\text{ and }\mathrm{h}_t\in\mathcal{H}_t\}\,.$$

注意，对于任意 $\mathfrak{S}^i\subseteq\mathcal{S}^i$，$\mathfrak{S}^i\setminus\mathrm{B}^i(\mathbf{w}_{\boldsymbol{v}},\ \varepsilon)$，其中包含玩家 i 在 $\mathfrak{S}^i$ 中所有显著性达到 ε 的偏离，也就是说，这些偏离与合作策略 $w^i_{\boldsymbol{v}}$ 之间的总变差距离至少为 ε 。

图1展示了这种情形。于是，对于任意 $\mathfrak{S}=\mathfrak{S}^1\times\ldots\times\mathfrak{S}^N\subseteq\mathcal{S}$，我们定义包含至少一个显著性达到 ε 的偏离的策略组合集合

$$\mathfrak{S}(\mathbf{w}_{\boldsymbol{v}},\varepsilon)=\left\{\mathbf{s}=(s^1,\ldots,s^N)\in\mathfrak{S},\text{ there exists } i\in[N]\colon\ s^i\in\mathfrak{S}^i\setminus\mathsf{B}^i(\mathbf{w}_{\boldsymbol{v}},\varepsilon)\right\}. \tag{8}$$

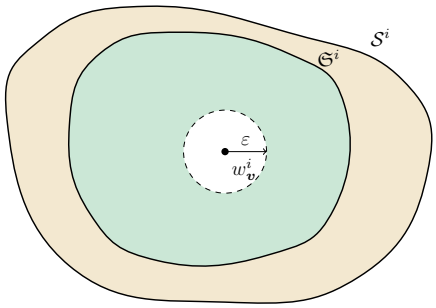

图1：如果某个玩家 $i\in[N]$ 在绿色子集中选择策略 s^i，那么 $\mathbf{s}=(s^i,s^{-i})\in\mathfrak{S}(\mathbf{w}_{\boldsymbol{v}},\varepsilon)$。

现在我们可以陈述第二个要求。直观地说，以下条件表明，任何包含 ε-显著偏离的策略组合 $\mathbf{s}\in\mathfrak{S}$ 都会在有限时间内触发惩罚。

条件 2。对于 $\varepsilon>0$，存在一个序列 $(\zeta_t^{(\varepsilon)})_{t\geqslant0}$，使得

$$\text{for any } t\geqslant0,\quad\sup_{\mathbf{s}\in\mathfrak{S}(\mathbf{w}_{\boldsymbol{v}},\varepsilon)}\mathbb{P}^{\mathbf{s}}(\tau_{\boldsymbol{\psi}}\geqslant t)\leqslant\zeta_t^{(\varepsilon)}\quad\text{and}\quad\sum_{t=0}^{\infty}\zeta_t^{(\varepsilon)}<\infty. \tag{9}$$

条件 2 确保，对于任意策略组合 $\mathbf{s}\in\mathfrak{S}(\mathbf{w}_{\boldsymbol{v}},\varepsilon)$（其中至少有一名玩家偏离 $w^i_{\boldsymbol{v}}$ 至少 ε），惩罚时间 $\tau_{\boldsymbol{\psi}}$ 以高概率不超过 t。尤其是，条件 2 意味着，当其中一名玩家发生 ε-显著偏离时，参与者会在有限的期望时间内进入惩罚阶段，因为

$$\sup_{\mathbf{s}\in\mathfrak{S}(\mathbf{w}_{\boldsymbol{v}},\varepsilon)}\mathbb{E}^{\mathbf{s}}[\tau_{\boldsymbol{\psi}}]=\sup_{\mathbf{s}\in\mathfrak{S}(\mathbf{w}_{\boldsymbol{v}},\varepsilon)}\sum_{t\geqslant0}\mathbb{P}^{\mathbf{s}}(\tau_{\boldsymbol{\psi}}\geqslant t)\leqslant\sum_{t\geqslant0}\sup_{\mathbf{s}\in\mathfrak{S}(\mathbf{w}_{\boldsymbol{v}},\varepsilon)}\mathbb{P}^{\mathbf{s}}(\tau_{\boldsymbol{\psi}}\geqslant t)\leqslant\sum_{t=0}^{\infty}\zeta_t^{(\varepsilon)}<\infty.$$

我们指出，条件 (2) 将 $\mathsf{B}^i(\mathbf{w}_{\boldsymbol{v}},\varepsilon)$ 中的“小幅”偏离搁置一旁：此类偏离不必在有限时间内触发惩罚。事实上，这并不值得担忧，因为此类偏离无法带来显著的效用增益，因此已由我们的近似均衡概念涵盖。以下命题将对此作出形式化论证。

命题1. 固定 $\varepsilon>0$。对任意 $i\in[N]$，若 $s^i\in\mathsf{B}^i(\mathbf{w}_{\boldsymbol{v}},\varepsilon)$，则 $U^i(s^i,s^{-i}_{\boldsymbol{v}})\leqslant v^i+\varepsilon$.

总之，条件 2 与命题1共同确保，玩家 i 若偏离到某个$s^i\neq s^i_{\boldsymbol{v}}$，要么会在有限时间内触发惩罚（若 $s^i\in\mathfrak{S}^i\setminus\mathsf{B}^i(\mathbf{w}_{\boldsymbol{v}},\varepsilon)$，依据条件 2），要么无法获得很大的效用增益（若 $s^i\in\mathfrak{S}^i\cap\mathsf{B}^i(\mathbf{w}_{\boldsymbol{v}},\varepsilon)$，依据命题1）。这一观察使我们能够在条件1和条件 2成立的前提下，基于先检验后惩罚策略 (5) 建立一个民间定理。

定理2. 设 $\mathfrak{S} = \mathfrak{S}^1 \times \ldots \times \mathfrak{S}^N \subseteq S$，并考虑一族满足条件*1*和条件 *2*的序贯检验。设 $\boldsymbol{v} \in U^1 \times \ldots \times U^N$是一个可行收益，且对于任意 $i \in [N]$

$$v^i \geqslant \underline{u}^i = u^i(\mathbf{b}) \,,$$

其中 $\mathbf{b} = (b^1, \ldots b^N) \in \Delta(\mathcal{A})$ 是某个参考混合策略纳什均衡，该均衡由纳什定理保证存在。记 $\tau(\varepsilon) =$上确界 $\mathbb{E}^{\mathbf{s}}[\tau_{\boldsymbol{\psi}}] < \infty$，并且 $\bar{u}^i =$最大值$_{a \in \mathcal{A}} u^i(a)$，若对任意 $i \in [N]_{\mathbf{s} \in \mathfrak{S}(\mathbf{w}_{\boldsymbol{v}}, \varepsilon)}$

$$\bar{\beta}^i = \frac{\bar{u}^i - (1-\gamma)v^i}{\bar{u}^i - \underline{u}^i} \in [0,1] \quad and \quad \beta^{\tau(\varepsilon)} \geqslant \bar{\beta}^i \,, \tag{10}$$

那么$\mathbf{s}_{\boldsymbol{v}}$ 在(5)中定义的量满足

(i) for any $i \in [N]$, $(1-\gamma)v^i \leqslant U^i(\mathbf{s}_{\boldsymbol{v}}) \leqslant v^i$,

(ii) $\mathbf{s}_{\boldsymbol{v}}$ *is a* $(\varepsilon + \gamma, \mathfrak{S})$*-NE.*

将定理2与其完全监控对应版本定理1进行比较很有启发性。在后者中，$U^i(\mathbf{s}_{\boldsymbol{v}})$ 恰好 等于目标收益 v^i。在前者中，它位于以 v^i 为中心、长度为 γ 的区间内。这一区别与第一类错误有关：即使所有参与者都坚持采用协作策略 $w_{\boldsymbol{v}}$，他们仍承担着以至多 γ 的概率错误进入惩罚阶段的风险。请注意，更好的检验会降低错误率 γ，从而缩小效用差距。

此外，注意到在定理1中，$\mathbf{s}_{\boldsymbol{v}}$ 是一个精确的子博弈完美纳什均衡。相反，在定理2中，它是一个 $(\varepsilon + \gamma, \mathfrak{S})$-纳什均衡。 $\varepsilon + \gamma$ 近似项源于第一类错误，以及参与者可以选择较小的偏离而不被检测到的可能性，正如命题1中所讨论的那样。此外，定理2仅表明$\mathbf{s}_{\boldsymbol{v}}$ 是纳什均衡，并未表明它是子博弈精炼的。正如我们将在第4节中看到的，通过采用不同的检验策略，可以实现子博弈完美，但代价是无法控制第一类和第二类错误。最后，注意在完美监控下（定理1），耐心条件直接取决于折扣因子 β，因为偏离会立即被检测到。相比之下，在定理2的先检验后惩罚策略下，偏离可能长达 $\tau^{(\varepsilon)}$ 期未被检测到，因此相关条件涉及 $\beta^{\tau^{(\varepsilon)}}$。因此，检测偏离所需的时间越长，对耐心的要求就越高。更有力的检验能够降低$\tau^{(\varepsilon)}$，从而放宽这一要求。

3.1 基于插件e-过程的任意时间检验

为了使定理2完全具有可操作性，我们仍需明确一种方法，使参与者能够设计出满足条件1和条件 2 的序贯检验族 $\boldsymbol{\psi}$ 。这可以通过利用任意时刻检验 和 e过程（Ramdas和Wang，2025）来实现。我们在附录A中简要回顾这一框架，而这里只介绍该框架的主要思想和概念。

给定一个原假设H_0，任意时刻有效检验建立在*e*过程 这一概念之上；对于属于 H_0 的任意概率分布，该过程通常是一个上鞅（M_t）$_{t\geqslant 0}$ 。在这一设定下，Ville 不等式（Ville，1939，见附录A）表明，（M_t）$_{t\geqslant 0}$ 对任意 $\mathbb{P}_0 \in H_0$,均成立

$$\mathbb{P}_0\left(\sup_{t\geqslant 0} M_t \geqslant 1/\gamma\right) \leqslant \gamma \qquad \text{for any } \gamma \in (0,1)\,. \tag{11}$$

换言之，在原假设下，事件 $\cup_{t\geqslant 0}\{M_t \geqslant 1/\gamma\}$ 发生的概率不可能大于 γ。因此，用于以 γ 水平检验 H_0 的一个自然检验统计量为 $\mathbb{1}\{M_t \geqslant 1/\gamma\}$。

考虑 $(\mathscr{H}_t)_{t\geqslant 0}$，与 $(h_t)_{t\geqslant 0}$, $\mathscr{H}_t = \sigma(h_s:\ s\leqslant t)$ 相关的自然滤过，我们现在说明玩家 $j\in[N]$ 如何为对手 $i\in[N]$ 设计一个 $(h_t)_{t\geqslant 0}$-上鞅 $\{E_t^i\}_{t\geqslant 0}$，以检验 H_0^i。由于在任意一轮 t，纯行动 $A_t^1,\ldots,A_t^N$ 都可观察到，参与者可以统计截至第 $t-1$ 步，每位玩家 $i\in[N]$ 采取行动 $a\in\mathcal{A}^i$ 的次数，记作 $N_t^i(a) = \sum_{s=0}^{t-1}\mathbb{1}\{A_s^i = a\}$。于是，可以将玩家 i 采取行动的（平滑）经验频率定义为

$$\hat{w}_t^i(a) = \frac{N_t^i(a)+1}{t+K} \qquad \text{for any } a\in\mathcal{A}\,, \tag{12}$$

以及 $\hat{w}_{-1}^i(a) = 1/K$。接着，我们可以构造 Ramdas 和 Wang (2025) 的插件e-过程，以如下方式检验玩家 i 的 H_0^i。令 $E_{-1}^i = 1$，并对任意 $t\geqslant 0$

$$E_t^i = \prod_{s=0}^{t}\frac{\hat{w}_s^i(A_s^i)}{w_{\boldsymbol{v}}^i(A_s^i)}\,. \tag{13}$$

我们在附录A的定理 6 中正式证明了 $(E_t^i)_{t\geqslant 0}$ 是一个 $(h_t)_{t\geqslant 0}$-上鞅，因此在 H_0^i 下是一个 e过程。直观地说，如果玩家 i 从 $w_{\boldsymbol{v}}^i$ 中抽取一个纯行动，那么当 s 足够大且任意 $a\in\mathcal{A}^i$ 时，$\hat{}w_s^i(a)/w_{\boldsymbol{v}}^i(a) \approx 1$，因此 E_t^i 随时间推移的平均增长不会太快。相反，如果玩家 i 偏离，那么 E_t^i 理应增加，并在首次超过 $1/\gamma$ 时触发检验。综合以上各项，我们便得到一个实用的检验程序，供参与者检验 H_0^i：他们可以跟踪 i 的行动经验频率 (12)，以构建 e过程 (13)，并检查它是否超过给定阈值。对于 $\gamma\in(0,\ 1)$，这便得到序贯检验 $\boldsymbol{\psi}^i = \{\psi_t^i\}_{t\geqslant 0}$，用于检验 H_0^i，其对任意 $t\geqslant 0$ 定义如下

$$\tilde{\psi}_t^i = \mathbb{1}\Big\{E_t^i \geqslant N/\gamma\Big\}\,. \tag{14}$$

根据 Ville 不等式 (11)，我们可以通过调整阈值来明确控制该检验的显著性水平。这立即意味着此类检验满足条件1，正式表述如下命题。

命题2。 设 $\gamma\in(0,\ 1)$，并对任意 $i\in[N]$，考虑在(14)中定义的检验序列 $(\tilde{\psi}_t^i)_{t\geqslant 1}$，以及在(5)中定义的相应策略$s_{\boldsymbol{v}}$ 。那么，条件*1*得到满足。

[2] 1 和 K 起到平滑作用，就像拉普拉斯估计量中的平滑一样。

现在我们将在一个简化假设下证明，上述检验序列也满足条件 2。在继续之前，我们先指出一个根本性的困难。对于 $\mathfrak{S} = \mathfrak{S}^1 \times \ldots \times \mathfrak{S}^N \subseteq \mathcal{S}$，条件 2 所要求的第二类错误控制必须在 (8) 中定义的集合 $\mathfrak{S}(\mathbf{w}_{\boldsymbol{v}}, \varepsilon)$ 上一致成立。也就是说，任何满足以下条件的策略 $s^i \in \mathfrak{S}^i$： $s^i(\mathrm{h}_t)$ 在某个时刻 $t \geqslant 0$ 偏离至少 ε，都应以较高概率最终触发惩罚。从统计学视角看，这相当于检验简单原假设 H_0^i，以对抗任意大的复合备择假设 H_1^i。众所周知，若不对备择假设施加额外的结构性限制，一致最强检验通常并不存在，而且在许多情况下，非平凡的第二类错误控制在信息论意义上是不可能实现的（Lehmann 和 Romano，2005年）。因此，我们通过将注意力限制在以下一组简单偏离上来简化问题

$$\tilde{\mathfrak{S}}^i = \left\{ s \in \mathcal{S}^i : \text{there exists } \tilde{w}^i \in \Delta(\mathcal{A}^i) \text{ such that for any } t \geqslant 0,\ \mathrm{h}_t \in \mathcal{H}_t,\ s(\mathrm{h}_t) = \tilde{w}^i \right\}. \tag{15}$$

简而言之，$\tilde{\mathfrak{S}} = \tilde{\mathfrak{S}}^1 \times \ldots \times \tilde{\mathfrak{S}}^N$包含平稳策略，这些策略始终采用单一混合策略 $\tilde{w}^i \in \Delta(\mathcal{A})$。这简化了统计分析，因为它确保在$\mathrm{H}_1^i$下，任何玩家$i \in [N]$ 在被发现之前都会从 $\tilde{w}^i$ 中抽取独立同分布的行动。然而，我们强调该问题仍然困难，因为参与者并不知道玩家 i选择了何种混合策略 $\tilde{w}^i$ 。换言之，H_1^i 仍然是一个复合备择假设。我们在以下命题中表明，当策略位于 $\mathfrak{S}(\mathbf{w}_{\boldsymbol{v}}, \varepsilon)$中时（其定义见方程(8)），条件2˜得到满足。

我们现在有s 统计杠杆，从而得到期望停止时间 $\tilde{\tau}^i$ 的一个界，其定义为 as

$$\tilde{\tau}_{\psi}^i = \inf\{t \geqslant 1 \colon \tilde{\psi}_t^i = 1\}, \tag{16}$$

其中$(\tilde{\psi}_t^i)_{t \geqslant 0}$ 在(14)中针对误差水平 $\gamma \in (0,1)$加以定义。

定理3。 考虑 $\varepsilon > 0$ 以及一个水平为 $\gamma \in (0,1)$的情形。考虑(14)中定义的序贯检验 $\{\tilde{\psi}^i\}_{i \in [N]}$。那么，对于 $\tilde{\mathfrak{S}} = \tilde{\mathfrak{S}}^1 \times \ldots \times \tilde{\mathfrak{S}}^N$，其中 $\tilde{\mathfrak{S}}^i$定义于(15)，且 $\tilde{\tau}_{\boldsymbol{\psi}} =$最小值$_{i \in [N]}\, \tilde{\tau}_{\psi}^i$，条件2成立，且

$$\zeta_t^{(\varepsilon)} = C_1 \left\{ t \exp\left(-\frac{2t\varepsilon^2}{C_2 \log(t)^2} \right) + \exp\left(-t \frac{4\varepsilon^4}{C_2 \log(t)} \right) + \exp\left(-\frac{t\varepsilon^4}{2|\log \underline{\mathbf{w}_{\boldsymbol{v}}}|} \right) \right\} + \mathbb{1}\{t \leqslant 10 \log(1/\gamma)/\varepsilon^2\},$$

对于任意 $t \geqslant 2$、$\zeta_0^{(\varepsilon)} = 1$、$\zeta_1^{(\varepsilon)} = 1$，其中 C_1、$C_2 > 0$是普适常数，并令$\mathbf{w}_{\boldsymbol{v}}$ =最小值$_{i \in [N]}$最小值$_{a \in \mathcal{A}^i}\, w_{\boldsymbol{v}}^i[a]$。特别地，我们有 $\sum_{t \geqslant 0} \zeta_t^{(\varepsilon)} < \infty$。

˜**推论 1.** 在与**定理3**相同的设定下，对于任意 $\mathbf{s} \in \mathfrak{S}(\mathbf{w}_{\boldsymbol{v}}, \varepsilon)$，存在一个普适常数 $C > 0$ 使得

$$\mathbb{E}^{\mathbf{s}}[\tilde{\tau}_{\boldsymbol{\psi}}] \leqslant \tilde{\tau}_{\varepsilon} = \frac{10 \log(1/\gamma)}{\varepsilon^2} + C \frac{1 + |\log \underline{\mathbf{w}_{\boldsymbol{v}}}|}{\varepsilon^5}.$$

随着 γ 减小（即检验变得更加保守），$\mathbb{E}^{\mathbf{s}}[\tilde{\tau}_{\boldsymbol{\psi}}]$的上界增大，这体现了控制第一类错误与检验效能之间的经典权衡。根据命题2和定理*3*，式（14）中定义的序贯检验满足条件1和条件 2。因此，可以将定理*2*应用于检验族$\{(\tilde{\psi}^i)_{t\geqslant 1}\}_{i\in[N]}$，从而得到以下推论。

推论 2。考虑与定理*3*相同的设定，其中包含 $\gamma=\varepsilon$，以及由(5)定义、检验由(14)指定的策略$\mathbf{s}_{\boldsymbol{v}}$。此外，设 $\boldsymbol{v}\in U^1\times\ldots\times U^N$是一个可行收益剖面，满足 $v^i\geqslant u^i$，其中 $i\in[N]$。若对于任意 $i\in[N]$

$$\bar{\beta}^i=\frac{\bar{u}^i-(1-\varepsilon)v^i}{\bar{u}^i-\underline{u}^i}\in[0,1]\quad and\quad \beta^{\tilde{\tau}_\varepsilon}\geqslant\bar{\beta}^i\,, \tag{17}$$

其中 $\tilde{\tau}_\varepsilon$ 由推论 *1* 定义。则有

(i) 对任意 $i\in[N]$，$(1-\varepsilon)v^i\leqslant U^i(\mathbf{s}_{\boldsymbol{v}})\leqslant v^i$，

(ii) $\mathbf{s}_{\boldsymbol{v}}$ 是一个$(2\varepsilon,\mathfrak{S})$-纳什均衡，其中 $\mathfrak{S}$ 的定义与(15)中相同。

注意，在推论 2 中，取 $\varepsilon\to 0$ 可以使我们几乎重现定理1中的完美监控结果。换言之，当参与者使用我们的先检验后惩罚策略，并采用基于 e过程构造的序贯检验时，他们能够应对不完美监控，并维持合作，从而获得任意接近任一可行理性收益剖面 $\boldsymbol{v}$ 的收益。不过需要注意的是，由于 $\varepsilon\to 0$，耐心条件 (17) 要求 $\beta\to 1$。

4 不完美监控下通过批次检验实现均衡

在上一节中，我们看到，任意时刻检验可用于实施先检验后惩罚策略，并且能够保证第一类错误率任意低。然而，该方法有两个局限。首先，我们的分析仅限于平稳偏离，即属于 (15) 中定义的 $\mathfrak{S}$，以便获得条件 2 中给出的第二类错误保证。尽管从实践角度来看这一假设似乎合理，但也可以认为它具有限制性，因为它排除了参与者采用复杂自适应策略的对抗性行为。其次，定理2中的 $\boldsymbol{v}$ 只是一个纳什均衡，而我们的目标是获得一个子博弈精炼纳什均衡。

因此，我们要探究能否通过考虑偏离集合 $\mathfrak{S}=\mathcal{S}$ 中的所有偏离，以完全一般的方式实施先检验后惩罚策略，并确保它构成一个子博弈完美均衡。正如我们将要展示的，这一更具雄心的目标确实可以通过改用另一种检验策略来实现。然而，为此付出的代价是失去对第一类错误的显式控制。

概括而言，我们提出采用一种批次 检验程序，具体如下：首先，将轮次按预定长度分成批次；在每个批次中，参与者根据观察到的纯行动，检查是否有任何参与者 $i\in[N]$ 在该批次中的平均行动为混合行动 $w^i_{\boldsymbol{v}}$

；如果是，所有参与者都会继续各自行动 $\mathbf{w}_{\boldsymbol{v}}$ ；否则，他们会在所有后续批次中转而采取惩罚 $\boldsymbol{b}\in\Delta(\mathcal{A})$。

形式化地，假设参与者就可行收益 $\boldsymbol{v}\in\mathcal{U}^1\times\ldots\times\mathcal{U}^N$和混合策略组合 $\mathbf{w}_{\boldsymbol{v}}=(w_{\boldsymbol{v}}^1\ldots,w_{\boldsymbol{v}}^N)\in\Delta(\mathcal{A})$ 达成一致，使得 $v^i=u^i(\mathbf{w}_{\boldsymbol{v}})$ 对于 $i\in[N]$成立。固定批次长度 $L>0$，并对任意 $k\geqslant 0$ 定义批次 $\mathcal{B}_k$ 如下：

$$\mathcal{B}_k=\{Lk,\ldots,L(k+1)-1\}\,,$$

因此 $\mathbb{N}=\cup_{k\geqslant 0}\mathcal{B}_k$。在每个批次 $\mathcal{B}_k$结束时，参与者通过统计检验检查玩家 i在轮次 $t\in\mathcal{B}_k$中的行动是否为从 $w_{\boldsymbol{v}}^i$ 独立同分布抽取的。形式化地，他们对任意$k\geqslant 0$进行检验

$$\mathrm{H}_{0,k}^i\colon\quad (A_t^i)_{t\in\mathcal{B}_k}\overset{\text{i.i.d}}{\sim}w_{\boldsymbol{v}}^i\qquad\text{versus}\qquad\mathrm{H}_{1,k}^i\colon\quad (A_t^i)_{t\in\mathcal{B}_k}\overset{\text{i.i.d}}{\not\sim}w_{\boldsymbol{v}}^i\,.\tag{18}$$

对于 $k\geqslant 0$，记 $\mathbf{b}_k=\{(A_l^i)_{t\in\mathcal{B}_k}\colon\ i\in[N]\}\in\mathcal{A}^L$。那么，对假设 (18) 的检验是一个映射$\phi_k^i\colon\mathcal{A}^L\to\{0,1\}$，其中 $\phi_k^i=1$ 表示 $\mathrm{H}_{0,k}^i$ 在批次 k被拒绝。

现在我们可以介绍批量测试后惩罚策略了。对于任意 $t\geqslant 0$，我们将 $k_t=\lfloor t/L\rfloor$ 写作 t 所属批次的索引；这尤其意味着 $\cup_{k<k_t}\mathbf{b}_k\subset\mathbf{h}_t$。接着，我们将批量测试后惩罚策略定义为 $\bar{\mathbf{s}}_{\boldsymbol{v}}=(\bar{s}_{\boldsymbol{v}}^1,\ldots,\bar{s}_{\boldsymbol{v}}^N)\in\mathcal{S}$，如下所示。对于任意 $j\in[N]$ 和 $\mathbf{h}_t\in\mathcal{H}_t$，$\bar{s}_{\boldsymbol{v}}(\mathbf{h}_t)=w_{\boldsymbol{v}}^j$ 若 $k_t=0$，且对于任意 $k_t\geqslant 1$

$$\bar{s}_{\boldsymbol{v}}^j(\mathrm{h}_t)=\begin{cases}w_{\boldsymbol{v}}^j & \text{if}\quad\prod_{k=1}^{k_t-1}\prod_{i=1}^{N}\left(1-\phi_k^i(\mathrm{b}_k)\right)=1\,,\\ b^j & \text{otherwise}\,.\end{cases}\tag{19}$$

批次先检验后惩罚策略的执行方式与其任意时间对应策略相同。在批次$k\geqslant 0$，如果尚未检测到偏离——即对于任意 $i\in[N]$ 和任意步骤 $l\in\{1,\ldots,k-1\}$，$\phi_l^i=0$——参与者就继续采用合作策略 $\mathbf{w}_{\boldsymbol{v}}$。然而，一旦检测到偏离，也就是对于某个 i 和 k，$\phi_k^i=1$，他们便通过采用 $\boldsymbol{b}$ 永久转入惩罚阶段。与之前一样，我们可以为检验序列族 $\boldsymbol{\phi}=(\boldsymbol{\phi}^1,\ldots,\boldsymbol{\phi}^N)$ 定义相应的拒绝时刻，如下所示

$$\text{for any } i\in[N],\quad\kappa_{\boldsymbol{\phi}^i}=\inf\{k\geqslant 0\colon\phi_k^i=1\}\qquad\text{and}\qquad\kappa_{\boldsymbol{\phi}}=\min_{i\in[N]}\kappa_{\boldsymbol{\phi}^i}\,.\tag{20}$$

在式（20）中，$\kappa_{\boldsymbol{\phi}}$ 是第一个全局迭代 k，此时 $\mathrm{H}_{10,k}\cdots\mathrm{H}_0^N{}_{,k}$ 中的一个假设被拒绝。换言之，这就是惩罚阶段开始后的那个批次。图2展示了批次先测试后惩罚策略。偏离在批次 $\kappa_{\boldsymbol{\phi}}$结束时被检测到，随后参与者从批次 $\kappa_{\boldsymbol{\phi}}+1$开始进入惩罚阶段。

与任意时刻情形一样，我们可以问，是否存在一些条件，能够支撑批次检验序列 $\boldsymbol{\phi}=(\boldsymbol{\phi}^1、\ldots\ldots、\boldsymbol{\phi}^N)$ 的设计，使得 (19) 有效。我们提出了两个这样的条件。虽然它们在理念上类似于为任意时刻检验提出的条件（条件1和条件 2），但要求略低，也更容易针对任意偏离得到满足。

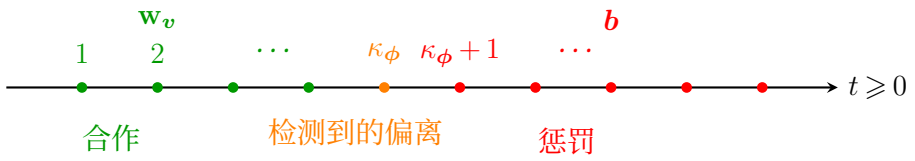

图2: 批次先测试后惩罚策略的时间线。圆点表示批次的结束 es.

首先，对于大多数延续博弈，在某个给定批次上错误检出偏离的概率应当很小。回顾一下，对于任意 $t \geqslant 0$，我们将 $k_t = \lfloor t/L \rfloor$ 定义为 t 所属的批次。

条件 3。 存在 $p_L \in (0,1)$ 和 $q_L \in (0,1)$，使得对于任意 $t \geqslant 0$，都存在一个历史集合 $\mathrm{H}_t \subset \mathcal{H}_t$，其中 $\mathbb{P}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}(\mathrm{h}_t \in \mathrm{H}_t) \geqslant 1 - q_L$ 满足：对于任意 $\mathrm{h}_t \in \mathrm{H}_t$

$$\mathbb{P}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}(\kappa_{\boldsymbol{\phi}} \geqslant k+1 \,|\, \kappa_{\boldsymbol{\phi}} \geqslant k\,, \mathrm{h}_t) \geqslant 1 - p_L \qquad \text{for any} \quad k \geqslant k_t\,.$$

非正式地说，条件 3 确保在任意时刻，如果从在 $\bar{\mathbf{s}}_{\boldsymbol{v}}$ 下很可能出现的某段历史开始（概率至少为 $1-q_L$），且玩家仍按照合作策略 $\mathbf{w}_{\boldsymbol{v}}$ 行动（给定 $\kappa_{\boldsymbol{\phi}} \geqslant k$），那么事件 $\{\kappa_{\boldsymbol{\phi}} = k\}$（对应于错误触发惩罚阶段）不太可能发生（概率至多为 p_L）。通常应将概率 p_L 和 q_L 视为较小的值。请注意，这一假设弱于条件1，因为我们只要求每个批次内的第一类错误较低。我们令 p_L 和 q_L 依赖于批次大小 L，因为我们预计该概率会随着批次大小的增加而降低。

其次，对于任意玩家 $i \in [N]$，一次相对于 $w^i_{\boldsymbol{v}}$、且在某个批次上仍未被检测到的偏离，不应带来过多的效用增益。

条件4. 对于 $\beta \in (0,1)$，存在 Δ_L，使得对任意 $k \geqslant 0$，玩家 $i \in [N]$ 的行动序列 $(a^i_t)_{t \in \mathcal{B}_k} \in (\mathcal{A}^i)^L$，

$$\mathbb{1}\{\phi^i_k = 0\} \sum_{t \in \mathcal{B}_k} \beta^t\, u^i(a^i_t, w^{-i}_{\boldsymbol{v}}) \leqslant \sum_{t \in \mathcal{B}_k} \beta^t (v^i + \Delta_L)\,,$$

其中 $u^i(a^i, w^{-i}_{\boldsymbol{v}}) = \sum_{a^{-i}} u^i(a^i, a^{-i})\, w^{-i}_{\boldsymbol{v}}[a^{-i}]$.

广义而言，条件4确保：如果在一个批次上未检测到偏离（如$\mathbb{1}\{\phi^i_k = 0\}$所刻画），那么当对手坚持 $w^{-i}_{\boldsymbol{v}}$时，玩家 i 的任何行动序列 $(a^i_t)_{t \in \mathcal{B}_k}$ 都不会在该批次上产生大于 $v^i + \Delta_L$ 的折扣收益。

回想一下，条件 2 要求偏离时未触发惩罚的概率有上界。正是由于这一要求，在随时框架中，我们将关注范围限制在$\tilde{}$平稳偏离上 $\mathfrak{S}$：要对任意偏离作出这样的保证并不可行。相比之下，条件4只要求偏离带来的效用增益有上界。正如我们稍后将看到的，对于第4.1节所考察的测试类，这一条件对任意偏离都成立。

这正是批次方法的关键优势，也解释了它为何能够容纳比任意时刻检验丰富得多的策略类别。顺带一提，请注意，如果参与者观察的是实现的收益而非行动，就可以直接检验条件 Condition 4。这一观察表明，我们的框架可以自然地扩展到基于收益的监控而非基于行动的监控的情形；我们将其留待今后研究。

下面我们将证明，当条件 3 和条件4 得到满足时，可以通过批次检验恢复民间定理。

定理2。令 $\varepsilon > 0$、$\beta \in (0,1)$，并考虑在 (19) 中定义的 $\bar{\mathbf{s}}_{\boldsymbol{v}}$，其中 ϕ 满足条件 *3* 和条件*4*。令 $\boldsymbol{v} \in \boldsymbol{U}^1 \times \ldots \times \boldsymbol{U}^N$ 为一个可行收益，使得 $v^i \geqslant u^i$ 对于 $i \in [N]$ 成立。进一步假设，对于任意 $i \in [N]$，

$$\beta^L \geqslant \frac{\bar{u}^i - v^i - \Delta_L}{\bar{u}^i - \underline{u}^i} \,. \tag{21}$$

则有

1. *For any* $i \in [N]$, $\beta^L \left(1 - \frac{p_L}{1-\beta^L}\right) v^i \leqslant U^i(\bar{\mathbf{s}}_{\boldsymbol{v}}) \leqslant v^i$,
2. $\bar{\mathbf{s}}_{\boldsymbol{v}}$ *is a* $\left(\beta^{-L}(1 - \beta^{2L} + \Delta_L) + \frac{p_L}{1-\beta^L}\,,\, q_L\right)$*-HP-SPNE.*

定理2表明，只要条件 3 和条件4 成立，批次检验就能得到一个近似民间定理。这一结果既与其任意时刻对应结果定理2存在相似之处，也有所不同。

关于相似之处，两项结果都保证 $U^i(\bar{\mathbf{s}}_{\boldsymbol{v}})$ 与 v^i 保持接近。在任意时刻情形下，随着第一类错误界 γ 变小，近似会更加精确；类似地，在批处理情形下，随着 p_L 减小，近似也会得到改善。此外，在这两种情形下，策略组合都构成一个近似均衡。在任意时刻设定中，近似项随着 γ 和 ε（即偏离的幅度）而缩小；类似地，在批处理情形下，随着每个批次中的最大效用增益 p_L 和 Δ_L 减小，该项也会消失。最后，耐心条件（21）涉及 β^L，正如（10）涉及 β^{τ_ε} 一样。其直觉完全相同：由于检验只在批次结束时进行，参与者在被检测到之前最多可以享有 L 轮次的偏离。

然而，定理4与定理2有两个重要区别。首先，批处理构造的 $\bar{\mathbf{s}}_{\boldsymbol{v}}$ 是针对所有可能偏离的均衡，而任意时间结果只对平稳偏离提供保证。这是因为，相较于条件1，条件4更容易在任意偏离下满足。其次，批次均衡是子博弈精炼的（除了少数比例为 q_L 的延续博弈），而任意时刻设定下的 $\mathbf{s}_{\boldsymbol{v}}$ 只是纳什均衡。这一区别的原因在于，对于任意时刻 $t \geqslant 0$，任意时间程序都会利用截至时间 t 观察到的所有行动来追踪经验频率 $\hat{w}_t^i(a)$。因此，可以在任意轮次 t 构造一段历史 $\mathbf{h}_t$，使其有力地支持某位玩家 i 的 $\mathbf{H}_0^i$，从而让该玩家能够偏离而不被迅速察觉。相反，批次检验程序会遗忘

过去的行动，只关注当前批次。在重复博弈文献中，这种遗忘特性通常称为有限记忆（Mailath and Samuelson，2006年），正是它使得该构造能够产生子博弈精炼均衡。对于从 $t \geqslant 0$开始的任意延续博弈，历史 h_t 只会影响当前批次的检验统计量，不会作为证据累积到后续批次中。总而言之，定理4中的均衡能够应对任意偏离且是子博弈精炼的，这一事实使得该结果从博弈论视角来看显著强于定理2。

4.1 通过集中性进行批次检验

为了实施批次先检验后惩罚策略，我们还需要确定满足条件 3 和条件4的检验序列$\boldsymbol{\phi} = (\boldsymbol{\phi}^1, \ldots\ldots, \boldsymbol{\phi}^N)$。我们提出一个非常简单的程序：在每个批次中，参与者收集玩家 i采取的纯行动，计算其经验频率，并检验这些频率是否与 $w^i_{\boldsymbol{v}}$相符。形式化地，记 $\mathcal{A}^i = \{a^i(1), \ldots\ldots, a^i(K)\}$，对于任意 $k \geqslant 0$和$i \in [N]$，令

$$\hat{w}^i_k = L^{-1}\left(\sum_{t\in\mathcal{B}_k} \mathbb{1}\{A^i_t = a^i(1)\}, \ldots, \sum_{t\in\mathcal{B}_k} \mathbb{1}\{A^i_t = a^i(K)\}\right)^{\intercal}, \tag{22}$$

表示 i 在第 k个批次中采取的纯行动的经验频率向量。于是，对于给定的阈值 $\delta \in (0, 1)$，我们用检验统计量检验 $\mathrm{H}^i_{0,\ k}$

$$\tilde{\phi}^i_k = \mathbb{1}\{\|\hat{w}^i_k - w^i_{\boldsymbol{v}}\|_1 \geqslant \delta\}\,. \tag{23}$$

(23) 背后的直觉很简单。如果 $\mathrm{H}^i_{0,\ k}$ 成立，那么经验分布 $\hat{}w^i_k$ 以很高的概率集中在 $w^i_{\boldsymbol{v}}$ 附近。其逆否命题立即告诉我们，如果 $\|\hat{}w^i_k - w^i_{\boldsymbol{v}}\|_1$ 很高，就可以怀疑原假设 $\mathrm{H}^i_{0,\ k}$ 不成立。我们将在下一个命题中证明，我们的检验 (23) 确实满足条件 3 和条件4中的要求。

命题 3. 令 $\beta, \delta \in (0, 1)$。对于任意 $i \in [N]$，考虑 (23) 中定义的检验，以及由此得到的、在 (19)中定义的策略 $\bar{}\mathbf{s}_{\boldsymbol{v}}$。那么，

1. 条件 *3* 在 $p_L = q_L = 2KN\exp(-2L\delta^2/K^2)$ 下成立，

2. 条件*4* 在 $\Delta_L = \delta + 3(1-\beta^L)$ 下成立。

命题 3 的第一点可直接由集中性论证得出。第二点中，Δ_L 包含两项。第一项源于玩家 i可以在 δ 范围内稍微偏离 $w^i_{\boldsymbol{v}}$，而不会触发玩家 i的检验。第二项则源于玩家 i 能够在一个批次内重新排列纯行动，而不改变 $\hat{}w^i_k$；同时，这种排列能让价值最高的行动在批次开始时执行，从而享受更高的折扣。

值得注意的是，L 与 δ 中的两种权衡出现在命题 3 中。一方面，较大的 L确保在 $\mathrm{H}^i_{0,k}$ 下，$\hat{}w^i_k$ 更紧密地集中于 $w^i_{\boldsymbol{v}}$ 附近，这在统计上是可取的（p_L

随 L减小）。另一方面，较大的 L给对手在一个批次内更多偏离空间，并可能获得可观的收益增益，因为检测只能在该批次结束时发生（Δ_L 随 L增大）。阈值 δ也存在类似的张力。较大的 δ 降低假阳性风险，即基于 $\|\hat{w}^i - w^i_{\boldsymbol{v}}\|_1$ 中仅仅是随机的波动而实施惩罚（p_L随 δ 增大而减小）。然而，将 δ 设得过高可能会鼓励对手偏离$w^i_{\boldsymbol{v}}$，因为此类偏离更不容易被检测到（Δ_L 随 δ增大）。

话虽如此，只要仔细调整 L 和 δ，就有可能兼顾二者，使 p_L 和 Δ_L 都很小。结合定理4的这一观察可得以下推论。

推论3。 令 $\boldsymbol{v} = (v^1, \ldots, v^N) \in U^1 \times \ldots \times U^N$，使得对任意 $i \in [N]$、$\varepsilon > 0$和$\beta \in (0,\ 1)$，都有 $v^i \geqslant u^i$。对任意 $i \in [N]$，考虑由(23)定义、水平为 $\delta = \varepsilon/16$ 且批次大小为

$$L = \left\lceil \frac{258K^2}{\varepsilon^2} \ln\left(\frac{126KN}{\varepsilon^2}\right) \right\rceil,$$

以及(19)中定义的相应策略 $\bar{\mathbf{s}}_{\boldsymbol{v}}$。若 $1 - \varepsilon/16 \leqslant \beta^L \leqslant 1 - \varepsilon/32$，则

(i) $(1-\varepsilon)v^i \leqslant U^i(\bar{\mathbf{s}}_{\boldsymbol{v}}) \leqslant v^i$,

(ii) $\bar{\mathbf{s}}_{\boldsymbol{v}}$ 是一个$(\varepsilon, \varepsilon)$-*HP-SPNE*。

该推论表明，只要 β 足够大，批量先检验后惩罚策略就是一种近似HP-SPNE，并且其收益始终非常接近目标 $\boldsymbol{v}$。特别地，只要 β 足够接近于一，就可以使 ε 趋于零。在这种情况下，尽管参与者只能观察从策略中抽取的纯行动，我们几乎重新获得了定理1中的完美监控结果。我们再次强调，该结果证明 $\bar{\mathbf{s}}_{\boldsymbol{v}}$ 是一个子博弈完美均衡，并且并未假设策略是平稳的。最后，我们注意到，推论3中关于 β 的耐心条件比原定理4中的条件略为严格。事实上，我们要求 $\beta^L \geqslant 1 - \varepsilon/16$，这比 $\beta^L \geqslant (\bar{u}^i - u^i - \varepsilon/16)(\bar{u}^i - u^i)^{-1}$更强，同时还要求一个上界 $\beta^L \leqslant 1 - \varepsilon/32$。后一个条件对于界定 β^L 中的项是必要的，因为这些项会出现在定理4的近似项中。尽管关于 β 的条件看起来可能具有限制性，但可以这样理解：固定 $\beta \in (0,1)$，对于任何满足 $1 - \varepsilon/16 \leqslant \beta^L \leqslant 1 - \varepsilon/32$ 的 $\varepsilon > 0$，该构造都会在适当选择 $L = L(\varepsilon)$的情况下得到一个$(\varepsilon, \varepsilon)$-HP-SPNE；然后选择可允许的最小 ε，即可得到最紧的保证。

尽管从博弈论视角来看，定理4比定理2的结果更强，但批量检验方法并不具备任意时刻检验的所有便利性质。特别是，任意时刻检验允许我们明确控制第一类错误和第二类错误（见命题2和定理3）。相反，批处理情形无法保证避免第一类错误。事实上，在批次先检验后惩罚策略下，当合作策略并非退化时，错误惩罚以概率1在有限时间内发生。

命题 4。 设 $\beta,\delta\in(0,\ 1)$。对于任意 $i\in[N]$，考虑(23) 中定义的检验，以及针对某个可行收益 $\boldsymbol{v}$ 而在(19) 中定义的相应策略 $\bar{\mathbf{s}}_{\boldsymbol{v}}$，使得 $v^i\geqslant u^i$ 对于 $i\in[N]$成立。假设

$$\min_{j\in[N]}\left\{\|w^j_{\boldsymbol{v}}\|_0=\sum_{\ell=1}^{K}\mathbb{1}\{w^j_{\boldsymbol{v},\ell}\neq 0\}\right\}>1\ \ and\ \delta<1-1/\min_{j\in[N]}\|w^j_{\boldsymbol{v}}\|_0\ \ . \tag{24}$$

然后 $\mathbb{P}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}(\kappa_{\boldsymbol{\phi}}<\infty)=1$。

我们强调，命题 4并不与推论3矛盾；相反，它仅意味着惩罚阶段以概率1在有限时间内发生。然而，惩罚平均而言发生得足够晚，因此不会对效用产生显著影响，从而仍然允许参与者进行合作，正如定理4所强调的那样。尽管如此，从公平性角度来看，或在参与者具有风险厌恶的环境中，这一特征可能并不可取，因为它意味着惩罚最终是不可避免的。在此类情形下，实施任意时间先检验后惩罚机制可能更为合适。这一比较凸显出，两种方法涉及不同的权衡，而最终应如何在两者之间进行选择，取决于制度背景以及参与者对风险和公平性的态度。

5 结论

民间定理长期以来一直是研究重复博弈中合作的经典方法，其依赖由偏离触发的惩罚。本文提出了一个通过统计检验分析多人重复博弈中如何维持合作的新框架：参与者使用混合策略，但只能观察到实现的纯行动；执行机制则采用统计方式，明确管理误报和漏检。我们引入了均衡概念，能够安全地忽略由监控过程生成的尾概率历史，并提出了一种一般性的先检验后惩罚策略，得到了两个民间定理类型的结果，表明任何可行且个体理性的收益组合，都可以在参与者足够有耐心时作为近似均衡得到支持。据我们所知，本研究首次利用假设检验框架所提供的灵活性，分析具有不完美公共监控的重复博弈。

我们还提供了两种实现及相应的检验。任意时间构造采用序贯推断，提供可实施的策略，在整个时域内实现统一的第一类错误控制，并保证有限样本检测；而批处理构造则放弃任意时间第一类错误控制，以处理更丰富的偏离模式，并获得子博弈完美的高概率类比。

更广泛而言，这项分析凸显了当互动由数据与推断所中介时，重复博弈的推理方式会如何改变。未来值得探索的方向包括更丰富的信息结构——例如私人信号、异质主体，以及自适应或非平稳环境等。

参考文献

Dilip Abreu、David Pearce 和 Ennio Stacchetti。《不完美监控下折扣重复博弈理论初探》。《计量经济学》，58(5)：1041–1063，1990年。

世界反兴奋剂机构。世界反兴奋剂机构技术文件，TD2023APMU。20 23.

Katsuhiko Aiba。针对私人近乎完美监控随机博弈的民间定理。《博弈与经济行为》，86:58–66，2014。ISSN 0899-8256。doi：https://doi.org/10.1016/j.geb.2014.03.007。URL https://www.sciencedirect.com/science/article/pii/S0899825614000517。

Sandra Andraszewicz、Benjamin Scheibehenne、Jörg Rieskamp、Raoul Grasman、Josine Verhagen 和 Eric-Jan Wagenmakers。面向管理学研究的贝叶斯假设检验入门。《管理学杂志》，41(2):521–543，2015。

Eshwar Ram Arunachaleswaran、Natalie Collina 和 Jon Schneider。学习与未知对手对弈。载于第*26*届*ACM*经济学与计算会议论文集，第478–504页，2025。

Andrea Attar、EloisaCampioni、Thomas Mariotti 和 Gwenaël Piaser。竞争机制与民间定理：两个例子。《博弈与经济行为》，125:79–93，2021年。ISSN 0899-8256。doi: https://doi.org/10.1016/j.geb.2020.10.006。URL https://www.sciencedirect.com/science/article/pii/S0899825620301561。

Robert J Aumann、Michael Maschler 和 Richard E Stearns。《不完全信息重复博弈》。麻省理工学院出版社，1995年。

Stephen Bates、Michael I Jordan、Michael Sklar 和 Jake A Soloff。委托代理假设检验。《arXiv预印本 arXiv:2205.06812》，2022年。

Yoav Benjamini 和 Yosef Hochberg。控制假发现率：一种实用且有效的多重检验方法。《皇家统计学会杂志：B辑（方法论）》，57(1):289–300，1995年。

Dimitris Bertsimas 和 Georgia Perakis。《动态定价：一种学习方法》。载于拥堵收费的数学与计算模型，第 101 卷，第 45–79 页。施普林格，2006年。

Felix Brandt、Vincent Conitzer 和 Ulle Endriss。计算社会选择。多智能体系统，2:213–284，2012。

Emilio Calvano、Giacomo Calzolari、VincenzoDenicolò 和SergioPastorello。人工智能、算法定价与合谋。《美国经济评论》，110(10):3267–97，2020年10月。doi: 10.1257/aer.20190623。URLhttps://www.aeaweb.org/articles?id=10.1257/aer.20190623。

Aymeric Capitaine、Etienne Boursier、Antoine Scheid、Eric Moulines、Michael Jordan、El-Mahdi El- Mhamdi 和Alain Durmus。协作学习中的解缠。神经信息处理系统进展，37:97231–97260，2024年。

Nicolo Cesa-Bianchi 和 Gábor Lugosi。《预测、学习与博弈》。剑桥大学出版社，2006年。

Herman Chernoff。基于观测值总和的假设检验渐近效率度量。《数学统计年刊》，23(4):493–507，1952年。

Natalie Collina、Varun Gupta 和 Aaron Roth。与多个非短视主体的重复签约：策略遗憾与有限责任。载于第*25*届*ACM*经济学与计算会议论文集，第640–668页，2024年。Olivier Compte。不完美私人监控下重复博弈中的通信。《计量经济学》，66(3):597–626，1998年。Thomas M Cover 和 Joy A. Thomas。《信息论基础》。John Wiley & Sons，1999年。Joyee Deb、Julio González-Díaz 和 Jérôme Renault。重复匿名随机匹配博弈中的一致民间定理。《博弈与经济行为》，100: 1–23，2016年。ISSN 0899-8256。doi: https://doi.org/10.1016/j.geb.2016.08.006。URLhttps://www.sciencedirect.com/science/article/pii/S0899825616300835。Jeffrey C. Ely 和 Juuso Välimäki。囚徒困境的稳健民间定理。《经济理论杂志》，102(1):84–105，2002年。Jeffrey C Ely、Johannes Hörner 和 Wojciech Olszewski。重复博弈中的无信念均衡。《计量经济学》，73(2):377–415，2005年。Ido Erev 和 Alvin E Roth。预测人们如何进行博弈：在具有唯一混合策略均衡的实验博弈中的强化学习。《美国经济评论》，88 (4):848–881，1998年。

Alireza Fallah 和 Michael Jordan。安全检查中的契约设计。载于第*25*届*ACM*经济学与计算会议论文集，第616–638页，2024年。

Zhe Feng、Chara Podimata 和 Vasilis Syrgkanis。不了解自身估值也能学会竞价。载于*2018年ACM经济学与计算会议论文集*，第505–522页，2018年。Ronald Aylmer Fisher。面向研究人员的统计方法。Oliver and Boyd，1925年。James W Friedman。超博弈的非合作均衡。《经济研究评论》，38(1):1–12，1971年。Drew Fudenberg 和 Eric Maskin。折扣或不完全信息下重复博弈的民间定理。《计量经济学》，54(3):533–554，1986年。Drew Fudenberg 和 Jean Tirole。博弈论。麻省理工学院出版社，1991年。Drew Fudenberg、David K. Levine 和 Eric Maskin。不完美公共信息下的民间定理。《计量经济学》，62(5):997–1039，1994年。Olivier Gossner。有限次重复博弈（采用混合策略）的民间定理。《国际博弈论杂志》，24:95–107，1995年。Johannes Hörner 和 Wojciech Olszewski。私人近乎完美监控博弈的民间定理。《计量经济学》，74(6):1499–1544，2006年。Johannes Horner 和 Jérôme Renault。公共监控下有限次重复博弈的民间定理。工作论文或预印本，2025年1月。URL https://hal.science/hal-04875858。Steven R Howard、Aaditya Ramdas、Jon McAuliffe 和 Jasjeet Sekhon。时间一致的非参数、非渐近置信序列。《统计年鉴》，49(2):1055–1080，2021年。Ramesh Johari、Vijay Kamble 和 Yash Kanoria。边学习边匹配。《运筹学》，69(2):655–681，2021年。Michael I Jordan。关于人工智能的集体主义经济学视角。*arXiv*预印本 *arXiv:2507.06268*，2025年。Yoav Kolumbus 和 Noam Nisan。如何操纵自己的智能体，以及为何要这么做：学习型智能体用户的激励。神经信息处理系统进展，35:28080–28094，2022。Yoav Kolumbus、Jon Schneider、Inbal Talgam-Cohen、Emmanouil-Vasileios Vlatakis-Gkaragkounis、Joshua R Wang 和 SM Weinberg。与学习型智能体签约。神经信息处理系统进展，37:77366–77408，2024年。Marie Laclau。网络上进行的重复博弈的民间定理。《博弈与经济行为》，76(2):711–737，2012。ISSN 0899-8256。doi: https://doi.org/10.1016/j.geb.2012.08.008。URL https://www.sciencedirect.com/science/article/pii/S0899825612001285。

Erich Leo Lehmann 和 Joseph P Romano。《统计假设检验》。施普林格，2005年。Lydia T Liu、Horia Mania 和 Michael Jordan。匹配市场中的竞争性赌博机。载于国际人工智能与统计会议，第 1618–1628 页。PMLR，2020年。George J Mailath 和 Stephen Morris。近乎公共监控下的重复博弈。《经济理论杂志》，102(1):189–228，2002年。George J Mailath 和 Larry Samuelson。《重复博弈与声誉：长期关系》。牛津大学出版社，2006年。Yishay Mansour、Aleksandrs Slivkins 和 Vasilis Syrgkanis。贝叶斯激励相容的赌博机探索。《运筹学》，68(4):1132–1161，2020年。John Moore 和 Rafael Repullo。子博弈精炼实施。《计量经济学》，56 (5):1191–1220，1988年。Jamie Morgenstern 和 Tim Roughgarden。简单拍卖学习。载于学习理论会议，第 49 卷，第 1298–1318 页。PMLR，2016年。Jaouad Mourtada。相对熵下离散分布的估计，以及缺失质量的偏离。*arXiv* 预印本 *arXiv:2504.21787*，2025。Jonas W Mueller、Vasilis Syrgkanis 和 Matt Taddy。用于高维动态定价的低秩赌博机方法。《神经信息处理系统进展》，32，2019年。Jerzy Neyman 和 Egon Sharpe Pearson。论统计假设最有效检验的问题。《伦敦皇家学会哲学汇刊》*A*辑，231(694-706): 289–337，1933年。Iosif Pinelis。巴拿赫空间中鞅分布的最优界。《概率年刊》，22 (4):1679–1706，1994年。美国注册会计师协会、上市公司会计监督委员会（PCAOB）。审计准则2305：实质性分析程序。《审计工作人员实务提示》，2023。Aaditya Ramdas 和 Ruodu Wang。使用 e 值进行假设检验。基础与趋势® 统计学，1(1-2):1–390，2025。Aaditya Ramdas、Johannes Ruf、Martin Larsson 和 Wouter M Koolen。交换性检验：叉凸性、上鞅与 e 过程。《国际近似推理杂志》，141:83–109，2022年。Tim Roughgarden。算法博弈论。《美国计算机协会通讯》，53(7):78–86，2010年。

Ariel Rubinstein。超博弈中的强完美均衡。《国际博弈论杂志》，9(1):1–12，1980年。Antoine Scheid、Aymeric Capitaine、Etienne Boursier、Eric Moulines、Michael Jordan 和 Alain Durmus。学习如何缓解外部性：具有后见理性的科斯定理。《神经信息处理系统进展》，37:15149–15183，2024年。Antoine Scheid、Etienne Boursier、Alain Durmus、Eric Moulines 和 Michael I Jordan。具有转移支付的树状多智能体博弈中的在线决策。*arXiv* 预印本 *arXiv:2501.19388*，2025。Tadashi Sekiguchi。私人监控下重复囚徒困境中的效率。《经济理论杂志》，76(2):345–361，1997年。Glenn Shafer 和 Vladimir Vovk。《概率与金融：这只是一场游戏！》John Wiley & Sons，2005年。Juliet Popper Shaffer。多重假设检验。《心理学年度评论》，46(1):561–584，1995年。Takuo Sugaya。私人监控下重复博弈的民间定理。《经济研究评论》，89(4):2201–2256，2022年。Vasilis Syrgkanis、Alekh Agarwal、Haipeng Luo 和 Robert E Schapire。博弈中正则化学习的快速收敛。《神经信息处理系统进展》，28，2015年。Yevgeny Tsodikovich、Xavier Venel 和 Anna Zseleva。具有转换成本的重复博弈中的民间定理。《博弈与经济行为》，146:137– 159，2024年。ISSN 0899-8256。doi: https://doi.org/10.1016/j.geb.2024.05.006。URL https://www.sciencedirect.com/science/article/pii/S0899825624000745。Jean Ville。《集体概念批判研究》。巴黎，*Gauthier-Villars*，1939。Abraham Wald。统计假设的序贯检验。《数学统计年刊》，16 (2):117–186，1945年。Ruodu Wang 和 Aaditya Ramdas。使用 e 值控制假发现率。《皇家统计学会会刊*B*辑：统计方法论》，84(3):822–852，2022年。Álvaro Cartea、Patrick Chang、José Penalva 和 Harrison Waldon。算法共谋与有限理性学习导出的民间定理。《博弈与经济行为》，157:1–21，2026。ISSN 0899-8256。doi: https://doi.org/10.1016/j.geb.2025.11.012。URL https://www.sciencedirect.com/science/article/pii/S0899825625001745。

统计学背景

定义 3（滤过）。 在共同的概率空间 $(\Omega, \mathcal{F}, \mathbb{P})$ 上，滤过 是由一列 σ-域构成的序列$\{\mathcal{F}_t\}_{t\geq 0}$，满足

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}\,.$$

其含义是，$\mathcal{F}_t$ 表示截至并包括时间 t 时可获得的所有信息。

定义 4（上鞅）。 适应于滤过$\{\mathcal{F}_t\}_{t\geqslant 0}$ 的实值随机过程 $(M_t)_{t\geq 0}$，若满足以下条件，则称为上鞅

$$\mathbb{E}[|M_t|] < \infty \quad \text{for any } t \geqslant 0 \quad \text{and} \quad \mathbb{E}[M_t \mid \mathcal{F}_{t-1}] \leqslant M_{t-1} \text{ a.s.}$$

若等号成立，则 $(M_t)_{t\geqslant 0}$ 是一个鞅。上鞅描述的是条件期望不会随时间增加的过程。

定义 5（e过程）。 e过程（或检验鞅）是一个非负的、适应于$\{\mathcal{F}_t\}_{t\geqslant 0}$的过程$\tilde{}$（$E_t$）$_{t\geqslant 0}$，满足

$$\mathbb{E}[\tilde{E}_t] \leqslant 1 \quad \text{for any } t \geqslant 0\,.$$

$\tilde{}$在原假设下，$(E_t)_{t\geqslant 0}$ 通常是一个非负上鞅，其均值为1或小于1。对于任意 $\gamma \in (0,1)$，停止规则“在首次出现满足 $E_t \geqslant 1/\gamma$的 t 时拒绝”给出一个序贯检验，该检验在所有停止时间下都能将第一类错误控制在水平 γ。

$\tilde{}$**定理5**（Ville 不等式，Ville（1939））。 设$(E_t)_{t\geq 0}$ 是一个非负上鞅，且满足$\tilde{}\mathbb{E}[E_0] \leqslant 1$。那么，对于任意阈值 $c > 0$，有

$$\mathbb{P}\left(\sup_{t\geqslant 0} \tilde{E}_t \geqslant c\right) \leqslant 1/c\,,$$

$\tilde{}$这也可以改写为 $\mathbb{P}(\exists t \geqslant 0 \text{ 使得 } E_t \geqslant c) \leqslant 1/c$。特别地，令 $c = 1/\gamma$，即可得到不等式 $\mathbb{P}(\sup_{t\geqslant 0} E_t \geqslant 1/\gamma) \leqslant \gamma$。这一结果保证了基于e过程的检验具有任意时刻有效的第一类错误控制。

引理1. *[Chernoff-Stein 引理，参见 Chernoff (1952)、Cover 与 Thomas (1999)]* 设 $(A_t)_{t\geqslant 0}$ 为一列取值于离散集合 $\mathcal{A}$ 的独立同分布随机变量。考虑简单假设

$$\mathrm{H}_0 \colon A_t \sim \tilde{w} \qquad vs \qquad \mathrm{H}_1 \colon A_t \sim q\,,$$

其中 $q(a) > 0$ 且 $\tilde{w}(a) > 0$ 对所有 $a \in \mathcal{A}$ 均成立，并且 $\mathrm{KL}(q\|\tilde{w}) = \sum_{a\in\mathcal{A}} q(a)\log\, {}^{q(a)}\!/_{\tilde{w}(a)} < \infty$。令 ϕ_n 表示时刻 n的检验，该检验取决于历史 h_n；令 $\phi_n^{-1}(\{1\}) \subseteq \mathcal{H}$ 表示由 ϕ_t诱导的假设 H_1 的接受域。定义第一类错误概率

$$\gamma_n^{(1)} = \mathbb{P}_{\tilde{w}}(\phi_n = 1)\,,$$

以及

$$\gamma_n^{(2)} = \min_{h_n \subseteq \sigma(A_0,\dots,A_n):\, \mathbb{P}_{\tilde{w}}(\phi_n=1)\leqslant\gamma^{(1)}} \mathbb{P}_q(\phi_n = 0)\,.$$

于是，我们有

$$\lim_{n\to\infty} 1/n \log \gamma_n^{(2)} = -\mathrm{KL}(\tilde{w}\|q)\,.$$

引理1刻画了在第一类错误（误报）受限于某个固定常数时，第二类错误（漏检）可达到的最佳指数衰减率。

混合检验鞅。 最后，我们重点介绍混合 e 过程，它们可能有助于将我们的设定扩展到不可观测效用。固定一个对称先验密度 π，其在 $\mathbb{R}$ 上严格为正且连续（通常为正态分布 $\mathcal{N}(0,\tau^2)$），使得 $\int_{\mathbb{R}} \pi(\lambda)\, d\lambda = 1$。将随机过程 $\{E_t\}_{t\geqslant 0}$ 定义为 $E_0 = 1$，并对任意 $t \geqslant 1$

$$\tilde{E}_t = \int_{\mathbb{R}} \exp\{\lambda S_t - \lambda^2\sigma^2\, t/2\}\, \pi(\lambda)\, d\lambda\,, \tag{25}$$

以及检验

$$\phi_t = \mathbb{1}\{\tilde{E}_t \geqslant 1/\gamma\} \qquad \text{for any } t \geqslant 1\,. \tag{26}$$

对于预先指定的水平 $\gamma \in (0,1)$，一旦满足 $\phi_t = 1$，即可拒绝 H_0。

任意时刻有效性。 在 H_0，下，式（25）中的过程 $(E_t)_{t\geqslant 0}$ 是一个非负的 $\{\mathcal{F}_t\}$-上鞅，其中 $\mathbb{E}[E_t] \leqslant 1$ 对任意 $t \geqslant 0$ 均成立。因此，根据 Ville 不等式，我们有

$$\mathbb{P}_{\mathrm{H}_0}\left(\sup_{t\geq 0} \tilde{E}_t \geqslant 1/\gamma\right) \leqslant \gamma\,.$$

因此，停止规则“在首次出现 t 且满足 $\phi_t = 1$ 时拒绝”的第一类错误任意时刻有效，并且小于 γ。

B 完美监控下民间定理的证明

定理1。 假设完美监控。固定一个收益组合 $\boldsymbol{v} = (v^1, \dots, v^N) \in \boldsymbol{U}^1 \times \dots \times \boldsymbol{U}^N$，使得 $v^i \geqslant u^i$ 对任意 $i \in [N]$ 均成立。若 $\beta \geqslant (\bar{u}^i - v^i)(\bar{u}^i - u^i)^{-1}$ 对任意玩家 i 均成立，则按式 (3) 定义的 $\mathbf{s}_{\boldsymbol{v}}$ 满足：（*i*） $\underline{U}^i(\mathbf{s}_{\boldsymbol{v}}) = v^i$ 对任意 $i \in [N]$ 均成立；（*ii*） $\mathbf{s}_{\boldsymbol{v}}$ 是一个子博弈完美纳什均衡。

定理1的证明。 令 $t_0 \geqslant 0$ 和 $\mathrm{h}_{t_0} \in \mathcal{H}$ 为在时间 t_0观察到的 $\mathbf{s}_{\boldsymbol{v}}$ 所生成的历史。我们考虑两种情况。固定某个 $i \in [N]$。

首先，假设存在 $t < t_0$，使得$\mathbf{w}_t \in \mathrm{h}_{t_0}$ 且 $w_t^j \neq w_{\boldsymbol{v}}^j$ 对某个 $j \in [N]$成立，因此我们已经处于惩罚阶段。根据式 (3) 中$\mathbf{s}_{\boldsymbol{v}}$的定义，从时间 $t+1$起，参与者只进行 $\boldsymbol{b} = (b^1, \dots, b^N)$。从 t_0 开始，玩家 i 的延续收益为

$$U^i(\mathbf{s}_{\boldsymbol{v}}\,;\mathrm{h}_{t_0}) = (1-\beta)\sum_{\ell=t_0}^{\infty} \beta^{\ell-t_0} u^i(b^i, b^{-i}) = u^i(b^i, b^{-i})\,.$$

现在，考虑从 t_0开始的延续博弈中的任意偏离 $s \in \mathcal{S}$。我们有

$$\begin{aligned} U^i(s, s_{\boldsymbol{v}}^{-i}\,;\mathrm{h}_{t_0}) &= (1-\beta)\sum_{\ell=t_0}^{\infty}\beta^{\ell-t_0}u^i(s(\mathrm{h}_{\ell-1}), b^{-i}) \\ &\leqslant (1-\beta)\sum_{\ell=t_0}^{\infty}\beta^{\ell-t_0}u^i(b^i, b^{-i}) = u^i(b^i, b^{-i}) = U^i(\mathbf{s}_{\boldsymbol{v}}\,;\mathrm{h}_{t_0})\,, \end{aligned}$$

因为 $\boldsymbol{b}$ 是一个纳什均衡。

其次，假设对于任意$\mathbf{w}_t = (w_{\boldsymbol{v}}^1, \ldots, w_{\boldsymbol{v}}^N)$，都有$\mathbf{w}_t \in \mathrm{h}^{t_0}$。根据$\mathbf{s}_{\boldsymbol{v}}$的定义，我们有

$$U^i(\mathbf{s}_{\boldsymbol{v}}\,;\mathrm{h}_{t_0}) = (1-\beta)\sum_{s=t_0}^{\infty}\beta^{s-t_0}u^i(w_{\boldsymbol{v}}^i, w_{\boldsymbol{v}}^{-i}) = v^i\,.$$

现在，考虑玩家 i的偏离 $s \in \mathcal{S}$，并记 $T = \min\{t \geqslant t_0\colon\ s(\mathrm{h}_t) \neq w_{\boldsymbol{v}}^i\} < \infty$。玩家$i$的延续收益为

$$\begin{aligned} U^i(s, s_{\boldsymbol{v}}^{-i}\,;\mathrm{h}_{t_0}) &= (1-\beta)\left(\sum_{s=t_0}^{T-1}\beta^{s-t_0}u^i(w_{\boldsymbol{v}}^i, w_{\boldsymbol{v}}^{-i}) + \beta^{T-t_0}u^i(s(\mathrm{h}_T), w_{\boldsymbol{v}}^{-i}) + \sum_{s=T+1}^{\infty}\beta^{s-t_0}u^i(b^i, b^{-i})\right) \\ &\leqslant (1-\beta)\left(\sum_{s=t_0}^{T-1}\beta^{s-t_0}v^i + \beta^{T-t_0}\bar{u}^i + \sum_{s=T+1}^{\infty}\beta^{s-t_0}\underline{u}^i\right) \\ &= (1-\beta^{T-t_0})v^i + (1-\beta)\beta^{T-t_0}\,\bar{u}^i + \beta\,\beta^{T-t_0}\,\underline{u}^i \end{aligned}$$

最后，由于 $\beta \geqslant (\bar{u}^i - v^i)(\bar{u}^i - u^i)^{-1}$，

$$U^i(s, s_{\boldsymbol{v}}^{-i}\,;\mathrm{h}_{t_0}) \leqslant \bar{u}^i - \beta^{T-t_0}((1-\beta)\bar{u}^i - \beta\underline{u}^i + v^i) \leqslant v^i = U^i(\mathbf{s}_{\boldsymbol{v}}\,;\mathrm{h}_{t_0})\,.$$

这证明$\mathbf{s}_{\boldsymbol{v}}$ 确实是一个子博弈完美纳什均衡。 □

C 任意时刻设定下的证明

命题1。 固定 $\varepsilon > 0$。对于任意 $i \in [N]$，如果 $s^i \in \mathsf{B}^i(\mathbf{w}_{\boldsymbol{v}}, \varepsilon)$，那么 $U^i(s^i, s_{\boldsymbol{v}}^{-i}) \leqslant v^i + \varepsilon$.

证明。 令 $i \in [N]$ $s^i \in \mathsf{B}^i(\mathbf{w}_{\boldsymbol{v}}, \varepsilon)$ 为玩家 i 的一个小偏离。由于对于任意 $t > \tau_{\boldsymbol{\psi}}$ 和 $w_t^i \in \Delta(\mathcal{A}^i)$，$u^i(w_t^i, w_{\boldsymbol{v}}^{-i}) \leqslant u^i(b^i, b^{-i}) = u^i$，我们有

$$\begin{aligned} U^i(s^i, s_{\boldsymbol{v}}^{-i}) - v^i &\leqslant (1-\beta)\,\mathbb{E}\left[\sum_{t=0}^{\tau_{\boldsymbol{\psi}}}\beta^t u^i(s^i(\mathrm{h}_t), w_v^{-i}) + \sum_{t=\tau_{\boldsymbol{\psi}}+1}^{\infty}\beta^t\underline{u}^i\right] - v^i \\ &= (1-\beta)\left(\mathbb{E}\left[\sum_{t=0}^{\tau_{\boldsymbol{\psi}}}\beta^t u^i(s^i(\mathrm{h}_t), w_v^{-i}) + \sum_{t=\tau_{\boldsymbol{\psi}}+1}^{\infty}\beta^t\underline{u}^i\right]\right) - (1-\beta)\,\mathbb{E}\left[\sum_{t=0}^{\infty}\beta^t\,v^i\right] \\ &= (1-\beta)\left(\mathbb{E}\left[\sum_{t=0}^{\tau_{\boldsymbol{\psi}}}\beta^t u^i(s^i(\mathrm{h}_t), w_{\boldsymbol{v}}^{-i})\right] - \mathbb{E}\left[\sum_{t=0}^{\tau_{\boldsymbol{\psi}}}\beta^t v^i\right]\right) \\ &\quad + (1-\beta)\underbrace{\left(\mathbb{E}\left[\sum_{t=\tau_{\boldsymbol{\psi}}+1}^{\infty}\beta^t\underline{u}^i\right] - \mathbb{E}\left[\sum_{t=\tau_{\boldsymbol{\psi}}+1}^{\infty}\beta^t v^i\right]\right)}_{\leqslant 0}, \end{aligned}$$

并且由于 $v^i = u^i(\mathbf{w}_{\boldsymbol{v}})$

$$U^i(s^i, s^{-i}_{\boldsymbol{v}}) - v^i \leqslant (1-\beta)\,\mathbb{E}\left[\sum_{t=0}^{\tau_{\boldsymbol{\psi}}} \beta^t \Big(u^i(s^i(\mathrm{h}_t), w^{-i}_{\boldsymbol{v}}) - u^i(w^i_v, w^{-i}_v) \Big)\right].$$

现在，定义 $\zeta^i = (u^i(a_1, w^{-i}_{\boldsymbol{v}}) \ldots, (a_K, w^{-i}_{\boldsymbol{v}})) \in \mathbb{R}^K$，使得 $u^i(w, \tilde{w}^{-i}) = \langle w, \zeta^i \rangle$ 对于任意 $w \in \Delta(\mathcal{A})$。我们可以将前一个表达式改写为

$$\begin{aligned} U^i(s^i, s^{-i}_{\boldsymbol{v}}) - v^i &= (1-\beta) \sum_{t=0}^{\tau_{\boldsymbol{\psi}}} \beta^t \Big\langle s^i(\mathrm{h}_t) - w^i_{\boldsymbol{v}}, \zeta_i \Big\rangle \\ &\leqslant (1-\beta) \sum_{t=0}^{\tau_{\boldsymbol{\psi}}} \beta^t \|\zeta_i\|_\infty \|s^i(\mathrm{h}_t) - w^i_{\boldsymbol{v}}\|_1 \\ &\leqslant (1-\beta)\,\varepsilon \sum_{t=0}^{\tau_{\boldsymbol{\psi}}} \beta^t \\ &\leqslant \varepsilon\,, \end{aligned}$$

其中 w 我们使用了 $\|\zeta^i\|_\infty \leqslant 1$ 和 $\|s^i(\mathrm{h}_t) - w^i_{\boldsymbol{v}}\|_1 \leqslant \varepsilon$，这是 $s^i \in \mathrm{B}^i(w_{\boldsymbol{v}}, \varepsilon$ 的推论). □

定理2。 设 $\mathfrak{S} = \mathfrak{S}^1 \times \ldots \times \mathfrak{S}^N \subseteq S$，并考虑一族满足条件*1*和条件 *2* 的序贯检验 $\boldsymbol{\psi}$。设 $\boldsymbol{v} \in U^{\,1} \times \ldots \times U^N$是一个可行收益，使得对于任意 $i \in [N]$

$$v^i \geqslant \underline{u}^i = u^i(\mathbf{b})\,,$$

其中 $\mathbf{b} = (b^1, \ldots b^N) \in \Delta(\mathcal{A})$ 是某个参考混合策略纳什均衡，它根据纳什定理必然存在。记作 $\tau(\varepsilon)$ =上确界 $\mathbb{E}^{\mathbf{s}}[\tau_{\boldsymbol{\psi}}] < \infty$，以及 $\bar{u}^i$ =最大值$_{a \in \mathcal{A}}\, u^i(a)$，若对于任意 $i \in [N]_{\mathbf{s} \in \mathfrak{S}(\mathbf{w}_{\boldsymbol{v}}, \varepsilon)}$

$$\bar{\beta}^i = \frac{\bar{u}^i - (1-\gamma)v^i}{\bar{u}^i - \underline{u}^i} \in [0,1] \quad \textit{and} \quad \beta^{\tau(\varepsilon)} \geqslant \bar{\beta}^i\,, \tag{10}$$

则$\mathbf{s}_{\boldsymbol{v}}$ 定义于(5)，满足

(i) for any $i \in [N]$, $(1-\gamma)v^i \leqslant U^i(\mathbf{s}_{\boldsymbol{v}}) \leqslant v^i$,

(ii) $\mathbf{s}_{\boldsymbol{v}}$ 是一个$(\varepsilon+\gamma, \mathfrak{S})$-纳什均衡。

证明 定理*2*的证明。 我们首先证明（i）。令 $i \in [N]$。由 $\tau = \min^{i \in [N]} \tau_{\boldsymbol{\phi}^i}$，并且我们有 e

$$\begin{aligned} U^i(\mathbf{s}_{\boldsymbol{v}}) &= (1-\beta)\,\mathbb{E}^{\mathbf{s}_{\boldsymbol{v}}}\left[\mathbb{1}\{\tau_{\boldsymbol{\psi}} < \infty\}\left(\sum_{t=0}^{\tau_{\boldsymbol{\psi}}} \beta^t v^i + \sum_{t=\tau_{\boldsymbol{\psi}}+1}^{\infty} \beta^t \underline{u}^i\right) + \mathbb{1}\{\tau_{\boldsymbol{\psi}} = \infty\} \sum_{t=0}^{\infty} \beta^t v^i\right] \\ &\geqslant (1-\beta)\,\mathbb{E}^{\mathbf{s}_{\boldsymbol{v}}}\left[\mathbb{1}\{\tau_{\boldsymbol{\psi}} = \infty\}(1-\beta)^{-1} v^i\right] = \mathbb{P}^{\mathbf{s}_{\boldsymbol{v}}}(\tau_{\boldsymbol{\psi}} = \infty)\, v^i\,. \end{aligned}$$

现在，根据并集界和条件1

$$\begin{aligned} \mathbb{P}^{\mathbf{s}_{\boldsymbol{v}}}(\tau_{\boldsymbol{\psi}} = \infty) &= 1 - \mathbb{P}^{\mathbf{s}_{\boldsymbol{v}}}(\{\tau_{\boldsymbol{\psi}^1} < \infty\} \cup \ldots \cup \{\tau_{\boldsymbol{\psi}^N} < \infty\}) \\ &\geqslant 1 - (\mathbb{P}^{\mathbf{s}}_{\boldsymbol{v}}(\tau_{\boldsymbol{\psi}^1} < \infty) + \ldots + \mathbb{P}^{\mathbf{s}_{\boldsymbol{v}}}(\tau_{\boldsymbol{\psi}^N} < \infty)) \\ &\geqslant 1 - N\frac{\gamma}{N} \\ &= 1 - \gamma\,, \end{aligned}$$

因此我们得到

$$U^i(\mathbf{s}_{\boldsymbol{v}}) \geqslant (1-\gamma)v^i\,. \tag{27}$$

关于上界，我们注意到，根据假设，$v^i \geqslant u^i$，因此

$$\begin{aligned}
U^i(\mathbf{s}_{\boldsymbol{v}}) &= (1-\beta)\,\mathbb{E}^{\mathbf{s}_{\boldsymbol{v}}}\left[\mathbb{1}\{\tau_{\boldsymbol{\psi}}<\infty\}\left(\sum_{t=0}^{\tau_{\boldsymbol{\psi}}}\beta^t v^i + \sum_{t=\tau_{\boldsymbol{\psi}}+1}^{\infty}\beta^t \underline{u}^i\right) + \mathbb{1}\{\tau_{\boldsymbol{\psi}}=\infty\}\sum_{t=0}^{\infty}\beta^t v^i\right] \\
&\leqslant (1-\beta)\,\mathbb{E}^{\mathbf{s}_{\boldsymbol{v}}}\left[\mathbb{1}\{\tau_{\boldsymbol{\psi}}<\infty\}\left(\sum_{t=0}^{\tau_{\boldsymbol{\psi}}}\beta^t v^i + \sum_{t=\tau_{\boldsymbol{\psi}}+1}^{\infty}\beta^t v^i\right) + \mathbb{1}\{\tau_{\boldsymbol{\psi}}=\infty\}\sum_{t=0}^{\infty}\beta^t v^i\right] \\
&= v^i.
\end{aligned}$$

现在我们关注第(ii)点。令 $\varepsilon>0$ 和 $s\in\mathfrak{S}$ 为玩家 i的一种偏离。我们考虑两种情况。首先，假设 $s\in\mathfrak{S}$（$\mathbf{w}_{\boldsymbol{v}}$，$\varepsilon$）。对于任意玩家 i，我们有

$$\begin{aligned}
U^i(s, s_{\boldsymbol{v}}^{-i}) &= (1-\beta)\,\mathbb{E}^{(s,s_{\boldsymbol{v}}^{-i})}\left[\mathbb{1}\{\tau_{\boldsymbol{\psi}}=\infty\}\sum_{t=0}^{\infty}\beta^t u^i(A_t^1,\ldots,A_t^N)\right] \\
&\quad + (1-\beta)\,\mathbb{E}^{(s,s_{\boldsymbol{v}}^{-i})}\left[\mathbb{1}\{\tau_{\boldsymbol{\psi}}<\infty\}\left(\sum_{t=0}^{\tau_{\boldsymbol{\psi}}}\beta^t u^i(A_t^1,\ldots,A_t^N) + \sum_{t=\tau_{\boldsymbol{\psi}}}^{\infty}\beta^t u^i(A_t^1,\ldots,A_t^N)\right)\right] \\
&\leqslant \mathbb{P}^{(s,s_{\boldsymbol{v}}^{-i})}(\tau_{\boldsymbol{\psi}}=\infty)\,\bar{u}^i + \mathbb{E}^{(s,s_{\boldsymbol{v}}^{-i})}[\mathbb{1}\{\tau_{\boldsymbol{\psi}}<\infty\}((1-\beta^{\tau_{\boldsymbol{\psi}}+1})\bar{u}^i + \beta^{\tau_{\boldsymbol{\psi}}+1}\underline{u}^i)]\,.
\end{aligned}$$

根据条件1，$\mathbb{E}^{(s,s_{\boldsymbol{v}}^{-i})}[\tau_{\boldsymbol{\psi}}]<\infty$，因此特别地，$\mathbb{P}^{(s,s_{\boldsymbol{v}}^{-i})}(\tau_{\boldsymbol{\psi}}=\infty)=0$。此外，$\mathbb{1}\{\tau_{\boldsymbol{\psi}}<\infty\}\leqslant 1$，所以我们得到

$$U^i(s, s_{\boldsymbol{v}}^{-i}) \leqslant \mathbb{E}^{(s,s_{\boldsymbol{v}}^{-i})}[(1-\beta^{\tau_{\boldsymbol{\psi}}+1})\bar{u}^i + \beta^{\tau_{\boldsymbol{\psi}}+1}\underline{u}^i] = \bar{u}^i - \mathbb{E}^{(s,s_{\boldsymbol{v}}^{-i})}[\beta^{\tau_{\boldsymbol{\psi}}+1}(\bar{u}^i-\underline{u}^i)]$$

由于 $\tau\mapsto -\beta^{\tau}(\bar{u}^i-\underline{u}^i)$是凹函数，根据詹森不等式，我们有

$$U^i(s, s_{\boldsymbol{v}}^{-i}) \leqslant \bar{u}^i - \beta^{\mathbb{E}^{(s,s_{\boldsymbol{v}}^{-i})}[\tau_{\boldsymbol{\psi}}]}(\bar{u}^i-\underline{u}^i) \leqslant \bar{u}^i - \beta^{\tau^{(\varepsilon)}}(\bar{u}^i-\underline{u}^i)\,,$$

其中我们记 $\tau^{(\varepsilon)} = \sup_{\mathbf{s}\in\mathfrak{S}(\mathbf{w}_{\boldsymbol{v}},\varepsilon)}\mathbb{E}^{\mathbf{s}}[\tau_{\boldsymbol{\psi}}]$。最后，

$$U^i(s, s_{\boldsymbol{v}}^{-i}) \leqslant (1-\gamma)v^i \leqslant U^i(\mathbf{s}_{\boldsymbol{v}}) \leqslant U^i(\mathbf{s}_{\boldsymbol{v}}) + (\gamma+\varepsilon)\,,$$

其中，我们在第一个不等式中使用了假设 $\beta^{\tau^{(\varepsilon)}} \geqslant (\bar{u}^i-(1-\gamma)v^i)(\bar{u}^i-u^i)^{-1}$，在第二个不等式中使用了方程（27），并在最后一个不等式中使用了 $\gamma+\varepsilon\geqslant 0$。

现在，假设$\mathbf{s}\in\mathrm{B}^1(\mathbf{w}_{\boldsymbol{v}},\varepsilon)\times\ldots\times\mathrm{B}^N(\mathbf{w}_{\boldsymbol{v}},\varepsilon)$。根据命题1和（27），对于任意玩家 $i\in[N]$，我们有

$$U^i(s, s_{\boldsymbol{v}}^{-i}) \leqslant v^i+\varepsilon = (1-\gamma)v^i + (\gamma\, v^i+\varepsilon) \leqslant U^i(\mathbf{s}_{\boldsymbol{v}}) + (\gamma+\varepsilon)\,. \tag{28}$$

因此，$\mathbf{s}_{\boldsymbol{v}}$ 是一个 $(\gamma+\varepsilon)$-纳什均衡，从而完成了证明。 □

引理 2 （下界关于 $\kappa_{\boldsymbol{\phi}}$）。 假设条件 *3* 对某个 $(p_L, q_L) \in (0,1)^2$ 成立。固定 $t \geqslant 0$以及一个公共历史 $h_t \in \mathsf{H}$。则存在一个序列 $(U_k)_{k \geqslant k_t}$，由独立同分布的 均匀分布 $([0,1])$ 随机变量组成，并且独立于所有 $i \in [N]$ 的过程 $(\hat{w}_k^i)_{k \geqslant k_t}$，使得如下结论成立。

定义，对于 $k \geqslant k_t$,

$$\bar{\varphi}_k = \mathbb{1}\{U_k \leqslant p_L\}, \qquad \bar{\kappa} = \inf\{k \geqslant k_t \colon \bar{\varphi}_k = 1\},$$

并回顾 $\kappa = \min\{{}^{\kappa}\phi^1, \ldots, {}^{\kappa}\phi^N\}$。那么，在给定 h_t的条件下

1. $\kappa_{\boldsymbol{\phi}} \geqslant \bar{\kappa}$ 几乎必然成立，且 $\kappa_{\boldsymbol{\phi}}$ 与 $\bar{\kappa}$独立；
2. $\bar{\kappa}$ ~服从 Geometric(p_L) 分布。

证明。 固定 $t \geqslant 0$ 和 $\mathrm{h}_t \in \mathsf{H}$。对于每个玩家 $i \in [N]$ 和每个 $k \geqslant k_t$，定义

$$\phi_k^i = \mathbb{1}\{\|\hat{w}_k^i - w_{\boldsymbol{v}}^i\|_1 \geqslant \delta\}.$$

根据构造，给定 h_t, ϕ_k^i 时是参数为

$$p_k = \mathbb{P}\Big(\|\hat{w}_k^i - w_{\boldsymbol{v}}^i\|_1 \geqslant \delta \,\Big|\, \mathrm{h}_t\Big).$$

根据条件 3，对于所有 $k \geqslant k_t$，我们有 $p_k \leqslant p_L$。现在我们构造一个耦合。设 $(U_k)_{k \geqslant k_t}$ 是一列独立同分布的 Uniform($[0,1]$) 随机变量，且对所有 $i \in [N]$，它们与 $(\hat{w}_k^i)_{k \geqslant k_t}$ 相互独立。定义

$$\varphi_k^i = \mathbb{1}\{U_k \leqslant p_k\}, \qquad \bar{\varphi}_k = \mathbb{1}\{U_k \leqslant p_L\}.$$

那么，给定 h_t，我们有

$$\phi_k^i \overset{\mathrm{d}}{=} \varphi_k^i, \qquad \text{and} \qquad \varphi_k^i \leqslant \bar{\varphi}_k \quad \text{almost surely}.$$

定义相应的停时

$$\kappa_{\boldsymbol{\phi}^i} = \inf\{k \geqslant k_t \colon \phi_k^i = 1\}, \qquad \bar{\kappa} = \inf\{k \geqslant k_t \colon \bar{\varphi}_k = 1\}.$$

根据上述耦合不等式，我们得到

$$\kappa_{\boldsymbol{\phi}^i} \overset{\mathrm{d}}{=} \inf\{k \geqslant k_t \colon \varphi_k^i = 1\} \geqslant \bar{\kappa} \qquad \text{almost surely}.$$

由于这对每个 $i \in [N]$ 都成立，因此可得

$$\kappa = \min_{i \in [N]} \kappa_{\boldsymbol{\phi}^i} \geqslant \bar{\kappa} \quad \text{almost surely}.$$

此外，$\bar{\kappa}$ 仅是 $(U_k)_{k \geqslant k_t}$ 的可测函数，而 κ 是 $(\hat{w}_k^i)_{k \geqslant k_t}$ 的可测函数。根据构造，这些集合相互独立，因此 κ 和 $\bar{\kappa}$ 相互独立。

最后，由于 $(\bar{\varphi}_k)_{k \geqslant k_t}$ 是独立同分布的 Bernoulli(p_L) 随机变量，立即可得

$$\bar{\kappa} \sim \mathrm{Geometric}(p_L),$$

至此即完成证明。 □

推论 2. 考虑与定理*3*相同且带有 $\gamma=\varepsilon$的设定，以及所得策略$\mathbf{s}_{\boldsymbol{v}}$，其定义见(5)，检验见(14)。此外，设 $\boldsymbol{v}\in U^1\times\ldots\times U^N$为一个可行收益剖面，使得 $v^i\geqslant u^i$ 对 $i\in[N]$成立。若对任意 $i\in[N]$

$$\bar{\beta}^i=\frac{\bar{u}^i-(1-\varepsilon)v^i}{\bar{u}^i-\underline{u}^i}\in[0,1]\quad and\quad \beta^{\tilde{\tau}_\varepsilon}\geqslant\bar{\beta}^i\,, \tag{17}$$

其中 $\tilde{\tau}_\varepsilon$ 如推论 *1*所定义。于是，成立

(i) 对任意 $i\in[N]$，$(1-\varepsilon)v^i\leqslant U^i(\mathbf{s}_{\boldsymbol{v}})\leqslant v^i$，

(ii) $\mathbf{s}_{\boldsymbol{v}}$ 是一个$(2\varepsilon,\mathfrak{S})$-纳什均衡，其中 $\mathfrak{S}$ 的定义如(15)所示。

推论 *2*的证明。 该结果直接来自以 $\gamma=\varepsilon$实例化定理2，并利用定理3中对 $\tau^{(\varepsilon)}$ 的界。

□

C.1 基于e过程的先检验后惩罚策略的证明

命题2。 令 $\gamma\in(0,1)$，并对任意 $i\in[N]$，考虑在(14)中定义的检验序列$(\tilde{\psi}_t^i)_{t\geqslant1}$以及由(5)定义的相应策略$\mathbf{s}_{\boldsymbol{v}}$ 。则条件*1*成立。

证明。 这是 Ville 不等式（定理5）的直接推论。 □

C.1.1 第一类错误。

回顾公式（4），每个玩家 j 都旨在对任意 $i\neq j$进行检验

$$\mathrm{H}_0^i\colon\, s^i=\sigma^i_{\boldsymbol{v}}\qquad\text{versus}\qquad \mathrm{H}_1^i\colon\, s^i\neq\sigma^i_{\boldsymbol{v}}\,.$$

H_0^i 为真意味着玩家 i 坚持采用合作策略 $w^i_{\boldsymbol{v}}$。相反，拒绝 H_0^i 意味着认定玩家 i 没有采用约定策略。还要回顾一下，我们用 $\mathbb{E}_0^i$ 和 $\mathbb{P}_0^i$ 表示在 H_0^i下的概率和期望。现在，对任意玩家 $i\in\{1,\ldots,N\}$，我们引入滤过$\{\mathcal{F}_t^i\}_{t\geqslant0}$，如下

$$\mathcal{F}_t^i=\sigma(\{A_s^i\}_{0\leqslant s\leqslant t})\,.$$

按照惯例，我们还定义$\mathcal{F}_{-1}^i=\{\varnothing,\ \Omega\}$。

定理 6. 在H_0^i 所定义的原假设下，$(E_t^i)_{t\geq0}$ 中(13) 定义的过程是一个非负的 ***{F_t^i}***-鞅，其中 $\mathbb{E}_0^i[E_t^i]=1$对任意 $t\geqslant0$均成立。因此，根据 *Ville* 不等式

$$\mathbb{P}_0^i\left(\sup_{t\geq0}E_t^i\ \geqslant1/\gamma\right)\leqslant\gamma\,.$$

特别地，停止规则“在第一个满足 $E_t^i\geq1/\gamma$ 的 t 时拒绝 H_0^i ”将第一类错误控制在水平 γ。

定理 6 的证明。由于 $\widehat{w}_t^i$ 是 $\mathcal{F}_{t-1}^i$-可测的，利用事件$\{A_t^i = a\}_{a\in\mathcal{A}}$下的全概率公式，我们得到

$$\mathbb{E}_0^i\left[\frac{\widehat{w}_t^i(A_t^i)}{w_{\boldsymbol{v}}(A_t^i)} \,\middle|\, \mathcal{F}_{t-1}^i\right] = \sum_{a\in\mathcal{A}} \frac{\widehat{w}_t^i(a)}{w_{\boldsymbol{v}}(a)} \mathbb{P}_0^i\left(A_t^i = a \mid \mathcal{F}_{t-1}^i\right) = \sum_{a\in\mathcal{A}} \widehat{w}_t^i(a) = 1\,,$$

其中，第二个等号利用了在 H_0^i, $\mathbb{P}_0^i(A_t^i = a \mid \mathcal{F}_{t-1}^i)$ 下 $= w_{\boldsymbol{v}}(a)$ 这一事实。因此，由于 E_{t-1}^i 是 $\mathcal{F}_{t-1}^i$-可测的，我们有

$$\mathbb{E}_0^i\left[E_t^i \mid \mathcal{F}_{t-1}^i\right] = E_{t-1}^i\ \mathbb{E}_0^i\left[\frac{\widehat{w}_t^i(A_t^i)}{w_{\boldsymbol{v}}(A_t^i)} \,\middle|\, \mathcal{F}_{t-1}^i\right] = E_{t-1}^i\,,$$

这表明 $\{E_t^i\}_{t\geqslant 0}$ 是一个单位均值的非负鞅（因为 $E_0^i = 1$）。对非负上鞅应用 Ville 不等式，可得 $\mathbb{P}_0^i(\sup_{t\geqslant 0} E_t^i \geqslant 1/\gamma) \leqslant \gamma$，这意味着该阈值规则具有所声称的任意时刻有效的第一类错误控制。

□

C.1.2 期望停止时间上界

现在，我们将关于序贯检验的结果表述为一个一般版本，涵盖了本文所考虑的特定博弈情形。设 $(A_t)_{t\geqslant 0}$ 是一列分布为 π、取值于 $\{1,\ldots,K\}$ 且在 $\mathbb{P}_\pi$ 下独立同分布的随机变量。我们用 $\mathbb{E}_\pi$ 表示相应的期望。对于 π_0，即定义在 $\{1,\ldots,K\}$ 上的一个分布，我们旨在检验

$$\mathrm{H}_0 \colon \pi = \pi_0 \quad \text{versus} \quad \mathrm{H}_1 \colon \pi \neq \pi_0\,. \tag{29}$$

定义 e过程（E_t）$_{t\geqslant -1}$，其中 $E_{-1} = 1$，且对于任意 $t \geqslant 0$

$$E_t = \prod_{s=0}^{t} \frac{\widehat{\pi}_s(A_s)}{\pi_0(A_s)}\,, \tag{30}$$

其中，对于任意 $a \in \{1,\ldots,K\}$

$$\hat{\pi}_t(a) = \frac{N_t(a)+1}{t+K} \quad \text{with} \quad N_t(a) = \sum_{s=0}^{t-1} \mathbb{1}\{A_s = a\}\,, \tag{31}$$

约定对于任意 $a \in \{1,\ldots,K\}$，$\hat{}\pi_{-1}(a) = 1/K$。我们在定理 6 中证明，$(E_t)_{t\geqslant -1}$ 在原假设 H_0 下是一个 e过程。对于 $\{1,\ldots,K\}$ 上的任意分布 π，我们将该分布的支撑集定义为

$$\mathrm{supp}(\pi) = \{a \in \{1,\ldots,K\} \text{ such that } \pi(a) > 0\}\,,$$

该分布的最小值定义为

$$\underline{\pi} = \min_{a\in\{1,\ldots,K\}} \pi(a)\,. \tag{32}$$

最后，令 $\mathcal{F}_t = \sigma(A_1、\ldots\ldots、A_t)$，并对任意 $n \in \mathbb{N}$ 定义

$$S_n^1 = \sum_{t=0}^{n-1} \left(\log \widehat{\pi}_{t-1}(A_t) - \mathbb{E}_\pi[\log \widehat{\pi}_{t-1}(A_t) \mid \mathcal{F}_{t-1}]\right).$$

引理 3. 存在普适常数 A_1 和 A_2，它们仅取决于 K、π，使得对于任意 $\delta > \underline{0}$、$n \geqslant 0$，都有

$$\mathbb{P}_\pi\Big(|S_n^1|/n \geqslant \delta\Big) \leqslant A_1 \exp\left(-\frac{n\delta^2}{8A_2\log(n)}\right).$$

引理 *3* 的证明。注意，$(S_n^1)_{n\geqslant 0}$ 是关于$(\mathcal{F}_n)_{n\geqslant 0}$的鞅。根据 Pinelis（1994年，定理4.1），存在一个普适常数 $C_1 > 0$，使得对任意 $p \geqslant 2$

$$\begin{aligned}\mathbb{E}_\pi[|S_n^1|^p]^{1/p} &\leqslant C_1\Big(p\,\mathbb{E}_\pi[|\sup_{0\leqslant t\leqslant n-1}\Delta_t|^p]^{1/p} + \sqrt{p}\,\mathbb{E}_\pi[\Big(\sum_{t=0}^{n-1}\mathbb{E}_\pi[\Delta_t^2|\mathcal{F}_{t-1}]\Big)^{p/2}]^{1/p}\Big)\\ &\leqslant C_1\Big(p\log(n) + \sqrt{p}\,\mathbb{E}_\pi\Big[\Big(\sum_{t=0}^{n-1}\mathbb{E}_\pi[\Delta_t^2|\mathcal{F}_{t-1}]\Big)^{p/2}\Big]^{1/p}\Big),\end{aligned} \tag{33}$$

其中，我们将鞅增量记为 $\Delta_t = \log \hat{}\,\pi_{t-1}(A_t) - \mathbb{E}_\pi[\log \hat{}\,\pi_{t-1}(A_t) \mid \mathcal{F}_{t-1}]$。对任意 $t \geqslant 0, a \in \{1,\dots,K\}$，都有 $N_{t-1}(a) \in \{0,\dots,t-1\}$ 几乎必然成立，因此

$$\hat{\pi}_{t-1}(a) = \frac{N_{t-1}(a)+1}{(t-1)+K} \in \Big[\frac{1}{(t-1)+K}, \frac{t}{(t-1)+K}\Big],$$

因此，$\log\big(1/(K+t-1)\big) \leqslant \log \hat{}\,\pi_{t-1}(A_t) \leqslant \log(t/(K+t-1))$几乎必然成立。对于 $\mathbb{E}_\pi[\log \hat{}\,\pi_{t-1}(A_t) \mid \mathcal{F}_{t-1}]$，同样成立。因此，

$$|\Delta_t| = |\log\hat{\pi}_{t-1}(A_t) - \mathbb{E}_\pi[\log\hat{\pi}_{t-1}(A_t) \mid \mathcal{F}_{t-1}]| \leqslant \log t\,.$$

现在，我们将 $N_t(a)$ 视为参数为 $\pi(a)$、在 $\mathbb{P}_\pi$ 下由 t 个伯努利随机变量之和，并应用 Hoeffding 不等式，得到对于任意 $\delta_a > 0$

$$\mathbb{P}_\pi(\{|N_t(a) - tq(a)| > \delta_a\}) \leqslant 2\exp(-2\delta_a^2/t)\,,$$

因此

$$\mathbb{P}_\pi(|N_{t-1}(a) - (t-1)q(a)| > (t-1)\pi(a)/2 \text{ for any } a \in \mathrm{supp}(\pi)) \leqslant 2K\exp(-(t-1)\underline{\pi}^2/2)\,.$$

因此，以至少 $1 - 2K\exp(-\pi^2(t-1)/2)$ 的概率，对于任意 $a \in \{1,\dots,K\}$，我们同时有 $N_{t-1}(a) \geqslant (t-1)\,\pi(a)/2$。在此事件下，对于任意 $t \geqslant 0$，我们有

$$1 \geqslant \hat{\pi}_{t-1}(a) = \frac{N_{t-1}(a)+1}{t-1+K} \geqslant \frac{(t-1)\pi(a)/2}{t-1+K} \geqslant \pi(a)/4K\,,$$

因此

$$|\log\hat{\pi}_{t-1}(a)| \leqslant \max\{0, |\log(\underline{\pi}/4K)|\} = |\log(\underline{\pi}/4K)|\,.$$

因此，在该事件仍然成立时，我们有

$$|\Delta_t|^2 = |\log\hat{\pi}_{t-1}(A_t) - \mathbb{E}_\pi[\log\hat{\pi}_{t-1}(A_t) \mid \mathcal{F}_{t-1}]|^2 \leqslant 4|\log(\underline{\pi}/4K)|^2\,.$$

在相反事件下，由于对任意 $a \in \mathrm{supp}(\pi)$， $1 \leqslant N_{t-1}(a)+1 \leqslant t$，我们确定性地有$|\log \hat{}\pi_{t-1}(a)| \leqslant \log\, t$，因而 $|\Delta_t|^2 \leqslant 4\log(t)^2$。因此，我们有

$$\mathbb{E}_\pi[|\Delta_t|^2] \leqslant 8\log(t)^2 K e^{-\underline{\pi}^2 t/2} + 4|\log(\underline{\pi}/4K)|^2 .$$

由于对任意 $x \geqslant 0, c>0$，且对任意 $t \geqslant 1$ 和 $\log(1)=0$，均有 $x^2 \mathrm{e}^{-cx} = (x\mathrm{e}^{-cx/2})^2 \leqslant (1/\mathrm{e}c)^2$ 以及 $\log\, t \leqslant t$，因此 $8\log(t)^2 K \mathrm{e}^{-\pi^2 t/2} \leqslant \underline{8}K(2/\mathrm{e}\pi^2)^2 = \underline{32}K/\mathrm{e}^2\pi^4$，从而

$$\mathbb{E}_\pi[|\Delta_t|^2] \leqslant 32K/\mathrm{e}^2\underline{\pi}^4 + 4|\log(\underline{\pi}/4K)|^2 = KC_2^2/\underline{\pi}^4 ,$$

其中 $C_2>0$ 是一个普适常数。回到公式 (33)，我们有，对任意 $p \geqslant 2$

$$\mathbb{E}_\pi[|S_n^1|^p]^{1/p} \leqslant C_1\Big\{p\log(n) + \sqrt{p}\Big(\Big(\sum_{t=0}^{n-1} KC_2^2/\underline{\pi}^4\Big)^{p/2}\Big)^{1/p}\Big\} = C_1\Big\{p\log(n) + \sqrt{K}C_2\sqrt{pn}/\underline{\pi}^2\Big\} . \quad (34)$$

利用马尔可夫不等式，我们可以写出

$$\mathbb{P}_\pi\Big(|S_n^1|/n \geqslant \delta\Big) = \mathbb{P}_\pi\Big(|S_n^1|^p \geqslant \delta^p n^p\Big) \leqslant \mathbb{E}_\pi[|S_n^1|^p]/\delta^p n^p \leqslant \frac{C_1^p\Big\{p\log(n) + \sqrt{K}C_2\sqrt{pn}/\underline{\pi}^2\Big\}^p}{\delta^p n^p} ,$$

由公式 (34) 可得

$$\mathbb{P}_\pi\Big(|S_n^1|/n \geqslant \delta\Big) \leqslant \Big(\frac{C_1\log(n)p}{\delta n} + \frac{\sqrt{K}C_2\sqrt{p}}{\delta\underline{\pi}^2\sqrt{n}}\Big)^p .$$

我们定义 $C = \max\{1, C_1, C_2^2K/q_{\underline{}}^4\}$，并且有

$$\mathbb{P}_\pi\Big(|S_n^1|/n \geqslant \delta\Big) \leqslant \Big(\frac{Cp\log(n)}{\delta n} + \frac{\sqrt{C}\sqrt{p}}{\delta\sqrt{n}}\Big)^p .$$

现在，对于一个 $n \geqslant 8C\log(n)/\delta^2$，取 $p = \delta^2 n/(4C\log n)$，我们有 $p \geqslant 2$，并得到 in

$$\begin{aligned} &C\log(n)p/\delta n = 1/4 , \\ &\sqrt{C}\sqrt{p}/\sqrt{n}\delta = \frac{\sqrt{C}\delta\sqrt{n}}{2\delta\sqrt{n}\sqrt{C\log(n)}} \leqslant 1/4 , \end{aligned} \quad (35)$$

这意味着

$$\mathbb{P}_\pi(|S_n^1|/n \geqslant \delta) \leqslant (1/2)^p = \exp(-p\log 2) = \exp\Big(-\frac{\delta^2 n\log(2)}{4C\log(n)}\Big) \leqslant \exp\Big(-\frac{\delta^2 n}{8C\log(n)}\Big) , \quad (36)$$

对于任意满足 $n \geqslant 8C\log(n)/\delta^2$ 的 n，证明至此完成。 □

备注 1。 注意，对于任意 $t \geqslant 1$，直接应用 Azuma-Hoeffding不等式，并将增量界定为 $\log(t)$，所得结果与引理 3 相当，但不够紧。由于$0 + \sum_{t=1}^{n-1}\log(t)^2 \leqslant n\log(n)^2$（其中第一项取 0）–当 $\varepsilon = 4\sqrt{n}\log(n)^3,$时，这表明对于任意 $\delta > 0$

$$\mathbb{P}_\pi\Big(|S_n^1|/n \geqslant \delta\Big) = \mathbb{P}_\pi\Big(|S_n^1| \geqslant n\delta\Big) \leqslant \exp\left(-\frac{n^2\delta^2}{8n\log(n)^2}\right) = \exp\left(-\frac{n\delta^2}{8\log(n)^2}\right) .$$

引理4. 考虑定义在 $\{1,\dots,K\}$ 上的分布 π，以及相应的拉普拉斯估计量 $\hat{}\pi$，其定义见公式 (31)。对于任意满足 $n \geqslant 32C_{\mathrm{KL}}/(\delta^2\log(8C_{\mathrm{KL}}/\delta))$ 的 n，有

$$\mathbb{P}_\pi\left(\frac{1}{n}\sum_{t=0}^{n-1}\mathrm{KL}(\pi\|\hat{\pi}_{t-1}) \geqslant \delta\right) \leqslant 4n\exp\left(-\frac{n\delta}{8C_{\mathrm{KL}}\log(n)^2}\right).$$

引理4的证明。 我们的目标是利用Mourtada（2025，定理1）来控制 $\sum_{t=0}^{n-1}$KL散度$(\pi\|\hat{}\pi_{t-1})$，该定理指出，对于任意 $t \geqslant 2$ 以及任意 $\delta \in (0,\ 1/4)$，存在常数 $C_{\mathrm{KL}} > 0$，使得对于任意$\eta \in (\exp(-t/6),\ \exp(-2))$

$$\mathbb{P}_\pi\left(\mathrm{KL}(\pi\|\hat{\pi}_{t-1}) > C_{\mathrm{KL}}\,\frac{K+\log(1/\eta)\,\log\log(1/\eta)}{t-1}\right) \leqslant 4\eta\,. \tag{37}$$

令 $x = (t-1)\delta/2C_{\mathrm{KL}}$，并假设 $t \geqslant \max\{2C_{\mathrm{KL}}/(\delta\exp(-3\delta/C_{\mathrm{KL}}))\ 4C_{\mathrm{KL}}K/\delta\}$ 以及$K \geqslant 10 > e^2$成立。这意味着 $x \geqslant K = \max\{K,\ e^2\}$，或等价地$(t-1)\delta/2C_{\mathrm{KL}} \geqslant \max\{K,\ e^2\} = K$，且 $t \geqslant 8C_{\mathrm{KL}}/\delta$。由此可得，一旦我们定义 $\eta = \exp(-x/\log\ x)$，便有 $\eta \in (\exp(-t/6),\ \exp(-2))$，从而可对该 η应用公式（37）。现在，由于此处$\log(x/\log(x)) \leqslant \log(x)$，我们有

$$\begin{aligned}
C_{\mathrm{KL}}\frac{K+\log(1/\eta)\,\log\log(1/\eta)}{t-1} &= C_{\mathrm{KL}}\frac{K+x/\log(x)\times\log(x/\log(x))}{t-1}\\
&\leqslant C_{\mathrm{KL}}\frac{K+x}{t-1}\\
&\leqslant 2C_{\mathrm{KL}}\frac{x}{t-1}\\
&= \frac{2C_{\mathrm{KL}}}{t-1}\frac{(t-1)\delta}{2C_{\mathrm{KL}}}\\
&= \delta\,,
\end{aligned}$$

因此，我们得到

$$\mathbb{P}_\pi\big(\mathrm{KL}(\pi\|\hat{\pi}_{t-1}) \geqslant \delta\big) \leqslant \mathbb{P}_\pi\left(\mathrm{KL}(\pi\|\hat{\pi}_{t-1}) > C_{\mathrm{KL}}\frac{K+\log(1/\eta)\,\log\log(1/\eta)}{t-1}\right) \leqslant 4\eta = 4\exp(-x/\log x)\,.$$

因此，对于任意 $t \geqslant t_{\min} = \max\{2C_{\mathrm{KL}}\exp(3\delta/C_{\mathrm{KL}})/\delta,\ 4C_{\mathrm{KL}}K/\delta\} = 4KC_{\mathrm{KL}}/\delta$ for $\delta \leqslant C_{\mathrm{KL}}/3$, 我们有

$$\mathbb{P}_\pi\big(\mathrm{KL}(\pi\|\hat{\pi}_{t-1}) \geqslant \delta\big) \leqslant 4\exp\left(-\frac{(t-1)\delta/(2C_{\mathrm{KL}})}{\log\left((t-1)\delta/(2C_{\mathrm{KL}})\right)}\right), \tag{38}$$

因此，我们定义

$$t_{\min} = 4KC_{\mathrm{KL}}/\delta\,. \tag{39}$$

根据平滑拉普拉斯估计量的定义，对于任意 $t \in \mathbb{N}$，$\mathrm{KL}(\pi\|\hat{}\pi_{t-1}) \leqslant \log(t)$，因此

$$\begin{aligned}
\left\{\frac{1}{n}\sum_{t=0}^{n-1}\mathrm{KL}(\pi\|\hat{\pi}_{t-1}) \geqslant \delta\right\} &\subseteq \left\{\frac{1}{n}\sum_{t=t_0}^{n-1}\mathrm{KL}(\pi\|\hat{\pi}_{t-1}) \geqslant \delta - t_0\log(t_0)/n\right\}\\
&\subseteq \bigcup_{t=t_0}^{n-1}\{\mathrm{KL}(\pi\|\hat{\pi}_{t-1}) \geqslant \delta_t\}\,,
\end{aligned}$$

对于任意序列 $(\delta_t)_{t\in[t_0,n-1]} \in (0,\ 1)^{n-t_0}$，满足 $\sum_{t=t_0}^{n-1}\delta_t = n\delta - t_0\log(t_0)$。因此，根据并集界可得

$$\mathbb{P}_\pi\left(\frac{1}{n}\sum_{t=0}^{n-1}\mathrm{KL}(\pi\|\hat{\pi}_{t-1}) \geqslant \delta\right) \leqslant \sum_{t=t_0}^{n-1}\mathbb{P}_\pi(\mathrm{KL}(\pi\|\hat{\pi}_{t-1}) \geqslant \delta_{t-1})\,.$$

由公式 (38) 中的逐点界，对于任意 $t_0 \in [t_{\min}, n]$，我们得到

$$\mathbb{P}_\pi\left(\frac{1}{n}\sum_{t=0}^{n-1}\mathrm{KL}(\pi\|\hat{\pi}_{t-1}) \geqslant \delta\right) \leqslant 4\sum_{t=t_0}^{n-1}\exp\left(-\frac{(t-1)\delta_{t-1}}{4C_{\mathrm{KL}}\log((t-1)\delta_{t-1}/(4C_{\mathrm{KL}}))}\right).$$

定义 $B(t_0) = (n\delta - t_0\log(t_0))/4C_{\mathrm{KL}}H(t_0,\ n)$，其中 $H(t_0,\ n) = \sum_{t=t_0}^{n-1}1/t$ 为调和级数。现在选择序列 $(\delta_t)^{t\in[t_0+1,n]}$，使得

$$t\delta_t/4C_{\mathrm{KL}} = B(t_0) \text{ and hence } \sum_{t=t_0}^{n-1}\delta_t = 4C_{\mathrm{KL}}B(t_0)/\sum_{t=t_0}^{n-1}1/t = n\delta - t_0\log(t_0)\,.$$

即 $\delta_t = 4C_{\mathrm{KL}}B(t_0)/t$。因此，我们有

$$\exp\left(-\frac{(t-1)\delta_{t-1}}{4C_{\mathrm{KL}}\log((t-1)\delta_{t-1}/(4C_{\mathrm{KL}}))}\right) = \exp(-B(t_0)/\log(B(t_0))\,.$$

因此，我们有

$$\mathbb{P}_\pi\left(\frac{1}{n}\sum_{t=0}^{n-1}\mathrm{KL}(\pi\|\hat{\pi}_{t-1}) \geqslant \delta\right) \leqslant 4(n-t_0)\exp(-B(t_0)/\log(B(t_0)) \leqslant n\exp(-B(t_0)/\log(B(t_0))\,.$$

现在我们选择 t_0，使得 $t_0\log(t_0) = n\delta/2$，也就是说

$$\exp(\log(t_0))\log(t_0) = n\delta/2 \text{ which gives } \log(t_0) = W(n\delta/2)\,,$$

其中 W 为朗伯函数。利用恒等式 $\exp(W(z)) = zW(z)$，可得

$$t_0 = \frac{n\delta}{2W(n\delta/2)}\,,$$

以及满足以下条件的 n

$$n \geqslant \frac{16C_{\mathrm{KL}}}{\delta^2\log(8C_{\mathrm{KL}}/\delta)}\,,$$

这保证了 $t_0 \geqslant t_{\min}$。我们有 $B(t_0) = (n\delta - t_0\log(t_0))/4C_{\mathrm{KL}}H(t_0,\ n) = n\delta/(8C_{\mathrm{KL}}H(t_0,\ n)) \geqslant n\delta/(8C_{\mathrm{KL}}\log(n))$，由此可得

$$\mathbb{P}_\pi\left(\frac{1}{n}\sum_{t=0}^{n-1}\mathrm{KL}(\pi\|\hat{\pi}_{t-1}) \geqslant \delta\right) \leqslant 4n\exp\left(-\frac{n\delta}{8C_{\mathrm{KL}}\log(n)\log(n\delta/(8C_{\mathrm{KL}}\log(n)))}\right),$$

对于任意 $n \geqslant 16C_{\mathrm{KL}}/(\delta^2\log(8C_{\mathrm{KL}}/\delta))$。因此，由于 $\log(n\delta/(8C_{\mathrm{KL}}\log(n))) \leqslant \log(n)$，可得

$$\mathbb{P}_\pi\left(\frac{1}{n}\sum_{t=0}^{n-1}\mathrm{KL}(\pi\|\hat{\pi}_{t-1}) \geqslant \delta\right) \leqslant 4n\exp\left(-\frac{n\delta}{8C_{\mathrm{KL}}\log(n)^2}\right),$$

因此，结论得证。 □

引理 5. 对于定义在 $\{1,\dots,K\}$ 上的任意分布 π_0,π，满足 $\mathrm{supp}(\pi)\subseteq\mathrm{supp}(\pi_0)$，定义 $S_n^2=\sum_{t=0}^{n-1}(\mathbb{E}_\pi[\log\hat{}\pi_{t-1}(A_t)\mid\mathcal{F}_{t-1}]-\log\pi_0(A_t))$。对于任意 $\delta>0$ 和 $n\geqslant 3\delta/\mathrm{KL散度}(\pi\|\pi_0)$，有

$$\mathbb{P}_\pi(S_n^2\leqslant\delta)\leqslant 4n\exp\left(-\frac{n\mathrm{KL}(\pi\|\pi_0)}{32C_{\mathrm{KL}}\log(n)^2}\right)+2\exp\left(-\frac{2n\mathrm{KL}(\pi\|\pi_0)^2}{16|\log\underline{\pi}_0|}\right).$$

引理 5 的证明。我们分解S_n^2 为

$$\begin{aligned}
S_n^2&=\sum_{t=0}^{n-1}(\mathbb{E}_\pi[\log\hat{\pi}_{t-1}(A_t)\mid\mathcal{F}_{t-1}]-\mathbb{E}_\pi[\log\pi_0])+\sum_{t=0}^{n-1}(\mathbb{E}_\pi[\log\pi_0]-\log\pi_0(A_t))\\
&=\sum_{t=0}^{n-1}(\mathbb{E}_\pi[\log\hat{\pi}_{t-1}(A_t)\mid\mathcal{F}_{t-1}]-\mathbb{E}_\pi[\log\pi])+\sum_{t=0}^{n-1}(\mathbb{E}_\pi[\log\pi]-\mathbb{E}_\pi[\log\pi_0])\\
&\quad+\sum_{t=0}^{n-1}(\mathbb{E}_\pi[\log\pi_0]-\log\pi_0(A_t)))\\
&=\sum_{t=0}^{n-1}(\mathbb{E}_\pi[\log\hat{\pi}_{t-1}(A_t)\mid\mathcal{F}_{t-1}]-\mathbb{E}_\pi[\log\pi])+n\,\mathrm{KL}(\pi\|\pi_0)+\sum_{t=0}^{n-1}(\mathbb{E}_\pi[\log\pi_0]-\log\pi_0(A_t)).
\end{aligned}$$

对于任意满足 KL散度$(\pi\|\pi_0)\geqslant 2\delta/n$ 的 $n\in\mathbb{N}^\star$，我们因此有

$$\begin{aligned}
&\mathbb{P}_\pi(S_n^2\leqslant\delta)\\
&=\mathbb{P}_\pi\left(\frac{1}{n}\sum_{t=0}^{n-1}(\mathbb{E}_\pi[\log\hat{\pi}_{t-1}(A_t)\mid\mathcal{F}_{t-1}]-\mathbb{E}_\pi[\log\pi])+\mathrm{KL}(\pi\|\pi_0)+\frac{1}{n}\sum_{t=0}^{n-1}(\mathbb{E}_\pi[\log\pi_0]-\log\pi_0(A_t))\leqslant\delta/n\right) \qquad (40)\\
&=\mathbb{P}_\pi\left(\frac{1}{n}\sum_{t=0}^{n-1}(\mathbb{E}_\pi[\log\hat{\pi}_{t-1}(A_t)\mid\mathcal{F}_{t-1}]-\mathbb{E}_\pi[\log\pi])+\frac{1}{n}\sum_{t=0}^{n-1}(\mathbb{E}_\pi[\log\pi_0]-\log\pi_0(A_t))\leqslant\delta/n-\mathrm{KL}(\pi\|\pi_0)\right)\\
&\leqslant\mathbb{P}_\pi\left(\left|\frac{1}{n}\sum_{t=0}^{n-1}(\mathbb{E}_\pi[\log\hat{\pi}_{t-1}(A_t)\mid\mathcal{F}_{t-1}]-\mathbb{E}_\pi[\log\pi])+\frac{1}{n}\sum_{t=0}^{n-1}(\mathbb{E}_\pi[\log\pi_0]-\log\pi_0(A_t))\right|\geqslant\mathrm{KL}(\pi\|\pi_0)-\delta/n\right)\\
&\leqslant\mathbb{P}_\pi\left(\left|\frac{1}{n}\sum_{t=0}^{n-1}(\mathbb{E}_\pi[\log\hat{\pi}_{t-1}(A_t)\mid\mathcal{F}_{t-1}]-\mathbb{E}_\pi[\log\pi])+\frac{1}{n}\sum_{t=0}^{n-1}(\mathbb{E}_\pi[\log\pi_0]-\log\pi_0(A_t))\right|\geqslant\mathrm{KL}(\pi\|\pi_0)/2\right)\\
&\leqslant\mathbb{P}_\pi\left(\left|\frac{1}{n}\sum_{t=0}^{n-1}(\mathbb{E}_\pi[\log\hat{\pi}_{t-1}(A_t)\mid\mathcal{F}_{t-1}]-\mathbb{E}_\pi[\log\pi])\right|\geqslant\mathrm{KL}(\pi\|\pi_0)/4\right) \qquad (41)\\
&\quad+\mathbb{P}_\pi\left(\left|\frac{1}{n}\sum_{t=0}^{n-1}(\mathbb{E}_\pi[\log\pi_0]-\log\pi_0(A_t))\right|\geqslant\mathrm{KL}(\pi\|\pi_0)/4\right) \qquad (42)
\end{aligned}$$

将引理4应用于 π，可给出方程（41）的上界，如下

$$\begin{aligned}
\mathbb{P}_\pi\left(\left|\frac{1}{n}\sum_{t=0}^{n-1}(\mathbb{E}_\pi[\log\hat{\pi}_{t-1}(A_t)\mid\mathcal{F}_{t-1}]-\mathbb{E}_\pi[\log\pi])\right|\geqslant\mathrm{KL}(\pi\|\pi_0)/4\right)&=\mathbb{P}_\pi\left(\frac{1}{n}\sum_{t=0}^{n-1}\mathrm{KL}(\hat{\pi}_t\|\pi)\geqslant\mathrm{KL}(\pi\|\pi_0)/4\right)\\
&\leqslant 4n\exp\left(-\frac{n\mathrm{KL}(\pi\|\pi_0)}{32C_{\mathrm{KL}}\log(n)^2}\right). \qquad (43)
\end{aligned}$$

$(A_t)_{t\leqslant n}$ 在方程（42）右侧的是来自 π 的独立同分布抽样。因此，右侧是独立同分布随机变量之和。在满足 $\mathrm{supp}(\pi)\subseteq\mathrm{supp}(\pi_0)$ 的 π 下，我们有$|\log\pi_0(A_t)|\leqslant\max\{|\log\pi_0|,\ |\log 1|\}=|\log\pi_0|$，且各加项 $\mathbb{E}_\pi[\log\pi_0]-\log\pi_0(A_t)$ 满足

$|\mathbb{E}_\pi[\log\ \pi_0] - \log\ \pi_0(A_t)| \leqslant 2\|\log_\pi_0\|_\infty$。另请注意，$\mathbb{E}_\pi[\mathbb{E}_\pi[\log\ \pi_0] - \log\ \pi_0(A_t)] = 0$。因此，对任意 $\delta > 0$，将 Hoeffding 不等式应用于方程（42）可得

$$\mathbb{P}_\pi\left(\Big|\sum_{t=0}^{n-1}(\mathbb{E}_\pi[\log\pi_0] - \log\pi_0(A_t))\Big| \geqslant n\mathrm{KL}(\pi\|\pi_0)/2\right) \leqslant 2\exp\left(-\frac{2n\mathrm{KL}(\pi\|\pi_0)^2}{16|\log\underline{\pi}_0|}\right). \tag{44}$$

因此，将 S_n^2 的分解、方程（43）和方程（44）合并起来，我们得到，对于任意 $n \geqslant 2\delta/\mathrm{KL}(\pi\|\pi_0)$

$$\mathbb{P}_\pi(S_n^2 \leqslant \delta) \leqslant 4n\exp\left(-\frac{n\mathrm{KL}(\pi\|\pi_0)}{32C_{\mathrm{KL}}\log(n)^2}\right) + 2\exp\left(-\frac{2n\mathrm{KL}(\pi\|\pi_0)^2}{16|\log\underline{\pi}_0|}\right).$$

□

定义停止时间 $\tau^\infty = \inf\{t \geqslant 1:\ \ A_t \notin \mathrm{supp}(\pi_0)\}$。显然可以观察到$\tau^\infty \geqslant \tau_\gamma + 1$，因为在事件 $\{\tau^\infty < \infty\}$上，根据插件型 e-过程的构造，并且由方程（30）可知

$$E_{\tau^\infty+1} = \prod_{s=0}^{\tau^\infty+1}\frac{\hat{\pi}_{s-1}(A_s)}{\pi_0(A_s)} = \frac{\hat{\pi}_{\tau^\infty}(A_{\tau^\infty})}{\pi_0(A_{\tau^\infty})}\prod_{s=0}^{\tau^\infty}\frac{\hat{\pi}_{s-1}(A_s)}{\pi_0(A_s)} = +\infty\,,$$

因为 $A_{\tau^\infty} \notin \mathrm{supp}(\pi_0)$，因此 $\pi_0(A_{\tau^\infty}) = 0$，而 $\hat{\pi}_{\tau^\infty}(A_{\tau^\infty}) > 0$。因此，$S_{\tau^\infty+1} = +\infty \geqslant 1/\gamma$，这意味着 $\tau^\infty \geqslant \tau_\gamma + 1$。第二个观察是，如果 $\mathrm{supp}(\pi) \subseteq \mathrm{supp}(\pi_0)$，那么 $\tau^\infty = +\infty$ 几乎必然成立，因为在这种情况下，$(A_t)_{t\geqslant 0}$ 几乎从不取值于 $\{1,\ldots,K\}\setminus\mathrm{supp}(\pi_0)$。此外，我们看到$\tau^\infty$ 服从参数为 $\sum_{a\in\{1,\ldots,K\}\setminus\mathrm{supp}(\pi_0)}\pi(a)$ 的几何分布，因为 $(A_t)_{t\geqslant 1}$ 是独立同分布的，并且在每一步 t 以概率 $\sum_{a\in\mathrm{supp}(\pi_0)}\pi(a)$ 落入 $\mathrm{supp}(\pi_0)$，以概率 $\sum_{a\in\{1,\ldots,K\}\setminus\mathrm{supp}(\pi_0)}\pi(a)$ 落入 $\{1,\ldots,K\}\setminus\mathrm{supp}(\pi_0)$。因此，在 $\mathrm{supp}(\pi) \not\subseteq \mathrm{supp}(\pi_0)$ 的情况下，我们得到

$$\mathbb{E}_\pi[\tau^\infty] = \frac{1}{\sum_{a\in\{1,\ldots,K\}\setminus\mathrm{supp}(\pi_0)}\pi(a)} < \infty\,.$$

定理 7。设 π_0 为定义在 $\{1,\ldots,K\}$ 上的离散分布，且 $(A_t)_{t\geqslant 0}$ 是在 π 下进行的独立同分布抽样。我们考虑方程 (29) 中定义的原假设 H_0，以及方程 (30) 中定义的 e过程 $(E_t)_{t\geqslant 0}$。对于第一类错误率 $\gamma \in (0,1)$，存在普适常数 $C_1, C_2 > 0$，使得对于任意满足 $n \geqslant 9\delta/2\mathrm{KL}(\pi\|\pi_0)$ 的 n

$$\mathbb{P}_\pi(S_n/n \leqslant \delta) \leqslant C_1\pi(\mathrm{supp}(\pi_0))^{n+1}\left\{n\exp\left(-\frac{n\mathrm{KL}(\pi\|\pi_0)}{C_2\log(n)^2}\right) + \exp\left(-\frac{n\mathrm{KL}(\pi\|\pi_0)^2}{8|\log\underline{\pi}_0|}\right)\right\}$$
$$+ C_1\exp\left(-\frac{n\mathrm{KL}(\pi\|\pi_0)^2}{C_2\log(n)}\right).$$

定理 7 的证明。首先注意，当 $\mathrm{supp}(\pi_0) \cap \mathrm{supp}(\pi) = \varnothing$ 时，停止时间几乎必然等于 1。因此，在本证明的其余部分中，我们假设这一极端情形不成立。回顾一下，对于任意 $t \geqslant 0,\ \ a \in \{1,\ldots,K\}$，我们有

$$\hat{\pi}_{t-1}(a) = \frac{N_{t-1}(a)+1}{(t-1)+K} \qquad , \qquad E_t = \prod_{s=1}^{t}\frac{\hat{\pi}_{s-1}(A_s)}{\pi_0(A_s)} \qquad , \qquad \mathcal{F}_t = \sigma(A_1,\ldots,A_t)\,,$$

且 $\hat{}\pi_{-1}(a) = 1/K$。定义

$$\tau_\gamma = \inf\{t \geqslant 0 \colon\ E_t \geqslant 1/\gamma\}\,,$$
$$X_t = \log \hat{\pi}_{t-1}(A_t) - \log \pi_0(A_t) \qquad , \qquad S_t = \sum_{s=0}^{t-1} X_s = \log E_t\,.$$

那么，对于任意整数 $n \geqslant 0$

$$\begin{aligned}\mathbb{P}_\pi(\tau_\gamma \geqslant n) &= \mathbb{P}_\pi(\text{for any } t < n, E_t \leqslant 1/\gamma) \\ &= \mathbb{P}_\pi\Big(\max_{1\leqslant k\leqslant n} S_k < \log(1/\gamma)\Big) \\ &\leqslant \mathbb{P}_\pi\big(S_n < \log(1/\gamma)\big)\,, \end{aligned} \tag{45}$$

现在我们将 S_n 分解如下

$$\begin{aligned} S_n &= \sum_{t=0}^{n-1} (\log \hat{\pi}_{t-1}(A_t) - \log \pi_0(A_t)) \\ &= \underbrace{\sum_{t=0}^{n-1} (\log \hat{\pi}_{t-1}(A_t) - \mathbb{E}_\pi[\log \hat{\pi}_{t-1}(A_t) \mid \mathcal{F}_{t-1}])}_{S_n^1} + \underbrace{\sum_{t=0}^{n-1} (\mathbb{E}_\pi[\log \hat{\pi}_{t-1}(A) \mid \mathcal{F}_{t-1}] - \log \pi_0(A_t))}_{S_n^2}\,. \end{aligned}$$

对于任意 $n \in \mathbb{N}$，$\delta > 0$，我们有

$$\begin{aligned}\mathbb{P}_\pi(S_n \leqslant \delta) &= \mathbb{P}_\pi(S_n/n \leqslant \delta/n) \\ &\leqslant \mathbb{P}_\pi(|S_n^1|/n \geqslant \delta/2n, S_n/n \leqslant \delta/n) + \mathbb{P}_\pi(|S_n^1|/n \leqslant \delta/2n, S_n/n \leqslant \delta/n) \\ &\leqslant \mathbb{P}_\pi(|S_n^1|/n \geqslant \delta/2n) + \mathbb{P}_\pi(S_n^2/n \leqslant 3\delta/2n)\,. \end{aligned}$$

注意，根据 τ^∞ 的定义，我们有 $R^2_{\tau^\infty+1} = +\infty$，因此对于任意 $\delta > 0$，有

$$\begin{aligned}\mathbb{P}_\pi(S_n^2 \leqslant 3\delta/2) &= \mathbb{P}_\pi(\{\tau^\infty > n\} \cap \{S_n^2 < 3\delta/2\}) \\ &= \mathbb{P}_\pi(\{\tau^\infty > n\})\mathbb{P}_\pi(S_n^2 < 3\delta/2 | \{\tau^\infty > n\})\,. \end{aligned}$$

在给定 $\{\tau^\infty > n\}$ 的条件下，$(A_t)_{t\leqslant n}$ 服从独立同分布 $\tilde{\pi}$，它是在 π_0 的支撑集上对 π 进行重新缩放所得的分布。具体而言，根据贝叶斯定理，对于任意 $a \in \{\mathbf{1}, \ldots, K\}$，我们定义 $\tilde{\pi}$ 为

$$\tilde{\pi}(a) = \frac{\pi(a)\mathbb{1}\{a \in \operatorname{supp}(\pi_0)\}}{\sum_{a' \in \operatorname{supp}(\pi_0)} \pi(a')}\,.$$

由于 τ^∞ 服从参数为 $\sum_{a\in\operatorname{supp}(\pi)/\operatorname{supp}(\pi_0)} \pi(a)$ 的几何分布，我们有

$$\mathbb{P}_\pi(\tau^\infty > n) = (1 - \sum_{a\in\operatorname{supp}(\pi)/\operatorname{supp}(\pi_0)} \pi(a))^{n+1} = \pi(\operatorname{supp}(\pi_0))^{n+1}\,.$$

对于任意$\mathcal{F}_n$-可测的 F，我们有

$$\mathbb{P}_\pi(F(A_1, \ldots, A_n) \mid \tau^\infty > n) = \mathbb{P}_{\tilde{\pi}}(F(A_1, \ldots, A_n))\,, \tag{46}$$

因此，我们可以写成

$$\mathbb{P}_\pi(S_n^2/n \leqslant 3\delta/2n) \leqslant \mathbb{P}_\pi(\tau^\infty > n)\mathbb{P}_{\tilde{\pi}}(S_n^2/n \leqslant 3\delta/2n)\,.$$

由于 $\mathsf{supp}(\tilde{\pi}) \subseteq \mathsf{supp}(\pi)$，我们可以使用引理 5 控制右侧，从而得到：对于任意满足 $n \geqslant 9\delta/2\,\mathrm{KL}(\pi\|\pi_0)$ 的n，我们有

$$\mathbb{P}_\pi(S_n^2/n \leqslant 3\delta/2n) \leqslant \mathbb{P}_\pi(\tau^\infty > n)\left\{4n\exp\left(-\frac{n\mathrm{KL}(\pi\|\pi_0)}{32C_{\mathrm{KL}}\log(n)^2}\right) + 2\exp\left(-\frac{2n\mathrm{KL}(\pi\|\pi_0)^2}{16|\log \underline{\pi}_0|}\right)\right\}.$$

引理 3 保证存在 $A_1, A_2 > 0$，使得

$$\mathbb{P}_\pi\Big(|S_n^1|/n \geqslant \delta\Big) \leqslant A_1\exp\left(-\frac{n\delta^2}{32A_2\log(n)}\right), \tag{47}$$

因此，对于任意满足 $\mathrm{KL}(\pi\|\pi_0)$ 的 $n \geqslant 9\delta/2$，我们有

$$\begin{aligned}\mathbb{P}_\pi\Big(|S_n^1|/n \geqslant \delta/n\Big) &\leqslant \mathbb{P}_\pi\Big(|S_n^1|/n \geqslant \frac{2n\mathrm{KL}(\pi\|\pi_0)}{9}\frac{1}{n}\Big)\\ &= \mathbb{P}_\pi\Big(|S_n^1|/n \geqslant \frac{2\mathrm{KL}(\pi\|\pi_0)}{9}\Big)\\ &\leqslant A_1\exp\left(-\frac{n\mathrm{KL}(\pi\|\pi_0)^2}{648A_2\log(n)}\right),\end{aligned}$$

因此，重新整理各项后，我们得到

$$\begin{aligned}\mathbb{P}_\pi(S_n/n \leqslant \delta) \leqslant \pi(\mathrm{supp}(\pi_0))^{n+1}&\left\{4n\exp\left(-\frac{n\mathrm{KL}(\pi\|\pi_0)}{32C_{\mathrm{KL}}\log(n)^2}\right) + 2\exp\left(-\frac{n\mathrm{KL}(\pi\|\pi_0)^2}{8|\log \underline{\pi}_0|}\right)\right\}\\ &+ A_1\exp\left(-\frac{n\mathrm{KL}(\pi\|\pi_0)^2}{648A_2\log(n)}\right).\end{aligned}$$

因此，存在普适常数 $C_1, C_2 > 0$，使得对于任意满足 $n \geqslant 9\delta/2\mathrm{KL}(\pi\|\pi_0)$ 的 n

$$\begin{aligned}\mathbb{P}_\pi(S_n/n \leqslant \delta) \leqslant C_1\pi(\mathrm{supp}(\pi_0))^{n+1}&\left\{n\exp\left(-\frac{n\mathrm{KL}(\pi\|\pi_0)}{C_2\log(n)^2}\right) + \exp\left(-\frac{n\mathrm{KL}(\pi\|\pi_0)^2}{8|\log \underline{\pi}_0|}\right)\right\}\\ &+ C_1\exp\left(-\frac{n\mathrm{KL}(\pi\|\pi_0)^2}{C_2\log(n)}\right).\end{aligned}$$

□

推论 4. 对于任意 $\gamma \in (0,1)$，考虑停止时间 τ_γ，存在 $C_1, C_2 > 0$，使得对于任意 $n \in \mathbb{N}$

$$\begin{aligned}\mathbb{P}_\pi(\tau_\gamma \geqslant n) \leqslant C_1&\left\{n\exp\left(-n\left[\frac{2\mathrm{TV}(\pi\|\pi_0)^2}{C_2\log(n)^2} + |\log\pi(\mathrm{supp}(\pi_0))|\right]\right) + \exp\left(-n\frac{4\mathrm{TV}(\pi\|\pi_0)^4}{C_2\log(n)}\right)\right.\\ &\left.+\exp\left(-n\left[\frac{\mathrm{TV}(\pi\|\pi_0)^4}{2|\log \underline{\pi}_0|} + |\log\pi(\mathrm{supp}(\pi_0))|\right]\right)\right\} + \mathbb{1}\{n \leqslant 5\log(1/\gamma)/\mathrm{KL}(\pi\|\pi_0)\}\,.\end{aligned}$$

推论 *4* 的证明。 由定理 7 出发，我们可以使用 Pinsker不等式得到

$$\begin{aligned}\mathbb{P}_\pi(S_n/n\leqslant\delta)&\leqslant C_1\pi(\mathrm{supp}(\pi_0))^{n+1}\left\{n\exp\left(-\frac{2n\mathrm{TV}(\pi\|\pi_0)^2}{C_2\log(n)^2}\right)+\exp\left(-\frac{4n\mathrm{TV}(\pi\|\pi_0)^4}{8|\log\underline{\pi}_0|}\right)\right\}\\&\quad+C_1\exp\left(-\frac{4n\mathrm{TV}(\pi\|\pi_0)^4}{C_2\log(n)}\right)\\&\leqslant C_1\exp(n\log(\pi(\mathrm{supp}(\pi_0))))\left\{n\exp\left(-\frac{2n\mathrm{TV}(\pi\|\pi_0)^2}{C_2\log(n)^2}\right)+\exp\left(-\frac{4n\mathrm{TV}(\pi\|\pi_0)^4}{8|\log\underline{\pi}_0|}\right)\right\}\\&\quad+C_1\exp\left(-\frac{4n\mathrm{TV}(\pi\|\pi_0)^4}{C_2\log(n)}\right),\end{aligned}$$

由于 $\log(\pi(\mathrm{supp}(\pi_0)))<0$，该式变为

$$\begin{aligned}\mathbb{P}_\pi(S_n/n\leqslant\delta)\leqslant C_1\Bigg\{&n\exp\left(-n\left[\frac{2\mathrm{TV}(\pi\|\pi_0)^2}{C_2\log(n)^2}+|\log\pi(\mathrm{supp}(\pi_0))|\right]\right)+\exp\left(-n\frac{4\mathrm{TV}(\pi\|\pi_0)^4}{C_2\log(n)}\right)\\&+\exp\left(-n\left[\frac{\mathrm{TV}(\pi\|\pi_0)^4}{2|\log\underline{\pi}_0|}+|\log\pi(\mathrm{supp}(\pi_0))|\right]\right)\Bigg\}.\end{aligned}$$

因此，对于任意满足 $n\geqslant 5\log(1/\gamma)/\mathrm{KL}$散度$(\pi\|\pi_0)$ 的 n，我们可以利用方程（45）得到

$$\begin{aligned}\mathbb{P}_\pi(\tau_\gamma\geqslant n)&\leqslant\mathbb{P}_\pi(S_n<\log(1/\gamma))\\&\leqslant C_1\Bigg\{n\exp\left(-n\left[\frac{2\mathrm{TV}(\pi\|\pi_0)^2}{C_2\log(n)^2}+|\log\pi(\mathrm{supp}(\pi_0))|\right]\right)+\exp\left(-n\frac{4\mathrm{TV}(\pi\|\pi_0)^4}{C_2\log(n)}\right)\\&\quad+\exp\left(-n\left[\frac{\mathrm{TV}(\pi\|\pi_0)^4}{2|\log\underline{\pi}_0|}+|\log\pi(\mathrm{supp}(\pi_0))|\right]\right)\Bigg\},\end{aligned}$$

并且我们可以断定，存在普适常数 C_1，$C_2>0$，使得

$$\begin{aligned}\mathbb{P}_\pi(\tau_\gamma\geqslant n)\leqslant C_1\Bigg\{&n\exp\left(-n\left[\frac{2\mathrm{TV}(\pi\|\pi_0)^2}{C_2\log(n)^2}+|\log\pi(\mathrm{supp}(\pi_0))|\right]\right)+\exp\left(-n\frac{4\mathrm{TV}(\pi\|\pi_0)^4}{C_2\log(n)}\right)\\&+\exp\left(-n\left[\frac{\mathrm{TV}(\pi\|\pi_0)^4}{2|\log\underline{\pi}_0|}+|\log\pi(\mathrm{supp}(\pi_0))|\right]\right)\Bigg\}+\mathbb{1}\{n\leqslant 5\log(1/\gamma)/\mathrm{KL}(\pi\|\pi_0)\},\end{aligned}$$

从而得到该结果。 □

定理3. 考虑 $\varepsilon>0$ 以及水平 $\gamma\in(0,1)$。考虑在（14）中定义的序贯检验 $\{\tilde{\psi}^i\}_{i\in[N]}$。于是，对于 $\tilde{\mathfrak{S}}=\tilde{\mathfrak{S}}^1\times\ldots\times\tilde{\mathfrak{S}}^N$，其中 $\tilde{\mathfrak{S}}^i$在 （15）中定义，且 $\tilde{\tau}_{\boldsymbol{\psi}}=\min_{i\in[N]}\tilde{\tau}^i_\psi$，条件 2成立，且

$$\zeta_t^{(\varepsilon)}=C_1\left\{t\exp\left(-\frac{2t\varepsilon^2}{C_2\log(t)^2}\right)+\exp\left(-t\frac{4\varepsilon^4}{C_2\log(t)}\right)+\exp\left(-\frac{t\varepsilon^4}{2|\log\underline{\mathbf{w}}_{\boldsymbol{v}}|}\right)\right\}+\mathbb{1}\{t\leqslant 10\log(1/\gamma)/\varepsilon^2\},$$

对于任意 $t\geqslant 2,\zeta_0^{(\varepsilon)}=1,\zeta_1^{(\varepsilon)}=1$，其中 $C_1,C_2>0$ 是普适常数，并且我们设 $\mathbf{w}_{\boldsymbol{v}}=\min^{i\in[N]}\min_{a\in\mathcal{A}^i}w^i_{\boldsymbol{v}}[a]$。特别地，我们有 $\sum_{t\geqslant 0}\zeta_t^{(\varepsilon)}<\infty$。

定理3的证明。 设 $\varepsilon>0$ 且 $\mathbf{s}\in\mathfrak{S}(\mathbf{w}_{\boldsymbol{v}},\varepsilon)$，根据定义，存在一个 $i\in[N]$，其采用策略 π^i，使得 $\|\tilde{w}^i-w^i_{\boldsymbol{v}}\|_1\geqslant 2\varepsilon$。考虑这样的玩家 i 及其对应的过程 $(E^i_t)_{t\geqslant 0}$。根据定义，玩家 i 抽取的行动 $(A^i_t)_{t\geqslant 0}\sim\tilde{w}^i$ 独立同分布；由定理 6 可知，这保证了 $(E^i_t)_{t\geqslant 0}$ 是一个 e过程，且与方程（30）给出的表达式一致。因此，根据推论 4，对于任意 $n\geqslant 2$，期望停止时间 $\tilde{\tau}_{\boldsymbol{\psi}^i}$ 满足

$$\mathbb{P}_\pi(\tilde{\tau}_{\boldsymbol{\psi}^i}\geqslant n)\leqslant C_1\left\{n\exp\left(-n\left[\frac{2\mathrm{TV}(\tilde{w}^i\|w_{\boldsymbol{v}})^2}{C_2\log(n)^2}+|\log\tilde{w}^i(\mathrm{supp}(w_{\boldsymbol{v}}))|\right]\right)+\exp\left(-n\frac{4\mathrm{TV}(\tilde{w}^i\|w_{\boldsymbol{v}})^4}{C_2\log(n)}\right)\right.$$
$$\left.+\exp\left(-n\left[\frac{\mathrm{TV}(\tilde{w}^i\|w_{\boldsymbol{v}})^4}{2|\log\underline{w_{\boldsymbol{v}}}|}+|\log\tilde{w}^i(\mathrm{supp}(w_{\boldsymbol{v}}))|\right]\right)\right\}+\mathbb{1}\{n\leqslant 5\log(1/\gamma)/\mathrm{KL}(\tilde{w}^i\|w_{\boldsymbol{v}})\}\,.$$

利用平克斯克不等式对最右项中的 KL散度进行上界估计，并利用 $\mathrm{TV}(\tilde{w}^i-w^i_{\boldsymbol{v}})=2\|\tilde{w}^i-w^i_{\boldsymbol{v}}\|_1\geqslant\varepsilon$ 这一事实，可得我们需要的表达式 $(\zeta^\varepsilon_t)_{t\geqslant 0}$。由于 ζ^ε 是概率的上界，因此可得 $t=0$ 和 $t=1$ 的结果。

□

命题 5. *考虑任意 $\gamma\in(0,1)$ $\eta\in(0,1)$对应的停止时间 τ_γ，存在一个普适常数 $C_\eta>0$，使得*

$$\mathbb{E}_\pi[\tau_\gamma]\leqslant\frac{5\log(1/\gamma)}{\mathrm{KL}(\pi\|\pi_0)}+C_\eta\frac{1+|\log\underline{\pi}_0|^{1/(1-\eta)}}{\mathrm{TV}(\pi\|\pi_0)^{4/(1-\eta)}}\,.$$

命题 5 的证明。对于任意 $\gamma\in(0,1)$，我们有

$$\mathbb{E}_\pi[\tau_\gamma]=\sum_{n\geqslant 0}\mathbb{P}_\pi(\tau_\gamma\geqslant n)\,,$$

并回顾推论 4 可知

$$\mathbb{P}_\pi(\tau_\gamma\geqslant n)\leqslant C_1\left\{n\exp\left(-n\left[\frac{2\mathrm{TV}(\pi\|\pi_0)^2}{C_2\log(n)^2}+|\log\pi(\mathrm{supp}(\pi_0))|\right]\right)+\exp\left(-n\frac{4\mathrm{TV}(\pi\|\pi_0)^4}{C_2\log(n)}\right)\right.$$
$$\left.+\exp\left(-n\left[\frac{\mathrm{TV}(\pi\|\pi_0)^4}{2|\log\underline{\pi}_0|}+|\log\pi(\mathrm{supp}(\pi_0))|\right]\right)\right\}+\mathbb{1}\{n\geqslant 5\log(1/\gamma)/\mathrm{KL}(\pi\|\pi_0)\}\,.$$

因此，我们有

$$\mathbb{E}_\pi[\tau_\gamma]\leqslant 5\log(1/\gamma)/\mathrm{KL}(\pi\|\pi_0)+\sum_{n\geqslant 0}C_1\left\{n\exp\left(-n\left[\frac{2\mathrm{TV}(\pi\|\pi_0)^2}{C_2\log(n)^2}+|\log\pi(\mathrm{supp}(\pi_0))|\right]\right)\right.$$
$$\left.+\exp\left(-n\frac{4\mathrm{TV}(\pi\|\pi_0)^4}{C_2\log(n)}\right)+\exp\left(-n\left[\frac{\mathrm{TV}(\pi\|\pi_0)^4}{2|\log\underline{\pi}_0|}+|\log\pi(\mathrm{supp}(\pi_0))|\right]\right)\right\},$$

其中，第一项来自推论 4 上界中的示性函数。我们可以写成

$$\mathbb{E}_\pi[\tau_\gamma]\leqslant 5\log(1/\gamma)/\mathrm{KL}(\pi\|\pi_0)+C_1\sum_{n\geqslant 0}\exp\left(-n\frac{4\mathrm{TV}(\pi\|\pi_0)^4}{C_2\log(n)}\right)$$
$$+C_1\min\left\{\sum_{n\geqslant 0}\left\{n\exp\left(-n\frac{2\mathrm{TV}(\pi\|\pi_0)^2}{C_2\log(n)^2}\right)+n\exp\left(-n\frac{\mathrm{TV}(\pi\|\pi_0)^4}{2|\log\underline{\pi}_0|}\right),2\exp(-n|\log\pi(\mathrm{supp}(\pi_0))|)\right\}\right.\,.$$

对于任意 $\eta \in (0,1)$，存在 $N_\eta > 0$，使得对于任意 $n \geqslant N_\eta$，我们有

$$\exp\left(-n\frac{4\mathrm{TV}(\pi\|\pi_0)^4}{C_2\log(n)}\right) \leqslant \exp\left(-n^{1-\eta}\frac{4\mathrm{TV}(\pi\|\pi_0)^4}{C_2}\right),$$
$$n\exp\left(-n\frac{2\mathrm{TV}(\pi\|\pi_0)^2}{C_2\log(n)^2}\right) \leqslant \exp\left(-n^{1-\eta}\frac{2\mathrm{TV}(\pi\|\pi_0)^2}{C_2}\right),$$
$$n\exp\left(-n\frac{\mathrm{TV}(\pi\|\pi_0)^4}{2|\log \underline{\pi}_0|}\right) \leqslant \exp\left(-n^{1-\eta}\frac{\mathrm{TV}(\pi\|\pi_0)^4}{2|\log \underline{\pi}_0|}\right),$$

并且我们还可以找到一个常数 $C_\eta > 0$，使得对于任意 $n \in \mathbb{N}$——因为我们可以将常数 C_1 吸收到 C_η 中

$$\exp\left(-n\frac{4\mathrm{TV}(\pi\|\pi_0)^4}{C_2\log(n)}\right) \leqslant C_\eta\exp\left(-n^{1-\eta}\frac{4\mathrm{TV}(\pi\|\pi_0)^4}{C_2}\right),$$
$$n\exp\left(-n\frac{2\mathrm{TV}(\pi\|\pi_0)^2}{C_2\log(n)^2}\right) \leqslant C_\eta\exp\left(-n^{1-\eta}\frac{2\mathrm{TV}(\pi\|\pi_0)^2}{C_2}\right),$$
$$n\exp\left(-n\frac{\mathrm{TV}(\pi\|\pi_0)^4}{2|\log \underline{\pi}_0|}\right) \leqslant C_\eta\exp\left(-n^{1-\eta}\frac{\mathrm{TV}(\pi\|\pi_0)^4}{2|\log \underline{\pi}_0|}\right).$$

对于任意 $c > 0$，我们有

$$\sum_{n\geqslant 0}\mathrm{e}^{-cn^{1-\eta}} \leqslant 1 + \int_{x\geqslant 0}\mathrm{e}^{-cx^{1-\eta}}\mathrm{d}x = 1 + \frac{1}{1-\eta}c^{-1/(1-\eta)}\Gamma(1/(1-\eta)),$$

其中 Γ 表示伽马函数。因此，我们可以将这个不等式应用于总和中的每一项，并得到

$$\begin{aligned}
\mathbb{E}_\pi[\tau_\gamma] &\leqslant 5\log(1/\gamma)/\mathrm{KL}(\pi\|\pi_0) + C_\eta\sum_{n\geqslant 0}\exp\left(-n^{1-\eta}\frac{4\mathrm{TV}(\pi\|\pi_0)^4}{C_2}\right)\\
&\quad + C_\eta\min\left\{\sum_{n\geqslant 0}\left\{\exp\left(-n^{1-\eta}\frac{2\mathrm{TV}(\pi\|\pi_0)^2}{C_2}\right) + \exp\left(-n^{1-\eta}\frac{\mathrm{TV}(\pi\|\pi_0)^4}{2|\log \underline{\pi}_0|}\right), 2\exp(-n|\log\pi(\mathrm{supp}(\pi_0))|)\right\}\right.\\
&\leqslant 3 + 5\log(1/\gamma)/\mathrm{KL}(\pi\|\pi_0) + \frac{C_\eta\Gamma(1/(1-\eta))}{1-\eta}\left\{\frac{4\mathrm{TV}(\pi\|\pi_0)^4}{C_2}\right\}^{-1/(1-\eta)}\\
&\quad + \min\left\{\frac{C_\eta\Gamma(1/(1-\eta))}{1-\eta}\left(\left(\frac{2\mathrm{TV}(\pi\|\pi_0)^2}{C_2}\right)^{-1/(1-\eta)} + \left(\frac{\mathrm{TV}(\pi\|\pi_0)^4}{2|\log \underline{\pi}_0|}\right)^{-1/(1-\eta)}\right), \frac{2}{1-\mathrm{e}^{-|\log\pi(\mathrm{supp}(\pi_0))|}}\right\}\\
&\leqslant 3 + 5\log(1/\gamma)/\mathrm{KL}(\pi\|\pi_0) + \frac{C_\eta\Gamma(1/(1-\eta))C_2^{1/(1-\eta)}}{4^{1/(1-\eta)}(1-\eta)}\frac{1}{\mathrm{TV}(\pi\|\pi_0)^{4/(1-\eta)}}\\
&\quad + \min\left\{\frac{C_\eta\Gamma(1/(1-\eta))}{1-\eta}\left(\frac{C_2^{1/(1-\eta)}}{2^{1/(1-\eta)}}\frac{1}{\mathrm{TV}(\pi\|\pi_0)^{2/(1-\eta)}} + \frac{2^{1/(1-\eta)}|\log \underline{\pi}_0|^{1/(1-\eta)}}{\mathrm{TV}(\pi\|\pi_0)^{4/(1-\eta)}}\right), \frac{2}{1-\mathrm{e}^{-|\log\pi(\mathrm{supp}(\pi_0))|}}\right\}.
\end{aligned}$$

我们现在将所有依赖于 η 以及与该问题独立的普适常数的各项归并在一起。此外，由于我们采用约定全变差$(p\|q) = 1/2\int|p(x) - q(x)|dx$，故有全变差$(\pi\|\pi_0) \leqslant 1$，从而 $1/$全变差$(\pi\|\pi_0)^2 \leqslant 1/$全变差$(\pi\|\pi_0)^4$。我们还有 $\pi(\mathrm{supp}(\pi_0)) \leqslant 1$，因此$-|\log\ \pi(\mathrm{supp}(\pi_0))| = \log\ \pi(\mathrm{supp}(\pi_0))$，这简化了右端项。因此，我们有

存在一个普适常数 $C_\eta > 0$，使得（我们也可以将第一项3以及与该问题独立的乘数吸收到 C_η 中，因为全变差$(\pi\|\pi_0) \leqslant 1$）我们有

$$\mathbb{E}_\pi[\tau_\gamma] \leqslant \frac{5\log(1/\gamma)}{\mathrm{KL}(\pi\|\pi_0)} + \frac{C_\eta}{\mathrm{TV}(\pi\|\pi_0)^{4/(1-\eta)}} + C_\eta \min\left\{ \frac{1+|\log \underline{\pi}_0|^{1/(1-\eta)}}{\mathrm{TV}(\pi\|\pi_0)^{4/(1-\eta)}}, \frac{1}{1-\pi(\mathrm{supp}(\pi_0))} \right\}$$

现在我们终于可以简化最小值，因为表达式中的一项此前已经加上了。因此，对于任意 $\eta \in (0, 1)$，我们终于得到存在一个普适常数 C_η，使得

$$\mathbb{E}_\pi[\tau_\gamma] \leqslant \frac{5\log(1/\gamma)}{\mathrm{KL}(\pi\|\pi_0)} + C_\eta \frac{1+|\log \underline{\pi}_0|^{1/(1-\eta)}}{\mathrm{TV}(\pi\|\pi_0)^{4/(1-\eta)}} .$$

□

推论 1。在与定理*3*相同的设定下，对于任意 $\mathbf{s} \in \mathfrak{S}(\mathbf{w}_{\boldsymbol{v}}, \varepsilon)$，存在一个普适常数 $C > 0$，使得

$$\mathbb{E}^{\mathbf{s}}[\tilde{\tau}_{\boldsymbol{\psi}}] \leqslant \tilde{\tau}_\varepsilon = \frac{10\log(1/\gamma)}{\varepsilon^2} + C\frac{1+|\log \underline{\mathbf{w}}_{\boldsymbol{v}}|}{\varepsilon^5} .$$

推论 *1* 的证明。利用命题2得到停止时间。设 $\varepsilon > 0$，$s \in \mathfrak{S}(\mathbf{w}_{\boldsymbol{v}}, \varepsilon)$。正如我们在定理3的证明中所示，存在一个玩家 $i \in [N]$ 及其策略 π^i，使得 $\mathrm{TV}(\tilde{w}^i - w^i_{\boldsymbol{v}}) = 2\|\tilde{w}^i - w^i_{\boldsymbol{v}}\|_1 \geqslant \varepsilon$。由于 e过程 $(E^i_t)_{t\geqslant 0}$ 满足所有要求的性质，我们可以应用命题 5，并得到对于任意 $\eta \in (0, 1)$，存在一个普适常数 $C_\eta > 0$，使得

$$\mathbb{E}^{\mathbf{s}}\pi[\tau^{(\varepsilon)}] \leqslant \frac{5\log(1/\gamma)}{\mathrm{KL}(\tilde{w}^i\|w^i_{\boldsymbol{v}})} + C_\eta \frac{1+|\log \underline{w^i_{\boldsymbol{v}}}|^{1/(1-\eta)}}{\mathrm{TV}(\pi\|w^i_{\boldsymbol{v}})^{4/(1-\eta)}} \leqslant \frac{10\log(1/\gamma)}{\mathrm{TV}(\tilde{w}^i\|w^i_{\boldsymbol{v}})^2} + C_\eta \frac{1+|\log \underline{w^i_{\boldsymbol{v}}}|^{1/(1-\eta)}}{\mathrm{TV}(\pi\|w^i_{\boldsymbol{v}})^{4/(1-\eta)}} .$$

利用 $\mathrm{TV}(\pi\|w^i_{\boldsymbol{v}}) \geqslant \varepsilon$ 这一事实并选择 $\eta = 1/5$，即可得到结果 . □

关于我们检验最优性的讨论。对于我们的检验 $\phi^i_T = \mathbb{1}\left\{ E^i_T \geqslant 1/\gamma^{(1),\ i} \right\}$ 基于样本 A^i_1、.....、A^i_T，我们可以定义

$$\gamma^{(1),i}_T = \mathbb{P}_{w_{\boldsymbol{v}}}(\phi^i_T = 1), \qquad \gamma^{(2),i}_T = \mathbb{P}_q(\phi^i_T = 0) ,$$

其中 $q \neq w_{\boldsymbol{v}}$。$\gamma^{(2),\ i}_T$ 对应于第二类错误概率（在备择假设q下接受 H^i_0）。注意，对定理 6 中期望停止时间上界的渐近分析表明，在 H^i_1 下，$\log(E^i_T)/T \xrightarrow[T\to\infty]{a.s.} \mathrm{KL}(q\|w_{\boldsymbol{v}})$。因此，对每个 $\varepsilon > 0$，都存在 T_ε，使得对所有 $T \geqslant T_\varepsilon$

$$\mathbb{P}\left(\log E^i_T \geq T\big(\mathrm{KL}(q\|w_{\boldsymbol{v}}) - \varepsilon\big) \right) \underset{T\to\infty}{\longrightarrow} 1 .$$

选择足够大的 T，使得 $\log(1/\gamma) \leq T(\mathrm{KL}(q\|w_{\boldsymbol{v}}) - \varepsilon)$，我们还得到

$$\beta_T = \mathbb{P}(\phi^i_T = 0) = \mathbb{P}\Big(\log E^i_T < \log(1/\gamma) \Big) \leqslant \mathbb{P}\left(\log E^i_T < T\big(\mathrm{KL} - \varepsilon\big) \right) \underset{T\to\infty}{\longrightarrow} 0 ,$$

这意味着

$$\liminf_{T\to\infty} -\frac{\log \beta_T}{T} \geqslant \mathrm{KL}(q\|w_{\boldsymbol{v}}) - \varepsilon\,.$$

由于 $\varepsilon>0$ 是任意的，我们有

$$\liminf_{T\to\infty} -\frac{\log \beta_T}{T} \geqslant \mathrm{KL}(q\|w_{\boldsymbol{v}})\,.$$

根据控制第二类错误概率衰减的 Chernoff-Stein 引理（见引理1），没有水平为 γ 的检验能够达到更大的指数，因此

$$\lim_{T\to\infty} -\frac{1}{T}\log \beta_T = \mathrm{KL}\left(q\|w_{\boldsymbol{v}}\right),$$

这表明 e过程阈值检验是渐近最优的。

C.2 批处理情形下的证明

定理 4. 令 $\varepsilon>0$, $\beta\in(0,1)$，并考虑在 (19) 中定义的 $\bar{\mathrm{s}}_{\boldsymbol{v}}$，其中 $\boldsymbol{\phi}$ 满足条件 *3* 和条件*4*。令 $\boldsymbol{v}\in U^1\times\ldots\times U^N$ 为一个可行收益，使得对于 $i\in[N]$, $v^i\geqslant u^i$。此外，假设对于任意 $i\in[N]$,

$$\beta^L \geqslant \frac{\bar{u}^i - v^i - \Delta_L}{\bar{u}^i - \underline{u}^i}\,. \tag{21}$$

那么，结论成立

1. *For any* $i\in[N]$, $\beta^L\left(1-\frac{p_L}{1-\beta^L}\right)v^i \leqslant U^i(\bar{\mathrm{s}}_{\boldsymbol{v}}) \leqslant v^i$,
2. $\bar{\mathrm{s}}_{\boldsymbol{v}}$ *is a* $\left(\beta^{-L}(1-\beta^{2L}+\Delta_L)+\frac{p_L}{1-\beta^L}\,,\, q_L\right)$*-HP-SPNE.*

定理*4*的证明。 我们先证明第 (2) 点。设 $t\geqslant 0$。下文中，我们记 $k_t=\lfloor t/L\rfloor$，因此 $k_tL\leqslant t<(k_t+1)L$。根据条件 3，存在 $\mathrm{H}_t\subset\mathcal{H}_t$，满足 $\mathbb{P}^{\bar{\mathrm{s}}_{\boldsymbol{v}}}(\mathrm{h}_t\in\mathrm{H}_t)\geqslant 1-q_L$，使得对任意 $\mathbb{P}^{\bar{\mathrm{s}}_{\boldsymbol{v}}}(\kappa_{\boldsymbol{\phi}}\geqslant k+1\,|\,\kappa_{\boldsymbol{\phi}}\geqslant k\,,\mathrm{h}_t)\geqslant 1-p_L$。在证明的其余部分，我们选取一个 $\mathrm{h}_t\in\mathrm{H}_t$。

首先，假设 $\mathrm{H}_t\subset\{\kappa_{\boldsymbol{\phi}}<k_t\}$，因此我们已经处于惩罚阶段。根据式 (3) 中 $\bar{\mathrm{s}}_{\boldsymbol{v}}$ 的定义，参与者从时间 t 起只进行 $\boldsymbol{b}=(b^1,\ldots,b^N)$。一方面，玩家 i 坚持 $\bar{s}^i_{\boldsymbol{v}}$ 会得到

$$U^i(\bar{\mathrm{s}}_{\boldsymbol{v}}\,;\mathrm{h}_t) = (1-\beta)\sum_{\ell=t}^{\infty}\beta^{\ell-t}u^i(b^i,b^{-i}) = u^i(b^i,b^{-i}) = \underline{u}^i\,.$$

另一方面，偏离到某个 $s\in\mathcal{S}$ 会产生收益

$$\begin{aligned}
U^i(s,\bar{s}^{-i}_{\boldsymbol{v}}\,;\mathrm{h}_t) &= (1-\beta)\,\mathbb{E}^{(s,\bar{s}^{-i}_{\boldsymbol{v}})}\left[\sum_{\ell=t}^{\infty}\beta^{\ell-t}u^i(w^i_\ell,b^{-i})\,|\,\mathrm{h}_t\right]\\
&\leqslant (1-\beta)\sum_{\ell=t}^{\infty}\beta^{\ell-t}u^i(b^i,b^{-i})\\
&= u^i(b^i,b^{-i})\\
&= \underline{u}^i\\
&= U^i(\bar{\mathrm{s}}_{\boldsymbol{v}}\,;\mathrm{h}_t)\,.
\end{aligned}$$

最后一个不等式源于 (b^i, b^{-i}) 是纳什均衡。因此，满足 HP-SPNE 条件。

现在，假设 $\mathrm{H}_t \subset \{k_t \leqslant \kappa_{\boldsymbol{\phi}}\}$，因此参与者尚未进入惩罚阶段。我们的证明分为三个步骤：(i) 我们将 $U^i(\bar{\mathbf{s}}_{\boldsymbol{v}}; \mathrm{h}_t)$ 的下界表示为一个依赖于 p_L 的项；(ii) 将 $U^i(s, \bar{s}^{-i}_{\boldsymbol{v}}; \mathrm{h}_t)$ 的上界表示为另一个依赖于 Δ_L 的项；(iii) 证明在定理的条件下，下界大于上界，允许存在一个近似项。

我们首先求 $U^i(\bar{\mathbf{s}}_{\boldsymbol{v}}; \mathrm{h}_t)$ 的下界。我们有

$$\begin{aligned} U^i(\bar{\mathbf{s}}_{\boldsymbol{v}}; \mathrm{h}_t) &= (1-\beta)\,\mathbb{E}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}\left[\mathbb{1}\{\kappa_{\boldsymbol{\phi}} = k_t\}\sum_{\ell=t}^{\infty}\beta^{\ell-t}u^i(w^i_\ell, w^{-i}_\ell) + \mathbb{1}\{\kappa_{\boldsymbol{\phi}} > k_t\}\sum_{\ell=t}^{\infty}\beta^{\ell-t}u^i(w^i_\ell, w^{-i}_\ell)\,|\,\mathrm{h}_t\right] \\ &\geqslant (1-\beta)\,\mathbb{E}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}\left[\mathbb{1}\{\kappa_{\boldsymbol{\phi}} > k_t\}\sum_{\ell=t}^{\infty}\beta^{\ell-t}u^i(w^i_\ell, w^{-i}_\ell)\,|\,\mathrm{h}_t\right]. \end{aligned}$$

由于 $t < (k_t+1)L$ 且 $\beta^{t-\ell} \geqslant \beta^{t-k_tL}$（因为 $t \geqslant k_tL$），并且收益为正，我们有

$$U^i(\bar{\mathbf{s}}_{\boldsymbol{v}}; \mathrm{h}_t) \geqslant (1-\beta)\mathbb{E}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}\left[\mathbb{1}\{\kappa_{\boldsymbol{\phi}} > k_t\}\sum_{\ell=(k_t+1)L}^{\kappa_{\boldsymbol{\phi}}L-1}\beta^{\ell-k_tL}u^i(w^i_\ell, w^{-i}_\ell)\,|\,\mathrm{h}_t\right]. \tag{48}$$

根据引理 2，我们可以扩充概率空间，并从一个由 U_k 组成的独立同分布序列 $(U_k)_{k\geqslant k_t}$ 构造 $\bar{\kappa}$，该序列独立于由 $\bar{\mathbf{s}}_{\boldsymbol{v}}$ 诱导的博弈过程。我们用 $\widetilde{\mathbb{E}}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}$ 表示在该乘积空间上的期望。根据引理 2，$\bar{\kappa} \leqslant \kappa_{\boldsymbol{\phi}}$ 几乎处处成立（在给定 h_t 的条件下），特别地，$\mathbb{1}\{\bar{\kappa} > k_t\} \leqslant \mathbb{1}\{\kappa_{\boldsymbol{\phi}} > k_t\}$。因此，我们有

$$\begin{aligned} U^i(\bar{\mathbf{s}}_{\boldsymbol{v}}; \mathrm{h}_t) &\geqslant (1-\beta)\,\widetilde{\mathbb{E}}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}\left[\mathbb{1}\{\bar{\kappa} > k_t\}\sum_{\ell=(k_t+1)L}^{L\bar{\kappa}-1}\beta^{\ell-k_tL}u^i(w^i_\ell, w^{-i}_\ell)\,|\,\mathrm{h}_t\right] \\ &= (1-\beta)\,\widetilde{\mathbb{E}}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}\left[\mathbb{1}\{\bar{\kappa} > k_t\}\sum_{k=k_t+1}^{\bar{\kappa}-1}\beta^{(k-k_t)L}\sum_{\ell=0}^{L-1}\beta^{\ell}u^i(w^i_{kL+\ell}, w^{-i}_{kL+\ell})\,|\,\mathrm{h}_t\right] \\ &= (1-\beta)\,\widetilde{\mathbb{E}}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}\left[\mathbb{1}\{\bar{\kappa} > k_t\}\sum_{k=1}^{\bar{\kappa}-k_t-1}\beta^{kL}\sum_{\ell=0}^{L-1}\beta^{\ell}v^i\right] \\ &= (1-\beta)\,\widetilde{\mathbb{E}}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}\left[\mathbb{1}\{\bar{\kappa} > k_t\}\sum_{k=1}^{\bar{\kappa}-k_t-1}\beta^{kL}\frac{1-\beta^L}{1-\beta}v^i\right] \\ &= \beta\,\mathbb{E}\left[\mathbb{1}\{\bar{\kappa} > k_t\}\,\beta^L\frac{1-\beta^{(\bar{\kappa}-k_t-1)L}}{1-\beta^L}(1-\beta^L)\,v^i\right] \\ &= \beta^L\,v^i\,\mathbb{E}\Big[\mathbb{1}\{\bar{\kappa} > k_t\}(1-\beta^{L(\bar{\kappa}-k_t-1)})\Big]. \end{aligned} \tag{49}$$

根据引理 2，我们知道 $\bar{\kappa} \sim \mathsf{Geometric}(p_L)$。因此，我们可以计算期望项

的具体表达式，并得到

$$\begin{aligned} U^i(\bar{\mathbf{s}}_{\boldsymbol{v}}\,;\mathrm{h}_t) &\geqslant \beta^L\, v^i \sum_{k\geqslant k_t+1} (1-p_L)^{k-k_t-1} p_L (1-\beta^{L(k-k_t-1)}) \\ &= \beta^L\, v^i \left(1 - p_L \sum_{k\geqslant 0} [(1-p_L)\beta^L]^k\right) \\ &= \beta^L\, v^i \left(1 - p_L (1-(1-p_L)\beta^L)^{-1}\right) \\ &\geqslant \beta^L \left(v^i - (1-\beta^L)^{-1} p_L\right), \end{aligned} \tag{50}$$

其中，最后一行用到了 $p_L \in (0,1)$、 $v^i \in [0,1]$ 和 $\beta^L \leqslant 1$。

现在我们来求 $U^i(s, \bar{s}_{\boldsymbol{v}}^{-i}; \mathrm{h}_t)$ 的一个上界，其中 $s \in \mathcal{S}^i$ 是玩家 i 的任意偏离。下文中，对于任意 $A^i \in \mathcal{A}^i$，我们记 $u^i(A^i, w^{-i}) = \sum_{a^{-i}} u^i(A^i, A^{-i}) w^{-i}[a^{-i}]$。我们有

$$U(s, \bar{s}_{\boldsymbol{v}}^{-i}\,;\mathrm{h}_t) = \mathbb{E}^{(s,\bar{s}_{\boldsymbol{v}}^{-i})}\left[\sum_{\ell=t}^{\infty} \beta^{\ell-t} u^i(A_\ell^i, w_\ell^{-i}) \,|\, \mathrm{h}_t\right]..$$

由于 $\ell \geqslant t \geqslant k_t L$，从 $k_t L$ 开始求和会得到一个上界。此外，鉴于 $t < (k_t+1)L$，对于任意 $\ell \geqslant t$，我们有 $\beta^{\ell-t} \leqslant \beta^{\ell-(k_t+1)L}$。因此

$$\begin{aligned} U(s, \bar{s}_{\boldsymbol{v}}^{-i};\mathrm{h}_t) &\leqslant (1-\beta)\mathbb{E}^{(s,\bar{s}_{\boldsymbol{v}}^{-i})}\left[\sum_{\ell=k_tL}^{\infty} \beta^{\ell-(k_t+1)L} u^i(A_\ell^i, w_\ell^{-i}) \,|\, \mathrm{h}_t\right] \\ &= (1-\beta)\beta^{-(k_t+1)L}\,\mathbb{E}^{(s,\bar{s}_{\boldsymbol{v}}^{-i})}\left[\mathbb{1}\{\kappa_{\boldsymbol{\phi}} > k_t\}\left(\sum_{k=k_t}^{\kappa_{\boldsymbol{\phi}}-1}\sum_{\ell=kL}^{(k+1)L-1} \beta^\ell u^i(A_\ell^i, w_{\boldsymbol{v}}^{-i})\right) \,|\, \mathrm{h}_t\right] \\ &\quad + (1-\beta)\,\beta^{-(k_t+1)L}\,\mathbb{E}^{(s,\bar{s}_{\boldsymbol{v}}^{-i})}\left[\sum_{\ell=\kappa_{\boldsymbol{\phi}}L}^{(\kappa_{\boldsymbol{\phi}}+1)L-1} \beta^\ell u^i(A_\ell^i, w_{\boldsymbol{v}}^{-i}) \,|\, \mathrm{h}_t\right] \\ &\quad + (1-\beta)\,\beta^{-(k_t+1)L}\,\mathbb{E}^{(s,\bar{s}_{\boldsymbol{v}}^{-i})}\left[\sum_{\ell\geqslant(\kappa_{\boldsymbol{\phi}}+1)L} \beta^\ell u^i(A_\ell^i, b^{-i}) \,|\, \mathrm{h}_t\right] \\ &\leqslant \underbrace{(1-\beta)\beta^{-(k_t+1)L}\,\mathbb{E}^{(s,\bar{s}_{\boldsymbol{v}}^{-i})}\left[\mathbb{1}\{\kappa_{\boldsymbol{\phi}} > k_t\}\sum_{k=k_t}^{\kappa_{\boldsymbol{\phi}}-1}\sum_{\ell=kL}^{(k+1)L-1} \beta^\ell u^i(A_\ell^i, w_{\boldsymbol{v}}^{-i}) \,|\, \mathrm{h}_t\right]}_{(\star\star)} \\ &\quad + (1-\beta^L)\,\mathbb{E}\Big[\beta^{(\kappa_{\boldsymbol{\phi}}-k_t-1)L} \,|\, \mathrm{h}_t\Big]\bar{u}^i + \mathbb{E}\Big[\beta^{(\kappa_{\boldsymbol{\phi}}-k_t)L} \,|\, \mathrm{h}_t\Big]\underline{u}^i \end{aligned} \tag{51}$$

现在我们对（$\star\star$）进行界定。由于$\{k < \kappa_{\boldsymbol{\phi}}\} \subset \{\phi_k^i = 0\}$，因此$\mathbb{1}\{k < \kappa_{\boldsymbol{\phi}}\} = \mathbb{1}\{k < \kappa_{\boldsymbol{\phi}}\}\mathbb{1}\{\phi_k^i = 0\}$。由此

可得

$$
\begin{aligned}
&\mathbb{E}^{(s,\bar{s}_{\boldsymbol{v}}^{-i})}\left[\mathbb{1}\{\kappa_{\boldsymbol{\phi}}>k_t\}\sum_{k=k_t}^{\kappa_{\boldsymbol{\phi}}-1}\sum_{\ell=kL}^{(k+1)L-1}\beta^{\ell}\,u^i(A_\ell^i,w_{\boldsymbol{v}}^{-i})\,|\,\mathrm{h}_t\right]\\
&\quad=\mathbb{E}^{(s,\bar{s}_{\boldsymbol{v}}^{-i})}\left[\mathbb{1}\{\kappa_{\boldsymbol{\phi}}>k_t\}\sum_{k=k_t}^{\infty}\mathbb{1}\{k<\kappa_{\boldsymbol{\phi}}\}\sum_{\ell=kL}^{(k+1)L-1}\beta^{\ell}u^i(A_\ell^i,w_{\boldsymbol{v}}^{-i})\,|\,\mathrm{h}_t\right]\\
&\quad\leqslant\mathbb{E}^{(s,\bar{s}_{\boldsymbol{v}}^{-i})}\left[\mathbb{1}\{\kappa_{\boldsymbol{\phi}}>k_t\}\sum_{k=k_t}^{\infty}\mathbb{1}\{k<\kappa_{\boldsymbol{\phi}}\}\mathbb{1}\{\phi_k^i=0\}\sum_{\ell=kL}^{(k+1)L-1}\beta^{\ell}u^i(A_\ell^i,w_{\boldsymbol{v}}^{-i})\,|\,\mathrm{h}_t\right]\\
&\quad\leqslant\mathbb{E}^{(s,\bar{s}_{\boldsymbol{v}}^{-i})}\left[\mathbb{1}\{\kappa_{\boldsymbol{\phi}}>k_t\}\sum_{k=k_t}^{\infty}\mathbb{1}\{k<\kappa_{\boldsymbol{\phi}}\}\sum_{\ell=kL}^{(k+1)L-1}\beta^{\ell}(v^i+\Delta_L)\,|\,\mathrm{h}_t\right]\qquad\text{by } \textit{Condition } 4\\
&\quad\leqslant\mathbb{E}^{(s,\bar{s}_{\boldsymbol{v}}^{-i})}\left[\mathbb{1}\{\kappa_{\phi}>k_t\}\sum_{k=k_t}^{\kappa_{\boldsymbol{\phi}}-1}\beta^{kL}\frac{1-\beta^L}{1-\beta}(v^i+\Delta_L)\,|\,\mathrm{h}_t\right]\\
&\leqslant\beta^{k_tL}\frac{1-\mathbb{E}^{(s,\bar{s}_{\boldsymbol{v}}^{-i})}[\beta^{(\kappa_{\boldsymbol{\phi}}-k_t-1)L}\,|\,\mathrm{h}_t]}{1-\beta^L}\frac{1-\beta^L}{1-\beta}(v^i+\Delta_L)\,,
\end{aligned}
$$

因此

$$
(\star\star)\leqslant\beta^{-L}(1-\mathbb{E}^{(s,\bar{s}_{\boldsymbol{v}}^{-i})}[\mathbb{1}\{\kappa_{\boldsymbol{\phi}}>k_t\}\beta^{(\kappa_{\boldsymbol{\phi}}-k_t-1)L}\,|\,\mathrm{h}_t])(v^i+\Delta^L)\,.
$$

因此，将其代回（51）可得

$$
\begin{aligned}
U(s,\bar{s}_{\boldsymbol{v}}^{-i};\mathrm{h}_t)&\leqslant\beta^{-L}\,(v^i+\Delta_L)+\mathbb{E}^{(s,\bar{s}_{\boldsymbol{v}}^{-i})}\Big[\beta^{(\kappa_{\boldsymbol{\phi}}-k_t-1)L}\,|\,\mathrm{h}_t\Big]\Big((1-\beta^L)\bar{u}^i+\beta^L\underline{u}^i-\beta^{-L}(v^i+\Delta_L)\Big)\\
&\leqslant\beta^{-L}\,(v^i+\Delta_L)+\mathbb{E}^{(s,\bar{s}_{\boldsymbol{v}}^{-i})}\Big[\beta^{(\kappa_{\boldsymbol{\phi}}-k_t-1)L}\,|\,\mathrm{h}_t\Big]\Big((1-\beta^L)\bar{u}^i+\beta^L\underline{u}^i-(v^i+\Delta_L)\Big)\,.
\end{aligned}\tag{52}
$$

由于我们假设 th 在 $\beta^L\geqslant(\bar{u}^i-v^i-\Delta_L)(\bar{u}^i-\underline{u}^i)^{-1}$ 处，第二项为负，因此剩下

$$
U(s,\bar{s}_{\boldsymbol{v}}^{-i};\mathrm{h}_t)\leqslant\beta^{-L}(v^i+\Delta_L)\,.\tag{53}
$$

最后，由（50）和（53）

$$
\begin{aligned}
U^i(s,\bar{s}_{\boldsymbol{v}}^{-i};\mathrm{h}_t)&\leqslant\beta^{-L}(v^i+\Delta_L)\\
&=\beta^Lv^i-\beta^L\frac{p_L}{1-\beta^L}+\beta^{-L}(1-\beta^{2L})v^i+\beta^{-L}\Delta_L+\beta^L\frac{p_L}{1-\beta^L}\\
&\leqslant U^i(\bar{\mathbf{s}}_{\boldsymbol{v}};\mathrm{h}_t)+\beta^{-L}(1-\beta^{2L})v^i+\beta^{-L}\Delta_L+\beta^L\frac{p_L}{1-\beta^L}\\
&\leqslant U^i(\bar{\mathbf{s}}_{\boldsymbol{v}};\mathrm{h}_t)+\beta^{-L}(1-\beta^{2L}+\Delta_L)+\frac{p_L}{1-\beta^L}
\end{aligned}
$$

由于 $\mathrm{h}_t\in\mathrm{H}_t$，且 $\mathbb{P}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}(\mathrm{h}_t\in\mathrm{H}_t)\geqslant1-q_L$，这证明 $^{-}\mathbf{s}_{\boldsymbol{v}}$ 是一个 $(\beta^{-L}(1-\beta^{2L}+\Delta_L)+p_L(1-\beta^L)^{-1},q_L)$-HP-SPNE。

接下来仍需证明第（1）点。按照从 $t=0$ 出发、与推导（50）时完全相同的计算步骤，我们得到 $U^i(^{-}\mathbf{s}_{\boldsymbol{v}};\mathrm{h}_t)\geqslant\beta^L(1-p_L(1-\beta^L)^{-1})v^i$。关于上界，我们

注意到根据假设， $v^i \geqslant u^i$，因此

$$\begin{aligned} U^i(\bar{\mathbf{s}}_{\boldsymbol{v}}) &= (1-\beta)\,\mathbb{E}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}\left[\sum_{\ell=0}^{(\kappa_{\boldsymbol{\phi}}+1)L-1}\beta^\ell v^i + \sum_{\ell=(\kappa_{\boldsymbol{\phi}}+1)L}^{\infty}\beta^\ell \underline{u}^i\right] \\ &\leqslant (1-\beta)\,\mathbb{E}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}\left[\sum_{\ell=0}^{(\kappa_{\boldsymbol{\phi}}+1)L-1}\beta^\ell v^i + \sum_{\ell=(\kappa_{\boldsymbol{\phi}}+1)L}^{\infty}\beta^\ell v^i\right] \\ &= v^i\,. \end{aligned}$$

证明至此结束。

□

在下列引理中，对于 $i \in [N]$，我们将 $U_k^i = u^i(a_k^i,\ w_{\boldsymbol{v}}^{-i}) = \sum_{a^{-i}} u^i(a_k^i,\ a^{-i}) w^{-i}[a^{-i}]$ 记为玩家 i 在对手采取 $w_{\boldsymbol{v}}^{-i}$ 时，与行动 $a_k^i \in \mathcal{A}^i$ 相关联的效用。为简化记号，我们假设 $U_1^i \geqslant \ldots \geqslant U_K^i$，并将行动 a_k^i 与其索引 k 对应。对于任意行动索引 $\mathbf{A} = (A_0,\ \ldots\ldots,\ A_{L-1}) \in [K]^L$，我们定义

$$V^i(\mathbf{A}) = \sum_{s=0}^{L-1}\beta^s\, U_{A_s}^i\,.$$

同理，f 对于任意 $\sigma \in \Sigma(\mathbf{A})$，其中 $\Sigma(\mathbf{A})$ 表示 $\mathbf{A}$ 上的排列集合，我们定义

$$V_L^i(\sigma, \mathbf{A}) = \sum_{s=0}^{L-1}\beta^s U_{\sigma(A_s)}^i\,.$$

引理 6. 考虑任意偶数 $L \in \mathbb{N}$ 和序列 $\mathbf{A} = (A_0^i,\ \ldots\ldots,\ A_{L-1}^i) \in [K]^L$。我们有

$$\max_{\sigma\in\Sigma(\mathbf{A})} V^i(\sigma, \mathbf{A}) \leqslant V^i(\mathbf{A}) + \frac{(1-\beta^L)^2}{1-\beta}\,.$$

引理 6 的证明。我们固定一个长度为 L 的序列 $\mathbf{A}$，其元素取自 $[K]$。为便于表述，在以下证明中我们省略对 i 的依赖。

首先，假设 L 是偶数。由于 $\beta \in (0,\ 1)$，$(\beta^t)_{t\in\{0,\ldots,L-1\}}$ 是递减的。因此，使效用最大化的排序是将 U_{A_t} 按非递增顺序排列，并定义 $\sigma^{\downarrow}$ 相应的置换。同理，使效用最小化的排序是将 U_{A_t} 按非递减顺序排列，并定义 $\sigma^{\uparrow}$ 相应的置换。因此

$$\begin{aligned} \max_{\sigma\in\Sigma(\mathbf{A})} V^i(\sigma, \mathbf{A}) - V^i(\mathbf{A}) &\leqslant \max_{\sigma,\sigma'\in\Sigma(\mathbf{A})} V^i(\sigma, \mathbf{A}) - V^i(\sigma', \mathbf{A}) \\ &= \sum_{t=0}^{L-1}\beta^t\, U_{\sigma^{\downarrow}(A_t)} - \sum_{t=0}^{L-1}\beta^t\, U_{\sigma^{\uparrow}(A_t)} \\ &= \sum_{t=0}^{L-1}\beta^t (U_{\sigma^{\downarrow}(A_t)} - U_{\sigma^{\uparrow}(A_t)})\,. \end{aligned}$$

由于 $\sigma^{\uparrow}(A_t)=\sigma^{\downarrow}(A_{L+1-t})$，它可以写成

$$\max_{\sigma,\sigma'\in\Sigma(\mathbf{A})}\left(V^i(\sigma,\mathbf{A})-V^i(\sigma',\mathbf{A})\right)=\sum_{t=0}^{L/2}(\beta^t-\beta^{L+1-t})\left(U_{\sigma^{\downarrow}(A_t)})-U_{\sigma^{\downarrow}(A_{L+1-t})}\right),$$

其中我们配对了对称索引。根据定义，对于任意 $t\in[L/2]$，都有 $U_{\sigma^{\downarrow}(A_t)}-U_{\sigma^{\downarrow}(A_{L+1-t})}\leqslant U_1-U_K\leqslant 1$，因此

$$\begin{aligned}\max_{\sigma,\sigma'\in\Sigma(\mathbf{A})}\left(V^i(\sigma,\mathbf{A})-V^i(\sigma',\mathbf{A})\right)&\leqslant\sum_{t=1}^{L/2}(\beta^t-\beta^{L+1-t})=(1-\beta^{L/2})\sum_{t=0}^{L/2}\beta^t\\&=(1-\beta^{L/2})(1-\beta^{L/2})/(1-\beta)\\&\leqslant\frac{(1-\beta^{L/2})^2}{1-\beta}\leqslant\frac{(1-\beta^L)^2}{1-\beta}.\end{aligned}$$

请注意，通过在求和中将与批次中段所执行的行动 $A_{(L-1)/2}$ 相关的项计算两次，证明可直接扩展到 L 为奇数的情形。 □

命题 3。令 $\beta,\delta\in(0,1)$。对于任意 $i\in[N]$，考虑在 (23) 中定义的检验以及在 (19)中定义的相应策略 $\bar{\mathbf{s}}_{\boldsymbol{v}}$。则，

1. 条件 *3* 成立，其中 $p_L=q_L=2KN\exp(-2L\delta^2/K^2)$，

2. 条件*4* 成立，其中 $\Delta_L=\delta+3(1-\beta^L)$。

证明。我们先证明第（1）点。令 $t\in\mathbb{N}$、$L>0$ 并记 $k_t=\lfloor t/L\rfloor$。我们要证明存在 $\mathbf{H}_t\subset\mathcal{H}_t$，使得

$$\mathbb{P}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}(\mathrm{h}_t\in\mathsf{H}_t)\geqslant 1-2KN\exp\left(-2\frac{L\delta^2}{K^2}\right),$$

并且对于任意 $\mathsf{h}_t\in\mathbf{H}_t$ 和 $k\geqslant k_t$，

$$\mathbb{P}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}(\kappa_{\boldsymbol{\phi}}\geqslant k+1\,|\,\kappa_{\boldsymbol{\phi}}\geqslant k\,,\mathrm{h}_t)\geqslant 1-2KN\exp\left(-2\frac{L\delta^2}{K^2}\right).$$

接下来，我们定义 $t_0=t-k_tL$。我们还定义向量 $w_{t_0}^{i\leftarrow}\in[0,1]^{t_0}$ 和 $w_{t_0}^{i\rightarrow}\in[0,1]^{L-t_0}$，使得对于任意 $\ell\in[K]$ 和 $i\in[N]$

$$w_{t_0,\ell}^{i\leftarrow}=t_0^{-1}\sum_{s=0}^{t_0-1}\mathbb{1}\{A_{k_tL+s}^i=a_\ell\}\quad\text{and}\quad w_{t_0,\ell}^{i\rightarrow}=(L-t_0)^{-1}\sum_{s=t_0}^{L-1}\mathbb{1}\{A_{k_tL+s}^i=a_\ell\},$$

它们分别表示玩家 i 在批次 k_t 开始至 t_0-1 期间，以及从 t_0 至批次 k_t 结束期间的行动 a_k 的经验频率。固定 $q_L\in(0,1)$，并将 $B\subset\mathcal{A}^{\mathbb{N}}$ 定义如下

$$B=B_0\times B_1\times\ldots\times B_{k_tL-1}\times\bar{B}\ldots,$$

其中，对于任意 $t\in\mathbb{N}\setminus\{k_tL,\ .....,\ k_tL+t_0-1\}$，$B_k=\mathcal{A}$，并且

$$\bar{B}=\left\{(A^i_{k_tL+s})_{\substack{1\leqslant i\leqslant N\\ 0\leqslant s\leqslant t_0-1}}\ :\ \text{for any } \ell\in[K] \text{ and } i\in[N],\ \left|w^{i\leftarrow}_{t_0,\ell}-w^i_{\boldsymbol{v}}[a^i_\ell]\right|\leqslant\sqrt{\frac{\ln(2KN/q_L)}{2t_0}}\right\},\qquad(54)$$

B 是从第 k_tL 轮到第 k_tL+t_0-1 轮的行动集合，其中每位玩家 $i\in[N]$ 的每种行动 $\ell\in[K]$ 在$\{k_tL,\ldots,k_tL+t_0\}$中的经验频率与$w^i_{\boldsymbol{v}}[a^i_\ell]$相差不大。注意，对每个 $(\ell,i)\in[K]\times[N]$，将Hoeffding不等式应用于独立且有界的变量（$\mathbb{1}\{A^i_{k_tL+s}=a_\ell\}$）$_{s\in\{0,\ldots,t_0-1\}}$，再对所有配对 (ℓ,i) 使用并集界，可得 $\mathbb{P}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}(\bar{B})\geqslant 1-q_L$；由从$\mathbf{w}_{\boldsymbol{v}}$进行的抽样相互独立可知

$$\mathbb{P}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}(B)=\mathbb{P}(B_0)\times\ldots\times P(B_{k_tL-1})\times\mathbb{P}(\bar{B})=1\,.\,\mathbb{P}(\bar{B})\geqslant 1-q_L\,.$$

We 证明以下历史集合满足我们所需的条件

$$\mathsf{H}_t=\{\kappa_{\boldsymbol{\phi}}<k_t\}\cup\{\,\{\kappa_{\boldsymbol{\phi}}\geqslant k_t\}\cap B\,\},\qquad(55)$$

其中回顾一下，$\kappa_{\boldsymbol{\phi}}=\min^{i\in[N]}\kappa_{\boldsymbol{\phi}^i}$。为理解为何 $\mathbb{P}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}(\mathsf{H}_t)\geqslant 1-q_L$，注意 $\mathsf{H}_t^c=\{\{\kappa_{\boldsymbol{\phi}}\geqslant k_t\}\cap\{\kappa_{\boldsymbol{\phi}}<k_t\}\}\cup\{\{\kappa_{\boldsymbol{\phi}}\geqslant k_t\}\cap B\}$，由此可得

$$\begin{aligned}\mathbb{P}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}(\mathsf{H}_t^c)&=\mathbb{P}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}(\{\{\kappa_{\boldsymbol{\phi}}<k_t\}\cap\{\kappa_{\boldsymbol{\phi}}\geqslant k_t\}\}\cup\{\{\kappa_{\boldsymbol{\phi}}\geqslant k_t\}\cap B^c\})\\&\leqslant 0+\mathbb{P}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}(\{\kappa_{\boldsymbol{\phi}}\geqslant k_t\}\cap B^c)\leqslant q_L\,,\end{aligned}$$

其中，最后一个不等式源于 $\mathbb{P}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}(B)\geqslant 1-q_L$。

令 $\mathrm{h}_t\in\mathsf{H}_t$。现在证明，当 $q_L=p_L=2KN\exp(-L\delta^2/K^2)$ 时，

$$\mathbb{P}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}(\kappa_{\boldsymbol{\phi}}\geqslant k+1\,|\,\kappa_{\boldsymbol{\phi}}\geqslant k\,,\mathrm{h}_t)\geqslant 1-p_L\,,$$

对任意 $k\geqslant k_t$。若 $\mathrm{h}_t\subset\{\kappa_{\boldsymbol{\phi}}<k_t\}$，则该命题显然成立，因为 $\{\kappa_{\boldsymbol{\phi}}\geqslant k\}\cap\mathrm{h}_t=\emptyset$。

因此，我们重点考虑 $\mathrm{h}_t\subset\{k_t\leqslant\kappa_{\boldsymbol{\phi}}\}\cap B$ 的情形。令 $k\geqslant 0$，使得 $k_t\leqslant k\leqslant\kappa_{\boldsymbol{\phi}}$。首先，假设 $k=k_t$。对于任意 $i\in[N]$，我们有

$$\begin{aligned}\|\hat{w}^i_{k_t}-w^i_{\boldsymbol{v}}\|_1&=\sum_{\ell=1}^K\left|\frac{1}{L}\sum_{s=0}^{L-1}\mathbb{1}\{A^i_s=a_\ell\}-w^i_{\boldsymbol{v}}[a_\ell]\right|\\&=\sum_{\ell=1}^K\left|\frac{t_0}{L}w^{i\leftarrow}_{t_0,\ell}+\frac{L-t_0}{L}w^{i\rightarrow}_{t_0,\ell}-\left(\frac{t_0}{L}w^i_{\boldsymbol{v}}[a_\ell]+\frac{L-t_0}{L}w^i_{\boldsymbol{v}}[a_\ell]\right)\right|\\&\leqslant\sum_{\ell=1}^K\left(\frac{t_0}{L}\left|w^{i\leftarrow}_{t_0,\ell}-w^i_{\boldsymbol{v}}[a_\ell]\right|+\frac{L-t_0}{L}\left|w^{i\rightarrow}_{t_0,\ell}-w^i_{\boldsymbol{v}}[a_\ell]\right|\right).\end{aligned}$$

由于 $\mathrm{h}_t\in B$

$$\|\hat{w}^i_{k_t}-w^i_{\boldsymbol{v}}\|_1\leqslant\sum_{\ell=1}^K\left(\frac{t_0}{L}\sqrt{\frac{\ln(2KN/q)}{2t_0}}+\frac{L-t_0}{L}\left|w^{i\rightarrow}_{t_0,\ell}-w^i_{\boldsymbol{v}}[a_\ell]\right|\right).$$

因此，对 $|w^{i\to}_{t_0,\ell} - w^i_{\boldsymbol{v}}[a_\ell]|$ 应用 Hoeffding 不等式，对于任意 $\ell \in [K]$ 和 $i \in [K]$，并结合所有配对 (ℓ, i) 上的并集界，可得

$$\mathbb{P}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}\left(\bigcap_{i\in[N]}\left\{\|\hat{w}^i_{k_t} - w^i_{\boldsymbol{v}}\|_1 \leqslant \sum_{k=1}^{K}\left(\frac{t_0}{L}\sqrt{\frac{\ln(2KN/q)}{2t_0}} + \frac{L-t_0}{L}\sqrt{\frac{\ln(2KN/q)}{2(L-t_0)}}\right)\right\} \,|\, \kappa_{\boldsymbol{\phi}} \geqslant k_t\,,\, \mathrm{h}_t\right) \geqslant 1 - q_L\,,$$

上述求和中的每一项都在 $t_0 = L/2$ 时取最大值，因此我们也有

$$\mathbb{P}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}\left(\bigcap_{i\in[N]}\left\{\|\hat{w}^i_{k_t} - w^i_{\boldsymbol{v}}\| \leqslant K\sqrt{\frac{\ln(2KN/q)}{L}}\right\} \,|\, \kappa_{\boldsymbol{\phi}} \geqslant k_t\,,\, \mathrm{h}_t\right) \geqslant 1 - q_L\,,$$

然后，取 $q_L = 2KN\exp\left(-2\frac{L\delta^2}{K^2}\right)$，得到

$$\mathbb{P}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}\left(\bigcap_{i\in[N]}\left\{\|\hat{w}^i_{k_t} - w^i_{\boldsymbol{v}}\|_1 \leqslant \delta\right\} \,|\, \kappa_{\boldsymbol{\phi}} \geqslant k_t\,,\, \mathrm{h}_t\right) \geqslant 1 - 2KN\exp\left(-\frac{L\delta^2}{K^2}\right). \tag{56}$$

由于在给定$\{\kappa_{\boldsymbol{\phi}} \geqslant k_t\}$的条件下，$\{\kappa_{\boldsymbol{\phi}} \geqslant k_t + 1\} = \cap_{i\in[N]}\{\phi^i_{k_t} = 0\} = \cap_{i\in[N]}\left\{\|\hat{} w^i_{k_t} - w^i_{\boldsymbol{v}}\|_1 \leqslant \delta\right\}$，由 (56) 可得

$$\begin{aligned}\mathbb{P}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}(\kappa_{\boldsymbol{\phi}} \geqslant k_t + 1 \,|\, \kappa_{\boldsymbol{\phi}} \geqslant k_t\,,\, \mathrm{h}_t) &= \mathbb{P}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}\left(\bigcap_{i\in[N]}\{\phi^i_{k_t} = 0\} \,|\, \kappa_{\boldsymbol{\phi}} \geqslant k\,,\, \mathrm{h}_t\right)\\ &= \mathbb{P}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}\left(\bigcap_{i\in[N]}\left\{\|\hat{w}^i_{k_t} - w^i_{\boldsymbol{v}}\|_1 \leqslant \delta\right\} \,|\, \kappa_{\boldsymbol{\phi}} \geqslant k_t\,,\, \mathrm{h}_t\right) \geqslant 1 - q_L\,,\end{aligned}$$

由此即得所需结果。

现在，假设 $k > k_t$。对有界独立随机变量$\{\mathbb{1}\{A^i_s = a_\ell\}\}^{s\in\mathcal{B}_k}$应用 Hoeffding不等式，对任意 $\ell \in [K]$,

$$\mathbb{P}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}\left(\|\hat{w}^i_{k,\ell} - w^i_{\boldsymbol{v}}[a^i_\ell]\|_1 \geqslant \sqrt{\frac{\ln(2KN/q_L)}{2t_0}} \,|\, \mathrm{h}_t\,,\, \kappa_{\boldsymbol{\phi}} \geqslant k\right) \leqslant \frac{q_L}{KN}\,,$$

这里我们去掉了对 h_t 的条件化，因为 $(A^i_s)_{s\in\mathcal{B}_k}$ 与 h_t 独立（因为 $k > k_t$）。接下来，对所有成对组合 $(i,\ \ell)$ 使用并集界，并按照前面相同的推理，即可得到所需结果。

我们现在证明第 2 点。为方便记号，我们省略 i 索引。接下来，我们定义

$$\xi^i = \left(\sum_{a^{-i}} u^i(a_1, a^{-i})w^{-i}_{\boldsymbol{v}}[a^{-i}], \ldots, \sum_{a^{-i}} u^i(a_K, a^{-i})w^{-i}_{\boldsymbol{v}}[a^{-i}]\right),$$

因此，对于任意 $w \in \Delta(\mathcal{A}^i)$，$u^i(w,\ w^{-i}_{\boldsymbol{v}}) = \langle w,\ \xi^i\rangle$。我们还引入以下记号

$$Z = \sum_{t=0}^{L-1}\beta^t \quad \text{and} \quad \hat{w}^{(\beta)}_k = Z^{-1}\left(\sum_{t=0}^{L-1}\beta^t\mathbb{1}\{A_{kL+t} = a_1\}, \ldots, \sum_{t=0}^{L-1}\beta^t\mathbb{1}\{A_{kL+t} = a_K\}\right).$$

注意，$w_k^{(\beta)}$是定义在$\{a_1,\ \ldots,\ a_K\}$上的概率测度。最后，对于在批次 k中采取的行动的任意置换$\sigma\colon\{1,\ \ldots,\ L\}\to\{1,\ \ldots,\ L\}$，我们记为

$$\hat{w}^i_{k,\sigma}=L^{-1}\left(\sum_{s=0}^{L-1}\mathbb{1}\{A^i_{kL+\sigma(1)}=a_1\},\ldots,\sum_{s=0}^{L-1}\mathbb{1}\{A^i_{kL+\sigma(L)}=a_K\}\right).$$

注意到 $\hat{}w^i_k=\hat{}w^i_{k,\ \sigma}$ 因此 $\mathbb{1}\{\|\hat{}w^i_{k,\ \sigma}-\tilde{w}^i\|_1\leqslant\delta\}\ =\mathbb{1}\{\|\hat{}w^i_k-\tilde{w}^i\|_1\leqslant\delta\}$。换句话说，玩家 i 可以在任意批次内调换行动，而不被发现，从而提高自己的收益。注意到

$$\begin{aligned}
&\mathbb{1}\{\|\hat{w}^i_k-w^i_{\boldsymbol{v}}\|_1\leqslant\delta\}\left(\sum_{t=kL}^{(k+1)L-1}\beta^t u^i(A^i_t,w^{-i}_{\boldsymbol{v}})-\sum_{t=kL}^{(k+1)L-1}\beta^t u^i(w^i_{\boldsymbol{v}},w^{-i}_{\boldsymbol{v}})\right)\\
&\quad=\mathbb{1}\{\|\hat{w}^i_k-w^i_{\boldsymbol{v}}\|_1\leqslant\delta\}\,\beta^{kL}\left(\sum_{t=0}^{L-1}\beta^t u^i(A^i_{kL+t},w^{-i}_{\boldsymbol{v}})-\sum_{t=0}^{L-1}\beta^t u^i(w^i_{\boldsymbol{v}},w^{-i}_{\boldsymbol{v}})\right)\\
&\quad\leqslant\max_\sigma\left\{\mathbb{1}\{\|\hat{w}^i_{\sigma,k}-w^i_{\boldsymbol{v}}\|_1\leqslant\delta\}\,\beta^{kL}\left(\sum_{t=0}^{L-1}\beta^t u^i(A^i_{kL+\sigma(t)},w^{-i}_{\boldsymbol{v}})-\sum_{t=0}^{L-1}\beta^t u^i(w^i_{\boldsymbol{v}},w^{-i}_{\boldsymbol{v}})\right)\right\}\\
&\quad=\mathbb{1}\{\|\hat{w}^i_k-w^i_{\boldsymbol{v}}\|_1\leqslant\delta\}\,\beta^{kL}\left(\max_\sigma\left\{\sum_{t=0}^{L-1}\beta^t u^i(A^i_{kL+\sigma(t)},w^{-i}_{\boldsymbol{v}})\right\}-\sum_{t=0}^{L-1}\beta^t u^i(w^i_{\boldsymbol{v}},w^{-i}_{\boldsymbol{v}})\right)\\
&\quad\leqslant\mathbb{1}\{\|\hat{w}^i_k-w^i_{\boldsymbol{v}}\|_1\leqslant\delta\}\,\beta^{kL}\left(\sum_{t=0}^{L-1}\beta^t u^i(A^i_{kL+t},w^{-i}_{\boldsymbol{v}})+\frac{(1-\beta^L)^2}{1-\beta}-\sum_{t=0}^{L-1}\beta^t u^i(w^i_{\boldsymbol{v}},w^{-i}_{\boldsymbol{v}})\right)\quad\text{by Lemma 6}\\
&\quad=\mathbb{1}\{\|\hat{w}^i_k-w^i_{\boldsymbol{v}}\|_1\leqslant\delta\}\beta^{kL}\,Z\left(\langle\hat{w}^{(\beta)}_k-w^i_{\boldsymbol{v}}\,,\xi^i\rangle+Z^{-1}\frac{(1-\beta^L)^2}{1-\beta}\right)\\
&\quad\leqslant\mathbb{1}\{\|\hat{w}^i_k-w^i_{\boldsymbol{v}}\|_1\leqslant\delta\}\beta^{kL}\,Z\Big(\|\hat{w}^{(\beta)}_k-w^i_{\boldsymbol{v}}\|_1\|\xi^i\|_\infty+(1-\beta^L)\Big)\\
&\quad\leqslant\mathbb{1}\{\|\hat{w}^i_k-w^i_{\boldsymbol{v}}\|_1\leqslant\delta\}\beta^{kL}\,Z\Big(\|\hat{w}^{(\beta)}_k-w^i_{\boldsymbol{v}}\|_1+(1-\beta^L)\Big)\qquad(\text{because } u^i(\mathbf{a})\leqslant 1\text{ for any }\mathbf{a}\in\mathcal{A})\\
&\quad\leqslant\mathbb{1}\{\|\hat{w}^i_k-w^i_{\boldsymbol{v}}\|_1\leqslant\delta\}Z\,\beta^{kL}\left(\|\hat{w}^{(\beta)}_k-\hat{w}\|_1+\|\hat{w}-w^i_{\boldsymbol{v}}\|_1+(1-\beta^L)\right)\\
&\quad\leqslant Z\,\beta^{kL}\left(\sum_{t=0}^{L-1}\left|\frac{\beta^t}{Z}-\frac{1}{L}\right|+\delta+(1-\beta^L)\right).
\end{aligned}\tag{57}$$

现在我们对 $\sum_{t=0}^{L-1}\left|\frac{\beta^t}{Z}=\frac{1}{L}\right|$进行界定。对于每个 $t\in\{0,\ \ldots,\ L-1\}$，应用三角不等式可得

$$\left|\frac{\beta^t}{Z}-\frac{1}{L}\right|\leqslant\left|\frac{\beta^t}{Z}-\frac{\beta^t}{L}\right|+\left|\frac{\beta^t}{L}-\frac{1}{L}\right|=\beta^t\left|\frac{1}{Z}-\frac{1}{L}\right|+\frac{1-\beta^t}{L}.$$

然后，对 $t=0,\ \ldots,\ L-1$ 求和可得

$$\sum_{t=0}^{L-1}\left|\frac{\beta^t}{Z}-\frac{1}{L}\right|\leqslant\left|\frac{1}{Z}-\frac{1}{L}\right|\sum_{t=0}^{L-1}\beta^t+\frac{1}{L}\sum_{t=0}^{L-1}(1-\beta^t)\leqslant\left|1-\frac{Z}{L}\right|+\frac{L-Z}{L},$$

因为 $\beta^t\leqslant 1$ 对所有 t，我们有 $Z\leqslant L$，因此 $\left|1-\frac{Z}{L}\right|=1-\frac{Z}{L}=\frac{L-Z}{L}$.故 ——

$$\sum_{t=0}^{L-1}\left|\frac{\beta^t}{Z}-\frac{1}{L}\right|\leqslant 2\,\frac{L-Z}{L}.$$

最后，注意到

$$L - Z = \sum_{t=0}^{L-1}(1-\beta^t) \leqslant \sum_{t=0}^{L-1}(1-\beta^L) = L(1-\beta^L)\,,$$

其中我们对 $0 \leqslant t \leqslant L-1$ 使用了 $\beta^t \geqslant \beta^L$。因此

$$\sum_{t=0}^{L-1}\left|\frac{\beta^t}{Z} - \frac{1}{L}\right| \leqslant 2(1-\beta^L)\,.$$

将其代入式 (57) 得到

$$\begin{aligned}
&\mathbb{1}\{\|\hat{w}_k^i - w_{\boldsymbol{v}}^i\|_1 \leqslant \delta\}\left(\sum_{t=kL}^{(k+1)L-1}\beta^t u^i(A_t^i, w_{\boldsymbol{v}}^{-i}) - \sum_{t=kL}^{(k+1)L-1}\beta^t u^i(w_{\boldsymbol{v}}^i, w_{\boldsymbol{v}}^{-i})\right)\\
&\qquad \leqslant Z\,\beta^{kL}\left(2(1-\beta^L) + \delta + \frac{(1-\beta^L)^2}{1-\beta}\right)\\
&\qquad = \sum_{t=kL}^{(k+1)L-1}\beta^t(2(1-\beta^L) + \delta + (1-\beta^L))\,, \qquad (58)
\end{aligned}$$

由此得证。 □

推论3. 设 $\boldsymbol{v} = (v^1, \ldots, v^N) \in U^1 \times \ldots \times U^N$，使得对于任意 $i \in [N]$、$\varepsilon > 0$ 和 $\beta \in (0,1)$，均有 $v^i \geqslant u^i$。对于任意 $i \in [N]$，考虑由式 (23) 定义的检验，其水平为 $\delta = \varepsilon/16$，批次大小为

$$L = \left\lceil \frac{258K^2}{\varepsilon^2}\ln\left(\frac{126KN}{\varepsilon^2}\right)\right\rceil,$$

以及由式 (19) 定义的相应策略 $\bar{\mathbf{s}}_{\boldsymbol{v}}$。若 $1-\varepsilon/16 \leqslant \beta^L \leqslant 1-\varepsilon/32$，则 n

(i) $(1-\varepsilon)v^i \leqslant U^i(\bar{\mathbf{s}}_{\boldsymbol{v}}) \leqslant v^i$,

(ii) $\bar{\mathbf{s}}_{\boldsymbol{v}}$ 是一个 $(\varepsilon, \varepsilon)$-*HP-SPNE*。

证明 推论*3*。固定 $\varepsilon \in (0,1]$。根据假设并结合我们的参数，我们有

$$1 - \frac{\varepsilon}{16} \leqslant \beta^L \leqslant 1 - \frac{\varepsilon}{32}\,, \qquad \delta = \frac{\varepsilon}{16}\,, \qquad L = \frac{256K^2}{\varepsilon^2}\ln\left(\frac{128K}{\varepsilon^2}\right). \qquad (59)$$

回顾可知

$$\Delta_L = \delta + 3(1-\beta^L)\,, \qquad p_L = 2KN\exp\left(-\frac{L\delta^2}{K^2}\right).$$

We 首先界定 $1-\beta^{2L}+\Delta_L$。利用 $1-\beta^{2L} = (1-\beta^L)(1+\beta^L) \leqslant 2(1-\beta^L)$，我们得到

$$1 - \beta^{2L} + \Delta_L \leqslant 2(1-\beta^L) + \delta + 3(1-\beta^L) = \delta + 5(1-\beta^L)\,. \qquad (60)$$

此外，由（59）可得 $1-\beta^L \leqslant \varepsilon/16$ 和 $\delta = \varepsilon/16$，因此

$$1 - \beta^{2L} + \Delta_L \leqslant \frac{\varepsilon}{16} + \frac{5\varepsilon}{16} = \frac{3\varepsilon}{8}\,. \qquad (61)$$

现在我们对前置因子 β^{-L} 进行界定。由于 $\beta^L \geqslant 1-\varepsilon/16$，我们立即得到

$$\beta^{-L} \leqslant \frac{1}{1-\varepsilon/16} \leqslant \frac{16}{15}, \tag{62}$$

其中，最后一个不等式用到了 $\varepsilon \leqslant 1$。结合 (61) 和 (62) 得到

$$\beta^{-L}(1-\beta^{2L}+\Delta_L) \leqslant \frac{16}{15}\cdot\frac{3\varepsilon}{8} = \frac{2\varepsilon}{5}. \tag{63}$$

现在我们对 $p_L/(1-\beta^L)$ 进行界定：令 $s = 2KN\exp(-L\delta^2/K^2)$，并利用 $1-\beta^L \geqslant \varepsilon/32$ 这一事实，得到

$$\frac{p_L}{1-\beta^L} = \frac{s}{1-\beta^L} \leqslant \frac{32s}{\varepsilon}. \tag{64}$$

现在利用我们对 L 和 δ 的选择，计算 s

$$s = 2KN\exp\left(-\frac{L\delta^2}{K^2}\right) = 2KN\exp\left(-\frac{L}{K^2}\cdot\frac{\varepsilon^2}{256}\right) = 2KN\exp\left(-\ln\left(\frac{128KN}{\varepsilon^2}\right)\right) = \frac{\varepsilon^2}{64}.$$

将其代入 (64) 得到

$$\frac{p_L}{1-\beta^L} \leqslant \frac{32}{\varepsilon}\cdot\frac{\varepsilon^2}{64} = \frac{\varepsilon}{2}. \tag{65}$$

因此，由 (63) 和 (65) 可得

$$\beta^{-L}(1-\beta^{2L}+\Delta_L) + \frac{p_L}{1-\beta^L} \leqslant \frac{2\varepsilon}{5}+\frac{\varepsilon}{2} = \frac{9\varepsilon}{10} \leqslant \varepsilon.$$

最后，注意到 $p_L = \varepsilon^2/64 \leqslant \varepsilon$。综上所述，我们得到 $\bar{\mathbf{s}}_{\boldsymbol{v}}$ 是一个$(\varepsilon,\varepsilon)$–子博弈完美纳什均衡。

对于第(ii)点，我们寻求一个下界，针对

$$\beta^L\left(1-\frac{p_L}{1-\beta^L}\right)v^i.$$

根据我们先前的计算，$p_L/(1-\beta^L) \leqslant \varepsilon/2$，因此我们有

$$\begin{aligned}
\beta^L\left(1-\frac{p_L}{1-\beta^L}\right)v^i &\geqslant \left(1-\frac{\varepsilon}{16}\right)\left(1-\frac{\varepsilon}{2}\right)v^i \\
&= \left(1-\frac{\varepsilon}{2}-\frac{\varepsilon}{16}+\frac{\varepsilon^2}{32}\right)v^i \\
&\geqslant \left(1-\frac{\varepsilon}{2}-\frac{\varepsilon}{16}\right)v^i \\
&= \left(1-\frac{9\varepsilon}{16}\right)v^i \\
&\geqslant (1-\varepsilon)v^i.
\end{aligned}$$

□

命题 4. 设 β、$\delta \in (0,\ 1)$。对任意 $i \in [N]$，考虑在(23) 中定义的检验，以及由此得到的策略 $\bar{\mathbf{s}}_{\boldsymbol{v}}$，该策略在(19) 中针对某个可行收益 $\boldsymbol{v}$ 定义，使得对 $i \in [N]$有 $v^i \geqslant \underline{u}^i$。假设

$$\min_{j\in[N]}\left\{\|w^j_{\boldsymbol{v}}\|_0 = \sum_{\ell=1}^{K}\mathbb{1}\{w^j_{\boldsymbol{v},\ell}\neq 0\}\right\} > 1 \ \textit{and}\ \delta < 1-1/\min_{j\in[N]}\|w^j_{\boldsymbol{v}}\|_0 \ . \tag{24}$$

于是 $\mathbb{P}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}(\kappa_{\boldsymbol{\phi}} < \infty) = 1$。

证明。 接下来，我们令

$$(j,r)\in \operatorname{argmin}_{i\in[N],\ell\in[K]}\left\{w^i_{\boldsymbol{v},\ell},\ w^i_{\boldsymbol{v},\ell}>0\right\}.$$

注意，记作 $\|w^j_{\boldsymbol{v}}\|_0=\sum_{\ell=1}^K \mathbb{1}\{w^j_{\boldsymbol{v},\ \ell}\neq 0\}$，

$$w^j_{\boldsymbol{v},r}\leqslant 1/\|w^j_{\boldsymbol{v}}\|_0\,, \tag{66}$$

因为否则 $\sum_{\ell\in[K]} w^j_{\boldsymbol{v},\ \ell}\geqslant \|w^j_{\boldsymbol{v}}\|_0 \underbrace{\frac{1}{\|w^j_{\boldsymbol{v}}\|_0}>1}_{\text{。}}$ 此外，

$$\mathbb{P}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}(\kappa_{\boldsymbol{\phi}}<\infty)=\mathbb{P}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}\left(\bigcup_{i\in[N]}\bigcup_{k\geqslant 0}\left\{\|\hat{w}^i_k-w^i_{\boldsymbol{v}}\|_1>\delta\right\}\right)\geqslant \mathbb{P}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}\left(\bigcup_{k\geqslant 0}\left\{\|\hat{w}^j_k-w^j_{\boldsymbol{v}}\|_1>\delta\right\}\right). \tag{67}$$

最后，注意对于任意 $k\geqslant 0$, $\cap_{t\in\mathcal{B}_k}\{A^j_t=a_r\}\subset\{\|\hat{}w^j_k-w^j_{\boldsymbol{v}}\|_1>\delta\}$，在此事件下

$$\|\hat{w}^j_k-w^j_{\boldsymbol{v}}\|_1=\sum_{\ell=1}^K\left|\hat{w}^j_{k,\ell}-w^j_{\boldsymbol{v},\ell}\right|=1-w^j_{\boldsymbol{v},r}\geqslant 1-1/\|w^j_{\boldsymbol{v}}\|_0>\delta\,,$$

其中，最后两个不等式分别由 (66) 和我们的假设得出。将这一观察与 (67) 结合，便得到

$$\begin{aligned}\mathbb{P}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}(\kappa<\infty)\geqslant \mathbb{P}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}\left(\bigcup_{k\geqslant 0}\left\{\|\hat{w}^j_k-w^j_{\boldsymbol{v}}\|_1>\delta\right\}\right)&\geqslant \mathbb{P}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}\left(\bigcup_{k\geqslant 0}\bigcap_{t\in\mathcal{B}_k}\left\{A^j_t=a_r\right\}\right)\\&=1-\mathbb{P}^{\bar{\mathbf{s}}_{\boldsymbol{v}}}\left(\bigcap_{k\geqslant 0}\bigcup_{t\in\mathcal{B}_k}\{A^j_t\neq a_r\}\right)\\&=1-\prod_{k=0}^{\infty}(1-(w^j_{\boldsymbol{v},r})^L)=1\,,\end{aligned}$$

最后一步中，我们利用了在 $\bar{}_{\mathbf{s}\boldsymbol{v}}$ 下的抽样独立性。 □