论文研读：对立偏好下的信息聚合

2024-2025 学年春夏学期计算经济学讨论班

郑涵文

zhwen@zju.edu.cn

2025 年 3 月 14 日

Introduction

考虑一个由 n 位居民组成的虚拟社区，需要决定是否征收住房税．社区居民分为两个利益对立的群体：房东和租户．房东倾向提高租金，租户则希望降低租金．社区通过居民投票决定政策，理想决策是当税收对租金的影响显现后，获得过半数居民支持的方案．若税收效果事前已知（例如确定会降低租金），决策将十分简单：若房东占多数则拒绝税收政策，若租户占多数则接受．

多数投票机制为每位居民提供"接受"或"拒绝"两个选项，以得票多者作为获胜方案，能有效反映多数人偏好．当税收效果已知时，此机制具有防策略性．然而政策效果在实施前具有不确定性，居民只能基于公共信息和个体经验形成对政策效果的直觉判断．若居民如实汇报判断，根据**孔多塞陪审团定理**，当多数人持有正确判断时，多数投票能以高概率正确聚合政策效果．

现实情况更为复杂：首先，当多数人持有错误判断时（如伪科学），多数投票机制将失效；其次，投票者通常具有异质性偏好．在住房税案例中，多数群体试图通过诚实投票选择最优方案，而持相反偏好的少数群体可能通过策略行为（非诚实）阻挠．

此类偏好对立现象广泛存在于薪资调整、总统选举、自由与安全的权衡等场景．政策影响(如英国脱欧后果）往往在实施前难以预判．核心问题是：在这些复杂场景下，是否仍能形成良好的集体决策？

Introduction

具体而言，论文研究了在以下情形下的投票博弈：

- **部分知情投票者**：投票者不知道世界的真实状态，只接收与世界状态相关的信号.
- **对立偏好**：每个投票者都有自己的类型，不同类型的偏好不同．在本论文中研究的是二元对立偏好，即投票者只有两种类型.

研究目标是揭示" 知情多数决策"——即当世界状态显现时，被多数类型投票者支持的方案. 怎样能够在不知道世界状态时，人们仍然能够做出 "知情多数决策"？换句话说，研究目标就是在世界状态未知的情况下，设计能够揭露知情多数决策的机制.

问题模型

- 世界状态：$\Omega = \{L, H\}$, 在这里 L 表明征收住房税使得租金降低，而 H 表明租金提高．
- 居民们需要投票表决这个税收提议，要么选择接受 ($\mathbf{A}$)，要么选择拒绝 ($\mathbf{R}$)．
- 居民并不知道真实世界的状态，而是自己有一个对世界状态的估计，被称为私人信号．信号用 S_i 表示，它属于 $\mathcal{S} = \{l, h\}$．如果 $S_i = l$，那么表明居民 i 相信这个世界状态是 L, 反之亦然．我们假设居民接收到的信号是独立同分布的，遵循某个依赖于真实世界状态 $\omega \in \Omega$ 的分布．所有的居民都知道世界状态和信号之间的共同分布．
- 我们使用 P_l^L 代表条件概率 $\Pr[S_i = l \mid \omega = L]$, 这表示了当现实世界状态为 L 时，居民接收到信号为 l 的概率．类似地定义 P_l^H，P_h^L，P_h^H. 显然有 $P_l^H + P_h^H = 1$ 和 $P_l^L + P_h^L = 1$.
- 最后我们使用 μ 表示真实世界状态为 H 的概率，相应地 $1-\mu$ 是 L 的概率．
- 论文研究的是信号与世界状态呈正相关的情况，即：

$$P_l^L > P_l^H \text{ and } P_h^H > P_h^L.$$

记 $\Delta = P_l^L - P_l^H = P_h^H - P_h^L$ 表示在世界状态下信号频率（概率）的差值．我们上面的假设等价于 $\Delta > 0$.

问题模型

- 每个居民 i 都有一个估值函数 $v_i : \Omega \times \{\mathbf{A}, \mathbf{R}\} \to \{0, 1, \ldots, B\}$
- 居民有两种类型：$\mathcal{T} = \{\mathcal{L}, \mathcal{H}\}$
  - An type-$\mathcal{L}$ agent (analogous to a tenant in the example) prefers alternative $\mathbf{A}$ in state L and prefers alternative $\mathbf{R}$ in state H. Formally,

$$v_i(L, \mathbf{A}) > v_i(L, \mathbf{R}) \text{ and } v_i(H, \mathbf{R}) > v_i(H, \mathbf{A}) \text{ for } i \text{ such that } t_i = \mathcal{L}.$$

  - An type-$\mathcal{H}$ agent (analogous to a landlord) prefers alternative $\mathbf{R}$ in state L and prefers $\mathbf{A}$ in state H. Formally,

$$v_i(L, \mathbf{R}) > v_i(L, \mathbf{A}) \text{ and } v_i(H, \mathbf{A}) > v_i(H, \mathbf{R}) \text{ for } i \text{ such that } t_i = \mathcal{H}.$$

- $\alpha_{\mathcal{L}}$ 表示 type-$\mathcal{L}$ 居民占总数的百分比, $\alpha_{\mathcal{H}}$ 表示 type-$\mathcal{H}$ 的. 显然 $\alpha_{\mathcal{L}} + \alpha_{\mathcal{H}} = 1$. $\alpha = \max\{\alpha_{\mathcal{L}}, \alpha_{\mathcal{H}}\}$ 表示多数群体所占比例. 当 $\alpha > \frac{1}{2}$ 时有确定的多数群体.

问题模型

An *instance* of our model includes:

- the number of agents, n;
- the common prior of the world state, $(\mu, 1-\mu)$, where μ is the probability of state H;
- the signal distributions, (P_h^H, P_l^H) and (P_h^L, P_l^L), where P_s^H and P_s^L denote the conditional probabilities of receiving signal $s \in \mathcal{S}$ in two possible states;
- the utility functions of agents, $\{v_i\}_{i=1}^n$ (assumed to be private);
- and the fraction of each type, $(\alpha_\mathcal{L}, \alpha_\mathcal{H})$ (consistent with the utilities).

We define a *configuration* of our model as parameters $\{(\alpha_\mathcal{L}, \alpha_\mathcal{H}), (\mu, P_h^H, P_h^L)\}$.

基本概念

Mechanism. A *mechanism* $M : \mathcal{R}^n \to \Delta(\{\mathbf{A}, \mathbf{R}\})$ is a decision-making process that outputs a distribution over alternatives based on the report profile $(r_1, r_2, \ldots, r_n) \in \mathcal{R}^n$ of voters. The *report space* $\mathcal{R}$ is specified by the questions posed to agents, which can be about their received signals, votes on alternatives, preferences in certain states, or beliefs about others' signals. The mechanism only sees agents' reports and does not know any component of the instance.

Strategy. A (mix) *strategy* for an agent is a function $\sigma : \mathcal{T} \times \mathcal{S} \to \Delta(\mathcal{R})$ that maps the type and the signal received by the agent to a distribution on possible reports. A strategy is said to be *truthful* if it honestly answers all questions based on the agent's knowledge and signal. When the type of the agents is clear from the context, we slightly abuse the notation and describe a strategy by the function $\sigma : \mathcal{S} \to \Delta(\mathcal{R})$ with the first function input (i.e., the type) omitted.

基本概念

Expected Utility Function. Given a strategy profile $\Sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$, we denote the expected utility of agent i under mechanism M as $u_i^{\mathcal{M}}(\Sigma)$, where the expectation is taken over the sampling of agents' signals (with agents' strategies, it decides the report distributions) and the mechanism's outcome. The expected utility is concerned with *ex-ante* outcomes, as opposed to the *ex-post* utility v_i, which is considered after the world state is realized. In this paper, we focus on *ex-ante* utilities when talking about any equilibrium solution concept.

Equilibrium. In a social choice setting with a large number of agents, it is often the case that any single agent's deviation has a negligible impact on the overall outcome. Therefore, rather than considering the typical Bayes Nash equilibrium, we consider a much stronger concept: the strong Bayes Nash equilibrium, where the collective deviations of subsets of agents are taken into account.

强贝叶斯纳什均衡

我们已经知道强贝叶斯纳什均衡的定义，接下来定义 ϵ-强贝叶斯纳什均衡

定义：ϵ-强贝叶斯纳什均衡

A strategy profile $\Sigma = (\sigma_1, \ldots, \sigma_n)$ is an ϵ-strong Bayes Nash equilibrium if no subset of agents D and alternative profile $\Sigma' = (\sigma'_1, \ldots, \sigma'_n)$ exist such that

1. $\sigma_i = \sigma'_i$ for each $i \notin D$,
2. $u_i(\Sigma') \geq u_i(\Sigma)$ for each $i \in D$,
3. there exist $i \in D$ such that $u_i(\Sigma') > u_i(\Sigma) + \epsilon$.

也就是说，没办法找到一个参与者集合 D，能够满足：这个集合中的参与者通过改变策略，使得对于集合中的每个人，改变策略后的收益不会减少，并且至少有一个人的收益会增加至少 ϵ.

ϵ-强贝叶斯纳什均衡，显然比强贝叶斯纳什均衡弱，它允许均衡策略下，部分人可以通过改变策略实现获利，但是对获利幅度进行了限制．当 $\epsilon = 0$ 时，它就是强贝叶斯纳什均衡．

多数投票机制

多数投票机制是常见且自然的机制，它能够很好地聚合多数人的意见．在多数投票机制中，每个人要么选择同意，要么选择反对．当同意票大于总人数的一半时，决议被通过；当反对票大于总人数一半时，决议被反对；当同意票和反对票相同时，有一半-一半的概率选择同意和反对．

在论文中，一个参与者的（混合）策略可以使用向量 (β_l, β_h) 表示：其中 $0 \leq \beta_s \leq 1$ 对应了当参与者收到信号 $s \in \mathcal{S}$ 时，投同意票 $\mathbf{A}$ 的概率．

实际上我们使用 (δ_l, δ_h) 去表示策略 (β_l, β_h)，这里 $\delta_s \in [-\frac{1}{2}, \frac{1}{2}]$ 表明 β_s 相对于 $\frac{1}{2}$ 的偏离．具体而言我们要求：

$$\beta_l = \frac{1}{2} - \delta_l \text{ and } \beta_h = \frac{1}{2} + \delta_h.$$

多数投票下的一些“诚实”机制

在多数投票机制下，”诚实”策略的定义并不明确：所谓”当接收到信号 l/h 时，如实报告 $\mathbf{A}/\mathbf{R}$ 能反映参与者的真实类型”究竟意味着什么？我们提出几种可能的”诚实”策略定义.

候选诚实策略一：信息性策略

最直观的“诚实”策略是将信号直接映射为投票．例如，对于类型 $\mathcal{H}$ 的参与者，它的策略会是 $(\beta_l, \beta_h) = (0, 1)$，或者说 $(\delta_l, \delta_h) = (\frac{1}{2}, \frac{1}{2})$. 即当收到信号 l 时投 $\mathbf{R}$，收到信号 h 时投 $\mathbf{A}$．这种策略被称为信息性策略.

此策略的局限性很显然：如果信号存在系统性偏差即 P_h^H 和 P_h^L 均大于 $\frac{1}{2}$，那么信息性策略将导致多数群体的错误决策.

候选诚实策略二：真诚策略

另一种”诚实”策略是以贝叶斯更新为基础的期望效用最大化策略．具体而言，参与者已知先验参数 μ，并根据自己接收到的信号进行贝叶斯信念更新，最后选择能最大化其期望效用的投票选项.

事实上，即使所有参与者类型相同，这两种”诚实”策略仍可能无法达成知情多数决策．要实现该目标，参与者需要采取比简单诚实策略更**复杂**的博弈策略.

单一类型参与者

我们将给出一个单一类型参与者情况下，能够导出多数知情决定的策略．这个策略将给更一般的二元对立参与者情况带来不少直觉．
我们假设所有参与者都是 $\mathcal{H}$．对称策略 $\{(\delta_l, \delta_h)\}$ 能够聚合多数知情决定，当它满足以下两个不等式：

$$\Delta_{\mathbf{A}}^{H}(\delta_l, \delta_h) := P_h^H \cdot \delta_h - P_l^H \cdot \delta_l > 0 \text{ and } \Delta_{\mathbf{A}}^{L}(\delta_l, \delta_h) := P_h^L \cdot \delta_h - P_l^L \cdot \delta_l < 0. \quad (1)$$

其中，$\Delta_{\mathbf{A}}^{H}(\delta_l, \delta_h)$ 表示在现实世界为 H 的时候，策略能够给 $\mathbf{A}$ 带来的期望票数增加量（相对于 $\frac{1}{2}$）；$\Delta_{\mathbf{A}}^{L}(\delta_l, \delta_h)$ 对于现实世界为 L 同理．
这两个不等式的直觉解释是：第一个不等式表明了，当现实世界是 H 时，参与者倾向于投 $\mathbf{A}$；而第二个不等式表明了，当现实世界是 L 时，参与者应该倾向于投 $\mathbf{R}$．无论现实世界是哪种情况，这个策略都指向了多数知情决定，即符合 $typeH$ 的利益．由于这是期望意义下，所以根据大数定理，当 $n \to \infty$ 时，满足以上两个不等式的策略能够以概率 1 收敛到多数知情决定．
这两个不等式将是我们后续研究二元类型参与者情况的基础．

二元类型参与者

现在我们转而研究多数投票机制在二元类型参与者下的情况．我们给出一个阈值 θ_{maj}，它决定了什么时候多数群体能够聚合知情多数决定．

定理

In the majority vote mechanism, a critical threshold exists for the majority proportion, denoted as $\theta_{\text{maj}} = \frac{1}{2M}$ *where*

$$M = \begin{cases} \frac{P_l^L}{P_l^L + P_l^H}, & \text{if } P_h^L + P_h^H \leq 1, \\ \frac{P_h^H}{P_h^L + P_h^H}, & \text{otherwise.} \end{cases}$$

二元类型参与者

定理

For this threshold, there exist functions $\epsilon(n)$, $p(n)$, and $\gamma(n)$ all dependent on the number of agents n, such that as $n \to \infty$, $\epsilon(n) \to 0$, $p(n) \to 1$ and $\gamma(n)$ does not approach to 0. For these functions, the following two statements hold:

- *If the majority proportion α exceeds this threshold θ_{maj}, i.e., $\alpha > \theta_{\mathrm{maj}}$, an $\epsilon(n)$-strong Bayes Nash equilibrium exists. Furthermore, any such equilibrium leads to the informed majority decision with a probability of at least $p(n)$.*
- *Conversely, if the majority proportion is below this threshold, i.e., $\alpha \leq \theta_{\mathrm{maj}}$, there is no $\gamma(n)$-strong Bayes Nash equilibrium.*

这个定理事实上在说，如果多数群体占比 α 超过了阈值，那么就能找到一个（近似）强贝叶斯纳什均衡，且这个均衡下能够以接近 1 的概率收敛到知情多数决策；反之，如果多数群体占比低于阈值，那么就不存在强贝叶斯纳什均衡．

二元类型参与者

从直觉上来说，如果多数人比例非常高，那他们就握有绝对的话语权，能够直接决定结果，少数人无论怎么做都无法造成影响．那么，在这样的情况下，只要多数人选择的策略使得他们获得了最优的效用，那么这就是一个强贝叶斯纳什均衡．因为多数人没有理由偏离，而少数人的力量实在过于弱小，它们怎么设计策略都没法改变最终的结果，因此也无法通过改变策略获利．并且多数人的最优效用也指向了多数知情决策，这似乎表明了强贝叶斯纳什均衡与知情多数决策之间的联系．

反过来说，如果多数人的比例相对较小，以至于和少数人的比例相差不大，那么少数人的力量虽然仍小于多数人，但是他们可以通过非常激进的策略（如永远投给 **A** 或 **R** 进行捣乱），以期待在某个世界状态下的胜利，而不是在两个世界状态下均输掉．

多数人的最优决策

不失一般性地我们假设 $typeH$ 是多数派．一个自然的想法是多数派想要最大化期望效用，无论现实世界是哪种情况．所以这里想要最大化在两种情况下的最小值．

引理 (Characterization of Optimal Strategy)

The strategy (δ_l^*, δ_h^*) *that maximizes the function* $P(\cdot,\cdot)$ *is*

$$\begin{cases} \delta_l^* = \frac{1}{2}\cdot\frac{P_h^L+P_h^H}{P_l^L+P_l^H},\ \delta_h^* = \frac{1}{2} & \text{if } \frac{P_h^L+P_h^H}{2} \leq \frac{1}{2}, \\ \delta_l^* = \frac{1}{2},\ \delta_h^* = \frac{1}{2}\cdot\frac{P_l^L+P_l^H}{P_h^L+P_h^H} & \text{otherwise.} \end{cases} \tag{2}$$

where $P(\delta_l,\delta_h) = \min\{p_{\mathbf{A}}^H(\delta_l,\delta_h), p_{\mathbf{R}}^L(\delta_l,\delta_h)\}$ *for* $p_{\mathbf{A}}^H(\delta_l,\delta_h)$ *and* $p_{\mathbf{R}}^L(\delta_l,\delta_h)$.
Here

$$p_{\mathbf{A}}^H(\delta_l,\delta_h) = \frac{1}{2} + \Delta_{\mathbf{A}}^H(\delta_l,\delta_h) \text{ and } p_{\mathbf{R}}^L(\delta_l,\delta_h) = \frac{1}{2} - \Delta_{\mathbf{A}}^L(\delta_l,\delta_h). \tag{3}$$

多数人的最优决策

以上引理实际上是在说，如果信号 h 的频率很低，那么参与者应该在接收到 h 的时候一定投 $\mathbf{A}$，而当接收到 l 的时候不是一定投 $\mathbf{R}$ 而是以一定的概率仍然投 $\mathbf{A}$. 反之亦然，在 l 频率很低的时候，接收到 l 一定投 $\mathbf{R}$，接收到 h 却不一定投 $\mathbf{A}$.

Example

例 (Optimal Strategy and Other Strategies)

Consider the signal distributions in Table 1, where signal h is generally more common.

	Signal h	Signal l
State H	0.75	0.25
State L	0.5	0.5

表: Signal distributions.

According to inequalities in (1), any strategy (δ_l, δ_h) satisfying $\delta_h < \delta_l < 3\delta_h$ will ensure the wish of type-$\mathcal{H}$ agents when no type-$\mathcal{L}$ agents exist. For example, (δ_l, δ_h) can be $(\frac{1}{6}, \frac{1}{8})$. Then agents vote for $\mathbf{A}$ with a probability of $\frac{1}{2} + \frac{1}{8} = \frac{5}{8}$ upon signal h and with a probability of $\frac{1}{2} - \frac{1}{6} = \frac{1}{3}$ upon signal l. Such a strategy makes the expected vote share of $\mathbf{A}$ in state H, and that of $\mathbf{R}$ in state L be $0.75 \times \frac{5}{8} + 0.25 \times \frac{1}{3} = \frac{53}{96}$ and $0.5 \times \frac{3}{8} + 0.5 \times \frac{2}{3} = \frac{25}{48}$, respectively. On the other hand, the optimal strategy for type-$\mathcal{H}$ agents is $(\delta_l^*, \delta_h^*) = (\frac{1}{2}, \frac{3}{10})$. Type-$\mathcal{H}$ agents deterministically vote for $\mathbf{R}$ upon the rarer signal l. The expected vote share of $\mathbf{A}$ in state H and that of $\mathbf{R}$ in state L are both $\frac{3}{5}$. □

多数人的最优决策

在上面的例子中，最优策略得到的结果是在两种世界状态下，多数决策的票数比例是一样的．这并不是偶然，两者相等是最优策略的必要条件．直觉上这很直观，我们想要两个值的最小值最大，那自然就是两个值相等的时候，否则永远可以把更大的那个值变小，更小的值变大，使得两者的最小值变大．

命题 1

The optimal strategy (δ_l^*, δ_h^*) satisfies $p_{\mathbf{A}}^H(\delta_l^*, \delta_h^*) = p_{\mathbf{R}}^L(\delta_l^*, \delta_h^*)$.

多数人的最优决策

证明.

We first refine the possible range for the optimal strategy. Taking arbitrary values of δ_l and δ_h with conditions $\delta_l > 0$, $\delta_h > 0$, and $\frac{P_l^H}{P_h^H} < \frac{\delta_h}{\delta_l} < \frac{P_l^L}{P_h^L}$, we have $p_{\mathbf{A}}^H(\delta_l, \delta_h) = \frac{1}{2} + P_h^H \cdot \delta_h - P_l^H \cdot \delta_l > \frac{1}{2}$ and $p_{\mathbf{R}}^L(\delta_l, \delta_h) = \frac{1}{2} + P_l^L \cdot \delta_l - P_h^L \cdot \delta_h > \frac{1}{2}$. Then the optimal values δ_l^*, δ_h^* must satisfy

$\min\{p_{\mathbf{A}}^H(\delta_l^*, \delta_h^*), p_{\mathbf{R}}^L(\delta_l^*, \delta_h^*)\} \geq \min\{p_{\mathbf{A}}^H(\delta_l, \delta_h), p_{\mathbf{R}}^L(\delta_l, \delta_h)\} > \frac{1}{2}$. Thus, we can deduce that δ_l^* and δ_h^* fall within the range $(0, \frac{1}{2}]$. (If they are outside this range, it is impossible to ensure that both $p_{\mathbf{A}}^H(\delta_l^*, \delta_h^*) > \frac{1}{2}$ and $p_{\mathbf{R}}^L(\delta_l^*, \delta_h^*) > \frac{1}{2}$.)

Then we show strategy (δ_l, δ_h) with $\delta_l, \delta_h \in (0, \frac{1}{2}]$ and $p_{\mathbf{A}}^H(\delta_l, \delta_h) \neq p_{\mathbf{R}}^L(\delta_l, \delta_h)$ is sub-optimal by providing a better strategy. □

多数人的最优决策

证明.

- If $p_{\mathbf{A}}^{H}(\delta_l, \delta_h) < p_{\mathbf{R}}^{L}(\delta_l, \delta_h)$, given that δ_l lies in the range $(0, \frac{1}{2}]$, we can adjust δ_l slightly to create a new strategy (δ_l^+, δ_h^+) where $\delta_l^+ = \delta_l - \epsilon$ (for some small ϵ such that $0 < \epsilon < \delta_l$), and $\delta_h^+ = \delta_h$. For a sufficiently small ϵ, it follows that $p_{\mathbf{A}}^{H}(\delta_l, \delta_h) < p_{\mathbf{A}}^{H}(\delta_l^+, \delta_h^+) < p_{\mathbf{R}}^{L}(\delta_l^+, \delta_h^+) < p_{\mathbf{R}}^{L}(\delta_l, \delta_h)$. Therefore, the minimum of $p_{\mathbf{A}}^{H}$ and $p_{\mathbf{R}}^{L}$ for (δ_l^+, δ_h^+) is greater than that for (δ_l^*, δ_h^*), indicating a better strategy.
- If $p_{\mathbf{A}}^{H}(\delta_l, \delta_h) > p_{\mathbf{R}}^{L}(\delta_l, \delta_h)$, we adjust δ_h to $\delta_h^+ = \delta_h - \epsilon$ (where $0 < \epsilon < \delta_h$) and keep $\delta_l^+ = \delta_l$. For a sufficiently small ϵ, we get $p_{\mathbf{A}}^{H}(\delta_l, \delta_h) > p_{\mathbf{A}}^{H}(\delta_l^+, \delta_h^+) > p_{\mathbf{R}}^{L}(\delta_l^+, \delta_h^+) > p_{\mathbf{R}}^{L}(\delta_l, \delta_h)$. Hence, the minimum of $p_{\mathbf{A}}^{H}$ and $p_{\mathbf{R}}^{L}$ for (δ_l^+, δ_h^+) is again greater than that for (δ_l, δ_h), proving the superiority of the new strategy.

□

有了这个结论，我们就能进行一些计算来推导出引理的结论.

多数人的最优决策

引理的证明

根据命题 1，有

$$\frac{1}{2} + P_h^H \cdot \delta_h^* - P_l^H \cdot \delta_l^* = \frac{1}{2} + P_l^L \cdot \delta_l^* - P_h^L \cdot \delta_h^*.$$

我们能够整理得到比例关系：

$$\frac{\delta_h^*}{\delta_l^*} = \frac{P_l^L + P_l^H}{P_h^L + P_h^H} = \frac{2 - (P_h^L + P_h^H)}{P_h^L + P_h^H} = \frac{1}{(P_h^L + P_h^H)/2} - 1.$$

因此能够把 $p_{\mathbf{A}}^H$ ($p_{\mathbf{R}}^L$ 也类似) 表示为：

$$\frac{1}{2} + P_h^H \cdot \delta_h^* - P_l^H \cdot \delta_l^* = \frac{1}{2} + P_h^H \cdot (\delta_h^* + \delta_l^*) - \delta_l^* = \frac{1}{2} + \delta_l^* \cdot \left[P_h^H \left(\frac{\delta_h^*}{\delta_l^*} + 1 \right) - 1 \right],$$

可以发现最大值由 δ_l^* 决定.

多数人的最优决策

引理的证明（续）

δ_l^* 和 δ_h^* 都在 $(0, \frac{1}{2}]$ 上，我们能够根据 $\frac{P_h^L + P_h^H}{2}$ 的值来确定最优解:

- If $\frac{P_h^L + P_h^H}{2} \le \frac{1}{2}$ (indicating that $\delta_h^* \ge \delta_l^*$), the optimal values are $\delta_h^* = \frac{1}{2}$ and $\delta_l^* = \frac{1}{2} \cdot \frac{P_h^L + P_h^H}{P_l^L + P_l^H}$.
- If $\frac{P_h^L + P_h^H}{2} > \frac{1}{2}$ (indicating that $\delta_h^* < \delta_l^*$), the optimal values are $\delta_l^* = \frac{1}{2}$ and $\delta_h^* = \frac{1}{2} \cdot \frac{P_l^L + P_l^H}{P_h^L + P_h^H}$.

无穷情况下的分析

在前面研究的单一类型参与者情况下，当 n 趋于无穷时，如果两个不等式被满足那么根据大数定理能直接（以概率 1 收敛）得到知情多数决策，受此启发我们可以在二元类型参与者情况下也从 n 趋于无穷的情形开始分析．
我们首先考虑，当多数群体使用上述的最优方案时，对他们而言，多数决策的投票占比是多少（期望上，更受欢迎的那一个选项的得票率），也就是在引理中我们想要最大化的那个值，我们记 $M = P(\delta_l^*, \delta_h^*)$，可以计算出

$$M = \begin{cases} \frac{P_l^L}{P_l^L + P_l^H} & \text{if } \frac{P_h^L + P_h^H}{2} \leq \frac{1}{2}, \\ \frac{P_h^H}{P_h^L + P_h^H} & \text{otherwise.} \end{cases} \tag{4}$$

当多数群体远大于少数群体时

当多数群体远大于少数群体时，仅由多数群体带来的某一选项投票率就已经超过一半 ($\alpha M > \frac{1}{2}$)，这时少数群体就算采取最激进的策略（全部投另一个选项）也无法扭转这个结果，因此没有任何改进方案的空间，容易发现这是一个强贝叶斯纳什均衡.

引理

For configuration $\{(\alpha_{\mathcal{L}}, \alpha_{\mathcal{H}}), (\mu, P_h^H, P_h^L)\}$ *where the majority proportion* $\alpha > \frac{1}{2M}$ *and the number of agents* $n \to \infty$*, any strategy profile* $(\sigma_1, \sigma_2, \ldots, \sigma_n)$ *such that the majority type-*$\mathcal{H}$ *agents adopt the optimal strategy forms a strong Bayes Nash equilibrium in the majority vote mechanism. Such equilibria lead to the informed majority decision with a probability of* 1.

根据大数定理这个引理的结论是显然的．以上引理是“好的”均衡，因为它确保了知情多数决策．那么是否存在“不好的”均衡？答案是否定的，在多数投票机制中，任何强贝叶斯纳什均衡都能确保知情多数决策.

引理

For the number of agents $n \to \infty$*, any strong Bayes Nash equilibrium in the majority vote mechanism leads to the informed majority decision with a probability of* 1.

强贝叶斯纳什均衡导出知情多数决策

直觉上无论少数派怎么决策，多数派都能派出等量的人数、使用相反的策略以抵消对方的投票（使得自己的投票加上对方的投票刚好在期望上是 50-50 的比例），剩下的人再使用能够确保自己想要的结果占优的策略，这样就能确保知情多数决策.

证明

We show the majority agents can benefit from deviation if the winning probability of the informed majority decision is less than 1.

Suppose the minority type-$\mathcal{L}$ agents take strategies (either symmetric or asymmetric) such that the expected fraction of $\mathbf{A}$ votes upon signal l and signal h are $\chi_l^{(\mathcal{L})}$ and $\chi_h^{(\mathcal{L})}$. In this case, the impact of the minority agents is equivalent to that when they adopt a symmetric strategy (δ_l, δ_h) with $\frac{1}{2} - \delta_l = \chi_l^{(\mathcal{L})}$ and $\frac{1}{2} + \delta_h = \chi_h^{(\mathcal{L})}$. Some type-$\mathcal{H}$ agents can conduct a "counter-strategy" to absorb the minority type's impact on votes. That is, the symmetric strategy $(-\delta_l, -\delta_h)$ so that $\beta_l = \frac{1}{2} - (-\delta_l) = 1 - \chi_l^{(\mathcal{L})}$ and $\beta_h = \frac{1}{2} + (-\delta_h) = 1 - \chi_h^{(\mathcal{L})}$. The remaining of the majority type-$\mathcal{H}$ agents can ensure the dominance of the informed majority decision adopting the optimal strategy (δ_l^*, δ_h^*), or any strategy satisfying inequalities in (1).

强贝叶斯纳什均衡导出知情多数决策

证明（续）

Formally, the majority type-$\mathcal{H}$ agents can take strategy $(-\delta_l, -\delta_h)$ with a probability of $\frac{1-\alpha}{\alpha}$ and the optimal strategy (δ_l^*, δ_h^*) otherwise. This strategy divides the majority into two groups. One group of size $(1-\alpha)\cdot n$ negates the minority's impact, while the other group of size $(2\alpha-1)\cdot n$ secures the informed majority decision in both states. The winning probability of the informed majority decision will increase to 1 when the majority-type agents take the above-mentioned strategy. Therefore, any strategy profile that does not lead to the informed majority decision cannot be a strong Bayes Nash equilibrium.

当多数群体规模较小时

从另一方面来说，当多数群体规模较小的时候，它们没办法做到无视少数派的意见.

引理

For configuration $\{(\alpha_{\mathcal{L}}, \alpha_{\mathcal{H}}), (\mu, P_h^H, P_h^L)\}$ *where the majority proportion* $\alpha \leq \frac{1}{2M}$ *and the number of agents* $n \to \infty$, *no strong Bayes Nash equilibrium exists in the majority vote mechanism.*

假设存在强贝叶斯纳什均衡，根据前面的引理知道得到知情多数决策的概率为 1. 接下来将使用和前面类似的方法证明少数派能够通过激进的策略来降低得到知情多数决策的概率.

当多数群体规模较小时

证明

Suppose the majority type-$\mathcal{H}$ agents take strategies (either symmetric or asymmetric) such that the expected fraction of $\mathbf{A}$ votes among them upon receiving signal l and signal h are $\chi_l^{(\mathcal{H})}$ and $\chi_h^{(\mathcal{H})}$. In this case, the impact of the majority agents is equivalent to that when they adopt a symmetric strategy (δ_l, δ_h) with $\frac{1}{2} - \delta_l = \chi_l^{(\mathcal{H})}$ and $\frac{1}{2} + \delta_h = \chi_h^{(\mathcal{H})}$. We can assume without loss of generality that the majority type-$\mathcal{H}$ agents play symmetric strategy (δ_l, δ_h).

By Lemma 3, the minimum vote share of the informed majority decision among the majority-type agents, $P(\delta_l, \delta_h) = \min\{p_{\mathbf{A}}^H(\delta_l, \delta_h), p_{\mathbf{R}}^L(\delta_l, \delta_h)\}$, is at most $P(\delta_l^*, \delta_h^*) = M$. Given that $\alpha M \leq \frac{1}{2}$, either $\alpha p_{\mathbf{A}}^H(\delta_l, \delta_h) \leq \frac{1}{2}$ or $\alpha p_{\mathbf{R}}^L(\delta_l, \delta_h) \leq \frac{1}{2}$ holds. That is, the majority type-$\mathcal{H}$ agents' votes are not enough to secure the informed majority decision. For strategy profiles that lead the winning probability of the informed majority decision to be 1, the minority type-$\mathcal{L}$ agents can form a coalition to vote strategically and increase their utilities. If $\alpha p_{\mathbf{A}}^H(\delta_l, \delta_h) \leq \frac{1}{2}$, type-$\mathcal{L}$ agents can deterministically vote for $\mathbf{R}$ to make the vote share of $\mathbf{R}$ in total be more than one half in state H. If $\alpha p_{\mathbf{R}}^L(\delta_l, \delta_h) \leq \frac{1}{2}$, type-$\mathcal{L}$ agents can deterministically vote for $\mathbf{A}$ to make $\mathbf{A}$ win in state L.

对于有限的 n

我们证明了对于 n 趋于无穷的时候，有很好的结论．从这一点出发，在 n 有限的情况下，仍然能够得到类似的且性质不错的结论，即近似强贝叶斯纳什均衡和较高概率的知情多数决策．

引理

For configuration $\{(\alpha_{\mathcal{L}}, \alpha_{\mathcal{H}}), (\mu, P_h^H, P_h^L)\}$ where the majority proportion $\alpha > \frac{1}{2M}$, any strategy profile $\Sigma = (\sigma_1, \ldots, \sigma_n)$ such that the majority type-$\mathcal{H}$ agents adopt the optimal strategy forms an ϵ-strong Bayes Nash equilibrium in the majority vote mechanism, where $\epsilon = 2B^2 \exp(-2c^2\lfloor \alpha n \rfloor)$ and c is a constant defined by $c = \frac{1}{3}(\alpha M - \frac{1}{2})$.
Meanwhile, the informed majority decision wins the majority vote with probability at least $1 - 2\exp(-2c^2\lfloor \alpha n \rfloor)$.

引理

Any ϵ-strong Bayes Nash equilibrium in the majority vote mechanism leads to the informed majority decision with a probability of at least $1 - 2\epsilon$, where $\epsilon = B\exp(-2c^2 n)$ and constant c is defined as $\frac{1}{2}(2\alpha - 1)(M - \frac{1}{2})$.

对于有限的 n

引理

For configuration $\{(\alpha_{\mathcal{L}}, \alpha_{\mathcal{H}}), (\mu, P_h^H, P_h^L)\}$ where the majority proportion $\alpha \leq \frac{1}{2M}$, no ϵ-strong Bayes Nash equilibrium exists in the majority vote mechanism. Here, ϵ is defined as $\epsilon = \frac{1}{4}\min\{\mu, 1-\mu\} - 2B^2\exp(-2c^2 n)$ and constant c is $\frac{1}{2}(2\alpha - 1)\min\{M - \frac{1}{2}, \frac{1}{2} - \alpha M\}$.

可以发现，在 n 有限的情况下，仍然能够得到类似的结论，阈值依然稳定.

新机制的提出

在前面，我们研究了多数投票机制下的投票博弈，并给出了一个阈值；现在提出一个新的机制，在该机制下能够得到一个更低的阈值，并且能够证明，当 α 低于这个阈值的时候，即使 $n \to \infty$，对于任意的匿名机制，都无法做到以概率 1 得到知情多数决策.

定理

There exists an anonymous mechanism $\{(\alpha_{\mathcal{L}}, \alpha_{\mathcal{H}}), (\mu, P_h^H, P_h^L)\}$ *with a critical threshold* $\theta^* = \frac{1}{\Delta+1}$ *for the majority proportion such that*

- *If majority proportion* α *exceeds this threshold* θ^**, i.e.,* $\alpha > \theta^*$*, then truthful reporting becomes an* $\epsilon(n)$*-strong Bayes Nash equilibrium and this equilibrium ensures the informed majority decision with probability at least* $p(n)$*, where* $\epsilon(n) \to 0$ *and* $p(n) \to 1$ *as* $n \to \infty$.

容易证明，这里的阈值 $\theta^* = \frac{1}{\Delta+1}$ 小于等于之前的阈值 $\theta_{\text{maj}} = \frac{1}{2M}$.

机制设计

接下来给出机制的具体设计：

Our Mechanism

1. **Information Collection.**
   Each agent i reports his/her private type ($\mathcal{L}$ or $\mathcal{H}$), the received signal (l or h), and a threshold value $\hat{\delta}_i$ defined as $\frac{P_l^L}{P_l^L + P_h^H}$.
2. **Majority Type Identification.**
   Determine the majority-type, $\hat{M}$, from the agents' reports.
3. **Threshold Calculation.**
   Calculate the collective threshold, $\hat{\delta}$, as the median of the reported $\hat{\delta}_i$ values.
4. **State Assessment.**
   Determine the true world state, $\hat{\omega}$, based on whether the reporting frequency of l exceeds the threshold $\hat{\delta}$. Set $\hat{\omega} = L$ if the threshold is exceeded; $\hat{\omega} = H$ otherwise.
5. **Outcome Selection.**
   Choose the preferred alternative of majority-type $\hat{M}$ in world state $\hat{\omega}$.

机制设计

新机制引出一个阈值 $\hat{\delta}$，它由每个参与者进行报告，这看起来需要参与者对现实世界有着精确的估计．但是我们可以通过询问参与者对世界状态的后验预测 (e.g., $\Pr[\omega = L \mid S_i = l]$) 和对其他人接收到的信号的估计 (e.g., $\Pr[S_j = l \mid s_i = h]$)，以一种比较自然的方式获得这个值．通过向所有人询问这个值并取中位数，机制期望能够得到 δ 的准确的值．$\delta = \frac{P_l^L}{P_l^L + P_h^H}$ 的以下特性起着关键作用：

命题 2

$$P_l^L > \frac{P_l^L}{P_l^L + P_h^H} > P_l^H \text{ and } P_h^L < \frac{P_h^H}{P_l^L + P_h^H} < P_h^H. \tag{5}$$

更进一步地，对于 $\alpha > \frac{1}{\Delta + 1}$，我们有以下不等式：

$$\alpha P_l^L > \frac{P_l^L}{P_l^L + P_h^H} > \alpha P_l^H + (1 - \alpha) \tag{6a}$$

$$\alpha P_h^L + (1 - \alpha) < \frac{P_h^H}{P_l^L + P_h^H} < \alpha P_h^H \tag{6b}$$

不等式性质证明

证明

It suffices to demonstrate that $\alpha P_l^L > \frac{P_l^L}{P_l^L+P_h^H}$ and $\frac{P_h^H}{P_l^L+P_h^H} < \alpha P_h^H$ hold for $1/(\Delta+1) < \alpha \leq 1$. The rest inequalities follow directly from them because each term is 1 minus a corresponding term, for example, $\alpha P_l^H + (1-\alpha)$ is $1 - \alpha P_h^H$.

Recall that $\Delta = P_l^L - P_l^H = P_h^H - P_h^L$, it can be derived that $P_l^L + P_h^H = P_l^L + P_h^L - P_h^L + P_h^H = 1 + \Delta$. Thus, we have $\alpha P_l^L > \frac{P_l^L}{\Delta+1} = \frac{P_l^L}{P_l^L+P_h^H}$ and $\frac{P_h^H}{P_l^L+P_h^H} = \frac{P_h^H}{\Delta+1} < \alpha P_h^H$, which concludes the proof.

不等式 (5) 实际上在说，当所有参与者采取诚实策略时，机制能正确聚合知情多数决策．因为它表示着当真实状态为 L 时，参与者收到信号 l 的比例高于阈值；当真实状态为 H 时，参与者收到信号 l 的比例低于阈值．这保证机制能通过信号分布准确判断真实状态．
不等式 (6) 实际上确保了诚实策略构成近似强贝叶斯纳什均衡，即少数派无法通过策略性投票扭曲结果．

高概率知情多数决策

引理

If all agents report truthfully, our mechanism will output the informed majority decision with probability at least $1 - 2\exp(-2c^2 n)$. The constant c is defined as

$$c = \frac{1}{3}\min\left\{P_l^L - \frac{P_l^L}{P_l^L + P_h^H}, \frac{P_l^L}{P_l^L + P_h^H} - P_l^H\right\}.$$

高于阈值带来的近似强贝叶斯纳什均衡

引理

For configuration $\{(\alpha_{\mathcal{L}}, \alpha_{\mathcal{H}}), (\mu, P_h^H, P_h^L)\}$ where the majority proportion $\alpha > \frac{1}{\Delta+1}$ and large enough n, truthful reporting forms an ϵ-strong Bayes Nash equilibrium in our mechanism. Here, we define ϵ as $\epsilon = \max\{B\exp(-2c^2\lfloor\alpha n\rfloor), 2B^2\exp(-2c^2 n)\}$, and define the constant c as

$$c = \frac{1}{3}\min\left\{\alpha P_l^L - \frac{P_l^L}{P_l^L + P_h^H}, \frac{P_l^L}{P_l^L + P_h^H} - \left[\alpha P_l^H + (1-\alpha)\right]\right\}.$$

当 n 趋于无穷时，根据前面的引理，我们能够得到知情多数决策的概率为 1．因此多数派没有偏离的动机．对于少数派，它们在这个情况下的力量很小，无法改变结果：

- In state L, even if all minority agents report signal h, the fraction reporting signal l will be αP_l^L. By the inequalities in (6), this lower bound of signal l's reporting rate, exceeds the threshold $\hat{\delta} = \frac{P_l^L}{P_l^L + P_h^H}$.
- In state H, even if all minority agents report signal l, the fraction reporting signal l will be $\alpha P_l^H + (1-\alpha)$. Inequalities in (6) show that this upper bound falls below the threshold.

低于阈值时的坏结果

我们研究当 $\alpha \leq \theta^*$ 时，关于强贝叶斯纳什均衡的满足性．研究结果表明没有一个合理的机制能够符合要求．

定理

For configuration $\{(\alpha_{\mathcal{L}}, \alpha_{\mathcal{H}}), (\mu, P_h^H, P_h^L)\}$ where the majority proportion $\alpha \leq \frac{1}{\Delta+1}$, $0 < \mu < 1, P_h^H = 1/2 + \Delta/2$, and $P_h^L = 1/2 - \Delta/2$, even the number of agents $n \to \infty$, in any anonymous mechanism M, truthful reporting cannot form an ϵ-strong Bayes Nash equilibrium that leads to the informed majority decision with probability 1. Here, ϵ is defined as $\epsilon = \frac{1}{4}\min\{\mu, 1-\mu\}$.

定理

For configuration $\{(\alpha_{\mathcal{L}}, \alpha_{\mathcal{H}}), (\mu, P_h^H, P_h^L)\}$ where the majority proportion $\alpha \leq \frac{1}{\Delta+1}$, $0 < \mu < 1, P_h^H = 1/2 + \Delta/2$, and $P_h^L = 1/2 - \Delta/2$, even the number of agents $n \to \infty$, in any anonymous mechanism M, no symmetric ϵ-strong Bayes Nash equilibrium can guarantee the informed majority decision with probability 1. Here, we define ϵ as $\epsilon = \frac{1}{4}\min\{\mu, 1-\mu\}$.

总结

1. 单一类型参与者 → 二元类型参与者
2. 知情多数决策 → 强贝叶斯纳什均衡
3. 多数投票机制
   1. 多数人的最优决策
   2. 以阈值为界的两种情况
4. 新机制的提出