21188142: 课程综合实践 II（数据要素交易基础） **2025-2026 学年短学期**

HW 1: 博弈论与多臂老虎机算法基础

教师：刘金飞，助教：吴一航 日期: 2025 年 6 月 27 日

1.1 占优策略均衡与纳什均衡的关系

证明如下关于占优策略均衡与纳什均衡的关系的结论：

1. 如果每个参与人 i 都有一个占优于其它所有策略的策略 s_i^*，那么 $s^* = (s_1^*, \cdots, s_n^*)$ 是纳什均衡；
2. 如果每个参与人 i 都有一个严格占优于其它所有策略的策略 s_i^*，那么 $s^* = (s_1^*, \cdots, s_n^*)$ 是博弈的唯一纳什均衡。

证明: 见 2024 年作业 2.2 题。 ■

1.2 N 人古诺竞争

假设在古诺竞争中，一共有 J 家企业。当市场中所有企业总产量为 q 时，市场价格为 $p(q) = a - bq$。且每个企业生产单位产品的成本都是同一个常数 c，即企业 i 的产量为 q_i 时该企业的成本为 $c_i(q_i) = c \cdot q_i$。假设 $a > c \geqslant 0$，$b > 0$，

1. 求纳什均衡下所有企业的总产量以及市场价格；
2. 讨论均衡价格随着 J 变化的情况，你有什么启示?
3. 讨论 $J \to \infty$ 的均衡结果，你有什么启示?

设企业 i 的产量为 q_i，除 i 外所有企业的总产量为 q_{-i}，所有企业总产量为 $q = q_i + q_{-i}$。对任意的企业 i，可以写出其利润为

$$q_i(a - b(q_i + q_{-i})) - cq_i,$$

不难得到一阶条件为

$$a - 2bq_i - bq_{-i} - c = 0,$$

将上式对所有的 i 求和可得 $Ja - 2bq - b(J-1)q - Jc = 0$，进而可以得到总产量

$$q = \frac{J}{J+1} \cdot \frac{a-c}{b},$$

对应的价格为

$$p = a - bq = \frac{a + Jc}{J + 1}.$$

可以看出，随着 J 增大，竞争程度增加，价格下降；$J \to \infty$ 时价格等于单位成本，对应完全竞争市场的情况。

1.3 公地悲剧

假设有 I 个农场主，每个农场主均有权利在公共草地上放牧奶牛。一头奶牛产奶的数量取决于在草地上放牧的奶牛总量 N：当 $N < \overline{N}$ 时，n_i 头奶牛产生的收入为 $n_i \cdot v(N)$；而当 $N \geqslant \overline{N}$ 时，$v(N) \equiv 0$。假设每头奶牛的成本为 c，且 $v(0) > c$，$v' < 0$，$v'' < 0$，所有农场主同时决定购买多少奶牛，所有奶牛均会在公共草地上放牧（注：假设奶牛的数量可以是小数，也就是无需考虑取整的问题）。

1. 将上述情形表达为策略式博弈；
2. 求博弈的纳什均衡下所有农场主购买的总奶牛数（可以保留表达式的形式，不用求出具体解）；
3. 求所有农场主效用之和最大（社会最优）情况下的总奶牛数（可以保留表达式的形式，不用求出具体解），并与上一问的结果比较，你能从中得到什么启示?

表达为策略式博弈太简单，略。下面求解均衡，首先注意到均衡情况下奶牛总量不可能超过 $\overline{N}$：否则此时存在农场主的效用为负数，这些农场主不如把策略改为一头奶牛也不买。设农场主 i 购买奶牛数量为 n_i，除 i 外所有农场主的总奶牛数为 n_{-i}，所有农场主总奶牛数为 $N = n_i + n_{-i}$。对任意的农场主 i，可以写出其利润为

$$n_i \cdot v(n_i + n_{-i}) - c \cdot n_i,$$

一阶条件为

$$v(n_i + n_{-i}) + n_i \cdot v'(n_i + n_{-i}) - c = 0,$$

对所有的 i 的一阶条件求和有

$$v(N) + \frac{1}{I} N \cdot v'(N) - c = 0. \tag{1.1}$$

下面求解社会最优，社会福利函数为

$$N \cdot v(N) - c \cdot N,$$

最大化的一阶条件为

$$v(N) + N \cdot v'(N) - c = 0. \tag{1.2}$$

记式1.1的解为 N_1，式1.2的解为 N_2，有 $N_1 > N_2$：否则如果 $N_1 \leqslant N_2$，根据 $v' < 0, v'' < 0$ 的性质，有 $v(N_1) \geqslant v(N_2)$ 和 $0 > N_1 v(N_1) \geqslant N_2 v(N_2)$，因此式1.1的等式左侧大于式1.2的等式左侧，产生了矛盾。启示：公地悲剧的囚徒困境本质，即自私的行动无法达到社会最优。

1.4 贝叶斯纳什均衡

考虑如下的不完全信息博弈：

- $I = \{1, 2\}$：1 和 2 分别是行、列参与人
- $T_1 = \{A, B\}, T_2 = \{C\}$：参与人 1 有两个类型，参与人 2 有一个类型
- $p(A, C) = \dfrac{1}{3}, p(B, C) = \dfrac{2}{3}$
- 每个参与人有两个可能的行动，下图所示矩阵给出了两种类型向量下的收益矩阵（左图为 $t = (A, C)$ 时的博弈，右图为 $t = (B, C)$ 时的博弈）：

	L	R
T	2, 0	0, 3
B	0, 4	1, 0

	L	R
T	0, 3	3, 1
B	2, 0	0, 1

求解该博弈的所有贝叶斯纳什均衡.

方法与 2024 年作业 4.2 基本一致，设 x 表示 A 选 T 的概率，y 表示 B 选 T 的概率，p 表示参与人 2 选 L 的概率。首先可以排除参与人 2 的纯策略，然后对参与人 2 利用无差异原则得到 $x = \dfrac{2 + 6y}{7}$。根据 $x, y \in [0, 1]$ 可知参与人 1 仅有的纯策略是 $x = 1$ 和 $y = 0$。考虑 A 选择混合策略的情况，根据无差异条件可以求出 $p = 1/3$，而 B 选择混合策略时同理有 $p = 3/5$，这表明参与人 1 的两种类型不可能同时选择混合策略。因此考虑如下两种情况：

1. A 选择纯策略 $x = 1$，B 选择混合策略 $y = 5/6$，$p = 3/5$，需要验证 $x = 1$ 是否是 A 的最优反应，不难验证的确是，因此构成均衡；
2. B 选择纯策略 $y = 0$，A 选择混合策略 $x = 2/7$，$p = 1/3$，但此时 B 不是最优反应，不构成均衡。

1.5 混合策略的不完全信息解释

考虑以下**抓钱博弈 (grab the dollar)**：桌子上放 1 块钱，桌子的两边坐着两个参与人，如果两人同时去抓钱，每人罚款 1 块；如果只有一人去抓，抓的人得到那块钱；如果没有人去抓，谁也得不到什么。因此，每个参与人的策略是决定抓还是不抓。

抓钱博弈描述的是下述现实情况：一个市场上只能有一个企业生存，有两个企业在同时决定是否进入。如果两个企业都选择进入，各亏损 100 万；如果只有一个企业进入，进入者盈利 100 万；如果没有企业进入，每个企业既不亏也不盈。

1. 求抓钱博弈的纯策略纳什均衡；

		参与人 2	
		抓	不抓
参与人 1	抓	$-1,-1$	$1,0$
	不抓	$0,1$	$0,0$

2. 求抓钱博弈的混合策略纳什均衡；

纯策略均衡：（抓，不抓），（不抓，抓）；混合策略均衡：每人以 1/2 概率抓/不抓。

现在考虑同样的博弈但具有如下不完全信息：如果参与人 i 赢了，他的利润是 $1+\theta_i$（而不是 1）。这里 θ_i 是参与人的类型，参与人 i 自己知道 θ_i，但另一个参与人不知道。假定 θ_i 在 $[-\varepsilon,\varepsilon]$ 区间上均匀分布。

		参与人 2	
		抓	不抓
参与人 1	抓	$-1,-1$	$1+\theta_1,0$
	不抓	$0,1+\theta_2$	$0,0$

由于两个参与人的情况完全对称，故考虑如下对称贝叶斯纳什均衡（两个人的策略相同）形式：参与人 $i(i=1,2)$ 的策略均为

$$s_i(\theta_i)=\begin{cases}\text{抓}, & \text{如果}\theta_i\geqslant\theta^*,\\ \text{不抓}, & \text{如果}\theta_i<\theta^*.\end{cases}$$

即 θ^* 是两个参与人抓或不抓的类型分界阈值，其中 θ^* 是一个待计算确定的参数。

3. 求 θ^*；

4. 当 $\varepsilon\to 0$ 时，上述贝叶斯纳什均衡会收敛于什么？从中你能得到怎样的启示。

给定参与人 j 的策略，参与人 i 选择抓的期望效用为

$$\left(1-\frac{\theta^*+\varepsilon}{2\varepsilon}\right)\cdot(-1)+\frac{\theta^*+\varepsilon}{2\varepsilon}\cdot(1+\theta_i),$$

显然 $\theta_i=\theta^*$ 时上式等于 0（抓的效用大于 / 小于不抓的效用 0 的分界线），故有

$$(2+\varepsilon)\theta^*+(\theta^*)^2=0,$$

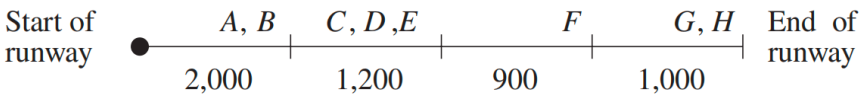
解得 $\theta^*=0$。故只要 $\theta_i\geqslant 0$ 选择抓，否则不抓。因为 $\theta_i\geqslant 0$ 和 $\theta_i<0$ 的概率各为 1/2，每个参与人在选择自己的策略时认为对方抓和不抓的概率均为 1/2，似乎面对的是一个混合策略均衡，但实际上是纯策略均衡。当 $\varepsilon\to 0$ 时，上述贝叶斯纯策略均衡就收敛于完全信息博弈的混合策略均衡。

1.6 飞机跑道成本分配的沙普利值计算

机场跑道的维护费用通常是向在那个机场降落飞机的航空公司来收取的。但是轻型飞机所需的跑道长度比重型飞机所需的跑道长度短，这就带来了一个问题，如何在拥有不同类型飞机的航空公司之间确定公平的维护费用分摊。

定义一个成本博弈 $(N;c)$（即每个联盟的效用是成本函数 c），这里 N 是降落在这个机场上的所有飞机的集合，$c(S)$（对每个联盟 S）是能够允许联盟中所有飞机降落的最短跑道的维护费用。如果用沙普利值来确定费用的分摊，**证明：每段跑道的维护费用由使用那段跑道的飞机均摊**。

下图描绘了一个例子，其中标号为 A, B, C, D, E, F, G 和 H 的八架飞机每天都要在这个机场降落。每架飞机所需的跑道的整个长度由图中的区间来表示。例如，飞机 F 需要前三个跑道区间。每个跑道区间的每周维护费用标示在图的下面。例如，$c(A, D, E) = 3200$，$c(A) = 2000$ 和 $c(C, F, G) = 5100$。在这一例子中，A 的沙普利值恰好等于 $2000/8 = 250$，而 F 的沙普利值等于 $2000/8 + 1200/6 + 900/3 = 750$。你的任务是将这一性质推广到一般的情形下给出证明（**提示：使用沙普利值的性质和公式的特点**）。



思路：以上图为例，将每个联盟 S 的效用（成本）都拆成四个部分：$c(S) = w(S) + x(S) + y(S) + z(S)$。其中 $w(S)$ 表示联盟 S 仅针对第一个跑道区间产生的成本，因此对所有的非空联盟 S 都有 $w(S) = 2000$，不难发现博弈 $(N;w)$ 中所有飞机都是对称的，根据沙普利值的对称性可知在博弈 $(N;w)$ 中每个人都要付出 $2000/8 = 250$ 的成本。类似地，$x(S)$ 表示联盟 S 仅针对第二个跑道区间产生的成本，则除了集合 $\{A, B\}$ 的子集之外的其他联盟 S 都满足 $x(S) = 1200$，而集合 $\{A, B\}$ 的子集 S 满足 $x(S) = 0$。根据零参与人性质，飞机 A, B 的成本为 0，根据对称性，其他飞机平摊 1200。类似地，$y(S)$ 和 $z(S)$ 分别表示仅针对第三、四个跑道区间产生的成本，重复上述讨论的理由可以证明结论（当然严谨的叙述需要跳出这一例子，给出更一般化的证明，但本质上没有区别）。

1.7 ε-贪心算法的遗憾分析

令 $\varepsilon_t = t^{-1/3}(K \log t)^{1/3}$，证明：$\varepsilon$-贪心算法的遗憾界为 $O(T^{2/3}(K \log T)^{1/3})$。

证明：分析时刻 $t+1$，在前 t 时刻中期望出现 $\sum_{i=1}^{t} \varepsilon_i$ 次探索，则每个臂被选中的平均次数为 $\sum_{i=1}^{t} \varepsilon_i / K$，则有

$$P(|\mu_{t+1}(a) - Q_{t+1}(a)| \leqslant \varepsilon) \geqslant 1 - 2\exp\left(-2\varepsilon^2 \frac{\sum_{i=1}^{t} \varepsilon_i}{K}\right)$$

令 $\varepsilon = \sqrt{\dfrac{K \log t}{\sum_{i=1}^{t} \varepsilon_i}}$，上式变为：

$$P\left(|\mu_{t+1}(a) - Q_{t+1}(a)| \leqslant \sqrt{\frac{K \log t}{\sum_{i=1}^{t} \varepsilon_i}} \right) \geqslant 1 - 2t^{-2}$$

将上述事件定义为 E，其补集为 $\bar{E}$。在事件 E 下，对于产生遗憾的时刻 $t+1$，有 $\mu(a^*) - \mu(a) \leqslant 2\sqrt{\dfrac{K \log t}{\sum_{i=1}^{t} \varepsilon_i}}$。假设 ε_t 非增（即 $\varepsilon_t \geqslant \varepsilon_{t+1}$），$t+1$ 时刻产生的遗憾为

$$\begin{aligned}
\mathbb{E}[R_{t+1}] &= \mathbb{E}[R_{t+1}^{E}] + \mathbb{E}[R_{t+1}^{\bar{E}}] \\
&\leqslant P(\text{exploration} \mid E) \times 1 + P(\text{exploitation} \mid E) \times (\mu(a^*) - \mu(a)) + P(\bar{E}) \times 1 \\
&\leqslant \varepsilon_{t+1} + (1 - \varepsilon_{t+1}) \times 2\sqrt{\frac{K \log t}{\sum_{i=1}^{t} \varepsilon_i}} + O(t^{-2}) \\
&\leqslant \varepsilon_{t+1} + 2\sqrt{\frac{K \log t}{\sum_{i=1}^{t} \varepsilon_i}} + O(t^{-2}) \leqslant \varepsilon_t + 2\sqrt{\frac{K \log t}{t\varepsilon_t}} + O(t^{-2})
\end{aligned}$$

令 $\varepsilon_t = \sqrt{\dfrac{K \log t}{t\varepsilon_t}}$（$\varepsilon_t = \left(\dfrac{1}{t} K \log t\right)^{1/3}$），则有

$$\mathbb{E}[R_{t+1}] \leqslant 3\left(\frac{1}{t} K \log t\right)^{1/3} + O(t^{-2})$$

因此，前 T 轮的整体遗憾界为

$$\begin{aligned}
\mathbb{E}[R(T)] &= 1 + \sum_{t=1}^{T-1} \mathbb{E}[R_{t+1}] \\
&\leqslant 1 + \sum_{t=1}^{T-1} 3\left(\frac{1}{t} K \log t\right)^{1/3} + O(t^{-2}) \\
&\leqslant 1 + 3(K \log T)^{1/3} \sum_{t=1}^{T-1} t^{-1/3} + O(t^{-2}) \\
&= O((T^{2/3})(K \log T)^{1/3})
\end{aligned}$$

最后一个等式可由 $\sum_{t=1}^{T} t^{-1/3} \leqslant \int_1^T t^{-1/3} dt = 3/2(T^{2/3} - 1)$ 证得，或用琴生不等式和调和级数求和（$\sum_{t=1}^{T} \frac{1}{t} = \ln T + \gamma$，其中 γ 为欧拉常数）证得。