第四讲：完全信息动态博弈

兴军亮

Junliang.Xing@ia.ac.cn

2020年10月15日

中国科学院大学
University of Chinese of Academy of Sciences

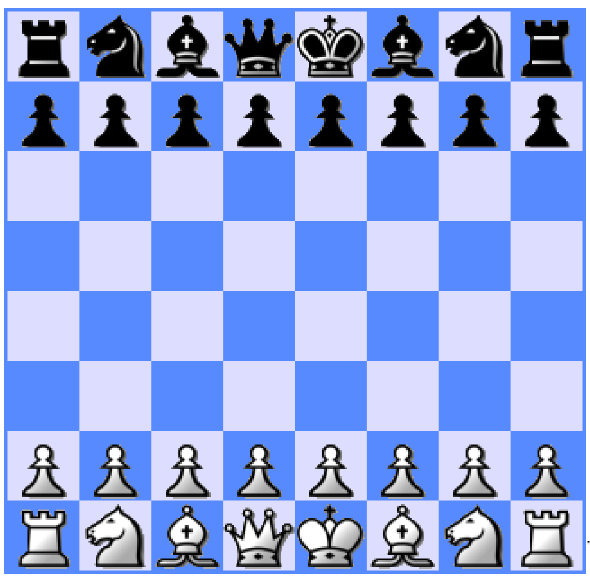
自动化研究所
CASIA
Institute of Automation

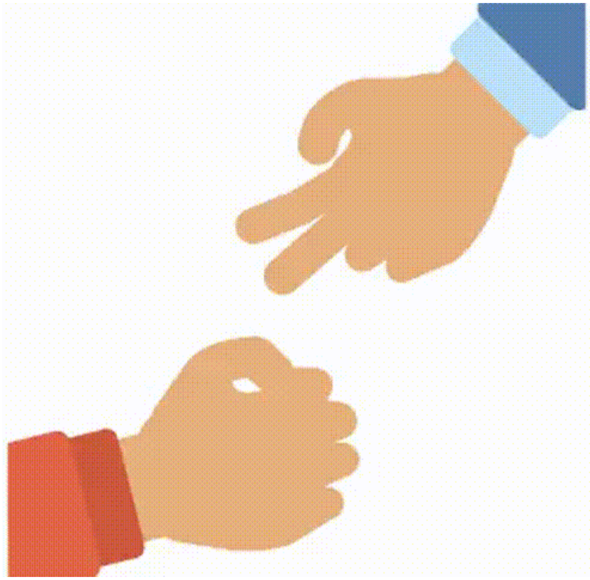
本讲提纲

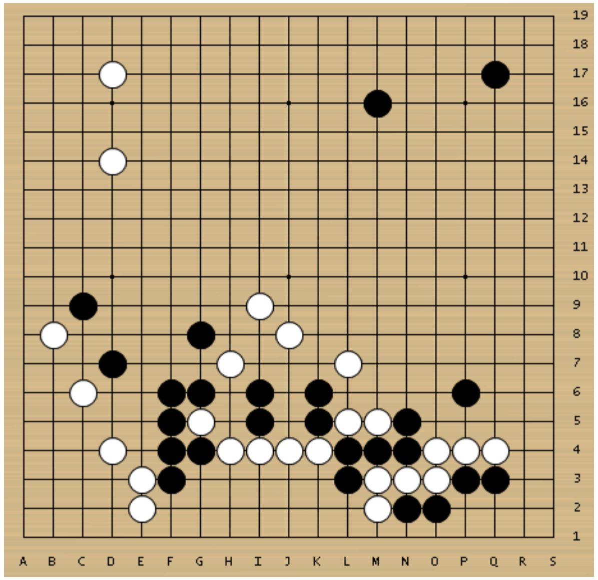
- 定义及其扩展式表示
- 扩展式博弈均衡分析
- 子博弈精炼纳什均衡
- 重复博弈的均衡问题

完全信息动态博弈

- 现实世界中博弈的很多实际应用问题中都存在着多阶段决策的动态结构，为了表示和分析这种动态结构，博弈研究者提出了动态博弈这种类型的博弈加以研究。
- 当动态过程中的博弈各方完全知道博弈的信息，就是我们这一讲要的完全信息动态博弈。







完全信息动态博弈

- 博弈定义：博弈过程中参与人预先知道关于博弈的所有信息，参与人的行动有先后顺序，且后行动者能够观察到先行动者行动的博弈。

- 完全信息动态博弈需要反应的要素
  - 参与人集合：$i \in N, N = (1,2,\ldots,n)$，参与人中可以包括自然
  - 参与人的行动顺序：谁在什么时候行动
  - 参与人的行动空间：每次行动时，有什么选择
  - 信息集合：每次行动结束，参与人知道些什么
  - 效益函数：每次行动时，参与人知道将得到什么
  - 外生事件（即自然选择）的概率分布

博弈的展开式表示

- 博弈树：树中每一对结点有且仅由一个树支与其相连，树的根是博弈的起始结点。
  - 结点：$x \in X \doteq \{x\}$，关系符号$\prec$表示结点的先后顺序关系。具体包括机会结点、决策结点和终结点；
  - 路径：结点x之前的全部结点集合称为x的前列集，又称为x的路径，记为$P(x)$;
  - 后续结点：结点x之后的所有结点的集合称为x的后续结点集，记为$T(x)$;
  - 跟随：当结点y位于结点x的路径上，即$y \in P(x)$，则称x跟随y;
  - 树枝：结点x到其直接跟随结点y的连线，树枝代表一个可供参与人选择的行动策略或者时间;
  - **信息集**：博弈决策结点集合的一个子集，是某个参与人无法区分的决策节点组成的集合，满足：1）每个决策结点都是同一个参与人的决策结点；2）该参与人知道博弈进入该节点子集，但是不知道自己具体在哪一个决策节点。

博弈的展开式表示举例

- 可观察情形下的硬币配对：有两个参与人（1和2），每人有一个硬币，其中一个人将自己的硬币的放下，正面刚朝上（以H表示）或反面朝上（以T表示），另一人**看到这个结果**，然后也将自己的硬币放下。如果两枚硬币配对成功（都正面朝上或者反面朝上），那么参与人2给1一元钱，反之参与人1给2一元钱。图1.1和图1.2是该博弈的两个版本，分别为参与人1先放硬币和参与人2先放硬币。

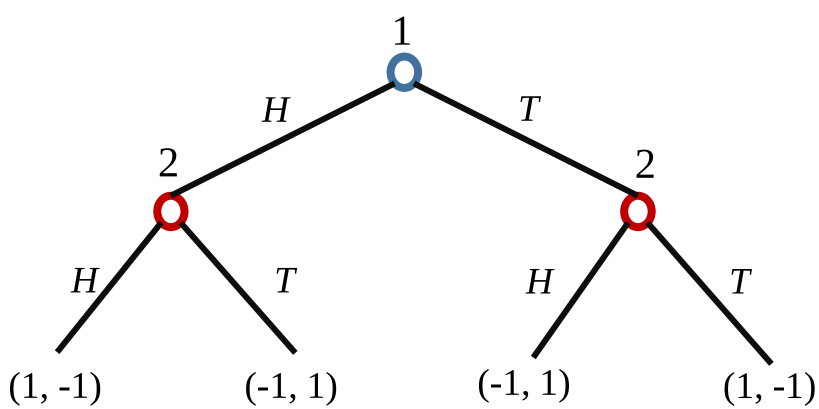


图1.1：可观察情形下的硬币配对（参与人1先行动）

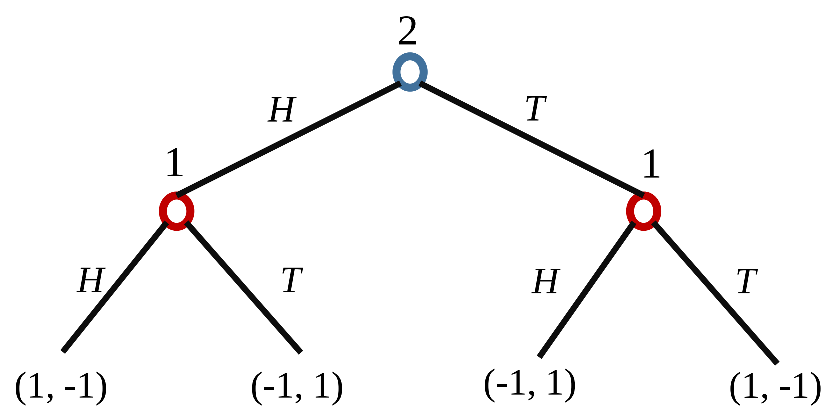


图1.2：可观察情形下的硬币配对（参与人2先行动）

博弈的展开式表示举例

- 不可观察情形下的硬币配对：有两个参与人（1和2），每人有一个硬币，其中一个人将自己的硬币的放下，正面刚朝上（以H表示）或反面朝上（以T表示），另一人<u>**看不到这个结果**</u>，然后也将自己的硬币放下。如果两枚硬币配对成功（都正面朝上或者反面朝上），那么参与人2给1一元钱，反之参与人1给2一元钱。图2.1和图2.2是该博弈的两个版本，分别为参与人1先放硬币和参与人2先放硬币。

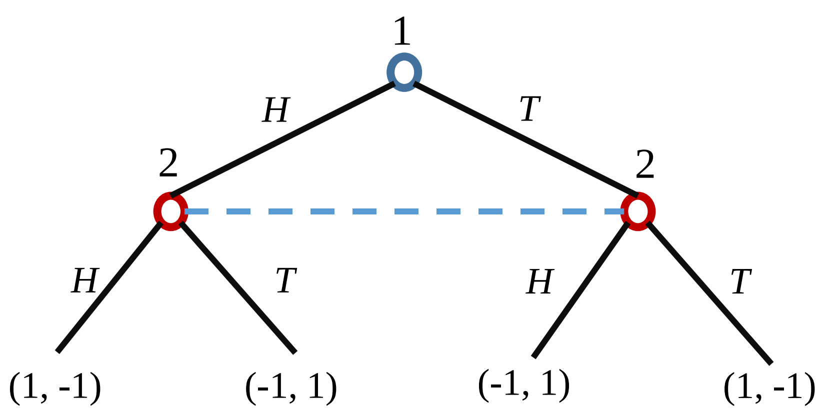


图2.1：不可观察情形下的硬币配对（参与人1先行动）

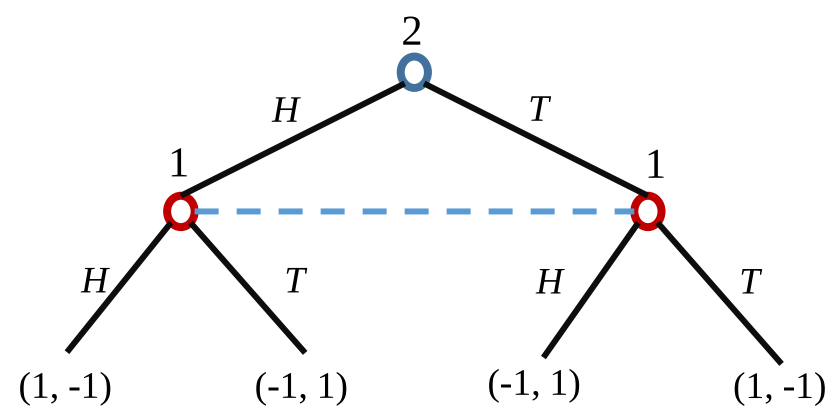


图2.2：不可观察情形下的硬币配对（参与人2先行动）

博弈的展开式表示举例

- 房地产开发博弈中的不同信息集
  - 图3.1：B知道A和N的选择
    - 共7个信息集
  - 图3.2：B知道A但不知N的选择
    - 共5个信息集
  - 图3.3：B知道N但不知A的选择
    - 共5个信息集

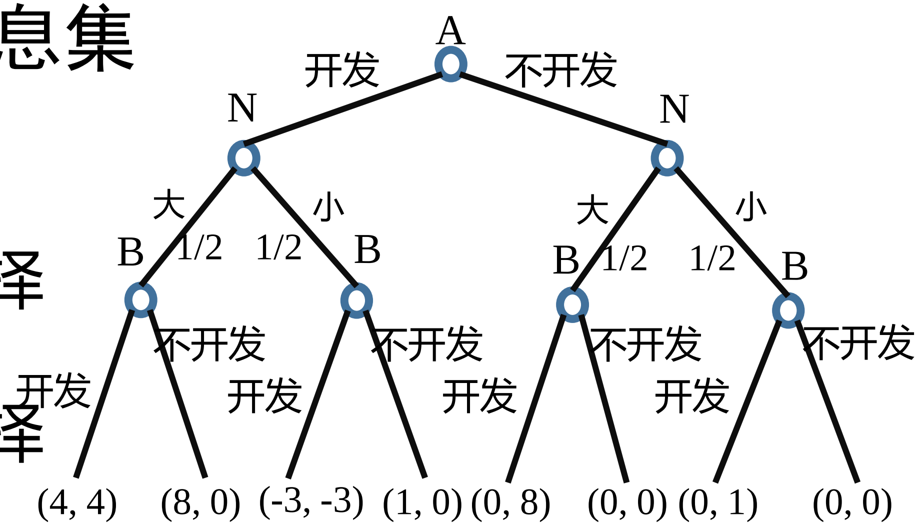


图3.1：房地产开发博弈（完全信息）

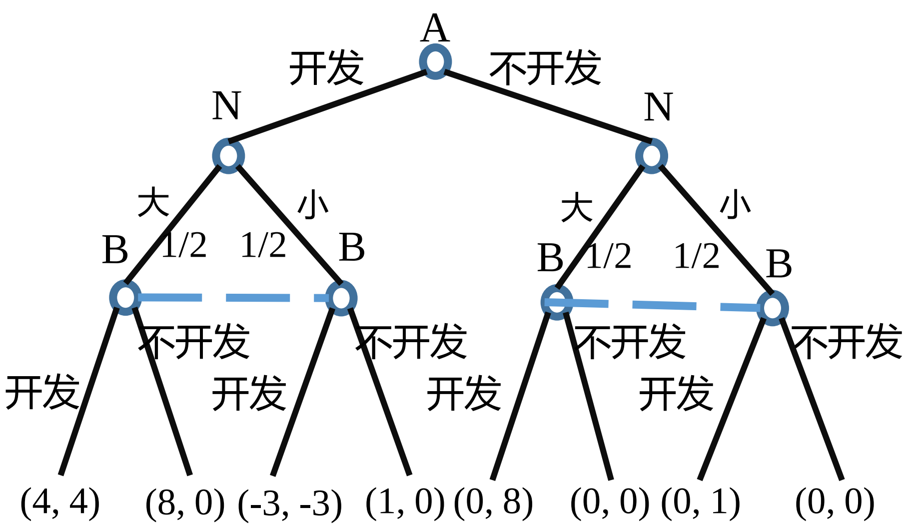


图3.2：房地产开发博弈（B知道A但不知道N）

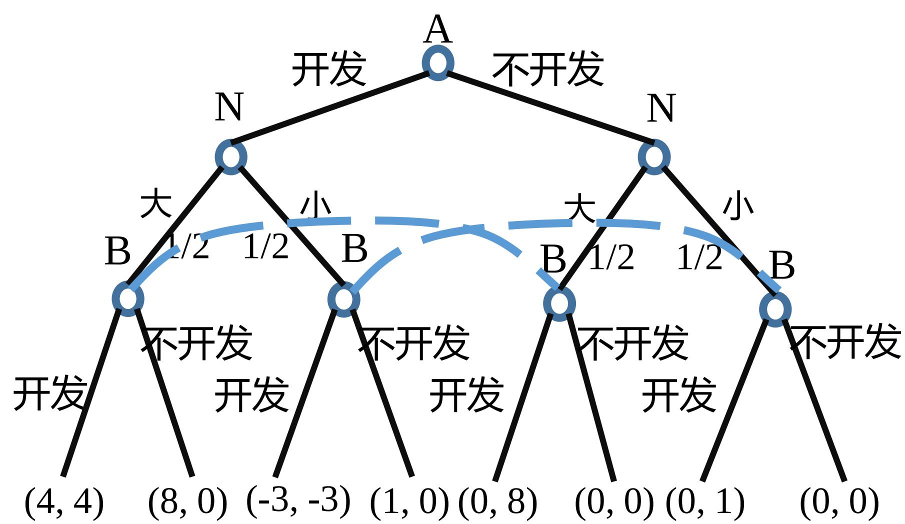
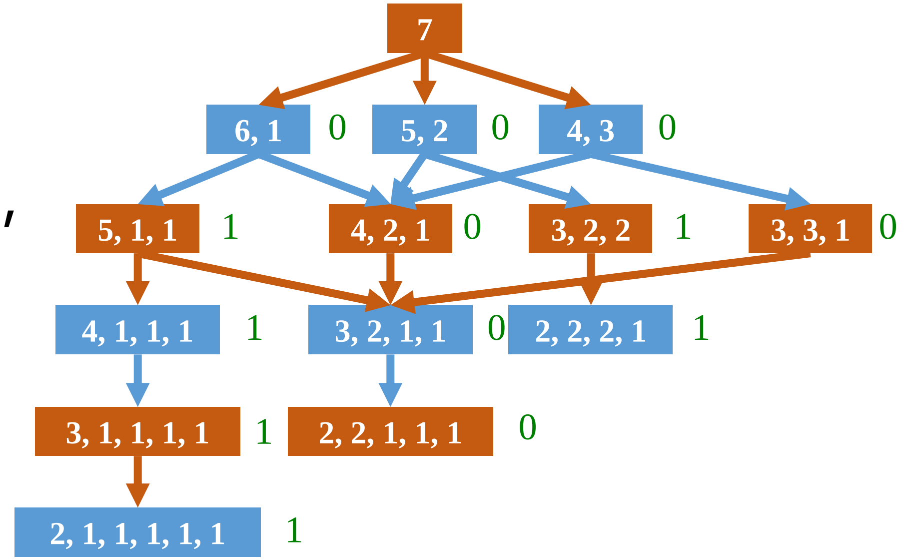


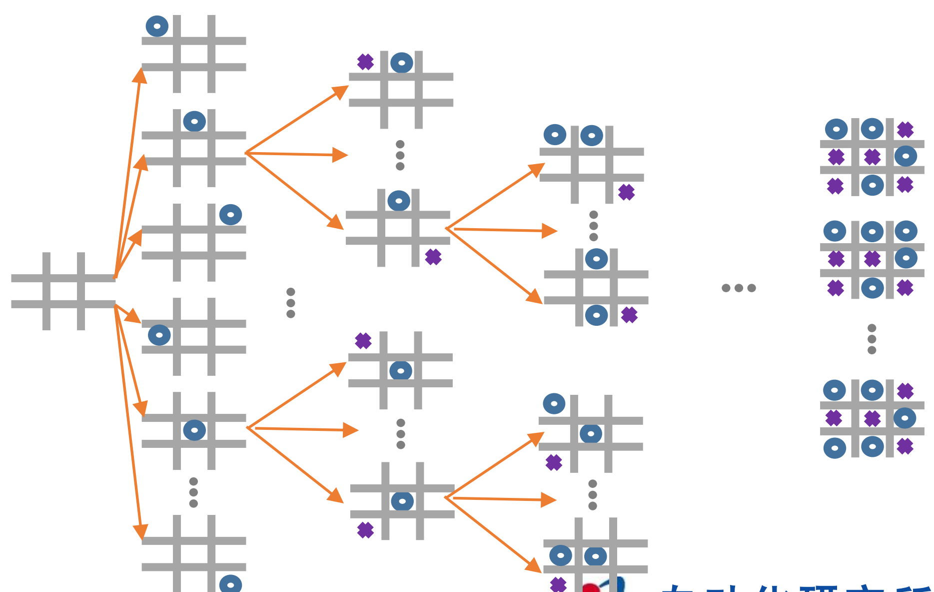
图3.3：房地产开发博弈（B知道N但不知道A）

完全信息动态博弈的展开式表示举例

- 分硬币游戏的博弈树表示
  - 有7枚硬币，只能将已分好的一堆硬币分成个数不等的两堆，当每堆只有1枚或2枚硬币则不能再分，红蓝双方轮流进行，直到谁不能再分，则为输。



- 井字棋游戏的博弈树表示
  - 一个三乘三的格子，两人相继放叉和圈圈两种子，谁先将三子连成一条线谁胜。



博弈的展开式表示

博弈展开式表示形式化描述

展开式博弈可以表示成一个多元组$\Gamma = \langle N, (A_i)_{i\in N}, \mathbb{H}, P, (\mathbb{I}_i)_{i\in N}, (\mu_i)_{i\in N}\rangle$，其中：

- $N = \{1,2,\cdots,n\}$是参与人集，这是一个有限集合；
- A_i是参与人i的行动集；
- $\mathbb{H}$是由所有终止历史（terminal history）组成的集合，其中一个终止历史是一条从根节点到终止节点的行动路径，使得它不是任何其他终止历史的的真子历史；以$S_{\mathbb{H}}$表示由所有终止历史的所有真子历史（包括空历史ϵ）组成的集合；
- $P: S_{\mathbb{H}} \to N$是一个参与人函数（player function），它将每个真子历史与相应的人联系在一起；
- $\mathbb{I}_i(i = 1,2,\cdots,n)$是由参与人$i$的所有信息集组成的集合；
- $\mu_i: \mathbb{H} \to \mathbb{R}\ (i = 1,2,\cdots,n)$ 给出了参与人 i 每个终止历史的效用。

博弈的展开式表示

- 以图1.1中的硬币配对博弈为例，说明上述形式化对应的含义： $\Gamma = \langle N, (A_i)_{i\in N}, \mathbb{H}, P, (\mathbb{I}_i)_{i\in N}, (\mu_i)_{i\in N}\rangle$
  - $N = \{1,2\}$;
  - $A_1 = A_2 = \{H, T\}$；
  - $\mathbb{H} = \{(H,H),(H,T),(T,H),(T,T)\}$，$S_{\mathbb{H}} = \{\epsilon, H, T\}$；
  - $P(\epsilon) = 1$，$P(H) = 2$，$P(T) = 2$；
  - $\mathbb{I}_1 = \{\{\epsilon\}\}$，$\mathbb{I}_2 = \{\{H\},\{T\}\}$；
  - $\mu_1(HH) = 1, \mu_1(HT) = -1,$ $\mu_1(TH) = -1, \mu_1(TT) = 1;$
  - $\mu_2(HH) = -1, \mu_2(HT) = 1,$ $\mu_2(TH) = 1, \mu_2(TT) = -1;$

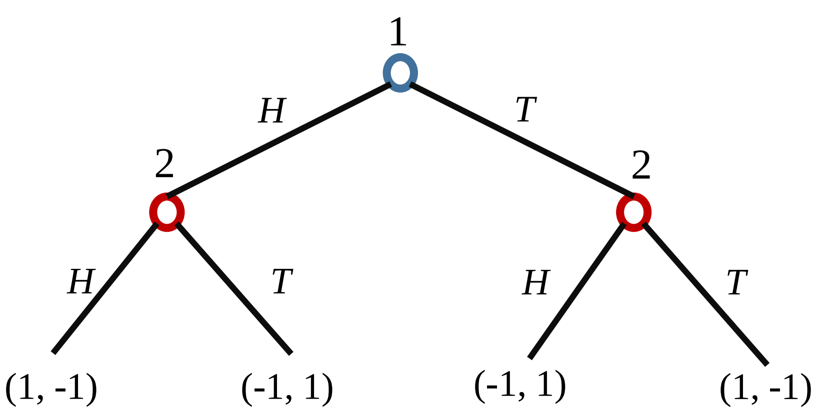


图1.1：可观察情形下的硬币配对（参与人1先行动）

完美信息的判断和定义

- 完美信息：与信息集相关的一个概念，它指的是没有参与人会忘记自己以前知道的事情，所有参与人都知道自己以前的选择。

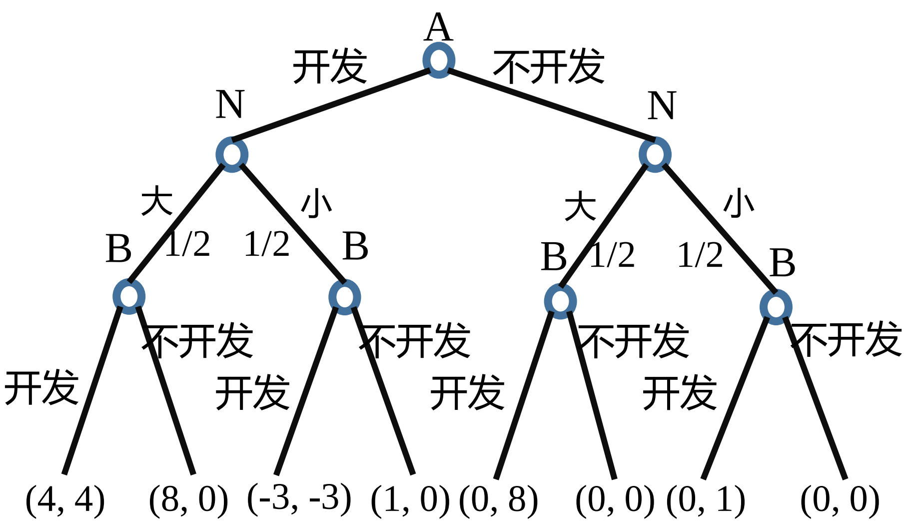
完美信息博弈与不完美信息博弈的一般判别方法

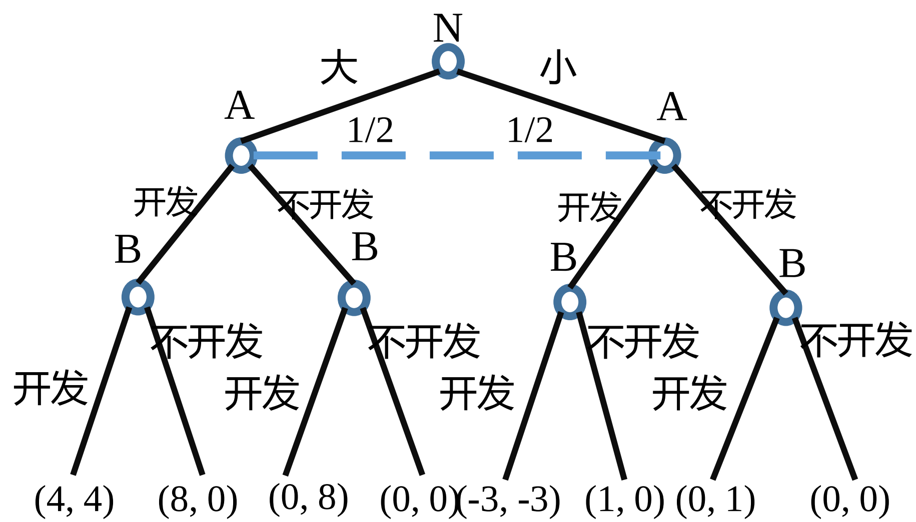
给定一个展开型博弈，如果博弈树的所有信息集都是单点集，该博弈称为完美信息博弈（Game of Perfect Information）；如果至少一个参与人至少有一个含有两个或多个元素的信息集，那么它是不完美信息博弈（Game of Imperfect Information）

- 完美信息博弈意味着：博弈中没有任何两个参与人同时行动，并且所有后行动者能够确切地知道先行动者选择了什么行动，所有参与人都可以观测到自然的决策。

完美信息的判断和定义

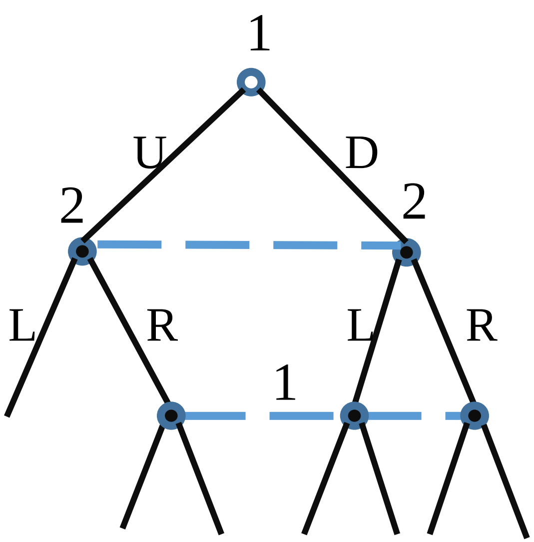
- 完美信息判断的一种特殊情况：“自然”的顺序
  - 房地产开发博弈为例（假设A不知道市场需求的大小）
  - A 在不知道自然的选择时，得到的博弈树如下图所示，是一个不完美信息博弈（A的一个信息集包含两个决策结）。

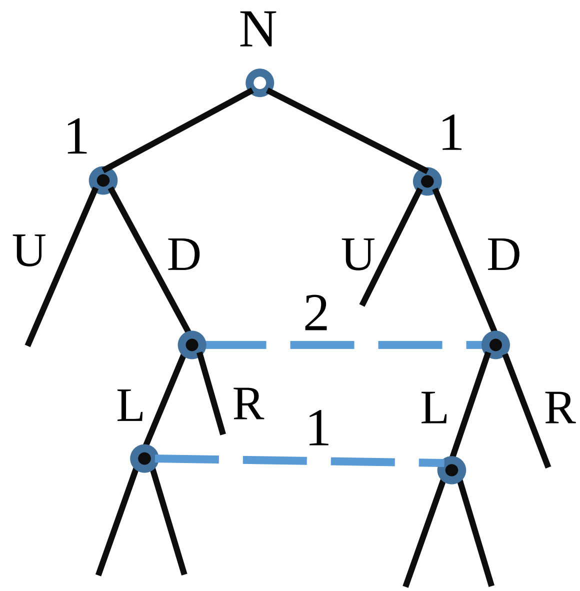


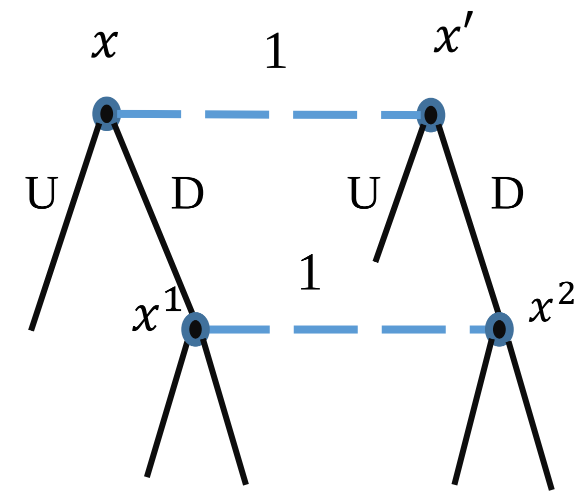


完美信息的判断和定义

- 完美信息判断的一种特殊情况：涉及到虚拟参与人“自然”的博弈。下图1和下图2是两个不完美信息的例子：







- 上图3是保证完美信息的条件：如果x^1和x^2属于同一个信息集，x是x^1的一个前列结，那么一定存在x^2的一个前列结x'和x属于同一个信息集。

扩展式表示与策略式表示的转化

- 策略：一个完整的行动方案

策略

参与人i 的策略(strategy)s_i 是映射$s_i\colon \mathbb{I}_i \to A_i$，使得$s_i(J) \in C(J), \forall J \in \mathbb{I}_i$，其中$C(J)$表示参与人$i$在信息集$J$上的所有可能行动组成的集合。

- 行为策略：扩展式博弈中的混合战略，指的是参与人在每一个信息集上随机地选择行动。

行为策略

令$\Delta(A(h_i))$为定义在行动集合$A(h_i)$上的概率分布，b_i为参与人i的一个行为战略，那么b_i是笛卡尔积$\times_{h_i \in H_i} \Delta(A(h_i))$中的一个元素。

- 如果h_i^1出现，将以$\Delta(A(h_i^1))$的概率选择$A(h_i^1)$

扩展式表示与策略式表示的转化

- 例：硬币配对博弈（可观测情形）
- 扩展式表示转化为策略式表示
  - 先表示所有终止历史组成的集合
    - $\mathbb{I}_1 = \{\{\epsilon\}\}$
    - $\mathbb{I}_2 = \{\{H\},\{T\}\}$
  - 参与人1的两个策略
    - $s_{11}: \{\epsilon\} \to H$
    - $s_{12}: \{\epsilon\} \to T$
  - 参与人2有四个策略
    - $s_{21}: \{H\} \to H; \{T\} \to H$
    - $s_{22}: \{H\} \to H; \{T\} \to T$
    - $s_{21}: \{H\} \to T; \{T\} \to H$
    - $s_{21}: \{H\} \to T; \{T\} \to T$

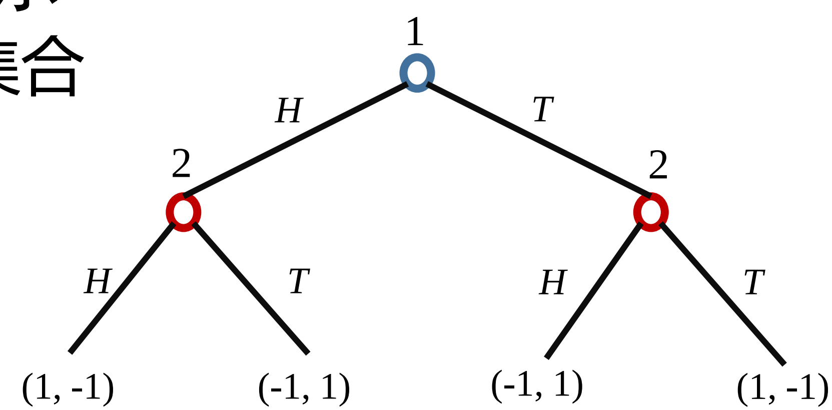


图1.1：可观察情形下的硬币配对（参与人1先行动）

1	2			
	s_{21}	s_{22}	s_{23}	s_{24}
s_{11}	1,-1	1,-1	-1,1	-1,1
s_{22}	-1,1	1,-1	-1,1	1,-1

扩展式表示与策略式表示的转化

- 例：硬币配对博弈（不可观测情形）
- 扩展式表示转化为策略式表示
  - 先表示所有终止历史组成的集合
    - $\mathbb{I}_1=\{\{\epsilon\}\}$
    - $\mathbb{I}_2=\{\{H\},\{T\}\}$
  - 参与人1的两个策略
    - $s_{11}:\{\epsilon\}\to H$
    - $s_{12}:\{\epsilon\}\to T$
  - 参与人2也有两个策略
    - $s_{21}:\{H,T\}\to \mathrm{H}$
    - $s_{22}:\{H,T\}\to T$

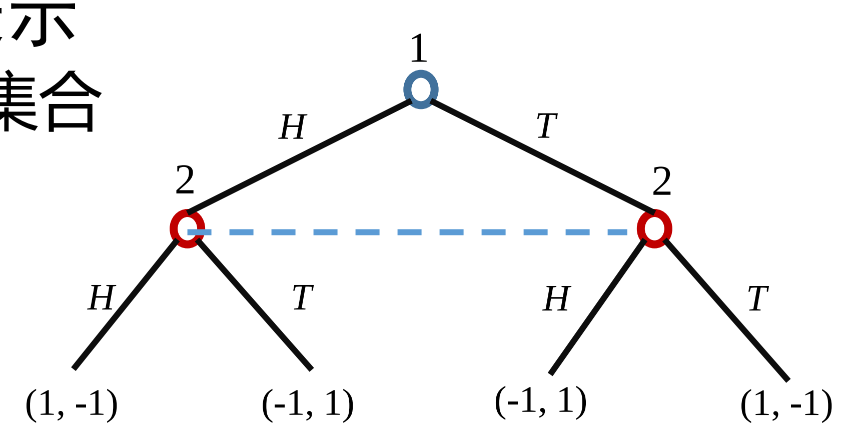
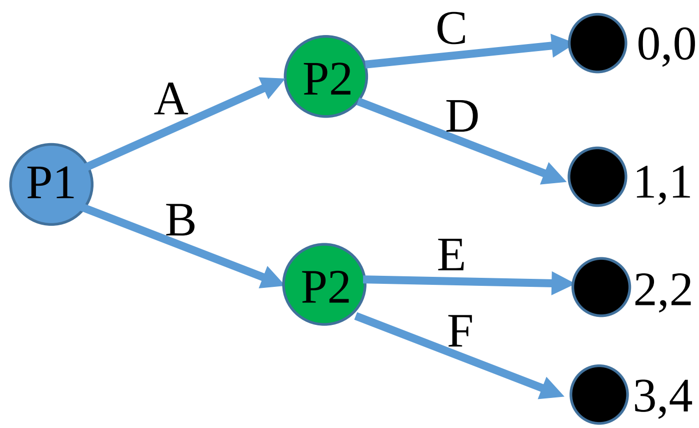


图2.1：不可观察情形下的硬币配对（参与人1先行动）

1	2	
	s_{21}	s_{22}
s_{11}	1,-1	-1,1
s_{22}	-1,1	1,-1

扩展式表示与策略式表示的转化

- 扩展式表示转换为策略式表示的基本方法：
  - 写出所有参与人的纯策略集合，集合大小确定了矩阵每一维度的大小，然后用收益函数对每个元素进行填充。



扩展式表示

$S_1 = \{A,B\}$

$S_2 = \{CE,CF,DE,DF\}$

	CE	CF	DE	DF
A	0,0	0,0	1,1	1,1
B	2,2	3,4	2,2	3,4

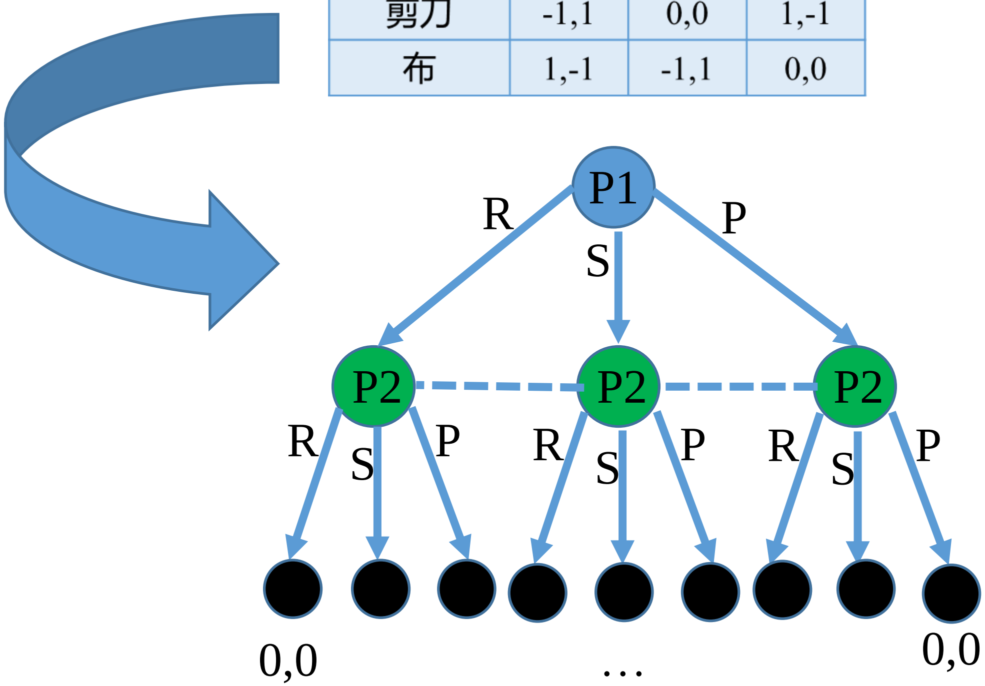
矩阵式表示

扩展式表示与策略式表示的转化

- 策略式表示转换为扩展式表示的基本方法：
  - 选取动作一方作为根节点，每一个可行策略作为一个分支；
  - 利用信息集来表示动作同时进行，叶节点写上相应的收益。

玩家1 \ 玩家2	石头	剪刀	布
石头	0, 0	1,-1	-1,1
剪刀	-1,1	0,0	1,-1
布	1,-1	-1,1	0,0

石头剪刀布的策略式表示

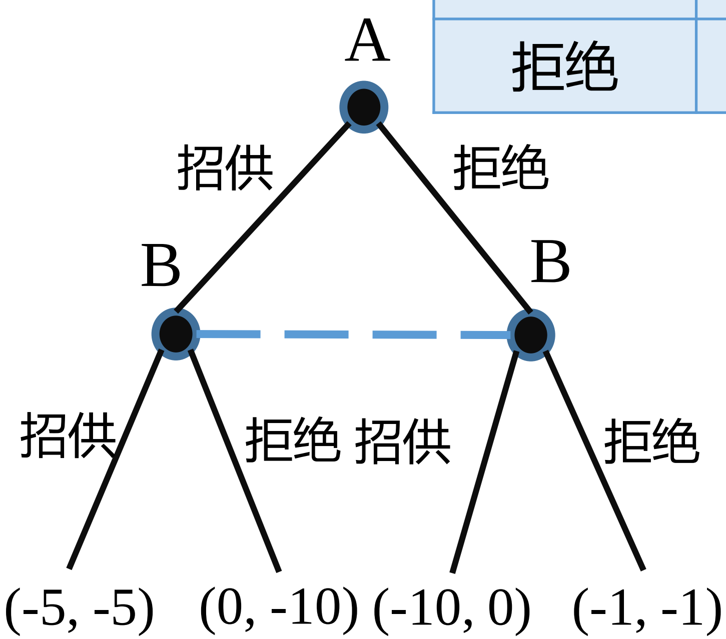


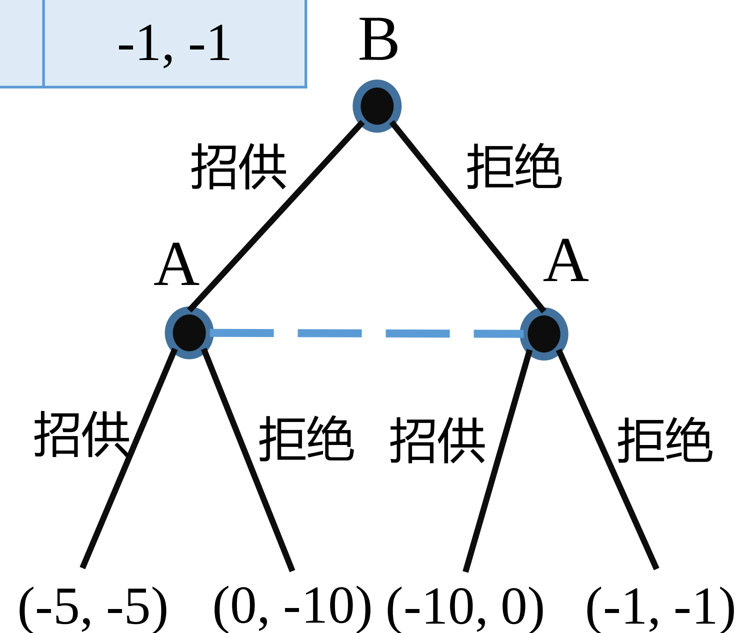
石头剪刀布的扩展式表示

展开式表示用于表示静态博弈

- 借助于信息集的概念，可以用展开式表示来表示静态博弈。以囚徒困境博弈为例：
  - 第二个决策的人不知道第一个人的决策

囚徒A \ 囚徒B	招供	拒绝
招供	-5, -5	-10, 0
拒绝	0, -10	-1, -1

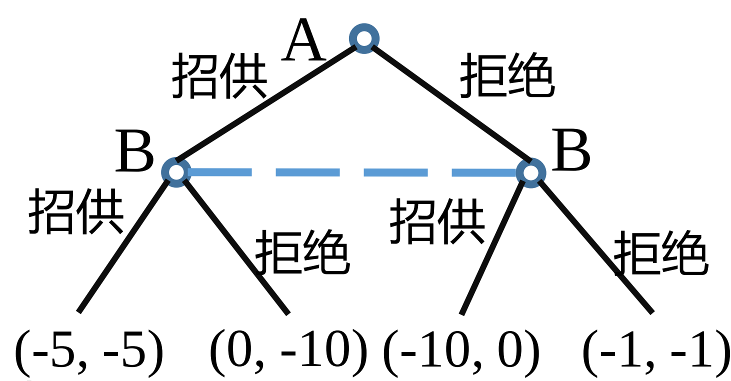




扩展式博弈的纳什均衡

- 分析思路：数学家的思维，尝试先把问题转化为一个已经解决的问题。
  - 将扩展式博弈表示为策略式博弈表示：关键要理解策略之间如何转化和之间的关系
    - 策略式表示的策略： “我选择我的策略，和你选择没关系”
    - 扩展式表示的策略： “如果…发生了，我将选择…的策略”
  - 举例说明：囚徒困境
    - 策略式表示的策略：
      - *A*的策略： “招供” ， “拒绝” ；
      - B的策略： “招供” ， “拒绝” 。
    - 扩展式表示的策略：
      - A的策略： “招供” ， “拒绝” ；
      - B的策略：
        - “无论A是否招供，我都招供” ，
        - “无论A是否招供，我都拒绝” ，
        - “A如果招供我就招供，否则拒绝” ，
        - “A如果招供我就拒绝，否则招供” 。

囚徒B \ 囚徒A	招供	拒绝
招供	-5, -5	-10, 0
拒绝	0, -10	-1, -1



扩展式博弈的纳什均衡

- 通过将扩展式博弈转换为策略式表示，每一个策略组合（也就是博弈树上的一条路径）决定了一个支付向量$u(u_1, \ldots, u_n)$，如果策略组合s^*对于所有的i，能够最大化$u_i(s_i, s_{-i}^*)$（或期望值），即：

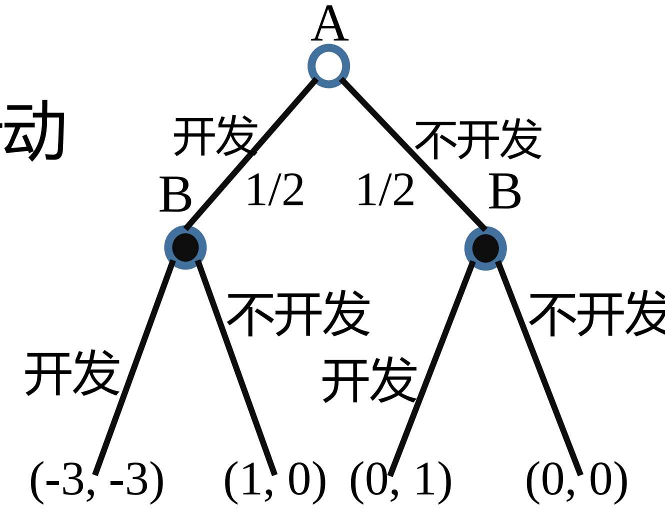
$$s_i^* \in \underset{s_i \in S_i}{\operatorname{argmax}}\, u_i(s_i, s_{-i}^*), \qquad \forall i$$

那么该策略组合就是扩展式博弈的纳什均衡。

- 运用类似上述的构造方式，对于扩展式表示混合策略（行为策略），同样可以构造出扩展式博弈的混合式策略。

扩展式博弈的纳什均衡

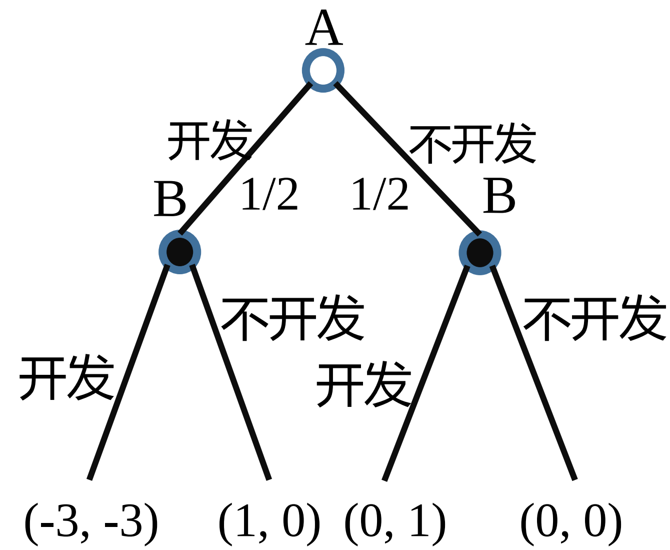
- 卖家市场的房地产开发博弈
  - 低需求，A先做行动，B看到A行动后行动
  - 如右图扩展式表示，完美信息博弈
- 利用信息集的概念进行构造
  - A有一个单节点信息集，两个可选行动
  - B有两个双节点信息集，每个有两个可选行动



- A的纯策略集很简单：（开发、不开发）
- B的纯策略集有四个
  - 不管A开发还是不开发，我都开发
  - A开发我开发，A不开发我不开发
  - A开发我不开发，A不开发我开发
  - 不管A开发还是不开发，我都不开发

扩展式博弈的纳什均衡

- A的2个纯策略：开发，不开发
- 将B的4个纯策略简记为
  - {开发，开发}：总是要开发
  - {开发，不开发}：跟着干
  - {不开发，开发}：唱反调
  - {不开发，不开发}：总是不开发
- 那么A和B之间博弈的标准式如下



开发商B / 开发商A	{开发，开发}	{开发，不开发}	{不开发，开发}	{不开发，不开发}
开发	-3, -3	-3, -3	**1**, **0**	**1**, **0**
不开发	**0**, **1**	0, 0	0, 1	0, 0

扩展式表示博弈的纳什均衡

- 扩展式博弈纳什均衡的存在性：有限完美信息博弈

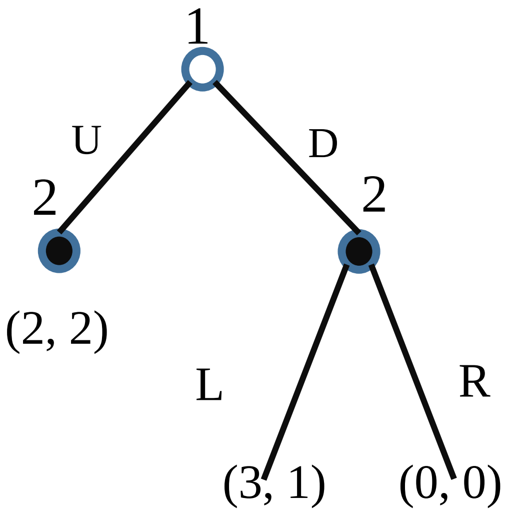
有限完美信息博弈纳什均衡（Zermelo,1913; Kuhn, 1953）

每一个有限完美信息博弈有一个纯策略纳什均衡。

证明：使用动态规划逆向归纳法。
博弈有限：存在一个最后决策结点的集合（倒数第二个结点）
在最后决策结点参与人决策，最大化自己的收益；
给定这个收益，倒数第二个决策结点参与人决策，最大化自己的收益；
如此重复，直到初始结点；
这个倒推过程结束，得到一个路径，给出每个参与人的特定策略，所有这些策略构成了一个纳什均衡。

扩展式表示博弈的纳什均衡

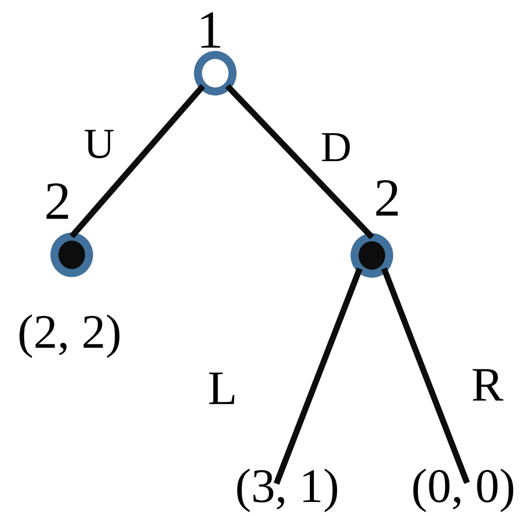
- 扩展式博弈的纳什均衡求取
  - 参与人1先行动，规则如右图如果参与人1选择U,那么无论参与人2选择什么，收益是不变的（2，2）
  - 求取这个博弈的纳什均衡



- 分析过程：
  - 当参与人1选择D时，参与人2有两个选择：如果选择L，参与人2得到1个单位奖励，如果选择R，他能得到0个单位的奖励。
  - 因此L是参与人2的最优选择。
  - 如果参与人2知道参与人1是理性的，那么参与人1将选择D。
  - 因此，(D,L)是一个纯策略纳什均衡

扩展式表示博弈的纳什均衡

- 扩展式博弈的纳什均衡求取
  - 将上一页的博弈转换为策略式表示：
  - 参与人1 有两个动作：U,D
  - 参与人2 有两个动作：L,R



- 均衡求取：
  - 重复剔出劣策略
  - 对参与人2，来说R是一个弱劣策略。
    - 即，无论参与人1选择U还是D, R 都不是参与人2的优策略。
  - 重复剔除之后即可得到和前面分析相同的结果

参与人2 \ 参与人1	L	R
U	2,2	2,2
D	**3,1**	0, 0

- 结论：
  - 逆向归纳法实际上是重复剔除劣策略方法在扩展式博弈中的应用。

扩展式表示博弈的纳什均衡

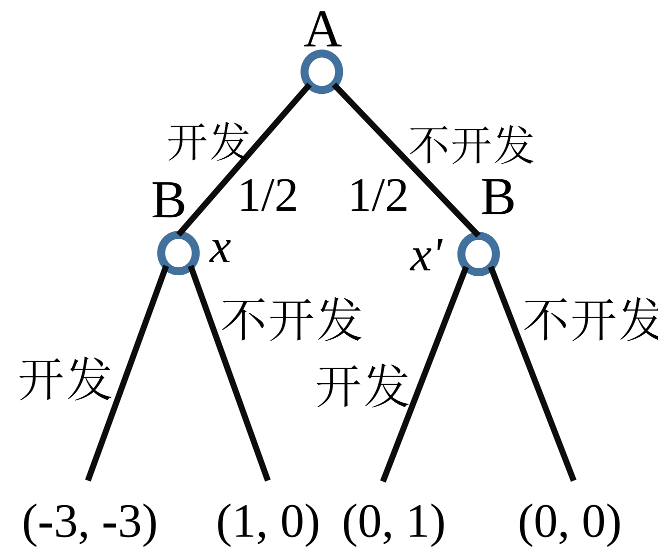
- 逆向归纳法并不适用于所有情况：
  - 无限博弈
    - 一个决策结有无穷多个后续结，如果支付函数不是特殊限制，那么最优选择不一定存在。
    - 或一个路径包含无穷多个决策结，不存在最后一个决策结。
  - 不完美信息博弈
    - 信息集不是单结的
- 只是逆向归纳法不适用，并不代表这两种情况的纳什均衡不存在。

子博弈精炼纳什均衡

- 动态博弈的扩展式表示都可以转换为策略式表示，从而转换为静态博弈。
- 完全信息静态博弈的纳什均衡因此同样适用于动态博弈。
- 但是这样得到的（多个）纳什均衡往往不是合理的均衡。
  - 纳什均衡会有多个均衡
  - 未考虑到参与者行动的先后顺序，使得很多均衡不合理
- 需要寻求能够改进（perfecting）和精炼（refining）纳什均衡概念，从而得到更为合理的博弈解。
- 精炼的含义：
  - 把动态博弈中的“合理纳什均衡”和“不合理纳什均衡”分开

子博弈精炼纳什均衡

- 以右图的房地产开发为例：
  - 首先转化为下图的策略式表示
  - 该策略式表示有三个纳什均衡
    - （不开发，{开发，开发}）
    - （开发，{不开发，不开发}）
    - （开发，{不开发，开发}）



只有（开发，{不开发，开发}）是唯一的合理均衡！

开发商B / 开发商A	{开发，开发}	{开发，不开发}	{不开发，开发}	{不开发，不开发}
开发	-3, -3	-3, -3	**1, 0**	1, 0
不开发	0, 1	0, 0	0, 1	0, 0

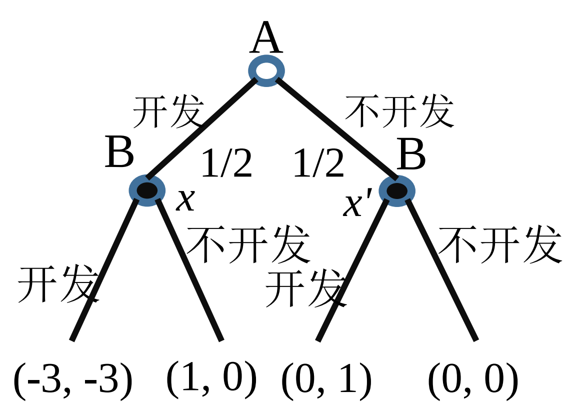
子博弈精炼纳什均衡

- 子博弈：原博弈的一部分，本身可以作为一个独立的博弈进行分析。

子博弈

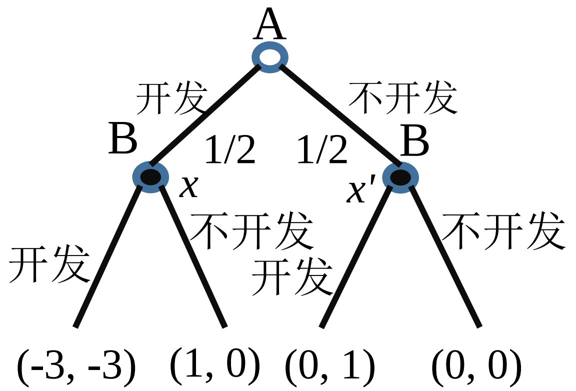
一个扩展式表述博弈的子博弈G由一个决策结x和所有该决策结的后续结$T(x)$（包括终结点）组成，它满足下列条件：1）x是一个单节点信息集，即$h(x)=\{x\}$；2）对于所有的$x' \in T(x)$，如果$x'' \in h(x')$，那么$x'' \in T(x)$。

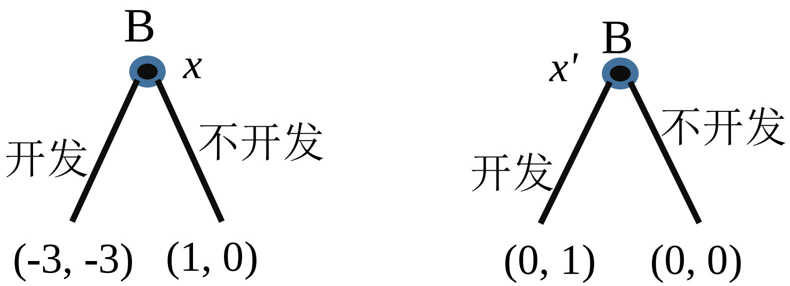
- 任何博弈本身都可以称为自身的一个子博弈

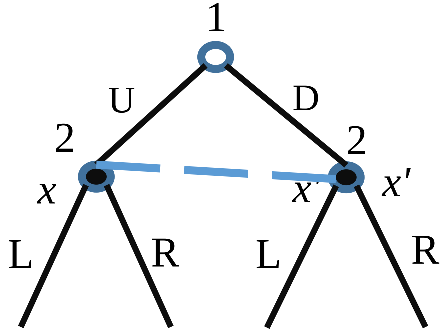


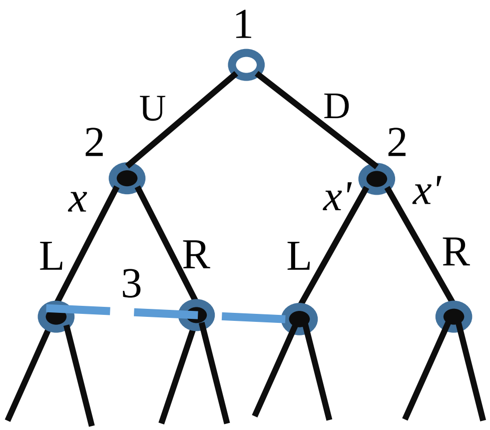
子博弈精炼纳什均衡

- 子博弈的两个条件分析
  - 条件一：一个子博弈必须从一个单结信息集开始
  - 条件二：子博弈信息集和支付向量都直接继承自原博弈









子博弈精炼纳什均衡

- 子博弈精炼纳什均衡：一个策略组合是子博弈精炼纳什均衡，当且仅当它在每一个子博弈（包括原博弈）上都构成一个纳什均衡。

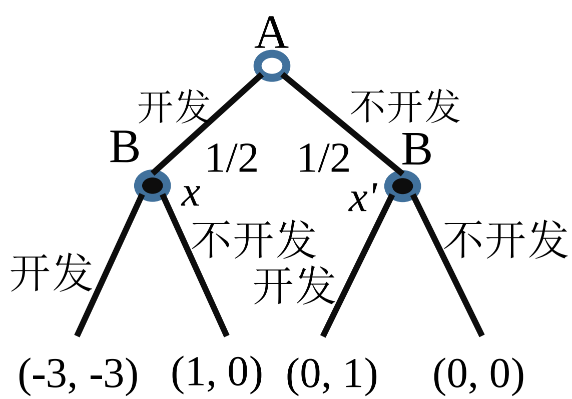
子博弈精炼纳什均衡（Selten, 1965）

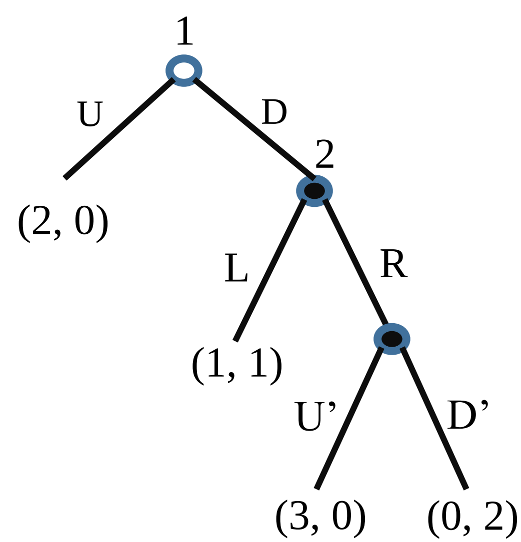
扩展式描述博弈的策略组合$s^* = (s_1^*, \ldots, s_i^*, \ldots, s_n^*)$是一个子博弈精炼纳什均衡，如果：1）它是原博弈的纳什均衡；2）它在每一个子博弈上给出纳什均衡。

- 在每一个子博弈上给出纳什均衡：无论过去发生什么事情，参与人在每一个决策结上都应该最优化自己的决策。

子博弈精炼纳什均衡

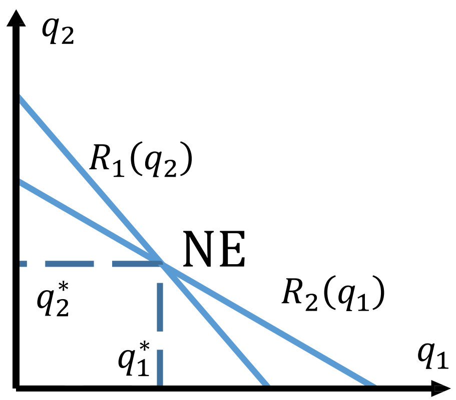
- 使用逆向归纳法求解子博弈精炼纳什均衡
  - 从博弈的最后一个决策结出发，该决策结上行动的参与人做最优决策
  - 倒回到第二个决策结，找出倒数第二个决策者的最优决策
  - 不断重复直到初始决策结
- 举例：
  - 房地产开发博弈
  - 三阶段行动博弈





子博弈精炼纳什均衡举例1

- 斯坦克尔伯格（Stackelberge）寡头竞争模型
  - 企业的行动有先后顺序，领头企业1首先选择产量$q_1 > 0$，尾随企业2根据领头企业的产量选择自己的产量$q_2 > 0$。
  - 博弈类型：完美信息**动态**博弈
  - 博弈树：领头企业作为初始决策结，尾随企业有无穷多个决策结（连续变量）
- 对比库诺特（Cournot）寡头竞争模型
  - 两个企业同时决策
  - 博弈类型：完全信息**静态**博弈
- $\begin{cases} q_1^* = R_1(q_2) = \dfrac{C_1'(q_1)}{P'(q_1+q_2)} - \dfrac{1}{P'(q_1+q_2)} P(q_1+q_2) \\ q_2^* = R_2(q_1) = \dfrac{C_2'(q_1)}{P'(q_1+q_2)} - \dfrac{1}{P'(q_1+q_2)} P(q_1+q_2) \end{cases}$ 反应函数



- 赋值：需求函数：$P = a - q_1 - q_2$ 单位成本：$c > 0$
  - 结果是：$q_1^* = q_2^* = (a-c)/3$

子博弈精炼纳什均衡举例1

- 斯坦克尔伯格（Stackelberge）寡头竞争模型
  - 企业的行动有先后顺序，领头企业1首先选择产量$q_1 \geq 0$，尾随企业2根据领头企业的产量选择自己的产量$q_2 \geq 0$。
  - 赋值：需求函数：$P = a - q_1 - q_2$ 单位成本：c>0
- 求解：
  - 首先，给定q_1，求企业2的最优选择
  - $\max\limits_{q_2 \geq 0} \pi_2(q_1, q_2) = q_2(a - q_1 - q_2) - q_2 c$
  - 最优化一阶条件（对q_2求导）得：$s_2(q_1) = (a - q_1 - c)/2$
  - 企业1知道企业2是理性的(领头知道尾随的最优决策是$s_2(q_1)$)
  - $\max\limits_{q_1 \geq 0} \pi_1(q_1, q_2) = q_1\big(a - q_1 - s_2(q_1)\big) - q_1 c$
  - 同样求一阶条件得：$q_1^* = (\mathrm{a} - \mathrm{c})/2$
  - 进而得：$q_2^* = s_2(q_1^*) = (\mathrm{a} - \mathrm{c})/4$

子博弈精炼纳什均衡举例1

- 对比斯坦克尔伯格均衡结果与库诺特均衡结果

	库诺特	斯坦克尔伯格
博弈类型	完全信息静态博弈	完全信息动态
行动先后	同时行动	领头企业先动
纳什均衡	$(q_1^* = r(q_2), q_2^* = r(q_1))$	$(q_1^*, s_2(q_1))$
均衡结果	$((a-c)/3, (a-c)/3)$	$((a-c)/2, (a-c)/4)$
总产量	$2(a-c)/3$	$3(a-c)/4$
价格（P）	$(a+2c)/3$	$(a+3c)/4$
利润	$((a-c)^2/9, (a-c)^2/9)$	$((\boldsymbol{a}-\boldsymbol{c})^2/\mathbf{8}, (a-c)^2/16)$

子博弈精炼纳什均衡举例1

- 结果分析：
  - 相较于库诺特均衡，企业1在斯坦克尔伯格均衡中的利润增加了，而企业2的利润减少了，我们称之为先手优势。
  - 这种先手优势的背后，企业2拥有更多的信息却使其在博弈中处于劣势。
  - 如果在博弈之前，企业1和企业2都不知道彼此的生产信息，那么企业1的产量威胁就是一种口头宣扬，不可置信。
  - 一旦在博弈之前，企业1知道了企业2会根据他的产量来做决策，那么他会把“口头”威胁，变成实际行动，这种威胁就是可置信的，迫使博弈走入斯坦克尔伯格均衡。

子博弈精炼纳什均衡举例2

- 鲁宾斯坦恩(Rubinstein)-斯塔尔(Stahl)议价模型
  - 参与人1和参与人2共同分享一块蛋糕
  - 第一个时期，参与人1先给出一个分享规则$(x_1, 1-x_1)$
  - 参与人2选择，接受还是拒绝
  - 若接受，则结束；若拒绝，则进入第二个时期
  - 第二个时期，参与人2给出一个分享规则$(x_2, 1-x_2)$
  - 参与人1选择，接受还是拒绝
  - 若接受，则结束；若拒绝，则进入第三个时期
  - ……
  - 直到有人接受对方的规则，则停止
- 分析：
  - 博弈类型：无限期完美博弈，但有可能被中止
  - 收益函数：$\pi_i = \delta_i^{t-1} x_t$, 其中$\delta_i$是参与人$i$ 的贴现因子

子博弈精炼纳什均衡举例2

- 鲁宾斯坦恩(Rubinstein)-斯塔尔(Stahl)议价模型
  - 分析思路：先从有限次T内可以结束博弈开始($T < \infty$)，然后讨论与贴现因子δ和博弈次数T之间的关系。
- 用逆向递归法求解有限次情况下的子博弈精炼纳什均衡
  - $T = 2$，(只进行两次)：
    - 如果在$t = 2$时参与人2提出的规则是$(x_2, 1 - x_2)$, 其中$x_2 = 0$, 即(0,1), 参与人1也会接受，因为参与人1没有机会进行下一次提议了。
    - 而参与人2在$t = 2$时得到的1个单位的收益，贴现到$t = 1$时刻为δ_2。
    - 如果在$t = 1$时刻，参与人1提出的规则$(x_1, 1 - x_1)$中，只要$1 - x_1 \geq \delta_2$，参与人2就会接受。
    - 子博弈精炼均衡为$(1 - \delta_2, \delta_2)$.
  - $T = 3$, 最后做决策的是参与人1
    - 同样分析，得到子博弈精炼均衡为$(1 - \delta_2(1 - \delta_1), \delta_2(1 - \delta_1))$
  - ……
  - 接下来讨论δ和T的各种情况

子博弈精炼纳什均衡举例2

- 鲁宾斯坦恩(Rubinstein)-斯塔尔(Stahl)议价模型
  - 上一页均衡为$\left(1-\delta_2\left(1-\delta_1(1-\delta_2)\right),\quad \delta_2\left(1-\delta_1(1-\delta_2)\right)\right)$
- 如果$\delta_1=\delta_2=0$:
  - 两个人都没有耐心，均衡结果肯定是(1,0)
- 如果$\delta_2=0$:
  - 无论δ_1为多少，结果都是(1,0)。参与人1知道参与人2不会有耐心进行讨价还价，所以参与人给出最有利于自己的报价。
- 如果$\delta_1=0,\delta_2>0$:
  - 均衡结果是$(1-\delta_2,\delta_2)$，因为参与人2将在第二轮得到最大收益，折现到第一轮中就是δ_2，那么参与人1将报价$(1-\delta_2,\delta_2)$。
- 如果$\delta_1=\delta_2=1$,无限耐心:
  - $T=1,3,5,\ldots$，均衡为(1,0)；$T=2,4,6,\ldots$, 均衡为(0,1);
  - 后动优势，谁最后一轮出价，谁得到的最多。

子博弈精炼纳什均衡举例2

- 鲁宾斯坦恩(Rubinstein)-斯塔尔(Stahl)议价模型
  - 上一页均衡为$\left(1-\delta_2\left(1-\delta_1(1-\delta_2)\right),\quad \delta_2\left(1-\delta_1(1-\delta_2)\right)\right)$
- 如果$0<\delta_i<1$：
  - 均衡结果不仅依赖于贴现因子之间的比率，而且依赖于博弈的时间T和谁在最后一个阶段出价。
  - 这种依赖关系会随着时间的增大而缩小。
  - 当T趋于无穷的时候，我们将得到“先动优势”。
  - 当$\delta_1=\delta_2=\delta$时，我们得到唯一的均衡结果$\left(\frac{1}{1+\delta},1-\frac{1}{1+\delta}\right)$

定理(Rubinstein, 1982)

在无限期轮流出价博弈中，唯一的子博弈精炼纳什均衡结果是：

$$x^*=\frac{1-\delta_2}{1-\delta_1\delta_2}$$

子博弈精炼纳什均衡举例2

- 鲁宾斯坦恩(Rubinstein)-斯塔尔(Stahl)议价模型
- 定理的简单说明：
  - 当我们讨论无穷轮博弈的时候，可以从任意一个由参与人1起点的子博弈开始讨论。
  - 时刻t，参与人1能得到的最大收益是M；t时刻的M，等价于$t-1$时刻的$\delta_1 M$; 参与人2知道，在$t-1$时刻的任意$x \geq \delta_1 M$的报价都会被接受。
  - 所以$t-1$时刻，参与人2给出方案$(\delta_1 M, 1-\delta_1 M)$；$t-1$时刻得到的$1-\delta_1 M$等价于在$t-2$得到的$\delta_2(1-\delta_1 M)$
  - 所以$t-2$时刻，参与人1报价$\left(1-\delta_2(1-\delta_1 M), \delta_2(1-\delta_1 M)\right)$
  - 无限博弈中，t时刻看到的与$t-2$时刻看到的应该相同：
  - $M = 1-\delta_2(1-\delta_1 M)$
  - 解得 $M = \frac{1-\delta_2}{1-\delta_1\delta_2}$
  - 用最小份额进行分析，可以得到同样的结果。

什么是重复博弈？

- 当一个博弈被重复多次进行时就构成了一个特殊的动态博弈：重复博弈（Repeated Game）。
- 重复博弈是非常重要的一种动态博弈类型，是人们研究的最为透彻的一种博弈类型，也基于机器学习的智能博弈技术的基础。
- 重复博弈是将相同结构的博弈重复多次，其中的每次博弈称为“阶段博弈（Stage Game）”。
- 举例：1）每个囚徒被放出来继续偷窃，多次进行囚徒困境博弈；2）智能体多次尝试走出迷宫；3）阿尔法狗AI不断通过自我博弈生成数据；4）OpenAI Five不断进行群体对抗；5）机器不但进行图灵测试尝试骗过人类；6）人类在与大自然的不断对抗中生存下来；等等。

什么是重复博弈？

重复博弈特点

1）阶段博弈之间没有“物质上”的联系（no physical links）；

2）所有参与人都会观测到博弈过去的历史；

3）参与人的总收益是所有阶段博弈收益的和或加权和。

重点理解一下第2个特点：

- 重复博弈整体可能是不完美信息博弈，因为阶段博弈有可能是不完美的；
- 由于参与人过去行动历史可以观测，每个参与人的整体策略都是“……如果你干啥，我就干啥，如果你还敢干啥，我就再干啥……”，这样每个人的纯策略空间大小是指数级组合爆炸的；
- 重复博弈的均衡结果和每次博弈的均衡结果可能不一样。

什么是重复博弈？

- 影响重复博弈均衡结果的主要因素

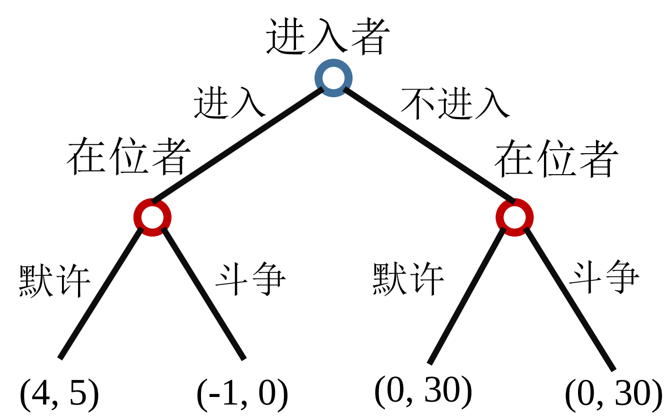
- 重复的次数
  - 一次性博弈
  - 较多次博弈
  - 无限次博弈
- 信息完备性
  - 双方都是完备信息
  - 我方有私有信息
  - 对方有私有信息
- 阶段博弈均衡的个数
  - 只有一个均衡
  - 有多个均衡

有限次重复博弈：连锁店悖论

- 考虑右图的市场进入阻挠博弈
  - 考虑一次博弈：唯一的子博弈精炼纳什均衡是（进入，默许）；
  - 假设在位者有20个市场（连锁店），那么在位者的均衡策略是什么呢？
  - 整个博弈还是存在唯一的子博弈精炼纳什均衡：每次都是（进入，默许）。

进入者 \ 在位者	默许	斗争
进入	4, 5	-1, 0
不进入	0, 30	0, 30

进入者先行动



- 考虑之前的多次囚徒困境博弈呢？
  - 两个囚徒总是都选择（坦白，坦白）
- 上述分析结果被称为连锁店悖论（Chainstore Paradox；Selten，1978）

有限次重复博弈：连锁店悖论

有限次重复博弈的均衡定理（Selten, 1978）

令Γ是阶段博弈，$\Gamma(T)$是重复T次的重复博弈（$T<\infty$）。那么，若Γ有唯一的纳什均衡，重复博弈$\Gamma(T)$的唯一子博弈精炼纳什均衡结果是阶段博弈Γ的纳什均衡重复T次（即每个阶段博弈结果都是单次博弈的均衡）。

证明：使用逆向递归分析法结合有限次重复博弈特点即可得证。

注意，阶段博弈均衡存在的唯一性是定理成立的重要条件。

对于阶段博弈存在多个纳什均衡的有限次重复博弈均衡怎么分析求解呢？

有限次重复博弈：连锁店悖论

- 考虑右图的抽象博弈
  - 总共有三个纳什均衡：(M, L), (U, C)和((3/7U, 4/7M), (3/7L, 4/7C))
  - 没法达到帕累托最优(D, R)

红方 \ 蓝方	L	C	R
U	0, 0	3, 4	6, 0
M	4, 3	0, 0	0, 0
D	0, 6	0, 0	5, 5

- 右图博弈若重复两次，均衡是（令贴现因子为δ）？

求解思路：绘制两阶段博弈的博弈树，根据子博弈精炼纳什均衡定义采用逆向递归分析法求解。

可以得到如下的策略是一个子博弈精炼纳什均衡：“在第一阶段选择(D, R)；如果第一阶段的结果是(D, R)，在第二阶段选择(M, L)，反之，选择混合战略((3/7U, 4/7M), (3/7L, 4/7C))”。

无限次重复博弈：无名氏定理

- 无限重复囚徒困境博弈：令两个囚徒的贴现因子（耐心）均为δ，则每个囚徒的效益为：$\mu_i = \sum_{k=1}^{\infty} u_i^k \delta^{k-1}$
  - 两个囚徒每次均选择坦白是不是一个子博弈精炼纳什均衡？
  - 还有其他子博弈精炼纳什均衡吗？
  - 冷酷战略：开始选择抵赖，一旦一方选择坦白，另一方一直选择坦白。

囚徒A \ 囚徒B	坦白	抵赖
坦白	-5, -5	0, -10
抵赖	-10, 0	-1, -1

计算验证：假设囚徒i先选择了坦白，他此阶段多收益1效益，他的机会注意出发对方冷酷战略，后面每次得到5惩罚：

$$0 + \delta(-5) + \delta^2(-5) + \cdots \le -1 + \delta(-1) + \delta^2(-1) + \cdots$$

$$-\frac{5\delta}{1-\delta} \le -\frac{1}{1-\delta}$$

当$\delta^* \ge \frac{1}{5}$时，给定j坚持冷酷战略并不先坦白，i就不会首先坦白；给定j首先坦白，i也一直会坦白，坚持冷酷战略。

无限次重复博弈：无名氏定理

无限次重复博弈的均衡定理（无名氏定理）（Friedman, 1971）

令Γ是一个n人阶段博弈，$\Gamma(\infty,\delta)$是以为阶段博弈的无限重复博弈，a^*是Γ的一个纳什均衡（纯战略或者混合策略），$e^*=(e_1,e_2,\cdots,e_n)$是a^*决定的效益向量，$v^*=(v_1,v_2,\cdots,v_n)$是任意一个可行的效益向量。V是可行效益向量的集合那么。那么，对于任意满足$v_i>e_i$的$v\in V(\forall i)$，存在一个贴现因子$\delta^*<1$，使得对于所有的$\delta\geq\delta^*$，$v=(v_1,v_2,\cdots,v_n)$是一个特定的子博弈精炼纳什均衡。

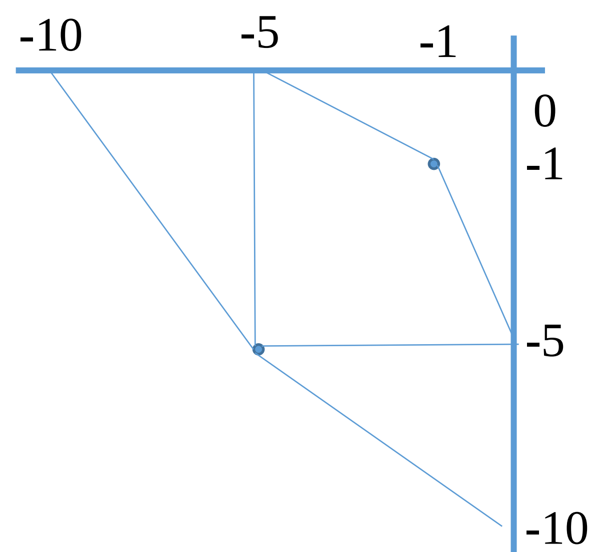
详细证明： 见参考文献Drew Fudenberg and Jean Tirole, Game Theory, MIT Press, 1991. 第5章。

无限次重复博弈：无名氏定理

- 无名氏定理的解释：在无限次重复博弈中，如果参与人有足够的耐心（即δ足够接近于1），那么任何满足个人理性的可行效益向量都可以通过特定的子博弈精炼纳什均衡得到。

囚徒A \ 囚徒B	坦白	抵赖
坦白	**-5, -5**	0, -10
抵赖	-10, 0	**-1, -1**

- 以囚徒困境博弈为例
  - a^*：（坦白，坦白）
  - e^*：（-5，-5）
  - v^*：（-1，-1）
  - V：由点（-5，-5）点及其到横纵坐标垂线交点加上点（-1，-1）构成的多边形区域。



- 其他参与人惩罚一个不合作者的办法是准星阶段博弈的纳什均衡。

参与人不固定的重复博弈

- 例如：消费品市场交易
  - 厂商是固定的参与人
  - 消费者是不固定的，只消费单次
  - 单次博弈：纳什均衡(不购买，低质量)

消费者 \ 厂商	高质量	低质量
购买	1,1	-1,2
不购买	0,0	0, 0

- 无限次重复博弈：
  - 结论：若$\delta \geq \frac{1}{2}$，下列策略组合是一个子博弈精炼纳什均衡：
    - 厂商从生产高质量产品开始，一直生产高质量产品，除非某次生产了低质量的产品，之后永远生产低质量的产品。
    - 消费者选择购买，只要厂商不曾生产过低质量产品。如果厂商生产过低质量产品，之后的其它消费者便不再购买。
  - 贴现因子δ的算法：
    - 厂商生产低质量产品得到的短期利润是2，如果一直生产高质量产品贴现值为$1/(1-\delta)$。那么如果$\frac{1}{1-\delta} > 2$，厂商没有意愿生产低质量产品。
  - 由于消费者是单次消费，所以他们没有贴现因子

本次课程作业

- 作业内容：寻找存在自博弈精炼纳什均衡的完全信息动态博弈习题两道，给出习题描述、问题分析、求解过程和最终结果。
  - 评分准则：选取习题的典型性、分析求解过程的完整性以及最终结果的指导性意义。
- 提交时间：2020年10月22日17:00之前
- 提交方法：在课程网站上提交，同时提交作业Word版到助教邮箱（kangyongxin2015@ia.ac.cn）
- 邮件发送规范
  - 邮件主题：计算博弈第三次作业_学号_姓名
  - 附件名称：计算博弈第三次作业_学号_姓名.docx

感谢聆听！

兴军亮

Junliang.Xing@ia.ac.cn

2020年10月15日

中国科学院大学
University of Chinese of Academy of Sciences

CASIA 自动化研究所
Institute of Automation