博弈论中的“囚徒困境”模型

■王家辉

“囚徒困境”模型是博弈论中的经典范例,它是1950年Tucker提出的,其完全信息下的静态博弈为广大博弈论的工作者和初学者所掌握,成为解释生活现象的有力工具。其实“囚徒困境”模型随着博弈论的深入发展,具有各种不同的形式,通常分为:完全信息的静态博弈,完全信息的动态博弈,不完全信息的静态博弈及不完全信息的动态博弈四种形式。

一、完全信息静态“囚徒困境”博弈

完全信息静态“囚徒困境”博弈部分地奠定了非合作博弈论的理论基础。它的基本模型是:警察抓住了两个合伙犯罪的罪犯,由于缺乏足够的证据指证他们的罪行,所以希望这两人中至少有一人供认犯罪,就能确认罪名成立。为此警察将这两个罪犯分别关押以防止他们串供,并告诉他们警方的政策是“坦白从宽,抗拒从严”:如果两人中只有一人坦白认罪,则坦白者立即释放,而另一人则将重判5年徒刑;如果两个同时坦白认罪,则他们将各判3年监禁。当然罪犯知道如果他们两人都拒不认罪,则警方只能以较轻的妨碍公务罪判处他们1年徒刑。用矩阵表示两个罪犯的得益如下(得益向量的第一个数字是囚徒1的得益,第二个数字是囚徒2的得益):

表1

		囚徒2	
		坦白	不坦白
囚徒1	坦白	(-3,-3)	(0,-5)
	不坦白	(-5,0)	(-1,-1)

假定两个罪犯熟悉彼此,这便是一个同时行动的完全信息静态博弈。容易看出,由于对于每个囚徒而言,无论对方选择什么策略,坦白都是自己的最优策略,所以(坦白,坦白)是博弈的Nash均衡。

二、完全信息动态“囚徒困境”博弈——重复“囚徒困境”博弈

研究重复博弈的意义在于基本博弈会重复进行,比如犯罪团伙会被警方多次审讯,日常生活中买卖会重复进行,国际间的战争此伏彼起。而且人们也发现基本博弈的重复进行并非基本博弈的简单累加,比如商业中的回头客问题。下面继续以表1所示的“囚徒困境”模型为例对多重博弈进行探讨。

首先观察“囚徒困境”的有限博弈,以T记基本博弈的重复次数。博弈重复进行所耗时间会比较长,支付的时间价值必须考虑,记r为折现因子。在有限博弈的情况下,可简化在r=1的情况下讨论,并采用动态博弈的逆向归纳法进行研究。先分析t=T阶段两博弈方的选择,这仍然是一个基本的囚徒困境博弈,此时前一阶段的结果已成为事实,又无后续阶段,因此不难得出结论,这一阶段的结果是(坦白,坦白),双方得益(-3,-3)。现在回到t=T-1阶段,理性的博弈方对于后一阶段的结局非常清楚,其结果必然是(坦白,坦白),因此不管现阶段的博弈结果是什么,双方在本阶段以后的最终得益都是在本阶段得益的基础上各加上-3,此时的得益矩阵是:

表2

		囚徒2	
		坦白	不坦白
囚徒1	坦白	(-6,-6)	(-3,-8)
	不坦白	(-8,-3)	(-4,-4)

容易看出,坦白仍是两博弈方的严格优超策略,即(坦白,坦白)是T-1阶段的唯一的纯Nash均衡。

以此往上类推,每阶段“囚徒困境”博弈的结果都是博弈双方采用坦白,所以T次重复博弈的子博弈精炼Nash均衡是每个博弈阶段双方都采用坦白。

再考虑“囚徒困境”博弈重复无数次。因为无限博弈没有最终阶段,所以不能运用逆向归纳法求解。考虑博弈双方都采用“冷酷战略”:(1)开始阶段选择抵赖;(2)选择抵赖直到有一方选择了坦白,为了报复对手的背叛,以后都选择坦白。

假定囚徒j严格执行上述冷酷战略,考察囚徒i的最优策略是否为冷酷战略?如果i在博弈的某个阶段首先选择了坦白,他在该阶段得到0,而不是-1,但他的这次背叛会遭到囚徒j的永远惩罚,因此i在随后每个阶段的支付都是-3。如果下列条件满足,给定j没有选择坦白,i将不会选择坦白:

$$0+r(-3)+r^2(-3)+\cdots \leq -1+r(-1)+r^2(-1)+\cdots \quad \text{即:} -\frac{3r}{1-r} \leq -\frac{1}{1-r}$$

解上述不等式得:$r^* \geq 1/3$(这个条件容易满足)。就是说,如果$r \geq 1/3$,给定j坚持冷酷战略并且j没有首先坦白,i不会选择首先坦白。

进一步假定j首先选择坦白,那么i是否有积极性坚持冷酷战略以惩罚j的不合作行为?如果i坚持冷酷战略,他随后每个阶段的支付是-3,但如果他选择其他战略,他在任何单一阶段的支付都不会大于-3,因此,无论r是多大,i都有积极性坚持冷酷战略。在博弈重复无数次的情况下,只要$r > 1/3$,子博弈精炼均衡是每个阶段博弈双方都采用抵赖进行合作。

三、不完全信息静态“囚徒困境”博弈

由于现实生活中许多博弈并不满足完全信息的要求,比如买卖双方都对彼此的信息掌握不完全,买者不知卖者产品的质量到底如何,卖者也不知道买者愿意付出多高的价格等等,因此研究不完全信息下的博弈有着重要的理论和现实意义。

假定囚徒1有两种类型,理性的(或称为不合作的)和非理性的(有意愿合作的),概率分别为1-p和p,又假定囚徒2只有一种类型——理性的。假定理性的囚徒可以选择任意的策略,而非理性的囚徒1只有一种策略“针锋相对”,即开

基金项目:国家自然科学基金项目(70272014)。

始阶段选择抵赖，随后的阶段以对方前一阶段的策略为自己现阶段的策略进行鼓励或报复。

由于博弈只进行一个回合，博弈双方没有合作可能，于是理性的囚徒 1 的最优策略是"坦白"。理性的囚徒 2 也会选择也"坦白"，因为对于一次博弈而言，不管囚徒 1 理性与否，坦白的策略总是囚徒 2 最优的，构成不完全信息静态博弈的 Bayes-Nash 均衡。

我们还可以按如下方法证明，由于博弈只进行一个阶段，则非理性的囚徒 1 选择抵赖，理性的囚徒 1 选择坦白，记囚徒 2 的选择为 X，博弈路径如下所示：

表 3

	t=1
非理性囚徒 1 (p)	抵赖
理性囚徒 1 (1-p)	坦白
囚徒 2	X

当 X=抵赖时，囚徒 2 的期望支付是：4p-5；当 X=坦白时，囚徒 2 的期望支付是：3p-3，

无论 p 为何值，3p-3 > 4p-5，故坦白是囚徒 2 的最优选择

四、不完全信息动态"囚徒困境"博弈

理论上在完全信息的情况下，T 次重复的"囚徒困境"博弈在每阶段博弈都选择"坦白"是两个囚徒的最优战略，然而这一结果并没有在现实生活中发生，我们常常看到屡次作案的犯罪团伙总是百般抵赖妄图逃脱法律的惩罚。国外实验经济学家作试验也表明，在有限次重复博弈中合作行为也频繁出现，因此需要将不完全信息引入重复博弈。

首先讨论"囚徒困境"博弈只重复两次的情况。在第二阶段，由于没有合作的空间，理性的囚徒 1 和囚徒 2 都会选择坦白，而非理性的囚徒 1 根据"针锋相对"策略要选择囚徒 2 在第一阶段的策略；在第一阶段，非理性的囚徒 1 选择抵赖，理性的囚徒 1 仍会选择坦白，因为它在该阶段的选择不会改变囚徒 2 在第二阶段选择坦白。现在考虑囚徒 2 在第一阶段的选择(X)如何影响非理性囚徒 1 在第二阶段的选择，如下表所示：

表 4

	t=1	t=2
非理性囚徒 1 (p)	抵赖	X
理性囚徒 1 (1-p)	坦白	坦白
囚徒 2	X	坦白

当 X="抵赖"，则囚徒 2 的期望支付是：p[(-1)+0]+(1-p)[(-5)+(-3)]=7p-8

当 X="坦白"，囚徒 2 此时的期望支付是：p[0+(-3)]+(1-p)[(-3)+(-3)]=3p-6

如果 7p-8≥3p-6，即 p≥1/2，囚徒 2 将会选择 X="抵赖"。

在 p≥1/2 的条件下，进一步考虑基本博弈重复三次的情况。在第三阶段理性的囚徒 1 和囚徒 2 会因为没有后续的合作机会而选择坦白；在第二阶段，由于理性的囚徒 1 知道囚徒 2 是理性的，自己在本阶段的选择不会改变囚徒 2 在下一阶段的选择，故仍会选择坦白。下面要说明理性囚徒 1 在第一阶段将会选择抵赖进行合作。尽管囚徒 1 在第一阶段选择坦白可能免于惩罚，但无疑向囚徒 2 宣示自己是理性的博弈方，于是囚徒 2 在第二阶段选择坦白，理性的囚徒 1 在第二阶段最大只能获得(-3)的支付；相反如果隐藏自己的真实情况，选择抵赖，那么可能在第一阶段获得(-1)的支付，第二阶段获得 0 的支付，无疑这将更为有利，所以理性的囚徒 1 的三阶段策略是(抵赖，坦白，坦白)。

就理性的囚徒 1 和 2 而言，第一阶段有合作的可能(双方都选"抵赖")，也有不合作的可能(理性囚徒 1 选择"抵赖"、囚徒 2 选择坦白)。

先看双方都选择"抵赖"的情形，那么博弈进入第二和第三阶段，即随后的阶段是表 4 所示的两阶段博弈，所以在给定 p≥1/2 的条件下，囚徒 2 在第二阶段选择抵赖。三次重复博弈的精练 Bayes 均衡如下表所示：

表 5

	t=1	t=2	t=3
非理性囚徒 1(p)	抵赖	抵赖	抵赖
理性囚徒 1 (1-p)	抵赖	坦白	坦白
囚徒 2	抵赖	抵赖	坦白

囚徒 2 选择(抵赖，抵赖，坦白)的期望支付为：

(-1)+p[(-1)+0]+(1-p)[(-5)+(-3)]=7p-9

再看双方不合作的情况。在不合作的情形下，囚徒 2 的策略有两种可能：(坦白，坦白，坦白)和(坦白，抵赖，坦白)。如果囚徒 2 选择(坦白，坦白，坦白)，博弈路径如下所示：

表 6

	t=1	t=2	t=3
非理性囚徒 1 (p)	抵赖	坦白	坦白
理性囚徒 1 (1-p)	抵赖	坦白	坦白
囚徒 2	坦白	坦白	坦白

囚徒 2 的期望支付是：(0)+(-3)+(-3)=-6。

如果囚徒 2 选择(坦白，抵赖，坦白)，博弈路径如下表所示：

表 7

	t=1	t=2	t=3
非理性囚徒 1 (p)	抵赖	坦白	抵赖
理性囚徒 1 (1-p)	抵赖	坦白	坦白
囚徒 2	坦白	抵赖	坦白

期望支付是：(0)+(-5)+p(0)+(1-p)(-3)=3p-8。

p≥1/2 在的条件下 7p-9≥-6，7p-9≥3p-8，因此(抵赖，抵赖，坦白)优于(坦白，坦白，坦白)和(坦白，抵赖，坦白)。

综合以上分析，只要囚徒 1 是非理性的概率 p≥1/2，表 5 所示的战略就是一个精炼 Bayes 均衡。类似可以进一步证明，如果 p≥1/2，对于 T > 3，下列战略组合构成一个精炼 Bayes 均衡：理性囚徒 1 在 t=1 至 t=T-2 阶段一直选择抵赖，在余下的两阶段选择坦白；囚徒 2 在 t=1 至 t=T-1 阶段选择抵赖，在最后阶段选择坦白。

我们清楚地看到，将不完全信息引入有限次"囚徒困境"重复博弈能很好地解释现实的社会现象——为什么有那么多的囚徒宁愿选择抵赖而不是选择优超策略坦白。

至于"囚徒困境"的不完全信息下的无数重复博弈的情况，我们应该容易得出：在相当宽松的条件下，每阶段选择合作是精炼 Bayes 均衡

(作者单位/厦门大学计统系)

(责任编辑/李友平)