中国科学 : 数学　2016 年　第 46 卷　第 10 期 : 1445 ∼ 1460
SCIENTIA SINICA　Mathematica

《中国科学》杂志社
SCIENCE CHINA PRESS

平均场随机对策: 单调成本函数与门限策略

献给陈翰馥教授 80 华诞

黄民懿[①*], 马琰[②]

① School of Mathematics and Statistics, Carleton University, Ottawa K1S 5B6, Canada;
② 郑州大学数学与统计学院, 郑州 450001
E-mail: mhuang@math.carleton.ca, mayan203@zzu.edu.cn

收稿日期: 2016-03-07; 接受日期: 2016-03-21; 网络出版日期: 2016-09-30; * 通信作者
加拿大自然科学与工程研究委员会基金 (批准号: RGPIN 2014-03827) 和郑州大学科研启动基金 (批准号: 129-51090091) 资助项目

摘要　本文考虑多参与者 Markov 决策过程框架下的平均场对策问题. 每个参与者具有连续状态与二值控制. 通过主动控制参与者状态达到一个重置点. 所有参与者由它们的成本函数耦合. 若所考虑的平均场对策有解, 个体策略的结构可通过门限策略刻画. 本文进一步引进对策的稳态方程, 并在正外部性条件下分析其解的唯一性.

关键词　平均场对策　Markov 决策过程　动态规划　门限策略　稳态分布
MSC (2010) 主题分类　60J05, 90C40, 91A10, 91A15

1　引言

平均场对策论研究涉及大量非合作参与者的随机决策问题, 在系统中每一个体的作用非常微弱, 但它们作为一个整体能产生重大的影响. 这个理论为降低策略分析和设计的复杂性提供了一个强有力的方法. 借助考虑一个无穷多参与者的辅助模型, 分析者可以使用具有一致性的平均场逼近, 为现实中的众多但有限的参与者构造出分布式策略, 并进一步证明其 ϵ-Nash 均衡性质[1–3]. 文献 [4] 独立提出了非常类似的方法. 另一相关的解的概念称作 oblivious 均衡, 适用于 Markov 决策模型[5]. 当模型由非线性扩散过程描述时 (参见文献 [3,4,6]), 平均场对策的分析所依赖的工具包括 Hamilton-Jacobi-Bellman (HJB) 方程、Fokker-Planck 方程和 McKean-Vlasov 方程. 在随机分析框架下, 更多文献可参见文献 [7,8]. 文献 [9–12] 研究了所谓的混合参与者模型, 其中的平均场作用涉及一个有重大影响的参与者. 对于平均场对策论的概述, 读者可参见文献 [9,13,14].

平均场对策在许多领域有着广泛应用, 包括电力系统[15]、大群体电动汽车充电控制[16,17]、经济和金融[18–20]、随机增长理论[21]、有生物背景的振荡子对策[22].

本文研究一类多参与者 Markov 决策过程 (MDP) 框架下的平均场对策问题. MDP 形式的动态对策这一经典领域由 Shapley 开创, 并将其称为随机对策[23,24]. 基于 MDP 的平均场对策, 参见文献 [5,18,25]. 本文中的参与者具有连续状态空间和二值行动空间, 并通过成本函数耦合. 状态用于衡

引用格式: Huang M Y, Ma Y. Mean field stochastic games: Monotone costs and threshold policies (in Chinese). Sci Sin Math, 2016, 46: 1445–1460, doi: 10.1360/N012016-00057

　

量风险水平, 如无主动控制, 其数值会随机增长. 参与者的单步成本依赖于自身状态、控制和群体平均状态. 自然而然, 参与者的成本是其状态的增函数. 这一建模框架受到实际应用问题的启发, 后者包括网络安全投资对策和流感疫苗接种对策[26–29]. 当成本函数也是群体平均状态的增函数时, 它反映了正外部性.

本文考虑有限时间区间上的对策, 门限策略是求解这一平均场对策模型时得到的有趣结果. 我们随后研究了相关不动点问题有解的条件. 在进一步的分析中, 我们处理对策的稳态方程并在正外部性条件下分析解的唯一性.

虽然平均场对策为降低策略设计的复杂性提供了一个强有力的体系, 但是, 除线性二次 (LQ) 情形计算较为简单外 (参见文献 [1,30–32]), 一般非线性系统中的策略往往只能被间接确定, 缺少简单形式, 其数值解计算仍需很大的工作量. 本文目标之一是, 发展一个合适的建模框架使得解具有相对简单的结构.

本文结构如下: 第 2 节引入有关的 Markov 决策过程框架; 第 3 节给出门限策略形式的最优响应; 第 4 节建立一个 ϵ-Nash 均衡结果; 第 5 节分析平均场方程组的解的存在性; 第 6 节引入稳态方程并分析解的唯一性; 第 7 节总结全文.

2 Markov 决策模型

2.1 系统动态

所考虑系统包括 N 个参与者, 表示为 $\mathcal{A}_i$, $1 \leqslant i \leqslant N$. 在时刻 $t \in \mathbb{Z}_+ = \{0,1,2,\ldots\}$, 记 $\mathcal{A}_i$ 的状态为 x_t^i, 行动为 a_t^i.

为简便起见, 我们考虑一群对称的参与者. 每位参与者具有状态空间 $S = [0,1]$. S 中的值可解释为风险程度. 所有参与者具有同一个行动空间 $A = \{a_0, a_1\}$, 其中 a_0 表示不行动, a_1 表示采取主动措施. 系统中每位参与者的状态演化可描述为一个 Markov 决策过程, 并且只受自身行动的影响. 以 $\mathcal{B}(I)$ 表示区间 I 上的 Borel σ- 代数.

当 $t \geqslant 0$ 且 $x \in S$, 参与者的状态具有如下转移核:

$$P(x_{t+1}^i \in B \mid x_t^i = x, a_t^i = a_0) = Q_0(B \mid x), \tag{2.1}$$

$$P(x_{t+1}^i = 0 \mid x_t^i = x, a_t^i = a_1) = 1, \tag{2.2}$$

其中 $Q_0(B \mid x)$ 是一个随机核, $B \in \mathcal{B}(S)$, 且 $Q_0([0,x) \mid x) = 0$. Q_0 的结构表明, 对给定的 $x_t^i = x$, $a_t^i = a_0$, 状态转移到 $[x,1]$ 中. 也就是说, 如果没有主动控制, 那么参与者的状态将会变差. 以下称 $a_t^i = a_1$ 为重置行动, $0 \in S$ 为重置点.

向量过程 $(x_t^1, \ldots, x_t^N)$ 构成了一个高维受控 Markov 过程, 其转移核有如下乘积测度形式:

$$P(x_t^i \in B_i, i = 1,\ldots,N \mid x_t^i = x^{[i]}, a_t^i = a^{[i]}, i = 1,\ldots,N) = \prod_{i=1}^{N} P(x_t^i \in B_i \mid x_t^i = x^{[i]}, a_t^i = a^{[i]}),$$

其中 $B_i \in \mathcal{B}(S)$, $x^{[i]} \in S$, $a^{[i]} \in A$. 此乘积测度表明这 N 个受控 Markov 过程具有独立转移.

2.2 个体成本

定义群体平均状态 $x_t^{(N)} = \frac{1}{N}\sum_{i=1}^{N} x_t^i$. $\mathcal{A}_i$ 的单步成本为

$$c(x_t^i, x_t^{(N)}, a_t^i) = R(x_t^i, x_t^{(N)}) + \gamma 1_{\{a_t^i = a_1\}},$$

其中 $\gamma > 0$, $\gamma 1_{\{a_t^i=a_1\}}$ 是措施成本. 函数 $R \geqslant 0$ 定义在 $S \times S$ 上, 用以度量风险相关的成本. 对 $0 < T < \infty$ 和折扣因子 $\rho \in (0,1)$, 定义以下成本函数:

$$J_i = E\sum_{t=0}^{T} \rho^t c(x_t^i, x_t^{(N)}, a_t^i), \quad 1 \leqslant i \leqslant N. \tag{2.3}$$

我们引入以下假定:

(A1) $\{x_0^i, i \geqslant 1\}$ 是独立同分布的随机变量, 取值于 S 且 $Ex_0^i = m_0$.

(A2) $R(x,z)$ 是 $S \times S$ 上的连续函数. 对给定的 z, $R(\cdot, z)$ 是严格增函数.

(A3) 存在取值于 S 的随机变量 ξ, 使得 $Q_0(\cdot \mid x)$ 等于随机变量 $x+(1-x)\xi$ 的分布. 另外, $P(\xi = 1) < 1$.

ξ 的分布函数记为 F_ξ. 为非平凡起见, (A3) 中假定 $P(\xi = 1) < 1$.

我们给出一些引进 (A3) 的原因. 对 $S = [0,1]$, 可称 $1-x$ 为距离最大状态 1 的状态裕量. 记 $m_t^i = 1 - x_t^i$, 若给定 m_t^i 及 $a_t^i = 0$, 则状态裕量衰减至 $m_{t+1}^i = m_t^i(1-\xi_t^i)$, 其中 ξ_t^i 与 ξ 同分布. 换言之, 若无主动控制, 则状态裕量会指数衰减.

例 1 实验室中有 N 台互相连接的计算机 M_i $(1 \leqslant i \leqslant N)$ 分别分配给了 N 个主用户 U_i $(1 \leqslant i \leqslant N)$, 每台机器可能会偶然由其他成员使用, 以方便提供特有资源. 机器 M_i 具有不适性状态 $x_t^i \in [0,1]$, 日常的使用或接触潜在的恶意软件都会随机地加重不适性. 用户 U_i 可在 M_i 上采取维护行动 a_1 (如安装或更新安全软件, 扫描和清理存储设备, 释放存储空间等) 来使机器恢复到理想状态 $x_t^i = 0$. U_i 的单步成本为 $R(x_t^i, x_t^{(N)}) + \gamma 1_{\{a_t^i=a_1\}}$, 其中对 $x_t^{(N)}$ 的依赖是由于机器共用和来自其他机器的可能威胁. 我们将以上模型称为实验室伙伴对策.

3 平均场极限模型

3.1 最优控制问题

本节假定 (A1)–(A3) 成立. 记序列 $(b_s, \ldots, b_t)$ $(s \leqslant t)$ 为 $b_{s,t}$. 设 x_t^i 由 (2.1) 和 (2.2) 给出, 并以确定值 z_t 逼近 (2.3) 中的 $x_t^{(N)}$. 定义

$$\bar{J}_i(z_{0,T}, a_{0,T}^i) = E\sum_{t=0}^{T} \rho^t c(x_t^i, z_t, a_t^i).$$

若 $a_t^i(x)$ 是由 S 到 A 的映射, 则称 a_t^i 为纯 Markov 策略. 我们称 a_t^i 是带有参数 $r \in [0,1]$ 的门限策略, 如果当 $x \geqslant r$ 时, $a_t^i(x) = 1$; 当 $x < r$ 时, $a_t^i(x) = 0$. 以上形式给出反馈策略. 以下分析将确定最优策略的形式.

3.2 动态规划方程

记 $a_{s,t}^i = (a_s^i, \ldots, a_t^i)$, $s \leqslant t$. 取定序列 $z_{0,T}$, 其中 $z_t \in [0,1]$. 对 $0 \leqslant s \leqslant T$ 和 $x \in S$, 定义

$$\bar{J}_i(s, x, z_{0,T}, a_{s,T}^i) = E\left[\sum_{t=s}^{T} \rho^{t-s} c(x_t^i, z_t, a_t^i) \;\middle|\; x_s^i = x\right].$$

值函数定义为 $V(t,x)=\inf_{a^i_{t,T}}\bar{J}_i(t,x,z_{0,T},a^i_{t,T})$, 其中 $a^i_{0,T}$ 选自于 Markov 策略集. 动态规划方程取以下形式:

$$\begin{cases} V(t,x)=\min\limits_{a^i_t}[c(x,z_t,a^i_t)+\rho E[V(t+1,x^i_{t+1})\mid x^i_t=x]], \\ V(T,x)=R(x,z_T), \quad 0\leqslant t<T. \end{cases} \tag{3.1}$$

(3.1) 等价于

$$\begin{cases} V(t,x)=\min\left[\rho\displaystyle\int_0^1 V(t+1,y)Q_0(dy\mid x)+R(x,z_t),\rho V(t+1,0)+R(x,z_t)+\gamma\right], \\ V(T,x)=R(x,z_T), \quad 0\leqslant t<T. \end{cases} \tag{3.2}$$

记

$$G_t(x)=\int_0^1 V(t,y)Q_0(dy\mid x), \quad 0\leqslant t\leqslant T. \tag{3.3}$$

引理 1 对每个 $0\leqslant t\leqslant T$, $V(t,x)$ 在 S 上连续.

证明 我们用归纳法证明. $V(T,x)$ 是 $x\in S$ 的连续函数. 假定对 $0<k\leqslant T$, $V(k,x)$ 关于 x 连续. 由条件 (A3),

$$G_k(x)=\int_0^1 V(k,x+(1-x)y)dF_\xi(y)=\int_0^1 V(k,(1-y)x+y)dF_\xi(y), \tag{3.4}$$

结合以上归纳假设即可得到 $\rho G_k(x)$ 是关于 x 连续的.

注意到 $R(x,z_{k-1})$ 关于 x 连续. 另一方面, 若 $g_1(x)$ 和 $g_2(x)$ 在 $[0,1]$ 上连续, 则 $\min\{g_1(x),g_2(x)\}$ 关于 x 连续. 由 (3.2) 可知, $V(k-1,x)$ 关于 x 连续.

通过归纳得出, 对所有的 $0\leqslant t\leqslant T$, $V(t,x)$ 关于 x 连续. □

引理 2 对每个 $0\leqslant t\leqslant T$, $V(t,x)$ 在 S 上是严格增的.

证明 对 $t=T$, 当 $x_1<x_2$ 时, $V(T,x_1)<V(t,x_2)$. 假定对 $0<k\leqslant T$,

$$V(k,x_1)<V(k,x_2), \quad \text{其中 } x_1<x_2. \tag{3.5}$$

对 $0\leqslant x_1<x_2\leqslant 1$,

$$R(x_1,z_{k-1})<R(x_2,z_{k-1}).$$

由 (3.4) 和 (3.5), 有

$$\rho G_k(x_1)+R(x_1,z_{k-1})<\rho G_k(x_2)+R(x_2,z_{k-1}).$$

对 $\alpha_1<\alpha_2$ 和 $\beta_1<\beta_2$, 有 $\min\{\alpha_1,\beta_1\}<\min\{\alpha_2,\beta_2\}$. 取

$$\alpha_i=\rho G_k(x_i)+R(x_i,z_{k-1}), \quad \beta_i=\rho V(k,1)+R(x_i,z_{k-1})+\gamma,$$

可得 $V(k-1,x_1)<V(k-1,x_2)$. 通过归纳得出, 对所有 $0\leqslant t\leqslant T$, $V(t,x)$ 是严格增的. □

引理 3 $G_t(x)$ 关于 x 连续且是严格增的.

证明 本引理由引理 1、2、(3.4) 和 (A3) 中 $P(\xi=1)<1$ 得出. □

引理 4 对 $t \leqslant T-1$, 若

$$\rho G_{t+1}(0) < \rho V(t+1,0) + \gamma < \rho G_{t+1}(1), \tag{3.6}$$

则存在唯一 $x^* \in (0,1)$ 使得 $\rho G_{t+1}(x^*) = \rho V(t+1,0) + \gamma$.

证明 由引理 3 和中值定理即可得证. □

定理 1 当 $t = T$ 时, 定义 $a_T^i = 0$. 当 $t \leqslant T-1$ 时, 定义策略 $a_t^i(x)$ 如下:

(1) 若 $\rho G_{t+1}(1) \leqslant \rho V(t+1,0) + \gamma$, 则对所有 $x \in S$, 取 $a_t^i(x) = 0$.

(2) 若 $\rho G_{t+1}(0) \geqslant \rho V(t+1,0) + \gamma$, 则对所有 $x \in S$, 取 $a_t^i(x) = 1$.

(3) 若 (3.6) 成立, 则取 a_t^i 为门限策略, 其参数 x^* 由引理 4 给出.

则 $a_{0,T}^i$ 是一个最优策略.

证明 容易看出 $a_T^i = 0$ 是最优的. 考虑 $t \leqslant T-1$, 由引理 3 和 4, 可验证当 a_t^i 按 (1)–(3) 选取时, (3.2) 中最小值可达到. □

4 平均场对策的解

本节假定 (A1)–(A3) 成立. 为得到平均场对策的解, 我们引进以下方程组:

$$\begin{cases} V(t,x) = \min\left[\rho \displaystyle\int_0^1 V(t+1,y) Q_0(dy \mid x) + R(x,z_t), \rho V(t+1,0) + R(x,z_t) + \gamma\right], \quad 0 \leqslant t < T, \\ V(T,x) = R(x,z_T), \\ z_t = E x_t^i, \quad 0 \leqslant t \leqslant T. \end{cases} \tag{4.1}$$

由 (A1), $z_0 = m_0$. 对 (4.1), 我们求解 $(\hat{z}_{0,T}, \hat{a}_{0T}^i)$ 使得 $\{x_t^i, 0 \leqslant t \leqslant T\}$ 由 $\{\hat{a}_t^i(x), 0 \leqslant t \leqslant T\}$ 产生, 其中后者满足取 $z_{0,T} = \hat{z}_{0,T}$ 后定理 1 的规则. 最后一个方程是平均场对策中标准的一致性条件.

考虑由 (2.1)–(2.3) 决定的 N 个参与者的对策. 记

$$a_{0,T}^{-i} = (a_{0,T}^1, \ldots, a_{0,T}^{i-1}, a_{0,T}^{i+1}, \ldots, a_{0,T}^N), \quad J_i = J_i(a_{0,T}^i, a_{0,T}^{-i}).$$

为进行性能估计, 我们考虑 a_t^i 在策略空间 $\mathcal{U}_t$ 中的摄动, 其中 $\mathcal{U}_t$ 包括所有依赖 $(x_t^1, \ldots, x_t^N)$ 的 Markov 策略.

定义 1 一组含 N 个参与者的策略 $\{a_{0,T}^i, 1 \leqslant i \leqslant N\}$, 被称为相对于成本 $\{J_i, 1 \leqslant i \leqslant N\}$ 的一个 ϵ-Nash 均衡, 其中 $\epsilon \geqslant 0$, 如果对任意 $1 \leqslant i \leqslant N$ 和任何 $b_{0,T}^i \in \prod_{t=0}^T \mathcal{U}_t$, 有

$$J_i(a_{0,T}^i, a_{0,T}^{-i}) \leqslant J_i(b_{0,T}^i, a_{0,T}^{-i}) + \epsilon.$$

定理 2 假定 (4.1) 具有解 $(\hat{z}_{0,T}, \hat{a}_{0,T}^i)$, 则 $(\hat{a}_{0,T}^1, \ldots, \hat{a}_{0,T}^N)$ 是一个 ϵ-Nash 均衡, 即

$$J_i(\hat{a}_{0,T}^i, \hat{a}_{0,T}^{-i}) - \epsilon \leqslant \inf_{a_{0,T}^i} J_i(a_{0,T}^i, \hat{a}_{0,T}^{-i}) \leqslant J_i(\hat{a}_{0,T}^i, \hat{a}_{0,T}^{-i}),$$

其中 $a_{0,T}^i \in \prod_{t=0}^T \mathcal{U}_t$, 且当 $N \to \infty$ 时 $\epsilon \to 0$.

证明 对 $(a_{0,T}^i, \hat{a}_{0,T}^{-i})$, 记相应的状态为 x_t^i 和 $\hat{x}_t^j$, $j \neq i$, 则

$$\lim_{N\to\infty} \max_{0 \leqslant t \leqslant T} |z_t^{(N)} - \hat{z}_t| = 0, \quad \text{a.s.} \tag{4.2}$$

其中 $z_t^{(N)} = \frac{1}{N}(\sum_{j\neq i}\hat{x}_t^j + x_t^i)$. 记

$$\epsilon_{1,N} = \sup_{a_{0,T}^i} |J_i(a_{0,T}^i, \hat{a}_{0,T}^{-i}) - \bar{J}_i(\hat{z}_{0,T}, a_{0,T}^i)|.$$

由 (4.2), 可得 $\lim_{N\to\infty}\epsilon_{1,N} = 0$. 进一步有

$$\begin{aligned} J_i(a_{0,T}^i, \hat{a}_{0,T}^{-i}) &= \bar{J}_i(\hat{z}_{0,T}, a_{0,T}^i) + J_i(a_{0,T}^i, \hat{a}_{0,T}^{-i}) - \bar{J}_i(\hat{z}_{0,T}, a_{0,T}^i) \\ &\geqslant \bar{J}_i(\hat{z}_{0,T}, a_{0,T}^i) - \epsilon_{1,N} \geqslant \bar{J}_i(\hat{z}_{0,T}, \hat{a}_{0,T}^i) - \epsilon_{1,N}. \end{aligned}$$

另一方面, 记 $\epsilon_{2,N} = |J_i(\hat{a}_{0,T}^i, \hat{a}_{0,T}^{-i}) - \bar{J}_i(\hat{z}_{0,T}, \hat{a}_{0,T}^i)|$, 则 $\lim_{N\to\infty}\epsilon_{2,N} = 0$. 于是,

$$J_i(a_{0,T}^i, \hat{a}_{0,T}^{-i}) \geqslant J_i(\hat{a}_{0,T}^i, \hat{a}_{0,T}^{-i}) - (\epsilon_{1,N} + \epsilon_{2,N}).$$

取 $\epsilon = \epsilon_{1,N} + \epsilon_{2,N}$, 则定理得证. □

5 存在性结果

记 $\mathcal{Z}_T^{m_0} = \{z_{0,T} \mid z_0 = m_0, z_t \in [0,1] \text{ 对 } 1 \leqslant t \leqslant T\}$, 引进以下假设:

(H1) ξ 具有概率密度函数 f_ξ.

(H2) 考虑具有成本函数 $\bar{J}_i(z_{0,T}, a_{0,T}^i) = E\sum_{t=0}^T \rho^t c(x_t^i, z_t, a_t^i)$ 的最优控制问题. 对任意 $z_{0,T} \in \mathcal{Z}_T^{m_0}$, 存在 $c > 0$ 使得最优策略对所有 $x \in [0,c]$ 和 $0 \leqslant t \leqslant T$ 满足 $a_t^i(x) = 0$.

我们称 (H2) 为这一族最优控制问题的一致正门限条件. 它意味着当一个参与者的状态取小值时, 措施成本超过主动控制进一步降低风险所带来的额外收益. 以上对 $z_{0,T}$ 一致成立.

定义 S 上的概率测度集 $\mathcal{P}_0$ 如下: $\nu \in \mathcal{P}_0$ 当且仅当若存在常数 $c_\nu \geqslant 0$ 和 $[0,1]$ 上的可测函数 $g(x) \geqslant 0$ 使得

$$\nu(B) = \int_B g(x)dx + c_\nu 1_B(0),$$

其中 $B \in \mathcal{B}(S)$, 1_B 是 B 的示性函数. 当限制在 $(0,1]$ 上时, ν 相对于 Lebesgue 测度 μ^{Leb} 是绝对连续的.

本节假设 (A1)–(A3)、(H1) 和 (H2) 成立, 且 x_0^i 的分布为 $\mu_0 \in \mathcal{P}_0$.

对给定的 $z_{0,T} \in \mathcal{Z}_T^{m_0}$, 最优控制下的状态 x_t^i 具有分布 μ_t. 定义 $w_t = \int_0^1 x\mu_t(dx)$ 和由 $[0,1]^T$ 到 $[0,1]^T$ 的映射 Φ:

$$(w_1, \ldots, w_T) = \Phi(z_1, \ldots, z_T).$$

引理 5 Φ 是连续的.

证明 取定 $z_{0,T} \in \mathcal{Z}_T^{m_0}$, 记最优策略为 $a_{0,T}^i$, 状态过程为 x_t^i. 取 $z'_{0,T} \in \mathcal{Z}_T^{m_0}$, 记相应的最优策略为 $b_{0,T}^i$, 状态过程为 y_t^i, x_t^i 和 y_t^i 的分布分别为 μ_t 和 μ'_t, 此处 $\mu_0 = \mu'_0$. 由引理 7 和 8, μ_t 和 μ'_t 属于 $\mathcal{P}_0$. 以上使得当门限参数经历小摄动时, μ_t 的摄动也很小. 由引理 9 和 10, 我们首先得出

$$\lim_{z'_{0,T}\to z_{0,T}} \sup_{B\in\mathcal{B}(S)} |\mu_1(B) - \mu'_1(B)| = 0.$$

重复估计, 进一步得

$$\lim_{z'_{0,T}\to z_{0,T}} \sup_{B\in\mathcal{B}(S)} |\mu_t(B) - \mu'_t(B)| = 0, \quad 0 \leqslant t \leqslant T.$$

于是,

$$\lim_{z'_{0,T}\to z_{0,T}}\int_0^1 x\mu'_t(dx)=\int_0^1 x\mu_t(dx),\quad 0\leqslant t\leqslant T.$$

连续性得证. □

定理 3 (4.1) 存在一个解 $(\hat{a}^i_{0,T},\hat{z}_{0,T})$.

证明 由引理 5 和 Brouwer 不动点定理即可得证. □

6 稳态方程

6.1 稳态形式

假定 (A1)–(A3) 成立. 本节引进 (4.1) 的稳态形式. 取 $z\in S$. 值函数不依赖时间, 故可记为 $V(x)$. 动态规划方程变为

$$V(x)=\min_{a^i}[c(x,z,a^i)+\rho E[V(x^i_{t+1})\mid x^i_t=x]],$$

以上给出

$$V(x)=\min\left[\rho\int_0^1 V(y)Q_0(dy\mid x)+R(x,z),\rho V(0)+R(x,z)+\gamma\right].\tag{6.1}$$

我们引入另一个方程

$$z=\int_0^1 x\pi(dx),\tag{6.2}$$

其中 π 是概率测度. 我们称 $(\hat{z},\hat{a}^i,\hat{\pi})$ 为 (6.1) 和 (6.2) 的稳态解, 如果 (i) 反馈策略 $\hat{a}^i$ 是相对于 (6.1) 中 $\hat{z}$ 的最优响应; (ii) 在策略 $\hat{a}^i$ 下, $\{x^i_t,t\geqslant 0\}$ 具有稳态分布 $\hat{\pi}$; (iii) $(\hat{z},\hat{\pi})$ 满足 (6.2).

方程组 (6.1) 和 (6.2) 有如下解释. 在有限时间区间问题中, 让 T 趋于 ∞. 如果这一族解 (标记上不同的 T 取值) 能趋于稳态, 对非常大的 t, 可预计 $V(t,x)$ 和 z_t 会几乎不随时间变化. 以上启发我们引进 (6.1) 和 (6.2) 作为 (4.1) 的稳态形式.

6.2 带一般 z 的值函数

考虑一般 $z\in S$, 未必同时满足 (6.1) 和 (6.2). 记 $G(x)=\int_0^1 V(y)Q_0(dy\mid x)$.

引理 6 (1) 方程 (6.1) 具有唯一解 $V\in C([0,1],\mathbb{R})$.

(2) V 是严格增的.

(3) 最优策略可确定如下:

(i) 若 $\rho G(1)\leqslant\rho V(0)+\gamma$, $a^i(x)\equiv 0$;

(ii) 若 $\rho G(0)\geqslant\rho V(0)+\gamma$, $a^i(x)\equiv 1$;

(iii) 若 $\rho G(0)<\rho V(0)+\gamma<\rho G(1)$, 则存在唯一 $x^*\in(0,1)$ 且 a^i 是带参数 x^* 的门限策略.

证明 (1) 可由不动点方法证得. 为了证明 (2), 定义动态规划算子

$$(\mathcal{L}g)(x)=\min\left[\rho\int_0^1 g(y)Q_0(dy\mid x)+R(x,z),\rho g(0)+R(x,z)+\gamma\right].$$

当 $k\geqslant 0$ 和 $g_0=0$, 定义 $g_{k+1}=\mathcal{L}g_k$. 由归纳法可证, g_k 在 $[0,1]$ 上是增函数. 由于 $\|V-g_k\|\to 0$, V 递增. 由 (6.1), 可得 V 是严格增的. 故 (2) 得证. 易证 $G(x)$ 是严格增的, 进一步可得到 (3). □

对给定的 z, 引理 6 给出了最优策略的结构. 若找到的最优策略是 $a^i(x) \equiv 0$, 我们将其记为带有参数 $\theta(z) = 1^+$ 的门限策略. 否则, 它是带有参数 $\theta(z) \in [0,1]$ 的一般门限策略.

6.3 给定门限策略下的稳态分布

假定 a^i 是具有参数 $\theta \in (0,1)$ 的门限策略. 记 $\{x_t^{i,\theta}, t \geqslant 0\}$ 为相应的状态过程, 这是一个 Markov 过程. 给定 $x_0^{i,\theta} = x \in S$, 令 $\mathcal{B}(S)$ 上的概率测度 $P^t(x,\cdot)$ 为 $x_t^{i,\theta}$ 的分布.

我们引入关于 ξ 的进一步条件.

(A4) ξ 有概率密度函数 $f_\xi(x)$, 且在 S 上, 有 $f_\xi(x) > 0$ a.e.

定理 4 对 $\theta \in (0,1)$, $\{x_t^{i,\theta}, t \geqslant 0\}$ 是一致遍历的且具有稳态概率分布 π_θ, 即对常数 $K > 0$ 和 $r \in (0,1)$, 有

$$\sup_{x \in S} \|P^t(x,\cdot) - \pi_\theta\| \leqslant K r^t, \tag{6.3}$$

其中 $\|\cdot\|$ 是符号测度的全变差范数.

定理 4 的证明见附录 B.

6.4 比较定理

记 $z(\theta) = \int_0^1 x \pi_\theta(dx)$. 我们有以下关于单调性的第一个比较定理.

定理 5 $z(\theta_1) \leqslant z(\theta_2)$, 其中 $0 < \theta_1 < \theta_2 < 1$.

定理 5 的证明见附录 D.

在进一步分析中, 我们考虑 R 具备乘积形式 $R(x,z) = R_1(x)R_2(z)$, 其中 R 仍满足 (A2), 且 $R_1 \geqslant 0$, $R_2 > 0$. 进一步假定:

(A5) $R_2 > 0$ 在 S 上严格增.

以上假定反映了正外部性, 因为每一个个体会从群体平均状态减少当中受益. 这一条件在唯一性分析中有重要作用.

给定了 R 的乘积形式后, (6.1) 取以下形式:

$$V(x) = \min\left[\rho \int_0^1 V(y) Q_0(dy \mid x) + R_1(x)R_2(z), \rho V(0) + R_1(x)R_2(z) + \gamma\right].$$

考虑 $0 \leqslant z_2 < z_1 \leqslant 1$ 和

$$V_l(x) = \min\left[\rho \int_0^1 V_l(y) Q_0(dy \mid x) + R_1(x)R_2(z_l), \rho V_l(0) + R_1(x)R_2(z_l) + \gamma\right]. \tag{6.4}$$

将 (6.4) 的最优策略表示为具备参数 θ_l 的门限策略, 其中 $\theta_l \in [0,1]$ 或取为 1^+. 以下给出关于不同平均场参数下, 门限参数的第二个比较定理.

定理 6 (6.4) 中的 θ_1 和 θ_2 由以下方式确定:

(1) 若 $\theta_1 = 0$, 则 $\theta_2 \in [0,1]$ 或 $\theta_2 = 1^+$.

(2) 若 $\theta_1 \in (0,1)$, 则 (i) $\theta_2 \in (\theta_1, 1)$, 或 (ii) $\theta_2 = 1$, 或 (iii) $\theta_2 = 1^+$.

(3) 若 $\theta_1 = 1$, 则 $\theta_2 = 1^+$.

(4) 若 $\theta_1 = 1^+$, 则 $\theta_2 = 1^+$.

证明 因为 $R_2(z_1) > R_2(z_2) > 0$, 可在 (6.4) 两边除以 $R_2(z_l)$, 并记 $\gamma_l = \frac{\gamma}{R_2(z_l)}$, 则 $0 < \gamma_1 < \gamma_2$. 动态规划方程归结为 (C.1). 从而, 最优策略由引理 14 确定. □

6.5 唯一性

所求解 (z, a^i, π) 将限制于解集 $\mathcal{C}$, 其中 $z \in S$, a^i 是带参数 $\theta \in [0,1]$ 或 $\theta = 1^+$ 的门限策略.

定理 7 假设 (A1)–(A5) 成立, 其中 $R(x,z) = R_1(x)R_2(z)$, 则方程组 (6.1) 和 (6.2) 在 $\mathcal{C}$ 中至多有一个解.

证明 假定存在两个不同的解

$$(z_1, a^i, \pi) \neq (z_2, b^i, \nu). \tag{6.5}$$

若 $z_1 = z_2$, 则 (6.1) 保证了 $a^i = b^i$, 从而 $\pi = \nu$. 以上与两个不同解矛盾. 现在假定

$$0 \leqslant z_2 < z_1 \leqslant 1. \tag{6.6}$$

我们考察定理 6 中所列各情节. 若 $\theta_1 \in (0,1)$, $\theta_2 \in (\theta_1, 1)$, 则定理 5 表明 $z_1 \leqslant z_2$, 与 (6.6) 矛盾. 对其他情形, 易得 $z_1 \leqslant z_2$, 同样与 (6.6) 矛盾. 因此, 所假设的 (6.5) 不成立. 唯一性得证. □

7 结论

本文在多参与者 Markov 决策过程框架下考虑平均场对策. 每个参与者具有单调成本函数, 可对其状态过程采取重置控制. 本文得到分布式门限策略, 进一步研究了此平均场对策的稳态方程及正外部性条件下解的唯一性.

参考文献

1 Huang M, Caines P E, Malhamé R P. Individual and mass behaviour in large population stochastic wireless power control problems: Centralized and Nash equilibrium solutions. In: Proceedings of the 42nd IEEE Conference on Decision and Control. Maui: IEEE, 2003, 98–103

2 Huang M, Caines P E, Malhamé R P. Large-population cost-coupled LQG problems with nonuniform agents: Individual-mass behavior and decentralized ε-Nash equilibria. IEEE Trans Automat Control, 2007, 52: 1560–1571

3 Huang M, Malhamé R P, Caines P E. Large population stochastic dynamic games: Closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle. Commun Inf Syst, 2006, 6: 221–251

4 Lasry J M, Lions P L. Mean field games. Jpn J Math, 2007, 2: 229–260

5 Weintraub G Y, Benkard C L, Van Roy B. Markov perfect industry dynamics with many firms. Econometrica, 2008, 76: 1375–1411

6 Cardaliaguet P. Notes on mean field games. Http://www.science.unitn.it/˜bagagiol/NotesByCardaliaguet.pdf, 2012

7 Carmona R, Delarue F. Probabilistic analysis of mean-field games. SIAM J Control Optim, 2013, 51: 2705–2734

8 Kolokoltsov V N, Li J, Yang W. Mean field games and nonlinear Markov processes. ArXiv:1112.3744, 2011

9 Bensoussan A, Frehse J, Yam P. Mean Field Games and Mean Field Type Control Theory. New York: Springer, 2013

10 Huang M. Large-population LQG games involving a major player: The Nash certainty equivalence principle. SIAM J Control Optim, 2010, 48: 3318–3353

11 Nguyen S L, Huang M. Linear-quadratic-Gaussian mixed games with continuum-parametrized minor players. SIAM J Control Optim, 2012, 50: 2907–2937

12 Nourian M, Caines P E. ϵ-Nash mean field game theory for nonlinear stochastic dynamical systems with major and minor agents. SIAM J Control Optim, 2013, 51: 3302–3331

13 Caines P E. Mean field games. In: Encyclopedia of Systems and Control. Berlin: Springer-Verlag, 2014, 706–712

14 Gomes D A, Saude J. Mean field games models—a brief survey. Dynam Games Appl, 2014, 4: 110–154

15 Kizilkale A C, Malhamé R P. Mean field based control of power system dispersed energy storage devices for peak load relief. In: Proceedings of the 52nd IEEE Conference on Decision and Control. Florence: IEEE, 2013, 4971–4976

16 Ma Z, Callaway D, Hiskens I. Decentralized charging control for large populations of plug-in electric vehicles. IEEE Trans Control Systems Technol, 2013, 21: 67–78
17 Parise F, Colombino M, Grammatico S, et al. Mean field constrained charging policy for large populations of plug-in electric vehicles. In: Proceedings of the 53rd IEEE Conference on Decision and Control. Los Angeles: IEEE, 2014, 5101–5106
18 Adlakha S, Johari R, Weintraub G Y. Equilibria of dynamic games with many players: Existence, approximation, and market structure. J Econom Theory, 2015, 156: 269–316
19 Chan P, Sircar R. Bertrand and Cournot mean field games. Appl Math Optim, 2015, 71: 533–569
20 Lucas Jr R E, Moll B. Knowledge growth and the allocation of time. J Political Econ, 2014, 122: 1–51
21 Huang M. A mean field capital accumulation game with HARA utility. Dynam Games Appl, 2013, 3: 446–472
22 Yin H, Mehta P G, Meyn S P, et al. Synchronization of coupled oscillators is a game. IEEE Trans Automat Control, 2012, 57: 920–935
23 Filar J A, Vrieze K. Competitive Markov Decision Processes. New York: Springer, 1997
24 Shapley L S. Stochastic games. Proc Natl Acad Sci USA, 1953, 39: 1095–1100
25 Huang M, Malhamé R P, Caines P E. On a class of large-scale cost-coupled Markov games with applications to decentralized power control. In: Proceedings of the 43rd IEEE Conference on Decision and Control. Atlantis, Paradise Island: IEEE, 2004, 2830–2835
26 Bauch C T, Earn D J D. Vaccination and the theory of games. Proc Natl Acad Sci of USA, 2004, 101: 13391–13394
27 Jiang L, Anantharam V, Walrand J. How bad are selfish investments in network security? IEEE/ACM Trans Networking, 2011, 19: 549–560
28 Lelarge M, Bolot J. A local mean field analysis of security investments in networks. In: Proceedings of the 3rd International Workshop on Economics of Networked Systems. Seattle: ACM, 2008, 25–30
29 Manfredia P, Postab P D, et al. Optimal vaccination choice, vaccination games, and rational exemption: An appraisal. Vaccine, 2010, 28: 98–109
30 Li T, Zhang J F. Asymptotically optimal decentralized control for large population stochastic multiagent systems. IEEE Trans Automat Control, 2008, 53: 1643–1660
31 Tembine H, Zhu Q, Basar T. Risk-sensitive mean-field stochastic differential games. IFAC Proc Volumes, 2011, 44: 3222–3227
32 Wang B C, Zhang J F. Mean field games for large-population multiagent systems with Markov jump parameters. SIAM J Control Optim, 2012, 50: 2308–2334
33 Meyn S, Tweedie R L. Markov Chains and Stochastic Stability, 2nd ed. Cambridge: Cambridge University Press, 2009

附录 A　用于第 5 节的一些引理

令 X 为具有分布 $\nu\in\mathcal{P}_0$ 的随机变量. 取 $x_t^i=X$. 通过取 $a_t^i\equiv 0$, 定义 $Y_0=x_{t+1}^i$. 通过取带参数 $r\in(0,1)$ 的门限策略 a_t^i, 进一步定义 $Y_1=x_{t+1}^i$, 则

$$P(Y_0\in B)=\int_0^1 Q_0(B\mid x)\nu(dx),\quad B\in\mathcal{B}(S). \tag{A.1}$$

引理 7　Y_0 的分布属于 $\mathcal{P}_0$.

证明　我们可直接得到 Y_0 的概率密度函数如下:

$$g(y)=\int_{0<x<y}\frac{1}{1-x}f_\xi\left(\frac{y-x}{1-x}\right)\nu(dx),\quad y\in(0,1).$$

此情形下, $P(Y_0=0)=0$. □

引理 8　Y_1 的分布属于 $\mathcal{P}_0$.

证明　显然 $P(Y_1=0)=P(X\geqslant r)$. 限制在 $(0,1]$ 上, Y_1 的分布相对于 μ^{Leb} 绝对连续. 记

$$g(y)=\int_{0<x<y\wedge r}\frac{1}{1-x}f_\xi\left(\frac{y-x}{1-x}\right)\nu(dx),$$

则对 $B\in\mathcal{B}(S)$,

$$P(Y_1\in B)=\int_B g(y)dy+P(X\geqslant r)1_B(0).$$

引理得证. □

令 $z_{0,T}\in\mathcal{Z}_T^{m_0}$ 取固定值, 记相应的最优策略为 $a_{0,T}^i$. 取另一序列 $z'_{0,T}\in\mathcal{Z}_T^{m_0}$, 记相应的最优策略为 $b_{0,T}^i$. 定义 $d(z'_{0,T},z_{0,T})=\sum_{k=1}^T|z'_k-z_k|$. 记 $z'_{0,T}\to z_{0,T}$, 如果 $d(z'_{0,T},z_{0,T})\to 0$. 取定 $t\leqslant T-1$. 基于 (H2), 我们考虑两种情形.

情形 A $a_t^i(x)=0$ 对所有 $x\in S$.

引理 9 对于情形 A, 当 $d(z'_{0,T},z_{0,T})$ 充分小时, 我们有

(1) 对所有 $x\in S$ 有 $b_t^i(x)=0$, 或

(2) 存在 $r'\in(0,1)$ 使得 b_t^i 是带参数 r' 的门限策略. 进一步, 若 r' 存在, 当 $z'_{0,T}\to z_{0,T}$ 时, $r'\to 1$.

证明 当 $z'_{0,T}\in\mathcal{Z}_T^{m_0}$ 用于 (H2) 中最优控制问题时, 记值函数为 $V'(t,x)$. 定义 $G'_t(x)$ 以代替 $G_t(x)$, 则 $G'_t(x)$ 是连续且严格增的. 由于 V 连续依赖于 $z_{0,T}$, 当 $z'_{0,T}\to z_{0,T}$ 时, $\sup_{x\in S}|V'(t,x)-V(t,x)|\to 0$. 我们只须考虑以下两种情形:

情形 1 $\rho G_{t+1}(1)<\rho V(t+1,0)+\gamma$.

对所有使得 $d(z'_{0,T},z_{0,T})$ 充分小的 $z'_{0,T}$, 我们同样有

$$\rho G'_{t+1}(1)<\rho V'(t+1,0)+\gamma, \tag{A.2}$$

则 $b_t^i(x)=0$ 对所有 $x\in S$. 因此 (1) 成立.

情形 2 $\rho G_{t+1}(1)=\rho V(t+1,0)+\gamma$.

取定任意 $0<\epsilon<1$, 则

$$\rho G_{t+1}(1-\epsilon)<\rho V(t+1,0)+\gamma. \tag{A.3}$$

取小值 $\delta>0$ 和任意 $z'_{0,T}\in\mathcal{Z}_T^{m_0}$ 使得 $d(z'_{0,T},z_{0,T})\leqslant\delta$. 对此 $z'_{0,T}$, 如果 $\rho G'_{t+1}(1)\leqslant\rho V'(t+1,0)+\gamma$, 取 $b_t^i(x)=0$, 对所有 $x\in S$. 如果对充分小的 δ, $z'_{0,T}$ 导致

$$\rho G'_{t+1}(1)>\rho V'(t+1,0)+\gamma, \tag{A.4}$$

则可找到 $r'\in(0,1)$ 使得

$$\rho G'_{t+1}(r')=\rho V'(t+1,0)+\gamma,$$

以上进一步确定 b_t^i 为具有参数 r' 的门限策略. 我们论证以上 r' 存在. 对充分小的 δ, 由 (A.3),

$$\rho G'_{t+1}(1-\epsilon)<\rho V'(t+1,0)+\gamma.$$

于是, 只要 (A.4) 成立, 对所有小值 δ, 都能找到 $r'\in(1-\epsilon,1)$.

由于 $\epsilon\in(0,1)$ 可以任取, 可知当 $z'_{0,T}\to z_{0,T}$ 时 $r'\to 1$. □

情形 B 存在 $r\in(0,1)$ 使得 a_t^i 是带参数 r 的门限策略.

引理 10 对于情形 B, 当 $d(z'_{0,T},z_{0,T})$ 充分小时, b_t^i 是带参数 $r'\in(0,1)$ 的门限策略, 且当 $z'_{0,T}\to z_{0,T}$ 时, $r'\to r$.

证明 我们有 $\rho G_{t+1}(r) = \rho V(t+1,0) + \gamma$. 固定小值 $\epsilon > 0$, 则

$$\rho G_{t+1}(r-\epsilon) < \rho V(t+1,0) + \gamma, \quad \rho G_{t+1}(r+\epsilon) > \rho V(t+1,0) + \gamma.$$

对充分小的 $\delta > 0$ 和使得 $d(z'_{0,T}, z_{0,T}) \leqslant \delta$ 的 $z'_{0,T}$, 我们有

$$\rho G'_{t+1}(r-\epsilon) < \rho V'(t+1,0) + \gamma, \quad \rho G'_{t+1}(r+\epsilon) > \rho V'(t+1,0) + \gamma.$$

可找到唯一的 $r' \in (r-\epsilon, r+\epsilon)$ 使得

$$\rho G'_{t+1}(r') = \rho V'(t+1,0) + \gamma,$$

其中 r' 依赖于 $z'_{0,T}$. 相应地, 这可得到带参数 r' 的门限策略 b^i_t. 由于 ϵ 可任意小, 引理得证. □

附录 B 定理 4 的证明

考虑 $0 < \theta < 1$. 不可约、非周期和小集的定义可参见文献 [33]. 令 δ_x 为在 $x \in \mathbb{R}$ 的 dirac 测度. 记 $\varphi := \delta_0$, 于是对 $B \in \mathcal{B}(S)$, $\delta_0(B) = 1_B(0)$.

在本附录中, 为简化符号, 记 $x_t := x^{i,\theta}_t$.

引理 11 $\{x_t, t \geqslant 0\}$ 是 φ- 不可约的.

证明 可直接验证

$$P(x_2 = 0 \mid x_0 = x) > 0, \quad x \in [0, \theta),$$
$$P(x_1 = 0 \mid x_0 = x) = 1, \quad x \in [\theta, 1].$$

通过设 x_0 的分布为 dirac 测度 δ_x 可计算得到以上概率. 以上表明 $\{x_t, t \geqslant 0\}$ 是 φ- 不可约的. □

引理 12 $\{x_t, t \geqslant 0\}$ 是非周期的.

证明 定义 $C_s = \{0\}$. 记 $\epsilon_0 = \int_\theta^1 f_\xi(y)dy > 0$ 和测度 $\nu = \epsilon_0 \delta_0$, 则

$$\begin{aligned} P(x_2 = 0 \mid x_0 = 0) &\geqslant P(x_2 = 0, x_1 \geqslant \theta \mid x_0 = 0) \\ &= P(x_1 \geqslant \theta \mid x_0 = 0) \\ &= \epsilon_0. \end{aligned}$$

对任意 $B \in \mathcal{B}(S)$, 可得

$$P(x_2 \in B \mid x_0 = 0) \geqslant \nu(B). \tag{B.1}$$

于是, 可取 C_s 为小集, 且 $\nu(C_s) = \epsilon_0$. 对 $x_0 = 0 \in C_s$, 进一步考察

$$\begin{aligned} P(x_3 = 0 \mid x_0 = 0) &\geqslant P(x_3 = 0, x_2 \geqslant \theta, x_1 < \theta \mid x_0 = 0) \\ &= P(x_2 \geqslant \theta, x_1 < \theta \mid x_0 = 0). \end{aligned}$$

令 ξ_1、ξ_2 和 ξ 为独立同分布的随机变量, 则

$$P(x_2 \geqslant \theta, x_1 < \theta \mid x_0 = 0) = P(\xi_1 + (1-\xi_1)\xi_2 \geqslant \theta, \xi_1 < \theta)$$

$$\geqslant P(\xi_2 \geqslant \theta, \xi_1 < \theta).$$

于是,

$$P(x_3 = 0 \mid x_0 = 0) \geqslant \int_0^\theta f_\xi(y)dy \int_\theta^1 f_\xi(y)dy. \tag{B.2}$$

记 $\epsilon_1 = \int_0^\theta f_\xi(y)dy$, 则对任意 $B \in \mathcal{B}(S)$,

$$P(x_3 \in B \mid x_0 = 0) \geqslant \epsilon_1 \nu(B). \tag{B.3}$$

由于 (B.1) 和 (B.3) 中的时间指标 2 和 3 具有最大公约数 1, 因此, $\{x_t, t \geqslant 0\}$ 是非周期的 (参见文献 [33, 第 112–114 页]). □

Markov 过程 $\{x_t, t \geqslant 0\}$ 满足 Doeblin 条件, 如果存在 $\mathcal{B}(S)$ 上的概率测度 ϕ 和 $\epsilon < 1$, $\eta > 0$, $m \geqslant 0$, 使得 $\phi(B) > \epsilon$ 蕴含

$$\inf_{x \in S} P(x_m \in B \mid x_0 = x) \geqslant \eta.$$

引理 13 Doeblin 条件对 $\{x_t, t \geqslant 0\}$ 成立.

证明 取 $\phi = \delta_0$, 这种情形下, $\phi(B) > 0$ 蕴含 $0 \in B$. 只需证明

$$\inf_{x \in S} P(x_4 = 0 \mid x_0 = x) \geqslant \eta.$$

对 $x \in [\theta, 1]$,

$$\begin{aligned} P(x_4 = 0 \mid x_0 = x) &\geqslant P(x_4 = 0, x_3 \geqslant \theta, x_2 < \theta, x_1 = 0 \mid x_0 = x) \\ &= P(x_3 = 0, x_2 \geqslant \theta, x_1 < \theta \mid x_0 = 0) \\ &= P(x_2 \geqslant \theta, x_1 < \theta \mid x_0 = 0) \\ &\geqslant \epsilon_0 \epsilon_1. \end{aligned} \tag{B.4}$$

令 ξ 和 ξ_k $(k = 1, 2, 3)$ 为独立同分布的随机变量. 对 $x \in [0, \theta)$,

$$\begin{aligned} P(x_4 = 0 \mid x_0 = x) &\geqslant P(x_4 = 0, x_3 \geqslant \theta, x_2 = 0, x_1 \geqslant \theta \mid x_0 = x) \\ &= P(x_3 \geqslant \theta, x_2 = 0, x_1 \geqslant \theta \mid x_0 = x) \\ &= P(\xi_3 \geqslant \theta) P(x_2 = 0, x_1 \geqslant \theta \mid x_0 = x) \\ &= P(\xi_3 \geqslant \theta) P(x + (1 - x)\xi_1 \geqslant \theta) \\ &\geqslant P(\xi_3 \geqslant \theta) P(\xi_1 \geqslant \theta) = \epsilon_0^2. \end{aligned} \tag{B.5}$$

由 (B.4) 和 (B.5) 可知, 取 $\epsilon = \frac{1}{2}$, $m = 4$, $\eta = \epsilon_0^2 \epsilon_1 > 0$, Doeblin 条件成立. □

定理 4 的证明 因为 $\{x_t, t \geqslant 0\}$ 是非周期的且满足 Doeblin 条件, 由文献 [33, 定理 16.0.2] 知, (6.3) 成立. □

附录 C 一个最优控制引理

考虑 $0 < \gamma_1 < \gamma_2$ 和动态规划方程

$$v_l(x) = \min \left\{ \rho \int_0^1 v_l(y) Q_0(dy \mid x) + R_1(x), \rho v_l(0) + R_1(x) + \gamma_l \right\}, \quad l = 1, 2, \quad x \in S. \tag{C.1}$$

记最优策略为 $a_l(x)$. 若 $\rho\int_0^1 v_l(y)Q_0(dy\mid 1)<\rho v_l(0)+\gamma_l$, 则 $a_l(x)\equiv 0$, 我们形式上记其为带参数 $\theta_l=1^+$ 的门限策略. 否则, $a_l(x)$ 是带参数 $\theta_l\in[0,1]$ 的门限策略, 即若 $x\geqslant\theta_l$, 则 $a_l(x)=1$; 若 $x<\theta_l$, 则 $a_l(x)=0$.

引理 14 (1) 若 $\theta_1=0$, 则 $\theta_2\in[0,1]$ 或者 $\theta_2=1^+$.

(2) 若 $\theta_1\in(0,1)$, 则 (i) $\theta_2\in(\theta_1,1)$, 或者 (ii) $\theta_2=1$, 或者 (iii) $\theta_2=1^+$.

(3) 若 $\theta_1=1$, 则 $\theta_2=1^+$.

(4) 若 $\theta_1=1^+$, 则 $\theta_2=1^+$.

证明 每一种情形都可通过 (C.1) 来证明. 这里我们只证明 (2). 注意到 v_l 在 $[0,1]$ 上连续且严格增. 记 $g_l(x)=\int_0^1 v_l(y)Q_0(dy\mid x)$, 则 g_l 严格增. 通过连续逼近方法 (类似引理 6 中的证明), 可得到

$$v_1\leqslant v_2\leqslant v_1+\delta_{21},$$

其中 $\delta_{21}=\gamma_2-\gamma_1$. 以上得出

$$g_1\leqslant g_2\leqslant g_1+\delta_{21}.$$

我们有 $\rho g_1(\theta_1)=\rho g_1(0)+\gamma_1$, 因 $0<\rho<1$, 可得 $\rho g_2(\theta_1)<\rho g_2(0)+\gamma_2$.

如果 $\rho g_2(1)>\rho g_2(0)+\gamma_2$, 则存在唯一的 $\theta_2\in(\theta_1,1)$ 使得 $\rho g_2(\theta_2)=\rho v_2(0)+\gamma_2$, 从而得到情形 (i) 中的门限参数. 情形 (ii) 和 (iii) 可类似考察. □

附录 D 定理 5 的证明

我们将结合折扣消除法和动态规划方程给出证明. 对 Markov 过程 $\{x_t^{i,\theta},t\geqslant 0\}$, 定义值函数

$$v_\alpha^\theta(x)=E\Big[\sum_{t=0}^{\infty}\alpha^t x_t^{i,\theta}\;\Big|\;x_0^{i,\theta}=x\Big],$$

其中 $x\in S$, $\alpha\in(0,1)$, 则 v_α^θ 满足动态规划方程

$$v_\alpha^\theta(x)=x+\alpha E[v_\alpha^\theta(x_1^{i,\theta})\mid x_0^{i,\theta}=x].$$

上式给出

$$v_\alpha^\theta(x)=\begin{cases}x+\alpha\displaystyle\int_0^1 v_\alpha^\theta(x+(1-x)y)dF_\xi(y), & x<\theta,\\ x+\alpha v_\alpha^\theta(0), & x\geqslant\theta.\end{cases}\tag{D.1}$$

我们在函数集 $B[0,1]$ 中求解 v_α^θ, 其中 $B[0,1]$ 由 $[0,1]$ 上的有界 Borel 可测函数构成. 对 $w\in B[0,1]$, 定义范数 $\|w\|=\sup_{0\leqslant x\leqslant 1}|w(x)|$. 以上值函数有界. 另外, 由随机核 Q_0 的可测性知 v_α^θ 是 Borel 可测的.

引理 15 v_α^θ 是 (D.1) 在 $B[0,1]$ 中的唯一解.

证明 定义算子

$$(\mathcal{L}^\theta w)(x)=\begin{cases}x+\alpha\displaystyle\int_0^1 w(x+(1-x)y)dF_\xi(y), & x<\theta,\\ x+\alpha w(0), & x\geqslant\theta,\end{cases}\tag{D.2}$$

其中 $w \in B[0,1]$, 则 $\mathcal{L}^\theta$ 是由 $B[0,1]$ 到自身且是压缩的, 因而具有唯一不动点. □

定义 $B[0,1]$ 中的一个序列: $w_0 = x$, $w_{k+1} = \mathcal{L}^\theta w_k$, $k \geqslant 0$, 其中 $\theta \in (0,1)$.

引理 16 对每个 $k \geqslant 0$,

(1) $w_k \leqslant w_{k+1}$;

(2) $w_k(x)$ 在 S 上严格增.

证明 (1) 可由归纳法得证. 我们用归纳法证明 (2). 它对 $k = 0$ 成立. 假定 w_k 是严格增的. 通过考虑 $0 \leqslant x_1 < x_2 \leqslant \theta$ 或 $\theta \leqslant x_1 < x_2 \leqslant 1$, 我们利用 (D.2) 得到 $w_{k+1}(x_1) < w_{k+1}(x_2)$. 因此, w_{k+1} 关于 x 严格增. □

引理 17 若 $0 \leqslant x_1 < x_2 \leqslant 1$, 则 $v_\alpha^\theta(x_1) < v_\alpha^\theta(x_2)$.

证明 由引理 15 的证明, $\lim_{k\to\infty} \|w_k - v_\alpha^\theta\| = 0$. 由引理 16 已得到 w_k 严格增, 因此,

$$v_\alpha^\theta(x_1) \leqslant v_\alpha^\theta(x_2), \quad \text{对 } x_1 < x_2.$$

结合上式与 (D.1) 右端, 则可得 $v_\alpha^\theta(x_1) < v_\alpha^\theta(x_2)$. □

考虑 $0 < \theta_1 < \theta_2 < 1$, 再定义值函数 $v_\alpha^{\theta_1}(x)$ 和 $v_\alpha^{\theta_2}(x)$, $x \in S$. 对 $l = 1, 2$, 定义 $B[0,1]$ 中的序列如下: $w_0^{\theta_l} = x$, $w_{k+1}^{\theta_l} = \mathcal{L}^{\theta_l} w_k^{\theta_l}$, $k \geqslant 0$, 则

$$\lim_{k\to\infty} \|w_k^{\theta_l} - v_\alpha^{\theta_l}\| = 0. \tag{D.3}$$

引理 18 对每个 $x \in S$, $w_k^{\theta_1}(x) \leqslant w_k^{\theta_2}(x)$.

证明 我们有如下表示:

$$w_{k+1}^{\theta_1}(x) = \begin{cases} x + \alpha w_k^{\theta_1}(0), & \theta_2 \leqslant x \leqslant 1, \\ x + \alpha w_k^{\theta_1}(0), & \theta_1 \leqslant x < \theta_2, \\ x + \alpha \displaystyle\int_0^1 w_k^{\theta_1}(x + (1-x)y) dF_\xi(y), & 0 \leqslant x < \theta_1, \end{cases} \tag{D.4}$$

$$w_{k+1}^{\theta_2}(x) = \begin{cases} x + \alpha w_k^{\theta_2}(0), & \theta_2 \leqslant x \leqslant 1, \\ x + \alpha \displaystyle\int_0^1 w_k^{\theta_2}(x + (1-x)y) dF_\xi(y), & \theta_1 \leqslant x < \theta_2, \\ x + \alpha \displaystyle\int_0^1 w_k^{\theta_2}(x + (1-x)y) dF_\xi(y), & 0 \leqslant x < \theta_1. \end{cases} \tag{D.5}$$

下面用归纳法证明. 对 $k = 0$, 引理成立. 假设引理对 $k \geqslant 0$ 成立, 即

$$w_k^{\theta_1}(x) \leqslant w_k^{\theta_2}(x), \quad x \in S.$$

我们继续考察 $k+1$ 的情形. 根据 x 分为 3 种情形.

(i) $\theta_2 \leqslant x \leqslant 1$. 由 (D.4) 和 (D.5) 可得 $w_{k+1}^{\theta_1}(x) \leqslant w_{k+1}^{\theta_2}(x)$.

(ii) $\theta_1 \leqslant x < \theta_2$. 因 $w_k^{\theta_1}$ 严格增, 则 $w_k^{\theta_1}(0) \leqslant w_k^{\theta_1}(x + (1-x)y)$, 对任意 $x, y \in S$. 由归纳假设知,

$$w_k^{\theta_1}(x + (1-x)y) \leqslant w_k^{\theta_2}(x + (1-x)y), \tag{D.6}$$

可得

$$w_{k+1}^{\theta_1}(x) = x + \alpha \int_0^1 w_k^{\theta_1}(0) dF_\xi(y) \leqslant x + \alpha \int_0^1 w_k^{\theta_2}(x + (1-x)y) dF_\xi(y).$$

于是, $w_{k+1}^{\theta_1}(x) \leqslant w_{k+1}^{\theta_2}(x)$.

(iii) $0 \leqslant x < \theta_1$. 由 (D.4)–(D.6) 得 $w_{k+1}^{\theta_1}(x) \leqslant w_{k+1}^{\theta_2}(x)$.

结合 (i)–(iii), 可断定引理对 $k+1$ 成立. 证毕. □

引理 19 对 $0 < \theta_1 < \theta_2 < 1$, $v_\alpha^{\theta_1}(x) \leqslant v_\alpha^{\theta_2}(x)$, 其中 $x \in S$.

证明 由引理 18 和 (D.3), 本引理即得证. □

定理 5 的证明 由 $\{x_t^{i,\theta}, t \geqslant 0\}$ 的遍历性, 可得到 $z(\theta_l) = \lim_{\alpha\uparrow 1}(1-\alpha)v_\alpha^{\theta_l}(x)$, 其右端不依赖于 x. 引理 19 蕴含 $z(\theta_1) \leqslant z(\theta_2)$. □

Mean field stochastic games: Monotone costs and threshold policies

HUANG MinYi & MA Yan

Abstract This paper considers mean field games in a multi-agent Markov decision process framework. Each player has a continuum state and binary action. By active control, a player can bring its state to a resetting point. All players are coupled through their cost functions. The structural property of the individual strategies is characterized in terms of threshold policies when the mean field game admits a solution. We further introduce a stationary equation system of the mean field game and analyze uniqueness of its solution under positive externalities.

Keywords **mean field game, Markov decision process, dynamic programming, threshold policy, stationary distribution**

MSC(2010) **60J05, 90C40, 91A10, 91A15**

doi: 10.1360/N012016-00057