中国科学: 数学 2024 年 第 54 卷 第 12 期: 1963 ～ 1978
SCIENTIA SINICA Mathematica
《中国科学》杂志社
SCIENCE CHINA PRESS

论 文

Borel 状态空间中平均零和随机博弈的新条件

献给中山大学建校百年暨中山大学数学学科建设 100 周年

郭先平[1], 廖景浩[1], 谭梓祺[1], 温馨[2*]

1. 中山大学数学学院, 广州 510275;
2. 中山大学管理学院, 广州 510275
E-mail: mcsgxp@mail.sysu.edu.cn, liaojh6@mail3.sysu.edu.cn, tanzq5@mail.sysu.edu.cn, wenx33@mail3.sysu.edu.cn

收稿日期: 2024-03-06; 接受日期: 2024-05-15; 网络出版日期: 2024-09-12; * 通信作者
国家重点研发计划 (批准号: 2022YFA1004600) 和国家自然科学基金 (批准号: 11931018, 72342006 和 72301304) 资助项目

摘要 本文研究 Borel 状态空间的离散时间 Markov 平均博弈. 对报酬函数可以无界的一般情形, 本文用平均最优双不等式取代相应的 Shapley 方程, 提出比现有的几何遍历性条件更弱的新条件. 在此新的条件下, 本文建立上述平均最优双不等式的可解性, 并由此证明平均博弈的值和 Nash 均衡策略的存在性. 进而, 在较强的几何遍历性条件下, 用本文的最优双不等式, 证明 Shapley 方程的可解性. 最后, 用电力系统与金融保险中的例子验证本文的条件, 阐明本文的结果.

关键词 零和平均随机博弈 最优性条件 平均最优双不等式 Shapley 方程 Nash 均衡策略

MSC (2020) 主题分类 91A15, 90D10, 60J10

1 引言

众所周知, 双人非合作零和平均随机博弈已被广泛研究. 为证明平均零和随机博弈 Nash 均衡策略的存在性, 平均 Shapley 方程起着关键作用. 例如, 文献 [12, 14, 15, 21, 28]、[5–8]、[13, 16, 22, 24, 27] 和 [3, 9, 29] 分别对离散时间随机折扣与平均博弈、连续时间折扣与平均博弈、半 Markov 折扣博弈以及随机微分博弈的相应 Shapley 方程展开了大量研究. 本文致力于离散时间随机平均博弈的研究, 以下仅对离散时间随机平均博弈的相关工作加以简述. 在 20 世纪 80 年代, Mertens 和 Neyman [17] 对有限状态和有限行动的离散时间折扣与平均博弈, 证明了相应博弈的值和 Nash 均衡策略的存在性. 然而, Puterman [25] 用例子表明, 当状态空间为可数无限时, 离散时间平均博弈的 Nash 均衡策略可能不存在. 为此, 许多学者致力于离散时间平均博弈 Nash 均衡策略存在性条件的研究. 例如, 文献 [1, 2, 4, 6, 17, 20, 26] 对随机博弈模型中的状态空间、行动空间、报酬函数或转移概率等需要的条件进行了广泛探讨, 其中, 文献 [1, 4, 20] 关于可数状态离散时间平均博弈的研究需要 w- 几何遍历性条件, 且文献 [20] 中的主要结果还需要比 w- 几何遍历性更强的条件. 另外, Borkar 和 Ghosh [2] 在 Liapunov

英文引用格式: Guo X P, Liao J H, Tan Z Q, et al. New conditions for zero-sum stochastic games with average criteria in Borel space (in Chinese). Sci Sin Math, 2024, 54: 1963–1978, doi: 10.1360/SSM-2024-0055

稳定性条件下, 对于可数状态离散时间的多人非零和平均博弈, 不仅证明了 Nash 均衡策略的存在性, 而且还证明了双人零和博弈的值和 Nash 均衡策略的存在性, 但其所用的 Liapunov 稳定性条件本质上等价于 w- 几何遍历性. Sennott [26] 对状态空间可数的离散时间平均博弈, 在有限行动和非负报酬的情形下, 证明了平均 Shapley 方程的解与平均 Nash 均衡策略的存在性. 特别地, 为证明 Shapley 方程解的存在性, 文献 [12, 14, 15, 21] 需要相应的齐次 Markov 过程满足一致几何遍历性和一致 λ- 不可约性. 此外, 基于折扣博弈值的相对差, 文献 [28] 提出了新的比一致几何遍历性更弱的条件下, 确立了平均 Shapley 方程的可解性, 证明了平均博弈的值和 Nash 均衡策略的存在性, 但其讨论仍需要状态空间可数的条件. 然而, 有大量博弈模型中的状态空间是不可数的, 例如以电量为状态的电力系统中电能供给的博弈模型, 以资产为状态的保险与再保险博弈模型, 其状态空间均是不可数的. 源于若干现实问题的驱动, 本文致力于一般状态随机平均博弈的研究. 本文不仅将文献 [28] 的结果拓展到一般状态空间的情形, 而且用平均最优双不等式取代文献 [12, 14, 15, 21] 中的 Shapley 方程, 还基于折扣博弈值的相对差, 提出了比文献 [12, 14, 15, 21] 中几何遍历性和一致 λ- 不可约性更弱的新条件. 在此新的条件下, 本文建立上述平均最优双不等式的可解性, 并由此证明平均博弈的值和 Nash 均衡策略的存在性. 进而, 在标准的几何遍历性条件下, 不同于文献 [12] 的证明方法, 利用本文的最优双不等式, 更简洁地证明了 Shapley 方程的可解性. 最后, 应用电力系统与金融保险中的例子验证本文的条件, 并阐明本文条件与以往相应条件的不同之处.

本文余下内容的结构如下. 第 2 节介绍离散时间随机博弈模型和基本概念. 第 3 节提出新的最优性条件, 并证明平均博弈的值和 Nash 均衡策略的存在性. 第 4 节在几何遍历性条件下, 通过本文的平均最优双不等式, 建立平均 Shapley 方程. 第 5 节用两个应用例子阐明本文条件的优越性.

2 离散时间双人零和随机博弈模型

离散时间双人零和随机博弈模型如下:

$$\{X, A, B, A(x), B(x), Q, r\},$$

其中

- X 表示状态空间, 假定是一个 Borel 空间;
- A 和 B 分别为玩家 1 和 2 的 Borel 行动空间;
- $A(x) \subseteq A$ 和 $B(x) \subseteq B$ 分别表示玩家 1 和 2 在状态 x 处的允许行动集, 假设均为紧的;
- Q 为给定 $\mathbb{K}$ 于 X 上的随机核, 其中 $\mathbb{K} := \{(x, a, b) \mid x \in X, a \in A(x), b \in B(x)\}$ 为 Borel 空间;
- $r(x, a, b)$ 为实值可测函数, 表示玩家 1 的单阶段报酬, 即为玩家 2 支付给玩家 1 的费用;
- 对给定的 $x \in X$, $r(x, a, b)$ 关于 (a, b) 在 $A(x) \times B(x)$ 上连续.

该离散时间零和博弈的演化过程, 可简述如下: 假设系统在初始时间 0 时处于状态 $x_0 \in X$, 玩家 1 和 2 根据系统目前状态 x_0, 分别独立地选择行动 $a_0 \in A(x_0)$ 和 $b_0 \in B(x_0)$. 作为该行动的选择的结果, 会发生以下事件: (1) 玩家 1 获得报酬 $r(x_0, a_0, b_0)$ (即玩家 2 支付给玩家 1 的费用); (2) 系统以概率 $Q(x_1 \mid x_0, a_0, b_0)$ 转移到新的状态 x_1. 根据新的当前状态 x_1 及状态与行动历史 x_0, a_0 和 b_0, 玩家们各自采取新的行动 $a_1 \in A(x_1)$ 和 $b_1 \in B(x_1)$. 玩家 1 获得新的报酬 $r(x_1, a_1, b_1)$. 如此过程的不断演变, 直至时刻 n, 得到可允许的历史

$$h_n = (x_0, a_0, b_0, \ldots, x_{n-1}, a_{n-1}, b_{n-1}, x_n),$$

其中 $(x_m, a_m, b_m) \in \mathbb{K}$, $m = 0, 1, \ldots, n-1$, $x_n \in X$, h_n 称为第 n 阶段的博弈历史. 在此博弈过程中, 玩家 1 希望最大化其报酬. 自然地, 玩家 2 需要最小化其成本.

为准确定义相应的优化准则, 需要给出策略的定义. 为此, 令 $H_0 := X$,

$$H_n := \mathbb{K}^n \times X = \mathbb{K} \times H_{n-1}, \quad \forall n = 1, \ldots$$

定义 2.1　称序列 $\pi^1 = \{\pi_n^1\}$ 为玩家 1 的策略, 如果其中 π_n^1 为给定 H_n 下于 A 上的随机核, 且满足

$$\pi_n^1(A(x_n) \mid h_n) = 1(h_n \in H_n, n = 0, 1, \ldots).$$

用 Π_1 表示玩家 1 所有策略构成的集合.

令 Φ_1 表示给定 X 下在 A 上且满足 $\varphi(A(x) \mid x) = 1, \forall x \in X$ 的随机核 φ 组成的集合.

策略 $\pi^1 = \{\pi_n^1\} \in \Pi_1$ 称为平稳的, 如果存在 $\varphi \in \Phi_1$, 使得

$$\pi_n^1(\cdot \mid h_n) = \varphi(\cdot \mid x_n), \quad \forall h_n \in H_n, \quad n = 0, 1, \ldots$$

记 Π_1^S 为玩家 1 的所有平稳策略族.

相应地, 用 $B(x)$ 和 B 分别代替 $A(x)$ 和 A, 可以定义玩家 2 的策略 $\pi^2 = \{\pi_n^2\}$、平稳策略 ψ、策略族 Π_2 和平稳策略族 Π_2^S. 显然有 $\Pi_i^S \subset \Pi_i$, $i = 1, 2$.

用 $\mathcal{B}(E)$ 表示给定集合 E 的对应 Borel σ- 代数. 对于任意随机策略对 $(\pi^1, \pi^2) \in \Pi_1 \times \Pi_2$ 和 X 上的任意概率测度 ν, 根据 Ionescu-Tulcea 定理 (参见文献 [10, 附录 C 中的命题 C.10 和注记 C.11]), 存在唯一的概率测度空间 $(\Omega, \mathcal{F}, P_\nu^{\pi^1,\pi^2})$ 及其上的随机过程 $\{x_n, a_n, b_n\}$, 使得对于任意 $G \in \mathcal{B}(X)$, $C \in \mathcal{B}(A)$, $D \in \mathcal{B}(B)$, $n \geqslant 0$, 都有

$$\begin{aligned}
&P_\nu^{\pi^1,\pi^2}(x_0 \in G) = \nu(G), \\
&P_\nu^{\pi^1,\pi^2}(x_{n+1} \in G \mid x_0, a_0, b_0, \ldots, x_n, a_n, b_n) = Q(G \mid x_n, a_n, b_n), \\
&P_\nu^{\pi^1,\pi^2}(a_n \in C, b_n \in D \mid x_0, a_0, b_0, \ldots, x_n) = \pi_n^1(C \mid x_0, a_0, b_0, \ldots, x_n)\pi_n^2(D \mid x_0, a_0, b_0, \ldots, x_n).
\end{aligned} \tag{2.1}$$

相应于 $P_\nu^{\pi^1,\pi^2}$ 的期望算子记为 $E_\nu^{\pi^1,\pi^2}$. 特别地, 当 $\nu(\cdot)$ 为集中于 $x \in X$ 的 Dirac 测度时, $P_\nu^{\pi^1,\pi^2}$ 和 $E_\nu^{\pi^1,\pi^2}$ 可分别记为 $P_x^{\pi^1,\pi^2}$ 和 $E_x^{\pi^1,\pi^2}$.

此外, 用 $\mathbb{P}(E)$ 表示 Borel 集合 E 上的所有概率测度族, 对于任意 $G \in \mathcal{B}(X)$ 以及概率测度 $\varphi \in \mathbb{P}(A(x))$ 和 $\psi \in \mathbb{P}(B(x))$, 对于报酬函数 r 和随机核 Q, 定义

$$\begin{aligned}
r(x, \varphi, \psi) &:= \int_{A(x)} \int_{B(x)} r(x, a, b)\psi(db)\varphi(da), \\
Q(G \mid x, \varphi, \psi) &:= \int_{A(x)} \int_{B(x)} Q(G \mid x, a, b)\psi(db)\varphi(da),
\end{aligned}$$

且对于任意 $(\pi^1, \pi^2) \in \Pi_1^S \times \Pi_2^S$, $x \in X$, $n = 1, \ldots$, 可以定义

$$Q^n(G \mid x, \pi^1, \pi^2) := P_x^{\pi^1,\pi^2}(x_n \in G), \quad r(x, \pi^1, \pi^2) := r(x, \pi^1(\cdot \mid x), \pi^2(\cdot \mid x)).$$

特别地,

$$Q(G \mid x, \pi^1, \pi^2) := Q^1(G \mid x, \pi^1(\cdot \mid x), \pi^2(\cdot \mid x)).$$

当 $n = 0$ 时, 令 $Q^0(G \mid x, \pi^1, \pi^2) = 1_G(x)$, 其中 $1_G(x)$ 表示 G 上的示性函数.

下面给出平均准则的定义.

给定初始状态 $x \in X$ 以及策略对 $(\pi^1, \pi^2) \in \Pi_1 \times \Pi_2$, 其 n 步期望准则定义为

$$J_n(x, \pi^1, \pi^2) := E_x^{\pi^1, \pi^2}\left[\sum_{t=0}^{n-1} r(x_t, a_t, b_t)\right], \quad \forall n \geqslant 1,$$

(长期期望) 平均准则定义为 $J(x, \pi^1, \pi^2) := \liminf_{n\to\infty} \frac{J_n(x, \pi^1, \pi^2)}{n}$.

进而, 分别定义平均准则的下值和上值 (函数):

$$\underline{J}(x) := \sup_{\pi^1 \in \Pi_1} \inf_{\pi^2 \in \Pi_2} J(x, \pi^1, \pi^2), \quad \overline{J}(x) := \inf_{\pi^2 \in \Pi_2} \sup_{\pi^1 \in \Pi_1} J(x, \pi^1, \pi^2),$$

显然 $\underline{J}(x) \leqslant \overline{J}(x)$. 若对于任意 $x \in X$, 都有 $\underline{J}(x) = \overline{J}(x)$, 则称这个共同的值为平均准则下博弈的值, 记为 $J(x)$. 在此基础上, 定义玩家的 Nash 均衡策略.

定义 2.2 设平均准则下博弈的值 $J(x)$ 存在, 若策略 $\pi_*^1 \in \Pi_1$ 满足

$$\inf_{\pi^2 \in \Pi_2} J(x, \pi_*^1, \pi^2) = J(x), \quad \forall x \in X,$$

则称 π_*^1 为玩家 1 在平均准则下的最优策略. 类似地, 若策略 $\pi_*^2 \in \Pi_2$ 满足

$$\sup_{\pi^1 \in \Pi_1} J(x, \pi^1, \pi_*^2) = J(x), \quad \forall x \in X,$$

则称 π_*^2 为玩家 2 在平均准则下的最优策略. 若 $\pi_*^i \in \Pi_i$ 为玩家 i $(i = 1, 2)$ 平均准则下的最优策略, 则称 (π_*^1, π_*^2) 为平均准则下的 Nash 均衡策略 (以下简称为平均 Nash 均衡策略) .

本文的目标是给出平均 Nash 均衡策略新的存在性条件.

3 平均最优双不等式和 Nash 均衡策略的存在性

本节建立平均最优双不等式, 提出新的 Nash 均衡策略的存在性条件. 为此, 引入以下记号: 对于任意 Borel- 可测函数 (权函数) $w : X \mapsto [1, +\infty)$, 称 X 上的实值可测函数 v 是 w- 有界的, 若其 w- 范数 $\|v\|_w := \sup_{x\in X} \frac{|v(x)|}{w(x)}$ 是有限的. 记 $\mathbb{B}_w(X)$ 为 X 上所有实值 w- 有界可测函数 u 组成的赋范线性空间. 此外, 记 $\mathbb{B}(X)$ 为 X 上所有实值有界可测函数构成的 Banach 空间.

在给出本文的最优性条件之前, 需要折扣博弈的若干结论. 为此, 对于任意状态 $x \in X$、策略对 $(\pi^1, \pi^2) \in \Pi_1 \times \Pi_2$ 和折扣因子 $\alpha \in (0, 1)$, 在一定的条件下, 定义 α- 折扣准则为

$$V_\alpha(x, \pi^1, \pi^2) := E_x^{\pi^1, \pi^2}\left[\sum_{t=0}^{\infty} \alpha^t r(x_t, a_t, b_t)\right].$$

类似平均准则, 可定义折扣准则的下值和上值分别为

$$\underline{V_\alpha}(x) := \sup_{\pi^1 \in \Pi_1} \inf_{\pi^2 \in \Pi_2} V_\alpha(x, \pi^1, \pi^2), \quad \overline{V_\alpha}(x) := \inf_{\pi^2 \in \Pi_2} \sup_{\pi^1 \in \Pi_1} V_\alpha(x, \pi^1, \pi^2).$$

显然,

$$\underline{V_\alpha}(x) \leqslant \overline{V_\alpha}(x), \quad \forall x \in X.$$

若 $\underline{V_\alpha}(x) = \overline{V_\alpha}(x)$, 则称此共同值为 α- 折扣博弈的值, 记为 $V_\alpha(x)$.

为了建立 α- 折扣 Shapley 方程, 证明 α- 折扣博弈值的存在性, 本文引入以下条件.

假设 3.1 对任给定 $x \in X$, $u \in \mathbb{B}(X)$, $\int_X u(y)Q(dy \mid x,a,b)$ 关于 (a,b) 在 $A(x) \times B(x)$ 上连续.

假设 3.2 存在正常数 L、$\gamma < 1$ 和 η, 以及定义于 X 上的可测函数 $w_1 \geqslant 1$, 使得

(a) $|r(x,a,b)| \leqslant L w_1(x)$, $\forall\, (x,a,b) \in \mathbb{K}$;

(b) $\int_X w_1(y)Q(dy \mid x,a,b) \leqslant \gamma w_1(x) + \eta$, $\forall\, (x,a,b) \in \mathbb{K}$.

引理 3.1 在假设 3.1 和 3.2 条件下, 对于任意 $x \in X$, α- 折扣博弈的值 $V_\alpha(x)$ 存在, 且存在平稳策略对 $(\pi_\alpha^1, \pi_\alpha^2) \in \Pi_1^S \times \Pi_2^S$, 使得对所有的 $x \in X$, 有

$$\begin{aligned} V_\alpha(x,\pi_\alpha^1,\pi_\alpha^2) = V_\alpha(x) &= r(x,\pi_\alpha^1,\pi_\alpha^2) + \alpha \int_X V_\alpha(y)Q(dy \mid x,\pi_\alpha^1,\pi_\alpha^2) \\ &= \min_{\psi \in \mathbb{P}(B(x))} \left[r(x,\pi_\alpha^1,\psi) + \alpha \int_X V_\alpha(y)Q(dy \mid x,\pi_\alpha^1,\psi) \right] \\ &= \max_{\varphi \in \mathbb{P}(A(x))} \left[r(x,\varphi,\pi_\alpha^2) + \alpha \int_X V_\alpha(y)Q(dy \mid x,\varphi,\pi_\alpha^2) \right]. \end{aligned} \tag{3.1}$$

证明 在本文的条件下, 用文献 [28, 引理 3.2] 和 [10] 中的证明方法, 易知上述结论成立. □

基于 α- 折扣博弈的结果, 下面给出关于平均准则博弈的值和 Nash 均衡策略存在的新条件.

假设 3.3 存在某个状态 $x_0 \in X$, 令 $h_\alpha(x) := V_\alpha(x) - V_\alpha(x_0)$. 假设存在 $\mathbb{B}_{w_1}(X)$ 中两个函数 v_1 和 v_2, 使得

$$v_1(x) \leqslant h_\alpha(x) \leqslant v_2(x), \quad \forall\, x \in X, \quad \alpha \geqslant \gamma.$$

注 3.1 (a) 假设 3.3 不需要文献 [12,14,15,21] 中要求状态过程 $\{x_n\}$ 满足一致几何遍历性和一致 λ- 不可约性. 在此条件下, 也可以证明平均博弈的值和 Nash 均衡策略存在 (见定理 3.1).

(b) 假设 3.3 既是文献 [28] 中假设 C 向一般状态空间的拓展, 又是离散时间 Markov 决策过程[10]相应假设的延伸.

(c) 假设 3.3 中的函数 v_1 和 v_2 均可能无界, 因此函数 $h_\alpha(x)$ 也可能既无上界也无下界. 此外, 该假设也不要求函数族 $\{h_\alpha(\cdot), \alpha > 0\}$ 满足等度连续性.

(d) 下文给出的命题 4.1 将证明几何遍历性是假设 3.3 的充分条件, 从而说明假设 3.3 是常用几何遍历性[12,14,15,21]的拓展.

为叙述本节的主要定理, 需要下面引理.

引理 3.2 在假设 3.1 和 3.2 下, 有如下结论成立:

(1) 对于任意给定 $(\pi^1,\pi^2) \in \Pi_1 \times \Pi_2$, 若存在常数 ρ^* 和 $h \in \mathbb{B}_{w_1}(X)$, 使得对于任意 $x \in X$, 都有

$$\rho^* + h(x) \geqslant (\leqslant) r(x,\pi_t^1,\pi_t^2) + \int_X h(y)Q(dy \mid x,\pi_t^1,\pi_t^2), \quad \forall\, t = 0,1,\ldots,$$

则 $\rho^* \geqslant (\leqslant) J(x,\pi^1,\pi^2)$, $\forall x \in X$;

(2) 对于任意给定 $(\pi^1,\pi^2) \in \Pi_1^S \times \Pi_2^S$, 若存在常数 ρ^* 和 $h \in \mathbb{B}_{w_1}(X)$, 使得对于任意 $x \in X$, 都满足

$$\rho^* + h(x) = r(x,\pi_t^1,\pi_t^2) + \int_X h(y)Q(dy \mid x,\pi_t^1,\pi_t^2), \quad \forall\, t = 0,1,\ldots, \tag{3.2}$$

则 $\rho^* = J(x,\pi^1,\pi^2)$, $\forall\, x \in X$.

证明 由文献 [10, 引理 5.2.5] 可知结论成立. □

应用以上引理, 下面给出本节的主要结论.

定理 3.1 在假设 3.1–3.3 条件下, 如下结论成立:

(1) 存在常数 ρ^* 和函数 $h_1, h_2 \in \mathbb{B}_{w_1}(X)$, 使得以下平均最优双不等式成立:

$$\rho^* + h_1(x) \geqslant \inf_{\psi\in\mathbb{P}(B(x))} \sup_{\varphi\in\mathbb{P}(A(x))} \left[r(x,\varphi,\psi) + \int_X h_1(y)Q(dy \mid x,\varphi,\psi)\right], \quad \forall\, x\in X, \tag{3.3}$$

$$\rho^* + h_2(x) \leqslant \sup_{\varphi\in\mathbb{P}(A(x))} \inf_{\psi\in\mathbb{P}(B(x))} \left[r(x,\varphi,\psi) + \int_X h_2(y)Q(dy \mid x,\varphi,\psi)\right], \quad \forall\, x\in X; \tag{3.4}$$

(2) 存在平稳策略对 $(\pi_*^1, \pi_*^2) \in \Pi_1^S \times \Pi_2^S$, 使得

$$\rho^* + h_1(x) \geqslant r(x,\pi_*^1,\pi_*^2) + \int_X h_1(y)Q(dy \mid x,\pi_*^1,\pi_*^2), \quad \forall\, x\in X, \tag{3.5}$$

$$\rho^* + h_2(x) \leqslant r(x,\pi_*^1,\pi_*^2) + \int_X h_2(y)Q(dy \mid x,\pi_*^1,\pi_*^2), \quad \forall\, x\in X; \tag{3.6}$$

(3) 对于任意 $x\in X$, 都有 $\rho^* = \underline{J}(x) = \overline{J}(x) = J(x)$, 这意味着 ρ^* 为博弈在平均准则下的值.

(4) (2) 中的 (π_*^1, π_*^2) 为平均 Nash 均衡策略.

证明 (1) 在假设 3.1 和 3.2 条件下, 对于任意 $x_0 = x\in X$ 及策略对 $(\pi^1,\pi^2)\in\Pi_1\times\Pi_2$, 都有

$$\begin{aligned}
|V_\alpha(x,\pi^1,\pi^2)| &\leqslant L\sum_{t=0}^{\infty}\alpha^t E_x^{\pi^1,\pi^2}[w_1(x_t)] \\
&\leqslant L\sum_{t=0}^{\infty}\alpha^t\left[\gamma^t\omega_1(x) + \frac{1-\gamma^t}{1-\gamma}\eta\right] &(3.7)\\
&\leqslant \frac{Lw_1(x)}{1-\alpha\gamma} + \frac{L\eta}{(1-\gamma)(1-\alpha)}, &(3.8)
\end{aligned}$$

其中不等式 (3.7) 可由归纳法证明. 进而, 选取折扣因子序列 $\{\alpha(m)\}$, 满足 $\alpha(m)\uparrow 1$. 由 (3.8) 及引理 3.1 可知 $\{(1-\alpha(m))V_{\alpha(m)}(x_0)\}$ 有界. 故存在子列 $\{\alpha(n)\}$ 和常数 ρ^*, 其中 $\alpha(n)\uparrow 1$, 使得

$$\lim_{n\to\infty}(1-\alpha(n))V_{\alpha(n)}(x_0) = \rho^*.$$

对任给的 $x\in X$, 令

$$h_1(x) := \liminf_{n\to\infty} h_{\alpha(n)}(x) = \lim_{n\to\infty} H_n(x),$$

其中 $H_n(x) := \inf_{k\geqslant n}\{h_{\alpha(k)}(x)\}$. 由引理 3.1 及 (3.1) 可以推导出

$$\begin{aligned}
(1-\alpha)V_\alpha(x_0) + h_\alpha(x) &= \max_{\varphi\in\mathbb{P}(A(x))}\left[r(x,\varphi,\pi_\alpha^2) + \alpha\int_X h_\alpha(y)Q(dy\mid x,\varphi,\pi_\alpha^2)\right] \\
&\geqslant r(x,\varphi,\pi_\alpha^2) + \alpha\int_X h_\alpha(y)Q(dy\mid x,\varphi,\pi_\alpha^2), \quad \forall\,\varphi\in\mathbb{P}(A(x)). &(3.9)
\end{aligned}$$

因为 $\mathbb{P}(B(x))$ 是紧的, 故在 $\{\pi_{\alpha(n)}^2(x)\}$ 中存在子列 $\{\pi_{\alpha(n_k^x)}^2(x)\}$, 使得 $\pi_{\alpha(n_k^x)}^2(x)\to\psi'(x)$ $(k\to\infty)$, 且 $\psi'(x)\in\mathbb{P}(B(x))$. 由于序列 $\{H_{n_k^x}(x)\}$ 收敛, 所以有

$$\lim_{k\to\infty} H_{n_k^x}(x) = h_1(x).$$

再根据假设 3.3, 对所有的 $n_k^x\geqslant 1$, 都有

$$\|h_{\alpha(n_k^x)}\|_{w_1} \leqslant \|v_1\|_{w_1} + \|v_2\|_{w_1}.$$

因此, 将 (3.9) 的 π_α^2 替换为 $\pi^2_{\alpha(n_k^x)}$, 应用 $r(x,\cdot,\cdot)$ 的连续性以及文献 [11] 推广的 Fatou 引理 8.3.7, 可得

$$\begin{aligned}\rho^*+h_1(x)&\geqslant\liminf_{k\to\infty}\left[r(x,\varphi,\pi^2_{\alpha(n_k^x)}(x))+\alpha(n_k^x)\int_X h_{\alpha(n_k^x)}(y)Q(dy\mid x,\varphi,\pi^2_{\alpha(n_k^x)}(x))\right]\\&\geqslant r(x,\varphi,\psi'(x))+\int_X h_1(y)Q(dy\mid x,\varphi,\psi'(x)),\quad\forall\,\varphi\in\mathbb{P}(A(x)),\quad x\in X.\end{aligned}$$

由 φ 的任意性, 可得

$$\rho^*+h_1(x)\geqslant\sup_{\varphi\in\mathbb{P}(A(x))}\left[r(x,\varphi,\psi'(x))+\int_X h_1(y)Q(dy\mid x,\varphi,\psi'(x))\right],\quad\forall\,x\in X.$$

这便证明了 (3.3). 为证明 (3.4), 仅需要令 $h_2(x):=\limsup_{n\to\infty}h_{\alpha(n)}(x)$, 并同理证明即可.

(2) 根据 (3.3) 和文献 [19] 的可测选择定理, 存在平稳策略对 $\pi_*^2\in\Pi_2^S$, 使得

$$\begin{aligned}\rho^*+h_1(x)&\geqslant\inf_{\psi\in\mathbb{P}(B(x))}\sup_{\varphi\in\mathbb{P}(A(x))}\left[r(x,\varphi,\psi)+\int_X h_1(y)Q(dy\mid x,\varphi,\psi)\right]\\&=\sup_{\varphi\in\mathbb{P}(A(x))}\left[r(x,\varphi,\pi_*^2)+\int_X h_1(y)Q(dy\mid x,\varphi,\pi_*^2)\right]\\&\geqslant r(x,\varphi,\pi_*^2)+\int_X h_1(y)Q(dy\mid x,\varphi,\pi_*^2),\quad\forall\,x\in X,\quad\varphi\in\mathbb{P}(A(x)).\end{aligned}\tag{3.10}$$

同理可知, 存在平稳策略对 $\pi_*^1\in\Pi_1^S$, 使得

$$\rho^*+h_2(x)\leqslant r(x,\pi_*^1,\psi)+\int_X h_2(y)Q(dy\mid x,\pi_*^1,\psi),\quad\forall\,x\in X,\quad\psi\in\mathbb{P}(B(x)).$$

由此可得 (3.5) 和 (3.6).

(3) 根据 (3.10) 和引理 3.2, 有

$$\rho^*\geqslant J(x,\pi^1,\pi_*^2),\quad\forall\,\pi^1\in\Pi_1,\tag{3.11}$$

由 π^1 的任意性, 可得

$$\rho^*\geqslant\sup_{\pi^1\in\Pi_1}J(x,\pi^1,\pi_*^2),$$

于是有

$$\rho^*\geqslant\inf_{\pi^2\in\Pi_2}\sup_{\pi^1\in\Pi_1}J(x,\pi^1,\pi_*^2)=\overline{J}(x).$$

同理可证 $\rho^*\leqslant\underline{J}(x)$, 显然有 $\underline{J}(x)\leqslant\overline{J}(x)$, 这便证明了 $\rho^*=\underline{J}(x)=\overline{J}(x)=J(x)$.

(4) 根据 (3.11) 和定理 3.1(3), 可得 π_*^2 为玩家 2 在平均准则下的最优平稳策略, 同理可证 π_*^1 为玩家 1 在平均准则下的最优平稳策略, 即 (π_*^1,π_*^2) 为平均准则下的平稳 Nash 均衡策略. □

注 3.2　定理 3.1 确立了平均博弈的值和 Nash 均衡策略的存在性, 但其假设条件比一致几何遍历性更弱, 且不需要文献 [12,14,15,21] 中要求的 λ- 不可约性. 此外, 定理 3.1 也将文献 [28] 中可数状态空间的主要结果拓展到更一般的 Borel 状态空间情形.

4 平均 Shapley 方程

本节引入几何遍历性条件, 并应用前面建立的平均最优双不等式, 建立平均 Shapley 方程. 在前面的假设条件下, 本文已经得到平均博弈的值和 Nash 均衡策略存在性, 但仅建立平均最优双不等式 ((3.3) 和 (3.4)), 而没得到平均 Shapley 方程 (4.6). 要建立平均 Shapley 方程, 则需要增加条件, 见如下假设 4.1 和 4.2 (参见文献 [12, 注 5.5 及假设 5.7]).

假设 4.1 (w_1- 几何遍历性条件) 存在正常数 $\theta < 1$ 和 M, 使得对每一对平稳策略 $(\pi^1,\pi^2) \in \Pi_1^S \times \Pi_2^S$, 存在概率分布 $\mu_{\pi^1,\pi^2} \in \mathbb{P}(X)$, 使得对所有的 $x \in X$, $n = 0,1,\ldots$, 以及 $u \in \mathbb{B}_{w_1}(X)$, 有

$$\left|\int_X u(y)Q^n(dy \mid x,\pi^1,\pi^2) - \int_X u(y)\mu_{\pi^1,\pi^2}(dy)\right| \leqslant w_1(x)\|u\|_{w_1} M\theta^n. \tag{4.1}$$

命题 4.1 在假设 3.1 和 3.2 的条件下, 假设 4.1 蕴涵假设 3.3.

证明 由 $|r(x,a,b)| \leqslant Lw_1(x)$ 可知 $r(\cdot,\pi_\alpha^1,\pi_\alpha^2) \in \mathbb{B}_{w_1}(X)$. 根据引理 3.1 和 (4.1), 对于任意 $x, x_0 \in X$, 都有

$$\begin{aligned}
|V_\alpha(x) - V_\alpha(x_0)| &= \left|E_x^{\pi_\alpha^1,\pi_\alpha^2}\left[\sum_{t=0}^\infty \alpha^t r(x_t,a_t,b_t)\right] - E_{x_0}^{\pi_\alpha^1,\pi_\alpha^2}\left[\sum_{t=0}^\infty \alpha^t r(x_t,a_t,b_t)\right]\right| \\
&= \sum_{t=0}^\infty \alpha^t \left|\int_X r(y,\pi_\alpha^1,\pi_\alpha^2)Q^t(dy \mid x,\pi_\alpha^1,\pi_\alpha^2) - \int_X r(y,\pi_\alpha^1,\pi_\alpha^2)Q^t(dy \mid x_0,\pi_\alpha^1,\pi_\alpha^2)\right| \\
&\leqslant \sum_{t=0}^\infty \alpha^t \left[\left|\int_X r(y,\pi_\alpha^1,\pi_\alpha^2)Q^t(dy \mid x,\pi_\alpha^1,\pi_\alpha^2) - \int_X r(y,\pi_\alpha^1,\pi_\alpha^2)\mu_{\pi_\alpha^1,\pi_\alpha^2}(dy)\right|\right. \\
&\quad \left. + \left|\int_X r(y,\pi_\alpha^1,\pi_\alpha^2)Q^t(dy \mid x_0,\pi_\alpha^1,\pi_\alpha^2) - \int_X r(y,\pi_\alpha^1,\pi_\alpha^2)\mu_{\pi_\alpha^1,\pi_\alpha^2}(dy)\right|\right] \\
&\leqslant [w_1(x) + w_1(x_0)]\|r\|_{w_1} M \sum_{t=0}^\infty \alpha^t\theta^t \\
&\leqslant \frac{\|r\|_{w_1}M}{1-\theta}[1 + w_1(x_0)]w_1(x) =: v_2(x), \quad \forall\, \alpha \in (0,1),
\end{aligned}$$

然后令 $v_1(x) := -v_2(x)$, 可知假设 3.3 成立. □

下面给出假设 4.1 成立的充分条件.

命题 4.2 假设对任给的平稳策略对 $(\pi^1,\pi^2) \in \Pi_1^S \times \Pi_2^S$, 随机核 $Q(\cdot \mid \cdot,\pi^1,\pi^2)$ 存在唯一的不变概率分布 μ_{π^1,π^2}. 此外, 存在权函数 $w_1 \geqslant 1$、概率测度 $\nu \in \mathbb{P}(X)$、正数 ε 和 $\gamma < 1$, 使对所有的 $(\pi^1,\pi^2) \in \Pi_1^S \times \Pi_2^S$, 有相应的可测函数 $0 \leqslant l_{\pi^1,\pi^2}(\cdot) \leqslant 1$, 使得以下条件成立:

(a) $Q(B \mid x,\pi^1,\pi^2) \geqslant l_{\pi^1,\pi^2}(x)\nu(B)$, $\forall\, x \in X$, $B \in \mathcal{B}(X)$;

(b) $\nu(l_{\pi^1,\pi^2}) := \int_X l_{\pi^1,\pi^2}(x)\nu(dx) \geqslant \varepsilon$;

(c) $\nu(w_1) := \int_X w_1(x)\nu(dx) < \infty$;

(d) $\int_X w_1(y)Q(dy \mid x,\pi^1,\pi^2) \leqslant \gamma w_1(x) + l_{\pi^1,\pi^2}(x)\nu(w_1)$, $\forall\, x \in X$.

则存在常数 M 和 $0 < \theta < 1$, 使得 (4.1) 成立.

证明 证明可参见文献 [11, 命题 10.2.5]. □

假设 4.2 在 X 上存在 σ- 有限的测度 ν 和恒正的密度函数 $g(x,a,b,\cdot)$, 使得

$$Q(D \mid x,a,b) = \int_D g(x,a,b,y)\nu(dy),$$

对所有的 $D\in\mathcal{B}(X)$ 和 $(x,a,b)\in\mathbb{K}$ 都成立.

注 4.1　(1) 假设 4.1 和 4.2 将用于确立平均 Shapley 方程 (见定理 4.1 的证明). 不过, 本文证明平均博弈的值和 Nash 均衡策略存在性时不需要用到这两个假设.

(2) 假设 4.2 蕴涵转移概率

$$Q(\cdot\mid\cdot,\pi^1,\pi^2)(\pi^1,\pi^2\in\Pi_1\times\Pi_2)$$

是 ν- 不可约的.

引理 4.1　在假设 3.2 和 4.1 条件下, 对所有 $(\pi^1,\pi^2)\in\Pi_1^S\times\Pi_2^S$, 如下结论成立:

(1) 令 $j(\pi^1,\pi^2):=\int_X r(y,\pi^1,\pi^2)\mu_{\pi^1,\pi^2}(dy)$, 则

$$j(\pi^1,\pi^2)=\lim_{n\to\infty}\frac{J_n(x,\pi^1,\pi^2)}{n}=J(x,\pi^1,\pi^2),\quad\forall\,x\in X.\tag{4.2}$$

记 $h_{\pi^1,\pi^2}(x):=\lim_{n\to\infty}[J_n(x,\pi^1,\pi^2)-nj(\pi^1,\pi^2)]$, 则 $h_{\pi^1,\pi^2}\in\mathbb{B}_{w_1}(X)$ 且其 w_1- 范数满足

$$\|h_{\pi^1,\pi^2}\|_{w_1}\leqslant\frac{LM}{1-\theta},\quad\forall\,(\pi^1,\pi^2)\in\Pi_1^S\times\Pi_2^S,\tag{4.3}$$

其中 L 来自假设 3.2(a), M 和 θ 源于假设 4.1.

(2) 二元组 $(j(\pi^1,\pi^2),h_{\pi^1,\pi^2})\in\mathbb{R}\times\mathbb{B}_{w_1}(X)$ 满足 Poisson 方程, 即

$$j(\pi^1,\pi^2)+h_{\pi^1,\pi^2}(x)=r(x,\pi^1,\pi^2)+\int_X h_{\pi^1,\pi^2}(y)Q(dy\mid x,\pi^1,\pi^2),\quad\forall\,x\in X\tag{4.4}$$

且 $\int_X h_{\pi^1,\pi^2}(y)\mu_{\pi^1,\pi^2}(dy)=0$.

证明　(1) 对所有的 $n=1,\ldots$ 和 $x\in X$ 及任给 $(\pi^1,\pi^2)\in\Pi_1^S\times\Pi_2^S$, 由假设 4.1 可得

$$\begin{aligned}|J_n(x,\pi^1,\pi^2)-nj(\pi^1,\pi^2)|&=\left|\sum_{t=0}^{n-1}[E_x^{\pi^1,\pi^2}r(x_t,\pi^1,\pi^2)-j(\pi^1,\pi^2)]\right|\\&\leqslant\sum_{t=0}^{n-1}\left|\int_X r(y,\pi^1,\pi^2)Q^t(dy\mid x,\pi^1,\pi^2)-j(\pi^1,\pi^2)\right|\\&\leqslant\sum_{t=0}^{n-1}w_1(x)\|r\|_{w_1}M\theta^t\\&\leqslant\frac{w_1(x)\|r\|_{w_1}M}{1-\theta}\\&\leqslant\frac{w_1(x)LM}{1-\theta}.\end{aligned}\tag{4.5}$$

由 (4.5) 即可证明 (4.2) 和 (4.3).

(2) 由文献 [11] 中的定理 7.5.10(e) 或者定理 10.2.3(d) 即知引理 4.1(2) 也成立. □

引理 4.2　在假设 4.2 下, 对所有的 $(\pi^1,\pi^2)\in\Pi_1^S\times\Pi_2^S$, 有如下结论:

(1) ν 和不变概率测度 μ_{π^1,π^2} 相互等价;

(2) 对所有的 $D\in\mathcal{B}(X)$, 若 $\mu_{\pi^1,\pi^2}(D)=0$, 则 $Q(D\mid x,a,b)=0$, $\forall\,(x,a,b)\in\mathbb{K}$.

证明　(1) 根据假设 4.2, 若 $\nu(D)=0$, 则

$$Q(D\mid x,a,b)=\int_D g(x,a,b,y)\nu(dy)=0,\quad\forall\,(x,a,b)\in\mathbb{K}.$$

由于 μ_{π^1,π^2} 是不变概率测度, 所以有

$$\mu_{\pi^1,\pi^2}(D) = \int_X Q(D \mid x, \pi^1, \pi^2)\mu_{\pi^1,\pi^2}(dy) = 0,$$

即 $\mu_{\pi^1,\pi^2} \ll \nu$.

另外, 根据注 4.1(2), 可以发现转移概率 $Q(\cdot \mid x, \pi^1, \pi^2)$ 是 ν- 不可约的. 再根据文献 [23, 定理 7.2], 有 $\nu \ll \mu_{\pi^1,\pi^2}$. 这便证明了 ν 和 μ_{π^1,π^2} 相互等价.

(2) 根据 (1), 若 $\mu_{\pi^1,\pi^2}(D) = 0$, 则 $\nu(D) = 0$. 结合假设 4.2 可得 $Q(D \mid x, a, b) = 0$. □

下面利用已建立的最优双不等式来推导平均 Shapley 方程, 与文献 [12] 中的方法不同.

定理 4.1 在假设 3.1、3.2、4.1 和 4.2 的条件下, 以下结论成立.

(1) 存在常数 ρ^* 和可测函数 $h(x)$, 使得以下 Shapley 方程成立:

$$\rho^* + h(x) = \inf_{\psi\in\mathbb{P}(B(x))} \sup_{\varphi\in\mathbb{P}(A(x))} \left[r(x,\varphi,\psi) + \int_X h(y)Q(dy \mid x,\varphi,\psi)\right] \tag{4.6}$$

$$= \sup_{\varphi\in\mathbb{P}(A(x))} \inf_{\psi\in\mathbb{P}(B(x))} \left[r(x,\varphi,\psi) + \int_X h(y)Q(dy \mid x,\varphi,\psi)\right], \quad \forall\, x \in X; \tag{4.7}$$

(2) 存在平稳策略对 $(\pi^1_*, \pi^2_*) \in \Pi^S_1 \times \Pi^S_2$, 使得

$$\rho^* + h(x) = r(x, \pi^1_*, \pi^2_*) + \int_X h(y)Q(dy \mid x, \pi^1_*, \pi^2_*), \quad \forall\, x \in X; \tag{4.8}$$

(3) (2) 中的 (π^1_*, π^2_*) 为平均 Nash 均衡策略.

证明 (1) 由定理 3.1 及命题 4.1 可知, 存在常数 ρ^* 和 Nash 均衡策略 $(\pi^1_*, \pi^2_*) \in \Pi^S_1 \times \Pi^S_2$, 使得 (引理 4.1) $j(\pi^1_*, \pi^2_*) = \rho^*$, 且 $(j(\pi^1_*, \pi^2_*), h_{\pi^1_*,\pi^2_*})$ 满足 Poisson 方程, 即

$$j(\pi^1_*, \pi^2_*) + h_{\pi^1_*,\pi^2_*}(x) = r(x, \pi^1_*, \pi^2_*) + \int_X h_{\pi^1_*,\pi^2_*}(y)Q(dy \mid x, \pi^1_*, \pi^2_*), \quad \forall\, x \in X. \tag{4.9}$$

将 (3.5) 与 (4.9) 相减, 可得

$$\begin{aligned} h_1(x) - h_{\pi^1_*,\pi^2_*}(x) &\geqslant \int_X [h_1(y) - h_{\pi^1_*,\pi^2_*}(y)]Q(dy \mid x, \pi^1_*, \pi^2_*) \\ &\geqslant \int_X [h_1(y) - h_{\pi^1_*,\pi^2_*}(y)]Q^n(dy \mid x, \pi^1_*, \pi^2_*), \quad \forall\, x \in X, \quad n = 1, 2, \ldots \end{aligned}$$

然后令 $n \to \infty$, 结合 (4.1), 有

$$h_1(x) - h_{\pi^1_*,\pi^2_*}(x) \geqslant \int_X [h_1(y) - h_{\pi^1_*,\pi^2_*}(y)]\mu_{\pi^1_*,\pi^2_*}(dy), \quad \forall\, x \in X.$$

此时, 取 $k_1 = \inf_{x\in X}[h_1(x) - h_{\pi^1_*,\pi^2_*}(x)]$, 则有

$$h_1(x) - h_{\pi^1_*,\pi^2_*}(x) \geqslant \int_X [h_1(y) - h_{\pi^1_*,\pi^2_*}(y)]\mu_{\pi^1_*,\pi^2_*}(dy) \geqslant k_1, \quad \forall\, x \in X.$$

这意味着存在 Borel 集 $D_1 \subset X$ 且满足 $\mu_{\pi^1_*,\pi^2_*}(D_1) = 1$, 使得

$$h_1(x) = h_{\pi^1_*,\pi^2_*}(x) + k_1, \quad \forall\, x \in D_1. \tag{4.10}$$

同理, 将 (3.6) 与 (4.9) 相减, 可得存在 Borel 集 $D_2 \subset X$ 且满足 $\mu_{\pi_*^1,\pi_*^2}(D_2)=1$, 使得

$$h_2(x) = h_{\pi_*^1,\pi_*^2}(x) + k_2, \quad \forall\, x \in D_2,$$

其中

$$k_2 = \sup_{x\in X}[h_2(x) - h_{\pi_*^1,\pi_*^2}(x)].$$

此时, 令 $D = D_1 \cap D_2$, 则有 $\mu_{\pi_*^1,\pi_*^2}(D)=1$, 从而对于任意 $x \in D$, 由 (3.3) 及引理 4.2 可得

$$\rho^* + h_{\pi_*^1,\pi_*^2}(x) \geqslant \inf_{\psi\in\mathbb{P}(B(x))} \sup_{\varphi\in\mathbb{P}(A(x))} \left[r(x,\varphi,\psi) + \int_X h_{\pi_*^1,\pi_*^2}(y) Q(dy \mid x,\varphi,\psi) \right], \quad \forall\, x \in D.$$

同理, 由 (3.4) 及引理 4.2 可得

$$\rho^* + h_{\pi_*^1,\pi_*^2}(x) \leqslant \sup_{\varphi\in\mathbb{P}(A(x))} \inf_{\psi\in\mathbb{P}(B(x))} \left[r(x,\varphi,\psi) + \int_X h_{\pi_*^1,\pi_*^2}(y) Q(dy \mid x,\varphi,\psi) \right], \quad \forall\, x \in D.$$

这便证明了

$$\begin{aligned} \rho^* + h_{\pi_*^1,\pi_*^2}(x) &= \inf_{\psi\in\mathbb{P}(B(x))} \sup_{\varphi\in\mathbb{P}(A(x))} \left[r(x,\varphi,\psi) + \int_X h_{\pi_*^1,\pi_*^2}(y) Q(dy \mid x,\varphi,\psi) \right] \\ &= \sup_{\varphi\in\mathbb{P}(A(x))} \inf_{\psi\in\mathbb{P}(B(x))} \left[r(x,\varphi,\psi) + \int_X h_{\pi_*^1,\pi_*^2}(y) Q(dy \mid x,\varphi,\psi) \right], \quad \forall\, x \in D. \end{aligned} \tag{4.11}$$

接下来, 定义函数 h 如下:

$$h(x) := \begin{cases} h_{\pi_*^1,\pi_*^2}(x), & x \in D, \\ \inf_{\psi\in\mathbb{P}(B(x))} \sup_{\varphi\in\mathbb{P}(A(x))} r(x,\varphi,\psi) - \rho^*, & x \in D^c. \end{cases}$$

由于 $\mu_{\pi_*^1,\pi_*^2}(D^c)=0$, 根据引理 4.2(2), 有

$$\int_X h(y) Q(dy \mid x,\varphi,\psi) = 0, \quad \forall\, x \in D^c,$$

因此有

$$\rho^* + h(x) = \inf_{\psi\in\mathbb{P}(B(x))} \sup_{\varphi\in\mathbb{P}(A(x))} \left[r(x,\varphi,\psi) + \int_X h(y) Q(dy \mid x,\varphi,\psi) \right], \quad \forall\, x \in D^c.$$

结合 (4.11), 便证明了 (4.6) 和 (4.7), 即对于任意 $x \in X$, 都有

$$\begin{aligned} \rho^* + h(x) &= \inf_{\psi\in\mathbb{P}(B(x))} \sup_{\varphi\in\mathbb{P}(A(x))} \left[r(x,\varphi,\psi) + \int_X h(y) Q(dy \mid x,\varphi,\psi) \right] \\ &= \sup_{\varphi\in\mathbb{P}(A(x))} \inf_{\psi\in\mathbb{P}(B(x))} \left[r(x,\varphi,\psi) + \int_X h(y) Q(dy \mid x,\varphi,\psi) \right], \quad \forall\, x \in X. \end{aligned}$$

(2) 和 (3) 的证明同定理 3.1(2) 和 3.1(4) 的证明. □

5 例子

本节介绍可再生能源发电供给模型以及保险分红这两个不同类型的模型, 并验证这两个模型都满足本文的相关条件 (即满足假设 3.1–3.3), 且其状态空间均为不可数的.

例 5.1 (可再生能源发电供给模型) 设有一个配备电量存储装置的发电社区, 其拥有可再生能源 (如太阳能、风力和水力等), 并可通过可再生能源生产电能, 也会自身消耗电量, 还可能将生产的电量给社区的下级用户 (如用电的家庭和单位等). 发电社区及其下级用户都希望达到交易的平衡, 并委托第三方机构设计合适的电量供给方案, 并让第三方机构知晓存储设备的实时电量. 记发电社区为局中人 1, 而其下级用户为局中人 2, 因发电社区得到的报酬为下级用户的支出, 第三方机构自然考虑用零和博弈的方式设计方案. 不妨设时间间隔为一天, 若局中人 1 当日欲出售给下级用户 (通过第三方机构) 的电量为 a, 局中人 2 当天向发电社区可能购买的电量为 b, 则实际的交易量即为 $\min\{a,b\}$, 局中人 1 获得报酬 $r(x,a,b)$, 也是局中人 2 的支付费用. 此外, 由于可再生能源产生的电量和用户消费电量的不确定性, 社区第 t 天生产的电量用随机变量 ξ_t 表示, 自身消耗的电量用随机变量 η_t 表示. 另外, 对所有的 $t=0,1,\dots,$

- x_t 表示在第 t 天开始时设备存储的电量, M^* 表示设备最大存储电量, 则 $x_t \in [0,M^*] =: X$;
- a_t 表示局中人 1 在第 t 天可能出售给下级用户的电量, 其中 $a_t \in [0,x_t] =: A(x_t)$;
- b_t 表示局中人 2 在第 t 天当日向发电社区购买的可能电量, 其中 $b_t \in [0,x_t] =: B(x_t)$.

记 $z_t := \xi_t - \eta_t$, 假设 $\{z_t\}$ 是独立同分布的且与初始状态 x_0 独立, z_0 有连续的密度函数 g, 并记其分布函数为 G, 满足 $G(-M^*)>0$ 且 $\int_{-\infty}^{M^*} zg(dz) + M^*(1-G(0)) < 0$.

由此可以得到如下递推关系:

$$x_{t+1} = \min\{(x_t - \min\{a_t,b_t\} + z_t)^+, M^*\}, \quad \forall\, t=0,1,\dots,$$

其中 $y^+ := \max\{y,0\}$. 为描述交易间的费用, 用 s 表示单位电价, 若存储设备无电, 则需要对下级用户进行缺电补偿, 假设单位电量补偿价格为 c_0, 总的补偿限额为 C_1. 由此可得报酬函数

$$r(x,a,b) = s\cdot\min\{a,b\} - \min\{c_0\cdot(-x+\min\{a,b\}-Ez_0)^+, C_1\}.$$

命题 5.1 例 5.1 中的可再生能源发电供给模型存在平均 Nash 均衡策略.

证明 下面验证假设 3.1–3.3 条件成立. 首先, 验证假设 3.1. 对每一个状态 $x\in X$ 和任意的 $u\in\mathbb{B}(X)$, 根据状态转移方程, 可得

$$\begin{aligned}\int_X u(y)Q(dy\mid x,a,b) &= \int_{-\infty}^{+\infty} u(\min\{(x-\min\{a,b\}+z)^+, M^*\})g(z)dz\\ &= u(0)G(\min\{a,b\}-x) + u(M^*)[1-G(M^*+\min\{a,b\}-x)]\\ &\quad + \int_{\min\{a,b\}-x}^{M^*+\min\{a,b\}-x} u(x-\min\{a,b\}+z)g(z)dz\\ &= u(0)G(\min\{a,b\}-x) + u(M^*)[1-G(M^*+\min\{a,b\}-x)]\\ &\quad + \int_0^{M^*} u(t)g(t-x+\min\{a,b\})dt. \end{aligned} \tag{5.1}$$

由 g 的连续性可知 $\int_X u(y)Q(dy\mid x,a,b)$ 关于 (a,b) 在 $A(x)\times B(x)$ 上连续. 下面验证假设 3.2. 不难发现报酬函数有界, 仅需要取 $w_1(x)$ 为任意正常数即可.

下面验证假设 3.3. 根据命题 4.1 和 4.2, 仅需验证命题 4.2 的条件.

由于 $\int_{-\infty}^{M^*} zg(dz)+M^*(1-G(0))<0$, 所以存在 $x_0>0$ 及 $\delta<0$, 使得

$$\int_{-x_0}^{M^*} zg(dz)+M^*(1-G(0))<\delta.$$

则对所有的 $(\pi^1,\pi^2)\in\Pi_1^S\times\Pi_2^S$, 当 $x\geqslant x_0$ 时, 有

$$\begin{aligned}
\int yQ(dy\mid x,\pi^1,\pi^2)&=\int_{-\infty}^{+\infty}\int_0^x\int_0^x \min\{(x-\min\{a,b\}+z)^+,M^*\}\pi^1(da\mid x)\pi^2(db\mid x)g(z)dz\\
&\leqslant\int_{-\infty}^{+\infty}\min\{(x-\min\{0,0\}+z)^+,M^*\}g(z)dz\\
&=\int_{-x}^{M^*-x}(x+z)g(z)dz+M^*(1-G(M^*-x))\\
&\leqslant x+\int_{-x_0}^{M^*} zg(z)dz+M^*(1-G(0))\\
&<x+\delta.
\end{aligned}$$

因此, 由文献 [18, 定理 8.4.3], 仅需令 $V(x)=x$, $C=[0,x_0]$, 则有

$$\int V(y)Q(dy\mid x,\pi^1,\pi^2)-V(x)<\delta<0,$$

因此在 (π^1,π^2) 下, $\{x_t\}$ 为正常返的. 从而存在唯一的不变概率分布 μ_{π^1,π^2}.

接下来, 再逐一验证命题 4.2 的剩余条件.

(a) 令 $\nu(D)=1_D(0)$, $\forall\, D\in\mathcal{B}(X)$ 以及

$$l_{\pi^1,\pi^2}(x):=\int_0^x\int_0^x G(\min\{a,b\}-x)\pi^1(da\mid x)\pi^2(db\mid x).$$

则根据 (5.1), 令 $u(x)=1_B(x)$, 对于任意 Borel 集 $B\in\mathcal{B}(X)$, 都有

$$Q(B\mid x,\pi^1,\pi^2)\geqslant\int_0^x\int_0^x 1_B(0)G(\min\{a,b\}-x)\pi^1(da\mid x)\pi^2(db\mid x)=\nu(B)l_{\pi^1,\pi^2}(x),\quad\forall\, x\in X.$$

(b) 令 $\epsilon:=G(-M^*)>0$, 则有

$$\nu(l_{\pi^1,\pi^2})\geqslant G(-M^*)=\varepsilon,\quad\forall\,(\pi^1,\pi^2)\in\Pi_1^S\times\Pi_2^S.$$

(c) 因为 w_1 为常数, 根据命题 4.2(b) 的记号, 可得 $\nu(w_1)=w_1(0)<+\infty$;

(d) 令 $\gamma:=1-G(-M^*)<1$, 因为 w_1 为常数, 故有

$$\int_X w_1(y)Q(dy\mid x,\pi^1,\pi^2)=w_1(0)G(-M^*)+\gamma w_1(x)\leqslant\nu(w_1)l_{\pi^1,\pi^2}(x)+\gamma w_1(x).$$

由此证明了例子满足命题 4.2 的所有条件. 从而, 根据命题 4.1 及定理 3.1, 可知该博弈的值和平均 Nash 均衡策略均存在. □

下面, 给出保险分红的例子, 其状态转移方程及报酬函数的类型不同于上述例子, 同样能证明其能满足本文的假设条件.

例 5.2 (保险分红模型) 在金融市场中, 保险公司及其股东代表两方对公司资产进行管理决策. 股东代表决定当月的分红数额 (用 a 表示), 保险公司决定当月的投资金额 (用 b 表示). 股东代表获得月收益 (记为 $r(x,a,b)$, 即为保险公司的月支出). 保险公司还产生额外保费收入与索赔支出. 用 η_t 表示当月投资理财的单位收益, 假设恒为正, 用 ξ_t 表示当月保费收入与索赔发生支出的差. 对所有 $t=0,1,\dots,$

- x_t 表示第 t 个月开始时公司总资产, θ_1 表示公司最大资产, 则 $x_t\in[0,\theta_1]=:X$;
- a_t 表示股东代表在每月初决定用于分红的数额, $a_t\in[0,x_t]=:A(x_t)$;
- b_t 表示保险公司在每月初决定用于投资的数额, $b_t\in[0,x_t]=:B(x_t)$.

设 $\{\eta_t\}$ 和 $\{\xi_t\}$ 都是独立同分布的随机变量, 它们相互独立, 分别有相应的连续密度函数 g_1 和 g_2. 定义随机变量 $z_b:=b\eta_0+\xi_0$, 易得其密度函数

$$h_b(y)=\int_0^{+\infty}g_1(s)g_2(y-bs)ds,$$

并记其分布函数为 G_b. 假设其满足 $G_{\theta_1}(-\theta_1)>0$ 且

$$\int_{-\infty}^{\theta_1}zh_{\theta_1}(dz)+\theta_1(1-G_{\theta_1}(0))<0.$$

由上述模型的表述可得如下关系:

$$x_{t+1}=\min\{(x_t+b_t\eta_t-a_t+\xi_t)^+,\theta_1\},\quad \forall\, t=0,1,\dots,\tag{5.2}$$

以及如下报酬函数:

$$r(x,a,b)=s\cdot a-m\cdot\min\{(x+bE\eta_0-a+E\xi_0)^+,\theta_1\}-\min\{c_0\cdot(-x-bE\eta_0+a-E\xi_0)^+,C_2\},$$

其中 s 表示单位货币购买力, m 表示股东满意度, c_0 为资产为负时股东需要承担的单位借贷成本, 但总的借贷成本假设不超过 C_2.

命题 5.2 例 5.2 中的保险分红模型存在平均 Nash 均衡策略.

证明 下面验证假设 3.1–3.3. 首先验证假设 3.1.

对于任意 $u\in\mathbb{B}(X)$, 根据状态转移方程 (5.2), 可得

$$\begin{aligned}\int_X u(y)Q(dy\mid x,a,b)&=u(0)G_b(a-x)+u(\theta_1)[1-G_b(\theta_1+a-x)]+\int_{a-x}^{\theta_1+a-x}u(x-a+y)h_b(y)dy\\&=u(0)G_b(a-x)+u(\theta_1)[1-G_b(\theta_1+a-x)]+\int_0^{\theta_1}u(t)h_b(t+a-x)dt.\end{aligned}\tag{5.3}$$

由此可得上述函数关于 (a,b) 在 $A(x)\times B(x)$ 上连续.

接着验证假设 3.2 的条件. 显然报酬函数有界, 那么取 w_1 为任意正常数即可.

为验证假设 3.3, 根据命题 4.1 和 4.2, 仅需验证命题 4.2 的条件. 首先, 对所有的平稳策略对 $(\pi^1,\pi^2)\in\Pi_1^S\times\Pi_2^S$, 根据条件假设, 有 $E[\theta_1\eta_0+\xi_0]<0$; 根据文献 [18, 定理 8.4.3], 类似例 5.1 的证明, 可得 Markov 链 $\{x_t^{\pi^1,\pi^2}\}$ 是正常返的, 从而存在唯一的不变概率分布 μ_{π^1,π^2}.

接下来, 再逐一验证命题 4.2 的剩余条件.

(a) 令 $\nu(D)=1_D(0),\forall\, D\in\mathcal{B}(X)$ 以及

$$l_{\pi^1,\pi^2}(x):=\int_0^x\int_0^x G_b(a-x)\pi^1(da\mid x)\pi^2(db\mid x),$$

同样根据 (5.3), 对于任意 Borel 集 $B \in \mathcal{B}(X)$, 都有

$$Q(B|x,\pi^1,\pi^2) \geqslant \nu(B) l_{\pi^1,\pi^2}(x), \quad \forall x \in X;$$

(b)–(d) 取 $\varepsilon := G_{\theta_1}(-\theta_1) > 0$, $\gamma := 1 - G_{\theta_1}(-\theta_1) < 1$, 则类似于命题 5.1(b)–5.1(d) 的证明可证明结论成立.

此时, 根据命题 4.1 及定理 3.1, 可以证明博弈的值和平均 Nash 均衡策略存在. □

注 5.1　这两个例子的状态空间都是不可数的, 且不要求文献 [12,14,15,21] 中的 λ- 不可约性.

致谢　感谢审稿人以及所有给予过帮助的人.

参考文献

1 Altman E, Hordijk A, Spieksma F M. Contraction conditions for average and α-discount optimality in countable state Markov games with unbounded rewards. Math Oper Res, 1997, 22: 588–618
2 Borkar V S, Ghosh M K. Denumerable state stochastic games with limiting average payoff. J Optim Theory Appl, 1993, 76: 539–560
3 Buckdahn R, Cardaliaguet P, Rainer C. Nash equilibrium payoffs for nonzero-sum stochastic differential games. SIAM J Control Optim, 2004, 43: 624–642
4 Federgruen A. On N-person stochastic games by denumerable state space. Adv in Appl Probab, 1978, 10: 452–471
5 Guo X P, Hernández-Lerma O. Zero-sum games for continuous-time Markov chains with unbounded transition and average payoff rates. J Appl Probab, 2003, 40: 327–345
6 Guo X P, Hernández-Lerma O. Zero-sum continuous-time Markov games with unbounded transition and discounted payoff rates. Bernoulli, 2005, 11: 1009–1029
7 Guo X P, Hernández-Lerma O. Zero-sum games for continuous-time jump Markov processes in Polish spaces: Discounted payoffs. Adv in Appl Probab, 2007, 39: 645–668
8 Guo X P, Hernández-Lerma O. New optimality conditions for average-payoff continuous-time Markov games in Polish spaces. Sci China Math, 2011, 54: 793–816
9 Hamadéne S. Mixed zero-sum stochastic differential game and American game options. SIAM J Control Optim, 2006, 45: 496–518
10 Hernández-Lerma O, Lasserre J B. Discrete-Time Markov Control Processes: Basic Optimality Criteria. New York: Springer-Verlag, 1996
11 Hernández-Lerma O, Lasserre J B. Further Topics on Discrete-Time Markov Control Processes. New York: Springer-Verlag, 1999
12 Hernández-Lerma O, Lasserre J B. Zero-sum stochastic games in Borel spaces: Average payoff criteria. SIAM J Control Optim, 2000, 39: 1520–1539
13 Jaśkiewicz A. Zero-sum semi-Markov games. SIAM J Control Optim, 2002, 41: 723–739
14 Jaśkiewicz A, Nowak A S. On the optimality equation for zero-sum ergodic stochastic games. Math Methods Oper Res, 2001, 54: 291–301
15 Küenle H U, Schurath R. The optimality equation and ε-optimal strategies in Markov games with average reward criterion. Math Methods Oper Res, 2003, 56: 451–471
16 Lal A K, Sinha S. Zero-sum two-person semi-Markov games. J Appl Probab, 1992, 29: 56–72
17 Mertens J F, Neyman A. Stochastic games. Internat J Game Theory, 1981, 10: 53–66
18 Meyn S P, Tweedie R. Markov Chains and Stochastic Stability. New York: Springer-Verlag, 1993
19 Nowak A S. Measurable selection theorems for minimax stochastic optimization problems. SIAM J Control Optim, 1985, 23: 466–476
20 Nowak A S. Sensitive equilibria for ergodic stochastic games with countable state spaces. Math Methods Oper Res, 1999, 50: 65–76
21 Nowak A S. Optimal strategies in a class of zero-sum ergodic stochastic games. Math Methods Oper Res, 1999, 50: 399–419
22 Nowak A S. Some remarks on equilibria in semi-Markov games. Appl Math, 2000, 27: 385–394
23 Orey S. Limit Theorems for Markov Chain Transition Probabilities. London: van Nostrand, 1971
24 Połowczuk W. Nonzero-sum semi-Markov games with countable state spaces. Appl Math (Warsaw), 2000, 27: 395–402

25 Puterman M L. Markov Decision Processes: Discrete Stochastic Dynamic Programming. New York: John Wiley & Sons, 2014
26 Sennott L I. Zero-sum stochastic games with unbounded costs: Discounted and average cost cases. Z Oper Res, 1994, 39: 209–225
27 Vega-Amaya O. Zero-sum average semi-Markov games: Fixed-point solutions of the Shapley equation. SIAM J Control Optim, 2003, 42: 1876–1894
28 Yang J, Guo X P. Zero-sum stochastic games with average payoffs: New optimality conditions. Acta Math Sin (Engl Ser), 2009, 25: 1201–1216
29 Yeung D W K. Solution mechanisms for cooperative stochastic differential games. Int Game Theor Rev, 2006, 8: 309–326

New conditions for zero-sum stochastic games with average criteria in Borel space

Xianping Guo, Jinghao Liao, Ziqi Tan & Xin Wen

Abstract In this paper, we study the expected average criterion in discrete-time Markov games with Borel spaces. For the general case of unbounded reward functions, we first replace the corresponding Shapley equation for the average criterion with average-optimality two-inequalities. Then, by using the relative difference of the values of the discounted games, we give a new set of optimality conditions, which are weaker than the geometric ergodicity condition in the existing literature. Under these new conditions, we not only establish the solvability of the average-optimality two-inequalities but also show the existence of both the value and a Nash equilibrium of the game. Moreover, under the stronger geometric ergodicity condition, by the average-optimality two-inequalities, we also establish the solvability of the Shapley equation. Finally, we present two examples of renewable resources and financial insurance to verify the conditions and illustrate the results in this paper.

Keywords zero-sum average stochastic game, optimality conditions, average-optimality two-inequalities, Shapley equation, Nash equilibrium

MSC(2020) 91A15, 90D10, 60J10

doi: 10.1360/SSM-2024-0055