中国科学: 数学 2020 年 第 50 卷 第 1 期: 197～210
SCIENTIA SINICA Mathematica

论 文

《中国科学》杂志社
SCIENCE CHINA PRESS
CrossMark
click for updates

概率论中的最优耦合与博弈论中的最优合作

献给严士健教授 90 华诞

张绍义[1*], 张韧[2]

1. 湖北大学数学与统计学院, 武汉 430062;
2. 武汉华夏理工学院信息工程学院, 武汉 430223

E-mail: zhshaoyi@aliyun.com, 43862746@qq.com

收稿日期: 2018-06-15; 接受日期: 2018-10-23; 网络出版日期: 2019-11-08; * 通信作者
应用数学湖北省重点实验室资助项目

摘要 本文推广了二重最优耦合的概念, 得到结果 I: 设 X 和 Y 是 Polish 空间, $\varphi: X\times Y\to\mathbb{R}$ 可测, $\mu\in\mathscr{P}(X)$, $\nu\in\mathscr{P}(Y)$, (i) 如果 φ 是有下界的下半连续函数, 那么 φ 最优耦合 γ_φ 存在; (ii) 如果 φ 是有上界的上半连续函数, 那么 φ 上最优耦合 γ^φ 存在. 结果 II: 设 G_i $(i=1,2)$ 是从可测空间 $(\Omega_i,\mathscr{F}_i)$ 到 Polish 空间 $(X_i,\rho_i,\mathscr{B}(X_i))$ 上的转移概率测度序列, (i) 如果 $\varphi: X_1\times X_2\to\mathbb{R}$ 是有下界的下半连续函数, 则 G_1 和 G_2 的 φ 最优可测耦合存在; (ii) 如果 $\varphi: X_1\times X_2\to\mathbb{R}$ 是有上界的上半连续函数, 则 G_1 和 G_2 的 φ 上最优可测耦合存在. 本文提出一种带约束的 n 重最优耦合的概念并证明这种最优耦合的存在性, 由此定义了一种博弈论中的 Nash 均衡的最优合作均衡, 并举例说明这种新均衡优于 Nash 均衡.

关键词 二重最优耦合 带约束的 n 重最优耦合 Nash 均衡

MSC (2010) 主题分类 60A99, 91A12

1 引言

耦合方法在随机过程、随机场、无穷粒子 Markov 过程、流体动力学极限和扩散过程的第一特征值的估计等领域都有广泛应用 (参见文献 [1–3]). 尽管这些应用以各种不同的形式出现, 但基本上都是围绕一个中心问题, 就是 Polish 空间上的概率测度序列 (或概率簇) 的收敛问题. 例如, 随机场中的 Gibbs 态的存在性、离散时间 Markov 链和连续时随机过程的遍历性等问题.

设 $(X,\rho,\leqslant,\cdot)$ 是一个 Polish 空间, 并赋予度量、偏序和乘积运算.

第 2 节概述利用 ρ、$\leqslant$、$\cdot$ 和 $\mathscr{P}(X)$ 上的两个概率的耦合, 在 $\mathscr{P}(X)$ 上建立随机度量 $\omega_\rho: \mathscr{P}(X)\times\mathscr{P}(X)\to\mathbb{R}_+$、随机偏序 $\leqslant_{\mathrm{st}}$ 和随机乘积 (卷积) $*: \mathscr{P}(X)\times\mathscr{P}(X)\to\mathscr{P}(X)$, 以此帮助耦合方法的初学者认识耦合方法在概率论和基础数学中的重要意义. 特别地, 当 X 是一维 Euclid 空间 $\mathbb{R}$ 时, 在建

英文引用格式: Zhang S Y, Zhang R. Optimal coupling in probability theory and optimal cooperation in game theory (in Chinese). Sci Sin Math, 2020, 50: 197–210, doi: 10.1360/N012018-00161

立 ω_ρ 和 $\leqslant_{\mathrm{st}}$ 时, 自然地导出最优耦合的概念, 以此说明最优耦合的存在性的本质往往是要在概率集上建立某种数学框架.

第 3 节概述 Monge 提出的质量分布的最优运输问题及其弱问题: Kantorovich 问题最优解的存在性, 该问题本质上是 $\mathscr{P}(\mathbb{R}^2)$ 上的两个概率测度某种最优耦合的存在性, 因此, 该问题也是产生最优耦合的源头之一, 由此可知, 最优耦合的存在性在基础数学和应用数学中都有重要意义.

第 4 节的定理 4.1 和 4.2 推广先前类似的结果, 并给出详细的证明, 还对定理的证明的形成做了概述, 其目的有三. (1) 这两个定理比先前类似的结果更便于应用; (2) 定理 4.2 的证明本质上还是文献 [4] 中的定理 1.3 证明的思想, 专著 [1] 在证明定理 5.23 时, 对该定理作了简化证明, 但有两处含糊不清的地方 (笔者之一曾与文献 [1] 的作者讨论过该问题). 因此, 本节的证明既是对文献 [4] 中定理 1.3 证明的简化, 也是对文献 [1] 中定理 5.23 证明的更正. (3) 该定理的证明是作者通过多年的努力得出的结果, 有较高的技巧性, 通过对该定理及其证明的形成的概述, 使之成为一个在数学研究中 "先退后进"、"从有限认识可数无穷, 再认识不可数无穷" 研究方法的一个具体案例, 为做数学研究的新手提供参考.

第 5 节提出一种带约束的 n 重最优耦合的概念, 并证明这种最优耦合的存在性.

第 6 节利用第 5 节的结果, 定义一种 Nash 均衡的最优合作均衡, 并证明这种最优合作均衡的存在性, 给出例子说明这种最优合作均衡往往优于 Nash 均衡.

2 耦合与概率集上的数学结构

设 $(X,\rho,\mathscr{B}(X))$ 是完备度量空间, 并赋予 Borel σ 代数, 用 $\mathscr{P}(X)$ 表示 X 上的全体概率. 设 $\mu\in\mathscr{P}(X)$, f 是 X 上的可测函数, 下面总用记号 $\mu(f):=\int f d\mu$, $\{\mu_n\}_{n\geqslant 1}\subset\mathscr{P}(X)$, $\mu\in\mathscr{P}(X)$. 如果对 X 上的有界连续函数 f, 都有 $\lim_{n\to\infty}\mu_n(f)=\mu(f)$, 则称 $\{\mu_n\}_{n\geqslant 1}$ 弱收敛于 μ, 后面总记为 $\mu_n\to\mu$ $(n\to\infty)$, 在 $\mathscr{P}(X)$ 上赋予弱收敛拓扑后 $\mathscr{P}(X)$ 成为拓扑空间, 但仍记为 $\mathscr{P}(X)$. 本文只讨论这种拓扑, 因为这种拓扑是最弱的, 所以应用范围最广. 要判别 $\mu_n\to\mu$ 是否成立, 需要考虑 $\{\mu_n\}_{n\geqslant 1}$ 中元素之间的关系, 而 $\mathscr{P}(X)$ 中任何两个概率之间的 "关系" 都包含在它们的耦合之中.

定义 2.1 设 X 和 Y 是两个 Polish 可测空间, $X\times Y$ 为乘积 Polish 空间, π_X 和 π_Y 分别是从 $X\times Y$ 到 X 和 Y 的投影, 设 $\mu\in\mathscr{P}(X)$, $\nu\in\mathscr{P}(Y)$, $K(\mu,\nu)=:\{\gamma\in\mathscr{P}(X\times Y)\mid \gamma\pi_X^{-1}=\mu,\gamma\pi_Y^{-1}=\nu\}$, 则称 $K(\mu,\nu)$ 中的任一元素 γ 为 μ 与 ν 的耦合概率, 而 μ 与 ν 都称为 γ 的边缘概率.

在很多情形下, $Y=X$.

注意到判别实数序列 $\{a_n\}_{n\geqslant 1}$ 的收敛性主要有两个基本定理. (1) 若 $\{a_n\}_{n\geqslant 1}$ 单调有界, 则 $\{a_n\}_{n\geqslant 1}$ 收敛; (2) 若 $\{a_n\}$ 是 Cauchy 序列, 则 $\{a_n\}_{n\geqslant 1}$ 收敛. 这是因为实数空间上有序结构和度量 (或拓扑) 结构. 常常用 $\mathbb{R}\times\mathbb{R}$ 上的实函数来定义这两种结构, 对 $(x,y)\in\mathbb{R}\times\mathbb{R}$, 定义实映射 $s(x,y)=I_{\{x>y\}}(x,y)$, 则利用 s, 在 $\mathbb{R}$ 上定义序关系 "$\leqslant$" 为 $x\leqslant y\Leftrightarrow s(x,y)=0$; 定义实映射 $\rho(x,y)=|x-y|$, 则 $\rho(x,y)$ 是 x 与 y 的距离或度量. 设 $\{\mu_n\}_{n\geqslant 1}\subset\mathscr{P}(\mathbb{R})$, 自然想在 $\mathscr{P}(\mathbb{R})$ 建立序结构和度量结构来判别 $\{\mu_n\}_{n\geqslant 1}$ 的收敛性. 虽然有几种方式来定义 $\mathscr{P}(\mathbb{R})$ 上的序结构和度量结构, 但最自然的方式是类似于定义 $\mathbb{R}$ 上的相关结构, 用 $\mathscr{P}(\mathbb{R})\times\mathscr{P}(\mathbb{R})$ 上的实映射来定义 $\mathscr{P}(\mathbb{R})$ 上的相关结构.

对任意给定 $\mu,\nu\in\mathscr{P}(\mathbb{R})$, 先选一个 "好的" $\gamma_s\in K(\mu,\nu)$, 令 $\omega_s(\mu,\nu)=\gamma_s(s)$. 然后定义 $\mathscr{P}(\mathbb{R})$ 的随机序关系 "$\leqslant_{\mathrm{st}}$" 为 $\mu\leqslant_{\mathrm{st}}\nu\Leftrightarrow\omega_s(\mu,\nu)=0$. 类似的理由, 可选一个 "好的" $\gamma_\rho\in K(\mu,\nu)$. 用 $\omega_\rho(\mu,\nu)=\gamma_\rho(\rho)$ 表示 μ 与 ν 的距离. 问题是怎样选取 "好的" γ_s (γ_ρ)? 考虑 γ_s' $(\gamma_\rho')\in K(\mu,\nu)$, 得

到 $\omega'_s(\mu,\nu)\neq\gamma'_s(s)$ $(\omega'_\rho(\mu,\nu)\neq\gamma'_\rho(\rho))$, 且满足 $\omega_s(\mu,\nu)\leqslant\omega'_s(\mu,\nu)$ $(\omega_\rho(\mu,\nu)\leqslant\omega'_\rho(\mu,\nu))$. 按 $\omega'_s(\mu,\nu)$ 定义的随机关系为 "$\leqslant'_{\rm st}$", 那么 $\{\mu_n\}_{n\geqslant1}$ 若按 "$\leqslant'_{\rm st}$" 单调, 则一定按 "$\leqslant_{\rm st}$" 单调. 另外, 若 $\{\mu_n\}$ 按度量 $\omega'_\rho(\mu,\nu)$ 收敛, 则一定按度量 $\omega_\rho(\mu,\nu)$ 收敛. 由此启发, $\gamma_s(\gamma_\rho)$ 的选取应该是使 $\gamma_s(s)$ $(\gamma_\rho(\rho))$ 越小越好, 于是选择 γ_s 和 γ_ρ 分别为

$$\gamma_s(s)=\omega_s(\mu,\nu)=\inf\{\gamma(s)\mid\gamma\in K(\mu,\nu)\},$$
$$\gamma_\rho(\rho)=\omega_\rho(\mu,\nu)=\inf\{\gamma(\rho)\mid\gamma\in K(\mu,\nu)\}.$$

当然, γ_s 和 γ_ρ 的存在性需要证明, 见后面定理 4.1(i).

用 $\gamma_s(s)$ 和 $\gamma_\rho(\rho)$ 分别在 $\mathscr{P}(\mathbb{R})$ 上建立偏序结构 "$\leqslant_{\rm st}$" 和度量结构 ω_ρ (这两个结论也需证明, 本文略去证明, 可参见文献 [1,5]) 后, 则 $(\mathscr{P}_0(\mathbb{R}),\omega_\rho,\leqslant_{\rm st})$ 是一个具有偏序的完备度量空间, 且其上的序列按 ω_ρ 收敛等价于弱收敛. $\mathscr{P}_0(\mathbb{R})=\{\mu\in\mathscr{P}(\mathbb{R})\mid\mu(|x|)<\infty\}$, 且有下面关于 $\{\mu_n\}$ 收敛的判定准则:

(i) 若 $\{\mu_n\}_{n\geqslant1}\subset\mathscr{P}_0(X)$ 且单调 (即 $\mu_n\leqslant_{\rm st}\mu_{n+1}$, $n=1,2,\ldots$) 概率有界 (即 $\forall\varepsilon>0$, 存在紧集 K, 使 $\inf_{n\geqslant1}\{\mu_n(K)\}\geqslant1-\varepsilon$), 那么 $\{\mu_n\}_{n\geqslant1}$ 收敛.

(ii) 若 $\{\mu_n\}_{n\geqslant1}$ 是 $\mathscr{P}_0(\mathbb{R})$ 上的 Cauchy 序列, 则 $\{\mu_n\}_{n\geqslant1}$ 收敛.

设 $(X,\rho,\leqslant)$ 是具有闭偏序的 Polish 空间, 上述耦合方法也可在 X 的概率集 $\mathscr{P}(X)$ 的子集 $\mathscr{P}_0(X):=\{\mu\in\mathscr{P}(X)\mid\int\rho(x,x_0)\mu(dx)<\infty\}$ 上定义序结构和度量结构并建立在其上的概率测度序列收敛判别法则 (参见文献 [1,5]). 由此可知, 耦合方法在概率论得到广泛的应用是很自然的.

若拓扑空间 X 有一种乘法运算 "$\cdot$" 的代数运算, 使得 X 具有群 (或半群) 的代数结构, 且运算 "$\cdot$" 是可测的, 即 $\forall B\in\mathscr{B}(X)$, 集合 $\{(x,y)\in X\times X\mid x\cdot y\in B\}\in\mathscr{B}(X)\times\mathscr{B}(X)$, 则可利用 "$\cdot$" 和耦合在 $\mathscr{P}(X)$ 建立的一种卷积运算 "$*$", 使得 $\mathscr{P}(X)$ 具有代数结构群 (或半群).

设 $x\in X$, $A\subset X$. 记 $x^{-1}A=\{y\in X,x\cdot y\in A\}$, 又设 $\mu,\nu\in\mathscr{P}(X)$, 令 $\gamma(dx,dy)=\mu(dx)\times\nu(dy)$ 是 μ 与 ν 的独立耦合. 定义 μ 与 ν 的卷积运算 "$*$" 为

$$\mu*\nu(B)=\iint I_B(x\cdot y)\gamma(dx,dy)=\int\nu(x^{-1}B)\mu(dx),\quad\forall B\in\mathscr{B}(X)),$$

那么 $\mu*\nu\in\mathscr{P}(X)$, 则 "$*$" 建立了 $\mathscr{P}(X)$ 上的群结构 (或半群结构), 可参见文献 [6].

3 耦合与最优运输问题

前一节简要讨论了耦合方法在概率论的基本理论中的作用和思想起源, 本节简要讨论耦合方法在社会科学和经济学等应用领域的思想起源.

1781 年, 法国数学家 Monge 提出了物质分布的最优运输问题.

设 $h_0,h_1:\mathbb{R}^2\to\mathbb{R}^2$ 是可测映射, h_0 表示某种单位质量 $(\int h_0(x)dx=1)$ 在 $\mathbb{R}^2$ 的初始分布密度. 通过某种方式的运输 $(\psi:\mathbb{R}^2\to\mathbb{R}^2)$, 使该种物质的终极分布密度为 h_1, 由于 ψ 表示物质的运输, 它应该满足

$$\int_{\psi^{-1}(B)}h_0(x)dx=\int_B h_1(y)dy,\quad\forall B\in\mathscr{B}(\mathbb{R}^2).\tag{3.1}$$

用 $\mathscr{M}$ 表示使 (3.1) 成立的全体运输, 设 $\psi\in\mathscr{M}$, 记 $V(\psi)=\int\|\psi(x)-x\|_2h_0(x)dx$ ($\|\cdot\|_2$ 是 $\mathbb{R}^2$ 上的 Euclid 范数), Monge 优化问题就是

$$\inf\{V(\psi)\mid\psi\in\mathscr{M}\},\tag{3.2}$$

即在 $\mathscr{M}$ 中寻找 ψ_0, 使 $V(\psi_0) = \inf\{V(\psi) \mid \psi \in \mathscr{M}\}$, ψ_0 称为最优解. 若记 $\mu(dx) = h_0(x)dx$, $\nu(dx) = h_1(x)dx$, $\gamma_\psi(dx,dy) = h_0(x)I_{\{(x,\psi(x))\}}(x,y)dxdy$ ($\psi \in \mathscr{M}$), 则 $\mu,\nu \in \mathscr{P}(\mathbb{R}^2)$, $\gamma_\psi \in \mathscr{P}(\mathbb{R}^2 \times \mathbb{R}^2)$, 且由 (3.1) 可知, $\gamma_\psi \in K(\mu,\nu)$, 又记 $\rho(x,y) = \|x-y\|_2$ ($x,y \in \mathbb{R}^2$). 此时, (3.2) 可化为

$$\inf\{\gamma_\psi(\rho) \mid \psi \in \mathscr{M}\}. \tag{3.3}$$

一般说来, (3.3) 的最优解可能不存在. 1942 年, 苏联数学家 Kantorovich 提出了 Monge 问题弱解的优化问题

$$\min\{\gamma(\rho) \mid \gamma \in K(\mu,\nu)\}. \tag{3.4}$$

与 (3.3) 相比, (3.4) 只是把求极值点的范围从 $\{\gamma_\psi \in K(\mu,\nu) \mid \psi \in \mathscr{M}\}$ 扩大到 $K(\mu,\nu)$. 但令人兴奋的是 (3.4) 的最优解总是存在的. 我们可以通过一个简单的例子来说明两者之间求解的差别. 设

$$\mu(dx) = \delta_{(0,0)}(dx), \quad \nu(dy) = \frac{1}{2}\delta_{(0,1)}(dy) + \frac{1}{2}\delta_{(1,0)}(dy),$$

即物质的初始分布是将物质全部存放在 $(0,0)$ 处, 终极分布是在 $(0,1)$ 和 $(1,0)$ 处各自存放 $\frac{1}{2}$ 的物质. 由于对任意运输 $\psi: \mathbb{R}^2 \to \mathbb{R}^2$ 都不能满足 $\nu(\{(0,1)\}) = \mu(\{(0,0)\})$ (因为 $\frac{1}{2} = \nu(\{(0,1)\}) < 1 = \mu(\{(0,0)\})$), 故 (3.3) 是在一个空集中求极值点, 因此, (3.3) 当然不存在最优解. 但我们可在 $\mathbb{R}^2 \times \mathbb{R}^2$ 上构造 γ 使

$$\gamma(\{(0,0),(0,1)\}) = \frac{1}{2}, \quad \gamma(\{(0,0),(1,0)\}) = \frac{1}{2}.$$

易见 $\gamma \in K(\mu,\nu)$, 且 γ 是 (3.4) 的最优解. 究其原因, 是在问题 (3.3) 中, 由于受 (3.1) 的限制, $(0,0)$ 处的物质不能被打散运输, 而在问题 (3.4) 中, $(0,0)$ 处的物质没有 (3.1) 的限制, 因此可将 $(0,0)$ 处的物质分成两包, 每包的质量分别为 $\frac{1}{2}$, 然后分别送往 $(0,1)$ 和 $(1,0)$ 处.

Kantorovich 问题 (3.4) 可以推广. 设 X 和 Y 是两个 Polish 空间. $C: X \times Y \to \mathbb{R}$ 可测, 设 $\mu \in \mathscr{P}(X)$, $\nu \in \mathscr{P}(Y)$, 问题 (3.4) 推广为

$$\min\{\gamma(C) \mid \gamma \in K(\mu,\nu)\}. \tag{3.5}$$

C 称为运输费用函数, (3.5) 的解称为关于费用函数 C 的最优运输.

在问题 (3.5) 中, 如果用效益 (或利润) 函数 $u(x,y)$ 替换 $C(x,y)$, $u(x,y)$ 表示将 x 地的某种商品销往 y 地所产生的效益 (或利润), 人们希望效益越大越好, 因此出现问题

$$\max\{\gamma(u) \mid \gamma \in K(\mu,\nu)\}. \tag{3.6}$$

后面称 (3.6) 为最优销售问题, u 称为效益函数. 问题 (3.6) 的解称为关于效益函数的最优销售. 下一节将讨论这种最优解的存在性.

4 最优可测耦合的存在性

第 2 节要处理 γ_s 和 γ_ρ 的存在性问题; 第 3 节要解决问题 (3.5) 和 (3.6) 最优解的存在性问题. 本节来统一处理这些问题.

定义 4.1 设 X 和 Y 是两个 Polish 空间, $\varphi: X\times Y\to\mathbb{R}$ 可测, $\mu\in\mathscr{P}(X)$, $\nu\in\mathscr{P}(Y)$.

(i) 如果存在 $\gamma_\varphi\in K(\mu,\nu)$ 使得

$$\gamma_\varphi(\varphi)=\omega_\varphi(\mu,\nu):=\inf\{\gamma(\varphi)\mid\gamma\in K(\mu,\nu)\},\tag{4.1}$$

则称 γ_φ 是 (关于 μ 与 ν 的) φ 最优耦合.

(ii) 如果存在 $\gamma^\varphi\in K(\mu,\nu)$ 使得

$$\gamma^\varphi(\varphi)=\omega^\varphi(\mu,\nu):=\sup\{\gamma(\varphi)\mid\gamma\in K(\mu,\nu)\},\tag{4.2}$$

则称 γ^φ 是 (关于 μ 与 ν 的) φ 上最优耦合.

定理 4.1 设 X 和 Y 是 Polish 空间, $\varphi: X\times Y\to\mathbb{R}$ 可测, $\mu\in\mathscr{P}(X)$, $\nu\in\mathscr{P}(Y)$.

(i) 如果 φ 是有下界的下半连续函数, 那么 φ 最优耦合 γ_φ 存在;

(ii) 如果 φ 是有上界的上半连续函数, 那么 φ 上最优耦合 γ^φ 存在.

定理 4.1 是一个很一般的定理, 注意到在第 1 节中, s 是非负下半连续函数, ρ 是非负连续函数. 因此, 由于定理 4.1(i) 中 γ_s 和 γ_ρ 存在, 第 3 节中运输问题的弱问题的原始问题 (3.4) 中的 ρ 也是非负连续的, 故由定理 4.1 知, 问题 (3.4) 的最优解存在, 而且定理 4.1 还给出了问题 (3.5) 和 (3.6) 存在最优解的非常一般的充分条件, 这种最优耦合的存在性, 在理论上起着重要的作用. 例如, 要证明 $(\mathscr{P}_0(X),\omega_\rho)$ 是完备的度量空间, 需要 γ_ρ 的存在性来证明 ω_ρ 满足三角不等式 (参见文献 [1, 第 174 页引理 5.3]), 要证明 $\mathscr{P}_0(X)$ 上的单调且概率有界序列收敛, 也需 γ_s 的存在性 (参见文献 [5, 第 131–139 页]). 此外, 定理 4.1 与现有的有关结果相比更便于应用. 它不但把原有对 φ 的非负性限制换成了较宽松的有下界的限制, 而且既可处理最小化问题, 也可处理最大化问题.

虽然定理 4.1 在概率理论中有很多应用, 但在处理随机场、离散时间 Markov 链和跳过程等理论问题中需要比定理 4.1 更加深刻的结果: 转移概率测度的最优可测耦合 (有时也称 Markov 耦合) 的存在性. 例如, 随机场理论中的 Dobrushin-Shlosman 唯一定理 (参见文献 [1, 第 391 页定理 10.9])、跳过程最优耦合算子存在定理 (参见文献 [7])、跳过程保序耦合存在定理 (参见文献 [8]) 及较近的 Markov 链保序关系可测耦合定理 (参见文献 [9, 定理 3.3]).

定义 4.2 设 G_i $(i=1,2)$ 是从可测空间 $(\Omega_i,\mathscr{F}_i)$ 到 Polish 空间 $(X_i,\mathscr{B}(X_i))$ 上的转移概率测度, $\varphi: X_1\times X_2\to\mathbb{R}$ 可测. 如果存在从 $(\Omega_1\times\Omega_2,\mathscr{F}_1\times\mathscr{F}_2)$ 到 $(X_1\times X_2,\mathscr{B}(X_1\times X_2))$ 转移概率测度 G, 使得 $\forall(\omega_1,\omega_2)\in\Omega_1\times\Omega_2$, 有

$$G(\omega_1,\omega_2,\cdot)\in K(G_1(\omega_1,\cdot),G_2(\omega_2,\cdot)).\tag{4.3}$$

(i) 如果 $\forall(\omega_1,\omega_2)\in\Omega_1\times\Omega_2$, 有

$$G(\omega_1,\omega_2,\varphi)=\omega_\varphi(G_1(\omega_1,\cdot),G_2(\omega_2,\cdot)),\tag{4.4}$$

则称 G 是 G_1 与 G_2 的 φ 最优可测耦合.

(ii) 如果 $\forall(\omega_1,\omega_2)\in\Omega_1\times\Omega_2$, 有

$$G(\omega_1,\omega_2,\varphi)=\omega^\varphi(G_1(\omega_1,\cdot),G_2(\omega_2,\cdot)),\tag{4.5}$$

则称 G 是 G_1 与 G_2 的 φ 上最优可测耦合.

定理 4.2 设 G_i $(i=1,2)$ 是从可测空间 $(\Omega_i,\mathscr{F}_i)$ 到 Polish 空间 $(X_i,\rho_i,\mathscr{B}(X_i))$ 上的转移概率测度序列, 则

(i) 如果 $\varphi: X_1\times X_2\to\mathbb{R}$ 是有下界的下半连续函数, 则 G_1 与 G_2 的 φ 最优可测耦合存在;

(ii) 如果 $\varphi: X_1\times X_2\to\mathbb{R}$ 是有上界的上半连续函数, 则 G_1 与 G_2 的 φ 上最优可测耦合存在.

定理 4.2 是很一般的定理, 目前所碰到的最优 Markov (或可测) 耦合定理的存在定理都是它的特例. 例如, 2010 年, 文献 [9] 花了很大的篇幅证明文献中的定理 3.3 (转移概率保 "关系" 可测耦合的存在性) 就是 2000 年文献 [10, 定理 1.1] 的特例, 而该定理又是定理 4.2 的特例. 此外, 同定理 4.1 评注一样, 定理 4.2 比现有类似的结果更便于应用. 虽然如此, 但从下面的最优耦合性的定理 (定理 4.3) 可看出, 定理 4.1 和 4.2 各自的 (i) 和 (ii) 的证明本质上是相同的.

定理 4.3 设 X 和 Y 是两个 Polish 空间, $\mu\in\mathscr{P}(X)$, $\nu\in\mathscr{P}(Y)$.

(i) 设 $\varphi: X\times Y\to\mathbb{R}$ 可测. 如果 $-\varphi$ 最优耦合 $\gamma_{-\varphi}$ $(\in K(\mu,\nu))$ 存在, 那么 φ 上最优可测耦合 γ^{φ} 存在, 且 $\gamma^{\varphi}=\gamma_{-\varphi}$.

(ii) 设 $\varphi,\psi: X\times Y\to\mathbb{R}$ 可测, 且存在实数 $a>0$ 和实数 c, 使 $\psi=a\varphi+c$, 如果 φ 最优耦合 (φ 上最优耦合) $\gamma_{\varphi}(\gamma^{\varphi})$ 存在, 那么 ψ 最优耦合 (ψ 上最优耦合) $\gamma_{\psi}(\gamma^{\psi})$ 存在, 且 $\gamma_{\psi}=\gamma_{\varphi}$ $(\gamma^{\psi}=\gamma^{\varphi})$.

证明 (i) 如果 $\gamma_{-\varphi}$ 存在, 由定义 4.1 知, $\forall\gamma\in K(\mu,\nu)$, 有 $\gamma_{-\varphi}(-\varphi)\leqslant\gamma(-\varphi)$, 等价地, $\gamma_{-\varphi}(\varphi)\geqslant\gamma(\varphi)$. 由 γ 的任意性知, $\gamma^{\varphi}=\gamma_{-\varphi}$.

(ii) 由结论 (i), 只需证关于 φ 最优耦合的相关结果成立即可. 如果 γ_{ψ} 存在, 那么, 对任意的 $\gamma\in K(\mu,\nu)$, 有 $\gamma_{\varphi}(\varphi)\leqslant\gamma(\varphi)$, 由此可知,

$$\gamma_{\varphi}(\psi)=\gamma_{\varphi}(a\varphi+c)=a\gamma_{\varphi}(\varphi)+c\leqslant a\gamma(\varphi)+c=\gamma(\psi).$$

由 γ 的任意性即知, $\gamma_{\psi}=\gamma_{\varphi}$. □

命题 4.1 定理 4.1(i) 与 4.1(ii) 等价.

证明 若定理 4.1(ii) 的条件成立, 即 φ 是有上界的上半连续函数, 故 $-\varphi$ 是有下界的下半连续函数, 由定理 4.3(i) 知, γ^{φ} $(=\gamma_{-\varphi})$ 存在, 即证明了 (i)⇒(ii), 类似地可证 (ii)⇒(i). □

命题 4.2 定理 4.2(i) 与 4.2(ii) 等价.

证明 注意到定义 4.2 中关于两个转移概率测度的最优耦合的最优性是逐点定义的 (参见 (4.3)–(4.5)), 那么逐点应用命题 4.1 即知命题 4.2 成立. □

注意到定理 4.1 是定理 4.2 的特例, 由此及命题 4.2 知, 为证定理 4.1 和 4.2, 只需证定理 4.2(i) 即可. 为证定理 4.2, 先给出下面引理.

引理 4.1 在定理 4.2 的假设下, 由于 X_i $(i=1,2)$ 是可分的, 对每个 $n\geqslant1$, X_i 存在可数的可测划分 $\Delta_i^{(n)}=\{B_{ik}^{(n)}\}_{k\geqslant1}$, 使得 $\mathrm{diam}B_{ik}^{(n)}:=\sup\{\rho_i(x_i,y_i)\mid(x_i,y_i)\in B_{ik}^{(n)}\}<\frac{1}{n}$.

(i) 当 $\Omega_1=\Omega_2=X_1=X_2=X$, $G_1=G_2=P$ 且概率核 $\{P(x,dy)\mid x\in X\}$ 是一致胎紧的, 此时简化 $B_{ik}^{(n)}=B_k^{(n)}$ $(i=1,2)$, 那么存在 m_n, 令 $B_0^{(n)}=\bigcup_{k=m_n+1}B_k^{(n)}$, 取 $x_k^{(n)}\in B_k^{(n)}$, $k=0,1,\dots,m_n$, 那么有

$$P(x,B_0^{(n)})<\frac{1}{n},\quad x\in X,\quad \mathrm{diam}B_k^{(n)}<\frac{1}{n},\quad k=1,\dots,m_n.$$

令 $P_n(x,\{x_k^{(n)}\})=P(x,B_k^{(n)})$, $k=0,1,\dots,m_n$, 则 $\{P_n: n\geqslant1\}$ 是支撑为有限的概率核序列, 且 $\forall x\in X$, 有 $P_n(x,\cdot)\to P(x,\cdot)$ $(n\to\infty)$.

(ii) 在 (i) 的条件下 (但 $\{P(x,dy)\mid x\in X\}$ 没有一致胎紧性), 取 $x_k^{(n)}\in B_k^{(n)}$, $k=1,2,\ldots$; 给定 $n\geqslant 1$, 令 $P_n(x,\{x_k^{(n)}\})=P(x,B_k^{(n)})$, $k=1,2,\ldots$ 给定 $m\geqslant 1$, $n\geqslant 1$, 令

$$
\begin{aligned}
&P_{(n,m)}(x,\{x_k^{(n)}\})=P(x,B_k^{(n)}),\quad k=1,2,\ldots,m,\\
&P_{(n,m)}(x,\{x_{(m+1)}^{(n)}\})=P\bigg(x,\bigcup_{k=m+1}B_k^{(n)}\bigg),
\end{aligned}
$$

则 $\{P_{(n,m)}: n,m\geqslant 1\}$ ($\{P_n: n\geqslant 1\}$) 是支撑为有限 (可数) 集的概率测度序列, 且 $\forall x\in X$, 有 $P_{(n,m)}(x,\cdot)\to P_n(x,\cdot)$ $(m\to\infty)$ 和 $P_n(x,\cdot)\to P(x,\cdot)$ $(n\to\infty)$.

(iii) 取 $x_{ik}^{(n)}\in B_{ik}^{(n)}$, $i=1,2$, $k=1,2,\ldots$

(a) 给定 $n\geqslant 1$, 令 $G_i^{(n)}(\omega_i,\{x_{ik}^{(n)}\})=G_i(\omega_i,B_{ik}^{(n)})$, $\omega_i\in\Omega_i$, $i=1,2$, $k=1,2,\ldots$ 给定 $m\geqslant 1$, $n\geqslant 1$, 令

$$
\begin{aligned}
&G_i^{(n,m)}(\omega_i,\{x_{ik}^{(n)}\})=G_i(\omega_i,B_{ik}^{(n)}),\quad k=1,2,\ldots,m,\\
&G_i(\omega_i,x_{i(m+1)}^{(n)})=G_i\bigg(\omega_i,\bigcup_{k=m+1}^{\infty}B_{ik}\bigg),\quad \omega_i\in\Omega_i,\quad k=m+1,
\end{aligned}
$$

则 $\{G_i^{(n,m)}: n,m\geqslant 1\}$ ($\{G_i^{(n)}: n\geqslant 1\}$) 是支撑为有限 (可数) 集的转移概率测度序列, 且 $\forall\omega_i\in\Omega_i$, 有 $G_i^{(n,m)}(\omega_i,\cdot)\to G_i^{(n)}(\omega_i,\cdot)$ $(m\to\infty)$, $G_i^{(n)}(\omega_i,\cdot)\to G_i(\omega_i,\cdot)$ $(n\to\infty)$.

(b) 设 $\mu_i\in\mathscr{P}(X_i)$, $i=1,2$, $\mu\in K(\mu_1,\mu_2)$. 令

$$
\begin{aligned}
&\mu_i^{(n)}(\{x_{ik}^{(n)}\})=\mu_i(B_{ik}^{(n)}),\quad k=1,2,\ldots,\quad i=1,2,\\
&\mu_i^{(n,m)}(\{x_{ik}^{(n)}\})=\mu_i(B_{ik}^{(n)}),\quad k=1,2,\ldots,m,\quad i=1,2,\\
&\mu_i^{(n,m)}(\{x_{i(m+1)}^{(n)}\})=\mu_i\bigg(\bigcup_{k=m+1}^{\infty}B_{ik}^{(n)}\bigg),\\
&\mu^{(n)}(\{(x_{1k}^{(n)},x_{2j}^{(n)})\})=\mu(B_{1k}^{(n)}\times B_{2j}^{(n)}),\quad k,j=1,2,\ldots,\\
&\mu^{(n,m)}(\{(x_{1k}^{(n)},x_{2j}^{(n)})\})=\mu(B_{1k}^{(n)}\times B_{2j}^{(n)}),\quad k,j=1,2,\ldots,m,\\
&\mu^{(n,m)}(\{(x_{1k}^{(n)},x_{2(m+1)}^{(n)})\})=\mu\bigg(B_{1k}^{(n)}\times\bigcup_{j=m+1}^{\infty}B_{2j}\bigg),\quad k=1,2,\ldots,m,\\
&\mu^{(n,m)}(\{(x_{1(m+1)}^{(n)},x_{2j}^{(n)})\})=\mu\bigg(\bigg(\bigcup_{k=m+1}^{\infty}B_{1k}^{(n)}\bigg)\times B_{2j}^{(n)}\bigg),\quad j=1,2,\ldots,m,\\
&\mu^{(n,m)}(\{(x_{1(m+1)}^{(n)},x_{2(m+1)}^{(n)})\})=\mu\bigg(\bigg(\bigcup_{k=m+1}^{\infty}B_{1k}^{(n)}\bigg)\times\bigg(\bigcup_{j=m+1}^{\infty}B_{2j}^{(n)}\bigg)\bigg),\quad j=1,2,\ldots,m,
\end{aligned}
$$

则 $\mu^{(n,m)}\to\mu^{(n)}$ $(m\to\infty)$, $\mu^{(n)}\to\mu$ $(n\to\infty)$, 且 $\mu^{(n,m)}\in K(\mu_1^{(n,m)},\mu_2^{(n,m)})$, $\mu^{(n)}\in K(\mu_1^{(n)},\mu_2^{(n)})$.

证明 可直接验证, 也可参见文献 [10–12]. □

引理 4.2 设 $\{G_n\}_{n\geqslant 1}$ 是从可测空间 $(\Omega,\mathscr{F})$ 到 Polish 空间 X 上的转移测度序列, 假设对每个 $\omega\in\Omega$, 都有

(i) $\{G_n(\omega,\cdot)\}_{n\geqslant 1}$ 是一致胎紧的;

(ii) $\sup_{n\geqslant 1}G_n(\omega,\cdot)<\infty$,

则存在从 $(\Omega,\mathscr{F})$ 到 X 上的转移概率测度 G, 对每一个 $\omega\in\Omega$, 存在子列 $\{n(\omega)\}\subset\{n\}$, 使得 $G_{n(\omega)}(\omega,\cdot)\to G(\omega,\cdot)$ $(n(\omega)\to\infty)$.

引理 4.2 的证明参见文献 [10].

引理 4.3 设 $\varphi: X_1 \times X_2 \to \mathbb{R}$ 下半连续, 且 $0 \leqslant \varphi \leqslant 1$, 则存在有界连续函数族 $\{\varphi_m^{(n)}\}_{m\geqslant 1}^{n\geqslant 1}$ 和有界下半连续函数 φ_m, 使得 $\varphi_m^{(n)} \nearrow \varphi_m$ $(n \nearrow \infty)$ 和 $\varphi_m \nearrow \varphi$ $(m \nearrow \infty)$.

证明 对 $m \geqslant 1$, 令 $\varphi_m = \frac{1}{m}\sum_{k=1}^{m-1} I_{O_{mk}}$ (其中 $O_{mk} = \{\varphi > \frac{k}{m}\}$ 是开集), 则 $\varphi_m \nearrow \varphi$ $(m \nearrow \infty)$, 且 φ_m 是下半连续的. 又由文献 [1, 定理 1.36] 可知, 对 $1 \leqslant k \leqslant m-1$, 存在有界 Lipschitz 连续函数 $\{\varphi_{mk}^{(n)}\}_{n\geqslant 1}$, 使得 $\varphi_{mk}^{(n)} \nearrow I_{O_{mk}}$ $(n \nearrow \infty)$. 令 $\varphi_m^{(n)} = \frac{1}{m}\sum_{k=1}^{m-1}\varphi_{mk}^{(n)}$, 则 $\varphi_m^{(n)} \nearrow \varphi_m$ $(n \nearrow \infty)$. □

引理 4.4 (i) 设 $\mu_n, \mu \in \mathscr{P}(X)$, $\mu_n \to \mu$; $\nu_n, \nu \in \mathscr{P}(Y)$, $\nu_n \to \nu$, $\gamma_n \in K(\mu_n, \nu_n)$, 且 $\gamma_n \to \gamma$ $(n \to \infty)$, 则 $\gamma \in K(\mu, \nu)$.

(ii) 设 $\varphi_n : X \times Y \to \mathbb{R}$ 有界非负下半连续, $\varphi : X \times Y \to \mathbb{R}$ 非负下半连续, 且 $\varphi_n \nearrow \varphi$ $(n \nearrow \infty)$. 设 $\mu \in \mathscr{P}(X)$, $\nu \in \mathscr{P}(Y)$, 如果 γ_{φ_n} 是 φ_n 最优耦合, 且 $\gamma_{\varphi_n} \to \gamma_\varphi$ $(n \to \infty)$, 则 γ_φ 是 φ 最优耦合.

证明 (i) 设 f 是 X 上的有界连续函数, 则 $\pi_X^{-1} f$ 是 $X \times Y$ 上的有界连续函数. 由于 $\mu_n(f) = (\gamma_n \pi_X^{-1})(f) = \gamma_n(\pi_X^{-1} f)$, 令 $n \to \infty$, 则有 $\mu(f) = \gamma(\pi_X^{-1} f) = (\gamma\pi_X^{-1})(f)$. 由 f 的任意性有 $\mu = \gamma\pi_X^{-1}$. 同理有 $\nu = \gamma\pi_Y^{-1}$, 即 $\gamma \in K(\mu, \nu)$.

(ii) 对任意的 $\gamma \in K(\mu, \nu)$, 由题设知, $\gamma_{\varphi_n}(\varphi_n) \leqslant \gamma(\varphi_n)$, 由此及 φ_n 单调上升, 对固定的 m, 由文献 [1, 定理 4.5] 知,

$$\gamma_\varphi(\varphi_m) \leqslant \varliminf_{n\to\infty} \gamma_{\varphi_n}(\varphi_m) \leqslant \varliminf_{n\to\infty} \gamma_{\varphi_n}(\varphi_n) \leqslant \varliminf_{n\to\infty} \gamma(\varphi_n) = \gamma(\varphi).$$

再令 $m \to \infty$ 得 $\gamma_\varphi(\varphi) \leqslant \gamma(\varphi)$. 由 γ 的任意性知, γ_φ 是 φ 最优耦合. □

定理 4.2(i) 的证明 **第 1 步** 假设 φ 是有界连续函数. 按引理 4.1(iii) 可构造 $\{G_i^{(n,m)}\}_{m\geqslant 1}^{n\geqslant 1}$ 和 $\{G_i^{(n)}\}_{n\geqslant 1}$, 则它们都是转移测度, 且 $\{G_i^{(n,m)}\}$ $(i = 1, 2)$ 的支撑为 $m+1$ 个点. 于是, 根据随机线性规划最优解的理论可知, $\{G_1^{(n,m)}\}$ 与 $\{G_2^{(n,m)}\}$ 的 φ 最可测耦合 $G^{(n,m)}$ 存在. 先固定 n, 考察 $\{G^{(n,m)}\}_{m\geqslant 1}$. 它满足引理 4.2 的条件, 于是存在转移测度 $G^{(n)}$, 以及对每个 $(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$, 存在子列 $\{m(\omega_1, \omega_2)\} \subset \{m\}$, 使得

$$G^{(n,m(\omega_1,\omega_2))}(\omega_1, \omega_2; \cdot) \to G^{(n)}(\omega_1, \omega_2; \cdot), \quad m(\omega_1, \omega_2) \to \infty. \tag{4.6}$$

对固定的 $(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$, 应用引理 4.1(iii) 和 4.4(i) 知,

$$G^{(n)}(\omega_1, \omega_2; \cdot) \in K(G_1^{(n)}(\omega_1, \cdot), G_2^{(n)}(\omega_2, \cdot)),$$

即 $G^{(n)}$ 是 $G_1^{(n)}$ 与 $G_2^{(n)}$ 的可测耦合. 设 $\gamma \in K(G_1(\omega_1, \cdot), G_2(\omega_2, \cdot))$, 按引理 4.1(iii)(b) 构造 $\{\gamma^{(n,m)}\}_{m\geqslant 1}^{n\geqslant 1}$ 和 $\{\gamma^{(n)}\}_{n\geqslant 1}$, 那么有

$$G^{(n,m(\omega_1,\omega_2))}(\omega_1, \omega_2; \varphi) \leqslant \gamma^{(n,m(\omega_1,\omega_2))}(\varphi) \quad (\text{由 } G^{(n,m)} \text{ 的 } \varphi \text{ 最优性}).$$

在上式中令 $m(\omega_1, \omega_2) \to \infty$, 由 (4.6) 及 φ 的连续性和引理 4.1(iii) 知,

$$G^{(n)}(\omega_1, \omega_2; \varphi) \leqslant \gamma^{(n)}(\varphi). \tag{4.7}$$

像处理 $\{G^{(n,m)}\}_{m\geqslant 1}$ 一样处理 $\{G^{(n)}\}_{n\geqslant 1}$ 知, 存在转移测度 G, 以及对每个 $(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$, 存在子序列 $n(\omega_1, \omega_2) \subset \{n\}$, 使得

$$G^{(n(\omega_1,\omega_2))}(\omega_1, \omega_2, \cdot) \to G(\omega_1, \omega_2, \cdot), \quad n(\omega_1, \omega_2) \to \infty, \tag{4.8}$$

且 G 是 G_1 与 G_2 的可测耦合. 令 $n(\omega_1,\omega_2)\to\infty$, 由 (4.7) 和 (4.8) 知, $G(\omega_1,\omega_2;\varphi)\leqslant\gamma(\varphi)$. 由

$$\gamma\in K(G_1(\omega_1,\cdot),G_2(\omega_2,\cdot))$$

的任意性知, G 是 G_1 与 G_2 的 φ 最优可测耦合.

第 2 步 先假设 φ 下半连续, 且 $0\leqslant\varphi\leqslant 1$. 按引理 4.3 构造 $\{\varphi_m^{(n)}\}_{m\geqslant 1}^{n\geqslant 1}$ 和 $\{\varphi_m\}_{m\geqslant 1}$. 由于 $\varphi_m^{(n)}$ 有界连续, 由第 1 步的证明知, G_1 与 G_2 的 $\varphi_m^{(n)}$ 最优可测耦合 $G_m^{(n)}$ 存在. 先固定 m, 考察 $\{G_m^{(n)}\}_{n\geqslant 1}$. 由引理 4.2 知, 存在转移测度 G_m, 对 $\forall(\omega_1,\omega_2)\in\Omega_1\times\Omega_2$, 存在 $\{n(\omega_1,\omega_2)\}\subset\{n\}$, 使得

$$G_m^{(n(\omega_1,\omega_2))}(\omega_1,\omega_2;\cdot)\to G_m(\omega_1,\omega_2;\cdot),\quad n(\omega_1,\omega_2)\to\infty,\quad m=1,2,\ldots$$

由此及引理 4.4(ii) 知, G_m 是 G_1 与 G_2 的 φ_m 最优可测耦合. 同处理序列 $\{G_m^{(n)}\}_{n\geqslant 1}$ 一样处理序列 $\{G_m\}_{m\geqslant 1}$ 可知, G_1 与 G_2 的 φ 最优可测耦合 G_φ 存在. 由此及引理 4.4 知, 当 φ 为一般的有下界下半连续函数时, 令 $\psi_n=\min\{\varphi,n\}$, 则 ψ_n 是有界的下半连续函数, 故 G_1 与 G_2 的 ψ_n 最优可测耦合 G_{ψ_n} 存在, 注意到 $\psi_n\nearrow\varphi$ $(n\nearrow\infty)$. 同处理 $\{G_m^{(n)}\}_{n\geqslant 1}$ 一样处理 $\{G_{\psi_n}\}_{n\geqslant 1}$ 可得到 G_1 与 G_2 的 φ 最优可测耦合 G 存在. 定理证毕. □

注 4.1 定理 4.2(i) 来源于陈木法院士在专著 [1] 的第一版提出的一个公开问题 (猜想). 设 Polish 空间 $(X,\rho,\mathscr{B}(X))$ 上的概率核为 $P(x,dy)$, 他猜想 $P(x_1,dy_1)$ 和 $P(x_2,dy_2)$ 的 ρ 最优可测耦合 $P(x_1,x_2,dy_1,dy_2)$ 存在, 即 $P(x_1,x_2;\rho)=\omega_\rho(P(x_1,\cdot),P(x_2,\cdot))$. 文献 [11] 在假设 $\{P(x,\cdot)\mid x\in X\}$ 是一致胎紧和一致可积的条件下, 利用引理 4.1(i) 构造出具有有限支撑的概率核序列 $\{P_n\}_{n\geqslant 1}$, 然后利用随机线性规划最优解的可测性证明了对任意 $\varepsilon>0$, 存在 P 与 P 的 ρ 最优可测耦合 $P^{(\varepsilon)}$, 使得

$$P^{(\varepsilon)}(x_1,x_2;\rho)\leqslant\omega_\rho(P(x_1,\cdot),P(x_2,\cdot))+\varepsilon,$$

此结果为解决这个问题迈出了第一步. 随后在文献 [11] 工作的基础上, 文献 [13] 用证明引理 4.2 的方法 (引理 4.2 是文献 [7] 正式提出的, 它是解决该类问题的有力工具), 在一定条件下 (去掉一致可积性), 证明了存在可测耦合使得 $P(x_1,x_2;\rho)=\omega_\rho(P(x_1,\cdot),P(x_2,\cdot))$. 文献 [12] 为了去掉 $\{P(x,\cdot)\mid x\in X\}$ 是一致胎紧的假设, 构造出了引理 4.1(ii) 双指标和单指标概率核序列. 该构造看似平凡, 但它却成功地去掉了要求概率核是一致胎紧的假设, 从而最终解决了陈木法猜想.

下面针对专著 [1] 对定理 5.32 的简化证明中所出现的 $B_0^{(n)}$ (参见文献 [1, 第 206 页第 28 行]) 作些说明, 因为 $\{G_i(\omega_i,\cdot)\mid\omega_i\in\Omega_i\}$ 没有一致胎紧性. 因此, 构造 $B_0^{(n)}$ 时应为

$$B_0^{(n)}=\bigcup_{k=m_n(\omega_i)+1}^{\infty}B_k^{(n)},$$

即 $B_0^{(n)}$ 与 ω_i 有关. 由此需要证明 $m_n(\omega_i)$ 是 $\mathscr{F}_i$ 可测的. 即使证明了 $m_n(\omega_i)$ 是可测的, 由于 $m_n(\omega_i)$ 的维数是不定的, 因此需要讨论无穷维线性规划的可测性问题 (文献 [1] 的简化证明中对所提的两个问题都没有说明). 此外, 文献 [1, 第 207 页第 4 行] 对 $\tilde{\mu}^n$ 的构造是不够的, 因为只有当 $\tilde{\mu}$ 是 μ_1 与 μ_2 的独立耦合时, $\tilde{\mu}^n$ 才能弱收敛到 $\tilde{\mu}$, 与证明的要求不符 (证明中要求 $\tilde{\mu}$ 是 μ_1 与 μ_2 的任意的耦合). 本文引理 4.1(iii) 对定理 5.23 的简化证明中这两处不妥之处进行了更正. 因此, 从现在来看, 构造双指标序列来处理这个问题是最有效的 (它也符合人们认识无穷的过程, 有限 $\to$ 可数无穷 $\to$ 不可数无穷). 由于 ρ 最优可测耦合的存在性不能包括保序 Markov 耦合的存在性等问题, 文献 [10] 提出了 φ 最优耦合和 φ 最优可测耦合的概念, 并证明了这最优耦合的存在性, 此处, φ 是非负下半连续函数, 把 ρ 换成 φ

给证明最优耦合存在性增加了很多麻烦: (1) $\omega_\varphi(\mu,\nu)$ 不再是概率度量, 因此不能用三角不等式; (2) φ 不是连续的, 因此, 在取下极限时, 出现两头大中间小的情形. 为了克服这个问题, 第一作者在文献 [10] 中首次给出了引理 4.1(iii)(b). 另外, 引理 4.3 是专门处理 φ 的, 对 ρ 是不需要的. 到此证明定理 4.2(i) 的思路就清楚了. 该定理证明思想的主线是 "先退后进". "退", 从 G_i 退到 $G_i^{(n)}$, 再退到 $G_i^{(n,m)}$ (引理 4.1(iii)), 又从 φ 退到 ψ_n, 再退到 $0\leqslant\varphi\leqslant 1$, 再退到 φ_m, 再退到 $\varphi_m^{(n)}$ (引理 4.3). "进", 用引理 4.2 及一些辅助引理, 分别从 $G_i^{(n,m)}$ 和 $\varphi_m^{(n)}$ 逐步进到 G_i 和 φ.

5 带约束 n 重 φ 最优耦合

设 X_i $(i=1,\dots,n)$ 是 Polish 空间, $X=\prod_{i=1}^n X_i$, 则 X 是乘积 Polish 空间. 设 $\mu_i\in\mathscr{P}(X_i)$ $(i=1,\dots,n)$, 记 $K(\mu_1,\dots,\mu_n)=\{\mu\in\mathscr{P}(X)\mid \mu\pi_i^{-1}=\mu_i, i=1,\dots,n\}$ (其中 π_i 为 X 到 X_i 的投影). 若 $\mu\in K(\mu_1,\dots,\mu_n)$, 则称 μ 是 $\mu_1,\dots,\mu_n$ 的 n 重耦合概率, μ_i $(i=1,\dots,n)$ 称为 μ 的边缘概率.

引理 5.1 设 $\mu_i\in\mathscr{P}(X_i)$ $(i=1,\dots,n)$, 则 $K(\mu_1,\dots,\mu_n)$ 是 $\mathscr{P}(X)$ 中的紧子集.

证明 由于 X_i 是 Polish 空间, 则 μ_i 是内正则的, 故对任意的 $\varepsilon>0$, 存在 X_i 的紧集 K_i, 使得 $\mu_i(K_i^c)<\frac{\varepsilon}{n}$ $(i=1,\dots,n)$. 由于 $K_1\times\cdots\times K_n$ 是 X 上的紧子集, 且 $\forall\,\mu\in K(\mu_1,\dots,\mu_n)$, 有

$$\begin{aligned}\mu(K^c)&\leqslant\sum_{i=2}^{n-1}\mu(K_1\times\cdots\times K_{i-1}\times K_i^c\times K_{i+1}\times\cdots\times K_n)\\&\quad+\mu(K_1^c\times K_2\times\cdots\times K_n)+\mu(K_1\times\cdots\times K_{n-1}\times K_n^c)\\&<\frac{\varepsilon}{n}+\cdots+\frac{\varepsilon}{n}=\varepsilon,\end{aligned}$$

即 $K(\mu_1,\dots,\mu_n)$ 是一致胎紧的, 又设 $\{\mu^{(n)}\}\subset K(\mu_1,\dots,\mu_n)$, $\mu\in\mathscr{P}(X)$, 且 $\mu_n\to\mu$, 那么, 对于 X_i 上的任意有界连续函数 f_i, 则 $\pi_i^{-1}f_i$ 是 X 上的有界连续函数, 有

$$\mu_i(f_i)=(\mu^{(n)}\pi_i^{-1})(f_i)=\mu^{(n)}(\pi_i^{-1}f_i)\to\mu(\pi_i^{-1}f_i)=(\mu\pi_i^{-1})(f_i).$$

由 f_i 的任意性知, $\mu\pi_i^{-1}=\mu_i$ $(i=1,\dots,n)$. 故 $\mu\in K(\mu_1,\dots,\mu_n)$, 即 $K(\mu_1,\dots,\mu_n)$ 是闭集. 由上面讨论及 Prohorov 定理知, $K(\mu_1,\dots,\mu_n)$ 是 $\mathscr{P}(X)$ 中的紧子集. □

设 $f:X\to\mathbb{R}$ 可测映射, 记 $\widehat{f}(\mu)=\mu(f)$ $(\forall\,\mu\in\mathscr{P}(X))$.

引理 5.2 设 $f:X\to\mathbb{R}$ 是有上界的上半连续函数, 则 $\widehat{f}:\mathscr{P}(X)\to\mathbb{R}$ 也是上半连续函数, 且在 $\mathscr{P}(X)$ 的任何紧子集 K 上达到最大值, 即存在 $\bar\mu\in K$, 使得 $\widehat{f}(\bar\mu)=\sup\{\widehat{f}(\mu):\mu\in K\}$.

证明 记常数 $c=\sup_{x\in X}f(x)$, 令 $g=-f+c$, 则 g 是非负下半连续函数. 设 $\{\mu^{(n)}\}_{n\geqslant1}\subset\mathscr{P}(X)$, $\mu\in\mathscr{P}(X)$, 且 $\mu^{(n)}\to\mu$ $(n\to\infty)$. 由文献 [1, 定理 4.5] 知,

$$\varliminf_{n\to\infty}\mu^{(n)}(g)\geqslant\mu(g).$$

此即表明 $\widehat{g}(\mu)$ 在 $\mathscr{P}(X)$ 上下半连续, 那么 $\widehat{g}(\mu)$ 在紧子集 K 上可达到最小值 (参见文献 [14, 第 12 页推论 1.2]), 即存在 $\bar\mu\in K$, 使得 $\widehat{g}(\bar\mu)=\min\{\widehat{g}(\mu)\mid\mu\in K\}$. 又 $\widehat{f}(\mu)=\mu(f)=\mu(c-g)=c-\mu(g)=c-\widehat{g}(\mu)$, 故有 $\widehat{f}(\bar\mu)=\sup\{\widehat{f}(\mu)\mid\mu\in K\}$. □

定理 5.1 设 $u_i,\varphi:X\to\mathbb{R}$ $(i=1,\dots,n)$ 是有上界的上半连续函数, 设 $\mu_i\in\mathscr{P}(X_i)$, $i=1,\dots,n$, 如果存在实数 a_i $(i=1,\dots,n)$, 使得

$$H:=\{\mu\in K(\mu_1,\dots,\mu_n)\mid\mu(u_i)\geqslant a_i, i=1,\dots,n\}\neq\emptyset,$$

那么存在 $\mu_\varphi \in H$, 使得

$$\mu_\varphi(\varphi) = \sup\{\mu(\varphi) \mid \mu \in H\}. \tag{5.1}$$

证明 由于 u_i 和 φ 都是有上界的上半连续函数, 由引理 5.2 知, $\widehat{u}_i$ 和 $\widehat{\varphi}$ 都是 $\mathscr{P}(X)$ 上的上半连续函数. 于是, 为证本定理, 由引理 5.2 知, 只需证 H 是 $\mathscr{P}(X)$ 上的紧子集. 事实上, 由于 $\widehat{u}_i$ 上半连续, 故 $\{\mu \in \mathscr{P}(X) \mid \mu(u_i) \geqslant a_i\} = \{\mu \in \mathscr{P}(X) \mid \widehat{u}_i(\mu) \geqslant a_i\}$ $(i = 1, \ldots, n)$ 是闭集. 由引理 5.1 知, $K(\mu_1, \ldots, \mu_n)$ 是紧集, 故

$$H = K(\mu_1, \ldots, \mu_n) \cap \left(\bigcap_{i=1}^{n} \{\mu \in \mathscr{P}(X) \mid \mu(u_i) \geqslant a_i\} \right)$$

是 $\mathscr{P}(X)$ 上的紧子集. 证毕. □

定义 5.1 把满足 (5.1) 的 μ_φ 称为带约束 $\mu(u_i) \geqslant a_i$ $(i = 1, \ldots, n)$ 的 $\mu_1, \ldots, \mu_n$ 的 n 重 φ 最优耦合.

我们将在下节给出定理 5.1 在博弈论中的应用.

6 Nash 均衡的最优合作均衡

为了便于阅读, 本文先简要介绍博弈论的基本概念和基本思想 (详见文献 [15, 16]). 博弈论是研究多个决策参与人在行为直接作用时的决策及这种决策的均衡问题, 为了对这种博弈问题进行定量分析, 需要建立理想化的数学模型. 策略型博弈模型 $G = (N, (X_i)_{i=N}, (u_i)_{i \in N})$ 是博弈论的最基本模型之一. 该模型有三个基本要素: (1) n 个参与人所组成的集合 $N = \{1, \ldots, n\}$; (2) 每个参与人 i $(\in N)$ 有一个可供选择的集合 X_i 称为 i 的策略空间; (3) 每个参与人 i $(\in N)$ 有一个目标函数 u_i $(u_i : \prod_{i=1}^n X_i \to \mathbb{R})$ 称为 i 的支付函数. N、$(X_i)_{i \in N}$ 和 $(u_i)_{i \in N}$ 是每个参与人 i 都知道的, 称为他们共同的知识. $X = \prod_{i=1}^n X_i$ 称为组合策略空间, $x_i \in X_i$ 称为 i 的策略. $x = (x_1, \ldots, x_n) \in X$ 称为组合策略, 对于 $i \in N$, 引入 i 的对手 $-i = N - \{i\}$ 是很方便的. 记 $X_{-i} = \prod_{k \neq i} X_k$, 称为 i 的对手的策略空间. 设 $x = (x_1, \ldots, x_n) \in X$, 记

$$(x_i, x_{-i}) = (x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n).$$

要对博弈模型 G 进行分析, 除上述要素之外, 还要加上每个参与人都是理性的. 参与者 i $(\in N)$ 都是理性的, 就是参与人作出决策时, 要与所追求目标一致, 其数学意义就是当 i 的对手选择策略 $x_{-i} \in X_{-i}$ 时, 参与人 i 选择 x_i 要使自己的目标函数 u_i 最大化, 即

$$u_i(x) = u_i(x_i, x_{-i}) = \max_{x_i' \in X_i} u_i(x_i', x_{-i}). \tag{6.1}$$

博弈主要分为非合作博弈和合作博弈, 往往是两种博弈结合起来使用. 先看非合作博弈. 非合作博弈是 N 中的全体参与者根据共同知识和理性假设, 独立地作出决策, Nash 建议各参与者作出的组合决策 $x^* = (x_1^*, \ldots, x_n^*)$ 应使得下式成立:

$$u_i(x_i^*, x_{-i}^*) \geqslant u_i(x_i, x_i^*), \quad x_i \in X_i, \quad i = 1, \ldots, n. \tag{6.2}$$

满足 (6.2) 的组合决策 $x^*=(x_1^*,\dots,x_n^*)$ 称为博弈 G 的一个 Nash 均衡, 也称博弈 G 的一个非合作博弈的解.

比较 (6.1) 和 (6.2). (6.1) 中的决策是参与人 i 假设对手决策为 x_{-i} 时所作出的理性决策 $x_i\ (x_{-i})$ (没有独立性). 该决策不是真实的决策. 只是参与人 i 的一个预演, 而 (6.2) 是所有参与人逐步多次预演, 然后对其对手作出理性的预测后, 独立地作出的决定, 它调和了每个人的个人理性之间的矛盾, 从 (6.2) 中可看出, N 中任何一个参与者 i 在其对手都不改变策略的条件下, 若第 i 个参与者改变策略, 它的目标函数 u_i 的值不会增加. Nash 均衡在上述意义下是一种弱有效性, 有时 Nash 均衡是一种很差的解 (如囚徒困境问题 [16]). 虽然如此, Nash 均衡能同时满足每个人的个人理性. Nash 均衡 x^* 是一种自我实施 (self-enforcing) 的策略 (即每个人都自愿执行组合策略 $x^*=(x_1^*,\dots,x_n^*)$).

下面简要介绍 Nash 均衡的存在性问题, 即使 G 中的策略空间 $X_i\ (i=1,\dots,n)$ 都是有限集时, G 的 Nash 均衡也不一定存在, 为了解决 Nash 均衡的存在性问题, 可把 X_i 换成 $\mathscr{P}(X_i)$, 把 u_i 按下面方式换成从 $\prod_{i=1}^n\mathscr{P}(X_i)$ 到 $\mathbb{R}$ 的映射, 仍记为

$$u_i(\mu_1,\dots,\mu_n):=\int u_i(x_1,\dots,x_n)\mu_1(dx_1)\cdots\mu_n(dx_n).$$

若 $\mu=(\mu_1,\dots,\mu_n)\in\prod_{i=1}^n\mathscr{P}(X_i)$, 则 $u_i(\mu)$ 称为混合策略组合的 μ 的期望支付. 如果存在

$$\mu^*=(\mu_1^*,\dots,\mu_n^*)\in\prod_{i=1}^n\mathscr{P}(X_i),$$

满足

$$u_i(\mu^*)=u_i(\mu_i^*,\mu_{-i}^*)\geqslant u_i(\mu_i,\mu_{-i}^*),\quad \mu_i\in\mathscr{P}(X_i),\quad i=1,\dots,n. \tag{6.3}$$

把满足 (6.3) 的 μ^* 称为博弈 G 的 (混合) Nash 均衡, 而把满足 (6.2) 的 x^* 称为 G 的 (纯) Nash 均衡, 后面把两种均衡统称为 Nash 均衡.

下面定理是 1951 年由 Nash 得到的著名 Nash 均衡存在定理.

定理 6.1 (Nash 均衡)　设 $G=(N,(X_i)_{i\in N},(u_i)_{i\in N})$ 是一个有限博弈 (即 X_i 都是有限集), 则 G 的 Nash 均衡总存在.

该定理的证明用到核心数学中的 Kakutani 不动点定理. Glicksberg 把定理 6.1 推广为下面定理 6.2.

定理 6.2　设 $G=(N,(X_i)_{i\in N},(u_i)_{i\in N})$ 是一个连续型博弈, 即 $(X_i,\rho_i)\ (i\in N)$ 是紧度量空间; $u_i:\prod_{i=1}^n X_i\to\mathbb{R}\ (i\in N)$ 是连续函数, 那么博弈 G 的 Nash 均衡总是存在的.

关于这两个定理可参见文献 [15, 第 3 章].

注意到 Nash 均衡是由个人理性得到均衡, 其实我们还可在满足个人理性的条件下, 再满足集体理性. 在博弈 G 中令 $\varphi=\sum_{i=1}^n u_i$, 把 φ 作为集体 N 的目标函数, 我们提出 Nash 均衡的最优合作均衡 (所谓合作博弈就是集体 N 为了共同目标 φ, 采取一致的行动).

定义 6.1　设 G 是满足定理 6.2 的条件的博弈模型,

$$\mu^*=(\mu_1^*,\dots,\mu_n^*)\in\prod_{i=1}^n\mathscr{P}(X_i)$$

是 G 的一个 Nash 均衡, 记 $\mu^*(u_i)=a_i$. 如果存在

$$\bar{\mu}\in K(\mu_1^*,\dots,\mu_n^*)\subset\mathscr{P}\bigg(\prod_{i=1}^n X_i\bigg),$$

使得

$$\bar{\mu}(\varphi) = \max\{\mu(\varphi) \mid \mu \in K(\mu_1^*, \ldots, \mu_n^*), \mu(u_i) \geqslant a_i, i \in N\},$$

则称 $\bar{\mu}$ 为 Nash 均衡 μ^* 的最优合作均衡.

Nash 均衡的最优合作均衡是优先满足个人理性, 再满足集体理性. 作为定理 5.1 的特例, 下面的定理成立.

定理 6.3 设 G 是满足定理 6.2 的条件博弈, μ^* 是 G 的一个 Nash 均衡, 那么 μ^* 的最优合作博弈 $\bar{\mu}$ 总是存在的.

下面说明 $\bar{\mu}$ 往往严格优于 μ^*.

例 6.1 博弈 $G = (N, (X_i)_{i\in N}, (u_i)_{i\in N})$, $N = \{1,2\}$, $X_1 = \{U, D\}$, $X_2 = \{L, R\}$, $u_1(U, L) = 5$, $u_1(U, R) = 0$, $u_1(D, L) = 4$, $u_1(D, R) = 1$, $u_2(U, L) = 1$, $u_2(U, R) = 0$, $u_2(D, L) = 4$, $u_2(D, R) = 5$. 容易验证

$$\mu^* = (\mu_1^*, \mu_2^*) = \left(\frac{1}{2}\delta_u + \frac{1}{2}\delta_D, \frac{1}{2}\delta_L + \frac{1}{2}\delta_D\right)$$

是 G 的一个混合 Nash 均衡 (其中 δ_x 表示单点概率测度). 此时, Nash 均衡 μ^* 关于 u_1 和 u_2 的期望支付分别为 $\mu^*(u_1) = 2.5$ 和 $\mu^*(u_2) = 2.5$, 也可以算出 μ^* 的最优合作均衡

$$\bar{\mu} = \frac{1}{2}\delta_{(U,L)} + \frac{1}{2}\delta_{(D,R)},$$

并可算出

$$\bar{\mu}(\varphi) = 6, \quad \bar{\mu}(u_1) = \bar{\mu}(u_2) = 3 > 2.5 = \mu^*(u_1) = \mu^*(u_2),$$

因此, 决策 $\bar{\mu}$ 比决策 μ^* 好. 更难能可贵的是 $\bar{\mu}$ 同 μ^* 一样, 也是一个自我实施的决策. 实施这两个决策的不同之处在于, 实施 μ^* 时, μ_1^* 与 μ_2^* 是独立作出的, 但实施 $\bar{\mu}$ 时, 参与人 1 与 2 需要合作 (例如, 两个人可通过投一枚均匀硬币, 正面向上时, 1 采取决策 U, 同时 2 采取决策 L; 反面向上时, 1 采取决策 D, 同时 2 采取决策 R).

致谢 衷心感谢审稿人的中肯评判和修改意见.

参考文献

1 Chen M F. From Markov Chains to Non-Equilibrium Particle Systems, 2nd ed. Singapore: World Scientific, 2004
2 Chen M F. Eigenvalues, Inequalities and Ergodic Theory. New York: Springer-Verlag, 2004
3 严士健. 无穷粒子马尔科夫过程引论. 北京: 北京师范大学出版社, 1989
4 张绍义. 最优 Markov 耦合的存在性及其应用. 博士学位论文. 北京: 北京师范大学, 2000
5 Lindvall T. Lectures on the Coupling Method. New York: John Wiley & Sons, 1992
6 Mukherjea A, Tserperpes N A. Measures on Topological Semigroups. Lecture Notes in Mathematics, Vol. 1547. New York: Springer-Verlag, 1976
7 张绍义. 跳过程 ρ 最优耦合算子的存在性. 数学学报, 1998, 41: 393–398
8 Zhang Y H. Sufficient and necessary conditions for stochastic comparability of jump processes. Acta Math Sin Engl Ser, 2000, 16: 99–102
9 Leskelä L. Stochastic relations of random variables and processes. J Theoret Probab, 2010, 23: 523–546
10 张绍义. 关于非负下半连续函数最优可测耦合的存在性及其应用. 数学学报, 2000, 43: 773–780
11 张绍义. 转移概率的可测耦合与概率距离. 数学年刊, 1995, 16: 769–775
12 张绍义. 最优可测的存在性与 Markov 过程的遍历性. 中国科学 A 辑, 1998, 28: 999–1008
13 张绍义, 徐侃. 转移概率最优可测耦合的存在性. 数学学报, 1997, 40: 5–13
14 Aubin J P. Optima and Equilibria: An Introduction to Nonlinear Analysis. New York: Springer-Verlag, 1993
15 Roger B M. 博弈论: 矛盾冲突分析. 于寅, 费剑平, 译. 北京: 中国人民大学出版社, 2015

16 张维迎. 博弈论与信息经济学. 上海: 上海三联书店; 上海人民出版社, 1996

Optimal coupling in probability theory and optimal cooperation in game theory

Shaoyi Zhang & Ren Zhang

Abstract In this paper, we extend the concept of binary optimal coupling. We obtain the result I: Suppose X and Y are Polish spaces, $\varphi : X \times Y \to \mathbb{R}$ is measurable, $\mu \in \mathscr{P}(X)$, $\nu \in \mathscr{P}(Y)$. (i) If φ is a lower semi-continuous function with a lower bound, then φ optimal coupling γ_φ exists; (ii) if φ is an upper semi-continuous function with an upper bound, then φ upper optimal coupling γ^φ exists. In addition, we obtain the result II: Suppose G_i $(i = 1, 2)$ is a transition probability measure sequence. (i) If $\varphi : X_1 \times X_2 \to \mathbb{R}$ is a lower semi-continuous function with a lower bound, then φ optimal coupling of G_1 and G_2 exist; (ii) if $\varphi : X_1 \times X_2 \to \mathbb{R}$ is an upper semi-continuous function with an upper bound, then φ upper optimal coupling G_1 and G_2 exist. A concept of n-fold optimal coupling with constraints is presented and the existence of this optimal coupling is proved. Furthermore, an optimal cooperation equilibrium of Nash equilibrium in game theory is defined. This equilibrium is superior to Nash equilibrium as illustrated.

Keywords binary optimal coupling, n-fold optimal coupling, Nash equilibrium

MSC(2010) 60A99, 91A12

doi: 10.1360/N012018-00161