arXiv:2601.09891v1 [econ.TH] 14 Jan 2026

模型设定错误下的学习与均衡 *

Ignacio Esponda DemianPouzo

2026年1月16日

摘要

本章提出一个统一框架，用以研究模型设定错误下的学习情形：主体在对其环境的错误模型中进行优化并更新信念。我们回顾从设定错误的模型中学习的统计学基础，并将这些见解拓展至具有内生的、依赖行动的数据的环境，包括单个主体情形和策略情形。

“所有模型都是错的，但有些模型很有用。”*George Box*。

关键词：模型设定错误下的学习、伯克–纳什均衡、收敛

1 引言

主体对其所处环境持有错误信念的情形在经济学中十分普遍。这类信念扭曲主要有两种成因：一是会系统性地造成信念形成偏差的行为机制（例如过度自信、忽视相关性、遗漏变量偏误）；二是出于实用考虑而进行的简化，即真实环境过于复杂，难以加以表征，或只能观察到粗略、

*我们感谢 Mira Frick、Drew Fudenberg、Ryota Iijima、Yuhta Ishii、Ran Spiegler、Philipp Strack、Yuichi Yamamoto、一位匿名审稿人，尤其感谢 Giacomo Lanzani 提供的宝贵意见。Esponda：iesponda@ucsb.edu。Pouzo：dpouzo@berkely.edu。

有限的信号，或存在一定程度的有限理性。本章提出一个统一的概念框架，以同一分析视角涵盖模型设定错误下学习的两种来源。在这一框架中，主体被建模为在受约束的主观模型类中进行优化。该框架不预先设定信念，而是明确主体用以解释数据的模型类、信息结构和更新规则；信念随后在这一学习过程中内生形成，而行动则根据这些信念进行最优选择。

尽管我们聚焦于一类特定的重复环境，但其 underlying 方法更为一般。该框架旨在明确环境是什么、主体观察到和未观察到什么，以及主体使用哪种（可能存在模型设定错误的）模型来解释数据。给定这些原始要素，以及博弈过程中产生的反馈，信念和行为便会内生决定。我们认为，这种严谨性是该方法的主要价值所在，也为扩展到本章节所涵盖情形之外的环境提供了范本。

我们对单主体和策略情形下的模型设定错误下的学习进行了自洽的阐述。我们介绍了原始要素——主体的主观模型类、数据结构和更新规则——并展示这些要素如何随时间生成信念和选择。我们讨论了针对模型设定错误的均衡概念（伯克–纳什均衡），并概述证明收敛至均衡（或证明不收敛）以及界定长期博弈过程的方法。我们的写法是陈述形式化定义和命题，提供突出关键思想与直觉的证明概要，并指出完整论证的原始来源；与此同时，我们还强调文献中发展的扩展和变体。

本章节并非对经济学中模型设定错误问题的全面综述。我们不讨论模型设定错误下学习的宏观经济学文献（例如 Evans and Honkapohja（2001年）、Branch（2006年）和 Hansen and Sargent（2011年）），也省略了大量关于模型设定错误的微观经济学文献；相关综述见 Bohren and Hauser（2025）。我们的目标是与之互补：深入探讨一组更为有限的思想，统一符号，并

突出主要的证明策略，以阐明这些结果在何时以及为何适用。我们提供的参考文献指向原始论文，也指向包含更详尽参考文献、可供感兴趣读者进一步查阅的相关文献。

第2节从外生数据学习讲起，这是统计学家研究的经典环境。第3至5节则转向从内生数据学习：在这种环境中，一个或多个主体（或博弈参与者）选择行动，从而塑造自己观察到的数据。这一研究方向主要由经济学领域发展而来。尽管可以分析许多行动环境，我们将聚焦于一种简单且标准的重复环境：同一个决策问题在每个时期反复出现，并以此阐释主要思想。模型设定错误下的学习也在其他类型的问题中得到探讨——其中包括大量关于社会学习的文献，研究主体如何从前人的行动中得出（可能错误的）推断（例如，Eyster 和 Rabin（2010）；Bohren（2016）；Bohren 和 Hauser（2021）；Frick、Iijima 和 Ishii（2020））、马尔可夫决策过程（Esponda 和 Pouzo（2021）；Anderson 等（2024））、最优停止（He（2022））以及递归宏观经济模型（例如，Molavi 等（2019）；Fuster、Laibson 和 Mendel（2010））等。这些环境借鉴了为我们的主要环境开发的一些工具，但本章不会涉及它们。

我们通过将本章节的主题置于更广阔的背景下，为引言作结——简要追溯促成当前框架的思想发展路径，并指出近期的发展和未来研究方向——从而阐明我们是如何走到今天这一步的，并描述该领域文献下一步的发展方向。

1.1 经济学中模型设定错误下的学习发展的五条脉络

在本小节中，我们通过勾勒塑造现代框架的五条脉络，将本章节置于更广阔的历史发展脉络中。我们的目标并非面面俱到（我们略去了几项重要贡献），而是提供方向指引：勾勒一幅路线图，说明该领域如何发展出当前的工具体系，以及近期研究正朝何处迈进。

第一条脉络可追溯至20世纪70年代，它将学习与均衡联系起来。Hahn（1973年）认为，均衡要求“学习过程已经穷尽”：新到的数据不再促使主体修正其理论。在未发表的研究中，Arrow 和 Green（1973年）提出了一个早期框架，可将其解读为模型设定错误问题：他们探讨均衡何时存在、具有何种性质，以及从非均衡信念出发的学习过程能否达到均衡；同时要求实际实现的结果分布 与预期的 分布相符，而不仅仅是均值相符。Kirman（1975年）通过建立持有关于其环境的错误理论的企业模型，明确提出了模型设定错误这一视角。然而，这些早期研究缺少的是与当时关于从错设模型中学习的统计学文献的系统联系，尤其是 Berk（1966）。

第二条脉络兴起于20世纪80年代末和20世纪90年代初，发展出了自我确认均衡（SCE）的概念。在纳什均衡中，博弈参与者的信念在均衡路径上和路径外都是正确的；而在SCE中，相比之下，信念只需在博弈过程的路径上正确。更准确地说，每个博弈参与者的策略在其信念给定时都是最优的，并且这些信念与实际博弈过程产生的反馈相一致；但对于未到达的偶然情况（路径外信息集）的信念则不必正确。这体现了对Hahn“没有进一步学习”思想的一种理解：新获得的数据以这样的意义确认了主体的理论，即观察到的任何内容都没有与之矛盾，但这并不意味着该理论在整体上是正确的。Fudenberg 和 Levine（1993a）在博弈中发展了SCE（另见Battigalli（1987年）以及Fudenberg 和 Kreps（1988年)），并且与同期关于适应性学习的研究（例如，Fudenberg 和 Levine（1993b）；Fudenberg 和 Kreps（1995年））一道，分析了纳什均衡或SCE何时能够作为学习动态的长期结果出现——关于博弈中的学习，可参见Fudenberg 和 Levine（1998年）的教材。重要的是，SCE并非本质上与模型设定错误有关：主体可以拥有关于博弈的正确模型，却仅仅因为这些偶然情况从未产生数据，而对路径外博弈过程持有错误信念。相较之下，在模型设定错误下的学习中，主体的模型类本身排除了真实数据生成过程，

因此，即使是路径上反馈，也可能只能由一种错误理论得到最合理的解释。

第三条脉络兴起于 2000 年代初，引入了嵌入特定错误信念的均衡概念。诅咒均衡和行为均衡刻画了对他人行动信息内容的忽视（Eyster和Rabin 2005年；Esponda 2008年）。类比预期均衡将粗略推理形式化：主体将决策情境归为少数几个类比类别，形成类别层面的预期，并对这些预期作出最佳回应；这些预期在每个类别内必须平均而言是正确的（Jehiel和Koessler 2008年）。这些都是模型设定错误的特例，而且也是静态的：这些偏差被直接写入均衡定义，而非从数据生成过程和学习规则中明确推导出来；它们也尚未涉及模型设定错误学习的统计理论。

第四条脉络认识到，在统一框架内对多种形式的模型设定错误进行建模的价值，并明确引入模型设定错误学习统计学的工具，以发展一般性的解概念和学习基础。本章回顾的正是这部分研究。其出现恰逢其时：尽管早期行为和有限理性文献很少将对错误模型的关注称为“模型设定错误”，这类关注却再度兴起；与此同时，博弈中的学习文献为建立形式化基础提供了一条途径，却缺乏与模型设定错误下信念渐近刻画的统计结果之间的紧密联系。关键的一步是将这些统计学洞见——最初由统计学家针对外生数据推导得出——应用于数据由主体行动所塑造的内生环境。由此呈现的图景十分清晰：基于学习的均衡（纳什均衡和自我确认均衡）强调，信念受到主体观察到的内容所限制；同样重要的是，信念也受到解释观察结果所依据的模型的塑造。

第五条研究脉络正在兴起。本章节回顾的研究将主体的（可能存在模型设定错误的）模型视为外生的：根据假设，观测证据永远不会直接证伪模型，主体也不会考虑自己的模型可能有误。这种处理方式可能很自然——例如，在形式化一种未被主体察觉的行为偏差时——但在许多情境中，主体知道自己的模型只是近似。

这至少引出了两个相关问题。首先，模型如何被选定，又如何随时间演变？也就是说，哪些模型设定错误可能持续存在，原因何在？其次，对于承认模型可能存在设定错误、并希望防范此类问题（例如通过稳健决策规则）的主体，恰当的实证与规范性决策理论框架是什么？相关领域的研究正开始积极探讨这些问题；受篇幅所限，我们不在此介绍这些进展（有关第一个问题的参考文献，见 Bohren and Hauser（2025）；有关第二个问题，见 Lanzani（2025））。

2 从外生数据中学习

考虑从一系列外生观测值 Y_1，Y_2，... 中进行学习，这些观测值在真实概率律 Q 下独立同分布，且每个观测值都属于欧几里得空间的一个博雷尔子集 $(\mathbb{Y}, \mathcal{Y})$。主体采用关于观测值 Y_t 的独立同分布主观模型，该模型由 $\mathcal{Q}(\Theta) := \{Q_\theta : \theta \in \Theta\}$给出。该模型可能设定正确，即 $Q \in \mathcal{Q}(\Theta)$，也可能设定错误，即 $Q \notin \mathcal{Q}(\Theta)$。

我们假设存在一个支配 Q及所有 Q_θ的 a σ-有限测度 ν，其密度分别为 q 和 q_θ。不对 ν 作具体规定，便能同时涵盖连续（勒贝格）和离散（计数）情形。

给定数据 $Y_1, \ldots, Y_t$，贝叶斯后验 记为 $\mu_t(\cdot) := \mu(\cdot \mid Y_1, \ldots, Y_t)$；根据贝叶斯更新规则并经过简单代数运算，在其存在时可得 [1]

$$\mu_t(B) = \frac{\int_B e^{-\sum_{i=1}^t \log \frac{q(Y_i)}{q_\theta(Y_i)}} \mu_0(d\theta)}{\int_\Theta e^{-\sum_{i=1}^t \log \frac{q(Y_i)}{q_\theta(Y_i)}} \mu_0(d\theta)}, \text{ for any Borel set } B \subseteq \Theta, \tag{2.1}$$

其中，$\theta \mapsto -\sum_{i=1}^t \log \frac{q}{(Y_i)\, q_\theta(Y_i)}$ 是给定样本（$Y_1, ..., Y_t$）时的对数似然比函数。

作为参考，真实概率分布在无限历史（$Y_1, Y_2, \ldots$）上生成一个概率测度，记为$\mathbf{P}$。类似地，主观模型的每个元素

[1]下面给出确保贝叶斯后验存在的假设。

$\mathcal{Q}(\Theta)$都会在无限历史上生成一个主观概率测度，记为$\mathbf{P}_\theta$，而先验也会通过$\int_\Theta \mathbf{P}_\theta \mu_0(d\theta)$诱导出一个主观概率测度。

2.1 后验集中

贝叶斯更新的一项理想性质是，随着数据量充足，主体最终能够学到真相——或者在模型设定错误时，至少能学到真相的一个近似。为形式化这一思想，相关文献研究后验分布的渐近集中。本节的主要结果确立了这种收敛性。

为得到这类结果，我们需要以下正则条件

假设 2.1（参数空间）。 参数集 Θ 是某个欧几里得空间的紧致子集。

假设 2.2（支配与几乎处处连续）。 对于 ν-几乎处处的 y，映射 $\theta \mapsto q_\theta(y)$ 是连续的。存在可测函数 $M\colon \mathbb{Y} \to [1,\infty)$，使得对于所有 θ 和 ν-几乎处处的 y，

$$q_\theta(y)\, M(y)^{-1} \;\le\; q(y) \;\le\; q_\theta(y)\, M(y) \text{ and } \int M(y) q(y) \nu(dy) < \infty.$$

假设 2.3（全支持）。 先验密度在 Θ 上有正下界。

多布鞅收敛定理。 对于每个博雷尔集 $B \subseteq \Theta$，序列 $(\mu_t(B))_{t\ge 0}$ 是关于由测度 $\int_\Theta \mathbf{P}_\theta \mu_0(d\theta)$ 下的数据生成的滤过的有界鞅。因此，根据多布鞅收敛定理（ Doob （ 1953 年）），$\mu_t(\theta)$ 随着 $t \to \infty$ 几乎必然按 $\int_\Theta \mathbf{P}_\theta \mu_0(d\theta)$ 收敛。由于任何满足 $1 = \int_\Theta \mathbf{P}_\theta(H)\, \mu_0(d\theta)$ 的事件 H 都蕴含，对于 μ_0-几乎处处的 θ，$\mathbf{P}_\theta(H) = 1$，因此，在 $\int_\Theta \mathbf{P}_\theta \mu_0(d\theta)$ 下得到的这一收敛结果可转移至对 μ_0-几乎处处的 θ 成立的 $\mathbf{P}_\theta$。

然而，这一结果有几个重要的局限。首先，根据 Doob 鞅收敛定理得到的收敛仅对 μ_0-几乎所有参数值成立。因此，即使在正确设定情形下，也就是存在一个 $\theta^\star \in \Theta$，使得 $\mathbf{P} = \mathbf{P}_{\theta^\star}$，真实参数 μ_0 也可能

落在定理未能给出结论的例外 μ_0-零测集内。其次，对我们而言更重要的是，这一论证不适用于错设情形，因为相关的概率律是 $\mathbf{P}$ ，而不是推导鞅收敛时所依据的概率律 $\int_\Theta \mathbf{P}_\theta \mu_0(d\theta)$。

基于这些原因，标准的贝叶斯一致性理论转而依赖另一种组织原则，该原则以频率学派概率论证为基础（例如 Schwartz（1965）），它们证明了在$\mathbf{P}$ 下的收敛，从而保证后验分布集中于真实的数据生成分布，而不只是对 μ_0-几乎每个参数值成立。

这一方法中的一个关键概念，是构造主体模型与真实概率分布之间的“自然距离概念”。事实证明，这一距离概念可由 Kullback–Leibler（KL）散度刻画。

Kullback–Leibler（KL）散度。 对于任意 $\theta \in \Theta$， Q_θ 与Q之间的 KL 散度定义为

$$\mathrm{KL}(Q^\star \| Q_\theta) \;:=\; \mathbb{E}_Q\left[\log \frac{q(Y)}{q_\theta(Y)}\right] \;\in [0,\infty].$$

我们定义 KL最小化集合；在我们的条件下，该集合定义良好，记为 a s

$$\Theta^{\mathrm{m}} \;:=\; \arg\min_{\theta\in\Theta} \; \mathrm{KL}(Q\|Q_\theta).$$

信念的渐近集中。 下一个结果确定了真实概率律$\mathbf{P}$下贝叶斯后验分布的渐近支撑集，并给出了 Berk（1966）经典定理的定量版本。

命题2.1。 对于所有闭集 $C \subset \Theta \setminus \Theta^m$，存在常数 $L < \infty$ 使得 [2]

$$\mu_t(C) \le L e^{-t(\rho_C - o_{as}(1))}, \; \forall t \ge 0, \tag{2.2}$$

其中 $\rho_C := \min_{\theta\in C} \mathrm{KL}(Q\|Q_\theta) \; - \min_{\theta\in\Theta} \mathrm{KL}(Q\|Q_\theta) \; > 0$。

[2]符号 $o_{as}(1)$ 表示一个几乎处处收敛到零的正随机变量。

讨论。 </style}后验质量在“坏”集合上的这种指数衰减 C 可由 (2.1) 和三个关键洞见得出。首先，对数似然比渐近上由负的 Kullback–Leibler 散度主导，因此 KL 散度较大的参数会被赋予较低权重。其次， C 与 Θ^m“充分分离”，其 KL 间隙 ρ_C 严格为正，从而为 (2.1) 的分子给出指数界。第三，先验为 Θ^m的每个邻域都赋予正质量，确保 (2.1) 的分母保持正确的指数阶，不会趋于零。

命题*2.1*的证明概要。 根据表达式2.1，后验信念的集中由对数似然比函数 $\theta \mapsto \sum_{i=1}^{t} \log \frac{q_\theta}{(Y_i)\ q(Y_i)}$ 决定。

——

根据大数定律（LLN），对于每个 $\theta \in \Theta$，经缩放的对数似然比函数收敛到 $-\mathbb{E}_Q \left[\log \frac{q}{(Y)\, q_\theta(Y)}\right]$。在我们的条件下，函数族 $\left\{\theta \mapsto t^{-1} \sum_{i=1}^{t} \log \frac{q_\theta}{(Y_i)\, q(Y_i)} :\ t \in \mathbb{N}\right\}$是随机等度连续的（参见 Newey 和 McFadden（1994年）），因此根据阿尔泽拉-阿斯科利定理的一个版本，收敛可扩展为在 Θ上一致成立，即

$$\sup_{\theta \in \Theta} \left| t^{-1} \sum_{i=1}^{t} \log \frac{q_\theta(Y_i)}{q(Y_i)} + KL(Q \,||\, Q_\theta) \right| = o_{as}(1).$$

因此，对于任意 $C \subseteq \Theta$ 可测集，

$$\mu_t(C) \leq \frac{\int_C e^{t(-KL(Q||Q_\theta)+o_{as}(1))} \mu_0(d\theta)}{\int_\Theta e^{t(-KL(Q||Q_\theta)-o_{as}(1))} \mu_0(d\theta)} = \frac{\int_C e^{-t(KL(Q||Q_\theta)-KL^m+o_{as}(1))} \mu_0(d\theta)}{\int_\Theta e^{-t(KL(Q||Q_\theta)-KL^m-o_{as}(1))} \mu_0(d\theta)}$$

其中 $KL^m := \text{是}_{\theta \in \Theta} KL(Q \,||\, Q_\theta)$的最小值。

现在考虑一个闭且 $C \cap \Theta^m = \{\emptyset\}$的集合 $C \subseteq \Theta$。由于 C 和 Θ^m 都是紧集且互不相交，因此 C 上的 KL散度与极小值集合 Θ^m“严格分离”，即 $\rho_C := \inf_{\theta \in C} KL(Q \,||\, Q_\theta) - KL^m > 0$. 因此， $\int_C e^{-t(KL(Q||Q_\theta)-KL^m+o_{as}(1))} \mu_0(d\theta) \leq \int_C e^{-t(\rho_C+o_{as}(1))} \mu_0(d\theta) \leq e^{-t(\rho_C+o_{as}(1))}$。

此外，对于任意此类 $\rho_C > 0$，集合 $\Theta_{\rho_C} := \{\theta \in \Theta :\ KL(Q \,||\, Q_\theta) - KL^m < 0.5\rho_C\}$ 具有非空内部——它是连续函数下某个开集的原像——，并且根据全支撑条件，在先验下具有正概率。因此，对于任意

闭且 $C \cap \Theta^m = \{\emptyset\}$ 的集合 C，

$$\mu_t(C) \le \frac{e^{-t(\rho_C + o_{as}(1))}}{\int_{\Theta_{\rho_C}} e^{-t(KL(Q\|Q_\theta) - KL^m - o_{as}(1))}\mu_0(d\theta)} \le \frac{e^{-t(\rho_C + o_{as}(1))}}{e^{-t(0.5\rho_C - o_{as}(1))}\mu_0(\Theta_{\rho_C})} \le \frac{e^{-t(0.5\rho_C - o_{as}(1))}}{\mu_0(\Theta_{\rho_C})}.$$

因此，取常数为 $1/\mu_0(\Theta_{\rho_C}$ 即可得到所需结果). □

命题2.1的推论。 若存在唯一参数使 KL散度 最小，则后验分布会收敛并集中于该点。[3] 形式化地说，

推论2.1。 若 $\Theta^{\mathrm{m}} = \{\theta^{\mathrm{m}}\}$，则 $\mu_t \Rightarrow \delta_{\theta^{\mathrm{m}}}$ **P**-几乎处处

特别地，推论2.1直接表明，若模型设定正确，主体最终将学到生成数据的真实分布。

最后，若 Θ^m 不是单点集，则贝叶斯后验未必会收敛到一个点。在如此一般的情形下，我们只能断定，它最终会集中于最佳拟合参数的集合 Θ^m。

例 2.1（抛硬币（伯努利模型））。 设 $Y \in \{0,1\}$，并假设模型设定为$Q_\theta(Y = 1) = \theta$，其中 $\Theta \subseteq [0,1]$。真实成功概率 $\theta^\star$ 与模型 θ 之间的 Kullback–Leibler 散度为

$$\mathrm{KL}(\theta^\star\|\theta) := KL(Q_{\theta^\star}\|Q_\theta) \;=\; \theta^\star \log\Big(\frac{\theta^\star}{\theta}\Big) \;+\; (1-\theta^\star)\log\Big(\frac{1-\theta^\star}{1-\theta}\Big).$$

这个例子有助于直观展示两个事实：(i) KL($\theta^\star\|\theta$) 对其参数通常是非对称的（KL($\theta^\star\|\theta$) $\neq$ KL($\theta\|\theta^\star$)）；(ii) 它不是一种度量，其表现也不像 (0，1) 上的欧几里得距离（等值集向边界倾斜，反映了似然曲率）。

为了刻画 Θ^m，不难看出，$\theta \mapsto KL(\theta\|\theta^\star)$ 在 $\theta < \theta^\star$ 时递减，而在 $\theta > \theta^\star$ 时递增。因此，对于 $\Theta \ni \theta^\star$ 的正确设定情形，$\theta^\star$

[3] $\Rightarrow$表示概率测度的弱收敛；δ_θ 是位于 θ 处的点质量。

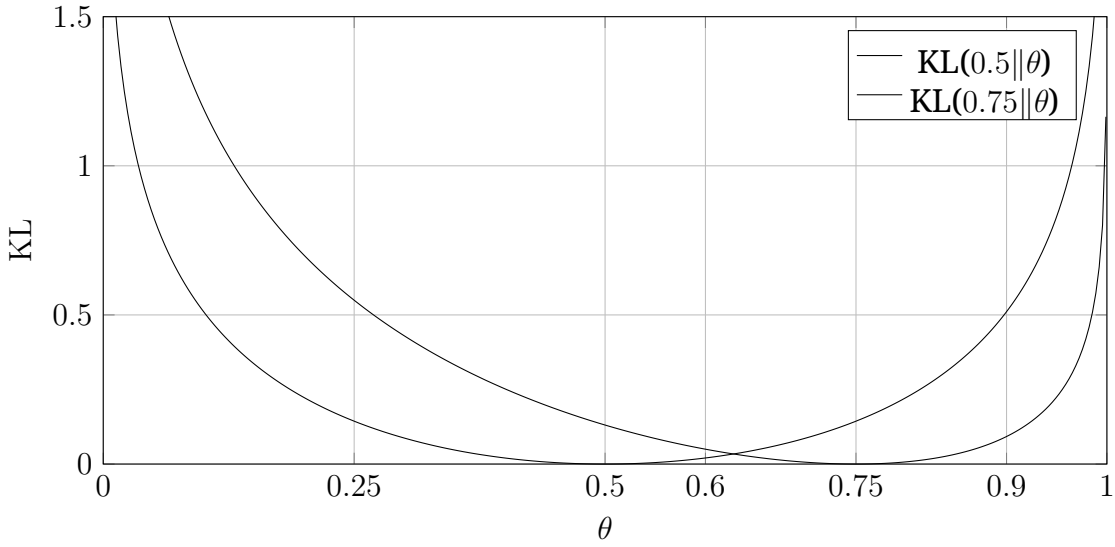


图 1：两个真值 $\theta^\star = 0.5$ 和 $\theta^\star = 0.75$ 对应的伯努利 KL 曲线。

是全局最小值，因此 $\Theta^m = \{\theta^\star\}$。另一方面，如果主体模型设定错误，即 $\theta^\star \notin \Theta$，情况在性质上就大不相同了。事实上， Θ^m 至多由两个元素组成： $\theta^\star$ 从右侧和左侧投影到 Θ 上所得的"投影"。例如，如果 $\theta^\star = 0.50$，那么 Θ^m 可能只包含一个元素——最接近真实参数 $\theta^\star$ 的投影。因此，在这种情况下，尽管模型设定错误，我们仍然有唯一的最小化参数，它代表了模型内的最佳拟合。另一方面，如果这些投影与真实参数的距离相等，则 Θ^m 由这两个元素组成。也就是说，在 KL散度 下，有两个参数对数据的拟合程度同样好。

这种差异对贝叶斯后验的集中有重要影响。在正确设定情形下，由于 $\Theta^{\mathrm{m}} = \theta^\star$，贝叶斯后验几乎必然集中于真值： $\mu_t \Rightarrow \delta_{\theta^\star}\ \mathbf{P}$。在具有唯一最佳拟合参数的设定错误情形下，后验仍会集中。例如，若 $\theta^\star = 0.5$ 但 $\Theta = \{0.25, 0.3, 0.6, 0.9\}$，则 $\Theta^{\mathrm{m}} = \{0.6\}$ 且 $\mu_t \Rightarrow \delta_{0.6}\ \mathbf{P}$。因此，贝叶斯后验仍会收敛，但其极限参数是 KL 投影（伪真值），而非真值。

另一方面，如果最佳拟合参数不具有唯一性，贝叶斯后验分布可能不会收敛。为说明这一点，令 $\theta^\star = 0.5$ 和 $\Theta = \{0.25, 0.75\}$。根据我们的论证

如上所述， $\Theta^m = \{0.25, 0.75\}$。在这种情况下，贝叶斯后验分布 $\mu_t(\cdot)$ 不收敛：

$$\liminf_{t\to\infty} \mu_t(0.25) = 0 \quad \text{and} \quad \limsup_{t\to\infty} \mu_t(0.25) = 1 \quad (\mathbf{P}\text{-a.s.}).$$

事实上，根据贝叶斯法则，

$$\log \frac{\mu_t(0.25)}{\mu_t(0.75)} = \log \frac{\mu_0(0.25)}{\mu_0(0.75)} + \sum_{i=1}^{t} \xi_i$$

其中 $\xi_i := \log\big(Q_{0.25}(Y_i)/Q_{0.75}(Y_i)\big)$。注意， ξ_i 是独立同分布的，且 $\mathbb{E}[\xi_i] = 0$ （这两个模型在 KL 意义下不相上下）。因此，后验分布的对数似然比是一个零漂移随机游走。

因此，根据随机游走的性质，$\big(\log \mu_t(0.25)/\mu_t(0.75)\big)_t$ 将以 $\mathbf{P}$-几乎必然的方式取到任意大的正值和任意大的负值，并且会无限多次回到零。这意味着，贝叶斯后验将在几乎把全部概率质量赋予 0.25 的时期与几乎把全部概率质量赋予 0.75 的时期之间不断振荡，始终无法稳定下来。此外，这一结果与具体的先验无关。 ♦

2.2 背景介绍

贝叶斯后验集中的研究源于 Schwartz（1965）的开创性工作。他在正确设定的条件下，为贝叶斯模型建立了首个一般性一致性定理。Schwartz 指出，如果先验对真实分布的每个 Kullback–Leibler（KL）邻域都赋予正质量，且存在能够将真实分布与其补集区分开来的一致一致检验，那么后验分布就会在弱拓扑下集中于数据生成分布附近。Schwartz 的结果明确表明，后验相合性不仅取决于先验的“支撑集”，也取决于频率学派检验性质；这一观点随后成为现代贝叶斯渐近理论的基础。

Berk（1966）将这一分析扩展到了模型设定错误的参数化情形。当数据生成分布不属于假定模型时，他证明后验分布仍会

集中，但此时会集中在真实分布的*KL*投影附近，也就是模型中使 $KL(Q||Q_\theta)$最小的元素。因此，模型设定错误会影响后验学习的极限点，但不会改变后验分布会集中的事实。抛硬币示例中信念的不收敛问题便源于这篇论文。后来，Bunke 和 Milhaud（1998）对 Berk 的参数模型设定错误情形作了更细致的处理，他们通过研究一般损失函数下贝叶斯估计量的渐近行为（不仅是后验分布本身），进一步强化了 Berk 的结论。Yamada（1976）还将 Berk 的结果推广到多维情形，并以更现代、清晰的方式证明了该结果，与此处给出的证明概要相似。Shalizi（2009）依靠信息论中的渐近均分性质（Shannon-McMillan-Breiman），将这些结果推广到相关数据。与 Berk 关于模型设定错误模型的结果相对应，White（1982）建立了频率学派的结果，证明最大化平均对数（准）似然 $\frac{1}{t}\sum_{i=1}^{t}\log q_\theta(Y_i)$ 的最大（准）似然估计量会收敛到 $\min_\theta \mathrm{KL}(Q||Q_\theta)$。因此，在独立同分布外生情形下，贝叶斯和频率学派方法以相同的参数为目标：即在 KL 散度下提供最佳拟合的参数。

–

命题2.1中使用的正则条件受到Schwartz（1965年）中条件的启发。连续性条件、支配条件和紧致性条件是建立对数似然比函数一致收敛于KL散度所必需的。

通过检查该命题的证明可以明显看出，参数空间的有限维要求并未被使用。事实上，Schwartz（1965年）的结果允许参数为无限维参数。[4] 然而，在无限维情形下，验证正则性假设并非易事。

正如 Diaconis 和 Freedman（1986）的开创性论文所指出的，Schwartz（1965）提出的“先验支撑集”条件在无限维模型中远非自动成立。原因在于，该条件要求先验对真值在 <style id='1'>KL散度下的</style>开邻域赋予正概率。

[4]关于无限维空间的其他较新参考文献包括Ghosal、Ghosh和van der Vaart（2000年），以及Shen和Wasserman（2001年）；关于更为详尽的讨论和其他参考文献，参见Ghosal和van der Vaart（2017年）。

然而，他们的例子表明，先验即使在弱拓扑下稠密，也可能无法对真实分布的足够小的 KL 邻域赋予任何质量。在这种情况下，先验被称为在 **P** 处薄，贝叶斯更新无法恢复真值，因为后验分布无法在真实概率分布周围的 KL 球中累积似然权重。与之相对，厚先验是指对数据生成律的每个 KL 邻域都赋予严格正质量的先验。Barron、Schervish和 Wasserman（1999）明确阐述了这一条件的必要性，他们证明，在温和的正则条件下，KL厚性不仅是充分条件（如 Schwartz 定理所示），而且对于非参数情形下的后验一致性而言，本质上也是必要的。由此，Diaconis 和 Freedman（1986）的例子澄清了 Schwartz 支持条件在数学上是精确的：非参数贝叶斯一致性根本上依赖于 KL厚性，而不仅仅是弱支撑或拓扑支持。

Kleijn和van der Vaart（2006年）将Schwartz的结果扩展到设定错误的模型，从而将Berk定理推广到模型设定错误下的非参数模型。与Schwartz（1965年）的正确设定情形不同，瓶颈不仅在于“KL厚性”条件，还在于极小值集合的“良好分离”——这种良好分离依赖于围绕伪真分布构造一致一致检验，而其局部几何可能更难以控制。

3 内生学习：单个主体

3.1 设定与均衡

单个主体选择一个行动 $a \in \mathbb{A}$。结果取值于 $\mathbb{Y}$。真实后果核为 $Q:\ \mathbb{A} \to \Delta\mathbb{Y}$。主体考虑一个（可能设定错误的）参数模型$\{Q_\theta:\ \theta \in \Theta\}$，其中每个 $Q_\theta:\ \mathbb{A} \to \Delta\mathbb{Y}$。每期收益为 $\pi:\ \mathbb{A} \times \mathbb{Y} \to \mathbb{R}$。

假设（单个主体）。

- 紧性。 $\mathbb{A}$ 和 Θ 是欧几里得空间中的非空紧致子集； $\mathbb{Y}$ 是欧几里得空间的一个博雷尔子集。

- 核的支配与几乎处处连续性。存在一个 σ-有限博雷尔测度 ν，定义在$\mathbb{Y}$ 上，使得对于所有(θ, a)， $Q(\cdot \mid a) \ll \nu$ 且 $Q_\theta(\cdot \mid a) \ll \nu$，其密度分别为 $q(\cdot \mid a)$和$q_\theta(\cdot \mid a)$。对于 ν-几乎处处的 y，映射 $a \mapsto q(y \mid a)$和$(\theta, a) \mapsto q_\theta(y \mid a)$均连续。

- 似然比包络。存在可测的 $M\colon \mathbb{Y} \to [1, \infty)$和 $r\colon \mathbb{Y} \to [0, \infty)$，使得对于所有$(\theta, a)$以及 ν-几乎处处的 y，

$$q_\theta(y \mid a)\, M(y)^{-1} \;\le\; q(y \mid a) \;\le\; q_\theta(y \mid a)\, M(y), \qquad q(y \mid a) \le r(y),$$

  且 $\int M(y)\; r(y)\; \nu(dy) < \infty$。

- 收益的支配性与几乎处处连续性。对于 ν-几乎处处的 y，$(a,\ y) \mapsto \pi(a,\ y)$ 连续；存在 $h_\pi\colon \mathbb{Y} \to [0, \infty)$，使得对所有 a 和 $\int h_\pi(y)\; M(y)\; r(y)\; \nu(dy) < \infty$，都有 $|\pi(a,\ y)| \le h_\pi(y)$。

这些条件适用于离散或连续的 $\mathbb{A}$ 和 $\mathbb{Y}$。它们确保 Kullback–Leibler 散度的极小值集合以及最优行动对应的存在，并且还为下文介绍的动态环境提供所需的一致大数定律。除非另有说明，本章节始终采用这些基本假设。

最优　　ty 对应。对于 $\theta \in \Theta$，定义行动 a 的期望收益

$$U(a, \theta) \;:=\; \int_{\mathbb{Y}} \pi(a, y)\, Q_\theta(dy \mid a).$$

给定信念 $\mu \in \Delta\Theta$，最优行动对应为

$$F(\mu) \;:=\; \arg\max_{a\in\mathbb{A}} \int_{\Theta} U(a,\theta)\,\mu(d\theta).$$

在基本假设下， F 非空、取紧致值，且关于 μ 上半连续（UHC）。此外，以下结果适用于任何具有这些性质的行动对应。因此，尽管模型设定错误学习文献通常通过最大化推导出 F ，但本分析也涵盖非最优选择规则。

依赖行动的 KL 散度。在行动 a 下，真实核与模型 Q_θ 之间的 Kullback–Leibler 散度为

$$K(\theta,a) \;:= KL(Q(\cdot \mid a)\|Q_\theta(\cdot \mid a)) = \int_{\mathbb{Y}} \log\!\Big(\tfrac{q(y|a)}{q_\theta(y|a)}\Big)\, q(y \mid a)\,\nu(dy).$$

对于任意行动分布 $\sigma \in \Delta\mathbb{A}$，定义 KL最小化集合

$$\Theta^{\mathrm{m}}(\sigma) \;:=\; \arg\min_{\theta\in\Theta} \int_{\mathbb{A}} K(\theta,a)\,\sigma(da).$$

直观地说， $\Theta^{\mathrm{m}}(\sigma)$ 收集了这样一些参数值：其对应的核最能拟合由长期行动频率 σ所诱导的真实数据。在我们的假设下， Θ^{m} 非空、取紧致值，且关于 σ 上半连续。

我们的目标是理解当主体反复面对这一决策问题并更新其关于参数空间的信念时，会发生什么。现在我们引入一个均衡概念，并在第 3.2 节正式构建动态框架，说明均衡代表这一学习过程的稳态。行动决定主体观察到的反馈；反馈决定其信念；而在模型设定错误的情况下，这些信念通常取决于行动的抽样方式，从而在行动与信念之间形成反馈循环。因此，这一稳态可以刻画为该过程的一个不动点，

即使是在单代理设定中。

定义 3.1。 如果存在一个信念 $\mu \in \Delta\Theta$，使得：概率测度 $\sigma \in \Delta\mathbb{A}$是该单代理问题的**伯克–纳什均衡**，

1. 最优性： $\mathrm{supp}(\sigma) \subseteq F(\mu)$。

2. 信念限制：$\mathrm{supp}(\mu) \subseteq \Theta^{\mathrm{m}}(\sigma)$。

例 3.1。 努力回报（过度自信）。主体选择 $a \in [0,\infty)$ 以最大化$\mathbb{E}[\pi(a,\ Y)]$，其中 $\pi(a,\ y) = y - c(a)$，其中 c 可微、严格凸，满足 $c(0) = c'(0) = 0$，且 $c'(a) \to \infty$。结果满足线性—高斯方程

$$Y = (\alpha^* + a)\theta^* + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0,1),$$

对于真实数据生成过程，以及

$$Y = (\alpha + a)\theta + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0,1), \quad \theta \in [0,\overline{\theta}],$$

对于感知模型，其中 $\alpha > 0$ 是固定的感知能力（不会被学习），且 θ 足够大（此处只需满足 $\theta \geq \theta^*$）。我们关注过度自信，即 $\alpha > \alpha^*$。这些方程导出核 $Q(\cdot \mid a) = \mathcal{N}((\alpha^*+a)\theta^*,\ 1)$ 和 $Q_\theta(\cdot \mid a) = \mathcal{N}((\alpha+a)\theta,\ 1)$。依赖行动的Kullback–Leibler散度为

$$K(\theta, a) = \tfrac{1}{2}\Big((\alpha + a)\theta - (\alpha^* + a)\theta^*\Big)^2.$$

对于任意行动分布 $\sigma \in \Delta[0,\infty)$，$\int K(\theta,a)\,\sigma(da)$ 关于 θ 严格凸，且具有唯一的最小化解；因此，我们可以只考虑退化信念（单个参数θ）。特别地，对于纯策略 $\sigma = \delta_a$，唯一的 KL 最小化解为

$$\theta^{\mathrm{m}}(\delta_a) = \theta^* + \theta^*\,\frac{\alpha^* - \alpha}{\alpha + a},$$

在过度自信（低估努力回报）的情况下，该值严格低于 θ^* ；偏差随 a 增大而缩小，并在 $a \to \infty$ 时消失。

给定信念 θ，最优性要求 $c'(a) = \theta$。信念限制要求该信念与由 a 导出的 KL 投影一致： $\theta = \theta^{\mathrm{m}}(\delta_a)$。因此，伯克–纳什均衡是满足以下条件的行动 a

$$c'(a) = \theta^{\mathrm{m}}(\delta_a) = \theta^* + \theta^* \frac{\alpha^* - \alpha}{\alpha + a}. \tag{3}$$

由于 $c'(0) = 0$ 且 $c'(a) \to \infty$，而（3）式右侧连续、递增且有上界 θ^*，因此解存在。由于（3）式两侧均递增，可能存在多个解。一个保证唯一性的简单充分条件是 c' 为凸函数。在过度自信的情况下，对所有 a，均有 $\theta^{\mathrm{m}}(\delta_a) < \theta^*$ ，因此（3）式的任意解 a 都满足 $a < (c')^{-1}(\theta^*)$（低于基于真实信念的基准）。随着错误认知消失（$\alpha \downarrow \alpha^*$） ， $\theta^{\mathrm{m}}(\delta_a) \uparrow \theta^*$ 逐点收敛，且伯克–纳什解（集）收敛至 $(c')^{-1}(\theta^*)$。 ♦

示例 3.2（逆向选择下的交易）。 买方报出价格 $a \in \mathbb{R}_+$。结果为$\mathbb{Y} = \{\square\} \cup \{v_1, \ldots, v_K\}$，其中 $\square$ 表示“不交易”，而 $v_j \in \mathbb{R}$ 表示交易发生时买方实现的价值。不交易时收益为 0，交易时收益为 $v - a$ ，因此交易以买方报出的价格进行。设 S 为卖方价值，$V \in \{v_1, \ldots, v_K\}$ 为买方价值，且 p 为它们的联合概率分布。卖方遵循索要其估值的弱占优策略，因此我们将重点放在买方身上。设

$$F_S(a) := \Pr(S \le a) \quad \text{and} \quad p_j(a) := \Pr\big(V = v_j \mid S \le a\big), \quad j = 1, \ldots, K.$$

真实后果核在 $\square$ 上赋予质量 $1 - F_S(a)$，并以概率 $F_S(a)$ 从所选 概率质量函数 $p(a) = (p_1(a) \ldots, p_K(a))$ 中抽取V。

模型设定错误。 买方将 $F_S(a)$ 视为正确（例如从要价中学习到的），但忽略了价值选择——等价地，他的行为仿佛 S 和 V 相互独立，因此

交易卖方的构成不包含关于 V 的任何信息。形式上，他假定存在一个在行动不变的 $\theta = (\theta_1, \ldots, \theta_K) \in \Delta(\{v_1, \ldots, v_K\})$，并在每次发生交易时使用它。因此，该感知核将 $1 - F_S(a)$ 分配给 $\square$，并以概率 $F_S(a)$ 以概率 θ_j 选择 $V = v_j$。

KL最小化解。 固定 a，满足 $F_S(a) > 0$。由于真实核与感知核在 $\square$ 上一致，KL散度可简化为

$$K(a, \theta) = F_S(a)\Big\{C(a) + \sum_{j=1}^{K} p_j(a) \log \frac{p_j(a)}{\theta_j}\Big\},$$

其中 $C(a)$ 与 θ 无关。对 θ 最小化可得唯一的 最小化解 $\theta^{\mathrm{m}}(\delta_a) = p(a)$。因此，

$$\mathbb{E}_{\theta^{\mathrm{m}}(\delta_a)}[V] = \sum_{j=1}^{K} v_j \, p_j(a) = \mathbb{E}[V \mid S \le a].$$

对于任意混合价格 $\sigma \in \Delta(\mathbb{R}_+)$，若 $\int F_S(a)\, \sigma(da) > 0$，则 KL 目标函数在 a 上以严格正权重线性加权，并且在单纯形上严格凸；因此，KL最小化解 $\theta^{\mathrm{m}}(\sigma)$ 是单元素集。所以，对于最优性和均衡，我们可以只考虑 θ 上的退化 信念。

退化信念下的最优定价。 给定 θ，买方感知到的预期收益为其报出价格 a时的

$$U(a, \theta) = F_S(a)\big(\mathbb{E}_\theta[V] - a\big).$$

具有严格正交易量的纯行动 伯克–纳什均衡是指满足$F_S(a^\star) > 0$ 且满足以下条件的价格 $a^\star$

$$a^\star \in \arg\max_{a \ge 0} \; F_S(a)\Big(\mathbb{E}_{\theta^{\mathrm{m}}(\delta_{a^\star})}[V] - a\Big).$$

因此，均衡是上述表达式的一个不动点，这反映了买方会使用由该价格本身所生成的选定分布，对 V形成平均信念。在考虑偏离时，他将这一信念视为固定，只比较偏离会如何改变成本和交易概率。在这种评估下， $a^\star$

至少带来与其他任何价格一样高的感知收益。 ♦

3.2 动态

动态设定。 固定一个全支撑先验 $\mu_0 \in \Delta\Theta$。对于时期 $t = 1$，2、.....：

1. 给定贝叶斯后验 μ_t，主体选择 $a_t \in F(\mu_t)$。如果 $F(\mu_t)$ 是多值的，则一个可测选择 $\phi_t(\cdot \mid \mu_t) \in \Delta\mathbb{A}$ 满足 $\phi_t(F(\mu_t) \mid \mu_t) = 1$。

2. 从真实核中抽取一个结果 y_t： $y_t \sim Q(\cdot \mid a_t)$。

3. 根据贝叶斯法则更新后验分布：

$$\mu_{t+1}(B) \;=\; \frac{\int_B q_\theta(y_t \mid a_t)\,\mu_t(d\theta)}{\int_\Theta q_\theta(y_t \mid a_t)\,\mu_t(d\theta)}, \qquad B \subseteq \Theta \text{ Borel}$$

在支配性／似然比界假设下，该更新定义良好。

概率律。 先验 μ_0、选择规则 $\{\phi_t\}$和真实核 Q在行动与结果的无限序列上诱导出一个概率测度 $\mathbf{P}$ 。下文所有几乎必然陈述均相对于$\mathbf{P}$而言。当某一结果取决于先验时，我们写作 $\mathbf{P}^{\mu_0}$以明确表示这种依赖关系。

经验行动分布。对于 $t \geq 1$，令

$$\sigma_t(B) \;:=\; \frac{1}{t}\sum_{\tau=1}^{t} \mathbf{1}_B(a_\tau), \qquad B \subseteq \mathbb{A} \text{ Borel}.$$

引理 3.1 （KL 最小化解附近的后验集中）。 几乎必然地，以下结论成立：如果 $\sigma_t \Rightarrow \sigma$，那么对于每个闭集 $C \subseteq \Theta$，使得 $\Theta^{\mathrm{m}}(\sigma) \cap C = \{\emptyset\}$， $\mu_t(C) \;\to\; 0$。

该引理的证明是将命题2.1改编到具有内生学习的情形。

命题 3.1。 几乎必然，以下结论成立：如果 $a_t \to a^\star$，那么 $\delta_{a^\star}$ 是一个伯克–纳什均衡；也就是说，存在 $\mu^\star \in \Delta\Theta$ 使得 $\mathrm{supp}(\mu^\star) \subseteq \Theta^{\mathrm{m}}(\delta_{a^\star})$ 满足 $a^\star \in F(\mu^\star)$。命题 *3.1* 的证明概要。 在事件 $\{a_t \to a^\star\}$ 上进行讨论。于是 $\sigma_t \Rightarrow \delta_{a^\star}$。根据引理 3.1，对于每个开集 $U \supset \Theta^{\mathrm{m}}(\delta_{a^\star})$，我们有 $\mu_t(U) \to 1$。因此，(μ_t) 的任意弱极限点$\mu^\star$ 都满足 $\mathrm{supp}(\mu^\star) \subseteq \Theta^{\mathrm{m}}(\delta_{a^\star})$。由于 $a_t \in F(\mu_t)$，且 F 非空、取值为紧致集并且上半连续，取极限可得 $a^\star \in F(\mu^\star)$。

□

3.3 混合行动中的稳态（有限 $\mathbb{A}$）

我们已经证明，如果行动收敛到纯行动 $a^\star$，那么 $\delta_{a^\star}$ 是伯克–纳什均衡。现在我们来讨论混合稳态：即作为混合行动博弈过程极限而出现的分布 $\sigma \in \Delta\mathbb{A}$ 。本小节中， $\mathbb{A}$是有限集，不过这一条件可以放宽。

路径 I：收益扰动与预期混合行动（有限 $\mathbb{A}$）。假设在每个 t下，收益都加入独立同分布的行动特定冲击（$\eta_t(a)$）$_{a\in\mathbb{A}}$ ，这些冲击在 t 之间相互独立，且独立于历史，并且服从在 $\mathbb{R}^{|\mathbb{A}|}$上绝对连续（具有连续密度）的共同分布。这保证了随机效用选择概率定义良好，且关于信念连续。

Give对于信念 $\mu_t \in \Delta\Theta$，定义预期混合行动（选择概率）

$$\sigma_t^{\mathrm{I}}(a) \;:=\; \mathbb{P}\Big(a \in \arg\max_{a'\in\mathbb{A}} \Big\{ \textstyle\int_\Theta U(a',\theta)\,\mu_t(d\theta) \;+\; \eta_t(a') \Big\}\Big), \qquad a \in \mathbb{A}.$$

等价地，存在一个连续映射 $\sigma^{\mathrm{I}}\colon\ \Delta\Theta \to \Delta\mathbb{A}$ ，使得对所有 t都有$\sigma_t^{\mathrm{I}} = \sigma^{\mathrm{I}}(\mu_t)$。

定义 3.2。 若存在 $\mu \in \Delta\Theta$ 使得 $\mathrm{supp}(\mu) \subseteq \Theta^{\mathrm{m}}(\sigma)$ 且 $\sigma = \sigma^{\mathrm{I}}(\mu)$，则分布 $\sigma \in \Delta\mathbb{A}$ 是预期策略下的伯克–纳什均衡。

当收益扰动较小时，意向策略中的伯克–纳什均衡近似于标准伯克–纳什均衡：随机效用固有的平滑效应会随着

噪声减小而消失，因此预期混合行动趋近于最佳回应，信念限制保持不变。[5]

命题 3.2。 几乎必然，以下蕴含关系成立：如果 $\sigma_t^{\mathrm{I}} \to \sigma \in \Delta\mathbb{A}$，那么 σ是预期策略下的伯克–纳什均衡。

命题 *3.2* 的证明概要。 (1) 从意向到经验。 给定历史，a_t 根据 σ_t^{I}抽取。鞅强大数定律给出 $\sigma_t - \frac{1}{t}\sum_{\tau\le t}\sigma_\tau^{\mathrm{I}} \to 0$ 几乎必然成立；因此， $\sigma_t^{\mathrm{I}} \to \sigma$蕴含 $_\sigma_t \to \sigma$ 几乎必然成立。

(2) 后验集中。 根据引理 3.1，以概率 1：如果 $\sigma_t \to \sigma$，那么对于每个开集 $U \supset \Theta^{\mathrm{m}}(\sigma)$，我们有 $\mu_t(U) \to 1$。因此，(μ_t) 的任意弱极限 $\mu^\star$ 都满足 $\mathrm{supp}(\mu^\star) \subseteq \Theta^{\mathrm{m}}(\sigma)$。

(3) 意向选择的连续性。 由于 σ^{I} 是连续的且 $\sigma_t^{\mathrm{I}} = \sigma^{\mathrm{I}}(\mu_t) \to \sigma$，沿着满足 $\mu_t \Rightarrow \mu^\star$ 的任意子序列，我们得到 $\sigma = \sigma^{\mathrm{I}}(\mu^\star)$。

结合 (2) 和 (3) 即得所述结论。 □

路径 II：无扰动，且由经验博弈过程得到稳态（有限 $\mathbb{A}$）。我们考虑经验行动分布 σ_t 收敛的路径，并刻画由此蕴含的稳态。

定义 3.3。若对于任意 $\theta, \theta' \in \Theta^{\mathrm{m}}(\sigma)$，均有$Q_\theta(\cdot \mid a) = Q_{\theta'}(\cdot \mid a)$ 对所有 $a \in \mathrm{supp}(\sigma)$ 成立，则称在 $\sigma \in \Delta\mathbb{A}$ 处满足弱识别。

在弱识别下，如果多个参数在 σ 下并列为 KL 最小化解，那么它们对 $\mathrm{supp}(\sigma)$ 中的每个行动都会作出相同的预测。因此，并列不会影响路径上预测或最佳回应——只有路径外反事实可能有所不同。弱识别

[5]形式化地，设 $\{\sigma_\varepsilon^{\mathrm{I}}, \mu_\varepsilon\}_{\varepsilon>0}$ 为扰动尺度 $\varepsilon \downarrow 0$下意向策略中的伯克–纳什均衡，且满足 $\sigma_\varepsilon^{\mathrm{I}} = \sigma_\varepsilon^{\mathrm{I}}(\mu_\varepsilon)$。若 $\mathbb{A}, \Theta$具有紧性， U连续，支配／包络条件确保 K连续， $\Theta^{\mathrm{m}}(\cdot)$上半连续且非空，并满足消失噪声性质$\sigma_\varepsilon^{\mathrm{I}}(\mu) \Rightarrow \mathrm{Prob}(F(\mu))$，则任何聚点 $(\sigma, \mu^\star)$，只要满足 $\sigma_\varepsilon^{\mathrm{I}} \Rightarrow \sigma, \mu_\varepsilon \Rightarrow \mu^\star$，就是一个（标准的）伯克–纳什均衡：$\mathrm{supp}(\sigma) \subseteq F(\mu^\star)$ 且 $\mathrm{supp}(\mu^\star) \subseteq \Theta^{\mathrm{m}}(\sigma)$。反过来，每个伯克–纳什均衡都能在 $\varepsilon \to 0$时由意向策略均衡逼近，这需要额外的正则性条件（例如扰动族的丰富性和并列情形处理）；这与 Harsanyi 净化定理相对应。

弱于识别，后者要求定义 3.3 中的性质对所有行动都成立，而不仅仅是对 σ 支撑集中的行动成立。

命题 3.3。几乎必然，以下蕴含关系成立：若 $\sigma_t \to \sigma \in \Delta\mathbb{A}$，且弱识别在 σ 处成立，则 σ 是一个伯克–纳什均衡。

命题 *3.3* 的证明概要。由引理 3.1，沿着 $\sigma_t \to \sigma$，对所有开集 $U \supset \Theta^{\mathrm{m}}(\sigma)$，都有 $\mu_t(U) \to 1$。因此，任何弱极限 μ^∞ 都满足 $\mathrm{supp}(\mu^\infty) \subseteq \Theta^{\mathrm{m}}(\sigma)$。固定 $a \in \mathrm{supp}(\sigma)$，并取满足 $a_t = a$ 的各期；沿该子序列，最优性给出 $a \in F(\mu_t)$，并由上半连续性传递到极限。在弱识别成立的条件下，所有支撑在 $\Theta^{\mathrm{m}}(\sigma)$ 上的信念，都会在 $\mathrm{supp}(\sigma)$ 上诱导出 相同的结果分布，因此期望收益也相同。所以，存在一个信念 μ^∞，对每个 $a \in \mathrm{supp}(\sigma)$ 都适用，从而得到伯克–纳什均衡。

□

正如证明概要所示，若不存在弱识别，我们就必须允许均衡中的信念依赖于行动。这促使我们对均衡作如下推广。**定义 3.4.** 若对每个$a \in \mathrm{supp}(\sigma)$，都存在一个信念 $\mu_a \in \Delta\Theta$，使得 $\mathrm{supp}(\mu_a) \subseteq \Theta^{\mathrm{m}}(\sigma)$ 且 $a \in F(\mu_a)$，则分布 $\sigma \in \Delta\mathbb{A}$ 是广义伯克–纳什均衡。**命题 3.4.** 几乎必然，以下蕴含关系成立：若 $\sigma_t \to \sigma \in \Delta\mathbb{A}$，则 σ是一个广义伯克–纳什均衡。

讨论。 路径 I（扰动）与 Harsanyi 对混合策略的解释一致，并导出意向策略中的伯克–纳什均衡（在噪声消失时，则导出标准 Berk–Nash）。路径 II（无扰动）将 σ 视为一种经验频率。在弱识别下，任何极限 σ 都是伯克–纳什均衡；若不存在弱识别，则支撑集中的每个行动都可以由（可能不同的）信念加以论证，从而得到广义伯克–纳什稳态。

例 3.3（需求未知的垄断者）。 一位垄断者选择价格 $a \in \mathbb{A} = \{2, 10\}$。这些价格下的真实平均销量为 $\phi(2) = 34$ 和 $\phi(10) = 2$。销量满足$Y = \phi(a) + \varepsilon_0$，其中 $\mathbb{E}[\varepsilon_0 \mid a] = 0$ 和 $Var(\varepsilon_0 \mid a) < \infty$。垄断者所感知的

模型是方差固定的高斯模型，$Y = \phi_\theta(a) + \varepsilon,\ \varepsilon \sim \mathcal{N}(0,1),\ \phi_\theta(a) = \alpha - \beta a, \theta = (\alpha, \beta) \in \Theta := [33, 40] \times [3, 3.5]$。模型设定错误的原因是，两个真实均值对应的完美拟合参数为 $\theta_0 = (42, 4) \notin \Theta$。

我们稍微滥用一下记号，令 σ 表示价格 10 的概率，$1-\sigma$ 表示价格 2 的概率。在方差固定的高斯感知似然下，KL准则可化简为

$$K(\sigma, \theta) = \tfrac{1}{2}(1-\sigma)\big(34 - (\alpha - 2\beta)\big)^2 + \tfrac{1}{2}\sigma\big(2 - (\alpha - 10\beta)\big)^2. \tag{3.1}$$

给定 $\theta = (\alpha,\ \beta)$，定义感知收益 $R_\theta(a) = a(\alpha - \beta a)$。

首先，我们证明纯价格不可能是均衡。如果 $\sigma = 0$，最小化 (3.1) 会强制满足 $\alpha - 2\beta = 34$，而在 Θ 中，这将得到 $(\alpha,\ \beta) = (40,\ 3)$。于是

$$R_\theta(10) - R_\theta(2) = 10(\alpha - 10\beta) - 2(\alpha - 2\beta) = 8\alpha - 96\beta = 32 > 0,$$

因此，无论对该极小值集合采用何种信念，10 的收益（利润）都高于 2。如果 $\sigma = 1$，最小化 (3.1) 会强制满足 $\alpha - 10\beta = 2$，这意味着 $\beta \in [3.1, 3.5]$ 且 $\alpha = 2 + 10\beta \in [33, 37]$；于是 $8\alpha - 96\beta = 16 - 16\beta < 0$，因此无论对 (3.1) 的极小值集合采用何种信念，2 的收益都高于 10。因此，任何均衡都必须在两种价格之间进行严格混合。

对于内部的 $0 < \sigma < 1$，令 $\theta^{\mathrm{m}}(\sigma) = (\alpha(\sigma)\ \beta(\sigma))$ 为 (3.1) 的（唯一）极小值点。由于两个价格都属于支撑集，最优性要求无差异：

$$R_{\theta^{\mathrm{m}}(\sigma)}(2) = R_{\theta^{\mathrm{m}}(\sigma)}(10) \iff 2(\alpha - 2\beta) = 10(\alpha - 10\beta) \iff \alpha = 12\,\beta. \tag{3.2}$$

当 σ 较小时，极小值点位于边界 $\alpha(\sigma) = 40$，其中

$$\beta(\sigma) = \frac{3 + 92\sigma}{1 + 24\sigma}. \tag{3.3}$$

在 $\alpha = 40$ 处施加 (3.2) 得到 $\beta = 10/3$。令 (3.3) 等于 $10/3$，可得

$$\frac{3 + 92\sigma}{1 + 24\sigma} = \frac{10}{3} \implies \sigma^* = \frac{1}{36}.$$

在此 σ^*, $\theta^{\mathrm{m}}(\sigma^*) = \big(40,\ \ 10/3\big)$ 以及 $R_{\theta^{\mathrm{m}}(\sigma^*)}(2) = R_{\theta^{\mathrm{m}}(\sigma^*)}(10)$，因此 σ^* 是唯一的伯克–纳什均衡。[6]

命题 3.1 和不存在纯行动均衡这一事实意味着，在此例中行动不会收敛。由于这里满足弱识别（事实上，我们有更强的条件：对于任意混合行动，KL散度的极小化解都是唯一的），根据命题 3.3，我们可以将该均衡解释为行动的稳态频率——垄断者在不同价格之间交替，但实际上从不进行混合。行动频率是否收敛到均衡则是另一回事，将在第5节讨论。

3.4 前瞻性主体（非短视）

现在我们考虑一位贴现因子为 $\delta \in (0,\ 1)$ 的前瞻性主体，他选择行动以最大化期望收益的贴现总和。

动态规划。 令 $\bar{Q}_\mu(\cdot \mid a) := \int_\Theta Q_\theta(\cdot \mid a)\,\mu(d\theta)$ 表示信念 μ 下的预测核，并令 $B(a,\ y,\ \mu)$ 表示观察到行动 a 所产生的结果 y 后的贝叶斯更新。价值函数满足贝尔曼方程

$$W(\mu) \ = \ \max_{a \in \mathbb{A}} \int_{\mathbb{Y}} \Big(\pi(a, y) + \delta\, W(B(a, y, \mu)) \Big)\, \bar{Q}_\mu(dy \mid a),$$

[6]当 σ 超过 $1/16$ 后，KL最小化解会移动到边界线段，其中对于$\sigma \in [1/16,\ 1/4]$，$(\alpha, \beta) = (40, 3.5)$；对于 $\sigma \geq 1/4$, $(\alpha, \beta) = (41 - 4\sigma, 3.5)$。在这两种情况下，$\alpha - 12\beta < 0$，这意味着 $R(10) - R(2) = 8(\alpha - 12\beta) < 0$；价格 2 严格占优，因此在这样的 ρ 下不可能存在均衡。

相应的最优行动对应为

$$F_\delta(\mu) \;:=\; \arg\max_{a\in\mathbb{A}} \int_{\mathbb{Y}} \Big(\pi(a,y) + \delta\, W(B(a,y,\mu))\Big)\, \bar{Q}_\mu(dy \mid a).$$

稳态视角（路径 II ＋ 弱识别）。我们考虑经验行动分布 σ_t 收敛的路径，并施加弱识别条件（定义 3.3）。

命题 3.5。几乎必然，以下蕴含关系成立：若 $\sigma_t \to \sigma$ 且在 σ 处满足弱识别条件，则 σ 是伯克–纳什均衡。

命题 *3.5* 的证明概要。根据命题 3.3 的证明（应用于满足 $\sigma_t \to \sigma$ 的前瞻路径），我们得到一个单一信念 $\mu^\infty \in \Delta\Theta$，其 $\mathrm{supp}(\mu^\infty) \subseteq \Theta^{\mathrm{m}}(\sigma)$，使得对于每个 $a \in \mathrm{supp}(\sigma)$，都有 $a \in F_\delta(\mu^\infty)$。接下来只需从 F_δ 推出 F。

根据在 σ 处的弱识别条件，$B(a,\ y,\ \mu^\infty) = \mu^\infty$，对于 $Q_{\mu^\infty}(\cdot \mid a)$-几乎所有 y 以及所有 $a \in \mathrm{supp}(\sigma)$ 均成立。因此，对于这样的 a，

$$\int_{\mathbb{Y}} \big(\pi(a,y) + \delta W(B(a,y,\mu^\infty))\big)\, \bar{Q}_{\mu^\infty}(dy \mid a) = \int_{\mathbb{Y}} \pi(a,y)\, \bar{Q}_{\mu^\infty}(dy \mid a) + \delta W(\mu^\infty),$$

因此，延续项 $\delta W(\mu^\infty)$ 对所有行动都是相同的。因此，最大化动态目标等同于在 $\mathrm{supp}(\sigma)$ 上最大化短视预期收益，即 $F_\delta(\mu^\infty) = F(\mu^\infty)$。由于 $a \in F_\delta(\mu^\infty)$，可得对所有 $a \in \mathrm{supp}(\sigma)$，都有 $a \in F(\mu^\infty)$。结合 $\mathrm{supp}(\mu^\infty) \subseteq \Theta^{\mathrm{m}}(\sigma)$，这表明 σ 是伯克–纳什。

□

解释。 在弱识别下，在假定的稳态处，学习没有边际价值：对于支持集上的行动，下一期信念几乎必然等于当前信念，因此探索不会改变激励。因此，在极限处，前瞻性最优性与短视最优性一致，且该稳态是伯克–纳什均衡。

3.5 文献注释

Esponda 和 Pouzo（2016）提出了一个多主体错设学习框架，既刻画系统性偏差，也刻画由复杂性或信息约束造成的准确环境表征能力的局限。他们引入了伯克–纳什均衡，并证明，只要行为收敛，就会收敛到伯克–纳什均衡。本文所用的定义和结果将他们的环境专门化为单个主体；正如我们在后文所示，这些定义和结果也可以直接推广到多个主体。Esponda 和 Pouzo（2016）将行动空间限定为有限集，并在与本文所采用假设相似的条件下证明了均衡的存在性；而 Anderson 等（2024）则将存在性结果推广到了紧致行动空间。Esponda 和 Pouzo（2016）通过收益扰动（路径 I）为伯克–纳什均衡提供依据；这一方法源自 Fudenberg 和 Kreps（1993）对混合策略纳什均衡收敛性的研究。路径 II——将混合策略解释为经验频率——源自 Esponda 和 Pouzo（2021），并由 Esponda、Pouzo 和 Yamamoto（2021）进一步发展；后者还强调了广义伯克–纳什概念的必要性。Murooka 和 Yamamoto（2023）将这一广义概念推广到了博弈中。

Spiegler（2016年）提出了“个人均衡”这一静态概念，可将其理解为伯克–纳什在无动态情形下针对单一主体的特例。该论文的独特贡献在于，利用有向无环图（DAG）刻画模型设定错误，从而为因果错误认知以及对相关结构的定性误解提供了一种简洁的图示语言，尤其是错误的条件独立性假设和归因错误。不过，该框架并非旨在刻画更具参数化特征的模型设定错误，例如函数形式错误。论文还引入了小扰动（颤抖）来处理零概率事件，为在模型中识别信念提供了一种务实途径。

例 3.1（努力）是 Heidhues、K˝oszegi 和 Strack（2018年）研究的模型的一个特例。他们证明，过度自信会使学习适得其反：主体在更新对外部基本面的信念时，会变得系统性地过于悲观，

行动也会朝错误的方向调整，最终离最优值更远——尤其是在错误的代价更高时。

例3.2（交易）源自 Esponda（2008）。主体表现出一种忽视相关性的倾向——他们将卖方的要价与买方的估值视为相互独立——这会导致选择忽视：他们正确地认识到成交物品的平均质量，却未能意识到，更高的价格会吸引质量更高的报价。因此，他们报出的价格过低，加剧了逆向选择。这一观点可推广到其他逆向选择环境。

例 3.3 改编自 Nyarko（1991），这是本章节所综述文献的早期先驱。他指出，当垄断者的模型设定错误时，价格不会收敛；而在模型设定正确时，价格则收敛至垄断最优水平。Esponda 和 Pouzo（2016）利用这个例子说明，纯行动 Berk–Nash 均衡可能不存在，而混合策略均衡却可能存在。通过随机逼近，他们证明，当垄断者在收益扰动下学习单一需求参数时，会收敛到这一混合均衡。后续研究利用这个例子阐释了更一般的（不）收敛结果（例如 Fudenberg、Romanyuk 和 Strack（2017）；Frick、Iijima 和 Ishii（2023）），以及模型设定错误的稳健性（例如 Fudenberg 和 Lanzani（2023））。Fudenberg、Romanyuk 和 Strack（2017）表明，当主体具有前瞻性时，他们可能会认为实验能持续带来收益，即使在短视主体的行动本会收敛的环境中，行为也可能不收敛。

4 博弈

为简便起见，我们考虑完全信息博弈。博弈参与者的集合 I 是有限的。对每个 $i \in I$，其个体行动空间为 $\mathbb{A}^i$；联合行动空间为 $\mathbb{A} := \times_{j \in I} \mathbb{A}^j$。参与者 i 的结果取值于 $\mathbb{Y}^i$。真实环境为 $Q^i : \mathbb{A} \to \Delta\mathbb{Y}^i$，它为每个行动组合指定一个关于 i 的结果分布。参与者 i 设想了一个参数化

模型 $\{Q^i_{\theta^i}:\ \theta^i \in \Theta^i\}$，其中 $Q^i_{\theta^i}:\ \mathbb{A} \to \Delta\mathbb{Y}^i$。收益为 $\pi^i:\ \mathbb{A}^i \times \mathbb{Y}^i \to \mathbb{R}$。我们采用与第3节相同的假设，上标表示参与者；例如，ν^i 是参与者 i 的支配测度，等等。

玩家 i 的**策略** 是一个概率分布 $\sigma^i \in \Delta\mathbb{A}^i$。一个（混合）**策略组合** 是 $(\sigma^i)_{i\in I} \in \times_{i\in I}\Delta\mathbb{A}^i$。我们将注意力限制在独立混合上：由此诱导的行动组合分布是其乘积

$$\sigma = \prod_{j\in I} \sigma^j \ \in\ \Delta\mathbb{A},$$

并记作 $\sigma^{-i} := \prod_{j\neq i}\sigma^j \in \Delta\mathbb{A}^{-i}$。

最优行动对应。给定 $\theta^i \in \Theta^i$ 和对手的独立混合 σ^{-i}，定义

$$U^i(a^i, \sigma^{-i}, \theta^i)\ :=\ \int_{\mathbb{A}^{-i}}\int_{\mathbb{Y}^i} \pi^i(a^i, y^i)\, Q^i_{\theta^i}(dy^i \mid a^i, a^{-i})\, \sigma^{-i}(da^{-i}). \tag{4.1}$$

给定信念 $\mu^i \in \Delta\Theta^i$，最佳回应对应为

$$F^i(\mu^i, \sigma^{-i})\ :=\ \arg\max_{a^i\in\mathbb{A}^i} \int_{\Theta^i} U^i(a^i, \sigma^{-i}, \theta^i)\, \mu^i(d\theta^i).$$

KL散度。对于 $i \in I$、$\theta^i \in \Theta^i$和 $a \in \mathbb{A}$，令

$$K^i(\theta^i, a)\ :=\ KL(Q^i(\cdot \mid a) \| Q^i_{\theta^i}(\cdot \mid a))\ =\ \int_{\mathbb{Y}^i} \ln\!\Big(\frac{q^i(y^i \mid a)}{q^i_{\theta^i}(y^i \mid a)}\Big)\, q^i(y^i \mid a)\, \nu^i(dy^i).$$

对于乘积策略组合 $\sigma = \prod_j \sigma^j$，玩家 i的KL最小化集合为

$$\Theta^{m,i}(\sigma)\ :=\ \arg\min_{\theta^i\in\Theta^i} \int_{\mathbb{A}} K^i(\theta^i, a)\, \sigma(da).$$

定义4.1。策略组合$(\sigma^i)_{i\in I}$是伯克–纳什均衡，如果对于每个 $i \in I$，都存在一个信念 $\mu^i \in \Delta\Theta^i$ 使得

$$\mathrm{supp}(\mu^i)\ \subseteq\ \Theta^{m,i}(\sigma) \qquad \text{and} \qquad \mathrm{supp}(\sigma^i)\ \subseteq\ F^i\big(\mu^i,\ \sigma^{-i}\big).$$

示例 4.1。 具有相互作用的双人努力（伯克–纳什）。博弈参与者 $i \in \{1,2\}$ 选择努力 $a^i \in [0,\infty)$，并获得 $\pi^i(a^i,\ y^i) = y^i - c(a^i)$。真实和感知的结果方程为

$$Y^i = \big(\alpha^* + a^1 + a^2 + a^1a^2\big)\theta^* + \varepsilon^i, \qquad Y^i = \big(\alpha + a^1 + a^2 + a^1a^2\big)\theta + \varepsilon^i,$$

具有共同的真实能力 $\alpha^* > 0$、感知能力 $\alpha > \alpha^*$（固定的过度自信，不会被学习）、$\theta^* > 0$以及 $\varepsilon^i \sim \mathcal{N}(0,\ 1)$。假设 c 连续可微且严格凸，$c(0) = c'(0) = 0$，并且当 $a^i \to \infty$趋于a^i时，$c'(a^i) \to \infty$。[7]

令 $s(a) := a^1 + a^2 + a^1a^2$。对于任意（可能是混合的）行动组合分布 $\sigma \in \Delta\mathbb{A}$，任一玩家的高斯 KL 目标函数为

$$\int_{\mathbb{A}} \tfrac{1}{2}\Big((\alpha + s(a))\theta - (\alpha^* + s(a))\theta^*\Big)^2 \sigma(da),$$

该目标函数关于 θ 严格凸（除非 σ 仅在满足 $\alpha + s(a) = 0$的组合上赋予正概率，但由于 $\alpha > 0$，这种情况不可能发生）。因此，对于每个 σ，KL 最小化解 $\theta^{\mathrm{m}}(\sigma)$都是唯一的。特别地，对于纯行动组合 a，有

$$\theta^{\mathrm{m}}(\delta_a) = \theta^* \,\frac{\alpha^* + s(a)}{\alpha + s(a)} \ (< \theta^* \text{ since } \alpha > \alpha^*).$$

因此，在均衡相关论述中，我们可以将注意力限制在退化信念（单一参数值）上。

给定点信念 θ，玩家 i 的短视一阶条件为 $c'(a^i) = \theta(1 + a^{-i})$。在伯克–纳什均衡（纯策略）中，信念等于实际策略组合下的 KL 最小化解，因此

$$c'(a^i) = (1 + a^{-i})\,\theta^* \,\frac{\alpha^* + s(a)}{\alpha + s(a)}, \qquad i = 1, 2. \tag{4.2}$$

[7]我们还施加渐近边际成本支配条件 $\liminf_{a^i\to\infty} \frac{c'(a^i)}{a^i} > \theta^*$。当 θ^* 已知时（真实基准），这保证了纯策略均衡的存在；同一界限为下文的伯克–纳什问题中的最佳响应提供了一个不变紧集。

对于每个固定的 a^{-i}，(4.2) 的右侧关于 a^i 连续，不超过 $(1+a^{-i})\theta^*$，并且当 $a^i \to \infty$ 时收敛至该上界。由于 c' 连续且严格递增，$c'(0) = 0$，并且 $c'(a^i) \to \infty$，对于固定的 a^{-i}，(4.2) 中的每个方程在 a^i 中都有唯一解（最佳响应为单值）。渐近支配条件 $\liminf_{a^i \to \infty} c'(a^i)/a^i > \theta^*$ 意味着最佳响应将某个盒集 $[0, M]^2$ 映射到自身，因此根据 Kakutani，纯策略伯克–纳什均衡存在。一般而言，均衡不一定唯一（两个递增的最佳响应可能相交多次）。♦

4.1 文献注释

博弈中的伯克–纳什均衡定义沿用 Esponda 和 Pouzo（2016）。上文只在两个特殊情形下陈述了该定义：（i）完全信息；（ii）对他人博弈过程分布的信念正确。Esponda 和 Pouzo（2016）提出了通用表述，尤其允许存在非对称（私人）信息。纳入这类信息或信号既符合现实，也很有用：它拓宽了应用范围，并有助于识别。例如，在前文关于过度自信的主体的例子中，我们通过对过度自信程度施加教条信念，实现了识别（KL最小化解的唯一性）。然而，在许多应用中，也可以利用参与者行动的变化来实现识别；当行动以可能与世界状态相关的信号为条件时，这种变化便会自然出现。由此，均衡方法能够给出具体预测：一旦相关参数得到（至少部分）识别，伯克–纳什均衡就能让我们直接刻画稳态结果，而无须从任意初始信念出发，求解完整的暂态学习与实验动态。

例4.1（努力博弈）出自 Ba 和 Gindin（2023年）。更一般地，他们将 Heidhues、K˝oszegi 和 Strack（2018年）的单主体模型扩展为两个博弈参与者；他们每个时期都会学习并做出最优回应，因此每位参与者的努力都会影响另一位参与者信号的信息量（信息外部性）。在正向外部性下，一位参与者付出更多努力会提高另一位参与者学习的价值，并促使对方相应地付出更多努力——也就是说，学习会相互促进；而在负向外部性下，学习则会相互制约。

由于一位主体付出的额外努力会通过这些外部性使另一位主体受益，轻微的过度自信便可能推动努力水平上升，从而改善福利。

与第3节（单个主体）的结果类似，这些结果对博弈同样成立：如果博弈过程收敛，就会收敛到伯克–纳什均衡。该节强调的两条路径也适用于此：在存在独立收益扰动时，可得到混合策略伯克–纳什均衡；没有扰动时，则必须采用广义均衡（参见 Esponda、Pouzo 和 Yamamoto（2021），以及 Murooka 和 Yamamoto（2023）对其在博弈中的扩展）。

如果每个时期的行动都被公开观察，且主体持续估计经验分布，那么，对博弈参与者拥有关于他人博弈过程分布的正确信念这一限制，便有了合理依据；对于有限行动集，可以像随机虚拟博弈那样，将信念建模为对经验频率的追踪（Fudenberg 和 Kreps, 1993）。关键在于，这又引入了另一种模型设定错误：主体学习仿佛 他人每个时期都采取一种平稳（可能是混合）策略——这通常并不成立，尽管策略收敛时渐近而言是准确的。这个视角揭示了纳什均衡——进而也包括伯克–纳什均衡——是相互最佳响应的不动点，而不是试图引导他人学习的努力所产生的结果：如果某个博弈参与者知道对手的学习规则，她就可以尝试操纵对手，而由此产生的稳态未必是纳什均衡。一旦博弈参与者对他人的学习规则进行建模（并试图影响这些规则），就还必须对他们关于他人模型的信念，以及他们关于他人如何看待其 模型的信念等进行建模，由此陷入无穷倒退。纳什均衡和伯克–纳什均衡通过假定最佳响应于（观察到的）博弈过程分布，终止了这种高阶倒退。

尽管前面的讨论强调了将高阶信念纳入（设定错误的）学习过程所面临的挑战，Murooka和Yamamoto（2023年）采取了一种具体的方法。[8]

[8]Ba 和 Gindin（2023）在研究博弈参与者如何在其博弈过程的动态版本中学习时，也采用了共同知识假设。

他们研究了结果服从以下条件时，通过均衡博弈进行的学习

$$y_t \;=\; Q\big(a_1^t,\, a_2^t;\, \theta^*,\, \alpha\big) \;+\; \varepsilon_t,$$

其中， θ^* 是一个标量未知的收益相关状态， α 是另一个收益参数，玩家对它持有教条且固定的高阶信念，而 ε 服从标准正态分布。每个时期，玩家根据 (i) 当前对 θ 的后验分布，以及 (ii) 对α的固定信念概况，计算在这些信念下阶段博弈的唯一纳什均衡，并据此进行博弈。之后，他们只观察公共结果 y_t （而非彼此的行动），并利用三项数据更新对 θ 的信念：实现的 y_t、自己的行动 a_i^t，以及均衡计算所隐含的预测的对手行动（对手的实际行动不可观察）。在完全投射[9], 下，每位玩家都假设其他人和自己对 θ 以及固定的 α 持有相同信念，且在所有更高阶上也是如此，因此预测值等于该信念概况下对手的均衡行动。对 θ的更新后验分布（对 α 的信念保持不变）被用于下一时期的均衡计算，这一过程不断重复。在某些技术假设下，他们发现，即使模型设定错误任意小，后验分布落在真实 θ^*的任何邻域内的概率也为零。

5 收敛

我们讨论三　种确定信念和行为收敛的一般方法　.

5.1 行动的收敛

如果行动收敛，那么根据命题 3.1，我们知道它们必定收敛到一个纯行动 Berk-Nash均衡。由此可推论，在不存在纯行动均衡的博弈中，行动不会收敛。例如，第3节中的未知需求垄断

[9]本文还考虑了关于（高阶）信念的更一般假设，遵循 Gagnon-Bartsch、Pagnozzi 和 Rosato（2021年）以及 Madar´asz、Danz 和 Wang（2023年）的思路。

案例就是如此。但即使存在纯行动 伯克–纳什 均衡，我们仍然不知道行动是否会收敛到其中某个均衡。

我们只考虑有限的行动空间和结果空间，以及一个可能是非短视的单个主体。

考虑以下对（纯行动）Berk-Nash均衡的精化 .

定义 5.1。 如果对于每个信念$\mu \in \Delta\Theta$，只要 supp(μ) $\subseteq \Theta^{\mathrm{m}}(\delta_a)$，就有 $F(\mu) = \{a\}$，那么行动 a就是一致严格的 Berk–Nash均衡。

虽然命题 3.1 的逆命题不成立，但事实证明，如果我们将均衡加强为一致严格的，那么就能得到以下稳定性概念的充要条件：

定义 5.2。 如果对于每个 $\kappa \in (0,1)$，都存在 $\varepsilon > 0$，使得对于每个满足 $\mu_0\Big(\widehat{\Theta}^{\varepsilon}(a)\Big) > 1-\varepsilon$ 的先验 $\mu_0 \in \Delta\Theta$，任何最优策略所规定的行动都以至少 $1-\kappa$ 的概率收敛到 a，则称行动 $a \in \mathbb{A}$ 是一致稳定的，其中

$$\widehat{\Theta}^{\varepsilon}(a) := \Big\{\theta \in \Theta : \ \exists\, \tilde{\theta} \in \Theta^{\mathrm{m}}(\delta_a) \text{ such that } ||Q_\theta(\cdot \mid a) - Q_{\tilde{\theta}}(\cdot \mid a)|| \leq \varepsilon\Big\}.$$

粗略地说，如果对于任意最优策略，以及从该行动对应的 KL最小化解的某个邻域内的每个信念出发，博弈过程都以高概率收敛于该行动，则称该行动是一致稳定的。

命题5.1。 行动 a是一致稳定的，当且仅当它是一致严格伯克–纳什均衡。

“如果”部分背后的想法是一个自我强化回路：如果 a 一致地严格，则存在一个邻域 B，其中包含其 KL 最小化解，以及一个差距 $m > 0$；只要后验分布在 B 上赋予足够大的概率， a 的短视优势就至少为 m，而偏离的前瞻性（实验）价值则一致地被某个 $m/2$ 所界定；因此，每个最优策略——无论是短视还是非短视——都会在该区域选择 a。进行 a 博弈过程会产生数据，而这些数据，

以高概率使后验分布保持在（或迅速返回） B内部，因此使 a 成为最优的条件往往会持续成立。特别是，反复被推离 B 足够远、以至于推翻严格裕度的可能性很小，而且是暂时的；至多经历有限次此类偏离后，信念将保持在 B 中，最优政策也会继续选择 a，从而实现一致稳定性。

现在我们要问，在任何全支持先验下，博弈过程最终稳定在某个行动a上的概率是否为正。

定义 5.3. 若对每个初始信念 $\mu_0 \in \Delta\Theta$（其支撑集为 $\mathrm{supp}(\mu_0) = \Theta$），都有$\mathbf{P}^{\mu_0}(\lim_{t\to\infty} a_t = a) > 0$，则行动 $a \in \mathbb{A}$ 是正吸引的。

获得正吸引性的一种方法，是找出这样的环境：从全支持先验出发，信念有正概率进入 KL 邻域，在该邻域内，一致严格行动 a 是唯一最优行动。根据上述一致稳定性结果，一旦信念进入该邻域，最优博弈过程就会停留在 a，因此在该事件发生时，路径会收敛到 a 。

定义 5.4. 若结果主观外生，则对每个 a， $a' \in \mathbb{A}$ 以及每个 $\theta \in \Theta$，都有$Q_\theta(\cdot \mid a) = Q_\theta(\cdot \mid a')$。也就是说，在主体模型下，结果分布不依赖于行动。

命题 5.2。 假设结果是主观外生的。如果 a是一致严格伯克–纳什均衡，那么 a 是正吸引的。

当结果是主观外生的，后验更新只取决于实际观察到的结果，而不取决于产生这些结果的行动。由于不同行动下的真实结果分布相互绝对连续，因此以正概率 会观察到一个有限的结果序列，其经验分布与行动 a 下的真实分布任意接近。对于任何这样的序列，都可以证明后验分布会在 KL 最小化集合 $\Theta^{\mathrm{m}}(\delta_a)$ 上赋予任意高的概率质量 。由于 a 是一致严格的，一旦信念足够集中，每个最优策略（包括前瞻性策略）都会唯一地选择 a；从

此后，先前的一致稳定性论证便适用：采取 a 会以高概率使信念保持在该区域内，因此，该过程以正概率 此后一直停留在 a ，也就是说，收敛到 a以正概率发生。

5.2 经验行动分布的收敛

我们继续假设行动空间有限，但不再要求结果空间有限。我们假设主体是短视的（在弱识别下，其中一些假设可以放宽）。我们的目标是刻画行动频率的渐近行为。即使行动序列不收敛，经验频率仍可能收敛，从而为长期行为和收益提供有意义的预测。

我们可以将经验行动频率 $\sigma_t \in \Delta\mathbb{A}$ 递归地表示为

$$\sigma_{t+1} = \sigma_t + \frac{1}{t+1}\big(\mathbf{1}(a_{t+1}) - \sigma_t\big), \tag{5}$$

其中，$\mathbf{1}(a_{t+1})$ 是行动指示向量。加上并减去给定当前信息的条件期望，

$$\sigma_{t+1} = \sigma_t + \frac{1}{t+1}\big(\mathbb{E}[\mathbf{1}(a_{t+1}) \mid \mu_{t+1}] - \sigma_t\big) + \frac{1}{t+1}\underbrace{\big(\mathbf{1}(a_{t+1}) - \mathbb{E}[\mathbf{1}(a_{t+1}) \mid \mu_{t+1}]\big)}_{\text{martingale difference}},$$

因此，演化取决于信念 μ_{t+1}。最后一项是相对于自然滤子的鞅差噪声，它一致有界，并乘以步长 $1/(t+1)$；因此，该项渐近可忽略。

关键步骤是借助 Berk定理和第3节引理 3.1 的思路，对极限信念作出刻画，从而重新表述这种依赖关系：当 t足够大时， μ_{t+1} 集中于那些参数值，其对应的 KL散度对于当前行动频率 σ_t下生成的数据而言近乎最小。

为处理映射 $\mu \mapsto \mathbb{E}[\mathbf{1}(a_{t+1}) \mid \mu]$可能存在的不连续性，我们将其替换为一个

对应 $\Delta F(\mu) \subseteq \Delta\mathbb{A}$，它包含该函数（即其图像包含该函数的图像），且具有良好的性质（上半连续，且取值非空、紧）。于是我们可以写出随机差分包含

$$\sigma_{t+1} = \sigma_t + \frac{1}{t+1}(r_{t+1} - \sigma_t), \qquad r_{t+1} \in \Delta F(\mu_{t+1}).$$

利用由 σ_t确定的 KL 最小化解周围 μ_{t+1} 的大-t 集中性，上述方程可以完全改写为关于 σ_t 的形式，另加渐近趋于零的近似误差。

接下来，将离散序列（σ_t）视为 $\Delta\mathbb{A}$ 中的点，并用直线段连接相邻点，沿“慢”时间尺度 $\tau_t = \sum_{i=1}^{t} 1/i$ 遍历这些线段（因此步长按 $1/(t+1)$缩小）。这就得到分段线性插值 $w(\cdot)$，它紧密跟踪 σ_t。吸收趋于零的误差项后，可以发现， w 可以很好地由微分包含的解近似：

$$\dot{\sigma}(t) \in \Delta F(\Delta\Theta^{\mathrm{m}}(\sigma(t))) - \sigma(t), \tag{DI}$$

它充当连续时间动力系统，引导离散过程（σ_t）的长期行为。

以下命题对此作出了形式化表述。

命题 5.3。令 $w:\ \mathbb{R}_+ \to \Delta\mathbb{A}$ 为经验行动频率（σ_t）$_{t\geq 1}$的连续时间插值，并考虑微分包含（*DI*）。对于初始条件 $\sigma(0) \in \Delta\mathbb{A}$，令 $\mathcal{S}_{\sigma(0)}$ 表示满足该初始条件的（DI）的绝对连续解 $\sigma(\cdot)$ 的集合，并令$\mathcal{S} \equiv \bigcup_{\sigma(0)\in\Delta\mathbb{A}} \mathcal{S}_{\sigma(0)}$ 为所有此类解构成的集合。

那么，几乎必然地，对每个 $T > 0$,

$$\lim_{t\to\infty} \inf_{\sigma(\cdot)\in\mathcal{S}} \sup_{s\in[0,T]} \left\| w(t+s) - \sigma(s) \right\| = 0.$$

要理解这一陈述，任取一个有限时域长度 $T > 0$。沿着过程向前看足够远（取 t 足够大）。那么，存在微分包含（DI）的一条解 $\sigma^t(\cdot)$，它从任意接近当前离散状态 $w(t)$的位置出发，并且其轨迹在此后的时间窗口 $[0, T]$ 内紧随插值路径。形式上：对每个 $\varepsilon > 0$ 和 $T > 0$，都存在 t_0，使得对所有 $t \geq t_0$，都可以找到微分包含（ DI ）的一条解 $\sigma^t(\cdot)$, 满足

$$\|\sigma^t(0) - w(t)\| \leq \varepsilon \quad \text{and} \quad \sup_{s\in[0,T]} \left\|\sigma^t(s) - w(t+s)\right\| \leq \varepsilon.$$

我们不 要求微分包含解在 $s = 0$ 时刻精确经过 $w(t)$；微小的随机误差和离散化误差通常会妨碍这种精确对齐。相反，该结论保证渐近地存在一个微分包含解，它从任意接近 $w(t)$ 的位置出发，并在任意固定的有限时间窗口内始终与插值后的离散路径保持一致地接近。

微分包含的静止点是广义 Berk-Nash 均衡（见定义 3.4）：对于每个 $a \in \text{supp}(\sigma)$, 都存在 $\mu_a \in \Delta\Theta(\sigma)$ 使得 $a \in F(\mu_a)$。不同于标准 Berk–Nash（它使用单个 信念来为所有支持集上的行动提供依据），广义版本允许不同的支持集上的行动由不同的 信念来解释, 而这些信念均来自与 σ 相容的 KL 最小化集。当关注对象是行动分布时，这一广义概念是自然的：如果信念没有趋于稳定，不同的支持集上的行动可能是在实质上不同（但与 σ 相容）的信念下选出的。与第 3 节的结果类似，可以证明，如果 $\sigma_t \to \sigma$，那么极限 σ 是一个广义伯克–纳什均衡。

接下来，我们来讨论收敛问题。定义 $d(\sigma,\ \Sigma) := \inf_{\tilde{\sigma}\in\Sigma} \|\sigma - \tilde{\sigma}\|$。

定义 5.5。 </style}集合 若存在一个开集，满足 int，且对每个，都存在，使得 (()) 对所有 成立，并且对于微分包含 (DI) 的每个解 ()，若其初始条件为 (0)，则称其为吸引的。最大的此类集合 U 称为吸引域U_Σ。若 $U_\Sigma = \Delta\mathbb{A}$，则称集合 Σ 为全局吸引的。

全局吸引意味着一致吸引：对于每个 $\varepsilon > 0$，都存在一个时间 $T(\varepsilon)$，使得对于从此处出发的微分包含的所有初始条件 $\sigma(0) \in \Delta\mathbb{A}$ 以及所有解 $\sigma(\cdot)$，以下结论同时成立

在那里，对于每个 $t \geq T(\varepsilon)$，都有 $d(\sigma(t)\ \Sigma) < \varepsilon$。因此，一个统一的时钟适用于所有情况——无需针对每个解或每个初始条件分别选择 T。这比仅仅声称每个解都收敛到 Σ 更强，因为后者允许 T （或收敛速率）取决于具体轨迹；全局吸引要求在整个状态空间上最终一致地接近。

命题 5.4。 如果 $\Sigma \subseteq \Delta\mathbb{A}$ 是全局吸引的，那么$\lim_{t\to\infty} d(\sigma_t, \Sigma) = 0$ 几乎必然。特别地，如果 $\Sigma = \{\sigma^*\}$ 是全局吸引的均衡，那么 $\sigma_t \to \sigma^*$ 几乎必然。

命题 5.4 证明的关键，是在固定长度窗口上将离散路径与微分包含解进行比较，再将这些比较串联起来。全局吸引性意味着存在一个统一的时间 $T(\varepsilon)$，适用于所有微分包含解，无论它们从何处出发。命题 5.3 使我们能够在 t 足够大时，选取一个微分包含解，使其在 $[0,T]$ 上与插值路径 $w(t+\cdot)$ 的距离保持在 ε 以内。由于每个微分包含解在经过时间 T 后都与 Σ 的距离不超过 ε，因此比较解在 T 时与 Σ 的距离不超过 ε；于是 $w(t+T)$ 与 Σ 的距离不超过 2ε，而（将插值过程还原后）对于邻近的 t'，$\sigma_{t'}$ 与 Σ 的距离不超过 3ε。在连续窗口上重复这一过程，全局吸引性的一致吸引会阻止过程逃出 Σ 的 ε 邻域；又由于 $\varepsilon > 0$ 是任意的， $d(\sigma_t,\ \Sigma) \to 0$ 几乎必然。

例5.1。现在我们重新讨论第3节中需求未知的垄断者这一例子。回顾一下，该例子有一个唯一的伯克–纳什均衡，即以混合概率$\sigma^* = \frac{1}{36}$ 选择价格 10。记作 $\sigma(t)$，微分包含简化为以下一维包含关系

$$\dot{\sigma}(t) \ \in \ \begin{cases} \{\,1-\sigma(t)\,\}, & \sigma(t) < \sigma^*, \\ [-\sigma(t),\, 1-\sigma(t)], & \sigma(t) = \sigma^*, \\ \{\,-\sigma(t)\,\}, & \sigma(t) > \sigma^*, \end{cases} \tag{DI-mono}$$

这简单地表示：价格 10 上的质量向当前的最佳回应移动——当 10 是唯一最优选择时，以速率$1-\sigma$ 上升；当 2 是唯一最优选择时，以速率 σ 下降；而在无差异时，则以这两个极值之间的任意速度演化。

定义 $I_\varepsilon = [L, U]$，其中 $L = \max\{0, \sigma^* - \varepsilon\}$，且 $U = \min\{1, \sigma^* + \varepsilon\}$。为证明 $\{\sigma^*\}$ 是全局吸引的，我们验证所有轨迹都会在某个有限时间 $T(\varepsilon)$ 之后进入并停留在 I_ε 内，而该时间不取决于初始条件。在边界处，向量场总是指向内部：若 $L = 0$，则 $\dot{\sigma} = 1 > 0$；若 $L = \sigma^* - \varepsilon > 0$，则 $\dot{\sigma} = 1 - \sigma \geq 1 - (\sigma^* + \varepsilon) > 0$；若 $U = 1$，则 $\dot{\sigma} = -1 < 0$；若 $U = \sigma^* + \varepsilon < 1$，则 $\dot{\sigma} = -\sigma \leq -(\sigma^* + \varepsilon) < 0$。因此，任何进入 I_ε 的轨迹都会留在其中。若 $\sigma(0) \leq \sigma^* - \varepsilon$（因此 $\varepsilon \leq 1/36$），那么在 $\sigma(t) < \sigma^*$ 期间，我们有 $\dot{\sigma}(t) = 1 - \sigma(t) \geq 1 - \sigma^* = 35/36$，这意味着到达 I_ε 所需的时间满足

$$T_\downarrow(\varepsilon) \leq \frac{(\sigma^* - \varepsilon) - \sigma(0)}{35/36} \leq \frac{36}{35}\left(\tfrac{1}{36} - \varepsilon\right) \leq \tfrac{1}{35}.$$

如果 $\sigma(0) \geq \sigma^* + \varepsilon$（因此 $\varepsilon \leq 35/36$），那么当 $\sigma(t) > \sigma^*$ 时，我们有 $\dot{\sigma}(t) = -\sigma(t) \leq -\sigma^* = -1/36$，因此

$$T_\uparrow(\varepsilon) \leq \frac{\sigma(0) - (\sigma^* + \varepsilon)}{1/36} \leq 35 - 36\varepsilon \leq 35.$$

因此，每条轨迹都会在 $T(\varepsilon) \leq \max\{1/35,\, 35 - 36\varepsilon\} \leq 35$ 内进入 I_ε，并永远停留在那里。因此，对于所有 $t \geq T(\varepsilon)$，$|\sigma(t) - \sigma^*| \leq \varepsilon$，根据定义，$\{\sigma^*\}$ 是全局吸引的。因此，由命题 5.4，经验动作频率几乎必然收敛到唯一的混合均衡 $\sigma^* = 1/36$。

命题 5.4 给出了关于吸引集的收敛结果，但它并未给出关于不稳定均衡的非收敛结果。连续时间近似保证，在任意固定时域内，都可以用某个微分包含解来逼近插值路径，而该解的初始条件只需接近当前状态。这个“接近”的邻域可能包含不稳定均衡本身，而该均衡存在一个保持不变的微分包含解。因此，仅凭这一结果，我们无法排除离散过程收敛到不稳定均衡的可能性。

5.3 信念的收敛

到目前为止，我们一直关注行动或行动分布的收敛。现在，我们转而考察一种互补的方法：信念的收敛。第2节传达的一个信息是，标准的鞅收敛定理在这里并无用处；但事实证明，可以采用一种更精细的论证来确立信念收敛，而这种论证依赖于鞅论证。

此前，我们研究了有限行动空间下的收敛。现在，我们考虑连续统行动空间，并进一步将注意力限制在最优行动对应 F 存在连续选择的情形。这通常排除了有限行动空间。一个充分条件是，对每个信念都存在唯一最优行动。我们将从 F 到 $\mu \mapsto a^*(\mu)$ 的连续选择记为 $\mu \mapsto a^*(\mu)$。

定义 5.6。 考虑任意 $\theta \in \Theta$。退化信念 δ_θ 为：

1. 若对任意 $\gamma < 1$，都存在一个邻域 $\mathcal{B} \ni \delta_\theta$，使得对于每个初始信念 $\mu \in \mathcal{B}$，$\mathbf{P}_\mu[\mu_t \to \delta_\omega] \geq \gamma$，则称其为局部稳定；

2. 若存在一个邻域 $\mathcal{B} \ni \delta_\omega$，使得对于每个初始信念 $\mu \in \mathcal{B}$，$\mathbf{P}_\mu[\exists t : \ \mu_t \notin \mathcal{B}] = 1$，则称其为不稳定 。

定义 5.7。 固定一个行动 a。对于 $\theta, \theta' \in \Theta$，定义 $\theta \succsim_a^{\mathrm{KL}} \theta' \iff K(\theta, a) \leq K(\theta', a)$。

我们将支持 Berk-Nash均衡的信念称为 Berk-Nash均衡信念。我们只考虑退化信念。

命题 5.5. δ_θ 是 *Berk–Nash*均衡信念，当且仅当 $\theta \succsim_{a^*(\delta_\theta)}^{\mathrm{KL}} \theta'$ 对所有 $\theta' \in \Theta$都成立；等价地，$\theta \in$ 使.....最小的参数$_{\tilde{\theta} \in \Theta} K(\theta, a^*(\delta_\theta))$。

以下命题与第3节的结果类似：如果行为收敛，那么它将收敛到一个 Berk-Nash均衡。

命题 **命题 5.6**。如果 δ_θ 不是伯克–纳什均衡信念，那么 δ_θ 就是不稳定的 .

预序 $\succsim_a^{\mathrm{KL}}$对于命题 5.6 中的刻画很有用，但正如下一个例子所示，它可能无助于得出局部稳定性结果。

例 5.2（缓慢学习）。 单个主体面临一个固定但未知的真实状态 $\theta^* \in \Theta = \{\theta_1, \theta_2\}$，其中 $\theta_2 > \theta_1 > 0$。每个时期，她选择精度／努力水平 $a \in [0, \bar{\gamma}]$，并以 $C(a)$ 为成本，随后以真实概率 $Q(1 \mid a) = a\theta^* + \beta^*$ 观察到 $y \in \{0, 1\}$。该主体低估了基准率，认为 $\beta < \beta^*$，这是一种“自我偏见”式更新：她对关于自身能力的利好消息反应过度，而对利空消息反应不足；她的感知信号模型为 $Q_\theta(1 \mid a) = a\theta + \beta$。假设对于每个信念，短视最优努力都是唯一的，并可由连续函数 $\mu \mapsto a^*(\mu)$ 表示。假设对于每个非退化信念 μ，均有 $a^*(\mu) > 0$（这反映了当真实参数不确定时，即使微小努力也具有正价值），并且对于任意退化信念 μ，有 $a^*(\mu) = 0$。（该假设可由关于原始要素的自然假设得到。）

如果 $a^*(\mu) = 0$，那么 z 具有无信息性（独立于 θ）；因此 $K(\theta_1,\ a^*(\mu)) = K(\theta_2,\ a^*(\mu))$。根据命题 5.5，两个点信念$\delta_{\theta_1}$ 和δ_{θ_2} 都是伯克–纳什均衡信念。因此，KL 预序 $\succsim_a^{\mathrm{KL}}$在此处无法进行局部区分，对于证明收敛毫无帮助。

定义 5.8。 固定一个行动 a 和 $q > 0$。对于 $\theta,\ \theta' \in \Theta$，定义

$$\theta \succsim_a^q \theta' \quad \Longleftrightarrow \quad \int \left(\frac{q_{\theta'}(y \mid a)}{q_\theta(y \mid a)} \right)^q Q(dy \mid a) \ \leq \ 1.$$

KL 预序 $\succsim_a^{\mathrm{KL}}$是完备的，而 $\succsim_a^q$ 通常是不完备的。族$\{\succsim_a^q\}_{q>0}$ 是嵌套的，并且随着 $q \downarrow 0$逼近 KL 支配。此外，若 $\theta \succ_a^{\mathrm{KL}} \theta'$，则存在 $q > 0$ 使得 $\theta \succ_a^q \theta'$。

命题 5.7。 考虑任意 $\theta \in \Theta$。点信念 δ_θ 是：

1. 局部稳定 当存在 $q > 0$ 和一个邻域 $\mathcal{B} \ni \delta_\theta$，使得对于每个$\mu \in \mathcal{B} \setminus \{\delta_\theta\}$、每个满足 $a = a^*(\mu)$ 的行动 a 以及每个 $\theta' \neq \theta$，都有 $\theta \succ_a^q \theta'$。

2. 不稳定 当存在 $q > 0$、一个邻域 $\mathcal{B} \ni \delta_\theta$ 以及某个 $\theta' \neq \theta$，使得对于每个 $\mu \in \mathcal{B} \setminus \{\delta_\theta\}$ 和每个满足 $a = a^*(\mu)$ 的行动 a，都有 $\theta' \succ_a^q \theta$。

例 5.3（续）。固定 $\theta^* = \theta_1$。由于 $Q(1 \mid a) = a\theta^* + \beta^*$ 且 $Q_\theta(1 \mid a) = a\theta + \beta$，我们有 $|Q(1 \mid a) - Q_{\theta_1}(1 \mid a)| = |\beta^* - \beta|$ 且 $|Q(1 \mid a) - Q_{\theta_2}(1 \mid a)| = |(\beta^* - \beta) - a(\theta_2 - \theta_1)|$。由于 $\theta_2 > \theta_1$ 且 $\beta^* > \beta$，定义 $\bar{a} := \frac{\beta^* - \beta}{\theta_2 - \theta_1} > 0$，使得第二个绝对值中的项在 $a \in (0, \bar{a}]$ 时为正，在 $a > \bar{a}$时为负。因此，对于 $a \in (0, \bar{a}]$，

$$\big|Q(1 \mid a) - Q_{\theta_2}(1 \mid a)\big| = (\beta^* - \beta) - a(\theta_2 - \theta_1) < \beta^* - \beta = \big|Q(1 \mid a) - Q_{\theta_1}(1 \mid a)\big|.$$

由此可得，对于所有 $a \in (0, \bar{a}]$，$K(\theta_2, a) < K(\theta_1, a)$。由于 $a^*(\delta_{\theta_1}) = 0$，$a^*(\mu) > 0$对非退化的 μ成立，且 a^* 连续，因此存在δ_{θ_1} 的一个邻域 $\mathcal{B}$，使得对于所有 $\mu \in \mathcal{B} \setminus \{\delta_{\theta_1}\}$，$0 < a^*(\mu) \leq \bar{a}$。因此，

$$K\big(\theta_2, a^*(\mu)\big) < K\big(\theta_1, a^*(\mu)\big) \quad \Longrightarrow \quad \theta_2 \succ^{\mathrm{KL}}_{a^*(\mu)} \theta_1,$$

因此，根据 (KL $\Rightarrow q$)，存在 $q > 0$，使得 $\theta_2 \succ^q_{a^*(\mu)} \theta_1$ 对所有 $\mu \in \mathcal{B} \setminus \{\delta_{\theta_1}\}$ 均成立。命题 5.7 表明 δ_{θ_1} 是不稳定的，而通过围绕δ_{θ_2}进行对称论证可知，δ_{θ_2} 是局部稳定的。

5.4 文献注释

第5.1节（行动收敛）的内容完全归功于 Fudenberg、Lanzani和Strack（2021年）。他们还提出了一致伯克–纳什均衡行动这一概念；在弱识别（定义 3.3）下，该概念与伯克–纳什均衡一致，但在识别失败时则更为精细。他们还通过证明：如果行动序列收敛，那么它必定收敛到一致伯克–纳什均衡，从而加强了命题 3.1。因此，如果一致伯克–纳什均衡不存在，行动就不会收敛。Fudenberg、Lanzani和Strack（2021年）还将他们的结果推广到这样一类环境：行动前信号能够影响实际结果和收益。

此外，缺乏逐路径收敛并不一定意味着频率不收敛：在极限下，经验分布可以将全部概率质量集中在 a^* 上，即使博弈过程从未稳定在该行动上，因为另一行动会被无限次采取，但其频率趋于零。Fudenberg、Lanzani和Strack（2021年）给出了一个排除此现象的条件。

第5.2节（行动分布收敛）中的内容完全基于 Esponda、Pouzo 和 Yamamoto（2021）。命题 5.3 除其他内容外，还基于 Bena¨ım、Hofbauer 和 Sorin（2005）的随机逼近结果。当行动集较大时，求解微分包含可能很困难；Esponda、Pouzo 和 Yamamoto（2021）表明，在适当条件下，可以改为研究参数空间上的微分包含 Θ。当 Θ 是一维时，这通常能简化分析，并可扩展至许多行动；不过，如果 Θ维数较高，分析又会变得困难。Murooka 和 Yamamoto（2023）将这些结果扩展到了具有连续统行动集的模型。正如前文所述，微分包含方法的一个局限是难以排除收敛到不稳定均衡的可能性：Pemantle（1990）早已强调了这一点，并引入噪声来得到不收敛结果（另见 Bena¨ım（1999））。近来，Murooka 和 Yamamoto（2025）给出了随机过程以零概率收敛到线性不稳定稳态的条件；重要的是，他们的结果涵盖了应用中常见的高斯噪声，而早期结果则未涵盖。

第5.3节（信念收敛）中的结果完全归功于 Frick、Iijima 和 Ishii（2023年）。为简便起见，我们假设行动选择关于 μ 连续；而他们在更弱的假设下建立了这些结果：行动选择在点质量信念附近关于 μ 连续，但在其他信念处可以不连续。因此，这一条件也可以与有限行动空间相容。在第2节中，我们指出，鞅收敛定理只能得出主观 信念收敛，因此在模型设定错误时没有信息性。Frick、Iijima 和 Ishii（2023年）通过转换分析对象，恢复了鞅式论证。他们的关键观察是，在任何满足 q-支配序的信念区域上，后验比率的 q 次幂会成为一个非负的

上鞅。这一洞见使他们能够借用正确设定情形中的鞅收敛技巧，并成为其结果的核心工具。

我们在行动选择的情境中介绍了他们的思想；Frick、Iijima 和 Ishii（2023）则在一个更一般、更抽象的环境中表述这些思想，不涉及明确的主体、行动或偏好。这一框架自然涵盖了某些设定错误的社会学习环境。尽管 Frick、Iijima 和 Ishii（2023）假设参数空间有限，但他们证明，部分结果可以推广到紧致有限维空间。

除了局部稳定性之外，Frick、Iijima 和 Ishii（2023）还证明了一个全局结果：信念收敛到经迭代剔除 KL 支配信念后保留下来的集合。该结果要求行动选择关于 μ 对所有信念都连续，而不仅仅是在点质量信念附近。在慢速学习示例中，他们将分析扩展到两个以上的参数值，并证明无论真实情况如何，系统都会全局收敛到最高参数值。特别是，他们表明，极其微小的模型设定错误也可能导致（正确）学习彻底失效。

总的来说，这三种收敛方法涵盖了广泛的应用场景，但彼此互补，且各自适用于不同的情形。行动收敛 方法假设行动集和结果集均为有限集，因此无法处理行动本身不收敛、但其长期频率收敛的情形。经验分布 方法刻画行动的渐近分布，无论行动是否收敛。然而，该方法需要求解微分包含，这在高维行动空间或参数空间中可能颇具挑战。信念收敛方法可以处理连续统行动集，不过若干应用依赖参数空间有限这一假设，因而可能受到限制。此外，该方法要求行动在点质量信念附近关于信念连续；对于全局稳定性结果，则要求处处连续。

对于应用研究而言，结合这些结果，或稍作调整，通常就能得出所需的（不）收敛结论，不过可能需要一些技术上的努力。最近，Esponda 和 Pouzo（2025年）提出了 *Berk–Nash*可理性化，

其定义直接类比于经典可理性化：(i) 存在一个算子 Γ，类似于最佳回应算子；(ii) 若 $A \subseteq \Gamma(A)$，则一组行动 A 是自我论证的。例如，在单主体情形下，

$$\Gamma(A) = \left\{ a \in \mathbb{A} \ : \ \exists \sigma \in \Delta A,\ \mu \in \Delta\Theta^m(\sigma) \text{ such that } a \in F(\mu) \right\}.$$

可理性化行动集合是最大的自我辩护集合，可以通过从完整行动集 $\Gamma \quad \mathbb{A}$迭代得到。他们证明，几乎必然地，每个被无限次采取（或接近）的行动都是可理性化的。从这个意义上说，它是 Frick、Iijima 和 Ishii（2023年）对全局稳定信念的迭代刻画在行动空间中的对应物；关键在于，它不要求对 F 进行连续选择，因此适用于有限行动集，也适用于给定信念对应多个最优行动的环境。当伯克–纳什可理性化集合为单元素时，这意味着几乎必然收敛到该行动。

除这些一般性方法之外，还有大量研究针对特定的模型设定错误情形，证明了（不）收敛结果。Nyarko（1991）在第3节中“需求未知的垄断”这一例子里表明，行动不会收敛。Schwartzstein（2014）证明了选择性注意模型中的信念收敛。Fudenberg、Romanyuk 和 Strack（2017）分析了带有布朗噪声的连续时间学习模型（不假设主体短视），并给出了行动收敛的条件。他们还表明，前瞻性主体可能会认为实验能带来持续收益，导致在短视主体的行动本会收敛的情形下，其行动无法收敛。Heidhues、K″oszegi 和 Strack（2018）运用压缩映射论证，证明在可加分离、超模的环境中，非短视主体会全局收敛到唯一的伯克–纳什均衡。He（2022）将这一方法应用于一个截尾数据、赌徒谬误的情形；Ba和Gindin（2023）则将基于压缩映射的方法推广到多主体学习，其中一个例子是第4节中的努力博弈。对于两点先验，Bohren and Hauser

（2021）基于 Kullback–Leibler 散度序，给出了行动/信念收敛的充要条件，并提出了一般性的局部和全局稳定性判据。最后，Heidhues、K˝oszegiandStrack（2021）证明，在正态先验和正态信号下，信念会收敛到一个伯克–纳什均衡。

此外，还有一支互为补充的模型设定错误的社会学习文献。Bohren（2016）研究了主体错误感知知情前辈占比的情形，并证明其会收敛至稳态；Frick、Iijima 和 Ishii（2020）分析了主体错误感知总体类型分布的环境，并证明了收敛。他们还指出，即使是微小的模型设定错误，也会使（正确的）社会学习变得脆弱。

参考文献

Anderson, Robert M、Haosui Duanmu、Aniruddha Ghosh 和 M Ali Khan。2024年。“关于无限空间中模型设定错误的马尔可夫决策过程的伯克–纳什均衡存在性。”《经济理论杂志》，217：105813。

Arrow, Kenneth J. 和 Jerry R. Green。1973年。“贝叶斯设定中的预期均衡札记。”社会科学数学研究所（IMSSS），斯坦福大学经济学系列工作论文第33号，加利福尼亚州斯坦福。

Ba、Cuimin 和 Alice Gindin。 2023年。“一个具有过度自信的设定错误学习多主体模型。”《博弈与经济行为》，142：315–338。

Barron、Andrew、Mark J. Schervish 和 Larry Wasserman。 1999。“非参数问题中后验分布的一致性。”《统计年刊》，27(2)：536–561。

Battigalli，Pierpaolo。 1987年。《博弈与社会情境中的理性行为与均衡》。论文（打字稿），博科尼大学，米兰。

Bena¨ım, Michel。 1999。“随机逼近算法的动态。”载于*S´eminairede Probabilit´es XXXIII*。第1709卷，数学讲义，由Jacques Az´ema、´Michel Emery、Michel Ledoux和Marc Yor编，1–68页。柏林、海德堡：施普林格。

Bena¨ım、Michel、Josef Hofbauer 和 Sylvain Sorin。2005年。“随机逼近与微分包含。”《*SIAM*控制与优化杂志》，44(1)：328–348。

Berk, Robert H. 1966年。“模型错误时后验分布的极限行为。”《数理统计年刊》，37(1)：51–58。

Bohren, J Aislinn. 2016年。“模型设定错误下的信息级联。”《经济理论杂志》，163：222–247。

Bohren, J Aislinn 和 Daniel N Hauser。 2021年。“异质错设模型下的学习：特征刻画与稳健性。”《计量经济学》，89(6)：3025–3077。

Bohren, J Aislinn 和 Daniel N Hauser。 2025年。“学习与博弈中的错设模型。”《经济学年度评论》，17。

Branch, William A. 2006年。《宏观经济学中的受限认知均衡与学习》。《后瓦尔拉斯宏观经济学：超越动态随机一般均衡模型》。剑桥大学出版社，纽约，135–160。

Bunke, Olaf, and Xavier Milhaud. 1998年。《错误模型下贝叶斯估计的渐近行为》。《统计年刊》，26(2)：617–644。

Diaconis, Persi, and David Freedman.1986年。《论贝叶斯估计的相合性》。《统计年刊》，14(1)：1–26。

Doob, Joseph L. 1953年。《随机过程》。纽约：威立。

Esponda, Ignacio。 2008年。《存在逆向选择的经济体中的行为均衡》。《美国经济评论》，98(4)：1269–1291。

Esponda, Ignacio 和 Demian Pouzo。 2016年。《伯克–纳什均衡：对设定错误的模型下主体进行建模的框架》。《计量经济学》，84(3)：1093–1130。

Esponda, Ignacio 和 Demian Pouzo。2021年。“设定错误的马尔可夫决策过程中的均衡。”《理论经济学》，16(2)：717–757。

Esponda 和 Ignacio 和 Demian Pouzo。2025年。“Berk–Nash可理性化。”工作论文。

Esponda, Ignacio、Demian Pouzo 和 Yuichi Yamamoto。2021年。“模型设定错误时贝叶斯学习者的渐近行为。”《经济理论杂志》，195：105260。

Evans, George W 和 Seppo Honkapohja。2001年。宏观经济学中的学习与预期。普林斯顿大学出版社。

Eyster, Erik 和 Matthew Rabin。2005年。“诅咒均衡。”《计量经济学》，73(5)：1623–1672。

Eyster, Erik 和 Matthew Rabin。2010年。“丰富信息环境中的盲从。”《美国经济学杂志：微观经济学》，2(4)：221–243。

Frick, Mira, Ryota Iijima 和 Yuhta Ishii。2020年。“误解他人与社会学习的脆弱性。”《计量经济学》，88(6)：2281–2328。

Frick, Mira, Ryota Iijima 和 Yuhta Ishii。2023年。“模型设定错误下的信念收敛：一种鞅方法。”《经济研究评论》，90(2)：781–814。

Fudenberg, Drew 和 David K Levine。1993*a*。“自我确认均衡。”《计量经济学：计量经济学会期刊》，523–545。

Fudenberg、Drew 和 David K Levine。 1993年。“稳态学习与纳什均衡。”《计量经济学：计量经济学会杂志》，547–573。

Fudenberg、Drew 和 David K Levine。1998年。《博弈学习理论》。 第 2 卷，麻省理工学院出版社。

Fudenberg、Drew 和 David M. Kreps。1988年。《博弈中的学习、实验与均衡理论》。打字稿专著（部分草稿 0.11），麻省理工学院和斯坦福大学。

Fudenberg、Drew 和 David M Kreps。1993年。《混合均衡学习》。《博弈与经济行为》，5(3)：320–367。

Fudenberg、Drew 和 David M Kreps。1995年。《扩展式博弈中的学习 I：自我确认均衡》。《博弈与经济行为》，8(1)：20–55。

Fudenberg 和 Giacomo Lanzani。2023年。《哪些模型设定错误会持续存在？》《理论经济学》，18(3)：1271–1315。

Fudenberg、Giacomo Lanzani 和 Philipp Strack。2021年。《内生错设学习的极限点》。《计量经济学》，89(3)：1065–1098。

Fudenberg, Drew, Gleb Romanyuk, and Philipp Strack.2017年。“基于错设先验的主动学习。”《理论经济学》, 12(3): 1155–1189.

Fuster, Andreas, David Laibson, and Brock Mendel. 2010年。“自然预期与宏观经济波动。”《经济展望杂志》, 24(4): 67–84.

Gagnon-Bartsch, Tristan, Marco Pagnozzi, and Antonio Rosato.2021年。“拍卖中的私人价值投影。”《美国经济评论》, 111(10): 3256–3298.

Ghosal, Subhashis, and Aad W. van der Vaart.2017年。《非参数贝叶斯推断基础》。剑桥统计与概率数学丛书，剑桥大学出版社。

Ghosal, Subhashis, Jayanta K. Ghosh 和 Aad W. van der Vaart。2000年。“后验分布的收敛速率。”《统计年刊》，28(2)：500–531。

Hahn, Frank Horace。1973年。论经济学中的均衡概念：就职演讲。剑桥大学出版社。

Hansen, Lars Peter, 和 Thomas J Sargent。2011年。《稳健性》。载于《稳健性》。普林斯顿大学出版社。

Heidhues, Paul, Botond K˝oszegi, 和 Philipp Strack。2018年。“不切实际的预期与错误引导的学习。”《计量经济学》，86(4)：1159–1214。

Heidhues, Paul, Botond K˝oszegi, 和 Philipp Strack。 2021年。“模型设定错误下学习模型的收敛。”《理论经济学》，16(1)：73–99。

He, Kevin。2022年。“从删失数据中错误学习：最优停止问题中的赌徒谬误及其他相关性错误。”《理论经济学》，17(3)：1269–1312。

Jehiel, Philippe, 和 Fr´ed´eric Koessler。2008年。“重新审视基于类比预期的不完全信息博弈。”《博弈与经济行为》，62(2)：533–557。

Kirman, Alan P. 1975年。“企业对需求状况的学习。”载于适应性经济模型。137–156。爱思唯尔。

Kleijn, Bas J. K., and Aad W. van der Vaart.2006年。“无限维贝叶斯统计中的模型设定错误。”《统计年刊》，34(2):837–877。

Lanzani, Giacomo. 2025年。“对模型设定错误的动态关注。”《计量经济学》，93(4): 1333–1370。

Madar´asz, Krist´of, David Danz, and Stephanie Wang.2023年。“投射性思维：模型、证据与应用。”证据与应用（*2023*年*5*月*31*日）。

Molavi、Pooya 等。 2019年。《具有学习和模型设定错误的宏观经济学：一般理论与应用》未发表手稿。

室冈武与 Yuichi Yamamoto。 2023年。《高阶模型设定错误下的收敛与稳态分析》。大阪大学国际公共政策研究科，大阪大学。

室冈武与 Yuichi Yamamoto。2025年。“博弈参与者对他人设定错误时的贝叶斯学习。”大阪公立大学社会经济研究所。

Whitney K. Newey 与丹尼尔·麦克法登。1994年。“大样本估计与假设检验。”载于《计量经济学手册》。第4卷，，编者为罗伯特·F·恩格尔与 Daniel L. McFadden，2111–2245。阿姆斯特丹：爱思唯尔。

Yaw Nyarko。1991年。“设定错误模型中的学习与周期的可能性。”《经济理论杂志》，55(2)：416–427。

Robin Pemantle。1990年。“罐模型与随机逼近中不收敛至不稳定点。”《概率年刊，698–712。

Lorraine Schwartz。 1965年。“论贝叶斯程序。” 概率论及相关领域杂志，4：10–26。

Schwartzstein, Joshua. 2014年。“选择性注意与学习。”《欧洲经济学会杂志》，12(6)：1423–1452。

Shalizi, Cosma Rohilla. 2009年。“相关数据与模型设定错误下贝叶斯更新的动态。”《电子统计学杂志》，3：1039 – 1074。

Shen, Xiaotong, and Larry Wasserman. 2001年。“后验分布的收敛速度。”《统计年刊》，29(3)：687–714。

Spiegler, Ran. 2016年。“贝叶斯网络与有限理性预期。”《经济学季刊》，131(3)：1243–1290。

White, Halbert. 1982年。“错误模型的最大似然估计。” 计量经济学，50(1): 1–25。

Yamada, Keigo. 1976。“错误模型下随机过程后验分布的渐近行为。” 数学分析与应用杂志，56(2): 294–308。