DOI: 10.6277/TER.202206_50(2).0003

混合均衡的隨機穩定: 期許水準於零和賽局之應用

袁國芝・王子豪・莊委桐*

在有唯一混合 Nash 均衡的兩人零和賽局中, 我們導入期許水準 (Aspiration) 建立了一個超賽局模型。參賽者於每期選定一個策略並重複玩 N 次零和賽局; 以此方式進行無限多期。參賽者透過檢視當前策略於前期的表現是否達到設定期許水準來決定繼續使用原策略或挑選新的策略。當策略都不能滿足時則降低期許水準。參賽者有微小的機率會進行新的嘗試或犯錯。依據 Young (1993, 1998) 的隨機隱定方法, 我們證明「兩個玩家期許水準爲零且採用混合均衡策略」爲隨機穩定狀態之一。因此我們部分解決了 Crawford 難題。並以兩性戰爭賽局爲例, 說明混合均衡也可以如純粹均衡一般穩定。

關鍵詞: 零和賽局, 混合 Nash 均衡, 期許水準, 隨機穩定
JEL 分類代號: C72, C73

1 緒論

在賽局理論裡, 經濟學家除了關心賽局的均衡是否存在, 也關心在動態調整過程中, 玩家的行爲是否會收斂到均衡。本文試圖研究以下議題: 給定一個兩人零和賽局, 且該賽局存在唯一的混合 Nash 均衡, 如果導入期許水準 (aspiration level) 作爲玩家調整行動的目標, 上述的混合均衡是否會

*作者分別爲東吳大學經濟學系副教授、愛卡拉互動媒體雲端科技事業部資料科學家與中央研究院經濟研究所副研究員與國立交通大學經營管理研究所合聘副教授。莊委桐爲通訊作者 (wjuang@econ.sinica.edu.tw)。我們感謝兩位匿名審查委員提供之寶貴意見與建議, 使得本文更加完善。

經濟論文叢刊 (*Taiwan Economic Review*), 50:2 (2022), 217–249。
國立台灣大學經濟學系出版

臺灣大學學術期刊資料庫

在動態過程中出現? 在現實生活中, 我們也常見到, 人們在諸如剪刀石頭布這類的賽局中採用混合均衡行動。然而, 相較於純粹均衡的行爲表現, 這類具隨機性質的行爲又是如何呈現出來的? 本文除了分析長期動態的調整是否會收斂到混合均衡之外, 也企圖爲人們採用混合均衡行動的行爲, 提供其背後的理論基礎。

1.1 研究背景

在零和賽局均衡的收斂問題上, Robinson (1951) 證明在兩人零和賽局中, 若雙方同時行動, 且皆採用 fictitious play — 統計對手使用各行動的相對次數, 針對此相對次數, 選取純粹行動作最適反應 — 長期下來, 雙方的平均報酬會收斂到混合均衡時的報酬。Shapley (1964) 提出一個簡單的例子, 說明同樣的方法應用在非零和賽局時, 雙方的平均報酬不會收斂到混合均衡時的報酬。Brown (1951) 則將兩人同時行動改爲輪流行動, 證明若一個零和賽局重複玩無限多次, 雙方的平均報酬會收斂到混合均衡時的報酬, 且雙方採用各行動的相對次數也會收斂到混合均衡。

雖然 Brown (1951) 的作法成功收斂到零和賽局的均衡解, 但有幾個缺憾。第一, 在多人賽局 (不必然是零和賽局) 裡, Stahl (1988) 發現, 若玩家採用更一般化的 fictitious play, Brown (1951) 的收斂性質只會在無限遠時發生。第二, 採用 fictitious play, 玩家必須知道, 在每一個結果 (outcome) 之下, 他能得到的報酬各自是多少, 才能針對對手各行動的相對次數, 選取預期報酬最高的純粹行動。最後一個, 在 Brown 的分析中, 雖然玩無限多次時, 玩家採用各行動的相對次數收斂到混合均衡, 看似玩家採用混合均衡行動, 但實際上玩家採用的還是純粹行動, 並非真的採用混合行動。

爲了處理上述的問題, Crawford (1974, 1985) 探討在不知道報酬矩陣的情況下, 玩家能否透過某些合理的學習法學到混合均衡。他假設玩家的理性是有限的, 只知道自己每一期所採取的行動以及報酬, 不知道賽局的報酬矩陣, 也不知道對手採用什麼行動, 因此玩家不會建立關於對手的信念來做最適反應。只會單純地調整自己的行動, 使報酬極大化。玩家每期選用一個混合行動, 與對手重複玩 N 次賽局, 並記錄每次行動的結果及其報酬。因此每期結束後, 玩家可以計算本期各純粹行動的期望報酬, 以

臺灣大學學術期刊資料庫

及本期整個混合行動的期望報酬。Crawford 提出的學習法爲: 若本期某個純粹行動的期望報酬不如採用混合行動的期望報酬時, 下一期玩家會降低玩此純粹行動的機率; 反之, 若某個純粹行動的期望報酬比混合行動高時, 下一期玩家會增加玩此純粹行動的機率。Crawford (1974) 證明在絕大部分兩人零和賽局裡,[1] 若玩家採用這種學習法, 將無法收斂到穩定的混合均衡。Crawford (1985) 將上述結論進一步延伸, 在絕大部分只有混合均衡的多人賽局裡 (不必是零和賽局), 若玩家採用上述的學習法, 將無法收斂到穩定的混合均衡。因此 Crawford 認爲, 能夠讓玩家學會合作的理性能力, 尚不足以讓玩家學到混合均衡, 經濟學家需要假設玩家有更複雜的理性能力, 才能解釋人們爲什麼會採用混合均衡。

針對只有一個均衡且爲混合均衡的兩人賽局, Conlisk (1993a,b) 提出四種可能的途徑來解決 Crawford 難題。第一種方法稱爲外插法 (the extrapolation process): 類似 Crawford 的學習法, 但假設每項純粹行動所得到的報酬有時間上的趨勢, 因此玩家可以利用本期與上期報酬資訊, 來預測下一期各純粹行動的報酬, 據此調整自己下一期的混合行動。Conlisk (1993a) 證明當上一期的資訊被充分利用, 且混合行動的調整速度不大時, 可以收斂到穩定的混合均衡。在這個方法中, 玩家所擁有的資訊並沒有超越 Crawford 所給定的範圍, 只是玩家比較充分利用資訊。

第二種方法稱爲追求滿足法 (the satisficing process): 玩家在進入賽局前, 會在心裡預設一個報酬水準, 稱爲玩家的期許水準。Conlisk (1993a) 將期許水準納入玩家的考慮, 若本期玩家採用的混合行動能滿足他的期許, 則玩家下期將採用同樣的混合行動; 若本期混合行動不能讓他滿足, 則玩家採用 Crawford 學習法調整混合行動。另外, 玩家每一期也會調整期許水準, 使其『略大於』本期期許水準與報酬之加權平均。在此學習法下, Conlisk 證明玩家的混合行動雖然無法收斂到混合均衡, 但是比 Crawford 學習法更接近均衡。

第三種方法稱爲搜尋選擇法 (search and selection): 面對一個陌生的賽局時, 玩家會先在前幾期嘗試各種可能的混合行動, 然後在從中選擇平均

[1] Crawford (1974) 證明只有在 measure zero 的零和賽局裡才有穩定的混合均衡, 其中一個例子是該賽局裡所有報酬矩陣皆爲零, 不過這樣的零和賽局顯然不具備經濟理論所著重的一般性。

臺灣大學學術期刊資料庫

報酬最高者。Conlisk (1993b) 證明在此學習法下, 長期而言, 玩家各期報酬的平均值會收斂到均衡報酬附近, 可惜並非當期報酬收斂, 且玩家所採用的混合行動也未必收斂到混合均衡。[2]

最後一個方法稱爲防禦調整法 (defensive adaptation): 每一期玩家除了知道自己各純粹行動的報酬資訊外, 也知道對手的報酬資訊, 這比 Crawford 學習法裡的玩家擁有更多資訊。玩家的目的是調整自己的混合行動, 使對手各純粹行動的報酬最高值降低。若目前所採用的混合行動能有效達成此目的, 則下一期將採用此混合行動的鄰域 (neighborhood); 若不能達成目的, 則任意選取新的混合行動。Conlisk (1993b) 證明此學習法下, 混合行動的鄰域會機率收斂 (converge in probability) 到混合均衡, 但混合行動本身並不會收斂到混合均衡。

總結 Conlisk 所提出的四種途徑, 雖然只能逼近混合均衡, 或者在更強的條件下收斂到混合均衡, 但已經提供不少收斂到混合均衡的可能性。

1.2 研究動機、方法與文獻比較

本論文試圖探討, 在上述討論的兩人零和賽局中, 如何在長期下收斂到混合均衡。我們所分析的是只有唯一混合均衡的兩人零和賽局。考慮兩人重複玩此類型賽局無限多期, 每一期玩家心中設有期許水準, 並選取行動希望能滿足他的期許。我們沿用 Crawford (1974, 1985) 的設定, 每一期玩家選取 (混合) 行動之後, 將與對手重複玩 N 次賽局, 且 N 夠大使得玩家足以應用大數法則, 明白此行動能帶給他多少平均報酬。

給定特定玩家以及玩家自己目前的期許水準, 若當前的行動所帶來的平均報酬大於期許水準, 則玩家感到滿足, 下一期將繼續沿用此行動; 反之, 玩家下一期將會從他的可選策略集合中選擇在該期許水準下尚未採用的行動; 如果在此期許水準下, 所有的行動都無法滿足玩家, 下一期他將降低他的期許水準, 並重新搜尋能令他滿足的行動。在此設定下, 期許水準具有僵固性, 並不會每期調整, 必須等到所有行動都用盡後才調整。因此

[2]Conlisk (1993b) 舉了一個例子, 說明在此學習法下, 其中一位玩家的行動收斂到混合均衡, 但另一名玩家採用其他混合行動。Conlisk 特別提醒這是經過巧妙設計的特例, 對於一般化的情況, 他並未進一步分析。

臺灣大學學術期刊資料庫

調整期許水準的頻率小於調整行動的頻率。此外, 在沒有突變 (mutation) 的情況下, 如果玩家所採用的行動已經能夠滿足他了, 他的期許水準與行動將不再變動。加入突變的可能性後, 每一期玩家有極小的正機率改變他的期許水準或行動。

在我們所設定的學習法下, 如果沒有加入突變 (mutation), 整個動態調整過程將收斂到某個穩定狀態 (steady state), 此處所謂的穩定狀態意指雙方不再變動期許水準與行動。沒有突變的情況下, 動態調整過程存在許多不同穩定狀態。例如當雙方期許水準小於或等於0時, 兩人都採用混合均衡策略、或是兩人都採用「每次都出布」, 都是穩定狀態。動態調整過程最終會收斂到哪個穩定狀態, 取決於整個系統的初始狀態 (initial state) 以及之後的隨機過程實現結果。因爲上述的動態過程可能產生的穩定狀態數量會很多, 因此我們將突變引入模型中, 來研究穩定狀態的選擇問題。這裡的「突變」, 指的是玩家背離他應該採取的策略 (包括期許水準和行動), 這可能是因爲玩家有意識地實驗新策略, 也可能是玩家無意間犯了錯誤, 或是其他未能納入模型考量的變數所產生的隨機影響。

加入突變之後, 原先的穩定狀態將不再穩定, 在突變機率大於0時, 每個狀態 (包括前述的穩定狀態) 都有正的機率發生。這將導致, 整個系統最終收斂的結果, 與初始狀態無關。但是每個穩定狀態對擾動的敏感度不一, 不同的穩定狀態可能需要不同數量突變事件來破壞它的穩定性。此處我們應用 Young (1993, 1998) 的方法, 來篩選最爲穩定的隨機穩定狀態 (stochastically stable states)。這牽涉到計算每個穩定狀態的隨機位能 (stochastic potential), 根據 Young (1993) 的定義, 所有穩定狀態中隨機位能最小者, 即稱之爲隨機穩定狀態。表示加入隨機擾動之後, 此狀態相對於其他狀態還要穩定。隨機位能的概念類似水往低處流, 其他穩定狀態受到擾動後, 將「流向」隨機位能最小的狀態; 而隨機位能最小的狀態, 較難流向其他穩定狀態, 故此狀態最爲穩定。

我們考慮的突變有三種不同的形式: 一、玩家提高他的期許水準, 並從所有策略中任意選取新的策略; 二、玩家期許水準不變, 並從策略集合中任意選取新的策略; 三、玩家降低他的期許水準, 並從所有策略中任意選取新的策略。我們假設發生突變的機率足夠小, 但不爲0。其中第一種突變的發

臺灣大學學術期刊資料庫

生機率最大, 第三種最小。在上述設定之下, 我們證明雙方都使用「混合均衡策略」且「期許水準爲 0」的狀態是隨機穩定狀態, 但並不是唯一的隨機穩定狀態, 這是本文的主要結論。

以下針對本文的主要設定有三點說明。首先, 針對「行動」、「策略」及「學習法」在本文中的定義加以區分說明。「行動」指的是在某一期的某一次互動玩零和賽局時, 玩家所能做的行爲; 以剪刀石頭布爲例, 行動可以是「出剪刀」、「出石頭」、「出布」、或是「以各 1/3 的機率出剪刀、石頭、布」。「策略」則是玩家在某一期開始時, 決定在該期中的每一次互動要採取何種行爲; 例如「單數次出布且雙數次出剪刀」或「每次以各 1/3 的機率出剪刀、石頭、布」。因此,「行動」指的是玩家在單一次互動的行爲, 而「策略」則是玩家在某一期內每一次互動的行爲選擇。「學習法」規範玩家在每一期中如何調整期許水準以及「策略」; 例如, 給定前期的行動報酬以及當前期望水準, 是否繼續採用原有的策略或選擇新的策略, 是否要調高或降低期許水準。我們可以將整個賽局架構—每期玩 N 次零和賽局, 重複玩無限多期—看成是一個龐大的超賽局 (super-game): 行動屬於單次互動的零和賽局層次; 策略則爲一期內各次互動要採用哪些行動的規範; 而學習法則規範玩家如何調整期許水準以及策略, 是屬於超賽局層次的。

其次, 在模型中, 我們假設玩家有一個可選擇的策略集合 (strategy set), 這個集合是有限的。當玩家爲有限理性時, 假設策略集合也是有限, 應該是合理的。雖然理論上玩家的策略集合可以是無限的, 但現實上, 個人理性所能處理的日常事務, 例如貨幣、長度、時間等, 都有其極限或最小單位。所以, 在離散化 (discretized) 的環境中, 玩家所能建構並執行的策略數量其實是有限的。

最後, 我們假設玩家的策略集合包括並包含「混合均衡策略」(每一次互動都玩混合均衡行動)。當我們容許「玩家不知道自己正在參與賽局」的可能性時, 這樣的假設似乎有所矛盾, 其實不然。主要是, 如果玩家不知道自己正在參與賽局, 自然無法知道什麼是混合均衡行動, 但這和玩家把該策略納入可選擇的策略集合中並不一定牴觸。理由是, 如果零和賽局的報酬是有理數, 那麼混和均衡分配給各個純粹行動的機率, 也會是有理數。特別是如果混合均衡中選擇各個行動的機率是簡單分數或小數 (例如百分位

臺灣大學學術期刊資料庫

的小數), 那麼策略集合「恰好」包括混合均衡策略的假設, 就不會過於強烈。所以, 我們所採用的假設是, 玩家盡其理性嘗試各種可能的混合策略時, 其中正好包含均衡策略。由於人類的有限理性, 而且我們實際處理日常事務時, 基本上都在離散化的世界中思考及運作, 因此, 只要玩家可選擇的策略集合足夠大, 納入混合均衡策略應該還是屬於合理的假設範圍。當然, 如果賽局的混合均衡中選擇各個行動的機率是無理數或複雜的分數, 我們就需要讓玩家能獲取更多訊息, 例如賽局的報酬矩陣, 這樣假設玩家的策略集合包括「賽局的混合均衡策略」才比較合理。

我們所提出的學習法與 Conlisk (1993a) 同樣將期許水準納入玩家的考慮, 並且玩家只有在不滿足時才會更換行動。不同的是 Conlisk (1993a) 每期都會調整期許水準, 但我們假設玩家必須嘗試過所有行動之後, 才會調整期許水準。這種調整方法在精神上與 Conlisk (1993b) 的搜尋選擇法相似, 玩家面對陌生問題時, 很自然地會先嘗試各種行動或策略, 再做最後的決定。不同的是 Conlisk (1993b) 的玩家在嘗試過所有行動之後, 會選擇平均報酬最高者。但在我們的設定下, 玩家搜尋到能滿足的行動之後就停止搜尋了, 如果所有行動都失敗, 他會降低期許水準, 並重複搜尋的動作。

另外文獻上也有許多運用期許水準討論賽局的文章, 大部分討論的是追求滿足者如何演化出合作的能力, 例如 Pazgal (1997)、Dieckmann (1998)、Karandikar et al. (1998)、Oechssler (2002) 和 Napel (2003)。少部分應用在有風險的決策問題, 例如 Gilboa and Schmeidler (1996)、Börgers and Sarin (2000)。

此外, 有關玩家的認知能力方面, Lambson and Probst (2004) 提出模式相符法 (pattern matching): 每一期玩家知道自己與對手過去所採用過的行動組合, 從這些歷史紀錄中, 玩家觀察固定長度的『模式』, 統計每次特定模式出現時, 下一期對手所採用的行動之相對次數, 針對此相對次數, 選取純粹行動作最適反應。模式相符法裡的玩家需要更多理性的能力: 報酬矩陣的知識、自己與對手所採用過的行動組合、能從歷史紀錄中辨識出模式以及能做極大化效用的計算等。但是此法應用在 Shapley (1964) 的賽局時, 雙方採用各行動的相對次數仍然不會收斂到混合均衡, 這是因爲模式相符法本質上還是保有 fictitious play 的精神。有關模式相符的實證研究

臺灣大學學術期刊資料庫

亦可參考 Spiliopoulos (2012) 一文。此外, Schade, Schroeder, and Krause (2010) 分析事前的得失經驗對於受試者在協調賽局中行爲的影響。他們以實驗來檢測展望理論 (prospect theory) 中的價值函數是否具有良好的預測能力。結果發現, 男性受試者大致符合理論預測, 而女性受試者則較男性容易採取混合行動。Chong, Ho, and Camerer (2016) 建立了一般化的認知層次 (GCH) 模型, 藉以去了解玩家在參與賽局時, 爲何會無法收斂到均衡行爲。Pangallo, Heinrich, and Farme (2019) 證明, 在複雜或具競爭性質的賽局中, 最適反應行爲時常出現循環現象, 而且在六種常見的學習演算法中也無法收斂到均衡狀態。

文獻中討論追求滿足者的長期行爲時, 在調整行動的設定上大致相同: 玩家如果對行動感到滿意, 則繼續沿用, 反之則採用新的行動。但是除了 Conlisk (1993a,b)、Dieckmann (1998)、Börgers and Sarin (2000) 允許玩家採用混合行動外, 其餘各篇只討論純粹行動。而在我們的設定之下, 由策略所引導的行動可以是混合行動。其次, 除了 Dieckmann (1998) 設定玩家不能調整期許水準外, 其餘各篇皆允許自由調整。只是模型中的玩家每期都會調整期許水準, 調整頻率相當頻繁。而在我們的設定下, 玩家調整期許水準的頻率小於調整行動的頻率, 這是比較符合現實的設定。[3] 尤其是在玩家的行動不能滿足期許時, 玩家調降期許水準之前應先考慮使用其他行動。

雖然 Conlisk (1993a) 首先考慮用期許水準解決 Crawford 難題, 但他並未深入分析。就我們所知, 文獻上也沒有其他將期許水準應用在混合均衡的例子。本文是首篇正式將期許水準納入零和賽局, 並應用 Young (1993, 1998) 的方法來分析隨機穩定狀態的文章。我們證明在加入突變的情況下, 其包含混合均衡策略的狀態可以是隨機穩定的, 這是本文的主要貢獻。在第 3 章我們進一步延伸討論兩性戰爭賽局 (battle of sexes), 該賽局有兩個純粹均衡與一個混合均衡, 我們將說明混合均衡在這類賽局裡, 混合均衡也可以如純粹均衡一般穩定。

本文的主要貢獻在於, 目前文獻關於玩家如何在參與賽局的動態過程中, 如何學習去玩混合均衡策略的議題, 仍有不盡完善的地方。如同前面

[3]「調整期許水準的頻率小於調整行動的頻率」這樣的設定, 類似「法則的調整頻率小於行動的調整頻率」的設定。關於後者, 可參考 Juang (2002)。

臺灣大學學術期刊資料庫

所提, 目前仍存在玩家在長期下, 未能收斂到混合策略的問題。雖然在 fictitious play 之下, 玩家採用各項行動的頻率長期下會收斂到混合策略的機率, 但它僅是在長期平均下來的頻率收斂到混合策略, 實際上玩家每一次採用的行爲仍然是純粹的行動, 而不是真正的混合策略。本文所提出的學習法, 玩家在長期下, 可以真正地去選取混合均衡的策略, 而非僅是長期平均下來的頻率收斂到混合策略。

本文其它章節安排如下: 第2章介紹模型, 其中2.2節討論沒有突變的調整, 2.3節討論有突變的調整; 第3章討論各定理與命題的意涵, 並延伸探討純粹均衡與混合均衡的隨機穩定性; 第4章爲結論, 並提出一些延伸的研究方向; 所有數學證明則置於附錄。

2 模型

2.1 基本設定

我們令 $\Gamma_Z = (I, A, u)$ 爲兩人零和賽局的標準式 (normal form)。其中 $I = \{1, 2\}$ 爲玩家的集合; $A = A_1 \times A_2$ 爲兩位玩家的有限純粹行動組合的集合, 給定任一玩家 $i \in I$, $a_i \in A_i$ 爲其可選擇的行動; $u = u_1 \times u_2$, 其中 $u_i \colon A \to R$ 爲玩家 i 的報酬函數。由於 Γ_Z 爲零和賽局, 對所有 $a \in A$, 我們有 $u_1(a) + u_2(a) = 0$ 的關係。

Γ_Z 的混合拓展形式 (mixed extension form) 爲 (I, Σ, U)。在此 $\Sigma = \Sigma_1 \times \Sigma_2$ 爲分佈在 $A_1 \times A_2$ 上之機率分配組合的集合, $\sigma_i \in \Sigma_i$ 爲玩家 i 的混合行動,[4] 其賦予每個 $a_i \in A_I$ 一個機率值 $\sigma_i(a_i) \geq 0$, 且 $\Sigma_{a_i \in A_i} \sigma_i(a_i) = 1$。$U_i : \Sigma \to R$ 爲 i 的期望報酬函數, 給定任意 $\sigma \in \Sigma$, 可以得到一組 A 的機率分佈, 並根據 u_i 計算出 i 的期望報酬 U_i, 意即 $U_i(\sigma) = \Sigma_{a \in A} \sigma_1(a_1) \cdot \sigma_2(a_2) \cdot u_i(a)$。給定玩家 i 的對手 j 的混合行動 $\sigma_j \in \Sigma_j$, i 的最適反應 (best response) 集合爲 $BR_i(\sigma_j) \equiv \arg\max_{\sigma_i \in \Sigma_i} U_i(\sigma_i, \sigma_j)$。

以下我們只考慮具有唯一混合 Nash 均衡, 均衡時雙方期望報酬爲0的零和賽局: Γ_M。其定義如下:

[4]在一般賽局中, 應該稱呼 σ 爲混合策略。然而從本文超賽局的角度, 我們改期稱呼爲混合行動以避免與超賽局的策略一詞混淆。

臺灣大學學術期刊資料庫

定義 1. Γ_M 爲符合下列兩條件的零和賽局:

1. Γ_M 只有唯一 Nash 均衡, 記爲 $\sigma^* \equiv (\sigma_1^*, \sigma_2^*) = \{\sigma \in \Sigma | \sigma \in BR(\sigma)\}$, 其中 $BR(\sigma) \equiv BR_1(\sigma_2) \times BR_2(\sigma_1)$。
2. 對所有 i, $U_i(\sigma^*) = 0$。

由於玩家需要足夠多回合的賽局結果, 才能對於所採用策略究竟可以爲其帶來多少報酬有一準確的資訊, 因此我們參考 Crawford (1974, 1985) 的作法, 把 Γ^∞ 稱爲一個超賽局 (super-game), 其結構爲: 每一期 (period) 兩人選取一個策略重複玩 N 次 Γ_M, 每一次玩 Γ_M 稱之爲一次互動 (interaction)。每期有 N 次互動, 如此重複玩無限多期。t_{mn} 表示第 m 期第 n 次互動的時間, $m = 1, \cdots, \infty$, $n = 1, \cdots, N$, Γ^∞ 的時間序列爲 $(t_{11}, t_{12}, \cdots, t_{1N}, t_{21}, \cdots, t_{2N}, \cdots, t_{mn}, \cdots)$[5]

i 在 Γ^∞ 的行爲可視爲在不同時間 t_{mn} 其擁有的資訊 I_i^{mn} 與可能行動 Σ_i 的一種對應關係。其中資訊 I_i^{mn} 裡可以包含例如像是 Γ_M 的報酬矩陣、對手的行動集合、i 對於參與賽局的理解程度等等內容。我們假設兩人在同時間的資訊可以不相等, 即 $I_1^{mn} \neq I_2^{mn}$, 用來反應兩個玩家資訊可能的差異。i 在 Γ^∞ 的所有可能的策略集合定義爲 χ_i。x_i^m 表示 i 在第 m 期當期所採用的策略。$\sigma_i^{mn} = x_i^m(I_i^{mn})$ 爲給定策略 x_i^m 和資訊 I_i^{mn} 時, i 在時間 t_{mn} 時採取的混合行動。因此給定兩人在第 m 期的策略 $x^m \equiv (x_1^m, x_2^m)$ 時, 兩人在時間 t_{mn} 時的混合行動組合可記爲 $\sigma^{mn}(x^m) \equiv (\sigma_1^{mn}(x_1^m), \sigma_2^{mn}(x_2^m))$。因此 i 在超賽局中的行爲就由三個層次構成: 同一期內的各次行動 (由同一策略 x_i^m 但各次互動時的資訊 I_i^{mn} 所決定)、不同期採取的各期策略、與如何轉換各期策略的學習法。

我們假設玩家的理性能力是有限的, $X_i \subset \chi_i$ 表示在玩家理性範圍內, 他所有可能採用的策略之集合, 又稱玩家 i 的理性策略集合。對所有 i, X_i 有兩個假設:

[5]雖然 Crawford (1974, 1985) 並沒有明確定義如本文中的超賽局結構, 但實際上兩者在過程上是相同的。考慮到在本文中, 除了策略可能調整之外, 還牽涉到期望水準的調整: 兩者都在每期玩 N 次賽局後, 根據結果來決定是否調整策略或是期許水準。因此, 我們藉由模型中超賽局的定義讓整個賽局過程結構更清楚且容易被理解。

臺灣大學學術期刊資料庫

假設 1. X_i 爲有限集合。

假設 2. X_i 包含混合行動策略 x_i^* 所形成的集合 $\{x_i \in \chi_i | x_i(I_i^{mn}) = \sigma_i^*, \forall I_i^{mn} \in I_i\}$。

x_i^* 其實就是玩家每次互動皆採用 Γ_M 的混合均衡策略, 同樣地, X_1 不必等於 X_2, 這反應兩個玩家在理性能力上可能存在的差異。

在模型中, 我們假設玩家的策略集合爲有限多個, 這基於以下兩點考量。第一、我們把模型中的玩家, 假設爲有限理性的個體。因爲玩家的理性是有限的, 因此能夠處理的訊息或是策略, 也會有所限制。現實環境中, 不論能力多強, 個人也不會是擁有無限理性的, 他能選擇的策略也不會是無限多的; 或者說, 即使知道可能的策略有無限多, 但實際在進行選擇時, 也不是真的在無窮多的選項中去進行選擇。第二、在運用隨機穩定 (stochastic stability) 的技術來進行均衡狀態的選擇時, 文獻都會假設個人的策略集合是有限的。這是因爲技術上, 狀態空間必須是有限多, 才能夠進一步去分析。在有限的狀態空間之下, 策略集合也必須是有限的, 這是屬於技術上的考量。

接著我們定義 $v_i^m \in \{\cdots, -2k, -k, 0, k, 2k, \cdots\}, k \in R^+$ 爲 i 在第 m 期的期許水準 (aspiration level)。定義域爲實數的等差級數, 級距爲 k。對 i 在第 m 期時, 定義 $pre(m)$ 爲 i 目前期許水準是在第 $pre(m)$ 期調整採用的, 即 $pre(m) \equiv \min\{1 \leq \alpha \leq m | v_i^\beta = v_i^m \forall m \geq \beta \geq \alpha\}$。$T^m \equiv \{pre(m), \cdots, m\}$ 爲 i 調整期許水準後至今的各期集合。在 T^m 期間裡, i 採用過的所有策略之集合記爲 $X_i^m \equiv \{x_i \in X_i | \exists \alpha \in T^m, s.t. x_i = x_i^\alpha\}$, 其補集爲 $(X_i^m)^c$, 表示從調整期許水準後, i 尚未使用過的策略集合。

每一期兩人重複玩 N 次零和賽局 Γ_M, 根據緒論的說明, 我們假設 N 夠大使玩家省去統計推論的瑣碎細節, 因此每期玩家得到的平均報酬等於期望報酬, 故可令 $\mu_{x_i}^m(x_j) \equiv \lim_{N\to\infty}(1/N)\Sigma_{n=1}^N u_i^{mn}$ 表示給定 j 採用策略 x_j 時, i 在第 m 期採用策略 x_i 的平均報酬。

最後定義 $\omega \equiv (v_1, v_2, x_1, x_2, (X_1)^c, (X_2)^c)$ 爲超賽局 Γ^∞ 的狀態 (state), 包含兩人的期許水準、使用的策略以及尚未使用過的策略集合, 各變數之定義如上, 若加上上標 m 則表示在第 m 期的狀態。Ω 表示所有可能的狀態聯集成的狀態空間 (state space), 爲一個有限 (finite) 的集合。

臺灣大學學術期刊資料庫

有了上述各定義之後, 下一小節我們將討論在沒有突變情況下, 玩家的策略與期許水準如何調整。

2.2 沒有突變時的調整學習法

給定玩家的期許水準, 每一期玩家選擇一個策略玩 N 次零和賽局 Γ_M, 若該策略帶來的平均報酬大於玩家的期許水準, 玩家感到滿足, 下一期將繼續使用此策略; 若平均報酬小於玩家的期許水準, 下一期玩家將從上次調整期許水準後, 尚未使用過的策略中, 隨機採用一個。若所有策略的平均報酬都無法滿足玩家, 下一期玩家將降低他的期許水準一個位階值 k, 並重新從所有策略中任意選取新的策略。正式的數學形式表示如下:

定義 2. Φ^0 表示沒有突變時, 玩家在超賽局 Γ^∞ 裡調整策略與期許水準的過程。定義 Φ^0 過程如下: 給定任意期 m 以及 $i \in \{1, 2\}$,

(1) 如果 $\mu^m_{x_i}(x_j) \geq v^m_i$, 則 $v^{m+1}_i = v^m_i$ 且 $x^{m+1}_i = x^m_i$。
(2) 如果 $\mu^m_{x_i}(x_j) < v^m_i$ 且 $(X^m_i)^c \neq \varnothing$, 則 $v^{m+1}_i = v^m_i$, $x^{m+1}_i \in (x^m_i)^c$ 且 $\mathbf{Pr}(x^{m+1}_i = x_i | x_i \in (X^m_i)^c) > 0$。
(3) 如果 $\mu^m_{x_i}(x_j) < v^m_i$ 且 $(X^m_i)^c = \varnothing$, 則 $v^{m+1}_i = v^m_i - k$, $x^{m+1}_i \in X_i$ 且 $\mathbf{Pr}(x^{m+1}_i = x_i | x_i \in X_i) > 0$。

本文對玩家爲有限理性的假設, 其實際涵義可從定義2-1看出。在整個超賽局架構下, 如果玩家的理性爲無窮大, 那麼他應該可以運用從第一期開始到最近一期的所有訊息, 而非僅是最近一期。同時, 一個無限理性的玩家, 甚至可以從對手在過去的行爲中, 去推測對手接下來可能如何行動。然而因爲我們假設玩家爲有限理性, 因此在上述動態過程中, 玩家在調整期許水準或策略時, 僅需要知道最近一期採用的策略、該策略最近一期的平均報酬、最近一期的期許水準、以及尚未採用的策略集合等。他也不需要知道對手過去做過什麼, 也不會去推測對手接下來會怎麼做。這是本文中, 玩家爲有限理性的假設和文獻中無限理性的差異。

在沒有突變的調整過程中, 一旦達到某狀態, 便一直停留在該狀態的話, 則稱該狀態爲穩定狀態 (steady state)。其定義爲:

臺灣大學學術期刊資料庫

定義 3. 在 Φ^0 過程裡, 一個狀態 ω 被稱之爲穩定狀態, 若且爲若 ω 符合下面條件: 對任意期 m_0, 如果 $\omega^{m_0} = \omega$, 則對所有 $m > m_0$, $\omega^m = \omega$。

爲了分析上的方便, 我們另外定義以下兩個符號:

定義 4. $\Omega_E \equiv \{\omega \in \Omega | \omega$ 爲穩定狀態$\}$表示 Φ^0 過程裡所有穩定狀態之集合。

定義 5. Ω_N 定義爲所有兩人期許水準皆爲 0, 且都採用 x_i^* 的狀態之集合, 意即

$$\Omega_N \equiv \Big\{\omega \in \Omega | \omega = \left(0, 0, x_1^*, x_2^*, (X_1)^c, (X_2)^c\right), \\ (X_i)^c \subseteq X_i \left\{x_i^*\right\} \forall i \in I\Big\}。$$

根據 Φ^0, 可以得到以下命題與定理。

命題 1. 在 Φ^0 過程裡, 若 $\omega = (v_1, v_2, x_1, x_2(X_1)^c, (X_2)^c) \in \Omega_E$, 則 $v_1 + v_2 \leq 0$。

命題 2. 在 Φ^0 過程裡, $\Omega_N \subseteq \Omega_E$。

命題 3. $\Omega_E \setminus \Omega_N \neq \emptyset$。

定理 1. 在超賽局 Γ^∞ 裡, 若玩家的理性策略集合 (X_1, X_2) 符合假設 1 跟假設 2, 則給定任意初始狀態 (initial state) ω^0, Φ^0 過程保證會收斂到穩定狀態。

上述命題與定理的完整數學證明請見附錄 A。直覺上, 一個狀態若是穩定狀態, 表示兩人所選定的策略, 每一期都能令他們滿足。因爲考慮的是零和賽局 Γ_M, 兩人報酬之和爲 0, 如果要讓兩人能同時滿足, 則兩人期許水準之和必然要小於等於 0, 此即命題 1。另外, 在零和賽局 Γ_M 裡, 當玩家每次互動都採用唯一的混合均衡行動 σ^* 時, 保證可以得到零平均報酬。所以在超賽局 Γ^∞ 裡, 當兩人期許水準爲 0, 且皆採用混合均衡策略 x_i^* 時, 必能在每一期同時滿足兩人, 因此達到穩定狀態, 此即命題 2。除了上述的穩定狀態外, 超賽局 Γ^∞ 還有許多其他的穩定狀態, 以剪刀石頭布的賽局

臺灣大學學術期刊資料庫

爲例, 若兩人期許水準皆爲0, 且兩人只出剪刀, 則兩人平均報酬亦爲0, 這也是一組穩定狀態, 更一般化的例子請見附錄 A 裡命題3的例證。

最後, 定理1保證在超賽局 Γ^{∞} 裡, Φ^0 過程會收斂到穩定狀態。概念上, 如果 Φ^0 不會收斂到穩定狀態, 則永遠有人要調整策略或降低期許水準。因爲策略爲有限個, 因此當所有策略皆無法滿足期許水準時, 接著就要調降期許水準。當兩人期許水準都在0以下時, 至少有混合均衡策略 x_i^* 可令 Φ^0 達到穩定狀態。定理1的證明, 大致上是依照上述的邏輯來推論。

上述定理及命題保證 Φ^0 會收斂到穩定狀態, 但不只有一個穩定狀態。下一小節我們將加入突變, 使得此處陳述的某些穩定狀態將變化爲不穩定 (unstable)。

2.3 有突變時的調整學習法

定理1告訴我們給定任何初始狀態, 最後一定會調整到某穩定狀態。現在我們考慮有突變 (mutation) 的情況。突變有三種不同的形式: 一、玩家提高期許水準, 並從所有策略中任意選取新的策略, 其發生機率爲 η^b (其中 $\eta \in (0, 1)$); 二、玩家期許水準不變, 並從所有策略中任意選取新的策略, 其發生機率爲 η^c; 三、玩家降低期許水準, 並從所有策略中任意選取新的策略, 其發生機率爲 η^d。我們假設發生各種突變的機率非常微小, 但不爲0, 其中第一種突變的發生機率最大, 第三種最小, 意即我們假設 $1 < b < c < d$。這假設隱含兩個意義: 玩家「犯錯」的機率比「實驗」的機率小, 且玩家之所以要「實驗」, 是因爲更大的企圖心。在穩定狀態下, 玩家選取的策略已經能滿足他的期許水準了, 此時玩家不應該調降期許水準, 所以發生第三種突變, 表示玩家「犯錯」, 這種突變的機率應該最小; 其次, 玩家有可能在穩定狀態時「實驗」新的策略, 在日常生活裡, 如果玩家這麼做, 通常是想試試能不能得到更高的報酬。如果玩家沒有這樣的企圖心, 既然策略已經能滿足他了, 似乎沒有必要實驗新的策略。換言之, 玩家在實驗新的策略時, 常常伴隨著更大的企圖心, 所以發生第一種突變的機率應比第二種高。正式的數學形式表示如下:

定義 2-2. Φ^{η} 表示加入突變時, 玩家在超賽局 Γ^{∞} 裡調整策略與期許水準的過程。該過程爲: 對所有 i, 在任意期 m,

臺灣大學學術期刊資料庫

(4) 有 $(1-\eta^b-\eta^c-\eta^d)$ 的機率: 依照 Φ^0 過程來調整。

如果 $\mu^m_{x_i}(x_j) \geq v^m_i$, 則 $v^{m+1}_i = v^m_i$ 且 $x^{m+1}_i = x^m_i$;

如果 $\mu^m_{x_i}(x_j) < v^m_i$ 且 $(X^m_i)^c \neq \varnothing$, 則 $v^{m+1}_i = v^m_i$, $x^{m+1}_i \in (X^m_i)^c$ 且 $\mathbf{Pr}(x^{m+1}_i = x_i | x_i \in (X^m_i)^c) > 0$;

如果 $\mu^m_{x_i}(x_j) < v^m_i$ 且 $(X^m_i)^c = \varnothing$, 則 $v^{m+1}_i = v^m_i - k$, $x^{m+1}_i \in X_i$ 且 $\mathbf{Pr}(x^{m+1}_i = x_i | x_i \in X_i) > 0$。

(5) 有 η^b 的機率: $v^{m+1}_i > v^m_i$, $x^{m+1}_i \in X_i$ 且 $\mathbf{Pr}(x^{m+1}_i = x_i | x_i \in X_i) > 0$。

(6) 有 η^c 的機率: $v^{m+1}_i = v^m_i$, $x^{m+1}_i \in X_i$ 且 $\mathbf{Pr}(x^{m+1}_i = x_i | x_i \in X_i) > 0$。

(7) 有 η^d 的機率: $v^{m+1}_i < v^m_i$, $x^{m+1}_i \in X_i$ 且 $\mathbf{Pr}(x^{m+1}_i = x_i | x_i \in X_i) > 0$。

其實 Φ^η 的 (4) 式即 Φ^0 過程。Φ^η 與 Φ^0 皆爲定義在狀態空間 Ω 的 Markov 過程, 其狀態個數是有限可數的。Φ^η 是由無擾動的過程 Φ^0 加入輕微擾動而得。我們可以選擇一個足夠小的數 $\eta^* > 0$, 讓 η 爲在區間 $(0, \eta^*]$ 的任意值。令 P^0、P^η 分別表示 Φ^0、Φ^η 的推移機率矩陣 (transition probability matrix), $P^0_{\omega\omega'}$、$P^\eta_{\omega\omega'}$ 分別表示在 Φ^0、Φ^η 過程裡, 狀態 ω 變化到狀態 ω' 的機率。

顯而易見 Φ^η 過程具有 aperiodic 與 irreducible 性質, 既然如此, Φ^η 對所有 $\eta \in (0, \eta^*]$, 皆存在唯一的定態分配 (stationary distribution) π^η。根據 Young and Foster (1991) 之定義, 在 Φ^η 過程裡, 對狀態 ω 而言, 若 $\lim_{\eta\to 0}\pi^\eta(\omega) > 0$, 則稱 ω 爲隨機穩定狀態 (stochastically stable state)。若一個狀態不是隨機穩定狀態, 則表示在有擾動的情況下, 長期而言, 此狀態在調整過程中出現的相對次數爲 0。以下我們採用 Young (1993, 1998) 的方法, 辨識出 Φ^η 裡的隨機穩定狀態。

首先根據 Φ^0、Φ^η 的定義, Φ^η 過程符合 Young (1993) 對 regular perturbed Markov process 的定義:

1. Φ^η 對所有 $\eta \in (0, \eta^*]$ 皆具有 aperiodic 與 irreducible 性質。

臺灣大學學術期刊資料庫

2. $\lim_{\eta\to 0} P^{\eta}_{\omega\omega'} = P^{0}_{\omega\omega'}$。
3. 若存在某些 $\eta > 0$ 使得 $P^{\eta}_{\omega\omega'} > 0$, 則存在某些 $r(\omega,\omega') \geq 0$ 使得 $0 < \lim_{\eta\to 0} \eta^{-r} P^{\eta}_{\omega\omega'} < \infty$。$r(\omega,\omega') \in 0$ 稱爲狀態 ω 變化到狀態 ω' 的阻力 (resistance), 給定任意一組狀態 (ω,ω'), 只能找到唯一的阻力 $r(\omega,\omega')$ 滿足上述條件。$r(\omega,\omega') = 0$ 若且爲若 $P^{0}_{\omega\omega'} > 0$。另外, 若對所有 $\eta \in (0,\eta^*]$, $P^{\eta}_{\omega\omega'} = P^{0}_{\omega\omega'} = 0$, 則定義 $r(\omega,\omega') = \infty$。我們在2.1節證明了在超賽局 Γ^{∞} 裡, Φ^0 過程存在許多穩定狀態, 且一定會收斂到某穩定狀態。考慮任意兩個穩定狀態 (ω,ω'), 若要從 ω 變化到 ω', 必然需要突變, 若突變機率爲 η^{θ}, 則令 $r(\omega,\omega') = \theta$。如此, 對任意一組狀態 (ω,ω'), $r(\omega,\omega')$ 皆有良好定義。

令 $(e_1, e_2, \cdots, e_L)$ 表示 Φ^0 過程裡所有的 recurrent classes, 顯然在本文的設定中, Φ^0 過程裡每一個 recurrent class 只包含一個狀態, 且必爲 Φ^0 的穩定狀態。對任意兩個不同的 recurrent classes: (e,e'), $\varsigma(e,e') = (e,\omega_1,\omega_2,\cdots,\omega_q,e')$ 表示以 e 爲起點, e' 爲終點的路徑。此路徑的阻力爲路徑上各邊阻力之總和, 意即 $r(\varsigma) = r(e,\omega_1) + r(\omega_1,\omega_2) + \cdots + r(\omega_q,e')$。令所有以 e 爲起點, e' 爲終點的路徑集合爲 $Z(e,e')$。因此, 從 e 到 e' 路徑的阻力 $r(e,e') = \min_{\varsigma\in Z(e,e')} r(\varsigma)$, 意即所有 $\zeta(e,e')$ 路徑中阻力的最小值。

接著定義一個有向圖 (Directed graph) G: 其各頂點爲 $\{e_1, e_2, \cdots, e_L\}$, 對任意一組頂點 (e,e'), 邊 $e \to e'$ 表示由穩定狀態 e 變化到另一個穩定狀態 e', 令邊 $e \to e'$ 的權數爲阻力 $r(e,e')$。接著定義 e-tree 爲有向圖 G 的一部份, 對所有頂點 $e' \neq e$, 都可以找到唯一路徑由 e' 變化到 e。因此每個 e-tree 都有 $|\Omega_E| - 1$ 個邊, 每個邊上都有一個阻力 r, 因此可以計算出該 e-tree 阻力的總和。給定穩定狀態 e 爲根 (root), 我們可以建構出不同的 e-tree, 令 γ_e 爲所有以 e 爲終點的 e-tree 中, 阻力總和最小的值, γ_e 稱爲穩定狀態 e 的隨機位能 (stochastic potential)。根據 Young (1993) 的定理 4, 因爲 Φ^{η} 爲 Φ^0 的 regular perturbed Markov process, 所以當 $\eta \to 0$ 時, 狀態 ω 爲 $\Phi^{\eta\to 0}$ 過程裡的隨機穩定狀態, 若且爲若 ω 爲隨機位能最小的穩定狀態。[6]

[6]關於 e-tree、隨機位能、隨機穩定狀態等概念, 可參考 Young (1993, 1998) 之說明。

臺灣大學學術期刊資料庫

應用 Young (1993) 的方法, 令 Ω_S 表示所有隨機穩定狀態之集合 ($\Omega_S \subseteq \Omega_E$)。另外定義 $e^* = (0, 0, x_1^*, x_2^*, X_1 \backslash \{x_1^*\}, X_2 \backslash \{x_2^*\})$, 表示雙方期許水準爲 0, 且在調整到 0 期許水準時, 立即採用混合均衡策略之狀態。根據 Φ^η, 可以得到以下命題與定理:

命題 4. 在 Φ^η 過程裡, $e^* \in \Omega_S$。

命題 5. 在 Φ^η 過程裡, 對所有 $e \in \Omega_E$, 若 $e = (0, 0, x_i^*, x_j, X_i \setminus \{x_i^*\}, X_j \setminus \{x_j\})$, 則 $e \in \Omega_S$。

命題 6. 在 Φ^η 過程裡, 若 $e = (v_1, v_2, x_1, x_2, (X_1)^c, (X_2)^c) \in \Omega_S$, 則 $v_1 = v_2 = 0$。

定理 2. 在超賽局 Γ^∞ 裡, 若玩家所能採用的策略集合 (X_1, X_2) 符合假設 1 跟假設 2, 則 Φ^η 過程保證 $e^* \in \Omega_S$, 但 $\Omega_S \setminus \{e^*\} \neq \varnothing$。

定理 2 可由命題 4、命題 5 得證, 各命題的完整數學證明請見附錄 B。命題 4 與定理 2 說明在超賽局 Γ^∞ 裡, 若加入突變的可能性, Φ^η 過程保證當雙方期許水準皆爲 0, 且都採用混合均衡策略 x_i^* 時, 必爲隨機穩定狀態。這是本文最重要的結論, 表示我們所提出的學習法, 部分解決了 Crawford 難題, 不過隨機穩定狀態並不唯一。命題 5 爲隨機穩定狀態的充分條件, 當雙方的期許水準都爲 0 時, 只要有一方採用混合均衡策略, 則不管對手採取什麼策略, 都是隨機穩定狀態。命題 6 爲隨機穩定狀態的必要條件, 在互有輸贏零和賽局裡, 長期而言, 雙方對賽局的期許都爲 0。

3 討論

3.1 關於定理 2 之討論

定理得到混合均衡並不是整個動態調整過程中唯一的隨機穩定狀態。如果雙方期許水準爲 0, 只要有一人採用混合均衡策略, 另一人不管採用什麼策略, 都是隨機穩定狀態。以剪刀石頭布爲例, 如果兩人的期許水準都是 0, i 以各 1/3 的機率出剪刀、石頭、布, 而 j 只出布, 也是一個隨機穩定狀態。這個發現跟 1.1 節註 2 裡 Conlisk (1993b) 所舉的例子相似, 我們可以從期許水準的角度進一步解釋這樣的結果。

臺灣大學學術期刊資料庫

在我們所設定的期許水準模型裡, 雖然假設玩家能採用的策略爲有限個, 且必須包含混合均衡策略, 但是對於玩家對賽局結構的資訊掌握不做出太多的要求。我們所設定的學習法, 只要求玩家選取的策略滿足自己的期許水準。如果他所採用的策略能夠滿足期許水準的要求, 他便會一直使用此策略。例如玩家可以從未被告知對手的行動, 只知道當他選擇以各1/3的機率出剪刀、石頭、布時, 一期下來他的平均報酬爲0。如果他的期許水準不大於0。那麼下期他就繼續同樣的策略。如果玩家考慮更多, 例如他分析在同一期內出剪刀的報酬是否平均而言比出石頭大, 那他可能做出一些推理來調整下一期的策略。所以期許水準模型下的玩家可以說是相當被動地處理其理性能力。這個特性是追求滿足者與極大化效用者最大的不同。此時就可能出現命題5的情況: 其中一位玩家使用混合均衡策略, 另一玩家使用任一策略。定理2雖然無法保證混合均衡是唯一隨機穩定狀態, 但從上述分析可知, 隨機穩定狀態的集合範圍與玩家的知識和理性能力有關。當玩家對賽局的認知越匱乏時, 命題5的情況就越容易出現。反之, 玩家擁有越充分的資訊, 就越能排除命題5的情況。

我們模型中的學習法可以擴展爲一個對理性能力較爲一般化的設定, 允許各種可能的理性能力。例如學習法改爲當某期某策略如果平均報酬高於期許水準, 則下一期將會調高期許水準至上一期的平均報酬並且繼續採取此策略; 若某期某策略帶來的平均報酬等於或低於期許水準, 就進行定義2-1中的行爲。這樣的修改下定理2仍然成立, 同時更能解釋對追求報酬更積極的玩家行爲。[7] 如果對照 Crawford (1974, 1985) 的設定中所有玩家僅擁有同一種理性策略的學習法, 因此其容許的穩定狀態就少於本文的可能穩定狀態。

雖然我們所提出的學習法目的在討論零和賽局的收斂性, 但也可以用來比較玩家所採用的各種學習法。例如1.2節提到過 Lambson and Probst (2004) 的模似相符法。在他們的設定裡, 玩家只能觀察固定長度的模式, 若 i 所觀察的模式長度大於 j, 則可能出現大部分時間由 i 獲勝的情況。但這在我們的設定之下, 將期許水準納入 j 的考慮, 且 j 的理性策略包括使用其他長度的模式, 就不會出現這種情況。對照股市技術分析的例子, 若

[7] 作者感謝匿名評審對此一擴展方向提供的寶貴建議。

臺灣大學學術期刊資料庫

分析師觀察月均線所做的投資皆失敗, 而他希望能獲得一定的報酬時, 他很自然會想採用其他均線作分析。因此, 玩家的理性策略應該包含不同長度的模式。當玩家的理性策略懂得使用不同長度的模式時, Lambson and Probst (2004) 所提出的例子就不會是穩定狀態。

3.2 兩性戰爭賽局的討論

在一個同時有純粹均衡與混合均衡的賽局裡, 我們可能認爲單純行動均衡應該比混合行動均衡更穩定, 意即純粹均衡的隨機位能較小。但是如果以下述的兩性戰爭賽局 (battle of sexes) 爲例, 我們可能得到不同的結果。考慮下面形式的兩性戰爭:

$1 \setminus 2$	a'	a''
a'	(2,1)	(0,0)
a''	(0,0)	(1,2)

此賽局有兩個純粹均衡: (a', a') 與 (a'', a''), 一個混合均衡: 玩家 1 玩 a' 的機率爲 2/3, 玩家 2 玩 a' 的機率爲 1/3, 均衡時雙方報酬爲 2/3。假設雙方只有三種理性策略: x' 爲每次皆採用 a'、x'' 爲每次皆採用 a''、x^* 爲每次皆採用混合均衡策略。

考慮 $e = (2, 1, x', x', \{x'', x^*\}, \{x'', x^*\})$ 與 $e^* = (2/3, 2/3, x^*, x^*, \{x', x''\}, \{x', x''\})$, 此兩狀態皆爲穩定狀態之一。前者爲純粹均衡, 後者爲混合均衡。根據 2.3 小節對隨機位能與隨機穩定狀態的定義, 我們將找出一個以 e 爲起點, e^* 爲終點的路徑 $\zeta(e, e^*)$, 該路徑阻力爲 b, 所以這兩個穩定狀態的阻力 $r(e, e^*)$ 亦爲 b。因此如果 e 是隨機穩定狀態, 則 e^* 必然也是隨機穩定狀態。

首先考慮 $e = (2, 1, x', x', \{x'', x^*\}, \{x'', x^*\})$, 令玩家 2 發生第一種突變: 提高期許水準爲 $(1 + k)$, 並採用策略 x''。在新狀態 $\omega_1 = (2, 1 + k, x', x'', \{x'', x^*\}, \{x', x^*\})$ 下, 雙方報酬爲 0, 兩人都不能滿足, 下一期將改變策略。令兩人改變策略時, 若 i 採用策略 x', 則 j 採用策略 x'', 意即兩人總是不能合作, 或者兩人同時採用 x^*。根據這樣調整策略的方式, 雙

臺灣大學學術期刊資料庫

方各期的報酬必爲 0 或 2/3, 策略用盡後, 兩人必須調降期許水準。若 i 的期許水準降到 2/3, 則令 i 採用策略 x^*, 否則維持上述的調整方式。

若 i 的期許水準爲 2/3 且採用策略 x^*, 則 i 發現每期的報酬都可令他滿足, i 將不再有任何變動。若此時 j 的期許水準大於 2/3, 則不管 j 採用什麼策略都不能滿足, 故長期而言, j 必然要調降期許水準爲 2/3, 此時可令 j 採用策略 x^*。因此動態過程到達穩定狀態 e^*。

根據上述的設定, 可以找到一個路徑 $\zeta(e, e^*) = (e, \omega_1, \omega_2 \cdots, \omega_q, e^*)$, 其各邊阻力除了 $r(e, \omega_1)$ 等於 b 之外 (發生第一種突變), 其餘各邊阻力皆爲 0 (未發生突變), 所以該路徑阻力爲 b, 因此此兩個穩定狀態的阻力 $r(e, e^*)$ 亦爲 b。因此, 混合均衡也可以如純粹均衡一般穩定。

從上述的分析可以發現, 期許水準可解釋爲玩家對於參與賽局預期或希望能夠獲取的報酬。當玩家 2 發現其在穩定狀態 e 雖然可獲得 1 單位的報酬, 但對手卻能得到 2 單位 (當對手報酬可觀察時) 或是單純基於想要更好的企圖心 (此時不需知道對手的訊息), 都可能引發玩家 2 提高其期許水準。這類的改變隨即引發策略的改變, 導致兩位玩家因爲合作失敗而報酬大減。在隨後的調整策略時, 由於協調失靈而持續合作失敗, 接著只好降低期許水準。即使降低期許水準, 但並不能排除雙方持續在策略上協調失靈 (x' 總是對上 x''), 因此繼續降低期許水準。這個階段的期許水準降低, 並非一開始企圖更好的動機 (藉由突變來達成), 而是現實未能符合理想下的自動調整 (不需突變)。直到雙方都調降期許水準到 2/3 且採用策略 x^* (這種情況發生的機率爲正), 此時兩位玩家所採用的策略恰好符合各自的期許水準, 而達到穩定狀態。而玩家的學習法, 則是玩家在參與賽局時, 如何調整期許水準和策略的規範: 給定各種突變發生的機率, 若當前策略可獲得不低於期許水準的報酬, 則策略不變; 否則改變策略。若所有策略都達不到當前的期許水準, 則調低期許水準。之所以會得到混合均衡和純粹均衡一樣穩定的結果, 一個主要原因是在上述學習法中, 突變發生的機率只和調高或調低期許水準及改變策略有關, 但是和所處狀態的報酬無關。這種突變機率和報酬無關的設定, 自從 Young (1993) 以來一直是相關文獻的主要設定。如果我們假設報酬愈高時突變發生的機率愈小, 那麼相信在兩性戰爭這個賽局裡, 混合均衡應該就不會和純粹均衡一樣穩定了。不過

臺灣大學學術期刊資料庫

因爲本文的重點爲混合均衡在零和賽局的穩定性, 因此假設突變機率和報酬成負相關的議題將留待未來探討。但是上述的討論也說明了, 學習法的設定對於分析的結論扮演著相當重要的角色。

此外, 這個例子有幾個值得注意的地方, 1.2 小節說明過 Dieckmann (1998) 考慮多人 2 × 2 賽局, 他證明若玩家期許水準固定不變, 則即使允許玩家採用混合行動, 最終仍會收斂到對野心大者有利的純粹均衡。期許水準固定不變是 Dieckmann (1998) 的重要假設, 但在兩性戰爭裡, 雖然合作比不合作好, 但是合作的方式會影響資源分配, 隱含有競爭的意味在內。現實生活中, 人們可能會因爲嫉妒或者偶發的進取心, 而提高期許水準。由我們所舉的例子可看出, 如果允許期許水準能調整, 穩定狀態可能比 Dieckmann 的結論來得多。

另外, 在本例中, 雙方在混合均衡的報酬小於純粹均衡, 顯示即使不具 Pareto 效率性的狀態, 也可能是隨機穩定狀態。[8] 這有幾種可能的解釋, 首先如上一小節之說明, 在我們的設定下, 玩家對賽局結構的資訊可能有限, 不過因爲玩家只要求滿足期許水準。在這種情況下, 使用混合均衡策略是穩定的。

其次, 如果將社會看成是兩性戰爭, 因爲雙方對合作有不同的偏好, 如果雙方互不相讓, 將無法達成合作協議。但是完全不合作是最糟糕的結果, 如果雙方只玩對自己有利的均衡行動 (玩家 1 玩 a', 玩家 2 玩 a''), 反而是兩敗俱傷。在雙方確定合作方式之前, 採用混合行動也可能是威脅對手合作的方式之一。玩家的做法可能是偶而採用對對方有利的均衡行動, 但多數時間還是採用對自己有利的均衡行動, 而混合均衡策略正好滿足這樣的

[8] 在 Young (1998) 第 7.3 小節裡, Young 設計了一個特殊的報酬矩陣:

1/2	A	B	C	D
A	(3,3)	(4,1)	(1,4)	(−1,−1)
B	(1,4)	(3,3)	(4,1)	(−1,−1)
C	(4,1)	(1,4)	(3,3)	(−1,−1)
D	(−1,−1)	(−1,−1)	(−1,−1)	(0,0)

本矩陣有一個純粹均衡: 兩人都玩 D; 一個混合均衡: 兩人都以各 1/3 的機率玩 A、B、C, 同樣地, 如果純粹均衡是隨機穩定狀態, 則混合均衡也是隨機穩定狀態, 但後者的報酬較高, 這與兩性戰爭剛好相反。

臺灣大學學術期刊資料庫

概念 (玩家 1 玩 a' 的機率大於玩 a'', 玩家 2 則相反)。

在原來的純粹均衡裡, 玩家 2 是因爲想提高報酬才改變策略, 最後卻落入更差的混合均衡, 直覺上, 如果玩家 2 在調整之後變得更差, 他應該會回復到調整前的狀況, 也就是說應該回到原先的純粹均衡, 而不會落到較差的混合均衡。

上述說法如果要成立, 必須建立在雙方可以互動的前提下。如果雙方可以談判, 或許可以在玩家 2 改變策略, 而讓雙方變得更差時, 藉由談判而回到原先對雙方都有利的純粹均衡。在這情況下, 雙方可以選擇回到原先的『均衡』。

但如果雙方並不知道自己正在參與賽局, 所謂的『回復到調整前的狀況』就只是採用原先純粹均衡策略而已。玩家能選擇的是『策略』而不是『均衡』, 即使玩家 2 採用了原先的純粹均衡策略, 也會因爲玩家 1 剛好選到不合作的策略, 使得報酬爲 0。對玩家 2 而言, 他發現他因爲想增加報酬而改用其他策略, 結果報酬變差, 此時再採用原來的純粹均衡策略, 報酬仍然很差, 因此下一期玩家 2 沒有理由繼續採用原先純粹均衡策略。如果玩家的策略一直無法滿足期許水準, 那他就要調降期許水準。長期而言, 他的期許水準與策略會落入混合均衡, 因爲只有在此狀態下, 他的策略才可以滿足他的期許水準。

如果要將兩性戰爭中的混合均衡排除在隨機穩定狀態之外, 可能需要對玩家的理性、策略、資訊、偏好 · · · 等做進一步的限制。例如玩家可以用賄賂的方式促成合作; 或者玩家在乎公平, 願意輪流合作; 或者其中一位玩家比較沒有耐性, 所以願意接受對他不利的合作方式, 以避免無止盡的談判。

4 結論

本文研究在特定類型的兩人零和賽局裡 — 雙方互有輸贏、具有唯一且爲混合的 Nash 均衡, 當玩家皆爲追求滿足者 (satisfier) 時, 其混合均衡如何被演化出來。我們所設定的超賽局結構爲: 每期玩家選擇策略, 重複玩 N 次賽局, N 夠大, 重複玩無限多期。我們所提出的學習法爲: 玩家盡量嘗試各種方法來達到期許水準, 如果不能滿足則降低期許水準, 並有微小的機

臺灣大學學術期刊資料庫

率允許玩家嘗試實驗或犯錯。我們假設理性策略爲有限個, 且包含混合均衡策略。在上述設定之下, 我們應用 Young (1993, 1998) 的方法證明雙方期許爲 0, 且皆採用混合均衡策略必爲隨機穩定狀態, 但同時存在其他隨機穩定狀態。這個結論部分解決了 Crawford 難題。

美中不足的是, 定理 2 表示還有其他隨機穩定狀態。其次, 在 3.2 節, 我們以兩性戰爭賽局爲例, 說明純粹均衡未必比混合均衡穩定。針對這兩點, 我們與 Crawford 有類似的看法, 如果要讓零和賽局中的混合均衡成爲唯一的隨機穩定狀態, 或者排除兩性戰爭賽局裡的混合均衡, 經濟學家必須對玩家的理性能力與知識做進一步的限制。

本文並未處理策略與資訊之間的關係, 原則上, 玩家的理性策略集合應該隨資訊的不同而不同, 當玩家對賽局的知識從一無所知變成充分瞭解時, 他的理性策略集合也應該跟著變化。例如當玩家不知道所在賽局的結構資訊時,『一直採用 A 行動』或許是個有意義的策略, 但如果玩家之後發現到自己其實正在參與剪刀石頭布賽局, 這就不再是理性的策略了。

本文的架構隱含假設了玩家的理性策略集合始終不變, 未來可以討論理性策略的演化問題, 或者深入討論理性策略與資訊之間的關係, 或許透過這樣的方式, 可以排除掉本文中許多看似不合理的隨機穩定狀態。

我們所提出的學習法有兩個核心觀念: 一、玩家的期許水準必須能變動; 二、玩家在調整期許水準前會先嘗試各種不同策略。我們相信實際生活中人們應該有類似的學習行爲, 因此我們所提出的學習法, 也許可以更廣泛的應用。例如我們在 3.2 節以兩性戰爭賽局爲例, 說明純粹均衡未必比混合均衡穩定。我們推測若賽局有唯一 Nash 均衡, 且爲混合均衡, 應用 3.2 節的方法, 定理 2 應該同樣成立。例如在 Shapley (1964) 賽局裡, 玩家期許水準爲均衡時的報酬, 且採用混合均衡策略, 應爲該賽局的隨機穩定狀態之一。

除了考慮如何透過新的學習法收斂到特定的均衡, 或是 Crawford 文中考慮玩家如何能學會採取混合均衡之外, 研究學習法的合理性也是一個重要的議題。如果一個學習法的合理性被認可, 那麼將此學習法應用在不同的賽局裡所得出的結論, 都將可以做爲一個『均衡』的合理標準。

臺灣大學學術期刊資料庫

附錄 A

A.1: 命題1之證明

引理 1. Φ^0 過程中, 若 $\omega^{m_0} = \omega = (v_1, v_2, x_1, x_2, (X_1)^c, (X_2)^c) \in \Omega_E$, 則對所有 i 與所有 $m \geq m_0$: $\mu^m_{x_i}(x_j) \geq v_i$。

証明. 若 ω 爲穩定狀態, 且存在某一期 $m \geq m_0$, 使得某一玩家 i 的 $\mu^m_{x_i}(x_j) < v_i$, 根據 Φ^0, i 必須調整策略或期許水準, 則 ω 不再是一個穩定狀態, 引理一得證。 □

命題 1. 在 Φ^0 過程裡, 若 $\omega = (v_1, v_2, x_1, x_2, (X_1)^c, (X_2)^c) \in \Omega_E$, 則 $v_1 + v_2 \leq 0$。

証明.

1. 假設 $\omega = (v_1, v_2, x_1, x_2, (X_1)^c, (X_2)^c) \in \Omega_E$, 且 $v_1 + v_2 > \varepsilon > 0$。
2. 若 Φ^0 在第 m_0 期到達 ω, 根據引理 1, 對所有 i 與所有 $m > m_0$, $\mu^m_{x_i}(x_j) \geq v_i$。因此, 對所有 $m > m_0$, $\mu^m_{x_1}(x_2) + \mu^m_{x_2}(x_1) \geq v_1 + v_2 > \varepsilon > 0$。
3. 根據 $\mu^m_{x_i}(x_j)$ 定義可知

$$\mu^m_{x_1}(x_2) + \mu^m_{x_2}(x_1) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} u_1^{mn} + \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} u_2^{mn}$$

$$= \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \left(u_1^{mn} + u_2^{mn}\right)。$$

再根據零和賽局之定義, 對所有 n, $u_1^{mn} + u_2^{mn} = 0$。因此對所有 m, $\mu^m_{x_1}(x_2) + \mu^m_{x_2}(x_1) = 0$。所以必存在一個 $m^* > m_0$, 使 $\mu^{m^*}_{x_1}(x_2) + \mu^{m^*}_{x_2}(x_1) < \varepsilon$。

□

以上 1.2.3. 點彼此矛盾, 所以若 ω 爲穩定狀態, 則 $v_1 + v_2 \leq 0$, 命題 1 得證。

臺灣大學學術期刊資料庫

A.2: 命題2之證明

引理 2. 在任意期 m, 若 j 採用 x_j^*, 則 i 採用 x_i^* 可以極大化第 m 期的平均報酬, 意即給定 $x_j^m = x_j^*$, 則 $x_i^* \in \arg\max_{x_i \in X_i} \mu_{x_i}^m(x_j^*)$。

証明. 在任意期 m, 給定 $x_j^m = x_j^*$, 根據 x_j^* 之定義, 對所有 n, $\sigma_j^{mn} = \sigma_j^*$。另外, 極大化第 m 期的平均報酬, 意同極大化每次互動的報酬期望值, 意即:

$$\begin{aligned}\arg\max_{x_i \in X_i} \mu_{x_i}^m\left(x_j^*\right) &= \arg\max_{x_i \in X_i} \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} u_i^{mn} \\ &= \arg\max_{x_i \in X_i} U_i^{mn}\left(\sigma_i^{mn}\left(x_i\right), \sigma_j^*\right), \ \forall n。\end{aligned}$$

因爲 $\sigma_i^* = \arg\max_{\sigma \in \Sigma_i} U_i(\sigma_i, \sigma_j^*)$, 所以 $x_i^* \in \arg\max_{x_i \in X_i} \mu_{x_i}^m(x_j^*)$。引理2得證。 □

引理 3. 在任意期 m, 不管對手 j 採用什麼策略, i 採用 x_i^* 可保證第 m 期的平均報酬至少爲0, 意即對所有 $x_j \in X_j$, $\mu_{x_i^*}^m(x_j) \geq 0$。

証明. 根據引理2, 在任意期 m 與所有 $x_j \in X_j$, $\mu_{x_j^*}^m(x_i^*) \geq \mu_{x_j}^m(x_i^*)$, 其中 $\mu_{x_j^*}^m(x_i^*) = \lim_{N\to\infty}(1/N)\Sigma_{n=1}^N u_j^{mn} = U_j^{mn}(\sigma_i^*, \sigma_j^*) = 0$, 因此 $\mu_{x_j^*}^m(x_i^*) \leq 0$。由命題1的證明中得知: $\mu_{x_i^*}^m(x_j) + \mu_{x_j^*}^m(x_i^*) = 0$, 故 $\mu_{x_i^*}^m(x_j) \geq 0$, 引理3得證。 □

引理 4. 在任意期 m_0, 若 i 的期許水準小於等於0, 且採用 x_i^*, 則從此之後 i 將不再調整期許水準與策略, 意即若 $v_i^{m_0} \leq 0$ 且 $x_i^{m_0} = x_i^*$, 則對所有 $m \geq m_0$, $v_i^m = v_i^{m_0}$, $x_i^m = x_i^{m_0} = x_i^*$。

証明. 由引理3得知, 若 $v_i^{m_0} \leq 0$ 且 $x_i^{m_0} = x_i^*$, 則對任意 $x_j^{m_0} = x_j \in X_j$, $\mu_{x_i^*}^{m_0}(x_j) \geq 0 \geq v_i^{m_0}$, 根據 Φ^0 過程, 在第 $m_0 + 1$ 期, i 仍會採用相同策略與期許水準, 意即 $v_i^{m_0+1} = v_i^{m_0} \leq 0$ 且 $x_i^{m_0+1} = x_i^{m_0} = x_i^*$, 由數學歸納法得知對所有 $m \geq m_0$, $v_i^m = v_i^{m_0}$, $x_i^m = x_i^{m_0} = x_i^*$, 引理4得證。 □

命題 2. 在 Φ^0 過程裡, $\Omega_N \subseteq \Omega_E$。

臺灣大學學術期刊資料庫

証明. 對任意 $\omega = (0, 0, x_1^*, x_2^*, (X_1)^c, (X_2)^c) \in \Omega_N$, 假設 Φ^0 在第 m_0 期到達 ω, 根據引理4, 從此之後兩人將不再調整期許水準與策略, 故 ω 爲穩定狀態, 意即 $\omega \in \Omega_E$。因此 $\Omega_N \subseteq \Omega_E$, 命題2得證。 □

A.3: 命題3之例證

命題 3. $\Omega_E \setminus \Omega_N \neq \emptyset$。

例證: 根據 Γ_M 的定義, 我們可以找到兩組行動組合 $(a', a'') \in A$, 其中 $u_1(a') \geq 0$, $u_2(a') \leq 0$, $u_1(a'') \leq 0$, $u_2(a'') \geq 0$, 對 $i = \{1, 2\}$, 令策略 $\hat{x}_i$ 爲 i 連玩 g 次 a_i' 後連玩 h 次 a_i'', 不斷重複之, 給定兩人玩 $(\hat{x}_1, \hat{x}_2)$, 則在任意期 m, 必有一人 i 的平均報酬 $\mu_{\hat{x}_i}^m(\hat{x}_j) \geq 0$, 另一人 j 的平均報酬 $\mu_{\hat{x}_j}^m(\hat{x}_i) \leq 0$, 且 $\mu_{\hat{x}_1}^m(\hat{x}_2)$、$\mu_{\hat{x}_1}^m(\hat{x}_2)$ 爲常數, 故可令 $c_1 = \mu_{\hat{x}_1}^m(\hat{x}_2)$, $c_2 = \mu_{\hat{x}_2}^m(\hat{x}_1)$, $\hat{\omega} = (c_1, c_2, \hat{x}_1, \hat{x}_2, (X_1)^c, (X_2)^c)$, 其中 $(X_i)^c \subseteq X_i \setminus \{\hat{x}_i\}$, 則 $\hat{\omega}$ 爲穩定狀態, 說明如下:

假設兩人在第 m_0 期到達 $\hat{\omega}$, 則對所有 i, $\mu_{\hat{x}_i}^{m_0}(\hat{x}_j) = c_i = v_i^{m_0}$, 根據 Φ^0 過程, 在第 $m_0 + 1$ 期, i 仍會採用相同策略與期許水準, 意即 $v_i^{m_0+1} = v_i^{m_0} = c_i$ 且 $x_i^{m_0+1} = x_i^{m_0} = \hat{x}_i$, 由數學歸納法得知對所有 $m \geq m_0$, $v_i^m = v_i^{m_0}$, $x_i^m = x_i^{m_0} = x_i^*$, 表示從第 m_0 期後, 雙方不再調整期許水準與策略, 故 $\hat{\omega}$ 爲穩定狀態。

故 $\hat{\omega} \in \Omega_E$ 但 $\hat{\omega} \notin \Omega_N$, 因此 $\Omega_E \setminus \Omega_N \neq \emptyset$, 命題3得證。

值得注意的是, 隨著 g 與 h 的不同組合, c_1 與 c_2 也會變化, 不管兩者都爲0, 或者一正一負, 都可以是穩定狀態。

A.4: 定理1之證明

引理 5. 在任意期 m_0, 若 Φ^0 過程不會收斂到穩定狀態, 長期而言, 必然有人要調降期許水準, 意即存在一個 $m^* > m_0$, 使得 $v_1^{m^*} < v_1^{m_0}$ 或 $v_2^{m^*} < v_2^{m_0}$。

証明. 假設 Φ^0 過程不會收斂到穩定狀態, 則可用序列 m_k, $k = 1, 2, \cdots, \infty$ 表示有人調整期許水準或策略的時期, 意即對所有 m_k, $\omega^{m_k} \neq \omega^{m_k-1}$。假設不存在這樣的 m^*, 意即兩人永遠不調降期許水準, 則對所有 m_k, 必然

臺灣大學學術期刊資料庫

有一人要調整策略, 且根據 Φ^0, 只能從上次調整期許水準後, 尚未使用過的策略中選用新策略。故在期許水準不變的情況下, 兩人尚未使用過的策略將越來越少, 意即對於 i、j 而言, $(X_i^{m_0})^c \supset (X_i^{m_1})^c \supset \cdots \supset (X_i^{m_k})^c$ 且 $(X_j^{m_0})^c \supset (X_j^{m_1})^c \supset \cdots \supset (X_j^{m_k})^c$。因爲策略爲有限個, 所以一定存在某一期 m_α, 使得 $(X_i^{m_\alpha})^c = (X_j^{m_\alpha})^c = \varnothing$, 表示在沒有人調整期許水準的情況下, 兩人策略將在第 m_α 期用盡。則在 $m_{\alpha+1}$ 期, 必然有人要調降期許水準。令 $m^* = m_{\alpha+1}$, 則在第 m^* 期, $v_1^{m^*} < v_1^{m_0}$ 或 $v_2^{m^*} < v_2^{m_0}$, 引理5得證。 □

引理 6. 在任意期 m_0, 若 Φ^0 過程不會收斂到穩定狀態, 則長期而言一定有一人期許水準在0以下, 且採用 x_i^*, 意即存在一個 i 與 $m^* > m_0$, $v_i^{m^*} \leq 0$ 且 $x_i^{m^*} = x_i^*$。

証明. 假設 Φ^0 過程不會收斂到穩定狀態, 由引理5, 可令序列 m_k, $k = 1, 2, \cdots, \infty$ 表示有人調降期許水準的時期, 意即對所有 m_k, $v_1^{m_k} < v_1^{m_{k-1}}$ 或 $v_2^{m_k} < v_2^{m_{k-1}}$。因爲永遠有人要調降期許水準, 故一定存在某一期 m_α, 使得 $v_1^{m_\alpha} \leq 0$ 或 $v_2^{m_\alpha} \leq 0$, 假設是$v_1^{m_\alpha} \leq 0$。 □

若在第 m_α 期後, 玩家1在某一期 $m_{\alpha'}$ 調降期許水準, 根據 Φ^0, 玩家1在調降期許水準之前必先使用過所有策略, 故一定存在某一期 $m^* \in [m_\alpha, m_{\alpha'}]$, 玩家1採用策略 x_1^*。

若在第 m_α 期後, 玩家1永遠不再調降期許水準, 這表示從此以後玩家2一定要調降期許水準, 意即對所有 $m_k > m_\alpha$, $v_2^{m_k} < v_2^{m_{k-1}}$。所以一定存在某一期 $m_{\alpha''} > m_\alpha$, 使得 $v_2^{m_{\alpha''}} \leq 0$, 且在第 $m_{\alpha''+1}$ 期, 玩家2調降期許水準。故一定存在某一期 $m^* \in [m_{\alpha''}, m_{\alpha''+1}]$, 玩家2採用策略 x_2^*。

故無論如何, 一定存在一個 i 與 $m^* > m_0$, 使得 $v_i^{m^*} \leq 0$ 且 $x_i^{m^*} = x_i^*$, 引理6得證。

定理 1. 在超賽局 Γ^∞ 裡, 若玩家所能採用的策略集合 (X_1, X_2) 符合假設1跟假設2, 則給定任意初始狀態 (initial state) ω^0, Φ^0 過程保證會收斂到穩定狀態。

臺灣大學學術期刊資料庫

証明. 假設 Φ^0 不會收斂到穩定狀態, 則根據引理6, 一定存在一個 i 與 m_0, 使得 $v_i^{m_0} \leq 0$ 且 $x_i^{m_0} = x_i^*$。根據引理4, i 從此之後不再調整期許水準與策略。則根據引理5, 在第 m_0 期後, 可找到一組序列 $m_k, k = 1, 2, \cdots, \infty$ 表示 j 調降期許水準的時期, 意即對所有 m_k, $v_j^{m_k} < v_j^{m_{k-1}}$。因此一定存在某一期 m_α 使得 $v_j^{m_\alpha} \leq 0$, 且在第 $m_{\alpha+1}$ 期, j 調降期許水準。故一定存在某一期 $m^* \in [m_\alpha, m_{\alpha+1}]$, j 採用策略 x_j^*, 根據引理4, 從此以後 j 不再調整期許水準與策略。 □

因此 Φ^0 至少會在第 m^* 期收斂到穩定狀態, 定理1得證。

附錄 B

B.1: 命題4之證明

命題 4. 在 Φ^η 過程裡, $e^* \in \Omega_S$。

証明. 給定任何 e-tree, 因爲從任何一個穩定狀態變化到另一個穩定狀態都需要突變, 且因爲我們假設所有突變的發生機率最大值爲 η^b, 因此各邊阻力至少大於 b, 隨機位能至少大於 $b(|\Omega_E| - 1)$。 □

令 $v_i(e), x_i(e), X_i^c(e)$ 分別表示穩定狀態 e 的 $v_i, x_i, (X_i)^c$。我們將所有穩定狀態切割成兩個子集: $\Omega_{00} \equiv \{e \in \Omega_E | v_1(e) = 0, v_2(e) = 0\}$ 表示兩人期許水準皆爲0的穩定狀態之集合; $\Omega_- \equiv \{e \in \Omega_E | v_i(e) < 0 \text{ for some } i\}$ 表示至少有一人期許水準小於0的穩定狀態之集合。最後定義 $\Omega_{0^*} \equiv \{e \in \Omega_E | v_1(e) = 0, x_i(e) = x_i^*, X_i^c(e) = X_i \setminus \{x_i^*\} \text{ for some } i\}$ 表示有一人期許水準爲0, 且採用混合均衡策略, 在此期許水準之下, 其餘策略皆未使用過。

由命題2得知, $e^* \in \Omega_N \subseteq \Omega_E$。以下我們將建構一個以 e^* 爲終點的 e^*-tree, 其各邊的阻力皆爲 b, 因此 e^* 的隨機位能必爲最小值, 故 e^* 爲隨機穩定狀態。

臺灣大學學術期刊資料庫

1. **考慮任意** $e \in \Omega_{00}$

令 i 發生第一種突變: 提高期許水準爲 k, 並採用任意策略, 因此到達新的狀態 ω。因爲 $v_i(\omega) + v_j(\omega) = k > 0$, 由命題1得 ω 不是穩定狀態, 長期而言, 必然有人要調降期許水準, 若是 j 調降期許水準, 而到達新的穩定狀態 e', 則 $e' \in \Omega_-$; 若是 j 不變期許水準, 則 i 必須將期許水準降爲0, 且可同時令 i 採用策略 x_i^*, 而到達新的穩定狀態 $e'' \in \Omega_{0^*}$。

故對所有 $e \in \Omega_{00}$, 必存在一個 $e' \in \Omega_{0^*} \cup \Omega_-$, 使得 $r(e, e') = b$。

2. **考慮任意** $e \in \Omega_-$, **其中** $v_i(e) < 0$

首先考慮 $v_i(e) < -k$ 的情況, 令 i 發生第一種突變: 提高期許水準爲 $v_i(e) + k < 0$, 並採用 x_i^*, 因此到達新的狀態 ω。若 $v_j(e) \leq 0$, 顯然不管 j 採用什麼策略, 兩人平均報酬皆爲0, 因此 $\omega \in \Omega_-$; 若 $v_j(e) > 0$, 則 ω 不是穩定狀態, 長期而言, i 不需做任何調整, 而 j 必然要調整期許水準至0, 因此到達另一新的穩定狀態 $e' \in \Omega_-$。

因此對任意 $e \in \Omega_-$, 如果 $v_i(e) < -k$, 則必然存在一個 $e' \in \Omega_-$, 使得 $r(e, e') = b$, 且 $v_i(e') = v_i(e) + k$。在此建構下, 任意以 $e \in \Omega_-$ 且 $v_i(e) < 0$ 的穩定狀態爲起點的路徑, 必然會經過另一個穩定狀態 $e'' \in \Omega_-$, 其中 $v_i(e'') = -k$。

接著考慮 $v_i(e) = -k$ 的情況, 同樣令 i 發生第一種突變: 提高期許水準爲0, 並採用 x_i^*, 因此到達新的狀態 ω'。若 $v_j(e) \leq 0$, 顯然不管 j 採用什麼策略, 兩人平均報酬皆爲0, 因此 $\omega' \in \Omega_{0^*}$; 若 $v_j(e) > 0$, 則 ω' 不是穩定狀態, 長期而言, i 不需做任何調整, 而 j 必然要調整期許水準至0, 因此到達另一新的穩定狀態 $e' \in \Omega_{0^*}$。因此對任意 $e' \in \Omega_-$, 如果 $v_i(e) = -k$, 則必然存在一個 $e'' \in \Omega_{0^*}$, 使得 $r(e, e'') = b$。

故此建構下, 所有以 $e \in \Omega_-$ 爲起點的任意路徑, 必然會經過另一個穩定狀態 $e' \in \Omega_{0^*}$, 且路徑上各邊阻力皆爲 b。

3. **考慮任意** $e \in \Omega_{0^*}$, **其中** $v_i(e) = 0$, $x_i(e) = x_i^*$, $X_i^c(e) = X_i \setminus \{x_i^*\}$

若 $v_j(e) < 0$, 令 j 發生第一種突變: 提高期許水準爲 $v_i(e) + k$, 且採用任意策略, 因此到達新的狀態 ω, 顯然 $\omega \in \Omega_{0^*}$, 且 $v_j(\omega) = v_j(e) + k$。在

臺灣大學學術期刊資料庫

此建構下, 所有以 $e \in \Omega_{0^*}$ 爲起點的任意路徑, 若 $v_i(e) = 0$, $x_i(e) = x_i^*$, $X_i^c(e) = X_i \setminus \{x_i^*\}$, 則必然會經過另一個穩定狀態 $e' \in \Omega_{0^*}$, 其中 $v_i(e') = v_j(e') = 0$, $x_i(e') = x_i^*$, $X_i^c(e') = X_i \setminus \{x_i^*\}$, 且路徑上各邊阻力皆爲 b。此時再令 j 發生第一種突變: 提高期許水準爲 k, 因此到達另一新的狀態 ω'', 由命題1知道 ω'' 不是穩定狀態, 長期而言, i 不需做任何調整, 而 j 必然要調整期許水準至0, 且可同時令 j 採用策略 x_j^*, 而到達穩定狀態 e^*。

故此建構下, 所有以 $e \in \Omega_{0^*}$ 爲起點的任意路徑, 必然會以 e^* 爲終點, 且路徑上各邊阻力皆爲 b。

由 1.2.3. 得, 必存在一個 e^*-tree, 其各邊的阻力皆爲 b, 因此 e^* 的隨機位能 $\gamma_{e^*} = b(|\Omega_E| - 1)$, 故 $e^* \in \Omega_S$, 命題4得證。

B.2: 命題5之證明

命題 5. 在 Φ^η 過程裡, 對所有 $e \in \Omega_E$, 若 $e = (0, 0, x_i^*, x_j, X_i \setminus \{x_i^*\}, X_j \setminus \{x_j\})$, 則 $e \in \Omega_S$。

証明. 首先由 e^* 出發, 令 j 發生第一種突變, 提高期許水準爲 k, 且採用任意策略, 因此到達新的狀態 ω。ω 不是穩定狀態, 長期而言, i 不需做任何調整, 而 j 必然要調整期許水準至0, 且可同時令 j 採用策略 x_j, 而到達新的穩定狀態 e。故 $r(e^*, e) = b$。 □

考慮各邊的阻力皆爲 b 的 e^*-tree, 將由狀態 e 出發的邊砍掉, 並增加由狀態 e^* 出發的邊 $(e^* \to e)$, 即可建構出一個各邊的阻力皆爲 b 的 e-tree, 故 $e \in \Omega_S$, 命題5得證。

B.3: 命題6之證明

命題 6. 在 Φ^η 過程裡, 若 $e = (v_1, v_2, x_1, x_2, (X_1)^c, (X_2)^c) \in \Omega_S$, 則 $v_1 = v_2 = 0$。

証明. 假設 $e \in \Omega_-$ 且 $e \in \Omega_S$, 其中 $v_i(e) < 0$。對任意以 e 爲終點的 e-tree, 考慮以 e^* 爲起點的任意路徑, 因爲 $v_1(e^*) = v_2(e^*) = 0$, 故 e^* 通往 e 的路徑中, 必須發生第三種突變, i 必須降低他的期許水準, 故此路徑

臺灣大學學術期刊資料庫

上必然有一邊的阻力爲 d, 因此 e 的隨機位能必然大過 e^*, e 不是隨機穩定狀態。 □

故任意穩定狀態若是隨機穩定狀態, 兩人期許水準必爲 0, 命題 6 得證。

參考文獻

Börgers, Tilman and Rajiv Sarin (2000), "Naïve Reinforcement Learning with Endogenous Aspirations," *International Economic Review*, 41(4), 921–950.

Brown, George William (1951), "Iterative Solution of Games by Fictitious Play," in Tjalling Charles Koopmans (ed.), *Activity Analysis of Production and Allocation*, Cowles Commission Monography, New York: John Wiley and Sons, Inc., 374–376.

Chong, Juin-Kuan, Teck-Hua Ho, and Colin Camerer (2016), "A Generalized Cognitive Hierarchy Model of Games," *Games and Economic Behavior*, 99, 257–274.

Conlisk, John (1993a), "Adaptation in Games: Two Solutions to the Crawford Puzzle," *Journal of Economic Behavior and Organization*, 22(1), 25–50.

——— (1993b), "Adaptive Tactics in Games: Further Solutions to the Crawford Puzzle," *Journal of Economic Behavior and Organization*, 22(1), 51–68.

Crawford, Vincent P. (1974), "Learning the Optimal Strategy in a Zero-Sum Game," *Econometrica*, 42(5), 885–891.

——— (1985), "Learning Behavior and Mixed-Strategy Nash Equilibria," *Journal of Economic Behavior and Organization*, 6(1), 69–78.

Dieckmann, Tone (1998), "Stochastic Learning and the Evolution of Convention," *Constitutional Political Economy*, 9(3), 187–212.

Gilboa, Itzhak and David Schmeidler (1996), "Case-Based Optimization," *Games and Economic Behavior*, 15(1)), 1–26.

Juang, Wei-Torng (2002), "Rule Evolution and Equilibrium Selection," *Games and Economic Behavior*, 39(1), 71–90.

Karandikar, Rajeeva, Dilip Mookherjee, Debraj Ray, and Fernando Vega-Redondo (1998), "Evolving Aspirations and Cooperation," *Journal of Economic Theory*, 80(2), 292–331.

Lambson, Val E. and Daniel A. Probst (2004), "Learning by Matching Patterns," *Games and Economic Behavior*, 46(2), 398–409.

臺灣大學學術期刊資料庫

Napel, Stefan (2003), "Aspiration Adaptation in the Ultimatum Minigame," *Games and Economic Behavior*, 43(1), 86–106.

Oechssler, Jörg (2002), "Cooperation as Result of Learning with Aspiration Levels," *Journal of Economic Behavior and Organization*, 49(3), 405–409.

Pangallo, Marco, Torsten Heinrich, and J. Doyne Farme (2019), "Best Reply Structure and Equilibrium Convergence in Generic Games," *Science Advances*, 5(2), DOI: 10.1126/sciadv.aat1328.

Pazgal, Amit (1997), "Satisficing Leads to Cooperation in Mutual Interest Games," *International Journal of Game Theory*, 26(4), 439–453.

Robinson, Julia (1951), "An Iterative Method of Solving a Game," *Annals of Mathematics*, 54(2), 296–301.

Schade, Christian, Andreas Schroeder, and Kai Oliver Krause (2010), "Coordination after Gains and Losses: Is Prospect Theory's Value Function Predictive for Games?" *Journal of Mathematical Psychology*, 54(5), 426–445.

Shapley, Lloyd Stowell (1964), "Some Topics in Two Person Games," in Melvin Dresher amd Lloyd Stowell Shapley and Albert William Tucker (eds.), *Advances in Game Theory*, New Jersey: Princeton Univ. Press, 1–28.

Spiliopoulos, Leonidas (2012), "Pattern Recognition and Subjective Belief Learning in a Repeated Constant-Sum Game," *Games and Economic Behavior*, 75(2), 921–935.

Stahl, Dale O. (1988), "On the Instability of Mixed-strategy Nash Equilibria," *Journal of Economic Behavior and Organization*, 9(1), 59–69.

Young, H. Peyton (1993), "The Evolution of Conventions," *Econometrica*, 61(1), 57–84.

——— (1998), *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*, New Jersey: Princeton University Press.

Young, H. Peyton and Dean Foster (1991), "Cooperation in the Short and in the Long Run," *Games and Economic Behavior*, 3(1), 145–156.

投稿日期: 2020年9月2日, 接受日期: 2021年5月3日

臺灣大學學術期刊資料庫

Learning through Aspiration to Play the Mixed Equilibrium in Zero-Sum Games

Kuo-Chih Yuan

Department of Economics, Soochow University

Tzu-Hao Wang

Cloud Division, iKala

Wei-Torng Juang

Institute of Economics, Academia Sinica

This paper aims to provide a theoretical foundation for players learning to play mixed strategies and analyze the stochastic stability of the unique mixed Nash equilibrium in zero-sum games. We construct a supergame in which each player selects a strategy to play a zero-sum game for N rounds in each period. Each player then compares the average payoffs received in the N rounds with her aspiration. If the former exceeds the latter, the player is satisfied and sticks to the same strategy for the next period; otherwise she randomly selects a new strategy from the set of feasible strategies that have not yet been adopted. If all strategies have been tried and none of them can fulfill the player's current aspiration, then the player lowers her aspiration. Players also have a small probability of making mistakes when adjusting their strategies or aspiration. We apply the stochastic stability approach proposed by Young (1993), combined with the aspiration hypothesis and show that the unique mixed Nash equilibrium outcome is stochastically stable in a zero-sum game. We also use Battle of the Sexes to illustrate that mixed equilibrium could be as stable as pure ones.

Keywords: zero-sum game, mixed Nash equilibrium, aspiration, stochastic stability

JEL classification: C72, C73

臺灣大學學術期刊資料庫