Repeated Games with Perfect Monitoring

Mihai Manea

MIT

Repeated Games

- normal-form stage game $G = (N, A, u)$
- players simultaneously play game G at time $t = 0, 1, \ldots$
- at each date t, players observe all past actions: $h^t = (a^0, \ldots, a^{t-1})$
- common discount factor $\delta \in (0, 1)$
- payoffs in the repeated game $RG(\delta)$ for $h = (a^0, a^1, \ldots)$: $U_i(h) = (1 - \delta) \sum_{t=0}^{\infty} \delta^t u_i(a^t)$
- normalizing factor $1 - \delta$ ensures payoffs in $RG(\delta)$ and G are on same scale
- behavior strategy σ_i for $i \in N$ specifies $\sigma_i(h^t) \in \Delta(A_i)$ for every history h^t

Can check if σ constitutes an SPE using the single-deviation principle.

Minmax

Minmax payoff of player i: lowest payoff his opponents can hold him down to if he anticipates their actions,

$$\underline{v}_i = \min_{\alpha_{-i} \in \prod_{j \neq i} \Delta(A_j)} \left[\max_{a_i \in A_i} u_i(a_i, \alpha_{-i}) \right]$$

- m^i: *minmax profile for i*, an action profile (a_i, α_{-i}) that solves this minimization/maximization problem
- assumes *independent* mixing by i's opponents
- important to consider mixed, not just pure, actions for i's opponents: in the matching pennies game the minmax when only pure actions are allowed for the opponent is 1, while the actual minmax, involving mixed strategies, is 0

Equilibrium Payoff Bounds

In any SPE—in fact, any Nash equilibrium—i's obtains at least his minmax payoff: can myopically best-respond to opponents' actions (known in equilibrium) in each period separately. Not true if players condition actions on correlated private information!

A payoff vector $v \in \mathbb{R}^N$ is *individually rational* if $v_i \geq \underline{v}_i$ for each $i \in N$, and *strictly individually rational* if the inequality is strict for all i.

Feasible Payoffs

Set of *feasible payoffs*: convex hull of $\{u(a) \mid a \in A\}$. For a *common discount factor* δ, normalized payoffs in $RG(\delta)$ belong to the feasible set.

Set of feasible payoffs includes payoffs not obtainable in the stage game using mixed strategies. . . some payoffs require correlation among players' actions (e.g., battle of the sexes).

Public randomization device produces a publicly observed signal $\omega^t \in [0, 1]$, uniformly distributed and independent across periods. Players can condition their actions on the signal (formally, part of history).

Public randomization provides a convenient way to convexify the set of possible (equilibrium) payoff vectors: given strategies generating payoffs v and v', any convex combination can be realized by playing the strategy generating v conditional on some first-period realizations of the device and v' otherwise.

Nash Threat Folk Theorem

Theorem 1 (Friedman 1971)

If e is the payoff vector of some Nash equilibrium of G and v is a feasible payoff vector with $v_i > e_i$ for each i, then for all sufficiently high δ, $RG(\delta)$ has SPE with payoffs v.

Proof.

Specify that players play an action profile that yields payoffs v (using the public randomization device to correlate actions if necessary), and revert to the static Nash equilibrium permanently if anyone has ever deviated. When δ is high enough, the threat of reverting to Nash is severe enough to deter anyone from deviating. □

If there is a Nash equilibrium that gives everyone their minmax payoff (e.g., prisoner's dilemma), then every strictly individually rational and feasible payoff vector is obtainable in SPE.

General Folk Theorem

Minmax strategies often do not constitute static Nash equilibria. To construct SPEs in which i obtains a payoff close to $\underline{v}_i$, need to threaten to punish i for deviations with even lower continuation payoffs. Holding i's payoff down to $\underline{v}_i$ may require other players to suffer while implementing the punishment. Need to provide incentives for the punishers... impossible if punisher and deviator have indetical payoffs.

Theorem 2 (Fudenberg and Maskin 1986)

Suppose the set of feasible payoffs has full dimension $|N|$. Then for any feasible and strictly individually rational payoff vector v, there exists $\underline{\delta}$ such that whenever $\delta > \underline{\delta}$, there exists an SPE of $RG(\delta)$ with payoffs v.

Abreu, Dutta, and Smith (1994) relax the full-dimensionality condition: only need that no two players have the same payoff function (equivalent under affine transformation).

Proof Elements

- Assume first that i's minmax action profile m^i is pure.
- Consider an action profile a for which $u(a) = v$ (or a distribution over actions that achieves v using public randomization).
- By full-dimensionality, there exists v' in the feasible individually rational set with $\underline{v_i} < v'_i < v_i$ for each i.
- Let w^i be v' with ε added to each player's payoff except for i; for small ε, w^i is a feasible payoff.

Equilibrium Regimes

- Phase I: play a as long as there are no deviations. If i deviates, switch to II_i.
- Phase II_i: play m^i for T periods. If player j deviates, switch to II_j. If there are no deviations, play switches to III_i after T periods.
  - If several players deviate simultaneously, arbitrarily choose a j among them.
  - If m^i is a pure strategy profile, it is clear what it means for j to deviate. If it requires mixing. . . discuss at end of the proof.
  - T independent of δ (to be determined).
- Phase III_i: play the action profile leading to payoffs w^i forever. If j deviates, go to II_j.

SPE? Use the single-shot deviation principle: calculate player i's payoff from complying with prescribed strategies and check for profitable deviations at every stage of each phase.

Deviations from I and II

Player i's incentives

- Phase I: deviating yields at most $(1-\delta)M + \delta(1-\delta^T)\underline{v}_i + \delta^{T+1}v_i'$, where M is an upper bound on i's feasible payoffs, and complying yields v_i. For fixed T, if δ is sufficiently close to 1, complying produces a higher payoff than deviating, since $v_i' < v_i$.
- Phase II_i: suppose there are $T' \leq T$ remaining periods in this phase. Then complying gives i a payoff of $(1-\delta^{T'})\underline{v}_i + \delta^{T'}v_i'$, whereas deviating can't help in the current period since i is being minmaxed and leads to T more periods of punishment, for a total payoff of at most $(1-\delta^{T+1})\underline{v}_i + \delta^{T+1}v_i'$. Thus deviating is worse than complying.
- Phase II_j: with T' remaining periods, i gets $(1-\delta^{T'})u_i(m^j) + \delta^{T'}(v_i' + \varepsilon)$ from complying and at most $(1-\delta)M + (\delta - \delta^{T+1})\underline{v}_i + \delta^{T+1}v_i'$ from deviating. For high δ, complying is preferred.

Deviations from III

Player i's incentives

- Phase III_i: determines choice of T. By following the prescribed strategies, i receives v_i' in every period. A (one-shot) deviation leaves i with at most $(1-\delta)M + \delta(1-\delta^T)\underline{v}_i + \delta^{T+1}v_i'$. Rearranging, i compares between $(\delta + \delta^2 + \ldots + \delta^T)(v_i' - \underline{v}_i)$ and $M - v_i'$. For any $\underline{\delta} \in (0,1)$, $\exists T$ s.t. former term is grater than latter for $\delta > \underline{\delta}$.
- Phase III_j: Player i obtains $v_i' + \varepsilon$ forever if he complies with the prescribed strategies. A deviation by i triggers phase II_i, which yields at most $(1-\delta)M + \delta(1-\delta^T)\underline{v}_i + \delta^{T+1}v_i'$ for i. Again, for sufficiently large δ, complying is preferred.

Mixed Minmax

What if minmax strategies are mixed? Punishers may not be indifferent between the actions in the support. . . need to provide incentives for mixing in phase *II*.

Change phase *III* strategies so that during phase II_j player i is indifferent among all possible sequences of T realizations of his prescribed mixed action under m^j. Make the reward ε_i of phase III_j dependent on the history of phase II_j play.

Dispensing with Public Randomization

Sorin (1986) shows that for high δ we can obtain any convex combination of stage game payoffs as a normalized discounted value of a deterministic path $(u(a^t))$... "time averaging"

Fudenberg and Maskin (1991): can dispense of the public randomization device for high δ, while *preserving incentives*, by appropriate choice of which periods to play each pure action profile involved in any given convex combination. Idea is to stay within ε^2 of target payoffs at all stages.