Принцип одноразового отклонения в повторяемых играх

One-shot deviation by player i at history k^t

Одноразовым отклонением i *-го игрока от (чистой) стратегии* s_i *при истории* k^t будем называть любую (чистую) стратегию $r_i \in S_i$, такую что $r_i\left(k^t\right) \neq s_i\left(k^t\right)$, но для любой другой истории k^τ , $r_i\left(k^\tau\right) = s_i\left(k^\tau\right)$.

Более литературно:

Одноразовое отклонение от стратегии s_i - это стратегия r_i , отличающаяся от s_i только в одном узле (или в одной партии для повторяемых игр).

Принцип применим для:

- конечных игр в экстенсивной форме с совершенной информацией
- конечно повторяемых игр
- бесконечно повторяемых игр

и некоторых других

Принцип одноразового отклонения(one-shotdeviationprinciple, OSDP)

Профиль чистых стратегий $s = (s_i, s_{-i})$ является равновесием по Нэшу, совершенным в подыграх, тогда и только тогда, когда для любого игрока i , для любой истории k^t и любого одноразового отклонения r_i от стратегии s_i при истории k^t , $u_i\left(s_i, s_{-i}|k^t\right) \geq u_i\left(r_i, s_{-i}|k^t\right)$

Более литературно:

Профиль чистых стратегий $s = (s_i, s_{-i})$ является равновесием по Нэшу, совершенным в подыграх, тогда и только тогда, когда ни один игрок не может увеличить свой выигрыш ни в одной подыгре, отклонившись от s_i лишь единожды.

Комментарий:

Чтобы убедиться в том, что некий профиль стратегий является совершенным в подыграх достаточно следующей процедуры:

Взяли некий узел. Убедились в том, что ни одно отклонение в этом узле не приводит к увеличению выигрыша игрока, делающего ход в этом узле, в подыгре, начинающейся с этого узла. Аналогичным образом перебрали остальные узлы.

Доказательство для случая конечной игры в экстенсивной форме с совершенной информацией

Необходимость очевидна.

Докажем достаточность.

Пусть $s = (s_i, s_{-i})$ - профиль, в котором ни у одного из игроков не существует выгодного одноразового отклонения. Предположим, что профиль не является равновесием по Нэшу, совершенным в подыграх. Тогда найдется подыгра B и стратегия $r_i \neq s_i$, такая что в подыгре B i -ый игрок предпочитает стратегию r_i , $u_i\left(r_i, s_{-i}|B\right) > u_i\left(s_i, s_{-i}|B\right)$.

Пусть подыгра B начинается с истории h^m , т.е. $M = G\left(h^m\right)$.

Заметим, что r_i и s_i отличаются лишь в конечном числе узлов (в бесконечно повторяемой игре это могло бы быть не так). Идея доказательства состоит в том, что отличия в r_i можно убирать по одному, начиная с самого последнего. При этом ни в одной подыгре платеж от измененной r_i уменьшаться не будет, а в конце концов измененная r_i совпадет с s_i .

Для начала рассмотрим подыгру M , начинающуюся с узла, где было самое последнее отличие (в произвольной игре с несовершенной информацией это могло бы быть невозможно). В подыгре M между r_i и s_i есть всего одно отличие, а одиночные отклонения от s_i не могут увеличить платеж игрока в силу принципа одноразового отклонения. Следовательно, убрав в r_i последнее отличие, мы не уменьшим платеж в подыгре M . Обозначим измененную r_i буквой $\tilde{r}_i$.

Что произойдет при этом с платежами в других подыграх? Для любой другой подыгры M' , отличной от M , выполнено одно из трех соотношений: $M' \subset M$ (подыгра M' начинается

внутри подыгры M), $M \subset M'$ (подыгра M начинается внутри подыгры M'), $M' \cap M = \emptyset$ (подыгры M' и M не пересекаются).
Случай $M' \subset M$. В M' стратегия $\tilde{r}_i$ никак не отличается от исходной r_i , т.к. внесенное нами в r_i изменение находится до начала подыгры M' .
Случай $M' \cap M = \emptyset$. В M' стратегия $\tilde{r}_i$ никак не отличается от исходной r_i , т.к. внесенное нами в r_i изменение находится вне подыгры M' .
Случай $M \subset M'$. В этом случае есть две возможности. Начав подыгру M' и используя стратегию r_i можно либо попасть в подыгру M , либо не попасть. Если при использовании r_i игрок попадал в подыгру M , то при использовании $\tilde{r}_i$ игрок также попадет в M (изменение находится в самой M). При этом платеж в подыгре M' совпадает с платежом в подыгре M , который при переходе от r_i к $\tilde{r}_i$ не уменьшился. Если же при использовании r_i игрок не попадал в подыгру M , то при использовании $\tilde{r}_i$ игрок также не попадет в M , и, следовательно, платеж в подыгре M' не изменится.
Таким образом, мы получили $\tilde{r}_i$, которая не хуже r_i ни в одной подыгре, однако имеет на одно отличие от s_i меньше. Аналогичным способом по одному убираем отличия между r_i и s_i . Изменяемая $\tilde{r}_i$ постепенно превращается в s_i . Рано или поздно изменяемая $\tilde{r}_i$ полностью совпадет с s_i в подыгре B . Получается противоречие, так как с одной стороны r_i была строго лучше s_i в подыгре B , а с другой стороны $\tilde{r}_i$ является последовательно улучшенной r_i , а платеж от s_i совпадает с платежом от $\tilde{r}_i$.
Доказательство для случая бесконечно повторяемых игр
Доказательство полностью аналогично предыдущему, с единственным отличием. Пусть снова r_i более выгодна чем s_i в некой подыгре B . Проблема заключается в том, что r_i может отличаться от s_i в бесконечном количестве партий. Требуется так изменить отклонение r_i , чтобы оно отличалось от s_i лишь в конечном числе узлов.
Это всегда возможно в силу того, что бесконечно повторяемые игры обладают свойством *трансверсальности.*
Условие трансверсальности (условие непрерывности на бесконечности).
(transversalitycondition, continuityatinfinity)
Если два профиля стратегий не отличаются в течение первых t партий, то различие в платежах от этих профилей должно стремиться к нулю при $t \to \infty$.

$$\lim_{t\to\infty} \sup_{\sigma(h^\tau)=\tilde{\sigma}(h^\tau),\forall\tau<t} |u_i(\sigma) - u_i(\tilde{\sigma})| = 0$$

Бесконечно повторяемые игры удовлетворяют этому условию в силу того, что существует разница $d = |u_{\max} - u_{\min}|$ между максимальным и минимальным платежами в базовой игре, и $\lim\limits_{t\to\infty} \delta^t \frac{d}{1-\delta} = 0$ Заметим, что игры, оканчивающиеся за конечное число ходов, очевидно, удовлетворяют этому условию. Если t превысило максимальное количество ходов, то платежи от двух профилей в точности совпадают.
В силу условия трансверсальности все ходы в r_i начиная с некоего N можно заменить на ходы из s_i . При этом N можно выбрать так, что платеж от измененной r_i будет сколь угодно мало отличаться от платежа исходной r_i в подыгре B . Стратегия r_i в подыгре B была строго лучше s_i , а значит и измененная r_i будет строго предпочитаться s_i в подыгре B . Однако измененная r_i имеет лишь конечное число отличий от s_i .
В остальном доказательство аналогично случаю игры, оканчивающейся за конечное число ходов.
Примеры использования
Пример 1.

Рассмотрим бесконечно повторяющуюся игру с базовой игрой $\begin{array}{|c|cc|} \hline & c & d \\ \hline c & (2;2) & (0;3) \\ d & (3;0) & (1;1) \\ \hline \end{array}$.

При каких значениях дисконт-фактора пара стратегий (стратегия переключения; в первой партии сыграть «c», в последующих повторять действия противника в предыдущей) будет равновесием по Нэшу, совершенным в подыграх?

Напомним, что с**тратегия переключения** (grim trigger) предписывает играть ход c в первой партии и далее до тех пор, пока оба игрока играют ход c :

Решение

Рассмотрим подыгры, начинающиеся с истории $\{(c, c), (c, c), ..., (c, c)\}$

Пусть в начале такой подыгре первый игрок делает одиночное отклонение.

Тогда в подыгре события развиваются так: $..., \underbrace{(d, c)}_{deviation\ period}, (d, d), (d, d), ...$

После периода отклонения первый игрок играет ход d , т.к. это предписывается стратегией переключения, к которой он вернулся.

Платеж первого игрока при таком отклонении составит: $3 + 1 \cdot \frac{\delta}{1-\delta}$

Необходимо неравенство $3 + 1 \cdot \frac{\delta}{1-\delta} \leq 2 \cdot \frac{1}{1-\delta}$.

Пусть в начале подыгры второй игрок делает одиночное отклонение:

Тогда события развиваются так: $..., \underbrace{(c, d)}_{deviation\ period}, (d, c), (d, d), (d, d) ...$

После периода отклонения первый игрок играет ход d , т.к. это предписывается стратегией переключения, а второй игрок копирует действия первого в предыдущей партии (т.к. он вернулся к этой стратегии).

Платеж второго игрока при таком отклонении составит: $3 + 0 \cdot \delta + 1 \cdot \frac{\delta^2}{1-\delta}$

Необходимо неравенство $3 + 0 \cdot \delta + 1 \cdot \frac{\delta^2}{1-\delta} \leq 2 \cdot \frac{1}{1-\delta}$

Рассмотрим подыгры, начинающиеся с истории отличной от $\{(c, c), (c, c), ..., (c, c)\}$

Есть четыре варианта окончания таких подыгр (нам интересно только то, чем оканчивается подыгра в силу стратегии второго игрока). Они могут оканчиваться на (c, c) , (c, d) , (d, c) или (d, d) .

Рассмотрим истории типа $\{(?, ?), (?, ?), ..., (?, ?), (c, d)\}$

Если игроки не отклоняются от своих стратегий в подыгре, следующей за такой историей, то события развиваются так: $(d, c), (d, d), (d, d), (d, d), ...$. Без вычислений видно, что второму игроку выгодно использовать одиночное отклонение в первой партии подыгры, чтобы события развивались $(d, d), (d, d), (d, d), (d, d), ...$

Рассмотрение остальных трех вариантов не имеет смысла, т.к. уже ясно, что предложенный профиль не является равновесием по Нэшу, совершенным в подыграх, ни при каких значениях дисконт-фактора.

Пример 2.

Рассмотрим бесконечно повторяющуюся игру с базовой игрой $\begin{array}{|c|cc|} \hline & c & d \\ \hline c & (2;2) & (0;3) \\ d & (3;0) & (1;1) \\ \hline \end{array}$.

При каких значениях дисконт-фактора пара стратегий двухходового возмездия будет равновесием по Нэшу, совершенным в подыграх?

Стратегия двухходового возмездия: в начале сыграть ход «c» и играть «c» до тех пор, пока играется (c, c) ; если было сыграно не (c, c) , то в течение двух последующих партий играть ход «d», затем действовать, как будто игра начиналась заново.

Решение

Игрок может использовать одноразовое отклонение либо находясь в фазе возмездия, либо находясь в стадии кооперации.
Рассмотрим одноразовое отклонение, совершаемое в фазе кооперации:
Пусть отклоняется первый игрок. События развиваются так:

$$...,\underbrace{(d,c)}_{deviation\ period},(d,d),(d,d),(c,c),(c,c),...$$

Во второй и третьей партиях от момента отклонения первый игрок уже вернулся к стратегии двухходового возмездия и наказывает сам себя согласно стратегии.
Платеж первого игрока будет равен $3+1\cdot\delta+1\cdot\delta^2+2\cdot\frac{\delta^3}{1-\delta}$.
Необходимо неравенство $3+1\cdot\delta+1\cdot\delta^2+2\cdot\frac{\delta^3}{1-\delta}\leq 2\cdot\frac{1}{1-\delta}$
Аналогичное неравенство возникнет, если одноразовое отклонение будет использовать второй игрок.
Рассмотрим одноразовое отклонение, совершаемое в фазе возмездия.
События развиваются либо: $...,\underbrace{(c,d)}_{deviation\ period},(d,d),(c,c),(c,c),...$ (если это начало фазы возмездия).
Либо: $...,\underbrace{(c,d)}_{deviation\ period},(c,c),(c,c),...$ (если отклонение происходит в конце фазы возмездия).
В любом случае, очевидно, что подобное отклонение не выгодно.
Решаем неравенство $3+1\cdot\delta+1\cdot\delta^2+2\cdot\frac{\delta^3}{1-\delta}\leq 2\cdot\frac{1}{1-\delta}$. Получаем $\delta^3-2\delta+1\leq 0$ и в результате $\delta\in\left[\frac{\sqrt{5}-1}{2};1\right)$.
Приложение. Названия некоторых стратегий в повторяющейся дилемме заключенного
Обозначения:
a^t - исход базовой игры с номером t ;
a_i^t - ход сделанный i -ым игроком в базовой игре с номером t .
s_i - стратегия i -го игрока;
h^t - предыстория игры к моменту времени t : $h^t=\{a^1,a^2,...a^{t-1}\}$;
$s_i(h^t)$ - ход, предписываемый стратегией s_i после истории h^t (в момент t);
$G(h^t)$ - подыгра, начинающаяся с истории h^t
Стратегия "Всегда кооперироваться" (always cooperate)
Предписывает всегда играть ход c : $s_i(h^t)=c,\quad\forall t$ **Наивная стратегия переключения** (na?ve grim trigger)
Предписывает играть ход c в первой партии и далее до тех пор, пока противник играет ход c : $s_i(h^t)=\begin{cases}c, & t=1\\ c, & t>1,\quad\forall\tau<t\Rightarrow a_j^\tau=c\\ d, & otherwise\end{cases}$
Стратегия переключения (grim trigger)
Предписывает играть ход c в первой партии и далее до тех пор, пока оба игрока играют ход c : $s_i(h^t)=\begin{cases}c, & t=1\\ c, & t>1,\quad\forall\tau<t\Rightarrow a^\tau=(c;c)\\ d, & otherwise\end{cases}$
Стратегия Зуб за зуб (Tit for Tat)
Предписывает играть ход c в первой партии и далее повторять ход противника в предыдущей партии: $s_i(h^t)=\begin{cases}c, & t=1\\ a_j^{t-1}, & t>1\end{cases}$
Стратегия Кнута и Пряника (Win-Stay, Lose-Shift; Pavlov strategy)

Предписывает играть ход c в первой партии и далее играть ход c , если в предыдущей партии действия игроков совпали: $s_i\left(h^t\right)=\begin{cases} c, & t=1 \\ c, & t>1, \quad a^{t-1}\in\{(c;c),(d;d)\} \\ d, & otherwise \end{cases}$

Тигр: Эта хитрая стратегия была внедренаизвестными специалистами по теории игр Кнутом Б.Б. и Пряником В.Л.

Стратегия ограниченного возмездия (limited retaliation)

Предписывает играть ход c , пока все игроки кооперируются. Если произошло нарушение, то в течение k ходов играть d , затем вернуться в исходное состояние. Состоит из трех фаз:

Фаза 1: сыграть ход c и переключиться в фазу 2;

Фаза 2: играть ход c до тех пор, пока все игроки играют ход c , в противном случае переключиться в фазу 3, положив $\tau := 0$;

Фаза 3: пока $\tau \leq k$, положить $\tau := \tau + 1$ и играть ход d , иначе переключиться в фазу 1.