Que valent les stratégies probabilistes au DIP

P. Mathieu et J.P. Delahaye
prenom.nom@univ-lille1.fr

Université Lille1, CRISTAL

JFSMA 2017 , Caen – 5 juillet 2017

Plan

Le Dilemme du Prisonnier

Matrice de gains

Jeu simultané, symétrique, à deux joueurs, à somme non nulle, non coopératif où chaque joueur doit choisir une carte parmi deux : la Coopération (C) et la Trahison (D).

		Player II	
		Cooperate	Defect
Player I	Cooperate	R=3 / R=3	S=0 / T=5
	Defect	T=5 / S=0	P=1 / P=1

$$T > R > P > S \text{ et } T + S < 2R$$

Avec Trahir on gagne toujours plus qu'avec Cooperer !
Il y a dilemme car l'intérêt individuel (Trahir) rapporte moins que l'intérêt collectif (Coopérer collectivement).

Plan

- Version itérée :
  - les joueurs se recontrent un certain nombre de fois
  - les joueurs ne connaissent pas le terme du jeu
- le score des joueurs est la somme des scores de chaque rencontre
- A chaque étape un joueur sait ce que son adversaire a joué dans les coups précédents
- Il est alors possible de définir un comportement prédéfini pour le jeu : **une stratégie**
- Quelques exemples de stratégies :
  - all_c
  - all_d
  - per_ccd
  - soft_majo
  - tit_for_tat
  - spiteful

Le Dilemme Itéré du Prisonnier

12 stratégies identifiables, reconnues, classiques, compréhensibles ...

- all_c (gentille : c*)
- all_d (méchante : d*)
- random (lunatique : c ou d au hasard)
- tft (donnant_donnant : coopère puis joue ce que l'autre a joué précédemment)
- spiteful (rancunier : coopère tant que l'autre coopère, sinon trahis)
- per_ddc (ddc*)
- per_ccd (ccd*)
- soft_majo (majorité_mou : joue ce que l'autre a majoritairement joué. En cas d'égalité et à la première partie il coopère)
- mistrust (méfiante : trahis puis joue ce que l'autre a joué précédemment)
- hard_majo (majorité_dur : joue ce que l'autre a majoritairement joué. En cas d'égalité et à la premiere partie il trahis)
- prober (sondeur : dcc, puis s'il a coopéré les 2 derniers coups, je trahis toujours, sinon tft)
- hard_tft (dd_dur : coopere sauf si l'autre a trahi lors de l'un des 2 coups précédents)

Le Dilemme Itéré du Prisonnier

Exemple de jeu itéré

	1	2	3	4	5	6	7	8	9	10	
	3	3	0	3	3	0	3	3	0	3	= 21
all_c	C	C	C	C	C	C	C	C	C	C	
per_ccd	C	C	D	C	C	D	C	C	D	C	
	3	3	5	3	3	5	3	3	5	3	= 36

	1	2	3	4	5	6	7	8	9	10	
	0	1	1	1	1	1	1	1	1	1	= 9
tft	C	D	D	D	D	D	D	D	D	D	
all_d	D	D	D	D	D	D	D	D	D	D	
	5	1	1	1	1	1	1	1	1	1	= 14

Université de Lille

Il existe deux types d'évaluation d'un ensemble de stratégies :

1. **Le tournoi** (round-robin)
   - rencontre 2 à 2 de tous les éléments d'un panel chaque stratégie est classée en fonction de son score total $\Sigma(V_{i,j})$
2. **Les évolutions écologiques** (Maynard-Smith,1982)
   - chaque stratégie est représentée par une sous-population d'agents
   - chaque agent rencontre tous les autres
   - chaque stratégie obtient un score (somme des scores totaux des agents de sa sous-population)
   - les stratégies sont redistribuées proportionnellement à leur score sur la population d'agents

	all_c	all_d	per_ccd	tft	
all_c	30	0	21	30	81
all_d	50	10	38	14	112
per_ccd	36	3	24	27	90
tft	30	9	27	30	96

Classement :

1. all_d
2. tft
3. per_ccd
4. all_c

Le Dilemme Itéré du Prisonnier

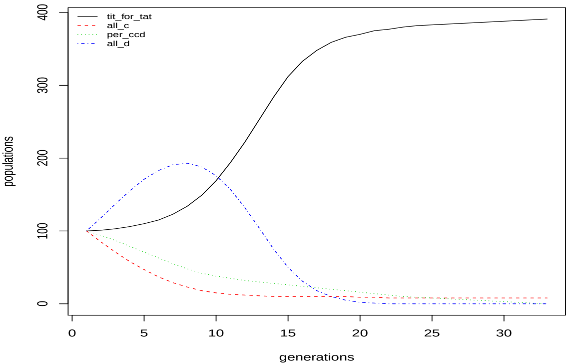
... et évolutions

	1	2	3	4	...
all_c	100	85	71	58	...
all_d	100	118	137	155	...
per_ccd	100	94	87	79	...
tft	100	101	103	106	...

Le Dilemme Itéré du Prisonnier

... et évolutions

	1	2	3	4	...
all_c	100	85	71	58	...
all_d	100	118	137	155	...
per_ccd	100	94	87	79	...
tft	100	101	103	106	...



Le Dilemme Itéré du Prisonnier

Quelle est la meilleure stratégie ?

- **Celle qui bat toutes les autres ?** `all_d` !
- **Celle qui fait le meilleur score possible face à toutes les autres ?** aucune, car on ne peut pas obtenir le score maximal à la fois contre `all_d` et `spiteful`
- Ne pas confondre "battre tout le monde" et "faire de bons scores"

Remarques : `all_d` gagne contre tout le monde. `tft` ne gagne jamais contre personne. `tft` ne perd jamais de plus de 5 points !

- Il n'y a pas de bonne stratégie dans l'absolu !
- On est bon dans une soupe donnée
- Il faut donc trouver des soupes objectives !

Etat final de coopération généralisée (hyper fréquent), alors qu'il n'y a pas d'autorité de contrôle et que la tentation de la trahison est présente pour tous à chaque coup.

Et c'est remarquable !

Le Dilemme Itéré du Prisonnier

gradual : une TRES bonne stratégie

Une nouvelle stratégie [ALIFE'5, 1996] : **gradual**

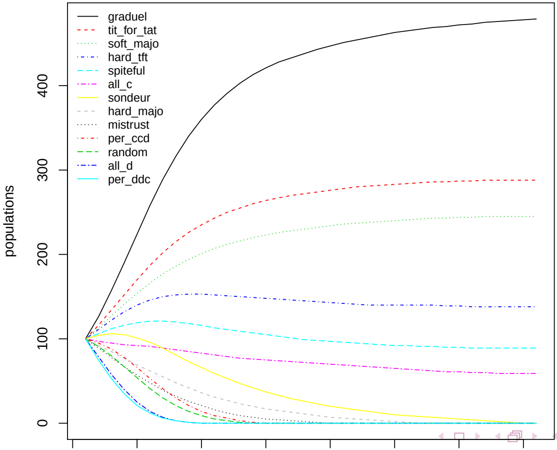
- Je commence par coopérer
- Tant que l'autre coopère, je coopère
- Quand il trahit, je trahis autant de fois qu'il m'a trahi par le passé
- Après ma réaction, je me calme 2 coups

	1	2	3	4	5	6	7	8	9	10	
	5	1	3	5	1	0	1	5	3	5	= 29
per_ddc	D	D	C	D	D	C	D	D	C	D	
gradual	C	D	C	C	D	D	D	C	C	C	
	0	1	3	0	1	5	1	0	3	0	= 14

Le Dilemme Itéré du Prisonnier

gradual : une TRES bonne stratégie

Les 12 de base + Gradual



Plan

Que valent les stratégies probabilistes

Le point de départ

Université de Lille


Proceedings of the National Academy of Sciences of the United States of America PNAS

CURRENT ISSUE // ARCHIVE // NEWS & MULTIMEDIA // AUTHORS // ABOUT COLLECTED ARTICLES // BROWSE BY TOPIC // EARLY EDITION // FRONT MATTER

> Current Issue > vol. 109 no. 26 > William H. Press, 10409–10413

Check for updates

Iterated Prisoner's Dilemma contains strategies that dominate any evolutionary opponent

William H. Press[a,1] and Freeman J. Dyson[b]

[a]Department of Computer Science and School of Biological Sciences, University of Texas at Austin, Austin, TX 78712; and

[b]School of Natural Sciences, Institute for Advanced Study, Princeton, NJ 08540

Contributed by William H. Press, April 19, 2012 (sent for review March 14, 2012)

Abstract

The two-player Iterated Prisoner's Dilemma game is a model for both sentient and evolutionary behaviors, especially including the emergence of cooperation. It is generally assumed that there exists no simple ultimatum strategy whereby one player can enforce a unilateral claim to an unfair share of rewards. Here, we show that such strategies unexpectedly do exist. In particular, a player X who is witting of these strategies can (*i*) deterministically set her opponent Y's score, independently of his strategy or response, or (*ii*) enforce an extortionate linear relation between her and his scores. Against such a player, an evolutionary player's best response is to accede to the extortion. Only a player with a theory of mind about his opponent can do better, in which case Iterated Prisoner's Dilemma is an Ultimatum Game.

This Issue

June 26, 2012
vol. 109 no. 26
Masthead (PDF)
Table of Contents

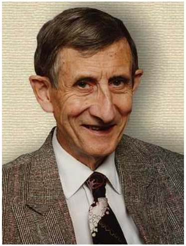
PREV ARTICLE NEXT ARTICLE

Don't Miss

PNAS Full-Text iOS App
Download the app for free from iTunes today!

Article Tools

Article Alerts
Alert me when this article is cited
Alert me if a correction is posted
Email this article to a colleague



Press & Dyson, PNAS 2012

Certaines stratégies probabilistes à mémoire de 1 coup sont dominantes dans l'absolu pour le dilemme itéré des prisonniers ! !

- Press & Dyson ont identifié des propriétés remarquables ... sur des confrontations 1 à 1
- Est-ce que ces résultats mathématiques ont vraiment une importance dans des confrontations généralisées ?
- Est-ce que leur resultat a un impact expérimental ?
- Parmi les stratégies probabilistes à 1 coup quelles sont celles qui se comportent bien quand on les met toutes ensemble ?

proba(p1,p2,p3,p4)

p1 : probabilité de jouer c lorsque le dernier coup a été `[c,c]`
p2 : probabilité de jouer c lorsque le dernier coup a été `[c,d]`
p3 : probabilité de jouer c lorsque le dernier coup a été `[d,c]`
p4 : probabilité de jouer c lorsque le dernier coup a été `[d,d]`

```
tit_for_tat = proba(1,0,1,0)
spiteful = proba(1,0,0,0)
pavlov = proba(1,0,0,1)
random = proba(1/2,1/2,1/2,1/2)
```

La stratégie `gradual` n'est pas une proba, elle a besoin de l'ensemble du passé, les `majoritaires` n'en sont pas non plus.

Que valent les stratégies probabilistes

La famille ZD(a,b,c)

Pour `a,b,c` donnés existe t-il une stratégie **`ZD(a,b,c)`** telle que $\forall$ `proba(p1,p2,p3,p4)` les gains entre ZD et proba sont liés linéairement par $\boxed{\texttt{aG1+bG2+c=0}}$

```
p1 = 1+3a+3b+c          p3 = 5a+c
p2 = 1+5b+c             p4 = a+b+c
```

$$\text{tit\_for\_tat} = \text{proba}(1,0,1,0) = \text{zd}(\tfrac{1}{5},\tfrac{-1}{5},0)$$

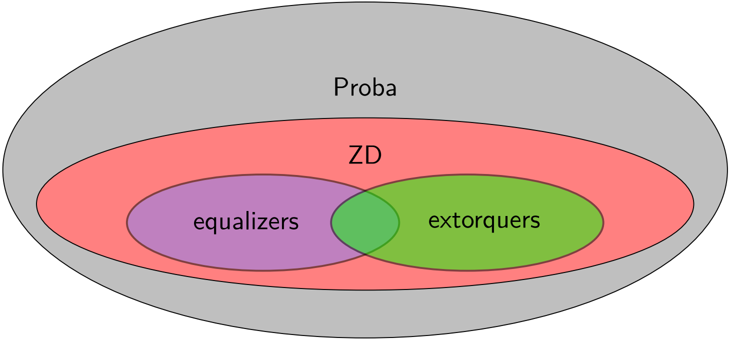
Le score de la proba peut donc être contrôlé par la ZD !

Que valent les stratégies probabilistes

deux sous-classes interessantes

Parmi les ZD, deux sous classes interessantes

- **Equalizer** : force tous ses adversaires à avoir le même résultat connu à l'avance
- **Extorquer** : exige de gagner x fois plus que l'autre par rapport à P



Que valent les stratégies probabilistes

Equalizers

$\boxed{\text{Si a=0 et b}\neq\text{0}}$ alors `G2=-c/b`. La `proba` a alors un gain moyen indépendant de ses propres params ! son gain ne dépend que de la `ZD` en face. Une telle ZD est appelée **Equalizer(b,c)**

Exemple : $\mathtt{ZD}(0, \frac{-1}{3}, \frac{2}{3})$ équivalent à $\mathtt{proba}(\frac{2}{3}, 0, \frac{2}{3}, \frac{1}{3})$, donc $G2 = \frac{-c}{b} = 2$,

```
equa =2,5 vs tit_for_tat=2
equa =3 vs gradual=2
equa =3 vs all_c=2
equa =2 vs per-ccd =2
```

Un equalizer force donc le résultat moyen de tout adversaire ...

Que valent les stratégies probabilistes

Equalizers

$\boxed{\text{Si a=0 et b}\neq\text{0}}$ alors `G2=-c/b`. La `proba` a alors un gain moyen indépendant de ses propres params ! son gain ne dépend que de la `ZD` en face. Une telle ZD est appelée **Equalizer(b,c)**

Exemple : `ZD(0,`$\frac{-1}{3}$`, `$\frac{2}{3}$`)` équivalent à `proba(`$\frac{2}{3}$`, 0, `$\frac{2}{3}$`, `$\frac{1}{3}$`)`, donc $G2 = \frac{-c}{b} = 2$,

```
equa =2,5 vs tit_for_tat=2
equa =3 vs gradual=2
equa =3 vs all_c=2
equa =2 vs per-ccd =2
```

Un equalizer force donc le résultat moyen de tout adversaire ...

mais parfois à ses dépens !

```
equa =1 vs all_d =2
equa =1 vs spiteful=2
equa =2 vs equa =2
```

Que valent les stratégies probabilistes

Extorquers

$\boxed{\text{Si } c = -(a+b).P}$, alors $G1 - P = X(G2 - P)$ avec $X = \frac{-b}{a}$

La stratégie **`Extorquer(a,b)`** force l'adversaire à lui donner plus que ce qu'il gagne par rapport à P

$$\text{tft} = \text{proba}(1{,}0{,}1{,}0) = \text{zd}(\tfrac{1}{5}, \tfrac{-1}{5}, 0) = \text{extorq}(\tfrac{1}{5}\ ,\ \tfrac{-1}{5}) \Rightarrow \text{X}{=}1$$

P est le repère (dans notre cas c'est 1). Si le coeff X est égal à 2, à chaque fois que celui qui est opposé voudra faire 1,5 , alors l'exorquer fera 2

Confronté à un extorqueur : ou bien je le laisse gagner plus que moi par rapport à P ou bien on gagnera tous les deux P

Que valent les stratégies probabilistes

Un exemple concret d'extorquer

`Extorq`$(\frac{1}{10},\ \frac{-1}{5})$ equivalent à `proba(4/5, 1/10, 3/5, 0)` est un `2-Extorqueur` : elle double l'écart à P de son adversaire.

Contre `Pavlov` qui gagne en moyenne $1,62$ ($0,62$ points au dessus de 1) cet extorqueur obtient en moyenne $2,24$ ($1,24$ points au dessus de 1).

`Extorq(`$\frac{1}{10}$`,` $\frac{-1}{5}$`)` equivalent à `proba(4/5, 1/10, 3/5, 0)` est un `2-Extorqueur` : elle double l'écart à P de son adversaire.

Contre `Pavlov` qui gagne en moyenne $1,62$ ($0,62$ points au dessus de 1) cet extorqueur obtient en moyenne $2,24$ ($1,24$ points au dessus de 1).

Un extorquer est une sorte de `all_d` généralisé

Il ne peut pas perdre (comme `all_d`)
Mais il joue mal contre lui même (comme `all_d`) ! !
Personne ne peut les battre, mais elles prennent le risque de gagner peu

Que valent les stratégies probabilistes

Engendrer de manière concrète ces familles

Equalizers(b,c)

```
p1 = 3b+c+1
p2 = 5b+c+1
p3 = c
p4 = b+c
```

Extorquers(a,b)

```
p1 = 2a+2b +1
p2 = 4b - a +1
p3 = 4a - b
p4 = 0
```

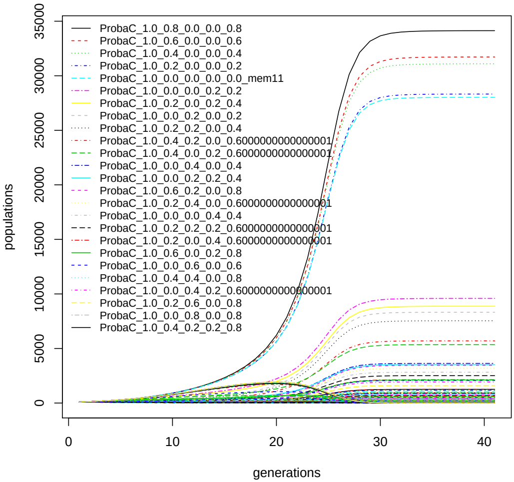
On peut à partir de ces formules et des paramètres a,b,c générer des ensembles massifs

mem11 $= 2^1 * 2^4 = 32$ stats , mem12 $= 2^2 * 2^8 = 1024$ strats , mem22 $= 2^2 * 2^{16} = 262144$ strats

probaCD_K1 $= 2.2^4 = 32$ strats , probaCD_k5 $= 2.6^4 = 2592$ strats, probaCD_K10 $= 2.11^4 = 29282$ strats

Que valent les stratégies probabilistes

Evolution écologique des probaCD_K5



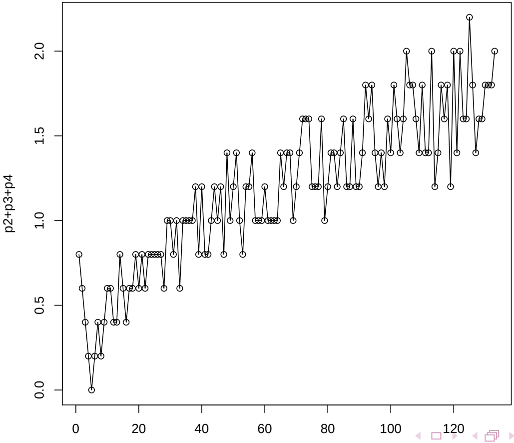
Aucune ZD dans les premières !

la première ZD est 29è ! c'est un extorq equivalent à tft

Que valent les stratégies probabilistes

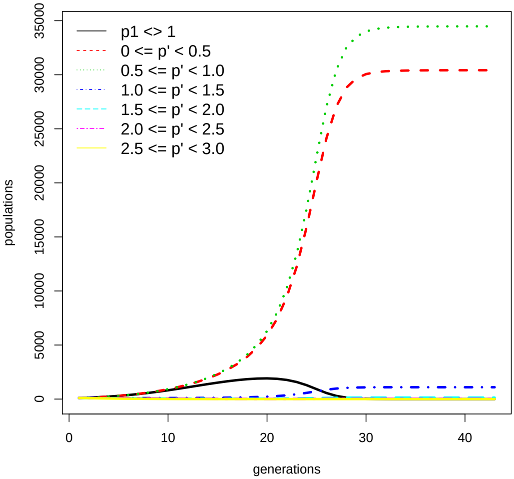
Corrélation entre le rang et les paramètres

$p' = p2 + p3 + p4$ $\qquad$ $Cor(rang, p') = 0,8805939$



Que valent les stratégies probabilistes

Affinement du résultat



Affinement par algorithme d'analyse canonique des corrélations (CCA)

$p'' = p2 + p3/2 + p4$ $\qquad$ $Cor(rang, p'') = 0,9411231$

$p''' = 0,266p2 + 0,138p3 + 0,277p4$ $\qquad$ $Cor(rang, p''') = 0,9413768$

Trois nouvelles stratégies introduites dans [AAMAS,2015]

- **winner12** : le gagnant des `mem(1,2)`
- **spiteful_cc** : joue comme `spiteful` sauf qu'elle commence en jouant CC (sorte de `spiteful` adouci)
- **tft_spiteful** : je joue `CC`. Dès que l'autre a trahi 2 fois consecutives je trahis toujours, sinon `tft`

Que valent les stratégies probabilistes

Robustesse : des probabilistes et des déterministes !

les 20 premières de ProbaCD_K5 + les 1024 mem(1,2) + les 21 stratégies de AAMAS

1	probaC_1.0_0.2_0.0_0.2	5255
2	tft_spiteful	4945
3	probaC_1.0_0.4_0.0_0.2	4877
4	probaC_1.0_0.2_0.0_0.4	4415
5	winner12	4331
6	mem12_CCCDCDDCDD	4331
7	probaC_1.0_0.6_0.0_0.2	4081
8	mem12_CCCDCDDDDD	3557
9	spiteful_cc	3557
10	probaC_1.0_0.8_0.0_0.0	3551
11	mem12_CCCCCDDDDD	2766
12	probaC_1.0_0.0_0.0_0.2	2614
13	probaC_1.0_0.6_0.0_0.0	2497
14	gradual	2469
15	mem12_CCCDDDDCDD	2368

Que valent les stratégies probabilistes

Conclusion

On le sait depuis longtemps !
les résultats mathématiques ne donnent pas grand chose de concret !

- Tout comme les équilibres classiques, n'impliquent pas de bonnes stratégies ...
- ... les récents résultats de Press&Dyson ne permettent d'hexiber aucune stratégie interessante en confrontations généralisées.
- Néanmoins nous avons identifié un paramètre qui permet d'identifier les bonnes stratégies proba : $\boxed{p1 = 1 \text{ et } p' = p2 + p3/2 + p4 \text{ faible}}$

Questions ?

Et si vous avez stratégies à tester, je suis à votre disposition ...