

ensemble de stratégies jouant au dilemme itéré du prisonnier. Nous n'envisagerons que des méthodes équitables, c'est-à-dire ne favorisant aucune équipe. Les systèmes que nous allons comparer seront stables: même si le hasard intervient, il ne modifie les classements que très rarement. La méthode de classement de la Coupe de France de football n'a pas cette propriété de stabilité, car il est assez probable qu'un second tournoi organisé la même année avec des tirages différents des rencontres conduirait à un autre classement final.

LE DILEMME DU PRISONNIER

Rappelons les règles du jeu. Deux individus jouent un dilemme du prisonnier si chacun doit choisir entre coopérer (c) et trahir (t). Ils sont récompensés par R points si chacun joue c, par P points si chacun joue t, et par, respectivement, T et S points si l'un joue t et l'autre c. On note: $[c, c] \rightarrow R+R$, $[t, t] \rightarrow P+P$, $[t, c] \rightarrow T+S$. La situation de jeu est celle d'un dilemme du prisonnier quand $T > R > P > S$ et $T+S \leq 2R$. Nous adopterons les valeurs de référence $T=5$, $R=3$, $P=1$, $S=0$.

Le nom «dilemme du prisonnier» provient d'un récit imaginaire. Deux individus armés sont arrêtés devant une banque. Un juge voudrait établir qu'ils étaient sur le point de mener une attaque. Il les interroge.

– Si l'un avoue, c'est-à-dire trahit son compère, et que l'autre poursuit la coopération avec son compère, situation notée $[t, c]$, celui qui avoue est libéré: rétribution de $T=5$ années de liberté. Celui qui n'a pas avoué va en prison 5 ans: rétribution nulle, $S=0$, il écope du maximum prévu pour une attaque de banque.

– Si les deux compères restent solidaires en n'avouant rien, $[c, c]$, ils vont 2 ans en prison chacun pour port illégal d'armes: rétribution de $R=3$ années de liberté par rapport aux 5 ans du pire cas.

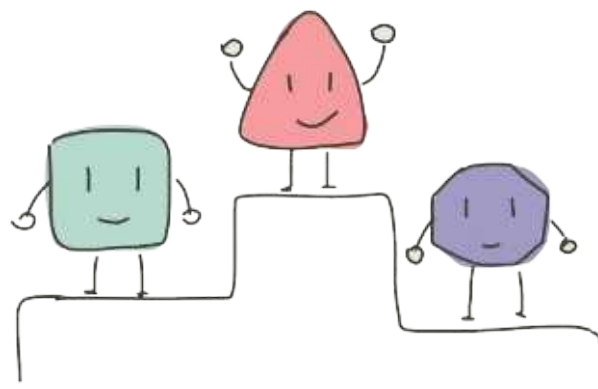
– Si les deux personnages avouent, $[t, t]$, ils font chacun 4 ans de prison: rétribution de 1 année de liberté par rapport au pire cas pour les remercier de leurs aveux, soit $P=1$.

Un comportement rationnel conduit à la trahison: elle produit toujours un meilleur résultat que la coopération. En effet, si l'individu A coopère, l'individu B gagne 5 points en jouant t et seulement 3 points en jouant c; si A trahit, B obtient 1 point en jouant t, et 0 en jouant c. Pour B, le choix t est ainsi toujours meilleur que le choix c. On parle de dilemme, car les deux >



Jean-Paul Delahaye a notamment publié: **Les Mathématiciens se plient au jeu**, une sélection de ses chroniques parues dans *Pour la Science* (Belin, 2017).

2



SIX STRATÉGIES ET QUELQUES COMBATS

Au dilemme du prisonnier, deux stratégies se rencontrent, chacune gagnant des points selon les règles explicitées dans l'article et précisées ici sur un exemple: – si les deux jouent c, chacune emporte 3 points: $[c, c] \rightarrow [3, 3]$ – si les deux jouent t, chacune remporte 1 point: $[t, t] \rightarrow [1, 1]$ – si A joue t et B joue c, alors A remporte 5 points et B aucun: $[t, c] \rightarrow [5, 0]$.

Lorsque le jeu est répété, le joueur adopte une stratégie qui indique comment il réagit au passé de la rencontre. Voici six stratégies possibles.

Gentille: je coopère quels que soient les coups passés.

Méchante: je trahis quels que soient les coups passés.

Donnant-donnant: je commence par coopérer, puis, au coup n , je joue ce que mon adversaire a joué au coup $n-1$.

L'idée est d'inciter l'adversaire à coopérer en réagissant à chacune de ses trahisons. On montre que **Donnant-donnant** ne perd jamais plus de 5 points dans une rencontre, quels qu'en soient la durée et l'adversaire, mais aussi qu'il ne fait jamais plus que son adversaire.

Méfiant: je commence par trahir, puis, au coup n , je joue ce que mon adversaire a joué au coup $n-1$.

Sondeur: je joue trahir-coopérer-coopérer pour les trois premiers coups, puis, si mon adversaire a coopéré aux coups 2 et 3, je trahis toujours, sinon je joue **Donnant-donnant**. L'idée est de tester les réactions de l'adversaire: s'il ne réagit

pas à la première trahison, je l'exploite au maximum, sinon je reste prudent.

Graduelle: je commence par coopérer, puis je coopère autant que c'est possible sauf si l'autre me trahit, auquel cas je réplique par une période de n trahisons successives, où n est le nombre de trahisons passées de mon adversaire, suivie de deux étapes de coopération. L'idée est d'être d'autant plus sévère que l'adversaire a souvent trahi, mais de tenter une réconciliation après chaque période de rétorsion en coopérant deux fois de suite pour calmer le jeu.

Voici quelques exemples de rencontre de durée 10.

1) **Donnant-donnant**

$0+5+0+5+0+5+0+5+0+5=25$
c t c t c t c t c t
t c t c t c t c t c
 $5+0+5+0+5+0+5+0+5+0=25$

Méfiant

Chaque stratégie emporte 25 points à l'issue des 10 coups.

2) **Sondeur**

$5+0+3+3+3+3+3+3+3+3=29$
t c c c c c c c c c
c t c c c c c c c c
 $0+5+3+3+3+3+3+3+3+3=29$

Graduelle

Chaque stratégie recueille 29 points.

3) **Méchante**

$5+1+5+5+1+1+1+1+5+5=30$
t t t t t t t t t t
c t c c t t t t c c
 $0+1+0+0+1+1+1+1+0+0=5$

Graduelle

Ici, **Méchante** emporte 30 points, et **Graduelle** seulement 5.

Le tableau des points gagnés pour des parties de longueur 100 est:

	Sondeur	Graduelle	Donnant-donnant	Méfiant	Gentille	Méchante	Total
Sond.	104	299	299	295	496	98	1 591
Grad.	299	300	300	296	300	89	1 584
Donn.	299	300	300	250	300	99	1 548
Méfi.	300	301	250	100	302	100	1 353
Gent.	6	300	300	297	300	0	1 203
Méch.	108	144	104	100	500	100	1 056

> individus emportent 6 points en jouant [c, c], alors qu'ils en gagnent 5 en jouant [c, t] et 2 en jouant [t, t]. L'intérêt collectif est de jouer [c, c], mais une analyse rationnelle égoïste conduit à [t, t], qui est collectivement le pire cas! Dans ce jeu à «somme non nulle», ce que certains gagnent n'est pas nécessairement pris à d'autres et donc bien s'entendre est payant.

Le dilemme est «itéré» quand la situation se présente de façon répétée aux deux individus. Jouer consiste alors à choisir une stratégie déterminant, chaque joueur étant informé du passé, comment jouer le coup suivant. Six stratégies sont définies dans l'encadré 2.

Nous décrirons cinq méthodes équitables et stables de classification d'un ensemble de stratégies. La durée (nombre de coups) des rencontres pour nos calculs est toujours 100, mais dès que le nombre de coups dépasse une dizaine, les résultats changent très peu.

Nous verrons que les méthodes aboutissent parfois à des classements différents. Il est difficile d'analyser les divergences entre méthodes, qui sont pourtant reproductibles, car si l'on refait les évolutions probabilistes, chaque méthode produit en pratique toujours le même classement.

Précisons qu'aucune de ces méthodes ne permet de donner un classement absolu entre toutes les stratégies envisageables: quand on change l'ensemble des compétiteurs, une stratégie bonne peut devenir mauvaise et inversement. Le problème posé n'est pas de construire un ordre absolu entre toutes les stratégies imaginables, mais de déterminer si, un ensemble de compétiteurs étant fixé, les divers systèmes de classifications entre ces compétiteurs divergent sensiblement, et si oui pourquoi.

Dans la suite, on se donne un ensemble de stratégies $S = \{s_1, s_2, \dots, s_n\}$ qu'on tente de classer par ordre de réussite quand elles jouent les unes avec les autres au dilemme du prisonnier.

Ce problème de la classification d'un ensemble de stratégies a été posé dès les premiers articles de Robert Axelrod et William Hamilton sur le sujet, en 1981. Dans son livre (*voir la bibliographie*), Robert Axelrod a lui-même largement pris la peine de comparer et analyser les différences entre les méthodes d'évaluation des gains que nous nommons M0 (nombres de parties gagnées), M1 (nombre de points gagnés), M2 (évolution déterministe).

Dans les recherches publiées sur le sujet, plusieurs autres méthodes ont été envisagées pour mener les classifications, mais le plus souvent les auteurs choisissent M1. Les méthodes M3 et M4 que nous décrivons sont des variantes probabilistes de M1 et M2; alors que nous pensions qu'elles devaient conduire à des classements identiques ou proches de ce que donnent M1 ou M2, nous avons découvert que les résultats sont plus compliqués et parfois étonnants.

M0: LE NOMBRE DE VICTOIRES

Notons M0 le tournoi des victoires. Son idée est simple: chaque paire s_i, s_j de stratégies est considérée, elles jouent une partie l'une contre l'autre et celle qui emporte le plus de points gagne. Le score de chaque stratégie est compté de la manière suivante: +1 pour chaque partie gagnée, -1 pour chaque partie perdue et 0 en cas d'égalité. Avec les six stratégies données en exemple à l'encadré 2, on obtient, en s'aidant du tableau de points gagnés: *Méchante* 4, *Méfiant* 3, *Donnant-donnant* -1, *Sondeur* -1, *Graduelle* -2, *Gentille* -3. On notera donc:

M0: Méc > Méf > Don = Son > Gra > Gen.

Le résultat n'est pas surprenant, car *Méchante* ne perd aucune partie: à chaque fois qu'elle joue, elle obtient au moins autant de points que son adversaire, quoi qu'il joue. Ce nombre de victoires est cependant une piètre mesure de la qualité des stratégies, car pour

3

CLASSEMENTS RESSEMBLANTS, MAIS DIFFÉRENTS

Nous avons recherché un cas amenant au désaccord le plus extrême. Aux stratégies *Méfiant* et *Sondeur*, nous avons associé *Pttc*, la stratégie qui joue périodiquement t, t, c, et *Pcct*, la stratégie qui joue périodiquement c, c, t. Les quatre stratégies ont toute la propriété de prendre l'initiative de trahir et c'est sans doute ce qui explique le désordre résultant. Les classements obtenus divergent sensiblement, même si l'on ne prend pas en compte M0. Aucun classement n'est meilleur que les autres.

M0 : Pttc = Méc > Son > Pcct

M1 : Pcct > Pttc > Méc > Son

M2 : Pttc > Pcct > Méc > Son

M3 : Méc > Son > Pcct > Pttc

M4 : Méc > Son > Pttc > Pcct

