Théorie de la décision et théorie des jeux – TD 1
Année 2006–07 – Corrigé partiel

Première séance : Les exercices au programme étaient les exercices 1 et 2 (manipulation de $\succ$), 4, 10 et 11 (information parfaite), 5, 6, et 9 (information imparfaite). Les exercices 3, 7, et 8 ne seront finalement pas abordés : ils mettent en jeu la notion d'aversion au risque, qui ne sera pas étudiée dans le cadre du cours de théorie des jeux.

Exercice 1 (Axiome de Houthakker). L'axiome de Houthakker peut être illustré comme suit : "Si le champion de l'île d'Irlande est originaire d'Ulster, alors tout champion du Royaume-Uni originaire d'Ulster est aussi champion d'Irlande".

On doit prouver qu'il est équivalent à la conjonction des axiomes de Sen, que l'on rappelle :

$$(\alpha) \quad : \quad \text{Si } x \in B \subset A \text{ et } x \in c(A)\text{, alors } x \in c(B)$$
$$(\beta) \quad : \quad \text{Soit } x{,}y \in c(A)\text{, } A \subset B\text{. Si } y \in c(B)\text{, alors } x \in c(B)$$

- Montrons que (α) et (β) entraînent l'axiome de Houthakker :
  Soient $x{,}y \in A \cap B$, $x \in c(A)$, $y \in c(B)$. Prouvons que $x \in c(B)$.
  Par (α), puisque $x \in A \cap B \subset A$ et $x \in c(A)$, on a $x \in c(A \cap B)$. De même, $y \in A \cap B \subset B$ et $y \in c(B)$ entraînent $y \in c(A \cap B)$.
  Et maintenant, de $x{,}y \in c(A \cap B)$, où $A \cap B \subset B$, et $y \in c(B)$, on déduit par (β) que $x \in c(B)$, ce qui conclut.
- Montrons que l'axiome de Houthakker entraîne (α).
  Soit $x \in B \subset A$ tel que $x \in c(A)$. $c(B)$ est non-vide parce que c est une fonction de choix. Soit donc $y \in c(B)$.
  Les prémisses de l'axiome d'Houthakker, $y \in c(B) \subset B = A \cap B$, $x \in B = A \cap B$, $x \in c(A)$, $y \in c(B)$ sont vérifiés. Il s'ensuit que $x \in c(B)$.
- Montrons que l'axiome de Houthakker entraîne (β).
  Soit $A \subset B$, $x{,}y \in c(A)$ et $y \in c(B)$. Comme $x, y \in c(A) \subset A = A \cap B$, $x \in c(A)$, et $y \in c(B)$, il suit par l'axiome de Houthakker que $x \in c(B)$.

Exercice 2 (Préférences lexicographiques).
1. Profitons-en pour souligner les liens entre relations d'ordre et relations de préférences. Une relation d'ordre (large) $\preccurlyeq$ est par définition une relation binaire réflexive, antisymétrique (i.e., $x \preccurlyeq y$ et $y \preccurlyeq x$ entraînent $x = y$), transitive. Un ordre strict $\succ$ est quant à lui une relation binaire irréflexive et transitive (donc également asymétrique). Un ordre strict $\succ$ correspond à un unique ordre large $\preccurlyeq$, selon $y \preccurlyeq x$ si et seulement si $\neg(x \succ y)$.

Il s'ensuit qu'une relation de préférences rationnelle $\succ$ (i.e., asymétrique et négativement transitive) est une relation d'ordre strict.

Réciproquement, lorsque $\succ$ est un ordre total strict (i.e., deux éléments distincts peuvent toujours être comparés), il est également négativement transitif : soit $\neg\,(x \succ y)$ et $\neg\,(y \succ z)$, pour x, y, z trois élements distincts. Alors, comme l'ordre est total, $y \succ x$ et $z \succ y$, soit par transitivité, $z \succ x$, ce qui montre, à nouveau par complétude, que $\neg\,(x \succ z)$. En conclusion, toute relation d'ordre strict et total est également une relation de préférences rationelle.

De manière encore plus immédiate, et par définition, une relation d'ordre large $\preccurlyeq$ totale est definit une relation de préférences rationnelle[1].

Ici, l'ordre lexicographique $\succ$ sur deux éléments est un ordre strict total, c'est donc également une relation de préférences rationnelle.

2. Par théorème, on sait qu'une telle représentation est possible si et seulement si l'ensemble en question contient un sous-ensemble dense $\succ$-ordonné. Ce n'est pas le cas ici, mais on va prouver l'impossibilité de la représentation sans ce résultat.
On raisonne par l'absurde et on suppose qu'une telle représentation u existe. Les contradictions vont être apportées à chaque fois par un argument de cardinalité.

Solution astucieuse (1) : Comme pour tout $r \in [0,1]$, $(r,1) \succ (r,0)$, on a $u((r,1)) > u((r,0))$, et $d(r) = u((r,1)) - u((r,0)) > 0$. Ainsi,

$$[0,1] = \bigcup_{n \in \mathbb{N}^*} \{r : d(r) > 1/n\} \ .$$

L'ensemble de gauche est non-dénombrable, et à droite nous avons affaire à une union dénombrable. L'union porte donc sur au moins un ensemble non-dénombrable, indexé par n^0. On note $K = u((1,1)) - u((0,0))$, et soit un entier N tel que $N > Kn^0 + 1$. On prend N éléments distincts dans $\left\{r \in [0,1] : d(r) > 1/n^0\right\}$, et on les ordonne, $r_1 < r_2 < \ldots < r_N$.
Comme pour tout $2 \leqslant n \leqslant N$, $(r_n, 0) \succ (r_{n-1}, 1)$, on a également $u((r_n, 0)) > u((r_{n-1}, 1))$. Ainsi,

$$u((r_n,0)) - u((r_{n-1},0)) > u((r_{n-1},1)) - u((r_{n-1},0)) > 1/n^0 \ .$$

Finalement,

$$\begin{aligned} K &= u((1,1)) - u((0,0)) \\ &= [u((1,1)) - u((r_N,0))] + [u((r_N,0)) - u((r_{N-1},0))] + \ldots \\ &\quad + [u((r_2,0)) - u((r_1,0))] + [u((r_1,0)) - u((0,0))] \\ &> 0 + 1/n^0 + 1/n^0 + \ldots + 1/n^0 + 0 > (N-1)/n^0 > K \end{aligned}$$

qui forme la contradiction recherchée.

Solution très astucieuse (2) : A tout x de [0,1], on associe un rationnel $r(x) \in I(x) =]u((x,0)), u((x,1))[$. (Il en existe car cet intervalle est non-vide.) Notons que comme les $I(x)$ sont disjoints (pour $x < x'$, $I(x)$ est strictement à gauche de $I(x')$), la fonction r définit une injection de [0,1] dans les rationnels, ce qui est impossible vu la cardinalité des ensembles en jeu.

Solution vue en TD (3) : On pose $f(x) = u((x,0))$. f est (strictement) croissante, donc admet au plus un nombre dénombrable de discontinuités (admet au plus $n(f(1) - f(0))$ sauts de discontinuité[2] plus grands que $1/n$). Il existe au moins un point de continuité $x_0 \in]0,1[$,

1. Au passage, notez que toute relation transitive et totale est réflexive.
2. Le saut de discontinuité en x vaut $f^+(x) - f^-(x)$, où $f^+(x)$ et $f^-(x)$ sont les limites à droite et à gauche de f en x (elles existent bien, par croissance de f).

et pour ce point et tout $\varepsilon > 0$,

$$u((x_0,0)) = f(x_0) < u((x_0,1)) < u((x_0+\varepsilon,0)) = f(x_0+\varepsilon)$$

soit, en faisant tendre ε vers 0, $f(x_0) = u((x_0,0)) = u((x_0,1))$, alors qu'on devrait avoir $u((x_0,0)) < u((x_0,1))$.

2(bis). Question subsidiaire : Au vu des questions 2 et 3, on peut se demander s'il existe une représentation de l'ordre lexicographique sur le simplexe de $\mathbb{R}^2$, i.e., les probabilités sur deux éléments. Oui, et c'est même évident, parce que ce simplexe est en bijection, via une bijection qui préserve l'ordre, à [0,1], muni de l'ordre habituel $<$. On peut proposer par exemple la première projection, $u((p,1-p)) = p$.

3. Un projet correspond à un triplet (p_1,p_2,p_3) appartenant au simplexe d'ordre 2 (on a $p_i \geqslant 0$ et $p_1+p_2+p_3 = 1$). Il n'y a que 2 degrés de liberté. Graphiquement, un projet est donc représenté de façon unique par un point de l'ensemble

$$S = \{(p_2,p_3) : p_2 \geqslant 0,\ p_3 \geqslant 0,\ p_2+p_3 \leqslant 1\} \ ,$$

qui est la projection du simplexe sur le plan (e_2,e_3). (S est un triangle rectangle isocèle de côtés les vecteurs unitaires.) Cette projection étant un homéomorphisme, i.e., une application bijective et bicontinue, qui en outre préserve l'ordre, on peut raisonner sur S de façon équivalente par rapport aux axiomes de VNM, qui ne mettent en jeu que la relation de préférences $\succ$ (que l'on considère ici comme une relation d'ordre strict total) et la topologie (via l'axiome de continuité).

Les axiomes de VNM sont les axiomes de rationalité, d'indépendance et de continuité. La rationalité est vérifiée (puisque, ici encore, l'ordre lexicographique considéré, qui certes n'est pas l'ordre lexicographique canonique sur $\mathbb{R}^3$, est une relation d'ordre strict et total).

L'axiome d'indépendance s'écrit : Soit p, q et $r \in S$, $\lambda \in]0,1]$, alors $p \succ q$ implique $\lambda p + (1-\lambda)r \succ \lambda q + (1-\lambda)r$. Il est vérifié dans le cas présent. En effet, supposons que $p \succ q$. Ou bien $p_3 < q_3$ et donc $\lambda p_3 < \lambda q_3$, puis $\lambda p_3 + (1-\lambda)r_3 < \lambda q_3 + (1-\lambda)r_3$; ou bien $p_3 = q_3$ et $p_2 < q_2$, alors $\lambda p_3 + (1-\lambda)r_3 = \lambda q_3 + (1-\lambda)r_3$ et $\lambda p_2 + (1-\lambda)r_2 < \lambda q_2 + (1-\lambda)r_2$.

L'axiome de continuité s'écrit : $\forall p \in S$, $\{q : q \succ p\}$ et $\{q : q \prec p\}$ sont ouverts. Or,

$$\begin{aligned} \{q \in S : q \succ p\} &= \{(q_2,q_3) \in S : q_3 < p_3\} \cup \{(q_2,q_3) \in S : q_3 = p_3,\ q_2 < p_2\} \\ &= (S\,(\cap[0,1] \times [0,p_3[)) \cup ([0,p_2[\times \{p_3\}) \ . \end{aligned}$$

Lorsque $p_2 > 0$, le second ensemble de la dernière égalité est non-vide, mais aucun de ses élements n'admet de voisinage dans $\{q \in S : q \succ p\}$. Ce dernier n'est donc pas ouvert.

Remarque : les axiomes de continuité et d'indépendance impliquent la propriété suivante. Pour tous $\forall p,q,r$ tels que $p \succ r \succ q$, il existe un (unique) $\lambda \in]0,1[$ tel que $\lambda p + (1-\lambda)q \sim r$. Lorsque l'on prend p, q, r avec des ordonnées strictement croissantes ($p \succ r \succ q$), sans que les trois points ne soient alignés, on note qu'une combinaison convexe de p et q passe d'être préférée à q à ne plus l'être, sans passer par une loterie équivalente (en effet, notez que la seule loterie équivalente à r est la loterie r elle-même, ... tout simplement, encore une fois, parce que l'ordre lexicographique est un ordre total et strict).

L'axiome de continuité est violé, on ne peut donc pas représenter les préférences lexicographiques sur le simplexe de $\mathbb{R}^3$ par une utilité de VNM.

Exercice 4 (Jeu à Las Vegas). (L'arbre du jeu est en annexe.) On procède par *backward induction*, en tout cas sur la branche qui correspond à avoir misé 1 jeton au premier tour. En effet, les cas où l'on mise 2 jetons ou plus au premier tour mènent à gagner le jeu avec probabilité 2/3. Nous calculons maintenant la probabilité de gagner le jeu lorsqu'un seul jeton a été joué au premier tour.
Les deux sous-arbres entourés en trait plein sur la figure correspondent au cas où le joueur a trois jetons et il ne lui reste plus qu'un tour à jouer. Alors, le joueur est obligé de miser 2 ou 3 jetons et de gagner le tour. Il gagne le jeu avec probabilité 1/3.
Reste à étudier les eventualités du deuxième tour de jeu. Lorsqu'au début de ce tour, le joueur a 4 jetons, s'il mise 2 jetons ou plus, sa probabilité de victoire est 2/3, et s'il mise 1 jeton, il gagne immédiatement le jeu avec probabilité 2/3 et avec probabilité 1/3 se retrouve dans le cas décrit au paragraphe précédent. Cette seconde stratégie est meilleure. Ainsi, lorsqu'il a 4 jetons, le joueur mise un jeton, et sa probabilité de gagner le jeu est de $2/3 + 1/3 \times 2/3 = 8/9$.
Lorsqu'il a 2 jetons et qu'il reste 2 tours de jeu, qu'il mise 2 d'abord puis au moins un, ou qu'il mise un jeton d'abord, puis au moins 2, il est obligé de gagner deux fois de suite pour remporter le jeu, d'où une probabilité de victoire de 4/9 dans ce cas.
Ainsi, au début du jeu, miser 1 jeton au premier coup amène à remporter le jeu avec une probabilité de $1/3 \times 4/9 + 2/3 \times 8/9 = 20/27$ de gagner le jeu (contre seulement 2/3 lorsqu'il mise 2 ou 3 jetons au premier tour).

La probabilité de victoire est ainsi 20/27, et nous avons décrit au cours de l'induction à rebours (*backward induction*) des stratégies optimales (il y en a plusieurs).

Exercice 5 (Valeur de l'information).
1., 2., et 3. Le joueur connaît p mais il ignore le résultat du lancer. Quand la pièce est tirée d'abord, dans la mesure où le joueur n'a pas accès au résultat du lancer, son ensemble d'information est vide. Il n'a que deux stratégies possibles, pile ou face, et ne peut moduler son choix en fonction d'aucune information extérieure. Cela est la même chose quand la pièce est tirée après l'annonce. Dans les deux cas, le gain espéré de la stratégie pile est pg_p et celui de face est $(1-p)g_f$.
Le joueur connaissant p, g_p et g_f jouera pile si $pg_p \geqslant (1-p)g_f$ et face sinon, et obtiendra $\max\{pg_p, (1-p)g_f\}$, qui est la valeur du jeu. (Ici, la valeur du jeu est le gain maximal que le joueur peut obtenir.)

5. Rappel : Le supremum d'un nombre quelconque de fonctions convexes est encore convexe, pour peu qu'il soit à valeurs finies. D'ailleurs, on a même que toute fonction convexe à valeurs réelles est l'enveloppe supérieure d'une famille de fonctions affines, celles dont le graphe est une droite d'appui du graphe de la fonction convexe considérée.

Pour prouver d'une autre manière la convexité de la valeur du jeu, on note que dans le premier cas, le joueur n'a que deux stratégies (pile et face), puisqu'il n'a pas d'information extérieure. Pile lui rapporte en espérance $(\lambda p' + (1-\lambda)p'')g_p = pg_p$ et de même, face lui rapporte $(1-p)g_f$.

Au total, la valeur ce jeu est

$$\max\{pg_p, (1-p)g_f\} = v(p) = v(\lambda p' + (1-\lambda)p'') \ .$$

Mais s'il est informé du premier tirage, le joueur a quatre stratégies possibles, pile–pile, pile–face, face–pile et face–face, le premier terme indiquant le choix si la probabilité est p' et le deuxième, celui si la probabilité est p''.
On sait que les valeurs dans chaque sous-jeu sont $\max\{p'g_p, (1-p')g_f\}$ et $\max\{p''g_p, (1-p'')g_f\}$ (obtenues pour la stratégie optimale dans le sous-jeu, voir questions 1 à 3).
La valeur espérée du jeu originel est donc (par induction à rebours)

$$\lambda \max\left\{p'g_p, (1-p')g_f\right\} + (1-\lambda)\max\left\{p''g_p, (1-p'')g_f\right\} = \lambda v(p') + (1-\lambda)v(p'') \ .$$

Cette valeur est supérieure à celle de l'autre jeu, car les stratégies pile–pile et face–face correspondent aux stratégies pile et face que nous y avions vues. On a $\lambda v(p') + (1-\lambda)v(p'') \geqslant v(p)$, d'où la convexité de v.
On voit ainsi que le joueur gagne plus lorsqu'il est informé du premier tirage. C'est bien naturel, puisqu'il peut adapter sa stratégie en fonction de l'information dont il dispose.
De plus, si cette information ne lui était pas profitable, il lui suffirait de ne pas en tenir compte, c'est-à-dire que les stratégies avec moins d'information peuvent être vues comme un sous-ensemble de celles utilisant plus d'information. On maximise donc le paiement par rapport à un ensemble plus grand de stratégies.
Remarque : Cet argument n'est valable que parce que le joueur est seul ! (Sinon, la non-utilisation de son information pourrait ne pas être crédible aux yeux des autres joueurs, et ceux-ci pourraient agir en conséquence, ce qui peut conduire à une valeur négative de l'information...)

Exercice 6 (Les trois portes). La figure se trouve dans les annexes. Elle présente l'arbre du jeu, avec les sept nœuds d'information, $J_1', J_{AB}', J_{AC}', J_{BA}', J_{BC}', J_{CA}', J_{CB}'$.

1. Après le choix par la nature de la porte, le joueur se trouve dans un premier ensemble d'information J_1' contenant trois nœuds, chacun ayant trois choix possibles $\{A, B, C\}$, ce choix étant le nom de la porte annoncé par le joueur.
Au deuxième niveau de l'arbre, il y a 6 ensembles d'information, correspondant à un choix du joueur et à une annonce du présentateur, et indexés par $\{AB, AC, BA, BC, CA, CB\}$. Chaque ensemble d'information contient deux nœuds (la voiture est derrière l'une des deux portes non annoncées par le présentateur), et à chaque nœud, le joueur a deux choix, les deux portes non annoncées par le présentateur.
Une stratégie (en information incomplète) associe une action à chaque ensemble d'information. Puisqu'il y a trois actions possibles en J_1' et deux pour chacun des J_{XY}', cela fait $3 \times 2^6 = 192$ stratégies possibles.

Remarque : Si l'on ne s'intéresse qu'aux stratégies équivalentes, i.e., aux choix qui se présentent vraiment, il s'agit de regarder le nombre d'ensembles d'information dans lesquels le joueur peut être après son premier choix. Il n'y en a que deux (sur les six), qui correspondent aux deux annonces possibles du présentateur. Cela ne fait plus que $3 \times 2^2 = 12$ stratégies. Celles-ci précisent le choix initial du joueur, puis sa décision de changer ou non de porte en fonction des deux annonces possibles du présentateur.

2. On se place dans l'un des ensembles d'information qui suivent l'annonce faite par le présentateur. Cet ensemble d'information correspond à un couple annonce joueur–annonce présentateur, et contient 2 sommets. L'un correspond à une annonce du joueur correcte (et le présentateur avait tiré une porte au hasard entre les deux mauvaises portes restantes), et l'autre à une annonce erronée (auquel cas le présentateur n'avait alors pas le choix de la porte à ouvrir). Par application de la formule de Bayes, on voit que la probabilité de se trouver au nœud correpondant à une annonce joueur erronée, conditionnellement au fait de se trouver dans le dit ensemble d'information, est de $2/3$, et que la probabilité de l'autre nœud est $1/3$. De là, l'espérance de gain si l'on applique la stratégie "changer" est $2/3$, contre $1/3$ si l'on conserve son annonce initiale.

On peut aussi remarquer que si l'on a raison à la première annonce (ce qui arrive avec probabilité $1/3$), on perd si l'on change, mais que si l'on avait tort (probabilité $2/3$), on gagne.

3. La stratégie optimale se calcule par *backward induction*, puisque l'on connaît maintenant la strategie optimale au second coup. On remarque alors que le premier choix est indifférent.

Exercice 9 (Valeur de l'information, bis théorique). Dans le jeu avec partition de l'information selon $\mathcal{P}$, une stratégie φ est une application qui à tout élément $B \in \mathcal{P}$ associe une action $a^B = \varphi(B)$. Le gain en espérance de φ vaut

$$g_\varphi = \sum_{B \in \mathcal{P}} \left(\sum_{k \in B} p_k \, g\,(k,\varphi(B)) \right) .$$

Lorsque la probabilité $\boldsymbol{p}$ sur K est connue, la stratégie optimale $\varphi^*_{\mathcal{P}}$ est donnée par

$$\varphi^*(B) = \operatorname*{argmax}_{a \in A} \sum_{k \in B} p_k \, g\,(k,a)$$

et son gain en espérance vaut la valeur $G(\boldsymbol{p},\mathcal{P})$ du jeu avec partition $\mathcal{P}$,

$$G(\boldsymbol{p},\mathcal{P}) = g_{\varphi^*} = \sum_{B \in \mathcal{P}} \max_{a \in A} \left(\sum_{k \in B} p_k \, g\,(k,a) \right) . \tag{1}$$

1. La valeur $G(\boldsymbol{p})$ du jeu sans partition correspond à la partition $\mathcal{Q}$ formée par le singleton $\{K\}$, est plus petite que toute valeur $G(\boldsymbol{p},\mathcal{P})$, car elle vaut selon (1),

$$G(\boldsymbol{p}) = \max_{a \in A} \sum_{k \in K} \big(p_k \, g\,(k,a) \big) .$$

2. On a $|\mathcal{P}|$ sous-jeux indexés par les éléments de $\mathcal{P}$, et à chacun d'entre eux on applique le résultat de la question précédente. On conclut en utilisant la linéarité de (1) en les éléments de la partition.

Exercice 10 (Décision dynamique en horizon infini).
1. On suppose que le joueur connaît g et T. Une stratégie σ est une suite de fonctions $\sigma_1, \sigma_2, \ldots$; la t-ième de ces fonctions associe à l'information h_t disponible au début du tour t (avant que le joueur ne joue), $h_t = (k_1, a_1, \ldots, k_{t-1}, a_{t-1}, k_t)$, une action $a_t = \sigma_t(h_t)$. En

fait, $k_1 = k$ et les états $k_2, k_3, \dots$ se déduisent des actions passées $a_1, a_2, \dots$ et de l'état initial k, puisque T est connue. On peut donc prendre simplement $h_t = (k, a_1, \dots, a_{t-1})$. Mais une simplification est encore possible : a_1 ne dépend que de k, de sorte que a_2, qui dépend de a_1 et k ne dépend également que de k, etc.
Finalement, une stratégie est donnée simplement par $|K|$ suites d'élements de A, où $|K|$ désigne le cardinal de K. On note une stratégie $\sigma : k \mapsto (a_1^k, a_2^k, \dots) \in A^{\mathbb{N}}$. Leur ensemble est donc $(A^{\mathbb{N}})^K$.
Notons qu'il y a un nombre infini (certes dénombrable) de stratégies, et qu'il n'est donc pas évident qu'il existe une stratégie optimale. Voir par exemple ce qui se passe dans le jeu où, jouant $x \in [0,1[\cap\mathbb{Q}$, on gagne x. Quelle serait une stratégie optimale ? (Pourtant, on peut définir que la valeur du jeu est 1.) Cependant, nous allons ici exhiber la stratégie optimale, et voir qu'elle a une forme très simple. Elle peut en fait être décrite par une application $K \to A$.

2. Montrons que F est β-contractante. Comme $\beta < 1$, le théorème de point fixe de Picard (qui vaut pour les espaces métriques complets non-vides) assurera l'existence d'un unique point fixe. Précisément, montrons que pour tous x, y,

$$\|F(x) - F(y)\|_\infty \leqslant \beta \, \|x - y\|_\infty \ .$$

Fixons la composante k. Soit $a_{x,k}$ l'élément de A atteignant l'argmax dans la définition de $F(x)_k$. On a

$$\begin{aligned} F(x)_k - F(y)_k &\leqslant \Big(g(k,a_{x,k}) + \beta \, x_{T(k,a_{x,k})}\Big) - \Big(g(k,a_{x,k}) + \beta \, y_{T(k,a_{x,k})}\Big) \\ &= \beta \Big(x_{T(k,a_{x,k})} - y_{T(k,a_{x,k})}\Big) \leqslant \beta \, \|x - y\|_\infty \ . \end{aligned}$$

Par symétrie, notre assertion est donc prouvée.

3. On définit notre stratégie sous forme extensive : au tour t, si l'état au début du tour est k_t, on joue l'argument maximum dans la définition de $F(v)_{k_t}$, i.e.,

$$a_t = \operatorname*{argmax}_{a \in A} \left\{ g(k_t,a) + \beta \, v_{T(k_t,a)} \right\} \ .$$

(Notons que comme le joueur connaît g et T, il connaît bien v.) Cette stratégie est particulièrement simple, elle est décrite par une application $K \to A$.
L'utilité de cette stratégie est v_k, comme on peut le voir par récurrence,

$$\underbrace{g(k_1,a_1) + \beta \Big(\underbrace{g(k_2,a_2) + \beta \quad \underbrace{(\dots)}_{=v_{T(k_2,a_2)}}}_{=F(v)_{k_2}=v_{k_2}=v_{T(k_1,a_1)}} \ \Big)}_{=F(v)_{k_1}=v_{k_1}=v_k} \ .$$

Un raisonnement propre consisterait à poser l'hypothèse de récurrence

$$\mathcal{P}_n : v_{k_1} = \sum_{t=1}^{n-1} \beta^t g(a_t,k_t) + \beta^{n-1} v_{k_n} \ .$$

4. Le fait que W soit un point fixe vient simplement du fait qu'une stratégie, comme le montre la question 1, peut être décomposée en un élément de A^K (on spécifie ce qui se passe au premier coup) et $|K|$ suites d'actions (ce qui se passe aux coups $t \geqslant 2$). Le fait que $F(W) = W$

vient simplement par *backward induction*: voir figure correspondante à la fin du corrigé.
On en conclut qu'à la question 3, on a explicitement décrit cette stratégie optimale, par unicité du point fixe de F.

5. Par la question 1, l'espace des stratégies est le produit de $|K|$ copies de l'espace métrique et compact des suites de A (qui est métrique et compact parce que produit d'un nombre dénombrable d'ensembles compacts et métriques, car finis). Plus précisément, on choisit sur A la distance de Hamming et on l'étend à $A^{\mathbb{N}}$ en considérant une série convergente. Ce faisant, on a métrisé la convergence ponctuelle des suites de A. (Consultez tout livre de topologie pour les détails, si vous n'êtes pas familiers de ce genre de raisonnements.)
L'application qui à une stratégie φ_k pour l'état $k \in K$ associe $U_k(\varphi_k)$ est continue, parce que séquentiellement continue (par théorème de convergence dominée) et définie sur un espace métrique. Comme l'espace de définition est en outre compact, on sait que cette application admet un maximum, dont l'antécédent est noté φ_k^*.
La stratégie optimale recherchée est alors $(\varphi_k^*)_{k\in K}$.

6. On a deux états, $K = \{J, D\}$, et on note $x = (x_D, x_J)$. Il s'agit de calculer le point fixe de l'application F définie par

$$\begin{aligned} F(x)_J &= \max\big\{10 + \beta x_D,\, 1 + \beta x_J\big\} \\ F(x)_D &= \max\big\{2 + \beta x_D,\, \beta x_J\big\} \end{aligned}$$

Si on avait pour point fixe x tel que $x_J = 1 + \beta x_J$, alors x_J vaudrait $1/(1-\beta)$, et pour x_D, on aurait

$$x_D = \max\left\{2 + \beta x_D,\, \frac{\beta}{1-\beta}\right\} = \frac{\beta}{1-\beta} \text{ ou } \frac{2}{1-\beta}$$

(puisque la solution de $x_D = 2 + \beta x_D$ est $2/(1-\beta)$). Mais alors,

$$F(x)_J \geqslant 10 + \beta x_D \geqslant 10 + \frac{\beta^2}{1-\beta} > 1/(1-\beta) = x_J$$

ce qui contredit la définition de x comme point fixe. (Pour obtenir l'inégalité stricte ci-dessus, voir qu'elle est équivalente à $9 - 10\beta + \beta^2 = (\beta-1)(\beta-9) > 0$). C'est donc que $x_J = 10 + \beta x_D$. Dès lors,

$$x_D = F(x)_D = \max\big\{2 + \beta x_D,\, \beta\,(10 + \beta x_D)\big\} = \frac{2}{1-\beta} \text{ ou } \frac{10\beta}{1-\beta^2}\ .$$

On a $x_D = 2/(1-\beta)$ si et seulement si

$$2 + \beta\frac{2}{1-\beta} = \frac{2}{1-\beta} \geqslant \beta\left(10 + \beta\frac{2}{1-\beta}\right)\ ,$$

c'est-à-dire, si et seulement si $2 \geqslant \beta(10 - 8\beta)$, ou encore, $4\beta^2 - 5\beta + 1 = 4\,(\beta-1)\,(\beta - 1/4) \geqslant 0$. Cela arrive si et seulement si $\beta \leqslant 1/4$ (vu $0 < \beta < 1$). Dans le cas où $\beta \geqslant 1/4$, $x_D = 10\beta/(1-\beta^2)$.

Exercice 11 (Exercice supplémentaire).
Enoncé : Un casino propose le jeu suivant. On lance deux fois de suite une pièce équilibrée et le joueur gagne autant de fois 100 € que de piles. Il possède un joker qu'il peut utiliser une fois au plus au cours du jeu, pour un coût de 40 €. Une telle utilisation force le prochain

lancer à être pile. Ecrire l'arbre de décision du jeu, compter les stratégies ; trouver la ou les stratégie(s) optimale(s) et le gain maximal.
Le corrigé se trouve à la fin des figures. Les stratégies optimales consistent à jouer son joker, soit au premier coup, soit au deuxième. (Il y a deux telles stratégies parmi les cinq stratégies équivalentes.)