Examen du module MODE

C. Gonzales / P. Perny

Durée : 3 heures

Exercice 1 (7 points — Processus décisionnels Markoviens –)

R^2 est un robot recycleur chargé de collecter des objets. Toutes les 10 mn il doit prendre une décision sur l'action qu'il va effectuer pendant la prochaine période (les 10 prochaines minutes), à choisir parmi les suivantes :
– rechercher activement des objets et les ramasser (action = search)
– rester immobile et attendre que quelqu'un lui donne des objets (action = wait)
– aller recharger sa batterie (action = recharge)
La meilleure façon pour R^2 d'obtenir des objets est de les chercher activement, mais cela consomme de l'energie ce qui n'est pas le cas s'il reste immobile. Lors d'une période de recherche active, la batterie peut se vider totalement, dans ce cas le robot R^2 doit s'éteindre et attendre un secours externe qui arrive en fin de période.
Pour simplifier on distinguera deux états distincts (*haut*, *bas*) concernant le niveau de la batterie au début d'une période et l'action *recharge* ne sera envisagée que quand le niveau de batterie est *bas*. Lorsque le niveau de batterie est *haut*, on supposera qu'une période de recherche active laisse la batterie au niveau *haut* avec une probablitité de $\alpha > 0$ et passe au niveau *bas* avec la probabilité $1 - \alpha$ (on considère que cette période ne peut vider

totalement la batterie). Lorsque le niveau de batterie est *bas*, on supposera qu'une période de recherche active laisse la batterie au niveau *bas* avec une probablitité $\beta > 0$ et vide la batterie avec une probabilité $1 - \beta$.
Chaque objet collecté par le robot induit une récompense de 1. On notera respectivement R_s et R_w les espérances du nombre d'objets collectés après une période de 10mn en mode *search* et en mode *wait* en supposant que ces quantités sont connues et vérifient $R_s > R_w \geq 1$. Si la batterie se vide au cours d'une période de recherche active, la récompense sur cette période prend la forme d'une pénalité (récompense de -3) car un secours externe est nécessaire. Précisons enfin que l'action *recharge* n'a pas d'impact sur les récompenses et recharge la batterie de manière certaine. L'action *wait* n'a aucun effet sur le niveau de la batterie.

Q 1.1 Modéliser ce problème comme un processus décisionnel Markovien à deux états. On prendra soin de préciser les actions possibles dans chaque état, les transitions possibles, leur probabilités et les récompenses associées dans un tableau de la forme suivante :

Etat	Action	Etat	Proba	Récompense
x	a	y	$T(x,a,y)$	$R(x,a,y)$
$\dots$			$\dots$	$\dots$

Q 1.2 Calculer les récompenses immédiates espérées $R(x,a)$ pour tout état x et pour toute action a envisageable dans l'état x.

Q 1.3 Soit π la politique qui consiste à chercher activement quand le niveau de la batterie est *haut* et à recharger quand le niveau est *bas*. Ecrire les équations de Bellman caractérisant la valeur de ces politiques pour un taux d'actualisation $0 < \gamma < 1$ (on notera respectivement h et l les valeurs de π dans les états $haut$ et bas).

Q 1.4 Soit π' la politique qui consiste à attendre en permanence ($wait$) sans jamais se recharger. Ecrire les équations de Bellman caractérisant la valeur de ces politiques pour un taux d'actualisation $0 < \gamma < 1$ (on notera respectivement h' et l' les valeurs de π' dans les états $haut$ et bas).

Application numérique : dans la suite de l'exercice on supposera que $\gamma = 0.5$, $\alpha = 0.8$, $\beta = 0.5$, $R_s = 5$, $R_w = 1$

Q 1.5 Evaluer les politiques π et π'. Laquelle préférez-vous ? pourquoi ?

Q 1.6 Ecrire les équations de Bellman qui caractérisent la politique optimale.

Q 1.7 Déterminer la politique optimale par une méthode de votre choix.

Exercice 2 (7 points — Arbres de décision et RDU –)

Un joueur est « largué » dans un labyrinthe dont il possède la carte. Dans ce dernier, des trésors de différentes valeurs ont été déposés et le jeu consiste à se déplacer de manière à trouver l'un des trésors. Dès que le joueur en a trouvé un, le jeu s'arrête. Au début, il ne sait pas exactement où il est, mais il peut observer que la pièce dans laquelle il est a la configuration de la figure 1.a, c'est-à-dire qu'il a une porte à gauche et une porte à droite. Dans ce cas, d'après son plan, il sait que s'il utilise la porte de gauche, il aura statistiquement 30% de chances de trouver un trésor de valeur 10 €, 40% de chances de trouver 70 € et 30% de trouver 50 €. De même, derrière la porte de droite, il a une chance sur deux qu'il y ait un couloir remontant vers le haut (figure 1.b) et une chance sur deux d'avoir un couloir vers le bas (figure 1.c). Dans le premier cas, on arrive obligatoirement à trois portes a, b, c, comme indiqué sur la figure 1.c. Derrière la porte de type a, le joueur a 30% de chances d'obtenir 200 € et 70% de chances d'obtenir 0 €, derrière la porte de type b, il a 50% de chances d'obtenir 150 € et 50% de chances d'obtenir 50 €, et derrière la porte de type c, il a 70% de chances d'obtenir 130 € et 30% de chances d'obtenir 100 €. Le couloir descendant, quant à lui, a deux couloirs adjacents (de types d et e comme indiqué sur la figure 1.b). Si le joueur suit le couloir d, il aura 70% de chances d'obtenir 100 € et 30% de chances d'avoir 40 €, alors que s'il suit le couloir e, il aura 20% de chances d'obtenir 130 € et 80% de chances d'obtenir 60 €. Dans la pièce de la figure 1.a, il n'est autorisé à ouvrir qu'une seule porte.

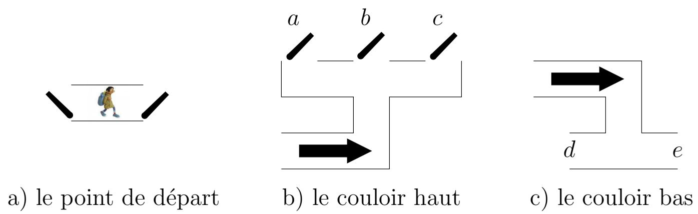

FIG. 1 – Le labyrinthe dans lequel se trouve le joueur.

Q 2.1 Tracez l'arbre de décision correspondant à ce problème.

Q 2.2 Déterminez la stratégie optimale du décideur à la racine de l'arbre si celui-ci a pour critère RDU, avec, pour fonction d'utilité sur l'espace des conséquences la fonction $f(x) = x$, et pour déformation de probabilité la fonction φ définie par :

x	0.15	0.20	0.25	0.30	0.35	0.45	0.50	0.60	0.65	0.70	0.75	0.85	1
$\varphi(x)$	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.87	0.90	1

Q 2.3 Considérons un espace de conséquences $\mathcal{X} = \{x_1, \ldots, x_n\}$, où $x_1 < x_2 < \cdots < x_n$. Les loteries sur $\mathcal{X}$ sont des tuples $\langle x_1, p_1; x_2, p_2; \ldots; x_n, p_n\rangle$, où p_i représente la probabilité d'obtenir la conséquence x_i et la somme des p_i est égale à 1. Soit deux loteries $P = \langle x_1, p_1; x_2, p_2; \ldots; x_n, p_n\rangle$ et $Q = \langle x_1, q_1; x_2, q_2; \ldots; x_n, q_n\rangle$ sur $\mathcal{X}$. On dit que P domine stochastiquement Q à l'ordre 1, ce que l'on note $P \succsim_{DS} Q$ si pour tout $i \in \{1, \ldots, n\}$, $\sum_{j \geq i} p_j \geq \sum_{j \geq i} q_j$.
Montrez que si $P \succsim_{DS} Q$, alors, pour toute loterie R sur $\mathcal{X}$, et tout $\lambda \in [0, 1]$, $\lambda P + (1 - \lambda)R \succsim_{DS} \lambda Q + (1 - \lambda)R$.

Q 2.4 Dans un arbre de décision, à chaque stratégie, l'on peut associer une loterie sur l'espace des résultats. On dira qu'une stratégie d_1 domine stochastiquement une stratégie d_2 à l'ordre 1 si la loterie associée à d_1 domine stochastiquement celle associée à d_2. En déduire un algorithme efficace pour calculer l'ensemble des stratégies stochastiquement non dominées à l'ordre 1 dans un arbre de décision.

Q 2.5 Déterminez l'ensemble des stratégies non stochastiquement dominées à la racine de l'arbre de décision obtenu à la question 1.

Q 2.6 La stratégie déterminée à la question 2 fait-elle partie de l'ensemble obtenu à la question précédente. Expliquez pourquoi.

Exercice 3 (5 points — Possibilités et probabilités –)

Soit $S = \{s_1, \ldots, s_n\}$ un ensemble fini d'états de la nature et Π une mesure de possibilité normalisée ($\Pi(S) = 1$) définie sur les sous-ensembles de S.

Q 3.1 Montrer qu'il existe une distribution de probabilité p sur S telle que : $\forall A \subseteq S, P(A) \leq \Pi(A)$ où $P(A) = \sum_{s \in A} p(s)$.

Q 3.2 Soit C_Π l'ensemble des distributions de probabilités sur S telles que $\forall A \subseteq S, P(A) \leq \Pi(A)$. Soit N la mesure de nécessité associée à Π. Montrer que toute distribution p de C_Π est telle que $\forall A \subseteq S, N(A) \leq P(A)$ où $P(A) = \sum_{s \in A} p(s)$.

Q 3.3 Dans le cas où S contient trois états et que Π est définie par :

A	$\emptyset$	$\{s_1\}$	$\{s_2\}$	$\{s_3\}$	$\{s_1, s_2\}$	$\{s_2, s_3\}$	$\{s_1, s_3\}$	S
$\Pi(A)$	0	0.4	1	0.2	1	1	0.4	1

Expliciter la mesure de nécessité associée à Π et donner deux distributions distinctes de C_Π.

Q 3.4 Etant donné une mesure de possibilité Π et deux éléments distincts p et p' de C_Π, est-il possible de trouver deux actes $f : S \to [0,1]$ et $g : S \to [0,1]$ tels que $EU_p(f) > EU_p(g)$ alors que $EU_{p'}(f) \leq EU_{p'}(g)$? (donner un exemple ou montrer que c'est impossible).

Note : la notation EU_p désigne le critère EU utilisé avec la distribution p.

Q 3.5 Etant donné une mesure de possibilité Π et un élément p de C_Π, est-il possible de trouver deux actes $f : S \to [0,1]$ et $g : S \to [0,1]$ tels que $EU_p(f) > EU_p(g)$ alors que $U_\pi^+(f) \leq U_\pi^+(g)$? (donner un exemple ou montrer que c'est impossible).

Note : la notation U_π^+ désigne le critère de l'utilité qualitative optimiste utilisé avec la distribution de possibilité π sur S associée à la mesure Π.

Exercice 4 (3 points –)

Un entrepreneur fabrique un produit dont la demande est incertaine. Elle peut être élevée (E), normale (N) ou bien encore insuffisante (I). L'information sur les probabilités des événements élémentaires E,N et I est que :

$$1/4 \leq P(E) \leq 1/2 \qquad 1/8 \leq P(N) \leq 1/2 \qquad 3/4 \leq P(E \cup I) \qquad P(I) \leq 1/2. \quad (1)$$

Q 4.1 Soit $\mathcal{P}$ l'ensemble des lois de probabilité sur $2^{\mathcal{X}}$, où $\mathcal{X} = \{E, N, I\}$, compatibles avec l'équation (1). Montrez que $\mathcal{P}$ est non vide.

Q 4.2 Déterminez $f = \inf\{P : P \in \mathcal{P}\}$.

Q 4.3 Déterminez l'inverse de Möbius de f.

Q 4.4 f est-elle une fonction de croyance ?