Informatique Mathématique
une photographie en 2015

X et Y et Z et ... (Eds.)

3 mars 2015

Table des matières

Table des figures

Liste des tableaux

Chapitre 5

Contrôle des modèles probabilistes partiellement observables

Nathalie Bertrand
Serge Haddad

De nombreux domaines d'application nécessitent l'analyse de modèles probabilistes partiellement observables. Considérons par exemple le diagnostic de dysfonctionnements dans les réseaux de télécommunications. Le système chargé de la surveillance d'un tel réseau n'a accès qu'à une connaissance partielle de l'état de chacun des composants, et les occurrences de pannes sont naturellement modélisées par une distribution de probabilité obtenue par des observations statistiques. De façon analogue, la gestion du trafic ferroviaire repose sur des capteurs de position des trains, et les informations qu'ils remontent doivent être interprétées pour contrôler au mieux la vitesse et l'espacement des trains. La modélisation de telles applications nécessite un cadre qui combine trois aspects importants : contrôle, probabilités, et observation partielle.

L'objectif de ce chapitre est de passer en revue les problèmes de contrôle pour plusieurs modèles formels qui combinent probabilités et observation partielle.

*Dans un premier temps, on s'intéressera aux automates probabilistes (*PA*), où le contrôleur, dépourvu d'observations, doit interagir avec un environnement afin d'atteindre un objectif avec une probabilité maximale. Nous étudierons d'abord les automates probabilistes du point de vue de la théorie des langages en établissant comment leur expressivité se compare à celle de*

familles classiques de langages. On présentera également leurs propriétés de clôture. Puis, on détaillera les problèmes de décision relatifs au contrôle des automates probabilistes. En particulier, on montrera que l'équivalence est décidable pour les automates probabilistes, mais que l'existence d'un contrôleur garantissant une probabilité supérieure à un seuil fixé est un problème indécidable. D'autres résultats plus récents d'indécidabilité seront aussi présentés.

*Nous introduirons ensuite les processus de décision markoviens partiellement observables (*POMDP*). Ce modèle généralise celui des* PA *puisque le contrôleur a accès au cours de l'exécution du système à une information partielle sur l'état courant. Les* POMDP *étendent également les processus de décision markoviens* MDP *pour lesquels l'observation est parfaite. On développera un algorithme de synthèse d'une politique optimale de contrôle à horizon fini, c'est-à-dire lorsque le nombre d'étapes est fini et fixé à l'avance. Puis, on étudiera des problèmes d'optimisation à horizon infini, en détaillant pour des propriétés* qualitatives *(c.-à-d. où on requiert une probabibilité positive ou égale à* 1 *d'atteindre un objectif), les résultats de décidabilité et de complexité. D'un point de vue applicatif, on montrera comment les* POMDP *permettent de spécifier et d'analyser le problème du diagnostic actif de panne, correspondant à la recherche d'un contrôleur garantissant la diagnosticabilité du système.*

Enfin, nous évoquerons le cas plus général des jeux stochastiques à signaux. Ce modèle classique en théorie des jeux étend les modèles précédents en exprimant l'interaction de deux joueurs partiellement informés dans un environnement probabiliste. Après avoir établi pour quels objectifs de gain ces jeux sont déterminés *(c.-à-d. que l'un des deux joueurs a une stratégie gagnante), nous présenterons les principaux résultats de décidabilité et d'indécidabilité pour les jeux stochastiques à signaux.*

5.1 Automates probabilistes

5.1.1 Présentation

Les automates probabilistes ont été introduits pour spécifier de nouvelles familles de langages formels [26]. Ils se différencient des automates non déterministes de la façon suivante : étant donné un caractère lu, l'état suivant est choisi de manière probabiliste. Ainsi après avoir lu un mot dans l'automate, chaque état a une probabilité d'être l'état atteint. En cumulant la probabilité des états finals, on obtient une valeur qu'on peut comparer avec un seuil pour déterminer si le mot appartient au langage spécifié par l'automate.

Un autre point de vue sur ces automates est lié à des problèmes de planification, et nous allons l'illustrer maintenant. Supposez que vous planifiez des vacances à l'étranger. Vous devez choisir avec quel train ou avion voyager, quelle maison ou chambre d'hôtel louer, quelles excursions entreprendre ou expositions visiter, etc. Chacune de ces actions peut être modélisée par un caractère d'un alphabet particulier et l'effet de ces actions étant incertain, le changement d'état déclenché par une action est naturellement spécifié par une distribution. Votre objectif est de maximiser la probabilité de réussir vos vacances. Par exemple, les états pourraient être : `départ`, `aéroport`, `hôtel`, `succès`, `échec`. Dans l'état `départ` vous avez le choix entre deux compagnies d'avion `lowcost` ou `highcost`. La compagnie `lowcost` vous conduit à l'état `aéroport` (respectivement `échec` en cas d'annulation de vol) avec probabilité 0.9 (respectivement 0.1). Une autre distribution est associée avec `highcost`. La réservation de l'hôtel peut être effectuée par `internet` ou `téléphone`. Une réservation `internet` vous conduit de l'état `aéroport` à l'état `hôtel` (respectivement `échec` en cas d'annulation de réservation) avec probabilité 0.7 (respectivement 0.3). Pour le choix des excursions, vous avez affaire à deux agences `toutvoir` et `nerienrater`. Le mot `lowcost.internet.toutvoir` est un plan possible et sa probabilité de réussite est la probabilité de terminer dans l'état `succès`. On peut alors chercher un mot qui maximise cette probabilité ou, plus modestement, un mot dont la probabilité est supérieure à un seuil.

Afin de définir les automates probabilistes (PA), nous adoptons la terminologie de la théorie des langages formels. Rappelons qu'étant donné un ensemble fini ou dénombrable E, une distribution sur E est une fonction π de E dans $\mathbb{R}_{\geq 0}$ telle que $\sum_{e \in E} \pi(e) = 1$, et on notera $\mathsf{Dist}(E)$ l'ensemble des distributions sur E.

Définition 5.1.1 (PA). *Un* automate probabiliste $\mathcal{A} = (Q, A, \{\mathbf{P}_a\}_{a \in A}, \pi_0, F)$ *est défini par :*

- *Q, l'ensemble fini des états ;*
- *A, l'alphabet fini ;*
- *pour tout $a \in A$, $\mathbf{P}_a$, une matrice stochastique indicée par Q, c.-à-d. telle que $\forall q \in Q \ \sum_{q' \in Q} \mathbf{P}_a[q, q'] = 1$ et $\forall q, q' \in Q \ \mathbf{P}_a[q, q'] \geq 0$;*
- *π_0, la distribution initiale des états et $F \subseteq Q$ l'ensemble des états finals.*

Lorsque la distribution initiale est une distribution de Dirac concentrée en q_0, on dit que q_0 est l'état initial.

Exemple 5.1.2 (Un exemple de PA). *La figure 5.1 décrit un* PA *avec un état initial q_0 et un état final q_1. À des fins de concision, un arc d'un état à un autre est*

étiqueté par un vecteur de probabilités de transition indicé par les caractères. Pour rendre ces indices explicites, le vecteur est représenté par une somme formelle. Par exemple, la transition qui boucle autour de q_0 est étiquetée par $1a + \frac{1}{2}b$ signifiant que lorsque a (respectivement b) est lue dans q_0, la probabilité que l'état suivant soit q_0, $\mathbf{P}_a[q_0, q_0]$ (respectivement $\mathbf{P}_b[q_0, q_0]$), est égale à 1 (respectivement $\frac{1}{2}$).

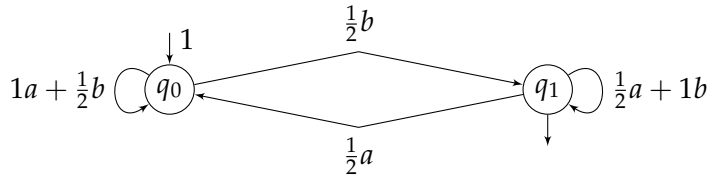


FIGURE *5.1 – Un exemple de* PA.

La définition suivante formalise la probabilité d'acceptation d'un mot, c'est-à-dire la probabilité d'atteindre un état final après avoir lu ce mot lorsque l'état initial est distribué selon π_0.

Définition 5.1.3. *Soit $\mathcal{A}$ un* PA *et $w = a_1 \dots a_n \in A^*$ un mot, la* probabilité d'acceptation *de w par $\mathcal{A}$ est définie par :*

$$\mathbf{Pr}_{\mathcal{A}}(w) \stackrel{def}{=} \sum_{q \in Q} \pi_0[q] \sum_{q' \in F} \left(\prod_{i=1}^{n} \mathbf{P}_{a_i} \right) [q, q'].$$

Notation. Etant donné un mot $w = a_1 \dots a_n$, la matrice stochatique $\mathbf{P}_w$ est définie par $\mathbf{P}_w \stackrel{\text{def}}{=} \prod_{i=1}^{n} \mathbf{P}_{a_i}$. En particulier $\mathbf{P}_\varepsilon = \mathbf{Id}$. Ainsi, $\mathbf{Pr}_{\mathcal{A}}(w) = \pi_0 \mathbf{P}_w \mathbf{1}_F^T$ où $\mathbf{1}_F$ est le vecteur indicateur du sous-ensemble F (et l'exposant T dénote la transposition).

Calculons $\mathbf{Pr}_{\mathcal{A}}(abba)$ dans l'exemple 5.1.2. Puisqu'il n'y a que deux états, il nous suffit de calculer successivement la probabilité d'acceptation des préfixes de *abba*. Pour le mot vide ε, $\mathbf{Pr}_{\mathcal{A}}(\varepsilon) = 0$. Puis :

— $\mathbf{Pr}_{\mathcal{A}}(a) = \frac{1}{2}\mathbf{Pr}_{\mathcal{A}}(\varepsilon) = 0$,
— $\mathbf{Pr}_{\mathcal{A}}(ab) = \mathbf{Pr}_{\mathcal{A}}(a) + \frac{1}{2}(1 - \mathbf{Pr}_{\mathcal{A}}(a)) = \frac{1}{2}$
— $\mathbf{Pr}_{\mathcal{A}}(abb) = \mathbf{Pr}_{\mathcal{A}}(ab) + \frac{1}{2}(1 - \mathbf{Pr}_{\mathcal{A}}(ab)) = \frac{3}{4}$
— $\mathbf{Pr}_{\mathcal{A}}(abba) = \frac{1}{2}\mathbf{Pr}_{\mathcal{A}}(abb) = \frac{3}{8}$

Plus généralement, on obtient les équations de récurrence suivantes :

$$\mathbf{Pr}_{\mathcal{A}}(wa) = \frac{1}{2}\mathbf{Pr}_{\mathcal{A}}(w) \text{ et } \mathbf{Pr}_{\mathcal{A}}(wb) = \frac{1}{2}(1 + \mathbf{Pr}_{\mathcal{A}}(w))$$

Justifions par exemple la deuxième équation. La probabilité d'atteindre q_1 après avoir lu wb est égale à la probabilité d'atteindre q_0 après avoir lu w et de passer de q_0 à q_1 par l'action b plus la probabilité d'atteindre q_1 après avoir lu w et de rester en q_1 par l'action b. En appliquant les résultats élémentaires sur les probabilités conditionnelles la première probabilité est égale à $\frac{1}{2}(1 - \mathbf{Pr}_{\mathcal{A}}(w))$ car $1 - \mathbf{Pr}_{\mathcal{A}}(w)$ est la probabilité d'atteindre q_0 après avoir lu w. La deuxième probabilité est égale à $\mathbf{Pr}_{\mathcal{A}}(w)$, d'où le résultat.

À partir de ces équations, on déduit une expression explicite de la probabilité d'acceptation :

$$\mathbf{Pr}_{\mathcal{A}}(a_1 \dots a_n) = \sum_{i=1}^{n} 2^{i-1-n} \cdot \mathbf{1}_{a_i=b}.$$

Observez que $\sup_{w \in A^*} \mathbf{Pr}_{\mathcal{A}}(w) = 1$ et que cette valeur n'est atteinte par aucun mot fini.

Étant donné un PA $\mathcal{A} = (Q, A, \{\mathbf{P}_a\}_{a \in A}, \pi_0, F)$ et un mot fini w, $\mathcal{A}$ se comporte comme une chaîne de Markov à temps discret (DTMC) $\mathcal{M}_{\mathcal{A}}^{w} = (S, \pi_0^S, \mathbf{P})$. Cette interprétation a encore plus de sens pour les mots infinis, comme nous le verrons par la suite. Si $w \stackrel{\text{def}}{=} w_1 \cdots w_n$, l'ensemble des états de cette DTMC est $S \stackrel{\text{def}}{=} Q \times [0, n]$; la distribution initiale est donnée par celle du PA : $\pi_0^S((q, 0)) = \pi_0(q)$ (et π_0^S est nulle partout ailleurs) ; et la matrice de transition est définie pour tout $q, q' \in Q$ et $i < n$ par $\mathbf{P}[(q, i), (q', i+1)] = \mathbf{P}_{w_{i+1}}[q, q']$, $\mathbf{P}[(q, n), (q, n)] = 1$ (les états correspondant à l'instant n sont absorbants) et $\mathbf{P}$ est nulle pour tous les autres coefficients.

Les *languages stochastiques* sont alors obtenus à l'aide de seuils et d'opérateurs de comparaison.

Définition 5.1.4. *Soit $\mathcal{A}$ un automate probabiliste, $\theta \in [0, 1]$ un* seuil *et $\bowtie \in \{<, \leq, >, \geq, =, \neq\}$ un opérateur de comparaison. Alors le langage $L_{\bowtie\theta}(\mathcal{A})$ est défini par :*

$$L_{\bowtie\theta}(\mathcal{A}) = \{w \in A^* \mid \mathbf{Pr}_{\mathcal{A}}(w) \bowtie \theta\}.$$

Les langages stochastiques de l'exemple 5.1.2 ont une interprétation naturelle. En définissant $v_a = 0$ et $v_b = 1$, la probabilité d'acceptation d'un mot $w = w_1 \dots w_n$ est alors le nombre binaire $0.v_{w_n} \dots v_{w_1}$. En particulier, $L_{\geq \frac{1}{2}}(\mathcal{A})$ est l'ensemble des miroirs des représentations des nombres binaires au moins égaux à $\frac{1}{2}$. Observez que la représentation d'un nombre n'est pas unique en raison de l'ajout de zéros en début du mot (c.-à-d. à la fin de la représentation).

Exemple 5.1.5 (Un PA qui compte)**.** *La figure 5.2 décrit un* PA *sur l'alphabet* $\{a, b\}$*. Il s'agit d'une représentation succincte où nous avons omis un état absorbant non final et toutes les transitions vers cet état. Par exemple, à partir de* q_0 *(respectivement* q_2*) en lisant* b *l'automate va vers l'état puits avec probabilité 1 (respectivement* $\frac{1}{2}$*).*

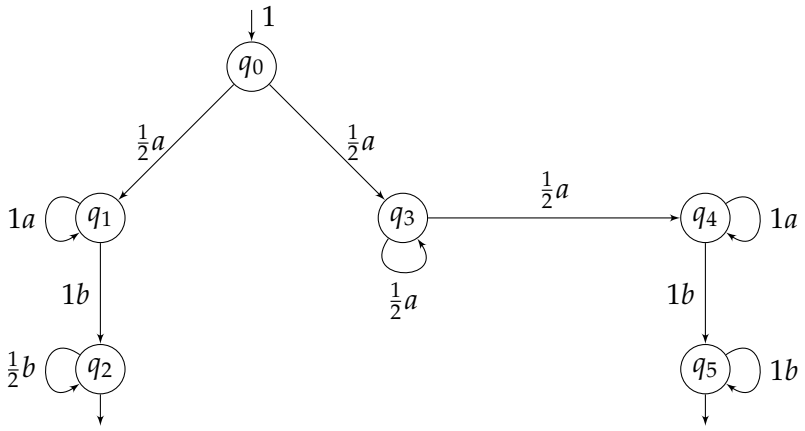


FIGURE *5.2 – Un* PA $\mathcal{A}$ *tel que* $L_{=\frac{1}{2}}(\mathcal{A}) = \{a^n b^n \mid n > 0\}$.

Tout mot w *différent de* $a^m b^n$ *avec* $m > 0, n > 0$ *ne peut être accepté et par conséquent,* $\mathbf{Pr}_{\mathcal{A}}(w) = 0$*. Soit donc* $w = a^m b^n$ *avec* $m > 0, n > 0$*. Le mot* w *peut être accepté par un chemin* $q_0 q_1^m q_2^n$ *ou par la famille de chemins* $q_0 q_3^r q_4^s q_5^n$ *avec* $0 < r, s$ *et* $r + s = m$*. La probabilité du premier chemin est* $\frac{1}{2} 1^{m+1} (\frac{1}{2})^{n-1} = \frac{1}{2^n}$ *tandis que la somme des probabilités des chemins de la famille est* $\sum_{r=2}^{m} \frac{1}{2^r} = \frac{1}{2} - \frac{1}{2^m}$*. En sommant, on obtient* $\mathbf{Pr}_{\mathcal{A}}(w) = \frac{1}{2} + \frac{1}{2^n} - \frac{1}{2^m}$*. Par conséquent,* $L_{=\frac{1}{2}}(\mathcal{A}) = \{a^n b^n \mid n > 0\}$.

Afin de pouvoir manipuler des PA de façon effective, nous devons imposer des restrictions sur leur définition, conduisant à la sous-classe suivante de PA.

Définition 5.1.6. *Un* PA *est dit* rationnel *si ses distributions de probabilité appartiennent à* $\mathbb{Q}^Q$ *: c.-à-d. pour tout* $a \in A$ *et tous* $s, s' \in Q$*,* $\mathbf{P}_a[s, s'] \in \mathbb{Q}$ *. Un* langage stochastique rationnel *est un langage stochastique spécifié par un* PA *rationnel avec un seuil rationnel.*

L'exemple 5.1.2 montre que $\{a^n b^n \mid n > 0\}$ est un langage stochastique rationnel.

5.1.2 Propriétés des langages stochastiques

Expressivité

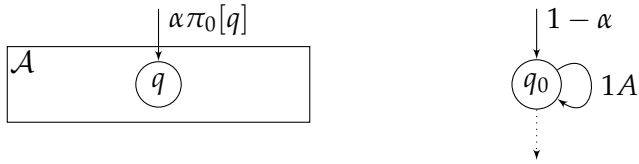
Une première question à traiter est liée à la définition des langages stochastiques : pouvons-nous restreindre les seuils possibles et les opérateurs de comparaison en conservant le même pouvoir d'expression ? Nous démontrons d'abord que nous pouvons toujours utiliser un seuil $\frac{1}{2}$ (ou toute valeur arbitraire strictement comprise entre 0 et 1).

Proposition 5.1.7. *Soit $\mathcal{A}$ un* PA*, $\theta \in [0,1]$ un seuil et $\bowtie \in \{<, \leq, >, \geq, =, \neq\}$ un opérateur de comparaison. Alors il existe un* PA *$\mathcal{A}'$ tel que :*

$$L_{\bowtie \frac{1}{2}}(\mathcal{A}') = L_{\bowtie \theta}(\mathcal{A}).$$

De plus si $\mathcal{A}$ est un PA *rationnel et $\theta \in \mathbb{Q}$ alors $\mathcal{A}'$ est également un* PA *rationnel.*

Démonstration. Étant donnés $\mathcal{A}$ et $\theta \neq \frac{1}{2}$, on construit $\mathcal{A}'$ comme représenté ci-dessous.



On ajoute un état q_0 qui est absorbant quelle que soit la lettre choisie. La distribution initiale est modifiée ainsi : $\pi'_0[q_0] = 1 - \alpha$ et pour tout $q \in Q$, $\pi'_0[q] = \alpha\pi_0[q]$ où $0 < \alpha < 1$. Le fait que q_0 soit final ou non, et la valeur α dépendent de θ de la façon suivante :

— Si $\theta > \frac{1}{2}$ alors $q_0 \notin F$ et $\alpha \stackrel{\text{def}}{=} \frac{1}{2\theta}$ de telle sorte que pour $w \in A^*$, $\mathbf{Pr}_{\mathcal{A}'}(w) = \frac{1}{2\theta}\mathbf{Pr}_{\mathcal{A}}(w)$. Ainsi, $w \in L_{\bowtie \frac{1}{2}}(\mathcal{A}')$ ssi $w \in L_{\bowtie \theta}(\mathcal{A})$.

— Si $\theta < \frac{1}{2}$ alors $q_0 \in F$ et $\alpha \stackrel{\text{def}}{=} \frac{1}{2(1-\theta)}$ de telle sorte que pour $w \in A^*$, $\mathbf{Pr}_{\mathcal{A}'}(w) = \frac{1-2\theta+\mathbf{Pr}_{\mathcal{A}}(w)}{2(1-\theta)}$. Ainsi, $w \in L_{\bowtie \frac{1}{2}}(\mathcal{A}')$ ssi $w \in L_{\bowtie \theta}(\mathcal{A})$.

Dans les deux cas on a donc bien $L_{\bowtie \frac{1}{2}}(\mathcal{A}') = L_{\bowtie \theta}(\mathcal{A})$. □

Nous établissons maintenant que les opérateurs d'égalité et de différence peuvent être omis.

Proposition 5.1.8. *Pour tout* PA *$\mathcal{A}$ il existe un* PA *$\mathcal{A}'$ tel que :*

$$L_{=\frac{1}{4}}(\mathcal{A}') = L_{\geq \frac{1}{4}}(\mathcal{A}') = L_{=\frac{1}{2}}(\mathcal{A}) \text{ et donc } L_{<\frac{1}{4}}(\mathcal{A}') = L_{\neq \frac{1}{2}}(\mathcal{A}).$$

De plus si $\mathcal{A}$ est un PA *rationnel et $\theta \in \mathbb{Q}$ alors $\mathcal{A}'$ est également un* PA *rationnel.*

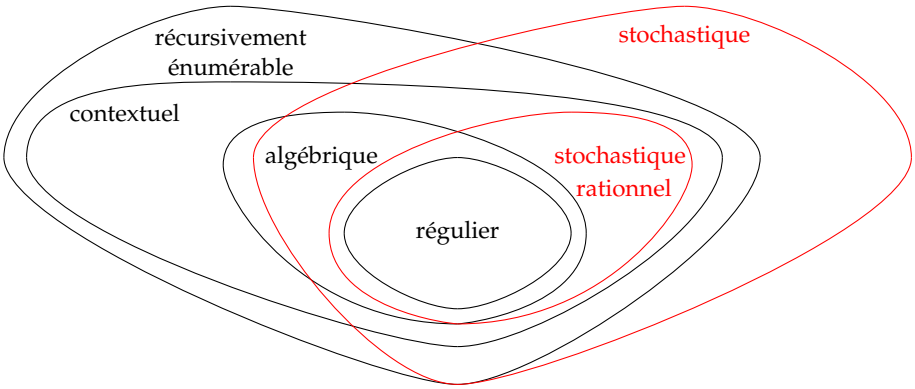


FIGURE 5.3 – *La hiérarchie de Chomsky complétée.*

Démonstration. On construit $\mathcal{A}' = (Q', A, \{\mathbf{P}'_a\}_{a\in A}, \pi'_0, F')$ ainsi :
— $Q' = Q \times Q$;
— $\mathbf{P}'_a[(q_1, q_2), (q'_1, q'_2)] = \mathbf{P}_a[q_1, q'_1]\mathbf{P}_a[q_2, q'_2]$;
— $\pi'_0[q_1, q_2] = \pi_0[q_1]\pi_0[q_2]$ et $F' = F \times (Q \setminus F)$.
Lorsqu'un mot w est lu, les deux composantes de l'automate se comportent de manière indépendante et par conséquent : $\mathbf{Pr}_{\mathcal{A}'}(w) = \mathbf{Pr}_{\mathcal{A}}(w)(1 - \mathbf{Pr}_{\mathcal{A}}(w))$. Ce qui entraîne $\mathbf{Pr}_{\mathcal{A}'}(w) \leq \frac{1}{4}$ quel que soit le mot w, et $\mathbf{Pr}_{\mathcal{A}'}(w) = \frac{1}{4}$ si et seulement si $\mathbf{Pr}_{\mathcal{A}}(w) = \frac{1}{2}$. □

Observons maintenant que l'on peut restreindre les opérateurs de comparaison utilisés pour définir les langages stochastiques. En effet, en complémentant les états finals, l'opérateur $\leq$ (respectivement $<$) peut être simulé par l'opérateur $>$ (respectivement $\geq$).

Proposition 5.1.9. *Soit $\mathcal{A}$ un* PA *et $\mathcal{A}'$ défini comme $\mathcal{A}$ excepté que $F' = Q \setminus F$. Alors :*

$$L_{\geq\theta}(\mathcal{A}') = L_{<\theta}(\mathcal{A}) \text{ et } L_{>\theta}(\mathcal{A}') = L_{\leq\theta}(\mathcal{A}).$$

La deuxième question d'expressivité que nous traitons est liée à la comparaison de la famille des langages stochastiques avec les familles classiques de langages, que l'on rappelle brièvement. Les langages récursivement énumérables sont ceux engendrés par une grammaire sans restriction sur la nature des règles ou, de manière équivalente, ceux reconnus par une machine de Turing. Les langages contextuels sont ceux engendrés par une grammaire dont le membre gauche de toute règle est au plus aussi long que le membre droit de la règle (à l'exception d'une unique règle pour générer le mot vide, s'il appartient au langage). Les langages algébriques sont ceux

engendrés par une grammaire dont le membre gauche de toute règle est un unique symbole. Les langages réguliers sont ceux engendrés par une grammaire dont le membre gauche de toute règle est un unique symbole et le membre droit a au plus deux symboles dont le premier est un caractère. Le théorème de Kleene établit que les langages réguliers sont aussi ceux reconnus par automate fini (que l'on peut choisir déterministe et complet). La figure 5.3 illustre le positionnement des langages stochastiques et des langages stochastiques rationnels vis à vis de ces familles de langages. Dans ce qui suit, nous détaillons ces comparaisons.

Considérons les langages $\{L_{>\theta}(\mathcal{A}) \mid 0 \leq \theta \leq 1\}$ où $\mathcal{A}$ est le PA présenté dans l'exemple 5.1.2. Étant donnés $\theta < \theta'$ il existe un nombre binaire b tel que $\theta < b < \theta'$, ce qui entraîne $L_{>\theta'}(\mathcal{A}) \subsetneq L_{>\theta}(\mathcal{A})$. Ainsi, la famille des langages $\{L_{>\theta}(\mathcal{A}) \mid 0 \leq \theta \leq 1\}$ n'est pas dénombrable, contrairement à la famille des langages récursivement énumerables. On en déduit :

Proposition 5.1.10. *Il existe un langage stochastique non récursivement énumerable.*

Un automate fini déterministe complet (de fonction de transition δ) est un cas particulier d'automate probabiliste dont la distribution associée à la paire (s, a) est une distribution de Dirac de support l'état $\delta(s, a)$. Ainsi, la famille des langages réguliers est incluse dans la famille des langages stochastiques rationnels. Il est facile de démontrer l'inclusion stricte puisque le langage de l'exemple 5.1.5 n'est pas régulier.

Proposition 5.1.11. *La famille des langages réguliers est strictement incluse dans la famille des langages stochastiques rationnels.*

Les deux propositions suivantes établissent que les langages algébriques et les langages stochastiques (rationnels) sont incomparables.

Proposition 5.1.12 ([23])**.** *Il existe un langage algébrique qui n'est pas un langage stochastique. Plus précisément* $L = \{a^{n_1}ba^{n_2}b \dots a^{n_k}ba^* \mid k \geq 2 \wedge \forall i,\ n_i \geq 0 \wedge \exists i > 1 :\ n_i = n_1\}$ *est un langage algébrique qui n'est pas stochastique.*

Démonstration. Soit $L = \{a^{n_1}ba^{n_2}b \dots a^{n_k}ba^* \mid \exists i > 1\ n_i = n_1\}$. L est algébrique car reconnaissable par un automate non déterministe à pile (voir par exemple [17]). En effet avec un compteur (c.-à.-d une pile sur un alphabet unaire), on compte n_1 le nombre de a jusqu'à la première occurrence de b. Puis on devine une occurrence de b et on décrémente le compteur par les occurrences de a jusqu'à la prochaine occurrence de b. Si le compteur est nul alors le mot est accepté.

Supposons que $L = L_{>\theta}(\mathcal{A})$ ou $L = L_{\geq\theta}(\mathcal{A})$ pour $\mathcal{A}$ un PA. Soit $\sum_{i=0}^{n} c_i x^i$ le polynôme minimal de la matrice $\mathbf{P}_a$. Puisque 1 est une valeur propre de $\mathbf{P}_a$, on obtient $\sum_{i=0}^{n} c_i = 0$ ce qui entraîne l'existence de coefficients strictement positifs et négatifs.

Par définition, $\sum_{i=0}^{n} c_i \mathbf{P}_{a^i} = 0$ et donc pour tout mot w, $\sum_{i=0}^{n} c_i \mathbf{P}_{a^i w} = 0$. Si $c_{i_1}, \ldots, c_{i_k}$ sont les coefficients strictement positifs du polynôme, choisissons $w = ba^{i_1}b \ldots ba^{i_k}b$, et soit $0 \leq i \leq n$.

Si $L = L_{>\theta}(\mathcal{A})$, par définition de L, $\pi_0 \mathbf{P}_{a^i w} \mathbf{1}_F^T > \theta$ ssi $i \in \{i_1, \ldots, i_k\}$. Donc : $0 = \sum_{i=0}^{n} c_i \pi_0 \mathbf{P}_{a^i w} \mathbf{1}_F^T > (\sum_{i=0}^{n} c_i)\theta = 0$, une contradiction.

Si maintenant $L = L_{\geq\theta}(\mathcal{A})$, on a $\pi_0 \mathbf{P}_{a^i w} \mathbf{1}_F^T \geq \theta$ ssi $i \in \{i_1, \ldots, i_k\}$. Donc : $0 = \sum_{i=0}^{n} c_i \pi_0 \mathbf{P}_{a^i w} \mathbf{1}_F^T > (\sum_{i=0}^{n} c_i)\theta = 0$, une contradiction.

Ainsi, L n'est pas un langage stochastique. □

Proposition 5.1.13. *Il existe un langage stochastique rationnel qui n'est pas algébrique. Plus précisément $L = \{a^n b^n c^n \mid n > 0\}$ est un langage stochastique rationnel qui n'est pas algébrique.*

Démonstration. En utilisant le lemme d'Ogden on prouve facilement que $L = \{a^n b^n c^n \mid n > 0\}$ n'est pas algébrique (voir par exemple [17]).

Observons que $L = L_1 \cap L_2$ avec $L_1 = \{a^n b^n c^+ \mid n > 0\}$ et $L_2 = \{a^+ b^n c^n \mid n > 0\}$. En utilisant la propriété de clôture énoncée par la proposition 5.1.20, il suffit de prouver que chaque L_i est un langage stochastique rationnel, tel que $L_i = L_{=\frac{1}{2}}(\mathcal{A}_i)$ avec $\mathcal{A}_i$ un PA rationnel. Ces deux automates sont des variantes de ceux de l'exemple 5.1.5. La figure 5.4 décrit $\mathcal{A}_1$ et nous laissons au lecteur le soin de spécifier $\mathcal{A}_2$. □

Observons que le problème du mot est décidable pour la famille des langages stochastiques rationnels. Étant donné un PA rationnel, un seuil rationnel, un opérateur de comparaison et un mot, on peut décider si le mot appartient au langage stochastique rationnel ainsi défini. En analysant la complexité d'une telle procédure, on obtient le résultat suivant.

Proposition 5.1.14 ([11])**.** *La famille des langages stochastiques rationnels est strictement incluse dans la famille des langages contextuels.*

Démonstration. Les langages contextuels sont exactement les langages pour lesquels le problème du mot peut être décidé par une procédure non déterministe en espace linéaire [18].

Nous démontrons que l'appartenance d'un mot à un langage stochastique rationnel peut être décidée par une procédure déterministe en espace linéaire. Cette complexité est loin d'être optimale [19], mais suffit pour

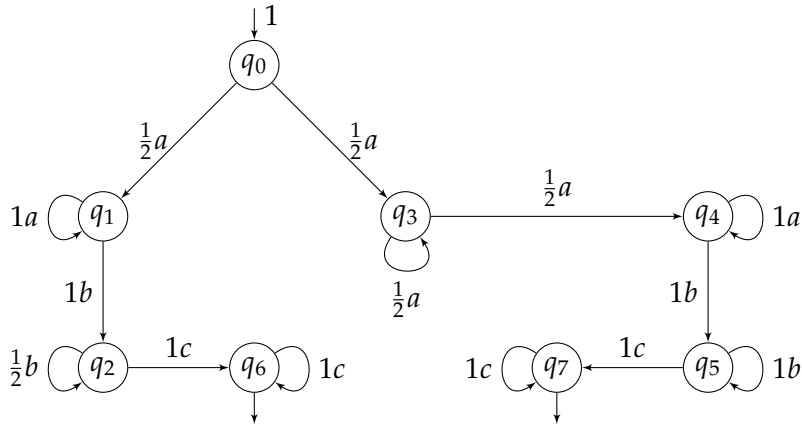


FIGURE 5.4 – *Un* PA $\mathcal{A}$ *tel que* $L_{=\frac{1}{2}}(\mathcal{A}) = \{a^n b^n c^+ \mid n > 0\}$.

prouver que les langages stochastiques rationnels sont contextuels. Tout d'abord on calcule le plus petit commun multiple, disons b, des dénominateurs de θ, des éléments des matrices $\{\mathbf{P}_a\}_{a\in A}$ et des éléments du vecteur π_0. Ceci se fait en espace constant (c.-à-d. indépendant de n, la taille du mot $w = a_1 \dots a_n$ à tester). Puis on construit les matrices entières $\mathbf{P}'_a = b\mathbf{P}_a$ et le vecteur entier $\pi'_0 = b\pi_0$ de nouveau en espace constant.

Le problème de l'appartenance de w à $L_{\bowtie\theta}(\mathcal{A})$ se traduit alors en : $\pi'_0(\prod_{i=1}^n \mathbf{P}'_{a_i})\mathbf{1}_F^T \bowtie \theta b^{n+1}$? Observez que l'espace nécessaire au calcul de θb^{n+1} est en $O(n)$. On calcule aussi $\mathbf{v} = \pi'_0(\prod_{i=1}^n \mathbf{P}'_{a_i})$ en initialisant $\mathbf{v}$ à π'_0 et en le multipliant successivement par $\mathbf{P}'_{a_i}$. A la $i$ème itération, la somme des coefficients de $\mathbf{v}$ est exactement b^{i+1}. Par conséquent, ceci s'effectue aussi en espace $O(n)$. Finalement, la comparaison requiert seulement des indices de bits à comparer eux aussi en espace $O(n)$. □

Nous terminons cette section par un résultat intéressant montrant qu'utiliser des « poids » en lieu et place des probabilités n'étend pas l'expressivité de tels automates. A cette fin, nous introduisons les PA generalisés et leurs langages.

Définition 5.1.15. *Un* PA generalisé $\mathcal{A} = (Q, A, \{\mathbf{P}_a\}_{a\in A}, \pi_0, \pi_f)$ *est défini par :*

- *Q, l'ensemble fini des états ;*
- *A, un alphabet fini ;*
- *Pour tout $a \in A$, $\mathbf{P}_a$, une matrice réelle indicée par $Q \times Q$;*

— π_0, *un vecteur initial de réels indicé par* Q *et* π_f, *un vecteur final de réels indicé par* Q.

Définition 5.1.16. *Soit* $\mathcal{A}$ *un* PA *généralisé et* $w \in A^*$ *un mot, le* poids d'acceptation *de* w *par* $\mathcal{A}$ *est défini par :*

$$\mathbf{Pr}_{\mathcal{A}}(w) \stackrel{def}{=} \pi_0 \mathbf{P}_w \pi_f^T$$

où comme précédemment $\mathbf{P}_w \stackrel{def}{=} \prod_{i=1}^{n} \mathbf{P}_{a_i}$ *pour* $w = a_1 \dots a_n$.

Nous définissons maintenant la famille des langages stochastiques (rationnels) généralisés.

Définition 5.1.17. *Soit* $\mathcal{A}$ *un* PA *généralisé,* $\theta \in \mathbb{R}$ *un seuil et* $\bowtie \in \{<, \leq, >, \geq, =, \neq\}$ *un opérateur de comparaison. Alors* $L_{\bowtie\theta}(\mathcal{A})$ *est défini par :*

$$L_{\bowtie\theta}(\mathcal{A}) = \{w \in A^* \mid \mathbf{Pr}_{\mathcal{A}}(w) \bowtie \theta\}$$

Lorsque les nombres apparaissant dans cette définition sont des rationnels, on parle de langage stochastique rationnel généralisé.

Théorème 5.1.18 ([33])**.** *Les familles des langages stochastiques (rationnels) généralisés et des langages stochastiques (rationnels) coïncident.*
De plus il existe un algorithme, opérant en temps polynomial, qui transforme un langage stochastique rationnel généralisé en un langage stochastique rationnel.

Clôture

La proposition suivante montre que comme pour la plupart des familles classiques de langages, les langages stochastiques sont clos par intersection et union avec un langage régulier. Ce qui est intéressant ici c'est que l'automate probabiliste construit dans la preuve est de taille linéaire en la taille des deux automates en entrée (contrairement au produit synchronisé utilisé dans d'autres constructions).

Proposition 5.1.19. *La famille des langages stochastiques (rationnels) est close par intersection et union avec les langages reguliers.*

Démonstration. Soient $L_{\bowtie\theta}(\mathcal{A}_1)$ un langage stochastique rationnel (avec $\bowtie \in \{>, \geq\}$) et $L_{=1}(\mathcal{A}_2)$ un langage régulier (c.-à-d. $\mathcal{A}_2$ est un automate probabiliste avec des distributions de Dirac). Sans perte de généralité, nous supposons que $\bowtie\ \theta$ est différent de > 1. Alors $\mathcal{A} = (Q, A, \{\mathbf{P}_a\}_{a\in A}, \pi_0, F)$ est défini par :

— $Q = Q_1 \uplus Q_2$ (où $\uplus$ dénote l'union disjointe) ;
— Pour tout $i \in \{1, 2\}$ et $q \in Q_i$, $\pi_0(q) = \frac{1}{2}\pi_{i,0}(q)$;
— Pour tout $a \in A$, $q_1, q_1' \in Q_1$, $q_2, q_2' \in Q_2$, $\mathbf{P}_a[q_1, q_1'] = \mathbf{P}_{1,a}[q_1, q_1']$, $\mathbf{P}_a[q_2, q_2'] = \mathbf{P}_{2,a}[q_2, q_2']$ et $\mathbf{P}_a[q_1, q_2'] = \mathbf{P}_a[q_2, q_1'] = 0$;
— $F = F_1 \uplus F_2$.

Montrons que $L_{\bowtie \frac{\theta}{2}}(\mathcal{A}) = L_{\bowtie\theta}(\mathcal{A}_1) \cup L_{=1}(\mathcal{A}_2)$. Considérons un mot w. Par construction, $\mathbf{Pr}_{\mathcal{A}}(w) = \frac{1}{2}(\mathbf{Pr}_{\mathcal{A}_1}(w) + \mathbf{Pr}_{\mathcal{A}_2}(w))$.
Cas $w \in L_{=1}(\mathcal{A}_2)$**.** $\mathbf{Pr}_{\mathcal{A}}(w) \geq \frac{1}{2} \bowtie \frac{\theta}{2}$. Donc $w \in L_{\bowtie \frac{\theta}{2}}(\mathcal{A})$.
Cas $w \notin L_{=1}(\mathcal{A}_2)$**.** $\mathbf{Pr}_{\mathcal{A}}(w) = \frac{1}{2}\mathbf{Pr}_{\mathcal{A}_1}(w)$.
Donc $w \in L_{\bowtie \frac{\theta}{2}}(\mathcal{A})$ ssi $w \in L_{\bowtie\theta}(\mathcal{A}_1)$.
Nous laissons le lecteur vérifier que $L_{\bowtie \frac{1+\theta}{2}}(\mathcal{A}) = L_{\bowtie\theta}(\mathcal{A}_1) \cap L_{=1}(\mathcal{A}_2)$. □

En utilisant une preuve similaire à celle de la proposition 5.1.8, on établit un deuxième résultat de clôture.

Proposition 5.1.20. *La famille des langages stochastiques (rationnels)* $\{L_{=\theta}(\mathcal{A})\}_{\mathcal{A},\theta}$ *est close par intersection.*

Afin de prouver des résultats négatifs de clôture, nous exhibons un langage stochastique particulier.

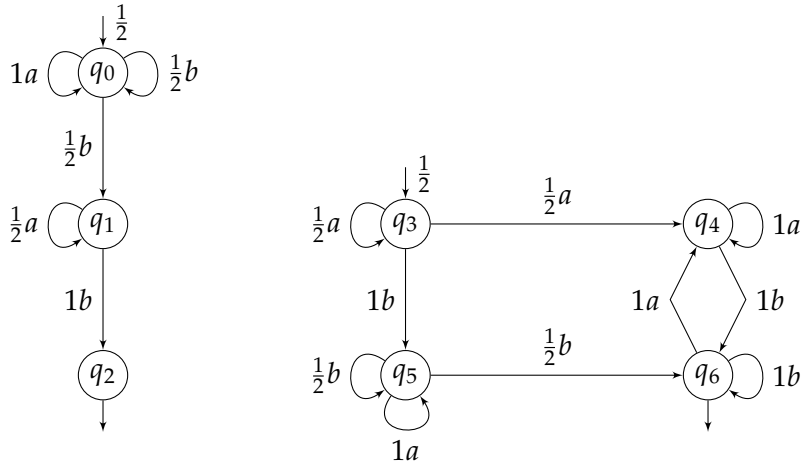


FIGURE 5.5 – *Un* PA *pour* $\{a^{m_1}b \dots ba^{m_k}b \mid k > 1 \wedge m_1 = m_k\}$.

Lemme 5.1.21. *Soit* $\mathcal{A}$ *l'automate de la figure 5.5. Alors :*

$$L_{=\frac{1}{2}}(\mathcal{A}) = \{a^{m_1}b \dots ba^{m_k}b \mid k > 1 \wedge \forall i,\ m_i \geq 0 \wedge m_1 = m_k\}.$$

Démonstration. Clairement, pour tout mot $w \in A^* \setminus \{a^{m_1}b \dots ba^{m_k}b \mid k > 1 \wedge m_i \geq 0\}$, on a $\mathbf{Pr}_{\mathcal{A}}(w) = 0$, car alors soit w se termine par a, soit w contient au plus un b.
Soit $w = a^{m_1}b \dots ba^{m_k}b$ avec $k > 1$. Ce mot peut être accepté soit par un chemin démarrant de q_0, soit par un chemin démarrant de q_3.

— Lorsque le chemin démarre de q_0, afin d'être accepté il doit rester dans q_0 pour tous les b excepté pour celui qui précède a^{m_k}. Puis il doit rester dans q_1 pour tous les a. Ceci conduit à une probabilité d'acceptation égale à $\frac{1}{2^{k+m_k}}$.
— Lorsque le chemin démarre de q_3, afin d'être rejeté il doit rester dans q_3 pour tous les a qui précèdent le premier b puis doit rester dans q_5 en lisant les autres b. Ceci conduit à une probabilité de rejet égale à $\frac{1}{2^{k+m_1}}$.

Ainsi, $\mathbf{Pr}_{\mathcal{A}}(w) = \frac{1}{2} - \frac{1}{2^{k+m_1}} + \frac{1}{2^{k+m_k}}$. Par conséquent, w is accepté si et seulement si $m_1 = m_k$. □

Proposition 5.1.22 ([34]). *La famille des langages stochastiques n'est close ni par concaténation avec un langage régulier, ni par étoile de Kleene, ni par homomorphisme.*

Démonstration. Nous établissons uniquement le premier résultat et laissons le soin au lecteur de s'appuyer sur le lemme 5.1.21 pour les autres résultats. Soit $L = \{a^{m_1}b \dots ba^{m_k}b \mid k > 1 \wedge m_1 = m_k\}$ le langage stochastique du lemme 5.1.21. Alors $LA^* = \{a^{m_1}ba^{m_2}b \dots a^{m_k}ba^* \mid \exists i > 1 :\ m_i = m_1\}$ n'est pas un langage stochastique comme établi par la proposition 5.1.12. □

Nous terminons cette section par un théorème valable uniquement pour les langages stochastiques, la clôture des langages stochastiques rationnels par intersection et union demeurant une question ouverte.

Théorème 5.1.23 ([12]). *La famille des langages stochastiques n'est close ni par intersection ni par union même pour un alphabet unaire.*

5.1.3 Décidabilité et indécidabilité dans les PA

Nous illustrons ici la frontière entre décidabilité et indécidabilité en étudiant deux problèmes voisins : l'équivalence d'automates probabilistes et l'égalité de langages stochastiques.

Définition 5.1.24. *Soient $\mathcal{A}$ et $\mathcal{A}'$ deux automates probabilistes sur le même alphabet A. $\mathcal{A}$ et $\mathcal{A}'$ sont* équivalents *si pour tout mot $w \in A^*$:*

$$\mathbf{Pr}_{\mathcal{A}}(w) = \mathbf{Pr}_{\mathcal{A}'}(w).$$

Algorithme 1 : Equivalence de deux automates probabilistes [29].

`Equivalence`$(\mathcal{A}, \mathcal{A}')$
Input : $\mathcal{A}, \mathcal{A}'$, deux PA sur l'alphabet A tels que $F \cup F' \neq \emptyset$
Output : le statut de l'équivalence avec un témoin en cas de non équivalence
Data : $\mathbf{v}, \mathbf{x}, \mathbf{y}, \mathbf{z}$ vecteurs de $\mathbb{R}^Q$ et $\mathbf{v}', \mathbf{x}', \mathbf{y}', \mathbf{z}'$ vecteurs de $\mathbb{R}^{Q'}$
Data : `Stack` dont les éléments sont des paires de vecteur de $\mathbb{R}^{Q \cup Q'}$ et de mot
Data : `Gen`, un ensemble de vecteurs orthogonaux (non nuls) de $\mathbb{R}^{Q \cup Q'}$, a un caractère

if $\pi_0 \cdot \mathbf{1}_F \neq \pi'_0 \cdot \mathbf{1}_{F'}$ **then return**$(\mathbf{false}, \varepsilon)$

`Gen` $\leftarrow \{(\mathbf{1}_F, \mathbf{1}_{F'})\}$; `Push`$(\mathtt{Stack}, ((\mathbf{1}_F, \mathbf{1}_{F'}), \varepsilon))$
repeat
  $((\mathbf{v}, \mathbf{v}'), w) \leftarrow$ `Pop`$(\mathtt{Stack})$
  for $a \in A$ **do**
    $\mathbf{z} \leftarrow \mathbf{P}_a \mathbf{v}$; $\mathbf{z}' \leftarrow \mathbf{P}'_a \mathbf{v}$
    if $\pi_0 \cdot \mathbf{z} \neq \pi'_0 \cdot \mathbf{z}'$ **then return**$(\mathbf{false}, aw)$
    $\mathbf{y} \leftarrow \mathbf{0}$; $\mathbf{y}' \leftarrow \mathbf{0}$
    for $(\mathbf{x}, \mathbf{x}') \in$ `Gen` **do**
      $\mathbf{y} \leftarrow \mathbf{y} + \frac{\mathbf{z} \cdot \mathbf{x}}{\mathbf{x} \cdot \mathbf{x}} \mathbf{x}$
      $\mathbf{y}' \leftarrow \mathbf{y}' + \frac{\mathbf{z}' \cdot \mathbf{x}'}{\mathbf{x}' \cdot \mathbf{x}'} \mathbf{x}'$
    if $(\mathbf{z}, \mathbf{z}') \neq (\mathbf{y}, \mathbf{y}')$ **then**
      `Push`$(\mathtt{Stack}, ((\mathbf{z}, \mathbf{z}'), aw))$
      `Gen` $\leftarrow$ `Gen` $\cup \{(\mathbf{z} - \mathbf{y}, \mathbf{z}' - \mathbf{y}')\}$
until `IsEmpty`$(\mathtt{Stack})$
return$(\mathbf{true})$

Sans perte de généralité, nous supposons que $F \cup F' \neq \emptyset$. Examinons l'algorithme 1. Il essaie d'établir la non équivalence en trouvant un contre-exemple par énumération des mots par longueur croissante (en démarrant avec le mot vide). Tant qu'il n'y parvient pas il maintient une pile de mots w pour lesquels il cherchera des contre-exemples de la forme aw. Afin d'éviter des calculs redondants, il conserve aussi dans la pile la paire de vecteurs $(\mathbf{P}_w\mathbf{1}_F, \mathbf{P}'_w\mathbf{1}_{F'})$.

Sans « élagage », on obtiendrait un semi-algorithme qui ne terminerait que si $\mathcal{A}$ et $\mathcal{A}'$ n'étaient pas équivalents. Aussi, l'algorithme maintient `Gen` un ensemble de vecteurs orthogonaux non nuls de $\mathbb{R}^{Q\cup Q'}$. Quand un mot w n'est pas un contre-exemple, l'algorithme teste si le vecteur $(\mathbf{P}_w\mathbf{1}_F, \mathbf{P}'_w\mathbf{1}_{F'})$ n'appartient pas à l'espace vectoriel engendré par `Gen`. Il effectue ce test en calculant la projection orthogonale de ce vecteur sur ce sous-espace et en le comparant avec le vecteur original. Si le vecteur n'y appartient pas alors le mot w est empilé. La différence entre le vecteur et sa projection orthogonale est ajoutée à `Gen` préservant ainsi la propriété d'orthogonalité.

Par construction, lorsque l'algorithme trouve un contre-exemple il a établi la non équivalence. Prouver l'équivalence lorsque l'algorithme n'a pas trouvé de contre-exemple est plus subtil.

Proposition 5.1.25. *L'algorithme 1 opère en temps $O(|A|n^3)$ avec $n = |Q| + |Q'|$ et décide si $\mathcal{A}$ et $\mathcal{A}'$ sont équivalents.*

Démonstration. Puisque la dimension de l'espace vectoriel engendré par `Gen` est au plus n, il y a au plus n itérations de la boucle principale. L'indice de la première boucle interne parcourt A tandis que celui de la boucle la plus interne parcourt `Gen` $\times$ `Gen`. Ceci induit une complexité temporelle en $O(n^3|A|)$.

Supposons, par l'absurde, que les automates ne sont pas équivalents et que l'algorithme ne l'a pas détecté. Soit u un mot tel que $\mathbf{Pr}_{\mathcal{A}}(u) \neq \mathbf{Pr}_{\mathcal{A}'}(u)$. u n'a pas été examiné par l'algorithme. Ecrivons $u = w'w$ avec w le plus grand suffixe examiné par l'algorithme. Parmi ces mots u, sélectionnons un mot tel que $|w'|$ soit minimal. Nous affirmons qu'il existe un mot w'' qui a été empilé avant w tel que $\mathbf{Pr}_{\mathcal{A}}(w'w'') \neq \mathbf{Pr}_{\mathcal{A}'}(w'w'')$.

En effet puisque w n'est pas empilé, considérons $w_1, \ldots, w_k$ les mots précédemment insérés dans la pile lorsque w est examiné, il existe donc $\lambda_1, \ldots, \lambda_k$ tels que :

$$\mathbf{P}_w\mathbf{1}_F = \sum_{i=1}^{k} \lambda_i \mathbf{P}_{w_i}\mathbf{1}_F \text{ et } \mathbf{P}'_w\mathbf{1}_{F'} = \sum_{i=1}^{k} \lambda_i \mathbf{P}'_{w_i}\mathbf{1}_{F'}.$$

Aussi :

$$\mathbf{Pr}_{\mathcal{A}}(w'w) \stackrel{\text{def}}{=} \pi_0 \mathbf{P}_{w'} \mathbf{P}_w \mathbf{1}_F = \sum_{i=1}^{k} \lambda_i \pi_0 \mathbf{P}_{w'} \mathbf{P}_{w_i} \mathbf{1}_F = \sum_{i=1}^{k} \lambda_i \mathbf{Pr}_{\mathcal{A}}(w'w_i).$$

De manière similaire :

$$\mathbf{Pr}_{\mathcal{A}'}(w'w) = \sum_{i=1}^{k} \lambda_i \mathbf{Pr}_{\mathcal{A}'}(w'w_i).$$

Donc $\sum_{i=1}^{k} \lambda_i \mathbf{Pr}_{\mathcal{A}}(w'w_i) \neq \sum_{i=1}^{k} \lambda_i \mathbf{Pr}_{\mathcal{A}'}(w'w_i)$ ce qui implique l'existence d'un indice i tel que $\mathbf{Pr}_{\mathcal{A}}(w'w_i) \neq \mathbf{Pr}_{\mathcal{A}'}(w'w_i)$. Par conséquent, nous avons établi notre affirmation.

Réécrivons $w' = w'''a$. Puisque w_i a été inséré dans la pile, aw_i est examiné par l'algorithme. Aussi, le mot $u' = w'w_i$ admet une décomposition $u' = z'z$ où z le plus grand suffixe examiné par l'algorithme a pour suffixe aw_i. Donc $|z'| < |w'|$, entraînant une contradiction avec la minimalité de w'. On a donc prouvé la correction de l'algorithme 1. □

Considérons à présent le problème de l'égalité des langages stochastiques $L_{\bowtie\theta}(\mathcal{A})$ et $L_{\bowtie'\theta'}(\mathcal{A}')$. Si $\mathcal{A}$ and $\mathcal{A}'$ sont équivalents et $\bowtie\ \theta$ est identique à $\bowtie'\ \theta'$, les langages sont égaux. Malheureusement, il ne s'agit que d'une condition suffisante et le problème plus simple du vide du langage est déjà indécidable.

Proposition 5.1.26 ([24]). *Soit $\mathcal{A}$ un automate probabiliste rationnel, le problème $L_{=\frac{1}{2}}(\mathcal{A}) = \emptyset$? est indécidable.*

Rappelons le problème de correspondance de Post (PCP). Étant donnés un alphabet A et deux morphismes φ_1, φ_2 de A vers $\{0,1\}^+$, existe-t-il un mot $w \in A^+$ tel que $\varphi_1(w) = \varphi_2(w)$? Ce problème est indécidable. Nous considérons ici une restriction de ce problème où les images des caractères appartiennent à $(10+11)^+$. Ce problème reste indécidable car en insérant un 1 avant chaque caractère de l'image on réduit le problème original à cette variante. Un mot $w = a_1 \dots a_n \in (10+11)^+$ définit une valeur $\mathsf{val}(w) \in [0,1]$ par : $\mathsf{val}(w) = \sum_{i=1}^{n} \frac{a_i}{2^{n-i}}$. De plus, puisque chaque mot démarre avec un 1, $\mathsf{val}(w) = \mathsf{val}(w')$ implique $w = w'$. Nous allons raffiner l'exemple 5.1.2 afin d'établir la preuve de la proposition 5.1.26 [13].

Démonstration. À partir d'une instance de PCP, on construit d'abord un PA $\mathcal{A}$ tel que $L_{=\frac{1}{2}}(\mathcal{A}) = \{\varepsilon\}$ si et seulement si le PCP n'a pas de solution.

Pour $w \in A^+$ et $i \in \{1,2\}$, nous définissons $\mathsf{val}_i(w) = \mathsf{val}(\varphi_i(a))$. Le PA $\mathcal{A} = (Q, A, \{\mathbf{P}_a\}_{a\in A}, \pi_0, F)$ est alors défini par :

— $Q = \{q_{10}, q_{11}, q_{20}, q_{21}\}$;
— $\pi_0[q_{10}] = \pi_0[q_{20}] = \frac{1}{2}$ et $\pi_0[q_{11}] = \pi_0[q_{21}] = 0$;
— Pour tout $a \in A$ et $i \in \{1, 2\}$, $\mathbf{P}_a[q_{i0}, q_{i1}] = 1 - \mathbf{P}_a[q_{i0}, q_{i0}] = \mathsf{val}_i(a)$, $\mathbf{P}_a[q_{i1}, q_{i1}] = 1 - \mathbf{P}_a[q_{i1}, q_{i0}] = \mathsf{val}_i(a) + 2^{-|\varphi_i(a)|}$, et tous les autres éléments des matrices stochastiques sont nuls ;
— $F = \{q_{11}, q_{20}\}$.

Par conséquent, pour tout $w \in A^*$ et $a \in A$

$$\mathbf{1}_{q_{i0}}\mathbf{P}_{wa}\mathbf{1}_{q_{i1}}^T = \mathbf{1}_{q_{i0}}\mathbf{P}_w\mathbf{1}_{q_{i1}}^T(\mathsf{val}_i(a) + 2^{-|\varphi_i(a)|}) + (1 - \mathbf{1}_{q_{i0}}\mathbf{P}_w\mathbf{1}_{q_{i1}}^T)\mathsf{val}_i(a)$$

$$= \mathsf{val}_i(a) + 2^{-|\varphi_i(a)|}\mathbf{1}_{q_{i0}}\mathbf{P}_w\mathbf{1}_{q_{i1}}^T.$$

Par induction nous obtenons que pour tout $w = a_1 \dots a_n$:

$$\mathbf{1}_{q_{i0}}\mathbf{P}_w\mathbf{1}_{q_{i1}}^T = \sum_{j=1}^{n} \mathsf{val}_i(a_j)2^{-\sum_{j<k\leq n}|\varphi_i(a_k)|} = \mathsf{val}_i(w).$$

Aussi, pour $w \in A^+$: $\mathbf{Pr}_{\mathcal{A}}(w) = \frac{1}{2}(\mathsf{val}_1(w) + 1 - \mathsf{val}_2(w))$. Donc $w \in L_{=\frac{1}{2}}(\mathcal{A})$ ssi $\mathsf{val}(\varphi_1(w)) = \mathsf{val}(\varphi_2(w))$. En raison de notre hypothèse sur les images, ceci est équivalent à $\varphi_1(w) = \varphi_2(w)$, ce qui signifie que w est une solution de l'instance du PCP.

A^+ est un langage régulier (reconnu par un automate à deux états). En utilisant la proposition 5.1.19, on construit $\mathcal{A}'$ tel que $L_{=\frac{1}{2}}(\mathcal{A}') = L_{=\frac{1}{2}}(\mathcal{A}) \setminus \{\varepsilon\}$. Par conséquent, $L_{=\frac{1}{2}}(\mathcal{A}') = \emptyset$ ssi le PCP n'a pas de solution. □

En s'appuyant sur la proposition 5.1.8 nous obtenons immédiatement le corollaire suivant.

Corollaire 5.1.27. *Soit $\mathcal{A}$ un automate probabiliste rationnel et $\theta \in]0, 1[$ un rationnel, le problème $L_{\geq\theta}(\mathcal{A}) = \emptyset$? est indécidable.*

Le prochain corollaire nécessite un peu plus de travail.

Corollaire 5.1.28. *Soit $\mathcal{A}$ un automate probabiliste rationnel et $\theta \in]0, 1[$ un rationnel, le problème $L_{>\theta}(\mathcal{A}) = \emptyset$? est indécidable.*

Démonstration. Les probabilités de l'automate correspondant à la réduction du PCP dans la preuve de la proposition 5.1.26 sont des multiples de 2^{-k} pour un certain k dépendant du PCP. La transformation de la proposition 5.1.8 produit un automate, $\mathcal{A}$, dont les probabilités sont des produits des probabilités originales. Ils sont donc des multiples de 4^{-k} et $L_{\geq\frac{1}{4}}(\mathcal{A}) = \emptyset$ ssi le PCP correspondant n'a pas de solution.

En raison des probabilités de transition, pour tout mot $w \in A^+$, $\mathbf{Pr}_{\mathcal{A}}(w) = \frac{d}{4^{k|w|}}$ où d est un entier dépendant de w. Donc $\mathbf{Pr}_{\mathcal{A}}(w) \geq \frac{1}{4}$ si et seulement si $\mathbf{Pr}_{\mathcal{A}}(w) > \frac{1}{4} - \frac{1}{4^{k|w|}}$.

Soit $\mathcal{A}' = (Q', A, \{\mathbf{P}'_a\}_{a\in A}, \pi'_0, F')$ défini par :

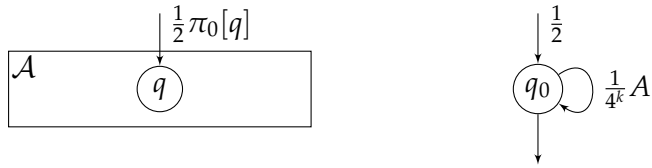
— $Q' = Q \cup \{q_0, q_1\}$;

— Pour tout $q \in Q$, $\pi'_0[q] = \frac{\pi_0[q]}{2}$, $\pi'_0[q_0] = \frac{1}{2}$ et $\pi'_0[q_1] = 0$;

— Pour tout $a \in A$ et $q, q' \in Q$, $\mathbf{P}'_a[q, q'] = \mathbf{P}'_a[q, q']$, $\mathbf{P}'_a[q_0, q_0] = 1 - \mathbf{P}'_a[q_0, q_1] \stackrel{\text{def}}{=} \frac{1}{4^k}$, $\mathbf{P}'_a[q_1, q_1] = 1$, et tous les autres éléments des matrices stochastiques sont nuls ;

— $F' = F \cup \{q_0\}$.

$\mathcal{A}'$ est décrit ci-dessous (où l'on a omis l'état puits q_1).



Ainsi, pour tout $w \in A^+$, $\mathbf{Pr}_{\mathcal{A}'}(w) = \frac{\mathbf{Pr}_{\mathcal{A}}(w)}{2} + \frac{1}{2\times 4^{k|w|}}$. Ce qui peut être réécrit comme : $\mathbf{Pr}_{\mathcal{A}}(w) = 2\mathbf{Pr}_{\mathcal{A}'}(w) - \frac{1}{4^{k|w|}}$. Par conséquent, $\mathbf{Pr}_{\mathcal{A}}(w) > \frac{1}{4} - \frac{1}{4^{k|w|}}$ ssi $\mathbf{Pr}_{\mathcal{A}'}(w) > \frac{1}{8}$. Donc $L_{>\frac{1}{8}}(\mathcal{A}') = L_{\geq\frac{1}{4}}(\mathcal{A}) \cup \{\varepsilon\}$. Comme précédemment, on peut construire $\mathcal{A}''$ tel que $L_{>\frac{1}{8}}(\mathcal{A}'') = L_{>\frac{1}{8}}(\mathcal{A}') \setminus \{\varepsilon\}$, ce qui conclut la preuve. □

Attirons l'attention sur le seuil $\theta \notin \{0, 1\}$ pour les résultats d'indécidabilité. En effet, lorsque $\theta \in \{0, 1\}$, étant donné un PA $\mathcal{A}$, on considère l'automate non déterministe $\mathcal{A}'$ dont l'ensemble des états initiaux est le support de la distribution initiale de $\mathcal{A}$, et dont la fonction de transition Δ est définie par $(q, a, q') \in \Delta$ ssi $\mathbf{P}_a[q, q'] > 0$. La spécification de l'ensemble des états finals dépend du problème considéré. En choisissant le même ensemble d'états finals F, $L_{>0}(\mathcal{A}) = L(\mathcal{A}')$. En choisissant le complémentaire de F, $L_{<1}(\mathcal{A}) = L(\mathcal{A}')$. Par conséquent, le problème du vide du langage pour les valeurs extrêmes de seuil est décidable.

Etant donné un automate probabiliste $\mathcal{A}$, le problème *de la valeur 1* consiste à décider s'il existe une suite de mots $\{w_n\}_{n\in\mathbb{N}}$ telle que $\lim_{n\to\infty} \mathbf{Pr}_{\mathcal{A}}(w_n) = 1$. Ce problème a des applications concrètes en planification. Le statut de ce problème n'a été résolu que très récemment.

Proposition 5.1.29 ([13]). *Le problème, étant donné $\mathcal{A}$ un* PA *rationnel, de l'existence d'une suite $\{w_n\}_{n\in\mathbb{N}}$ telle que $\lim_{n\to\infty} \mathbf{Pr}_{\mathcal{A}}(w_n) = 1$, est indécidable.*

Le problème de valeur 1 conduit assez naturellement à une autre sémantique des PA. Au lieu de considérer des mots finis, on s'intéresse aux mots infinis. De manière similaire au cas des mots finis, lorsqu'un mot infini $w = w_1 w_2 \dots$ est choisi, le PA $\mathcal{A}$ se comporte comme une DTMC $\mathcal{M}_{\mathcal{A}}^{w}$ dont les états sont les paires (q, n) avec $q \in Q$ et $n \in \mathbb{N}$. Dans l'état (q, n), les probabilités de transition sont définies par la matrice $\mathbf{P}_{w_{n+1}}$: $\mathbf{P}[(q, n), (q', n+1)] = \mathbf{P}_{w_{n+1}}[q, q']$ (les autres probabilités de transition étant nulles). L'ensemble des états finals est alors interprété comme un ensemble d'états que l'exécution aléatoire $\rho = q_0 q_1 \dots$ doit rencontrer infinimement souvent :

$$\mathbf{Pr}_{\mathcal{A}}(w) = \mathbf{Pr}_{\mathcal{M}_{\mathcal{A}}^{w}}(\{\rho \mid \rho = q_0 q_1 \dots \wedge \forall n \; \exists n' > n \; q_{n'} \in F\}).$$

Lorsqu'on considère cette sémantique on parle *d'automate de Büchi probabiliste* (PBA) [4, 3]. Définissons les langages associés.

Définition 5.1.30. *Soit $\mathcal{B} = (Q, A, \{\mathbf{P}_a\}_{a \in A}, \pi_0, F)$ un* PBA*, $\theta \in [0, 1]$ un* seuil *et $\bowtie \in \{<, \leq, >, \geq, =, \neq\}$ un opérateur de comparaison. Alors le langage $L^{\omega}_{\bowtie \theta}(\mathcal{B})$ est défini par :*

$$L^{\omega}_{\bowtie \theta}(\mathcal{B}) = \{w \in A^{\omega} \mid \mathbf{Pr}_{\mathcal{B}}(w) \bowtie \theta\}.$$

La preuve originale du résultat suivant [2, 3] s'obtient par réduction d'une variante du problème du vide pour les langages stochastiques. Cette preuve s'appuie sur le fait que le fait que le langage d'un PBA défini par une condition de stricte positivité soit vide ou non, ne dépend pas uniquement de la structure de l'automate, mais aussi des valeurs précises des probabilités. Nous fournissons ici une courte preuve alternative basée sur la proposition 5.1.29.

Théorème 5.1.31. *Le problème, étant donné un* PBA *$\mathcal{B}$, de savoir si $L^{\omega}_{>0}(\mathcal{B}) = \emptyset$, est indécidable.*

Démonstration. À partir d'un PA $\mathcal{A}$ sur l'alphabet A, on construit un PBA $\mathcal{B}$, comme illustré par la figure 5.6. Le PBA $\mathcal{B}$ étend $\mathcal{A}$ avec un état supplémentaire $f_\sharp$, qui est son état final, et des transitions depuis tout état final de $\mathcal{A}$ vers $f_\sharp$, ainsi que de $f_\sharp$ vers l'état initial de $\mathcal{A}$, toutes sur une nouvelle lettre $\sharp \notin A$. Cette construction assure que la valeur de $\mathcal{A}$ est 1 si et seulement si il existe un mot infini w accepté avec probabilité positive dans $\mathcal{B}$.

Pour se convaincre de la correction de cette réduction, supposons d'abord que la valeur de $\mathcal{A}$ est 1. Alors, à tout entier $k \in \mathbb{N}$, on peut associer un mot fini $v_k \in \Sigma^*$ tel que $\mathbf{Pr}_{\mathcal{A}}(v_k) \geq 1 - 2^{-k}$. On définit alors le mot infini

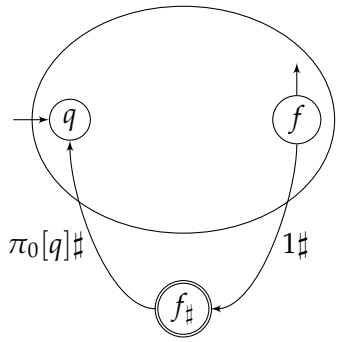


FIGURE 5.6 – *Réduction du problème de valeur 1 dans les* PA.

w comme la concaténation des v_k séparés par deux $\sharp$: $w = v_1\sharp\sharp v_2\sharp\sharp v_3\cdots$. La probabilité d'acceptation de w dans $\mathcal{B}$ est $\mathbf{Pr}_{\mathcal{B}}(w) \geq \prod_{k\in\mathbb{N}}(1-2^{-k}) > 0$. Ainsi, w est accepté par $\mathcal{B}$.
Supposons à présent que la valeur de $\mathcal{A}$ est strictement inférieure à 1. Ainsi, il existe $\varepsilon > 0$, tel que pour tout mot fini $v \in \Sigma^*$, $\mathbf{Pr}_{\mathcal{A}}(v) < 1-\varepsilon$. Pour qu'un mot soit accepté avec probabilité positive dans $\mathcal{B}$, il doit nécessairement contenir une infinité de facteurs $\sharp\sharp$. Il est donc de la forme $w = v_1\sharp\sharp v_2\sharp\sharp v_3\cdots$ avec $v_k \in \Sigma^*$ pour tout $k \in \mathbb{N}$. Alors, $\mathbf{Pr}_{\mathcal{B}}(w) = \prod_{k\in\mathbb{N}} \mathbf{Pr}_{\mathcal{A}}(v_k) < \prod_{k\in\mathbb{N}}(1-\varepsilon) = 0$, et donc, w n'est pas accepté avec probabilité positive par $\mathcal{B}$. □

5.2 Processus de décision markoviens partiellement observables

5.2.1 Présentation

Afin de motiver le modèle des processus de décision markoviens partiellement observables, reprenons notre problème de planification de vacances. Tel que nous l'avons présenté, le plan produit par l'analyse d'un automate probabiliste est le « meilleur » avant son exécution. Cependant la modélisation par automate probabiliste ne prend pas en compte des informations que nous pourrions acquérir durant les vacances. Par exemple, le choix de l'agence pour les excursions s'effectuerait à l'hôtel en ayant pris connaissance des tarifs en vigueur à cette date. Afin de prendre en compte l'interaction entre planification et exécution, il est nécessaire d'introduire un formalisme plus expressif : les processus de décision markoviens partiellement observables (POMDP) [1]. Dans la suite l'*agent* (ou *contrôleur*, ou

joueur) choisit les actions au cours de l'exécution.

Les POMDP étendent les automates probabilistes. De même que précédemment, étant donné un état courant s et une action a, l'état suivant s' est obtenu par la distribution $p(s'|s,a)$ (analogue des matrices de probabilité $\{\mathbf{P}_a\}_{a\in A}$ pour les PA). Comme les POMDP sont souvent employés pour résoudre des problèmes d'optimisation, une récompense $r(s,a)$ est associée à une paire constituée d'un état s et d'une action a. Dans le cas d'exécutions finies on associe une récompense finale $r_f(s)$ pour tout $s \in S$.

Un POMDP est muni d'un ensemble d'observations Ω. À chaque instant, l'agent reçoit une observation $\omega \in \Omega$. Les observations sont liées aux états et aux actions : l'agent observe ω lorsqu'il a exécuté l'action a et atteint l'état s avec probabilité $o(\omega|a,s)$. De manière cohérente, les récompenses obtenues ne sont pas connues de l'agent. Si elles étaient observables, il faudrait exiger que ces récompenses ne dépendent que des observations et des actions.

Définition 5.2.1 (POMDP)**.** *Un* processus de décision markovien partiellement observable $\mathcal{M} = (S, \Omega, A, o, p, r)$ *est défini par :*

- *S, l'ensemble fini des états ;*
- *Ω, l'ensemble fini des observations ;*
- *A, l'ensemble fini des actions ;*
- *À tout $a \in A$ et $s \in S$, on associe une distribution dont le domaine est Ω : $o(\omega|a,s)$ dénote la probabilité d'une observation ω pour cette distribution.*
- *À tout $a \in A$ et $s \in S$, on associe une distribution dont le domaine est S : $p(s'|s,a)$ dénote la probabilité du nouvel état s' pour cette distribution.*
- *À chaque état s on associe une récompense finale $r_f(s) \in \mathbb{Q}$ et à chaque action $a \in A$ on associe une récompense $r(s,a) \in \mathbb{Q}$.*

Exemple 5.2.2 (Un exemple de POMDP)**.** *Nous illustrons le formalisme des* POMDP *avec un exemple proposé par Sondik [31] et nous introduisons une représentation graphique en figure 5.7. Ce* POMDP *modélise un problème de marketing. Au début d'une période de vente, une société décide de mettre sur le marché soit un produit de luxe (L) soit un produit standard (S) ce qui constitue les actions du* POMDP*. Le comportement des consommateurs dépend de leur état (interne) inconnu de la société : préférant les marques (B) ou y étant indifférent ($\overline{B}$). Cet état peut changer d'une période à la suivante. Cependant la seule information dont la société dispose est le fait que le client achète le produit (P) ou non ($\overline{P}$).*
La représentation graphique inclut trois types de sommets : les cercles sont les états du POMDP*, les losanges correspondent aux instants où l'action est choisie et les rectangles correspondent aux instants où le nouvel état est sélectionné*

aléatoirement et l'observation n'est pas encore fournie. Par exemple, il y a un arc de B à (B, L) signifiant que la société a choisi de distribuer le produit de luxe. L'étiquette de l'arc correspond à la récompense $r(B, L)$. Il y a un arc de (B, L) à (L, B) étiquetée par 0.8 signifiant que $p(B|B, L) = 0.8$. Enfin l'étiquette $0.8P + 0.2\overline{P}$ de l'arc de (L, B) à B signifie qu'avec probabilité 0.8 le client achètera le produit, c.-à-d. $o(P|L, B) = 0.8$.

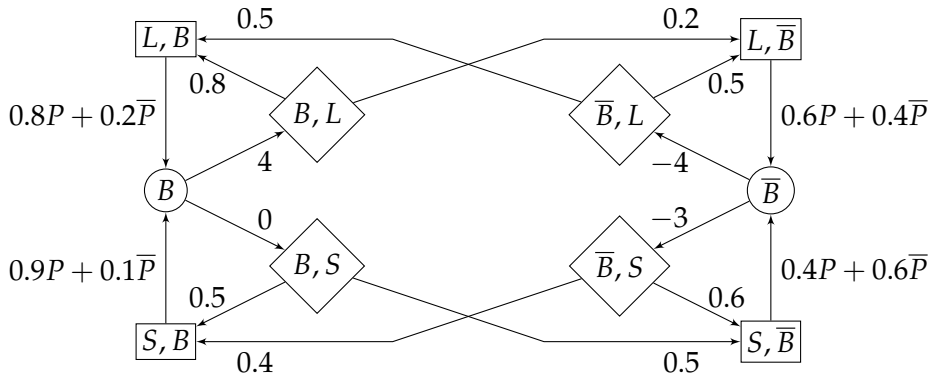


FIGURE 5.7 – *Un exemple de* POMDP.

Afin d'obtenir un processus stochastique à partir d'un POMDP, nous introduisons la notion de *politique* (aussi appelée *stratégie*). Une politique élimine le non déterminisme de la façon suivante. Étant donnée l'information disponible sur l'exécution courante, une politique sélectionne (de manière éventuellement aléatoire) la prochaine action à exécuter. L'information disponible correspond à ce qui est observé, et est donc un élément de $(A\Omega)^*$.

Définition 5.2.3. *Soit $\mathcal{M}$ un* POMDP. *Une* politique $\nu : (A\Omega)^* \to \mathsf{Dist}(A)$ *est une fonction de l'ensemble des* histoires $(A\Omega)^*$ *vers l'ensemble des distributions sur A. On note $\nu(\rho, a)$ la probabilité que l'action a soit choisie étant donnée l'histoire ρ.*

Une fois que la politique est fixée, un POMDP devient une chaîne de Markov à temps discret (DTMC) que nous formalisons ainsi.

Définition 5.2.4. *Soit $\mathcal{M}$ un* POMDP, *ν une politique et π une distribution sur S. Alors $\mathcal{M}_\pi^\nu$ est la* DTMC *définie par :*

- *$(A\Omega)^* \times S$, l'ensemble des états ;*
- *La distribution initiale π_0 est définie par : $\pi_0(\varepsilon, s) = \pi(s)$ et π_0 est nulle pour les autres états ;*

— *La matrice de transition* $\mathbf{P}$ *est définie par :* $\mathbf{P}[(\rho, s), (\rho a \omega, s')] = \nu(\rho, a) p(s'|s, a) o(\omega|a, s')$ *et* $\mathbf{P}$ *est nulle ailleurs.*

Lorsque l'ensemble des observations Ω est un singleton, l'observation ne fournit aucune information utile, on parle alors de POMDP aveugle. Ce modèle est très proche des PA : la seule différence est la possibilité de sélectionner l'action courante de manière aléatoire (ce point sera discuté dans la section 5.2.3). Lorsque $\Omega = S$ et pour toute paire (a, s) $o(s|a, s) = 1$, l'agent connaît exactement l'état du système, le POMDP devient un processus de décision markovien (MDP). Ce formalisme a fait l'objet de nombreuses recherches [25] mais il ne rentre pas dans le cadre de ce chapitre.

Considérons le cas d'une observation déterministe ne dépendant que de l'état courant : $\forall s \in S \; \exists \omega_s \in \Omega \; \forall a \in A \; o(\omega_s|a, s) = 1$. A priori, cette restriction semble diminuer le pouvoir d'expression des POMDP. En réalité, étant donné un POMDP $\mathcal{M}$ arbitraire, on peut construire un POMDP $\mathcal{M}'$ satisfaisant cette restriction et ayant le même comportement que $\mathcal{M}$. L'ensemble des états de $\mathcal{M}'$ est défini par $S' = S \times \Omega$ avec $o(\omega|a, (s, \omega)) = 1$ et $p'((s', \omega')|(s, \omega), a) = p(s'|s, a) o(\omega'|a, s')$. Nous laissons le soin au lecteur de compléter cette transformation et de vérifier sa correction. Avec des observations déterministes, la définition se simplifie ainsi.

Définition 5.2.5 (POMDP)**.** *Un* processus de décision markovien partiellement observable $\mathcal{M} = (S, \Omega, A, o, p, r)$ *est défini par :*

— S*, l'ensemble fini des états ;*
— Ω*, l'ensemble fini des observations ;*
— A*, l'ensemble fini des actions ;*
— $o : S \to \Omega$ *est la fonction d'observation ;*
— *À tout* $a \in A$ *et* $s \in S$*, on associe une distribution dont le domaine est* S *:* $p(s'|s, a)$ *dénote la probabilité du nouvel état* s' *pour cette distribution.*
— *À chaque état* s *on associe une récompense finale* $r_f(s) \in \mathbb{Q}$ *et à chaque action* $a \in A$ *on associe une récompense* $r(s, a) \in \mathbb{Q}$*.*

La notion de politique (ou stratégie) introduite en définition 5.2.3 pour la définition générale des POMDP s'adapte naturellement au cadre où l'observation est déterministe. De plus, la matrice de transition de la chaîne de Markov à temps discret induite est simplifiée : $\mathbf{P}[(\rho, s), (\rho a o(s'), s')] = \nu(\rho, a) p(s'|s, a)$, et $\mathbf{P}$ est nulle ailleurs.

Exemple 5.2.6 (Un exemple de POMDP avec observation déterministe)**.** *La figure 5.8 décrit un* POMDP *avec observation déterministe. L'ensemble des observations est* $\Omega = \{ry, ps\}$*. On représente les observations par des rayures (ry) et des pois (ps) :* $o(q_0) = o(q_1) = ry$ *et* $o(q_2) = ps$*. Les arcs sont ici étiquetés*

comme dans les PA. *Par exemple, l'étiquette de l'arc de q_1 vers q_0 signifie que $p(q_0|q_1, a) = \frac{1}{4}$ et $p(q_0|q_1, b) = \frac{1}{2}$. Toutes les récompenses sont nulles (et omises), exceptée la récompense finale de q_2, qui est égale à 1.*

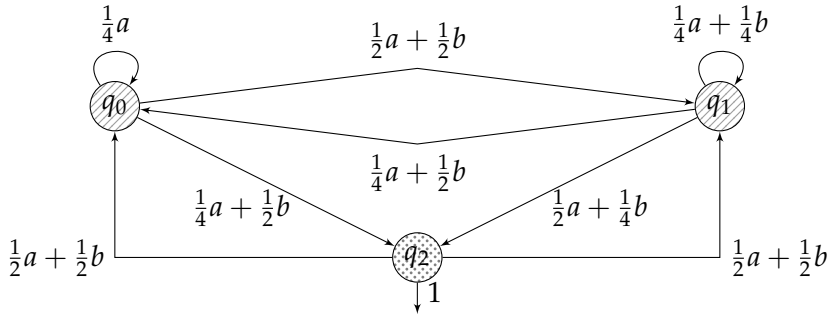


FIGURE 5.8 – *Un* POMDP *avec observation déterministe.*

On associe généralement trois types de récompenses à des exécutions.

Définition 5.2.7. *Soit $\mathcal{M}$ un* POMDP *et $\sigma = s_0 a_1 s_1 \dots a_i s_i \dots$ une exécution finie ou infinie. Lorsque σ est finie, n représente le nombre d'actions effectuées. Alors :*

- *Quand σ est finie, sa* récompense totale *$u(\sigma)$ est définie par $u(\sigma) = \sum_{i=1}^{n} r(s_{i-1}, a_i) + r_f(s_n)$;*
- *Quand σ est infinie, sa* récompense actualisée *$v_\lambda(\sigma)$ par rapport à $0 < \lambda < 1$ est définie par $v_\lambda(\sigma) = \sum_{i=1}^{\infty} r(s_{i-1}, a_i)\lambda^{i-1}$;*
- *Quand σ est infinie, sa* récompense moyenne par valeurs supérieures *$g_+(\sigma)$ est définie par $g_+(\sigma) = \limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} r(s_{i-1}, a_i)$.*

Notons X_n la variable aléatoire représentant l'état de $\mathcal{M}_\pi^\nu$ visité au $n^{ième}$ instant. De même, Y_n et $O_n = o(X_n)$ dénotent la variable aléatoire de l'action et de l'observation à cet instant.

Définition 5.2.8. *Soit $\mathcal{M}$ un* POMDP *avec π une distribution initiale sur S. Soit ν une politique et $0 < \lambda < 1$ un facteur d'actualisation. Alors :*

- *La* récompense totale espérée *u_t^ν au temps t sous la politique ν est définie par : $u_t^\nu = \sum_{i=0}^{n-1} \mathbf{E}^\nu(r(X_i, Y_i)) + \mathbf{E}^\nu(r_f(X_n))$;*
- *La* récompense actualisée espérée *v_λ^ν sous la politique ν est définie par :*
  $v_\lambda^\nu = \sum_{i=0}^{\infty} \mathbf{E}^\nu(r(X_i, Y_i))\lambda^i$;
- *La* récompense moyenne par valeurs supérieures espérée *g_+^ν sous la politique ν est définie par : $g_+^\nu = \limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{E}^\nu(r(X_i, Y_i))$.*

Nous sommes maintenant en mesure de spécifier les problèmes d'optimisation des POMDP. Soit $u_t^* = \sup_\nu(u_t^\nu)$, $v_\lambda^* = \sup_\nu(v_\lambda^\nu)$ et $g_+^* = \sup_\nu(g_+^\nu)$. On souhaite résoudre les problèmes suivants :

— Calculer u_t^* (respectivement v_λ^*, g_+^*).
— Trouver (si elle existe) une politique ν telle que $u_t^\nu = u_t^*$ (respectivement $v_\lambda^\nu = v_\lambda^*$, $g_+^\nu = g_+^*$).
— Etant donné $\varepsilon > 0$, calculer u_t (respectivement v_λ, g_+) tel que $|u_t^* - u_t| \leq \varepsilon$ (respectivement $|v_\lambda^* - v_\lambda| \leq \varepsilon$, $|g_+^* - g_+| \leq \varepsilon$) et trouver une politique correspondante.

5.2.2 Analyse des POMDP à horizon fini

Afin d'analyser le problème d'optimisation d'un POMDP à horizon fini, nous introduisons une distribution clé. Soit une histoire $h = a_0\omega_1 \dots a_{i-1}\omega_i$, alors $\pi_i^h(s)$ pour $s \in S$ dénote la probabilité que $X_i = s$, connaissant l'histoire h :

$$\pi_i^h(s) = \mathbf{Pr}(X_i = s \mid \bigwedge_{j \leq i} O_j = \omega_j \wedge \bigwedge_{j \leq i-1} Y_j = a_j).$$

Nous n'avons pas indiqué quelle politique ν avait été choisie. Tout d'abord, cette probabilité conditionnelle est bien définie pour toute politique ν telle que $x \stackrel{\text{def}}{=} \mathbf{Pr}^\nu(\bigwedge_{j \leq i} O_j = \omega_j \wedge \bigwedge_{j \leq i-1} Y_j = a_j) > 0$. Il est donc nécessaire que pour tout $0 \leq j < i$, $\nu(a_0\omega_1 \dots \omega_j)(a_j) > 0$. On observe alors que x et π_i^h (si ce vecteur est défini) sont indépendants du choix d'un tel ν. On peut donc sans perte de généralité supposer que $\nu(a_0\omega_1 \dots \omega_j)(a_j) = 1$. Pour simplifier les notations, on introduit la variable aléatoire de l'histoire à l'instant i : $H_i = Y_0O_1 \dots Y_{i-1}O_i$.

Le lemme suivant fournit un moyen de calculer la distribution π_i^h par valeurs de i croissantes.

Lemme 5.2.9. *Soient $\mathcal{M}$ un* POMDP *$h = a_0\omega_1 \dots a_{i-1}\omega_i$ un comportement observé, a_i une action et ω_{i+1} une observation pour lesquels il existe ν une politique telle que $\mathbf{Pr}^\nu(H_{i+1} = ha_i\omega_{i+1}) > 0$. Alors :*

— *Si $o(s) \neq \omega_{i+1}$, $\pi_{i+1}^{ha_i\omega_{i+1}}(s) = 0$;*
— *Si $o(s) = \omega_{i+1}$,*

$$\pi_{i+1}^{ha_i\omega_{i+1}}(s) = \frac{\sum_{s' \in S} \pi_i^h(s')p(s|s', a_i)}{\sum_{s''|o(s'')=\omega_{i+1}} \sum_{s' \in S} \pi_i^h(s')p(s''|s', a_i)}.$$

Démonstration. Nous établissons d'abord une équation intermédiaire.

$$\mathbf{Pr}(X_i = s' \mid H_i = h \wedge Y_i = a_i) = \frac{\mathbf{Pr}(X_i = s' \wedge Y_i = a_i \mid H_i = h)}{\mathbf{Pr}(Y_i = a_i \mid H_i = h)} = \pi_i^h(s') \quad (5.1)$$

puisque d'après nos observations sur le choix des politiques, le dénominateur est égal à 1. Établissons maintenant le lemme pour le cas $o(s) = \omega_{i+1}$ (l'autre cas étant trivial).

$$\pi_{i+1}^{ha_i\omega_{i+1}}(s) = \frac{\mathbf{Pr}(X_{i+1} = s \wedge O_{i+1} = \omega_{i+1} \mid H_i = h \wedge Y_i = a_i)}{\sum_{o(s'')=\omega_{i+1}} \mathbf{Pr}(X_{i+1} = s'' \wedge O_{i+1} = \omega_{i+1} \mid H_i = h \wedge Y_i = a_i)}$$

Le dénominateur est indépendant de s. Développons le numérateur N en fonction de l'état atteint au temps i.

$$\begin{aligned} N &\stackrel{\text{def}}{=} \mathbf{Pr}(X_{i+1} = s \wedge O_{i+1} = \omega_{i+1} \mid H_i = h \wedge Y_i = a_i) \\ &= \sum_{s' \in S} \mathbf{Pr}(X_i = s' \wedge X_{i+1} = s \wedge O_{i+1} = \omega_{i+1} \mid H_i = h \wedge Y_i = a_i). \end{aligned}$$

Nous ré-exprimons les termes non nuls de la somme à l'aide de probabilités conditionnelles. Afin d'alléger les notations, nous n'avons pas indiqué la restriction aux s' pour lesquels le terme associé est non nul.

$$\begin{aligned} N = \sum_{s' \in S} \Big(&\mathbf{Pr}(X_i = s' \mid H_i = h \wedge Y_i = a_i) \\ &\cdot \mathbf{Pr}(X_{i+1} = s \mid X_i = s' \wedge H_i = h \wedge Y_i = a_i) \\ &\cdot \mathbf{Pr}(O_{i+1} = \omega_{i+1} \mid X_{i+1} = s \wedge X_i = s' \wedge H_i = h \wedge Y_i = a_i) \Big). \end{aligned}$$

Le dernier facteur est égal à 1. À partir de l'équation (5.1) et de la sémantique d'un POMDP, on en déduit :

$$N = \sum_{s' \in S} \pi_i^h(s') p(s|s', a_i).$$

Il suffit alors de reporter dans le dénominateur pour conclure. □

Ce lemme motive l'introduction de $\mathbf{P}_{a,\omega}$, une matrice indicée par $S \times S$, et définie par $\mathbf{P}_{a,\omega}[s, s'] = p(s'|s, a)$ si $o(s') = \omega$ et $\mathbf{P}_{a,\omega}[s, s'] = 0$ sinon. Les matrices $\mathbf{P}_a$ sont définies comme précédemment. Observons que $\mathbf{P}_a = \sum_{\omega \in \Omega} \mathbf{P}_{a,\omega}$. Le lemme peut alors être réécrit ainsi :

$$\pi_i^h \cdot \mathbf{P}_{a_i,\omega_{i+1}} \cdot \mathbf{1} > 0 \quad \Longrightarrow \quad \pi_{i+1}^{ha_i\omega_{i+1}} = \frac{\pi_i^h \cdot \mathbf{P}_{a_i,\omega_{i+1}}}{\pi_i^h \cdot \mathbf{P}_{a_i,\omega_{i+1}} \cdot \mathbf{1}}.$$

Théorème 5.2.10. *Soit $\mathcal{M}$ un* POMDP *et k un entier représentant l'horizon. Alors l'algorithme 2 calcule un ensemble d'indices Z_k, une famille de vecteurs $\{\mathbf{r}_z\}_{z \in Z_k}$, et une famille de polyèdres $\{\mathbf{D}_z\}_{z \in Z_k}$ tels que $\bigcup_{z \in Z_k} \mathbf{D}_z$ soit égal à l'espace des distributions sur les états et tels que pour tout $\pi \in \mathbf{D}_z$, distribution initiale sur S, $u_k^*(\pi) = \pi \cdot \mathbf{r}_z$.*

Démonstration. Nous prouvons ce théorème par induction sur k. Par définition, pour $k = 0$, $u_0^* = \mathbf{r}_f \cdot \pi$. Par conséquent, $Z_0 = \{\varepsilon\}$, $\mathbf{r}_\varepsilon = r_f$ et $\mathbf{D}_\varepsilon$ est l'espace des distributions.

Supposons le résultat valide pour k, et introduisons le vecteur $\mathbf{b}_a$ défini par $\mathbf{b}_a[s] = r(s,a)$. On examine alors la première action à choisir et en s'appuyant sur le lemme précédent, on obtient :

$$u_{k+1}^*(\pi) = \max_{a\in A}\left(\pi\cdot\mathbf{b}_a + \sum_{\omega\in\Omega|\pi\cdot\mathbf{P}_{a,\omega}\cdot\mathbf{1}>0}(\pi\cdot\mathbf{P}_{a,\omega}\cdot\mathbf{1})u_k^*\left(\frac{\pi\cdot\mathbf{P}_{a,\omega}}{\pi\cdot\mathbf{P}_{a,\omega}\cdot\mathbf{1}}\right)\right).$$

Choisissons un polyèdre $\mathbf{D}_z = \{\pi \mid \forall 1 \le i \le n_z\ \pi\cdot\mathbf{c}_{z,i} \le k_{z,i}\}$ arbitraire de Z_k. Définissons le polyèdre :

$$\mathbf{D}_{a,\omega,z} = \{\pi \mid \pi\cdot\mathbf{1} = 1 \wedge \forall 1 \le i \le n_z\ \pi\cdot\mathbf{P}_{a,\omega}\cdot\mathbf{c}_{z,i} - k_{z,i}(\pi\cdot\mathbf{P}_{a,\omega}\cdot\mathbf{1}).\le 0\}.$$

Choisissons aussi une fonction $f : A\times\Omega \to Z_k$ et considérons une distribution π telle que pour tout (a,ω) soit $\pi\cdot\mathbf{P}_{a,\omega}\cdot\mathbf{1} = 0$ soit $\pi\in\mathbf{D}_{a,\omega,f(a,\omega)}$.

$$\begin{aligned}u_{k+1}^*(\pi) &= \max_{a\in A}\left(\pi\cdot\mathbf{b}_a + \sum_{\omega\in\Omega|\pi\cdot\mathbf{P}_{a,\omega}\cdot\mathbf{1}>0}(\pi\cdot\mathbf{P}_{a,\omega}\cdot\mathbf{1})\frac{\pi\cdot\mathbf{P}_{a,\omega}}{\pi\cdot\mathbf{P}_{a,\omega}\cdot\mathbf{1}}\cdot\mathbf{r}_{f(a,\omega)}\right)\\ &= \max_{a\in A}\left(\pi\cdot\mathbf{b}_a + \sum_{\omega\in\Omega}\pi\cdot\mathbf{P}_{a,\omega}\cdot\mathbf{r}_{f(a,\omega)}\right).\end{aligned}$$

On a pu enlever la restriction $\pi\cdot\mathbf{P}_{a,\omega}\cdot\mathbf{1} > 0$ pour les observations de la somme car les termes ajoutés sont nuls. Définissons enfin le polyèdre :

$$\mathbf{D}_{f,a} = \{\pi \mid \pi\in\bigcap_{a',\omega}\mathbf{D}_{a',\omega,f(a',\omega)} \wedge \pi\cdot(\mathbf{b}_a + \sum_{\omega\in\Omega}\mathbf{P}_{a,\omega}\cdot\mathbf{r}_{f(a,\omega)}) \ge \pi\cdot(\mathbf{b}_{a'} + \sum_{\omega\in\Omega}\mathbf{P}_{a',\omega}\cdot\mathbf{r}_{f(a',\omega)})\}$$

Notons $\mathbf{r}_{f,a} = \mathbf{b}_a + \sum_{\omega\in\Omega}\mathbf{P}_{a,\omega}\cdot\mathbf{r}_{f(a,\omega)}$. Nous laissons au lecteur le soin de vérifier que $\bigcup_{f,a}\mathbf{D}_{f,a}$ est l'ensemble des distributions et que pour tout $\pi\in\mathbf{D}_{f,a}$, $u_{k+1}^*(\pi) = \pi\cdot\mathbf{r}_{f,a}$. Autrement dit, $Z_{k+1} = Z_k^{A\times\Omega}\times A$. □

Comme la taille de l'ensemble des indices Z_k à considérer croît exponentiellement avec l'horizon k, il est intéressant de minimiser cette famille, en éliminant les domaines vides. En pratique cette minimisation est faite à

Algorithme 2 : Calcul des valeurs et politiques optimales à horizon fini.

`Optimal`$(\mathcal{M}, k)$
Input : $\mathcal{M}$, un POMDP ; k, un horizon

Output : $\mathbf{r}$, les formes linéaires à évaluer
pour les stratégies optimales à horizon $\leq h$
Output : $\mathbf{D}$, les domaines associés

Data : $\mathbf{P}_{a,\omega}$ des matrices sous-stochastiques ; ω, une observation
Data : a, une action ; z, une séquence d'actions ; s, s', des états
for $a \in A$ **do**
  for $s \in S$ **do** $\mathbf{b}_a[s] \leftarrow r(s,a)$

  for $\omega \in \Omega$ **do**
    for $s, s' \in S$ **do**
      if $o(s') = \omega$ **then** $\mathbf{P}_{a,\omega}[s,s'] = p(s'|s,a)$
      else $\mathbf{P}_{a,\omega}[s,s'] = 0$

$Z_0 \leftarrow \{\varepsilon\}$; $\mathbf{r}[\varepsilon] \leftarrow r_f$; $\mathbf{D}[\varepsilon] \leftarrow \{\pi \mid \pi \cdot \mathbf{1} = 1\}$
for i **from** 1 **to** k **do**
  $Z_i \leftarrow Z_{i-1}^{A\times\Omega} \times A$
  for $f \in Z_{i-1}^{A\times\Omega}$ **do**
    for $a \in A, \omega \in \Omega$ **do**
      Let $\mathbf{D}[f(a,\omega)] = \{\pi \mid \forall 1 \leq i \leq n\ \pi \cdot \mathbf{c}_i \leq k_i\}$
      $\mathbf{D}_{a,\omega,f(a,\omega)} \leftarrow \{\pi \mid \pi \cdot \mathbf{1} = 1 \wedge$
      $\forall 1 \leq i \leq n\ \pi \cdot \mathbf{P}_{a,\omega} \cdot \mathbf{c}_i - k_i(\pi \cdot \mathbf{P}_{a,\omega} \cdot \mathbf{1}) \leq 0\}$
    for $a \in A$ **do**
      $\mathbf{D}[f,a] \leftarrow \{\pi \mid \pi \in \bigcap_{a',\omega} \mathbf{D}_{a',\omega,f(a',\omega)} \wedge$
      $\pi \cdot (\mathbf{b}_a + \sum_{\omega\in\Omega} \mathbf{P}_{a,\omega} \cdot \mathbf{r}[f(a,\omega)]) \geq$
      $\pi \cdot (\mathbf{b}_{a'} + \sum_{\omega\in\Omega} \mathbf{P}_{a',\omega} \cdot \mathbf{r}[f(a',\omega)])\}$
      $\mathbf{r}[f,a] \leftarrow \mathbf{b}_a + \sum_{\omega\in\Omega} \mathbf{P}_{a,\omega} \cdot \mathbf{r}[f(a,\omega)]$
return$(\mathbf{D},\mathbf{r})$

la volée. Nous renvoyons à [30, 8, 16] pour plus de détails sur ces développements algorithmiques.

À titre d'exemple, illustrons le calcul de ces formes linéaires sur le POMDP de l'exemple 5.2.6. Décrivons d'abord les matrices $\mathbf{P}_{a,rg}$, $\mathbf{P}_{a,vt}$ $\mathbf{P}_{b,rg}$ et $\mathbf{P}_{a,vt}$.

$$\mathbf{P}_{a,rg} = \begin{pmatrix} \frac{1}{4} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix} \quad \mathbf{P}_{a,vt} = \begin{pmatrix} 0 & 0 & \frac{1}{4} \\ 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 \end{pmatrix} \quad \mathbf{P}_{b,rg} = \begin{pmatrix} 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix} \quad \mathbf{P}_{b,vt} = \begin{pmatrix} 0 & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{4} \\ 0 & 0 & 0 \end{pmatrix}$$

Les variables des formes linéaires seront notées $\mathbf{x} = (x_0, x_1, x_2)$. Par définition, $\mathbf{x} \cdot \mathbf{r}_\varepsilon = x_2$. Puisque Z_0 est un singleton, $Z_1 = A$. Plus précisément, $\mathbf{r}_a = \mathbf{b}_a + \mathbf{P}_{a,rg} \cdot \mathbf{r}_\varepsilon + \mathbf{P}_{a,vt} \cdot \mathbf{r}_\varepsilon = \mathbf{P}_{a,vt} \cdot \mathbf{r}_\varepsilon = (\frac{1}{4}, \frac{1}{2}, 0)$ (car les deux autres termes sont nuls). D'où $\mathbf{x} \cdot \mathbf{r}_a = \frac{1}{4}x_0 + \frac{1}{2}x_1$. Par un calcul analogue, $\mathbf{x} \cdot \mathbf{r}_b = \frac{1}{2}x_0 + \frac{1}{4}x_1$. On déduit aisément que $\mathbf{D}_a = \{\mathbf{x} \mid \mathbf{x} \cdot \mathbf{1} = 1 \wedge x_0 \leq x_1\}$ et $\mathbf{D}_b = \{\mathbf{x} \mid \mathbf{x} \cdot \mathbf{1} = 1 \wedge x_0 \geq x_1\}$. Soit $f \in A^{A \times \Omega}$ définie par $f(a, rg) = a$, $f(a, vt) = b$, $f(b, rg) = b$, $f(a, vt) = a$. Alors $\mathbf{r}_{f,a} = \mathbf{P}_{a,rg} \cdot \mathbf{r}_b + \mathbf{P}_{a,vt} \cdot \mathbf{r}_a = (\frac{1}{4}, \frac{3}{16}, \frac{3}{8}) + (0, 0, 0)$. Le calcul de $\mathbf{D}_{f,a}$ est déjà bien plus complexe...

5.2.3 Décidabilité et indécidabilité à horizon infini

Après avoir décrit comment optimiser le comportement d'un POMDP à horizon fini, examinons maintenant les questions d'optimisation à horizon infini. Comme l'analyse des PA l'a illustré, ces questions sont à la frontière de la décidabilité. Il est donc intéressant d'identifier des critères de difficulté des problèmes. Un critère naturel consiste à distinguer les problèmes *quantitatifs* dont l'énoncé comporte une valeur de probabilité ou d'espérance arbitraire (e.g. un seuil) et les problèmes *qualitatifs* dont l'énoncé ne comporte que des probabilités extrêmes (i.e. 0 ou 1). Les problèmes quantitatifs standard des POMDP sont indécidables. Dans le cas qualitatif, le critère déterminant est lié à la complexité de la condition de gain d'une exécution. Ainsi il est presque immédiat que le problème de l'accessibilité d'un état est plus simple que le problème de l'accessibilité répétée. Ceci peut se formaliser en considérant la hiérarchie qui définit la σ-algèbre des boréliens : l'accessibilité est spécifiée par un ouvert alors que l'accessibilité répétée est spécifiée par une intersection dénombrable d'ouverts. Une autre critère provient du fait qu'il existe deux valeurs extrêmes et qu'on peut ainsi définir des objectifs *multiples* vis à vis des seuils de probabilité. Nous détaillons maintenant les résultats les plus significatifs.

Problèmes quantitatifs

Le théorème suivant établit que la plupart des problèmes de décision relatifs aux problèmes d'optimisation sont indécidables.

Théorème 5.2.11 ([20]). *Soit $\mathcal{M}$ un* POMDP. *Les problèmes suivants sont indécidables.*

- *Étant donnée une valeur g_+, decider si $g_+^* > g_+$.*
- *Étant donnée une valeur v_λ, decider si $v_\lambda^* > v_\lambda$.*
- *Decider s'il existe une stratégie à mémoire finie ν telle que $v_\lambda^\nu = v_\lambda^*$.*
- *Pour $\varepsilon > 0$ fixé, et sachant que $g_+^* \in [0, \varepsilon] \uplus [1-\varepsilon, 1]$, décider si $g_+^* \in [0, \varepsilon]$.*

L'indécidabilité de ce dernier problème, appelé *ε-approximabilité*, montre que l'indécidabilité des problèmes quantitatifs est très robuste.

Le seul résultat de calculabilité connu pour les problèmes quantitatifs est une conséquence directe de l'analyse des POMDP à horizon fini et de la définition de la récompense actualisée espérée.

Proposition 5.2.12. *Soit $\mathcal{M}$ un* POMDP, $0 < \lambda < 1$ *un facteur d'actualisation et $\varepsilon > 0$. Alors on peut calculer une stratégie à mémoire finie ν telle que :*

$$v_\lambda^\nu \geq v_\lambda^* - \varepsilon$$

Problèmes qualitatifs

Nous considèrons dans la suite des problèmes qualitatifs. Plus précisément, on compare la probabilité maximale de satisfaire une condition borélienne aux constantes 0 et 1. Afin de spécifier les objectifs considérés, nous empruntons les notations de la logique temporelle linéaire LTL [21]. Étant donné $\mathcal{M}$ un POMDP, on s'intéresse aux conditions de gain des types suivants : accessibilité, sûreté, Büchi et co-Büchi. Commençons donc par introduire ces objectifs, qui sont toutes définies comme un sous-ensemble de Q^ω relativement à un sous-ensemble d'états $F \subseteq Q$.

- La condition d'*accessibilité*, notée $\Diamond F$ stipule que F doit être visité au moins une fois :

$$\Diamond F = \{q_0 q_1 q_2 \cdots \in Q^\omega \mid \exists n,\ q_n \in F\}$$

- La condition de *sûreté* $\Box F$ est duale de celle d'accessibilité ($\Box F = \neg\Diamond\neg F$) :

$$\Box F = \{q_0 q_1 q_2 \cdots \in Q^\omega \mid \forall n,\ q_n \in F\}$$

— La condition de *Büchi*, notée $\Box\Diamond F$ requiert que l'ensemble F soit visité un nombre infini de fois :

$$\Box\Diamond F = \{q_0 q_1 q_2 \cdots \in Q^\omega \mid \forall m\ \exists n \geq m,\ q_n \in F\}$$

— Enfin, la condition de *co-Büchi*, $\Diamond\Box F$, est comme son nom l'indique duale de celle de Büchi ($\Diamond\Box F = \neg\Box\Diamond\neg F$) :

$$\Diamond\Box F = \{q_0 q_1 q_2 \cdots \in Q^\omega \mid \exists m\ \forall n \geq m,\ q_n \in F\}$$

Avant de présenter les résultats concernant les POMDP pour les conditions de gain ci-dessus, on énonce un résultat général, dû à Chatterjee *et al.* [9].

Théorème 5.2.13. *Soit $\mathcal{M}$ un* POMDP *avec un espace d'états fini ou dénombrable et $\varphi \subseteq Q^\omega$ un objectif borélien. Pour toute stratégie ν, il existe une stratégie déterministe ν' telle que $\mathbf{Pr}^\nu(\mathcal{M} \models \varphi) \leq \mathbf{Pr}^{\nu'}(\mathcal{M} \models \varphi)$.*

Ce théorème exprime que, lorsqu'il existe une stratégie (*a priori* randomisée) pour un objectif donné avec un seuil de probabilité (par exemple accessibilité avec probabilité 1), alors il existe une stratégie déterministe qui réalise cet objectif avec ce même seuil.

Indécidabilité de problèmes qualitatifs

Théorème 5.2.14. *Le problème suivant est indécidable : étant donnés un* POMDP *$\mathcal{M}$ et un sous-ensemble de ses états $F \subseteq Q$, existe-t-il une stratégie ν assurant $\mathbf{Pr}^\nu(\mathcal{M} \models \Box\Diamond F) > 0$?*

L'indécidabilité peut déjà être établie pour la classe des POMDP *aveugles*, c'est-à-dire pour lesquels il y a une observation unique, indépendante de l'état courant. Il est même suffisant de prouver l'indécidabilité dans les POMDP aveugles restreints aux stratégies déterministes, grâce au théorème 5.2.13. Or, les POMDP aveugles munis d'une condition de Büchi et restreints aux stratégies déterministes (c.-à-d. des mots infinis) ne sont rien d'autre que des PBA. Le théorème 5.2.14 est alors une conséquence immédiate du théorème 5.1.31.

Intéressons nous à présent à la réalisation d'objectifs multiples. On cherche à décider de l'existence d'une stratégie assurant simultanément une propriété de Büchi presque sûrement et une propriété de sûreté avec probabilité positive. Au delà de son intérêt purement théorique, ce problème sera motivé par la suite dans le cadre du diagnostic de panne pour les systèmes contrôlables probabilistes.

Dans un premier temps, montrons que les stratégies assurant une telle combinaison d'objectifs peuvent être complexes. Plus précisément, sur l'exemple de la figure 5.9, il existe bien une stratégie qui satisfait les deux objectifs simultanément, mais aucune stratégie à mémoire finie ne convient. Le POMDP de la figure 5.9 n'a qu'une seule classe d'observation, le joueur est donc aveugle et ne prend ses décisions qu'en fonction du nombre d'étapes qui se sont produites. L'ensemble récurrent est $F = \{q_2, r_2\}$, et l'ensemble sûr est $I = \{q_1, q_2\}$.

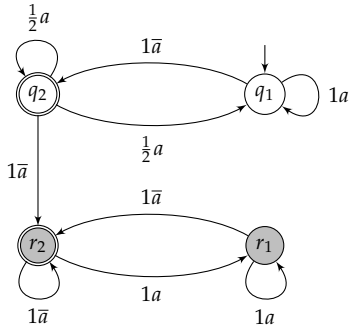


FIGURE 5.9 – *Un exemple nécessitant une mémoire infinie.*

Commençons par définir une stratégie, à mémoire infinie, qui garantit à la fois de visiter F infiniment souvent avec probabilité 1 et de rester dans I avec probabilité positive. Cette stratégie est déterministe et peut donc être décrite par un mot infini : $w = \overline{a}a^{k_1}\overline{a}a^{k_2}\overline{a}a^{k_3}\cdots$ où $\mathbf{k} = (k_i)_{i\in\mathbb{N}_{>0}}$ est une suite d'entiers naturels strictement positifs tels que $p_{\mathbf{k}} = \prod_{i>0}(1-2^{-k_i}) > 0$. Sous cette stratégie, la probabilité de rester dans la partie sûre est strictement positive, puisque c'est précisément $p_{\mathbf{k}}$. De plus, jouer infiniment souvent l'action $\overline{a}$ garantit de visiter infiniment souvent F presque sûrement.

Justifions maintenant que toute stratégie à mémoire finie échoue à satisfaire les deux objectifs simultanément. Soit ν une stratégie à mémoire finie. Puisque le POMDP est aveugle, ν est en fait une suite ultimement périodique de distributions d'actions : $\nu = \delta_1\cdots\delta_{n-1}(\delta_n\cdots\delta_m)^\omega$ avec $\delta_i \in \mathsf{Dist}(\{a,\overline{a}\})$ pour $i \in \{1,\cdots,m\}$ (où $\mathsf{Dist}(E)$ est l'ensemble des distributions à support dans E). Afin de satisfaire l'objectif de Büchi $\Box\Diamond F$ presque sûrement, l'action $\overline{a}$ doit être jouée infiniment souvent, avec probabilité 1. On en déduit qu'il existe $i \in \{n,\cdots,m\}$ telle que $\delta_i[\overline{a}] = p > 0$. Mais après $n+(k+1)m$ actions, la probabilité de rester dans I est inférieure ou égale à $(1-\frac{p}{2^{m-n}})^k$. Par conséquent, la probabilité de rester toujours dans

I est nulle. La combinaison des deux objectifs n'est donc pas réalisable par une stratégie à mémoire finie.

En s'appuyant sur cet exemple on établit le théorème suivant a priori surprenant, puisque indépendamment, les problèmes d'existence d'une stratégie assurant une condition de Büchi presque sûrement ou une condition de sûreté positivement, sont décidables pour les POMDP (voir plus loin le théorème 5.2.18 et la proposition 5.2.17). Cependant, imposer qu'une même stratégie satisfasse les deux conditions mène à un problème indécidable.

Théorème 5.2.15 ([5])**.** *Le problème suivant est indécidable : étant donnés un* POMDP $\mathcal{M}$ *et deux sous-ensembles de ses états* $F, I \subseteq Q$*, existe-t-il une stratégie* ν *telle que* $\mathbf{Pr}^{\nu}(\mathcal{M} \models \Box\Diamond F) = 1$ *et* $\mathbf{Pr}^{\nu}(\mathcal{M} \models \Box I) > 0$ *?*

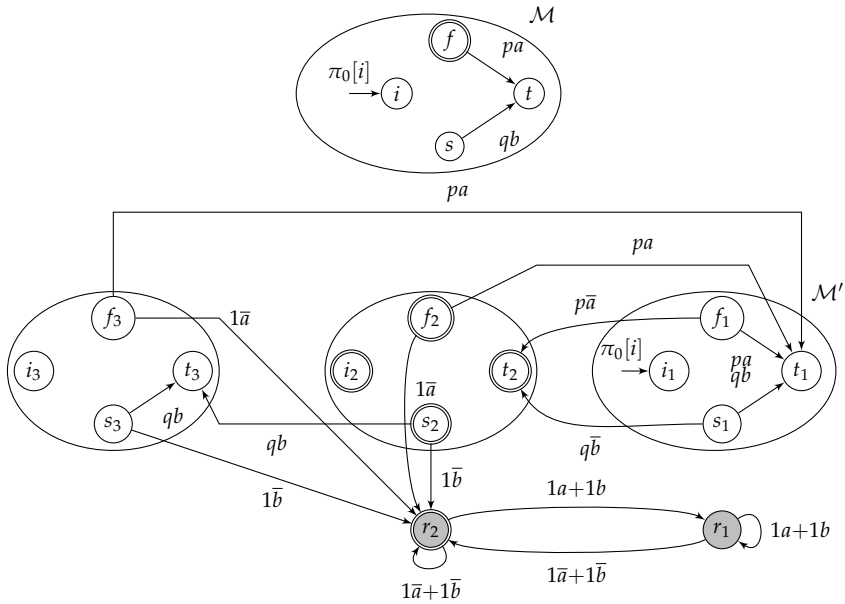


FIGURE *5.10 – Réduction pour le théorème 5.2.15.*

Démonstration. On réduit le problème d'existence d'une stratégie assurant un objectif de Büchi avec probabilité positive dans un POMDP aveugle. La réduction est inspirée de l'exemple présenté en figure 5.9 et est illustrée en figure 5.10.

$\mathcal{M}$ est un POMDP aveugle d'ensemble d'actions $\{a, b\}$ et pour lequel l'ensemble à visiter infiniment souvent est F (avec $f \in F$). À partir de $\mathcal{M}$,

on construit le POMDP aveugle $\mathcal{M}'$ en dupliquant l'ensemble des actions en $\{a, b, \overline{a}, \overline{b}\}$, et en utilisant trois copies de $\mathcal{M}$ (dont les sous-ensembles d'états sont notés Q_1, Q_2, Q_3) et deux états supplémentaires r_1, r_2. Dans $\mathcal{M}'$, l'ensemble récurrent est constitué de la deuxième copie de $\mathcal{M}$ (à gauche), ainsi que de l'état r_2, et est représenté par les états entourés deux fois ; l'ensemble I sûr regroupe tous les états non grisés, c'est-à-dire, tous sauf r_1 et r_2. Le passage de la première copie (à droite) à la deuxième se fait par les actions $\overline{\alpha}$ pour $\alpha \in A$. Ces mêmes actions, depuis les autres copies font quitter la zone sûre, en menant à l'état r_2. Depuis un état qui était récurrent pour $\mathcal{M}$ dans la deuxième et la troisième copies (ici f_2, f_3), les actions a et b permettent de retourner à la première copie. Depuis un état qui n'était pas récurrent pour $\mathcal{M}$ dans la deuxième copie (ici s_2), les actions a et b conduisent à la troisième copie. Enfin, dans la zone non sûre, les actions a et b mènent à r_1, tandis que les $\overline{a}$ et $\overline{b}$ ont pour destination r_2.

Nous observons d'abord les faits suivants. Une séquence infinie de $\mathcal{M}'$ visite infiniment souvent F' si et seulement si $\{\overline{a}, \overline{b}\}$ apparaît infiniment souvent. Cette séquence ne visite jamais $\{r_1, r_2\}$ si lorsque ces actions apparaissent, l'état courant appartient à la première copie. Enfin on rejoint la première copie depuis les autres copies uniquement par une action $\{a, b\}$ à partir d'une copie d'un état de F.

Supposons qu'il existe une stratégie ν dans $\mathcal{M}$ telle que $\mathbf{Pr}^{\nu}(\mathcal{M} \models \Box\Diamond F) = p > 0$. Cette stratégie peut être choisie déterministe. Par conséquent, elle est décrite par un mot infini $w = w_1 \dots w_n \dots$ Choisissons une suite infinie $(\beta_j)_{j\in\mathbb{N}}$ avec $0 < \beta_j < 1$ telle que $\prod_{j\geq 0} \beta_j > 0$. Nous construisons itérativement une séquence infinie strictement croissante d'entiers $(n_j)_{j\in\mathbb{N}}$ comme suit. On pose $n_0 = 0$, et si $n_0, \dots, n_j$ ont été définis, $n_{j+1} > n_j$ est le plus petit entier qui satisfait :

$$\mathbf{Pr}^{\nu}\Big(\mathcal{M} \models \Diamond^{[n_j+1, n_{j+1}]} F \mid \mathcal{M} \models \bigwedge_{k=0}^{i} \Diamond^{[n_k+1, n_{k+1}]} F \wedge \Box\Diamond F\Big) \geq \beta_j \ ,$$

où la notation $\Diamond^{[m,M]} F$ signifie que F est visité entre les instants m et M. En raison de la propriété de ν, et de l'hypothèse de récurrence, la probabilité conditionnelle est bien définie, et elle tend vers 1 lorsque n_{j+1} tend vers l'infini. Ainsi n_{j+1} est bien définie. Par construction :

$$\mathbf{Pr}^{\nu}\Big(\mathcal{M} \models \bigwedge_{j\geq 0} \Diamond^{[n_j+1, n_{j+1}]} F\Big) \geq p \prod_{j\geq 0} \beta_j > 0 \ .$$

Définissons la stratégie déterministe ν' décrite par $w' = w'_1 \dots w'_n \dots$ dans $\mathcal{M}'$. A un instant k différent de tout n_j, $w'_k = w_k$ et pour tout n_j, $w'_{n_j} = \overline{w}_{n_j}$.

Puisque $\{\overline{a}, \overline{b}\}$ est sélectionné infiniment souvent, F' est visité infiniment souvent par toute exécution engendrée par ν'. De plus la probabilité que pour tout j, à l'instant n_j l'état courant soit dans Q_1, est au moins égale à $p \prod_{j \geq 0} \beta_j$. Par conséquent, avec au moins cette probabilité, la séquence aléatoire dans $\mathcal{M}'$ engendrée par ν restera toujours dans $Q_1 \cup Q_2 \cup Q_3$. ν' atteint donc les objectifs du théorème.

Supposons maintenant qu'il existe ν' une stratégie de $\mathcal{M}'$ qui atteint les objectifs du théorème. Puisque $\mathcal{M}'$ est aveugle, ν' correspond à une séquence de distributions $\delta'_1, \delta'_2, \ldots$ avec $\delta'_i \in \mathsf{Dist}(\{a, b, \overline{a}, \overline{b}\})$. Alors ν est une stratégie aléatoire définie par $\delta_1, \delta_2, \ldots$ avec $\delta_i[\alpha] = \delta'_i[\alpha] + \delta'_i[\overline{\alpha}]$ pour $\alpha \in \{a, b\}$. Considérons l'ensemble des séquences de $\mathcal{M}'$ qui visitent infiniment souvent F' et jamais $\{r_1, r_2\}$. Cet ensemble a une probabilité positive sous la stratégie ν'. Si on « projette » l'ensemble de ces séquences en oubliant la copie et le surlignement, alors l'ensemble obtenu a même probabilité dans $\mathcal{M}$ sous la stratégie ν et il visite infiniment souvent F. □

Décidabilité de problèmes qualitatifs

Commençons par quelques résultats relativement faciles.

Proposition 5.2.16. *Le problème suivant est* NLOGSPACE*-complet : étant donnés un* POMDP $\mathcal{M}$ *et un sous-ensemble de ses états* $F \subseteq Q$*, existe-t-il une stratégie* ν *assurant* $\mathbf{Pr}^{\nu}(\mathcal{M} \models \Diamond F) > 0$ *?*

Démonstration. La preuve se base sur le fait que l'existence d'une stratégie assurant $\mathbf{Pr}^{\nu}(\mathcal{M} \models \Diamond F) > 0$ est équivalente à l'existence d'un chemin menant d'un état initial à un état cible. L'existence d'un tel chemin est clairement une condition nécessaire. C'est également une condition suffisante, car alors la simple stratégie qui consiste à jouer à chaque étape une distribution uniforme sur toutes les actions du POMDP assigne à ce chemin une probabilité non nulle. □

Pour ce qui concerne les objectifs de sûreté, on a :

Proposition 5.2.17. *Les problèmes suivants sont* EXPTIME*-complets : étant donnés un* POMDP $\mathcal{M}$ *et un sous-ensemble de ses états* $F \subseteq Q$

— *existe-t-il une stratégie* ν *assurant* $\mathbf{Pr}^{\nu}(\mathcal{M} \models \Box F) = 1$ *?*
— *existe-t-il une stratégie* ν *assurant* $\mathbf{Pr}^{\nu}(\mathcal{M} \models \Box F) > 0$ *?*

Démonstration. Considérons tout d'abord le cas d'une condition de sûreté avec probabilité 1. Pour cela, on construit le jeu à deux joueurs à observation parfaite associé à un POMDP (et à un ensemble d'états F) en utilisant les *croyances* déduites des observations. Étant donnés $\mathcal{M} = (Q, \Omega, A, o, p)$

et $F \subseteq Q$, on définit le jeu à deux joueurs et sous observation parfaite $\mathcal{G}_{\mathcal{M}} = (\mathsf{C}, A, \Delta)$ où $\mathsf{C} = 2^Q \setminus \{\emptyset\}$ est l'ensemble des croyances possibles. Pour $c \in \mathsf{C}$ une croyance et $a \in A$ une action, on note $\mathsf{Post}(c, A) = \{q' \in Q \mid \exists q \in c,\ p(q'|q, a) > 0\}$. La relation de transition $\Delta \subseteq \mathsf{C} \times A \times \mathsf{C}$ satisfait alors $(c_1, a, c_2) \in \Delta$ s'il existe une observation $o \in \Omega$ telle que $c_2 = \mathsf{Post}(c_1, a) \cap o$. Le jeu se déroule de la façon suivante : depuis un état $c_1 \in \mathsf{C}$, le premier joueur choisit une action $a \in A$, et son adversaire choisit un successeur c_2 tel que $(c_1, a, c_2) \in \Delta$. L'objectif de sûreté $\Box F$ dans le POMDP est transformé en un objectif de sûreté $\Box \tilde{F}$ dans le jeu à deux joueurs, avec $\tilde{F} = \{c \in \mathsf{C} \mid c \subseteq F\}$.
Cette construction fondée sur les croyances déduites des observations assure qu'il existe une stratégie ν dans $\mathcal{M}$ telle que $\mathbf{Pr}^{\nu}(\mathcal{M} \models \Box F) = 1$ depuis la distribution initiale π_0 si et seulement si le premier joueur a une stratégie dans $\mathcal{G}_{\mathcal{M}}$ depuis $\{q \mid \pi_0(q) > 0\}$ qui garantit $\Box \tilde{F}$ quels que soient les choix de son adversaire. Les jeux de sûreté à deux joueurs et sous observation parfaite sont résolubles en temps polynomial (cf, par exemple [14]). Par conséquent, le problème de l'existence d'une stratégie assurant $\mathbf{Pr}_{\nu}(\mathcal{M} \models \Box F) = 1$ est dans EXPTIME.
La borne inférieure de complexité est obtenue en observant que les jeux de sûreté partiellement observables sont EXPTIME-difficiles [7].

Venons en à présent aux conditions de sûreté avec probabilité positive. Pour cela, montrons que l'on peut utiliser le cas précédent comme sous-routine. Plus précisément, il existe une stratégie ν assurant $\mathbf{Pr}^{\nu}(\mathcal{M} \models \Box F) > 0$ dans $\mathcal{M}$ depuis π_0 si et seulement si il existe un état q tel que (a) il existe une stratégie ν^1 depuis π_0 assurant $\mathbf{Pr}^{\nu^1}(\mathcal{M} \models F \ \mathsf{Until}\ q) > 0$, et (b) il existe une stratégie ν^2 depuis q assurant $\mathbf{Pr}^{\nu^2}(\mathcal{M} \models \Box F) = 1$. La formule $F \ \mathsf{Until}\ q$ est satisfaite par les séquences telles que F reste vrai jusqu'à ce que q le devienne.

L'implication de droite à gauche est claire : il suffit de combiner ν^1 et ν^2 pour construire une stratégie assurant l'objectif de sûreté avec probabilité 1. On bascule de ν_1 dès qu'on atteint une croyance contenant q.

Détaillons donc l'autre implication, et montrons même que l'on peut prendre $\nu^1 = \nu$. Sans perte de généralité on suppose que les états hors de F sont absorbants. On raisonne alors par l'absurde en supposant que pour tout état q atteint avec probabilité positive depuis q_0 sous ν, il n'existe pas de stratégie ν^q depuis q telle que $\mathbf{Pr}^{\nu^q}(\mathcal{M} \models \Box F) = 1$. En particulier, pour $N = |Q|$, on peut borner, indépendamment de q et de ν^q, la probabilité de rester dans F pendant N étapes par une constante $1 - \gamma^N$, où γ est la plus petite valeur de probabilité qui apparaît dans le POMDP $\mathcal{M}$. La probabilité de rester dans F pendant kN étapes sous ν est donc bornée par $(1 - \nu^N)^k$,

qui tend vers 0 lorsque $k \in \mathbb{N}$ croît, ce qui contredit le fait que ν gagne positivement l'objectif de sûreté.

Pour décider si depuis la distribution initiale π_0 il existe une stratégie ν assurant $\mathbf{Pr}^\nu(\mathcal{M} \models \Box F) > 0$, on procède donc de la façon suivante : on commence par calculer l'ensemble $\mathsf{Win}_{=1}$ des états $q \in Q$ tels que depuis q il existe une stratégie réalisant l'objectif de sûreté $\Box F$ avec probabilité 1. On décide ensuite si depuis π_0, il existe une stratégie qui permet d'atteindre $\mathsf{Win}_{=1}$ avec probabilité positive tout en restant dans F. Puisque que l'on a supposé que l'ensemble $Q \setminus F$ est absorbant, pour cette deuxième étape, il suffit de vérifier que $\mathsf{Win}_{=1}$ est accessible depuis l'état initial, comme expliqué dans la preuve de la proposition 5.2.16. Ces deux étapes fournissent un algorithme EXPTIME pour décider si depuis q_0 il existe une stratégie ν telle que $\mathbf{Pr}^\nu(\mathcal{M} \models \Box F) > 0$.
Pour prouver la borne inférieure, on peut réduire le problème d'acceptation par une machine de Turing alternante en espace polynomial [10]. □

Intéressons nous pour terminer à la question de l'existence d'une stratégie assurant un objectif de Büchi presque sûrement.

Théorème 5.2.18. *Le problème suivant est* EXPTIME*-complet : étant donnés un* POMDP $\mathcal{M}$ *et un sous-ensemble de ses états* $F \subseteq Q$*, existe-t-il une stratégie* ν *assurant* $\mathbf{Pr}^\nu(\mathcal{M} \models \Box\Diamond F) = 1$ *?*

Démonstration. De façon similaire à la preuve de la proposition 5.2.17, on considère l'automate des croyances associé à un POMDP. Pour $\mathcal{M} = (Q, \Omega, A, o, p)$, on construit $\mathcal{A}_\mathcal{M} = (\mathsf{C}, A \times \Omega, \Delta)$ tel que $\mathsf{C} = 2^Q$ est l'ensemble des croyances possibles ($\emptyset$ est introduit pour faciliter la spécification de la procédure), et la relation de transition $\Delta \subseteq \mathsf{C} \times (A \times \Omega) \times \mathsf{C}$ est déterministe et définie par :

$$\Delta(c, (a, o)) = \bigcup_{q \in c} \{q' \in Q \mid p(q, a, q') > 0\} \cap o.$$

Dans ce qui suit, pour simplifier les notations, on écrit $\Delta(c, (a_1, o_1) \cdots (a_n, o_n))$ l'application successive de la fonction de transition, c'est-à-dire $\Delta(\cdots \Delta(c, (a_1, o_1)), \cdots (a_n, o_n))$.

En utilisant l'automate des croyances, on calcule l'ensemble Win des croyances gagnantes, c'est-à-dire l'ensemble des supports de distributions depuis lesquels il existe une stratégie assurant l'objectif de Büchi presque sûrement. Notons que cet ensemble est bien défini : pour deux distributions ayant même support, une stratégie commune est gagnante presque sûrement (resp. positivement) depuis les deux ou depuis aucune. En effet : $\mathbf{Pr}^\nu_{\pi_0}(\mathcal{M} \models \varphi) = \sum_{q \in Q} \pi_0(q) \cdot \mathbf{Pr}^\nu_q(\mathcal{M} \models \varphi)$. L'ensemble Win est défini

comme un plus grand point fixe, en partant de $\mathsf{Win}^0 = \mathsf{C}$, et en calculant Win^{i+1} à partir de Win^i comme suit :

$$\mathsf{Win}^{i+1} = \{c \in \mathsf{Win}^i \mid \exists (a_1, o_1) \cdots (a_n, o_n),\ \emptyset \neq \Delta(c, (a_1, o_1), \cdots (a_n, o_n)) \subseteq F \\ \text{et } \forall k,\ \forall o'_k,\ \Delta(c, (a_1, o_1), \cdots (a_k, o'_k)) \in \mathsf{Win}^i\}\}.$$

On se convainc assez facilement que si depuis une distribution initiale, il existe une stratégie gagnante, alors son support appartient au point fixe Win. Clairement, l'existence d'un chemin γ_c menant de c à F est nécessaire pour pouvoir atteindre F presque sûrement. De plus, quelles que soient les observations qui font dévier γ_c, depuis la distribution obtenue, la stratégie doit à nouveau permettre de satisfaire l'objectif de Büchi presque sûrement.

Définissons une stratégie ν, qui assure $\mathbf{Pr}^\nu(\mathcal{M} \models \Box\Diamond F) = 1$ depuis n'importe quelle distribution initiale ayant pour support une croyance de Win. Cette stratégie ne dépend pas des distributions précises, mais seulement de leur support. À partir de $c \in \mathsf{Win}$, idéalement, la stratégie cherche à réaliser le chemin $\gamma_c = c \xrightarrow{a_1,o_1} \cdots \xrightarrow{a_n,o_n}$ qui mène à F. La stratégie possède autant de « modes » qu'il y a de croyances dans Win, et dans le mode correspondant à c, elle cherche à réaliser γ_c. Depuis c, la première décision est donc de jouer a_1, et si l'observation est o_1, de continuer avec a_2, etc. tant que les observations sont conformes à celles décrites par le chemin menant à F. Si, à l'étape k, l'observation est o'_k au lieu de l'observation attendue o_k, par définition du point fixe, la nouvelle croyance c' est toujours dans Win. La stratégie passe donc dans le mode correspondant à c', et cherche à atteindre F par $\gamma_{c'}$, le chemin associé à c'. □

Le tableau qui suit récapitule les résultats de complexité dans les POMDP pour les différentes questions qualitatives, selon l'objectif considéré.

	Accessibilité	Sûreté	Büchi	co-Büchi
$\mathbf{Pr} > 0$	NLOGSPACE-c.	EXPTIME-c.	indécidable	EXPTIME-c.
$\mathbf{Pr} = 1$	EXPTIME-c.	EXPTIME-c.	EXPTIME-c.	indécidable

FIGURE 5.11 – *Complexité des problèmes qualitatifs pour les* POMDP.

Application au diagnostic de panne

Comme application des problèmes théoriques sur les POMDP étudiés dans cette section, nous présentons maintenant le diagnostic de panne.

En théorie du contrôle, le diagnostic consiste à détecter l'occurrence de fautes dans des systèmes partiellement observables. Plus précisément, nous considérons un système sujet aux pannes, et qu'un utilisateur extérieur peut observer, au moins de façon partielle, et qu'il cherche à contrôler pour pouvoir détecter les fautes. Pour cela, partons d'un modèle probabiliste de système à événement discret, dans lequel certains événements sont observables, d'autres non, et certains événements sont contrôlables et d'autres non.

Définition 5.2.19. *Un système probabiliste à événement discret* $\mathcal{S} = (S, s_0, \mathcal{E}, \Delta)$ *est défini par la donnée de* S *un ensemble fini d'états dont* $s_0 \in S$ *l'état initial,* $\mathcal{E}$ *un ensemble d'événements et* $\Delta : S \times \mathcal{E} \times S \to [0,1]$ *une fonction de transition probabiliste telle que : pour tout* $s \in S$, $\sum_{a \in \mathcal{E},\ s' \in S} p(s,a,s') = 1$.

Remarquons que les probabilités sortantes d'un état ont pour somme 1. Autrement dit, un système probabiliste à événement discret est une DTMC lorsqu'on oublie les actions.

Afin de spécifier le problème de diagnostic de panne, l'ensemble des événements $\mathcal{E}$, est partitionné de deux façons : entre les événements observables et inobservables $\mathcal{E} = \mathcal{E}_o \sqcup \mathcal{E}_{no}$ et entre les événements contrôlables et incontrôlables $\mathcal{E} = \mathcal{E}_c \sqcup \mathcal{E}_{nc}$ avec $\mathcal{E}_{no} \subseteq \mathcal{E}_{nc}$. De plus $\mathcal{E}$ contient un événement inobservable particulier, $\mathbf{f} \in \mathcal{E}_{no}$ qui désigne la faute (ou erreur, ou panne).

L'utilisateur du système n'observe que partiellement son comportement : étant donnée une exécution, il ne perçoit que la suite des événements observables qui la compose. On dit qu'une telle suite d'observations est *sûrement fautive* si toutes les exécutions du système qui peuvent la générer contiennent l'événement fautif $\mathbf{f}$. Symétriquement, elle est *sûrement correcte* si toutes les exécutions qui la génèrent ne contiennent pas $\mathbf{f}$. Dans les autres cas, elle est dite *ambiguë*.

Le problème de contrôle afin d'assurer la diagnosticabilité du système est intuitivement le suivant : peut-on dynamiquement interdire certains événements contrôlables, de façon à ce que les exécutions d'observation ambiguë soient de mesure nulle. De plus le contrôle ne doit pas bloquer le système : autrement dit depuis tout état, au moins une transition reste franchissable.

La figure 5.12 représente un exemple de système probabiliste à événement discret, pour lequel $\mathcal{E}_c = \mathcal{E}_o = \{a,b,c,d,e\}$, et $\mathcal{E}_{nc} = \mathcal{E}_{no} = \{\mathbf{f}, u\}$. Sur cet exemple, $adcb^\omega$ est une suite d'observations sûrement fautive, acb^ω est sûrement correcte, tandis que $aadcb^\omega$ est ambiguë puisqu'elle correspond à une exécution fautive et une exécution correcte ayant respectivement pour préfixes $s_0 \xrightarrow{\mathbf{f}} s_1 \xrightarrow{a} s_2 \xrightarrow{a} s_2$ et $s_0 \xrightarrow{u} s_3 \xrightarrow{a} s_4 \xrightarrow{a} s_2$.

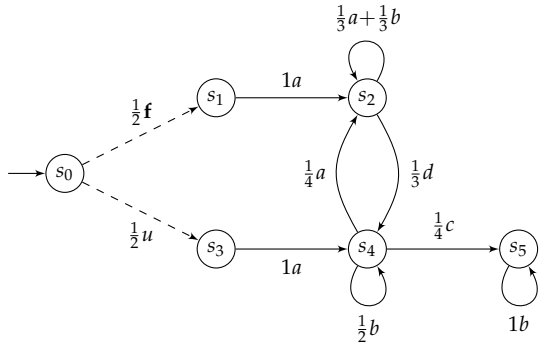


FIGURE 5.12 – *Un système probabiliste à événement discret.*

Les stratégies du contrôleur pour le problème de diagnostic probabiliste prennent la forme suivante : pour chaque suite finie d'événements observables, un choix est fait sur l'ensemble des événements contrôlables autorisés. Ce choix pouvant être « randomisé », formellement, une stratégie est une fonction $\nu : \mathcal{E}_o^* \to \mathsf{Dist}(\mathcal{E}_c)$.

Sur notre exemple, une stratégie possible de contrôle, particulièrement simple et non randomisée, est la suivante : $\nu(\varepsilon) = \mathcal{E}_c$, et $\nu(av) = \mathcal{E}_c \setminus \{a\}$ pout tout $v \in \mathcal{E}_o^*$. Cette stratégie interdit donc l'occurrence de l'événement a, sauf la toute première. Ainsi, elle exclut l'observation ambiguë $aadcb^\omega$ discutée plus haut. La seule observation ambiguë restante dans ce système contrôlé est ab^ω, et elle correspond à un ensemble d'exécutions de probabilité nulle. Remarquons que cette stratégie n'introduit pas non plus de blocage.

Le contrôle pour le diagnostic probabiliste a de fortes ressemblances avec les problèmes que l'on a étudiés pour les POMDP. En effet, dans les deux cas, le modèle est probabiliste, et un utilisateur partiellement informé cherche à contrôler le système pour assurer un objectif donné. Reformuler le diagnostic probabiliste pour les systèmes à événements discrets, en un problème de décision décidable sur les POMDP, soulève plusieurs difficultés. Tout d'abord, dans le cadre du diagnostic, le contrôleur a la possibilité d'interdire des événements, alors que dans un POMDP, la stratégie choisit quelle action effectuer. De plus, pour ce qui concerne le diagnostic, le contrôleur observe certains événements (précisément ceux qui sont observables) alors que dans un POMDP l'information est liée aux états visités. La première difficulté est levée en définissant comme actions du POMDP les sous-ensembles d'événements qui comprennent tous les événements non contrôlables. Pour s'affranchir de la seconde difficulté, on s'appuie sur un

automate déterministe avec condition d'acceptation de Büchi, qui reconnaît exactement les séquences observables non ambiguës [15]. Enfin une troisième difficulté vient du fait que le contrôleur ne reprend la main qu'après un événement observable. Aussi, une transition du POMDPcorrespond à une séquence d'événements inobservables suivis d'un événement observable.

Sans détailler sa correction, nous présentons brièvement la structure de cet automate déterministe de Büchi $\mathcal{B}_{\mathcal{S}}$ associé à un PLTS $\mathcal{S}$. Ses états sont des triplets (U, V, W) de sous-ensembles d'états de $\mathcal{S}$, tels que $U \cup V \cup W \neq \emptyset$ et $V \cap W = \emptyset$. L'alphabet est $\mathcal{E}_o$, de manière à ce que cet automate reconnaisse des séquences observables. Intuitivement, si une séquence $v \in \mathcal{E}_o^*$ mène à un état (U, V, W), la composante U représente l'ensemble des états accessibles par une exécution correcte d'observation v, tandis que $V \cup W$ est l'ensemble des états accessibles par une séquence fautive d'observation v. De plus, W contient les états correspondant à des fautes dont les plus anciennes, alors que V fait office de « salle d'attente » pour des fautes plus récentes. Les séquences observables menant à (U, V, W) avec $U \neq \emptyset$ et $V \cup W \neq \emptyset$ sont donc ambiguës, et l'automate cherche d'abord à résoudre l'ambiguïté entre U et W, puis celle posée par les fautes plus récentes (et donc stockées dans V). L'ensemble des états acceptants de $\mathcal{B}_{\mathcal{S}}$ est formé de ceux tels que $U = \emptyset$ ou $W = \emptyset$, qui correspondent donc respectivement à des séquences nécessairement fautives, ou des séquences dont l'ambiguïté la plus ancienne a été levée.

En construisant une forme de produit du PLTS $\mathcal{S}$ avec l'automate déterministe $\mathcal{B}_{\mathcal{S}}$, on obtient un POMDP $\mathcal{M}_{\mathcal{S}}$ dont on note Desamb l'ensemble des états dont la deuxième composante est un état acceptant de $\mathcal{B}_{\mathcal{S}}$ (c'est-à-dire pour lesquels l'une au moins des composantes U ou W est vide).

Proposition 5.2.20. *$\mathcal{S}$ est activement diagnosticable si et seulement si il existe une stratégie ν dans $\mathcal{M}_{\mathcal{S}}$ telle que* $\mathbf{Pr}_{\nu}(\mathcal{M}_{\mathcal{S}} \models \Box\Diamond\mathsf{Desamb}) = 1$

En utilisant le théorème 5.2.18, on en déduit un algorithme de décision pour le problème de diagnostic actif probabiliste. On peut même établir la complexité exacte de ce problème :

Théorème 5.2.21. *Le problème de contrôle pour le diagnostic de systèmes probabilistes à événements discrets est* EXPTIME*-complet.*

La borne supérieure en EXPTIME peut paraître surprenante. En effet, $\mathcal{M}_{\mathcal{S}}$ est un POMDP de taille exponentielle en la taille de $\mathcal{S}$, et l'algorithme de résolution pour une condition de Büchi presque sûre dans les POMDP est en EXPTIME, puisqu'il repose sur la construction de l'ensemble des

croyances. Cependant, dans le POMDP $\mathcal{M}_{\mathcal{S}}$ que nous considérons, les croyances du système $\mathcal{S}$ sont déjà présentes dans l'état $((U, V, W), s)$ sous la forme $U \cup V \cup W$. Ainsi, la deuxième exponentielle lors de la résolution du POMDP peut être évitée, et le problème du diagnostic actif reste dans EXPTIME. Pour la borne inférieure, on peut réduire la résolution de jeux de sûreté partiellement observables qui est un problème EXPTIME-difficile [7].

Une façon de contrôler un système pour qu'il soit diagnosticable, est de forcer (presque sûrement) l'occurrence d'une faute. On peut alors même prédire qu'une faute aura lieu avant qu'elle se produise. Évidemment, un tel contrôleur présente peu d'intérêt et il est pertinent de raffiner le problème de contrôle pour le diagnostic en *diagnostic actif sauf* : existe-t-il une stratégie qui assure que le système soit à la fois presque sûrement diagnosticable, et non fautif avec probabilité positive ? En utilisant la même approche que pour le diagnostic actif, on peut reformuler le problème du diagnostic actif sauf pour $\mathcal{S}$ en l'existence d'une stratégie dans le POMDP $\mathcal{M}_{\mathcal{S}}$ qui garantisse en même temps un objectif de Büchi avec probabilité 1 et un objectif de sûreté avec probabilité positive. L'objectif de Büchi est comme avant $\Box\Diamond\mathsf{Desamb}$, tandis que l'objectif de sûreté est $\Box\mathsf{Correct}$ où Correct correspond aux états $(q, (U, V, W))$ de $\mathcal{M}_{\mathcal{S}}$ tels que $V \cup W = \emptyset$. Malheureusement, le problème auquel on se réduit est indécidable, comme on l'a vu avec le théorème 5.2.15. En outre, en adaptant la preuve de ce théorème, on peut montrer que le problème de diagnostic actif sauf pour les systèmes probabilistes est indécidable [5]. Cependant en se limitant aux stratégies à mémoire finie (une condition nécessaire pour implémenter un contrôleur) ce problème devient EXPTIME-complet.

5.3 Jeux stochastiques partiellement observables

5.3.1 Présentation

Afin de généraliser le modèle des POMDP étudié dans la section précédente, on s'intéresse maintenant aux jeux stochastiques à signaux, un modèle courant en théorie des jeux pour les interactions à deux joueurs dans un environnement probabiliste et partiellement observable [32, 28, 27]. De façon analogue au cas des POMDP, les joueurs ne peuvent pas observer l'état courant du jeu : leur seule source d'information est la suite de signaux qu'ils recoivent au long la partie. À la différence des observations dans les POMDP, on suppose ici que les signaux ne dépendent pas uniquement de l'état courant, mais peuvent dépendre de l'action, et peuvent même varier pour une action et un état cible fixés. Illustrons la notion de jeux stochastiques à signaux par un premier exemple, à un seul joueur, pour se

familiariser avec le modèle et comprendre en quoi les signaux diffèrent des observations dans les POMDP.

Exemple 5.3.1 (Exemple de jeu stochastique à signaux à un joueur)**.**

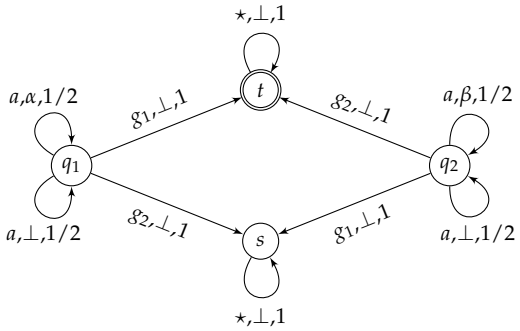


FIGURE *5.13 – Un exemple de jeu stochastique à signaux à un joueur.*

La figure 5.13 présente un exemple de jeu stochastique à un joueur, informé par des signaux. L'ensemble des signaux est $\{\alpha, \beta, \bot\}$*, et les actions possibles sont* a*,* g_1 *et* g_2*. La transition de* q_1 *à* q_1 *étiquetée par* $a, \alpha, 1/2$ *signifie que depuis* q_1*, si le joueur choisit l'action* a*, il recevra le signal* α*, et le nouvel état sera* q_1 *avec probabilité* $1/2$*. La notation* $\star$ *représente n'importe quelle action. Le jeu commence avec probabilité uniforme dans chacun des états* q_1 *et* q_2*, et l'objectif est d'atteindre l'état* t*. Pour cela, le joueur doit deviner correctement l'état initial : si le jeu a commencé en* q_i*, alors* g_i *lui permet d'atteindre* t*. Afin de faire son choix de façon sûre, il peut auparavant jouer l'action* a *jusqu'à ce que, presque sûrement, il recoive le signal* α *ou* β*, permettant de déterminer si le jeu se trouve dans l'état* q_1 *ou* q_2*. Dans ce jeu, l'unique joueur a donc une stratégie pour atteindre l'état* t *avec probabilité* 1*. Observez que même à un joueur, le modèle des jeux stochastiques à signaux généralise celui des* POMDP*, puisque, le signal reçu ne dépend pas uniquement de l'état, ni même de l'action et de l'état, mais aussi de la transition elle-même.*

Venons en maintenant à la définition de jeux stochastiques à deux joueurs à signaux.

Définition 5.3.2. *Un* jeu stochastique à deux joueurs à signaux *est un tuple* $(Q, \delta_0, I, J, C, D, \Delta)$ *où* Q *est un ensemble fini d'états,* $\delta_0 \in \mathsf{Dist}(Q)$ *une distribution initiale,* I *(resp.* J*) est l'ensemble des actions du joueur* 1 *(resp. du joueur* 2*),* C *(resp.* D*) est l'ensemble des signaux du joueur* 1 *(resp. du joueur* 2*), et* $\Delta : Q \times I \times J \to \mathsf{Dist}(Q \times C \times D)$ *est la fonction de transition.*

La sémantique d'un jeu stochastique à signaux est la suivante : initialement, le jeu commence dans l'état q_0, avec probabilité $\delta_0(q_0)$; puis, à

chaque étape $n \in \mathbb{N}$, chacun des joueurs choisit une action $i_n \in I$ et $j_n \in J$. Ils recoivent respectivement les signaux $c_n \in C$ et $d_n \in D$, et le nouvel état du jeu est q_{n+1}. Ceci arrive avec probabilité $\Delta(q_{n+1}, c_n, d_n | q_n, i_n, j_n)$, donnée par la fonction de transition Δ. Une partie est une suite finie ou infinie $q_0, i_0, j_0, c_1, d_1, q_1, \ldots, c_n, d_n, q_n \cdots$, telle que, pour tout $m \geq 0$, $\Delta(q_{m+1}, c_{m+1}, d_{m+1} | q_m, i_m, j_m) > 0$.

Dans les jeux stochastiques à signaux, comme pour l'étude qualitative des POMDP, les conditions de gain sont des sous-ensembles de Q^ω. On se concentrera ici sur des objectifs définis par les conditions d'accessibilité, de Büchi, et de leurs conditions duales respectives, sûreté et co-Büchi.

Les deux joueurs prennent leurs décisions en fonction de la suite des signaux qu'ils ont reçue jusqu'alors. Une stratégie est donc une fonction associant à chaque suite finie de signaux privés, une distribution de probabilité sur les actions. Formellement, une stratégie du joueur 1 est une fonction $\sigma : C^* \to \mathsf{Dist}(I)$. Quand le joueur 1 a reçu les signaux $c_1, \ldots, c_n$ il joue donc l'action i avec probabilité $\sigma(c_1, \ldots, c_n)(i)$. La notion de stratégie $\tau : D^* \to \mathsf{Dist}(J)$ pour le joueur 2 est définie de façon symétrique. Un profil de stratégies, c'est-à-dire une stratégie pour chacun des joueurs, définit une mesure de probabilité sur l'ensemble des parties, et après projection, une mesure de probabilité sur l'ensemble des séquences de Q^ω, notée $\mathbf{Pr}^{\sigma,\tau}$.

Étant donnée une condition de gain $V \subseteq Q^\omega$, on dit qu'une stratégie σ pour le joueur 1 est *presque sûrement gagnante* si quelle que soit la stratégie τ que choisit le joueur 2, $\mathbf{Pr}^{\sigma,\tau}(V) = 1$. Similairement, σ est *positivement gagnante* si quelle que soit la stratégie τ du joueur 2, $\mathbf{Pr}^{\sigma,\tau}(V) > 0$. Par extension, on dira qu'une distribution est presque sûrement gagnante (resp. positivement gagnante), si le joueur 1 possède une stratégie presque sûrement (resp. positivement) gagnante depuis cette distribution initiale. De la même façon que pour les POMDP, le fait que le joueur 1 possède une stratégie gagnante (que ce soit presque sûrement, ou positivement) ne dépend que du support de la distribution initiale. Ainsi, on parlera également de support presque sûrement ou positivement gagnant.

Exemple 5.3.3 (Exemple de jeu stochastique à signaux à deux joueurs)**.** *Un nouvel exemple de jeu stochastique à signaux, à deux joueurs, est donné en figure 5.14. Les transitions sont étiquetées par des triplets : le premier élément correspond aux actions choisies par chacun des joueurs, le deuxième élément décrit les signaux qu'ils reçoivent, et le dernier élément est la probabilité de cette transition.*

Le jeu commence dans l'état q_0, et chacun des joueurs choisit pile (p) ou face (f). Le nouvel état est $q_=$ ou $q_\neq$, selon qu'ils font le même choix ou non. Depuis q_1, l'évolution est la même, à la différence près que les signaux reçus par le joueur

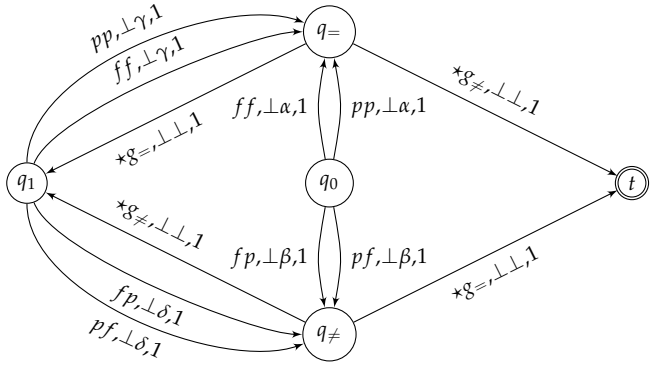


FIGURE 5.14 – *Un exemple de jeu stochastique à signaux à deux joueurs.*

2 *sont a priori différents. Le joueur* 1 *est aveugle (il reçoit tout le temps le même signal) et ne peut que compter les étapes. L'objectif du joueur* 1 *est d'atteindre la cible* t*, et cela arrive lorsque le joueur* 2 *ne devine pas correctement : soit il joue* $g_{\neq}$ *dans l'état* $q_=$*, soit il joue* $g_=$ *dans l'état* $q_{\neq}$*. Nous allons voir qu'en fonction des signaux du joueur* 2*, le jeu peut être presque sûrement gagnant pour le joueur* 1*, gagnant avec probabilité positive, ou même perdant.*
Dans un premier temps, supposons que tous les signaux α*,* β*,* γ *et* δ *sont différents. Le joueur* 2 *connaît alors toujours la position réelle du jeu, et peut choisir en conséquence* $q_=$ *ou* $q_{\neq}$ *pour éviter* t*. Il a donc une stratégie qui garantit d'éviter* t*, et le joueur* 1 *gagne avec probabilité* 0*.*
Supposons maintenant que $\alpha = \beta$*, mais* $\gamma \neq \delta$*. Intuitivement, à la première étape, le joueur* 2 *ne peut pas distinguer si le jeu se trouve dans l'état* $q_=$ *ou* $q_{\neq}$*. Puis, lorsque le jeu atteint* q_1*, le joueur* 2 *peut éviter d'atteindre* t *quoi que fasse le joueur* 1*. Pour chacun des joueurs, à la première étape, le meilleure stratégie est de jouer uniformément pile ou face. Ainsi, dans ce cas, le joueur* 1 *gagne avec probabilité* $1/2$*.*
Supposons enfin que $\alpha = \beta$ *et* $\gamma = \delta$*, et donc que le joueur* 2 *ne puisse jamais savoir si le jeu se trouve en* $q_=$ *ou* $q_{\neq}$*. La meilleure stratégie pour le joueur* 1 *est de toujours jouer uniformément pile ou face. Sous cette stratégie aléatoire, et quoi que fasse le joueur* 2*, toutes les deux étapes, la probabilité est* $1/2$ *d'atteindre* t*. Ainsi, le joueur* 1 *gagne presque sûrement.*

Pour terminer cette introduction aux jeux stochastiques à signaux, on définit la notion de croyance, qui représente l'information qu'un joueur a de la position actuelle du jeu, et que l'on a déjà introduite pour les POMDP. Étant donné un jeu $(Q, \delta_0, I, J, C, D, \Delta)$, initialement, la croyance du joueur 1, et du joueur 2 est $\mathsf{Supp}(\delta_0)$. Supposons qu'à une étape du jeu, la croyance

du joueur 1 est c_1. S'il choisit de jouer $i \in I$ et qu'il reçoit le signal $c \in C$, sa croyance est mise à jour en c_1', définie par :

$$c_1' = \{q' \in Q \mid \exists q \in c_1 \; \exists j \in J \; \exists d \in D, \; \Delta(q', c, d|q, i, j) > 0\}.$$

La mise à jour des croyances pour le joueur 2 est définie de façon tout à fait symétrique.

5.3.2 Caractère déterminé

Considérons un jeu dont *l'arène* est une séquence infinie d'états qui vérifient ou pas une propriété F. Le joueur 1 choisit un état de la séquence puis le joueur 2 choisit un état ultérieur de la séquence. Le joueur 1 (respectivement 2) gagne si l'état choisi par le second joueur ne satisfait pas F. L'issue d'une partie est fixée par les choix (i.e. les stratégies) des deux joueurs. On note $val_j(\nu_j, \nu_{\bar{j}}) \in \{0, 1\}$ le résultat du joueur $j \in \{1, 2\}$ avec adversaire $\bar{j}$ lorsque j (respectivement $\bar{j}$) choisit comme stratégie ν_j (respectivement $\nu_{\bar{j}}$). Dans un jeu de ce type, $val_j(\nu_j, \nu_{\bar{j}}) = 1 - val_{\bar{j}}(\nu_j, \nu_{\bar{j}})$. Une stratégie ν_j est gagnante si pour toute stratégie $\nu_{\bar{j}}$ de l'adversaire $val_j(\nu_j, \nu_{\bar{j}}) = 1$. Il ne peut y avoir simultanément une stratégie gagnante pour les deux joueurs.

Le jeu est *déterminé* si l'un des deux joueurs a une stratégie gagnante. Un résultat fondamental de la théorie des jeux énonce que les jeux à information parfaite et condition borélienne sont déterminés [22]. Dans notre exemple, le joueur 1 gagne si la séquence satisfait $\Diamond\Box\neg F$ et le joueur 2 gagne si la séquence satisfait $\Box\Diamond F$. On retrouve ici aussi la dualité des formules sous la forme d'un jeu.

Définissons une notion de jeu déterminé adaptée aux jeux stochastiques. Dans un tel jeu une fois fixées les stratégies des deux joueurs, le jeu stochastique devient un processus stochastique et la valeur du jeu pour un joueur est une variable aléatoire de Bernouilli (sous réserve que l'ensemble des chemins qui satisfont la condition de gain soit mesurable). Chaque joueur cherche alors à maximiser l'espérance de cette variable aléatoire c'est à dire la probabilité qu'un chemin aléatoire satisfasse la condition de gain.

Définition 5.3.4. *Une condition de gain (accessibilité, sureté, Büchi) est* déterminée *si pour tout jeu stochastique à signaux, toute distribution initiale δ_0 admet soit une stratégie presque sûrement gagnante pour le joueur* 1*, soit une stratégie gagnante avec probabilité positive pour le joueur* 2*.*

Formellement, le caractère déterminé s'exprime ainsi, pour $W \subseteq Q^\omega$:

$$\left(\exists \nu_1 \; \forall \nu_2 \; \mathbf{Pr}_{\delta_0}^{\nu_1,\nu_2}(W) = 1\right) \text{ ou } \left(\exists \nu_2 \; \forall \nu_1 \; \mathbf{Pr}_{\delta_0}^{\nu_1,\nu_2}(W) < 1\right)$$

En raison de l'observation partielle des jeux stochastiques à signaux le caractère déterminé dépend étroitement de la condition de gain [6].

Théorème 5.3.5. *Les jeux stochastiques à signaux avec conditions d'accessibilité, de sûreté ou de Büchi, sont déterminés.*

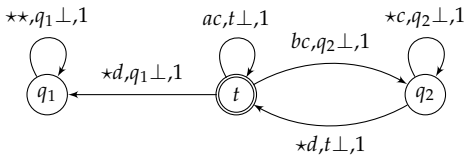


FIGURE 5.15 – *Un jeu non-déterminé avec condition de gain de co-Büchi.*

Théorème 5.3.6. *Les jeux stochastiques à signaux avec conditions co-Büchi ne sont pas déterminés.*

Démonstration. Nous allons démontrer ce résultat en analysant le jeu de la figure 5.15 où la distribution initiale est concentrée en t. Dans ce jeu, le joueur 1 observe tout, le joueur 2 est aveugle, et l'objectif du joueur 1 est de ne visiter l'état t qu'un nombre fini de fois. Puisque le joueur 2 est aveugle, une stratégie de ce joueur est une suite de distributions où l'on note c_i (respectivement d_i) la probabilité de choisir c (respectivement d) à l'instant i.

Fixons une stratégie arbitraire pour le joueur 2 et définissons la stratégie du joueur 1 ainsi. A l'instant i, si $c_{\geq i} = \prod_{j \geq i} c_i$ est nulle alors le joueur 1 choisit a sinon le joueur 1 choisit b. La suite $c_{\geq i}$ est croissante. Supposons qu'elle soit toujours nulle. Alors la probabilité de choisir d au moins une fois est égale à 1 et donc presque sûrement la transition $t \to q_1$ sera franchie (satisfaisant ainsi la condition de co-Büchi). Dans le cas contraire, à partir d'un certain i_0, il existe $\varepsilon > 0$ tel que pour tout $i \geq i_0$, $c_{\geq i} \geq \varepsilon$. La stratégie du joueur 1 conduit soit à franchir $t \to q_1$, soit à visiter infiniment souvent q_2 avec une probabilité au plus $1 - \varepsilon$ à chaque nouvelle visite depuis t de le quitter. Par conséquent, presque sûrement le chemin aléatoire restera indéfiniment dans q_1 ou q_2.

Fixons maintenant une stratégie arbitraire pour le joueur 1 et spécifions, par induction sur les instants, une stratégie déterministe du joueur 2 de la forme $c^{n_1}dc^{n_2}d\ldots$ ou $c^{n_1}dc^{n_2}d\ldots c^{n_i}dc^{\omega}$ à l'aide d'une suite $(p_i)_{i\in\mathbb{N}^*}$ telle que pour tout i, $0 < p_i < 1$ et $\prod_{i\in\mathbb{N}^*} p_i > 0$. Supposons-la définie jusqu'à l'instant $k = i + \sum_{j\leq i} n_j$, qu'elle soit de la forme $c^{n_1}dc^{n_2}d\ldots c^{n_i}d$ et que la probabilité d'être dans t à l'instant k soit supérieure ou égale à

$\prod_{j\leq i} p_j$. Cette hypothèse est satisfaite pour $i = 0$. Définissons p comme la probabilité qu'à partir de cet instant, la stratégie du joueur 1 choisisse indéfiniment a tant que le système est dans l'état t. Si p est non nul alors la stratégie du joueur 2 se termine par c^ω avec une probabilité au moins égale à $p \prod_{j\leq i} p_j > 0$ de rester indéfiniment dans t. Si p est nul alors il existe n_{i+1} tel qu'à partir de cet instant la probabilité induite par la stratégie du joueur 1 de choisir b au moins une fois dans les n_{i+1} prochains instants soit supérieure ou égale à p_i. La stratégie du joueur 2 se prolonge par $c^{n_{i+1}}d$ vérifiant à l'étape suivante l'hypothèse d'induction. Dans le cas d'une stratégie construite de la forme $c^{n_1}dc^{n_2}d\ldots$ la probabilité de visiter infiniment souvent t est au moins égale à $\prod_{i\in\mathbb{N}^*} p_i$. □

5.3.3 Résolution des jeux stochastiques à signaux

Les jeux stochastiques à signaux formant une extension des POMDP, les résultats d'indécidabilité obtenus dans la section précédente s'appliquent également ici. En particulier, on ne peut pas décider, étant donnés un jeu stochastique à signaux et un sous-ensemble F de ses états, si le joueur 1 a une stratégie garantissant avec probabilité positive de visiter F infiniment souvent, quoi que fasse le joueur 2. Si les résultats d'indécidabilité s'étendent des POMDP aux jeux stochastiques, il est remarquable que les résultats de décidabilité soient également valables dans ce cadre plus général. Commençons par un cas relativement simple où le joueur 1 cherche à garantir une condition d'accessibilité avec probabilité positive.

Proposition 5.3.7. *On peut calculer en* EXPTIME *l'ensemble des supports positivement gagnants pour le joueur* 1 *et la condition d'accessibilité* $\Diamond F$.

Démonstration. Puisque les jeux stochastiques à signaux et condition d'accessibilité sont déterminés, de façon équivalente, on peut calculer les supports depuis lesquels le joueur 2 a une stratégie qui assure de rester hors de F avec probabilité 1. Pour cette condition de sûreté, l'ensemble des supports $\mathcal{L} \subseteq 2^Q$ qui sont gagnants pour le joueur 2 peut être caractérisé comme un plus grand point fixe. On définit donc $\mathcal{L}_0 = \{L \in 2^Q \mid L \cap F = \emptyset\}$, et inductivement à partir de $\mathcal{L}_i$, on construit $\mathcal{L}_{i+1}$ comme l'ensemble des supports $L \in \mathcal{L}_i$ qui n'intersectent pas F et depuis lesquels le joueur 2 possède une action qui garantit que sa prochaine croyance est encore dans $\mathcal{L}_i$, et ce quelle que soit l'action choisie par le joueur 1, et quel que soit le signal reçu par le joueur 2.

$$\mathcal{L}_{i+1} = \{L \in \mathcal{L}_i \mid \exists j \in J\ \forall i \in I\ \forall d \in D,\ \mathcal{B}(d|L,i,j) \neq \emptyset \implies \mathcal{B}(d|L,i,j) \in \mathcal{L}_i\}$$

Pour l'ensemble limite $\mathcal{L}$ dans ce calcul de point fixe, on se convainc aisément que depuis tout support $L \in \mathcal{L}$, le joueur 2 a une stratégie pour assurer de rester hors de F. En effet, depuis L, il possède une action qui lui permet de rester sûrement dans $\mathcal{L}$, et on peut donc itérer ce raisonnement. De plus, si $L \notin \mathcal{L}$, on peut montrer que quelle que soit sa stratégie, le joueur 1 a une stratégie permettant de forcer la visite de F en temps fini, avec probabilité positive. □

Après ce cas relativement facile, concentrons nous sur des jeux avec condition de Büchi, et voyons comment le résultat du théorème 5.2.18 se généralise aux jeux stochastiques.

Théorème 5.3.8. *On peut calculer en* 2EXPTIME *l'ensemble des supports presque sûrement gagnants pour le joueur* 1 *et la condition de Büchi* $\Box\Diamond F$.

Démonstration. On se contente de donner ici l'idée de la preuve.

Pour décider si le joueur 1 possède une stratégie presque sûrement gagnante, on fait appel à l'algorithme de la proposition 5.3.7.

Supposons pour commencer que depuis tout support le joueur 1 peut garantir d'atteindre F une fois avec probabilité positive. Dans cette situation très favorable, en répétant toujours cette même stratégie qui cherche à atteindre F, le joueur 1 peut assurer la condition de Büchi $\Box\Diamond F$ presque sûrement.

Puisque les jeux stochastiques à signaux et condition d'accessibilité sont déterminés, dans le cas contraire, il existe un support L depuis lequel le joueur 2 possède une stratégie qui garantit $\Box\neg F$ presque sûrement, et donc sûrement. En fait, dès que le joueur 2 peut forcer la croyance du joueur 1 à être précisément L depuis un support L', alors L' est gagnant positivement pour le joueur 2. Ceci n'est pas complètement trivial, car le joueur 2 ne connaît pas a priori les croyances du joueur 1. Pour gagner positivement, il lui faut donc jouer de façon aléatoire jusqu'à un moment supposer que la croyance du joueur 1 est L, et à ce moment là jouer la stratégie qui gagne sûrement depuis L. Bien qu'une telle stratégie ne soit pas optimale (dans bien des cas, le joueur 2 supposera à tort que la croyance de son adversaire est L), elle permet au joueur 2 d'atteindre son objectif de co-Büchi avec probabilité positive.

De ces observations, on déduit que le joueur 1 doit éviter d'avoir pour croyance L, et même L', s'il veut gagner presque sûrement. En faisant cela, il peut cependant empêcher le jeu d'aller vers la cible F, et crée donc de nouveaux supports gagnants positivement pour le joueur 2, etc.

Le raisonnement qui précède suggère donc de calculer l'ensemble des supports presque sûrement gagnants pour le joueur 1 pour la condition

de Büchi comme le plus grand ensemble $\mathcal{W}$ de supports depuis lesquels le joueur 1 a une stratégie pour assurer une probabilité positive d'atteindre F tout en garantissant que sa croyance reste dans $\mathcal{W}$. On peut tirer de cela une caractérisation de $\mathcal{W}$ comme plus grand point fixe, en utilisant l'algorithme de la proposition 5.3.7 comme routine. □

Au delà du saut de complexité entre les POMDP et les jeux stochastiques à signaux à deux joueurs, soulignons une différence importante, qui concerne la mémoire nécessaire aux joueurs pour satisfaire leurs objectifs. Dans les POMDP, cette mémoire était au mieux nulle (pour gagner positivement un jeu avec condition d'accessibilité), et au pire exponentielle (par exemple pour gagner positivement un jeu avec condition de sûreté), essentiellement basée sur les croyances. Dans le cas des jeux stochastiques à signaux à deux joueurs, une mémoire exponentielle ne suffit pas. Plus précisément, on peut montrer que pour gagner positivement un jeu avec condition de sûreté, les croyances du joueur 2 ne sont pas suffisantes, et il peut devoir recourir à ses croyances sur les croyances de son adversaire, impliquant une mémoire doublement exponentielle.

Bibliographie

[1] K. J. Aström. Optimal control of Markov decision processes with incomplete state estimation. *Journal of Mathematical Analysis and Applications*, 10 :174–205, 1965.

[2] C. Baier, N. Bertrand, and M. Größer. On decision problems for probabilistic Büchi automata. In *Proceedings of FoSSaCS'08*, volume 4962 of *Lecture Notes in Computer Science*, pages 287–301. Springer, 2008.

[3] C. Baier, N. Bertrand, and M. Größer. Probabilistic ω-automata. *J. ACM*, 59(1), 2012.

[4] C. Baier and M. Größer. Recognizing omega-regular languages with probabilistic automata. In *Proceedings of LICS'05*, pages 137–146. IEEE Computer Society Press, 2005.

[5] N. Bertrand, E. Fabre, S. Haar, S. Haddad, and L. Hélouët. Active diagnosis for probabilistic systems. In *Proceedings of FoSSaCS'14*, volume 8412 of *Lecture Notes in Computer Science*, pages 29–42. Springer, 2014.

[6] N. Bertrand, B. Genest, and H. Gimbert. Qualitative determinacy and decidability of stochastic games with signals. In *Proceedings of LICS'09*, pages 319–328. IEEE Computer Society Press, 2009.

[7] D. Berwanger and L. Doyen. On the power of imperfect information. In *Proceedings of FSTTCS'08*, volume 2 of *LIPIcs*, pages 73–82. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2008.

[8] A. R. Cassandra, M. L. Littman, and N. L. Zhang. Incremental pruning : A simple, fast, exact method for partially observable Markov decision processes. In *Proceedings of UAI'87*, pages 54–61. Morgan Kaufmann, 1997.

[9] K. Chatterjee, L. Doyen, H. Gimbert, and T. A. Henzinger. Randomness for free. In *Proceedings of MFCS'10*, volume 6281 of *Lecture Notes in Computer Science*, pages 246–257. Springer, 2010.

[10] K. Chatterjee, L. Doyen, and T. Henzinger. Qualitative analysis of partially-observable Markov decision processes. In *Proceedings of*

MFCS'10, volume 6281 of *Lecture Notes in Computer Science*, pages 258–269. Springer, 2010.

[11] P. D. Diêu. On the languages representable by finite probabilistic automata. *Z. Math. Logik Grundlagen Math.*, 17 :427–442, 1971.

[12] M. Fliess. Propriétés booléennes des langages stochastiques. *Mathematical Systems Theory*, 7(4) :353–359, 1974.

[13] H. Gimbert and Y. Oualhadj. Probabilistic automata on finite words : Decidable and undecidable problems. In *Proceedings of ICALP'10*, volume 6199 of *Lecture Notes in Computer Science*, pages 527–538. Springer, 2010.

[14] E. Grädel, W. Thomas, and T. Wilke, editors. *Automata, Logics, and Infinite Games : A Guide to Current Research*, volume 2500 of *Lecture Notes in Computer Science*. Springer, 2002.

[15] S. Haar, S. Haddad, T. Melliti, and S. Schwoon. Optimal constructions for active diagnosis. In *Proceedings of FSTTCS'13*, volume 24 of *LIPIcs*, pages 527–539. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2013.

[16] E. A. Hansen. An improved policy iteration algorithm for partially observable MDPs. In *Proceedings of NIPS'97*. The MIT Press, 1997.

[17] J. Hopcroft, R. Motwani, and J. Ullman. *Introduction to Automata Theory, Languages, and Computation.* Addison-Wesley, Third edition, 2006.

[18] P. S. Landweber. Three theorems on phrase structure grammars of type 1. *Information and Control*, 6(2) :131–136, 1963.

[19] I. I. Macarie. Space-efficient deterministic simulation of probabilistic automata. *SIAM Journal on Computing*, 27(2) :448–465, 1998.

[20] O. Madani, S. Hanks, and A. Condon. On the undecidability of probabilistic planning and related stochastic optimization problems. *Artificial Intelligence*, 147(1-2) :5–34, 2003.

[21] Z. Manna and A. Pnueli. *The Temporal Logic of Reactive and Concurrent Systems*. Springer, 1992.

[22] D. A. Martin. Borel determinacy. *Annals of Mathematics*, 102(2) :363–371, 1975.

[23] M. Nasu and N. Honda. A context-free language which is not acceptable by a probabilistic automaton. *Information and Control*, 18(3) :233–236, 1971.

[24] A. Paz. *Introduction to probabilistic automata (Computer science and applied mathematics)*. Academic Press Inc., 1971.

[25] M. L. Puterman. *Markov Decision Processes : Discrete Stochastic Dynamic Programming*. Wiley, 2005.

[26] M. O. Rabin. Probabilistic automata. *Information and Control*, 6(3) :230–245, 1963.

[27] J. Renault. The value of repeated games with an informed controller. Technical report, CEREMADE, Paris, Jan. 2007.

[28] D. Rosenberg, E. Solan, and N. Vieille. Stochastic games with imperfect monitoring. Technical Report 1376, Northwestern University, July 2003.

[29] M.-P. Schützenberger. On the definition of a family of automata. *Information and Control*, 4(2-3) :245—270, 1961.

[30] R. D. Smallwood and E. J. Sondik. The optimal control of partially observable Markov processes over a finite horizon. *Operations Research*, 21(5) :1071–1088, 1973.

[31] E. J. Sondik. The optimal control of partially observable Markov processes over the infinite horizon : Discounted costs. *Operations Research*, 26(2) :282–304, 1978.

[32] S. Sorin. *A first course on zero-sum repeated games*. Springer, 2002.

[33] P. Turakainen. Generalized automata and stochastic languages. *Proceedings of the American Mathematical Society*, 21(2) :303–309, 1969.

[34] P. Turakainen. Some closure properties of the family of stochastic languages. *Information and Control*, 18(3) :253—256, 1971.