Contrôle, probabilités et observation partielle

Deuxième partie : POMDP

Nathalie Bertrand*, Serge Haddad**
* Inria Rennes Bretagne Atlantique
** LSV, ENS Cachan & CNRS & Inria

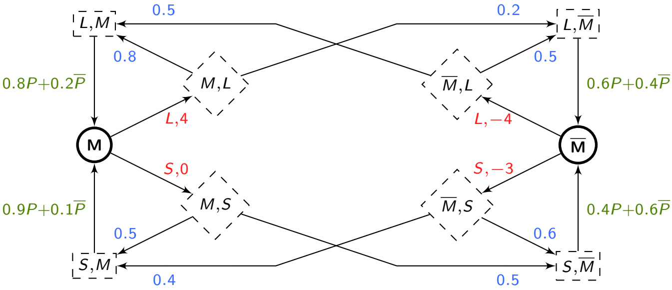
EJC IM 2015, Orléans

Un premier exemple de POMDP

Une entreprise commercialise un produit, de luxe (L) ou standard (S).
Les consommateurs sont sensibles aux marques (**M**) ou non ($\overline{\mathbf{M}}$)
mais l'entreprise ne connaît pas cette information...
... et sait uniquement si le produit est acheté (P) ou non ($\overline{P}$).

Un premier exemple de POMDP

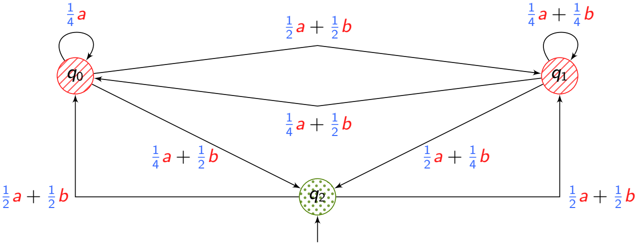
Une entreprise commercialise un produit, de luxe (L) ou standard (S).
Les consommateurs sont sensibles aux marques (**M**) ou non ($\overline{\mathbf{M}}$)
mais l'entreprise ne connaît pas cette information...
... et sait uniquement si le produit est acheté (P) ou non ($\overline{P}$).



États : **M**, $\overline{\mathbf{M}}$; Actions : L, S; Observations : P, $\overline{P}$;

- probabilités : $p(\mathbf{M}|\mathbf{M}, L) = 0.8$;
- récompenses : $r(\mathbf{M}, L) = 4$;
- observations :$o(P|L, \mathbf{M}) = 0.8$

Un deuxième exemple de POMDP



États : $\{q_0, q_1, q_2\}$; Actions : $\{a, b\}$; Observations : $\{$, $\}$
$p(q_1|q_0, a) = \frac{1}{2}$; $o(q_0) = o(q_1) =$; récompenses nulles partout

POMDP

POMDP (à observation déterministe)

Un POMDP $\mathcal{M} = (S, \Omega, A, o, p, r)$ est défini par :

- S, l'ensemble fini des états ;
- Ω, l'ensemble fini des observations ;
- A, l'ensemble fini des actions ;
- $o : S \to \Omega$, la fonction d'observation ; $o(s) \in \Omega$ est l'observation associée à l'état s ;
- $p : S \times A \to \mathrm{Dist}(S)$, la fonction de transition ; $p(s'|s, a)$ est la probabilité que le prochain état soit s' en faisant l'action a depuis s ;
- $r : S \times A \to \mathbb{Q}$, la fonction de récompense ; $r(s, a)$ est la récompense associée à l'action a depuis l'état s.

Stratégie

Pour obtenir un processus stochastique une *stratégie* élimine le non-déterminisme.

Stratégie

Une *stratégie* est une fonction $\nu : (A\Omega)^* \rightarrow \mathrm{Dist}(A)$ qui associe à chaque *histoire* $\rho \in (A\Omega)^*$ une distribution sur les actions ; $\nu(\rho, a)$ est la probabilité que a soit choisie étant donnée l'histoire ρ.

Stratégie

Pour obtenir un processus stochastique une *stratégie* élimine le non-déterminisme.

Stratégie

Une *stratégie* est une fonction $\nu : (A\Omega)^* \to \mathrm{Dist}(A)$ qui associe à chaque *histoire* $\rho \in (A\Omega)^*$ une distribution sur les actions ; $\nu(\rho, a)$ est la probabilité que a soit choisie étant donnée l'histoire ρ.

Chaîne de Markov associée

Soient $\mathcal{M}$ un POMDP, ν une stratégie et $\pi \in \mathrm{Dist}(S)$ une distribution initiale. La chaîne de Markov $\mathcal{M}_\nu^\pi$ associée à $\mathcal{M}$, ν et π est définie par :

- $(A\Omega)^* \times S$, l'ensemble (infini) des états ;
- π_0 la distribution initiale telle que $\pi_0(\varepsilon, s) = \pi(s)$ et π_0 est nulle pour les autres états ;
- $\mathbf{P}$ la matrice de transition telle que : $\mathbf{P}[(\rho, s), (\rho a o(s'), s')] = \nu(\rho, a)p(s'|s, a)$, et $\mathbf{P}$ est nulle ailleurs.

POMDP particuliers

Deux cas très particuliers :

- $\Omega = S$: l'agent connaît l'état du système ; on a un processus de décision markovien.
- $|\Omega| = 1$: l'observation est inutile ; on a un POMDP *aveugle*.

POMDP particuliers

Deux cas très particuliers :

- $\Omega = S$: l'agent connaît l'état du système ; on a un processus de décision markovien.
- $|\Omega| = 1$: l'observation est inutile ; on a un POMDP *aveugle*.

PA vs POMDP

Les automates probabilistes sont un cas particulier des POMDP.

mot automate probabiliste $\Longleftrightarrow$ stratégie déterministe POMDP aveugle

Conséquence : Les résultats d'indécidabilité des PA se transfèrent aux POMDP.

Problèmes à horizon infini

Objectifs

Accessibilité F visité au moins une fois :

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

Sûreté toujours rester dans F :

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

Büchi F visité un nombre infini de fois :

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

But : Pour φ un objectif, évaluer $\sup_\nu \mathbb{P}^\nu(\mathcal{M} \models \varphi)$.

Problèmes à horizon infini

Objectifs

Accessibilité F visité au moins une fois :

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

Sûreté toujours rester dans F :

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

Büchi F visité un nombre infini de fois :

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

But : Pour φ un objectif, évaluer $\sup_\nu \mathbb{P}^\nu(\mathcal{M} \models \varphi)$.

Les stratégies déterministes suffisent!

Soit $\mathcal{M}$ un POMDP, et $\varphi \subseteq S^\omega$ un objectif borélien. Pour toute stratégie ν, il existe une stratégie déterministe ν' telle que

$$\mathbb{P}^\nu(\mathcal{M} \models \varphi) \leq \mathbb{P}^{\nu'}(\mathcal{M} \models \varphi).$$

Indécidabilité analyse quantitative à horizon infini

Indécidabilité de l'accessibilité quantitative

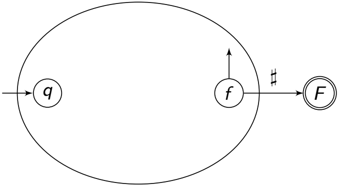
Le problème d'existence d'une stratégie assurant une probabilité $> p$ pour un objectif d'accessibilité $\Diamond F$ est indécidable pour les POMDP.

Indécidabilité analyse quantitative à horizon infini

Indécidabilité de l'accessibilité quantitative

Le problème d'existence d'une stratégie assurant une probabilité $> p$ pour un objectif d'accessibilité $\Diamond F$ est indécidable pour les POMDP.

On réduit le problème du vide pour les PA.
Seule subtilité : synchroniser les chemins!



stratégies déterministes de $\mathcal{M}$: $\nu_w = w\sharp$, où w mot dans le PA $\mathcal{A}$

$$\mathbb{P}^{\nu_w}(\mathcal{M} \models \Diamond F) = \mathbb{P}_{\mathcal{A}}(w)$$

Indécidabilité analyse qualitative à horizon infini

Indécidabilité de l'accessibilité répétée positive

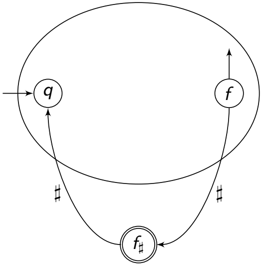
Le problème d'existence d'une stratégie assurant une probabilité > 0 pour un objectif d'accessibilité répétée $\square\diamond F$ est indécidable pour les POMDP.

Indécidabilité analyse qualitative à horizon infini

Indécidabilité de l'accessibilité répétée positive

Le problème d'existence d'une stratégie assurant une probabilité > 0 pour un objectif d'accessibilité répétée $\Box\Diamond F$ est indécidable pour les POMDP.

On réduit le problème de valeur 1 dans les PA.



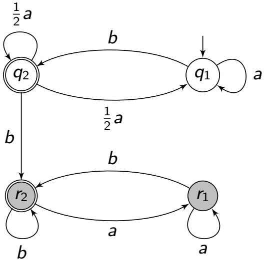
stratégies déterministes de $\mathcal{M}$: $\nu_{\mathbf{w}} = w_1\sharp\sharp w_2\sharp\sharp w_3\cdots$, où w_i mots pour le PA $\mathcal{A}$

$$\mathbb{P}^{\nu_{\mathbf{w}}}(\mathcal{M} \models \Box\Diamond f_\sharp) > 0 \iff \prod_i \mathbb{P}_{\mathcal{A}}(w_i) > 0$$

$$\mathrm{val}(\mathcal{A}) = 1 \iff \exists (w_i)_{i\in\mathbb{N}} \prod \mathbb{P}_{\mathcal{A}}(w_i) > 0$$

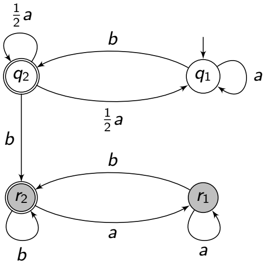
Combinaison d'objectifs à horizon infini

Besoin de mémoire infinie pour des combinaisons d'objectifs !
$\Box\Diamond\{q_2, r_2\}$ presque sûrement et $\Box\{q_1, q_2\}$ avec probabilité positive.



Combinaison d'objectifs à horizon infini

Besoin de mémoire infinie pour des combinaisons d'objectifs !
$\Box\Diamond\{q_2, r_2\}$ presque sûrement et $\Box\{q_1, q_2\}$ avec probabilité positive.



Indécidabilité de combinaison de garanties

Le problème d'existence d'une stratégie assurant

- une probabilité > 0 pour un objectif de sureté $\Box G$, et
- une probabilité $= 1$ pour un objectif de Büchi $\Box\Diamond F$

est indécidable pour les POMDP.

Décidabilité analyse qualitative à horizon infini

Décidabilité accessibilité positive

Le problème d'existence d'une stratégie assurant une probabilité > 0 pour un objectif d'accessibilité $\Diamond F$ est NLOGSPACE-complet pour les POMDP.

Décidabilité analyse qualitative à horizon infini

Décidabilité accessibilité positive

Le problème d'existence d'une stratégie assurant une probabilité > 0 pour un objectif d'accessibilité $\Diamond F$ est NLOGSPACE-complet pour les POMDP.

Équivalent au problème d'accessibilité dans les graphes.
La stratégie purement aléatoire convient : randomisation uniforme sur toutes les actions à chaque étape.

Décidabilité analyse qualitative à horizon infini (2)

Décidabilité sûreté presque sûrement

Le problème d'existence d'une stratégie assurant une probabilité $= 1$ pour un objectif de sûreté $\Box G$ est EXPTIME-complet pour les POMDP.

Décidabilité analyse qualitative à horizon infini (2)

Décidabilité sûreté presque sûrement

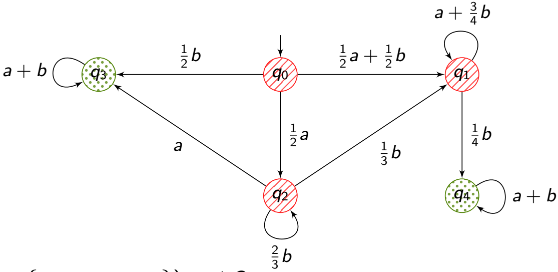
Le problème d'existence d'une stratégie assurant une probabilité $= 1$ pour un objectif de sûreté $\square G$ est EXPTIME-complet pour les POMDP.

Croyances

La *croyance* de l'agent est l'ensemble des états possibles étant donnée la suite d'observations jusqu'alors.

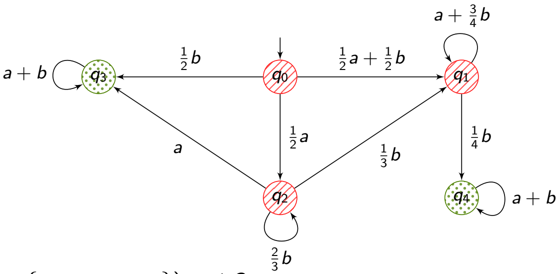
Il faut et il suffit que l'agent maintienne sa croyance incluse dans G.
On construit le *jeu des croyances*.

Jeu des croyances sur un exemple

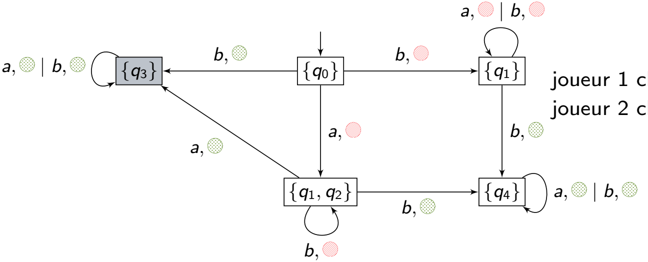


$\exists \nu \ \mathbb{P}^{\nu}(\mathcal{M} \models \Box\{q_0, q_1, q_2, q_4\}) = 1$?

Jeu des croyances sur un exemple



$\exists \nu \ \mathbb{P}^{\nu}(\mathcal{M} \models \Box\{q_0, q_1, q_2, q_4\}) = 1$?



joueur 1 choisit a ou b

joueur 2 choisit ou

Décidabilité analyse qualitative à horizon infini (3)

Décidabilité sûreté positive

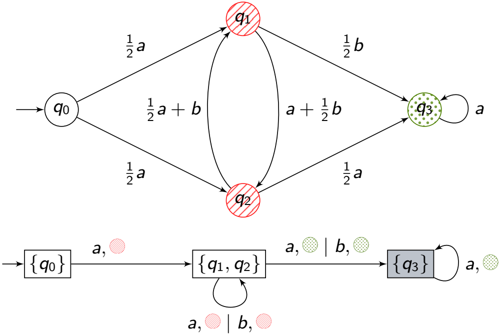
Le problème d'existence d'une stratégie assurant une probabilité > 0 pour un objectif de sûreté $\Box G$ est EXPTIME-complet pour les POMDP.

Décidabilité analyse qualitative à horizon infini (3)

Décidabilité sûreté positive

Le problème d'existence d'une stratégie assurant une probabilité > 0 pour un objectif de sûreté $\square G$ est EXPTIME-complet pour les POMDP.

Les stratégies positionnelles sur le jeu des croyances ne suffisent pas...



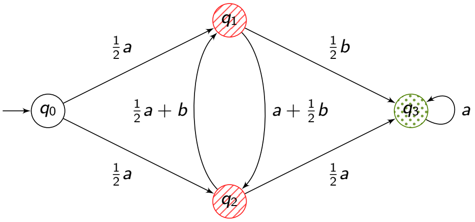
Pourtant, choisir a, puis faire le pari que l'on se trouve dans q_1, et alterner a et b pour toujours, garantit une probabilité $\frac{1}{2}$ pour $\square\{q_0, q_1, q_2\}$.

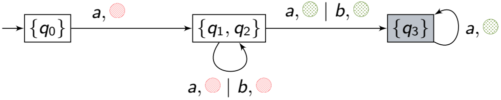
Décidabilité analyse qualitative à horizon infini (3)

Décidabilité sûreté positive

Le problème d'existence d'une stratégie assurant une probabilité > 0 pour un objectif de sûreté $\square G$ est EXPTIME-complet pour les POMDP.

Les stratégies positionnelles sur le jeu des croyances ne suffisent pas...





... mais presque ! Il faut et il suffit d'atteindre une croyance $C \subseteq S$ telle qu'il existe un état $s \in C$ et une stratégie qui assure de rester dans G depuis s.

Décidabilité qualitative à horizon infini (3)

Décidabilité accesssibilité (répétée) presque sûre

Le problème d'existence d'une stratégie assurant une probabilité $= 1$ pour un objectif d'accessibilité $\Diamond F$ ou de Büchi $\Box\Diamond F$ est EXPTIME-complet pour les POMDP.

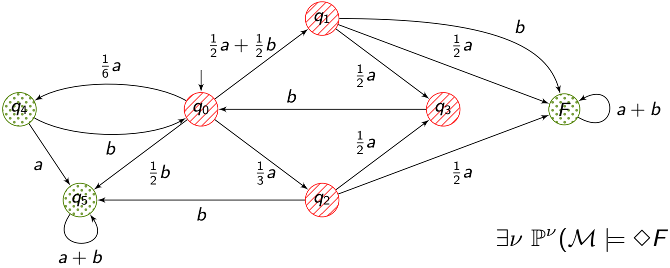
Idée : Il faut pouvoir atteindre une croyance incluse dans F ; toute observation qui ferait dévier de ce chemin doit mener encore à une croyance gagnante, pour pouvoir réessayer d'atteindre F.

Win est le plus grand ensemble de croyances tel que :

$$\mathrm{Win} = \{C \mid \exists C \xrightarrow{a_1,o_1} C_1 \cdots \xrightarrow{a_n,o_n} C_n \subseteq F$$

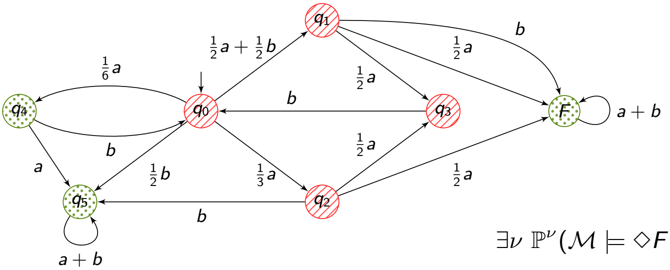
$$\text{et } \forall o'_k\, C \xrightarrow{a_1,o_1} C_1 \cdots \xrightarrow{a_k,o'_k} C'_k \in \mathrm{Win}\}$$

Algorithme de décision sur un exemple

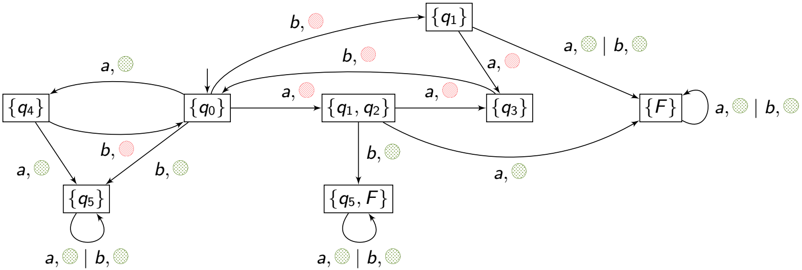


$\exists \nu \ \mathbb{P}^{\nu}(\mathcal{M} \models \Diamond F) = 1$?

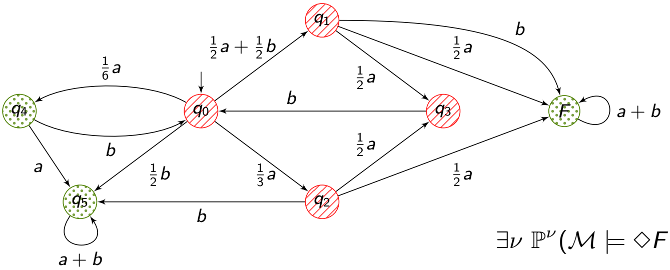
Algorithme de décision sur un exemple



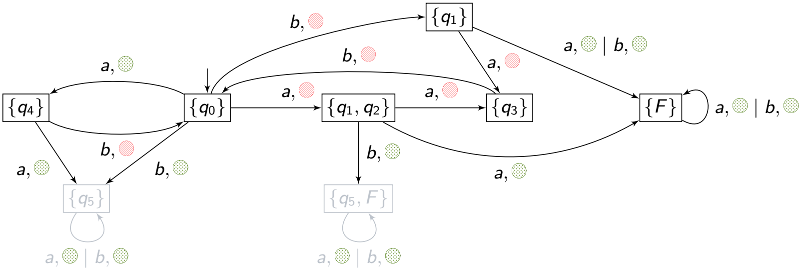
$$\exists \nu \ \mathbb{P}^{\nu}(\mathcal{M} \models \Diamond F) = 1 \ ?$$



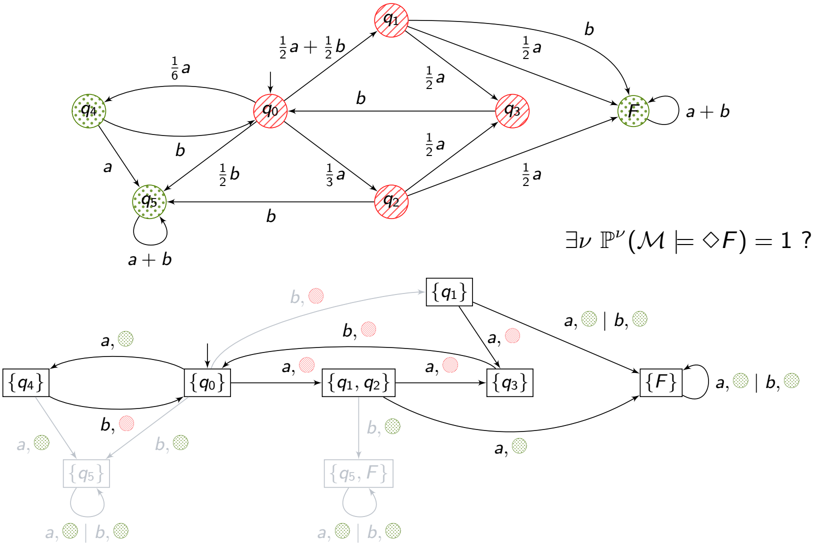
Algorithme de décision sur un exemple



$$\exists \nu \ \mathbb{P}^{\nu}(\mathcal{M} \models \Diamond F) = 1 \ ?$$



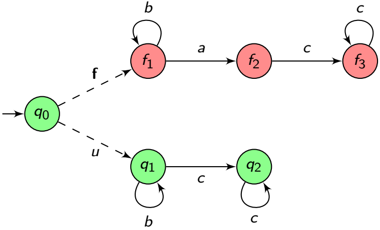
Algorithme de décision sur un exemple



$$\exists \nu \ \mathbb{P}^{\nu}(\mathcal{M} \models \Diamond F) = 1 \ ?$$

Diagnostic de pannes

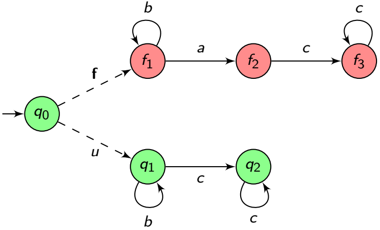
Objectif: savoir si une faute **f** s'est produite, à partir des événements observés.
$\Sigma_o = \{a, b, c\}$ observables ; $\Sigma_u = \{\mathbf{f}, u\}$ non-observables



Diagnostic de pannes

Objectif: savoir si une faute **f** s'est produite, à partir des événements observés.

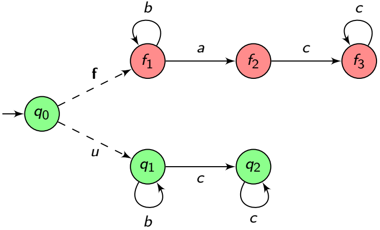
$\Sigma_o = \{a, b, c\}$ observables ; $\Sigma_u = \{\mathbf{f}, u\}$ non-observables



c^+	✓	correcte
ac^+	✗	fautive
b^+	?	ambiguë

Diagnostic de pannes

Objectif: savoir si une faute **f** s'est produite, à partir des événements observés.
$\Sigma_o = \{a, b, c\}$ observables ; $\Sigma_u = \{\mathbf{f}, u\}$ non-observables



c^+	✓	correcte
ac^+	✗	fautive
b^+	?	ambiguë

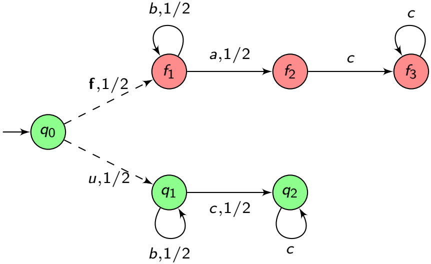
Diagnosticabilité

Un système est diagnosticable si toutes les séquences observées sont non-ambiguës.

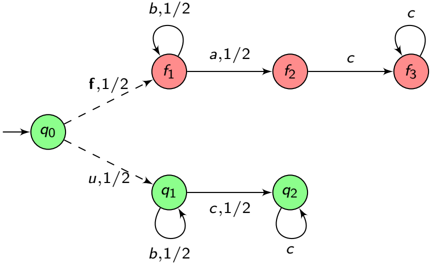
Décidabilité du diagnostic

Le problème du diagnostic est décidable en temps polynomial.

Diagnostic de systèmes probabilistes

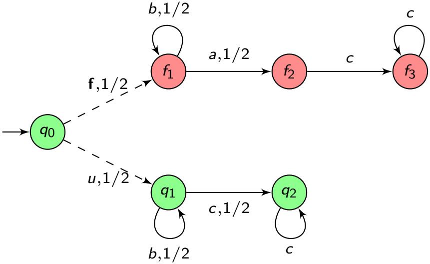


Diagnostic de systèmes probabilistes



b^+ ambiguë mais...

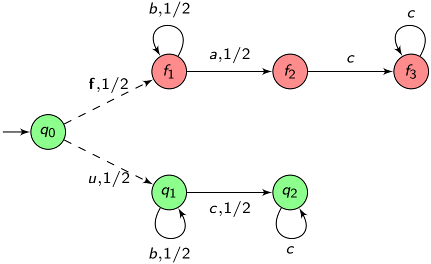
Diagnostic de systèmes probabilistes



b^+ ambiguë mais...

$$\lim_{n\to\infty} \mathbb{P}(\mathbf{f} b^n + u b^n) = 0$$

Diagnostic de systèmes probabilistes



b^+ ambiguë mais...

$$\lim_{n\to\infty} \mathbb{P}(\mathbf{f}b^n + ub^n) = 0$$

Diagnosticabilité probabiliste

Un système probabiliste est diagnosticable si la probabilité des séquences observées ambiguës est nulle.

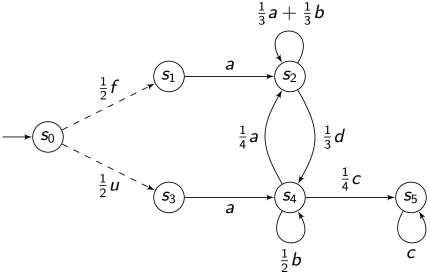
Décidabilité du diagnostic probabiliste

Le problème du diagnostic probabiliste est PSPACE-complet.

Diagnostic actif

Objectif : contrôler le système afin que les séquences ambiguës aient une probabilité nulle.

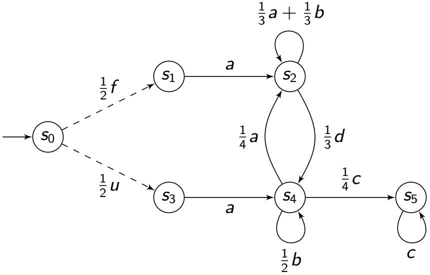
$\Sigma_o = \Sigma_c = \{a, b, c, d\}$ observables et controllables ;
$\Sigma_u = \Sigma_e = \{\mathbf{f}, u\}$ non-observables et non-controllables



Diagnostic actif

Objectif : contrôler le système afin que les séquences ambiguës aient une probabilité nulle.

$\Sigma_o = \Sigma_c = \{a, b, c, d\}$ observables et controllables ;
$\Sigma_u = \Sigma_e = \{\mathbf{f}, u\}$ non-observables et non-controllables

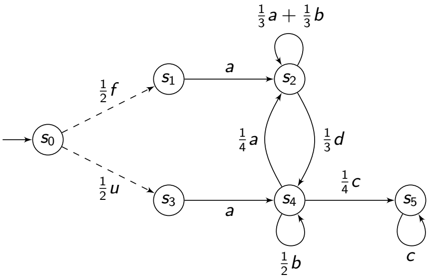


$aadc^\omega$ ambiguë
$\mathbb{P}(faadc^\omega + uaadc^\omega) > 0$

Diagnostic actif

Objectif : contrôler le système afin que les séquences ambiguës aient une probabilité nulle.

$\Sigma_o = \Sigma_c = \{a, b, c, d\}$ observables et controllables ;
$\Sigma_u = \Sigma_e = \{\mathbf{f}, u\}$ non-observables et non-controllables



$aadc^\omega$ ambiguë
$\mathbb{P}(faadc^\omega + uaadc^\omega) > 0$

interdire a après le premier a

Contrôleur : décide quelles actions sont autorisées à partir des observations
$\sigma : \Sigma_{\text{obs}}^* \to 2^{\Sigma_c}$

Résolution

Décidabilité du diagnostic actif probabiliste

Le problème de contrôle pour assurer la diagnosticabilité d'un système probabiliste est EXPTIME-complet.

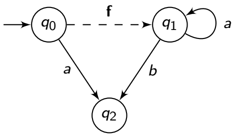
Résolution

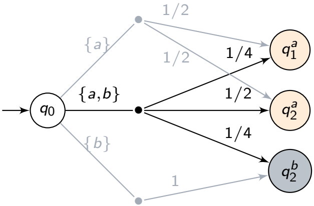
Décidabilité du diagnostic actif probabiliste

Le problème de contrôle pour assurer la diagnosticabilité d'un système probabiliste est EXPTIME-complet.

Idée de l'algorithme EXPTIME

- caractériser les séquences non-ambiguës par un automate déterministe de Büchi $\mathcal{B}$
- construire produit du LTS probabiliste avec $\mathcal{B}$: un nouveau pLTS
- le transformer en POMDP $\mathcal{P}$
  chaque action est un sous-ensemble des événements controllables
  les observations sont les événements observables

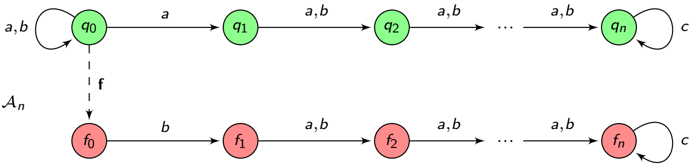




- decider s'il existe une stratégie assurant une probabilité $= 1$ pour la condition de Büchi dans $\mathcal{P}$.

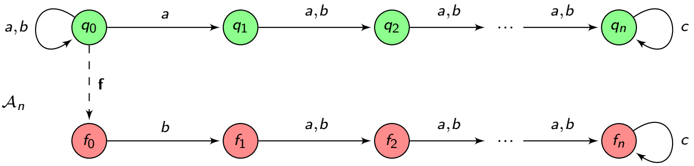
4 Exercices

Diagnostic (probabiliste)



- Donner un contrôleur pour le système ci-dessus.

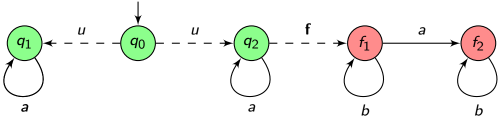
Diagnostic (probabiliste)

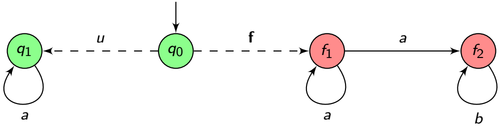


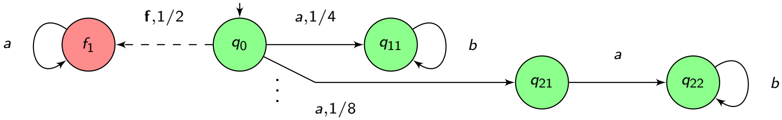
- Donner un contrôleur pour le système ci-dessus.
- Quelle quantité d'information (en fonction de n) un tel contrôleur doit posséder ?

Diagnostic probabiliste

Pour les systèmes suivants, dire s'ils sont diagnosticables.

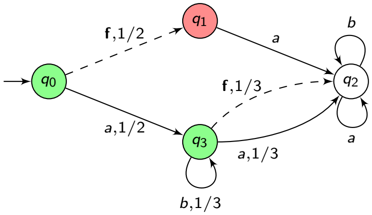






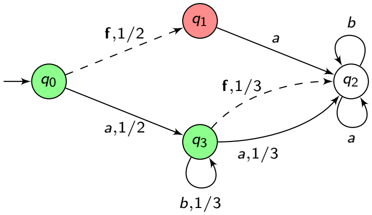
Diagnostic actif

- Donner un contrôleur pour le système suivant :



Diagnostic actif

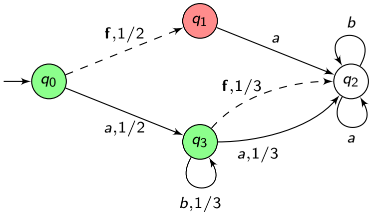
- Donner un contrôleur pour le système suivant :



- Quelle est la probabilité des exécutions non fautives sous ce contrôleur ?

Diagnostic actif

- Donner un contrôleur pour le système suivant :



- Quelle est la probabilité des exécutions non fautives sous ce contrôleur ?

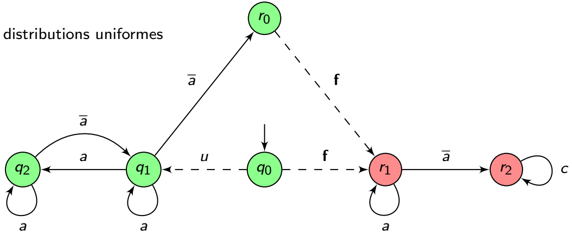
Diagnostic probabiliste sauf

existe-t-il un contrôleur qui rende le système diagnosticable tout en assurant une probabilité non nulle aux exécutions non fautives ?

Diagnostic actif sauf

Diagnostic probabiliste sauf

existe-t-il un contrôleur qui rende le système diagnosticable tout en assurant une probabilité non nulle aux exécutions non fautives ?

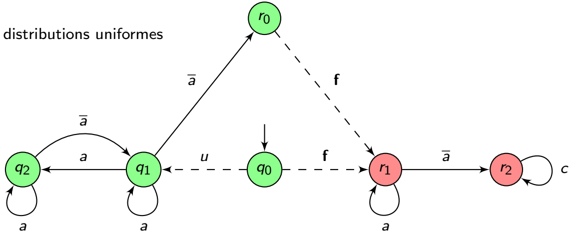


- Donner un contrôleur assurant un diagnostic sauf.

Diagnostic actif sauf

Diagnostic probabiliste sauf

existe-t-il un contrôleur qui rende le système diagnosticable tout en assurant une probabilité non nulle aux exécutions non fautives ?



- Donner un contrôleur assurant un diagnostic sauf.
- Peut-on concevoir un contrôleur *à mémoire finie* répondant au problème ?