Contrôle dynamique stochastique : une approche à base de modèles semi-Markov

Application à l'optimisation d'un traitement médical

Orlane Rossini [1], Alice Cleynen [1,2], Benoîte de Saporta [1] et Régis Sabbadin [3]

[1]IMAG, Univ Montpellier, CNRS, Montpellier, France
[2]John Curtin School of Medical Research, The Australian National University, Canberra, ACT, Australia
[3]Univ Toulouse, INRAE-MIAT, Toulouse, France

13 Mars 2024

Le contexte médical

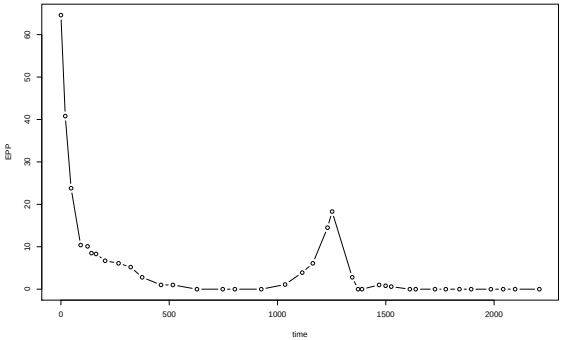


Figure: Exemple de données d'un patient[a]

- Des patients ayant eu un cancer bénéficient d'un suivi régulier;
- La concentration d'immunoglobuline clonale est mesurée dans le temps;
- Le médecin doit prendre de nouvelles décisions à chaque visite.

[a] IUCT Oncopole et CRCT, Toulouse, France

Le contexte médical

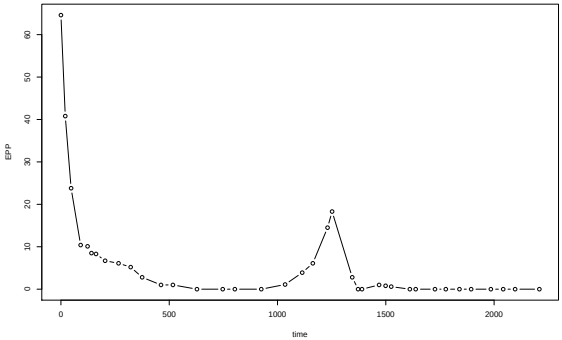


Figure: Exemple de données d'un patient[a]

- Des patients ayant eu un cancer bénéficient d'un suivi régulier;
- La concentration d'immunoglobuline clonale est mesurée dans le temps;
- Le médecin doit prendre de nouvelles décisions à chaque visite.

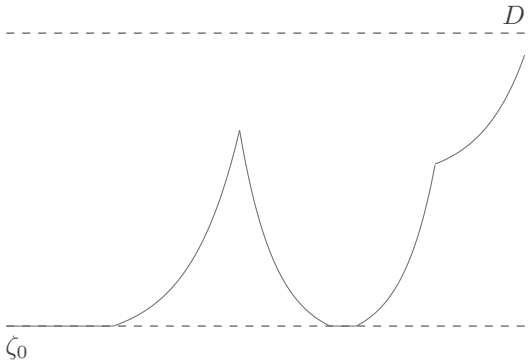
$\Longrightarrow$ **Contrôle dynamique stochastique**

[a] IUCT Oncopole et CRCT, Toulouse, France

Sommaire

Le modèle PDMP[1] contrôlé

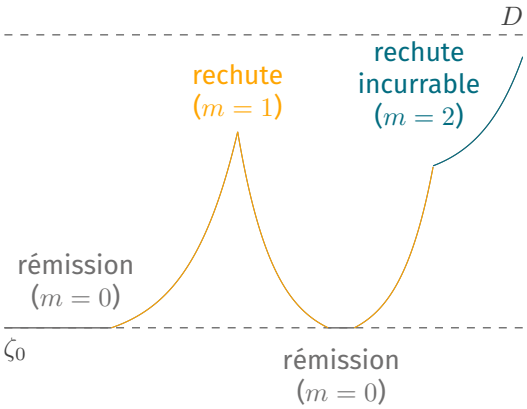
On passe aléatoirement d'un régime déterministe à un autre.



[1]Processus Markovien Déterministe par Morceaux

Le modèle PDMP[1] contrôlé

On passe aléatoirement d'un régime déterministe à un autre.



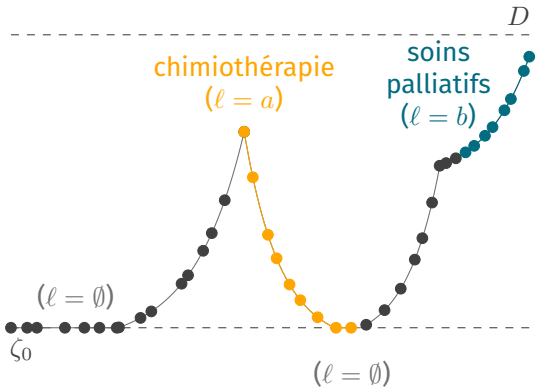
Soit l'état du patient $x = (m, k, \zeta, u)$:

- m le régime;
- k le nombre de rechute;
- ζ le biomarqueur;
- u le temps depuis le dernier saut.

[1]Processus Markovien Déterministe par Morceaux

Le modèle PDMP[1] contrôlé

On passe aléatoirement d'un régime déterministe à un autre.



Soit l'état du patient $x = (m, k, \zeta, u)$:

- m le régime;
- k le nombre de rechute;
- ζ le biomarqueur;
- u le temps depuis le dernier saut.

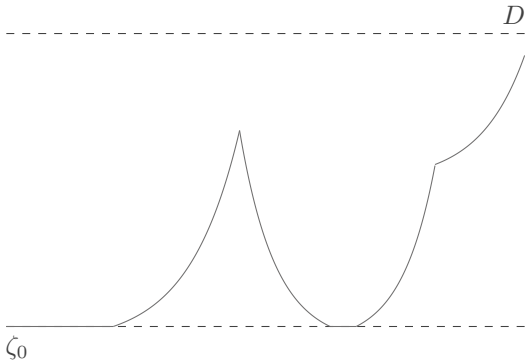
Soit d la décision telle que: $d = (\ell, r)$:

- ℓ le traitement;
- r le temps avant la prochaine visite.

[1] Processus Markovien Déterministe par Morceaux

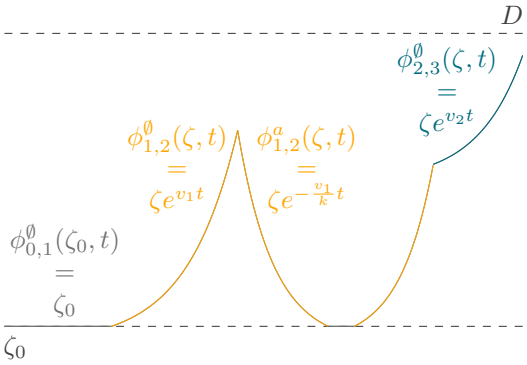
Caractérisation d'un PDMP

Un PDMP se définit par trois caractéristiques locales.



Caractérisation d'un PDMP

Un PDMP se définit par trois caractéristiques locales.



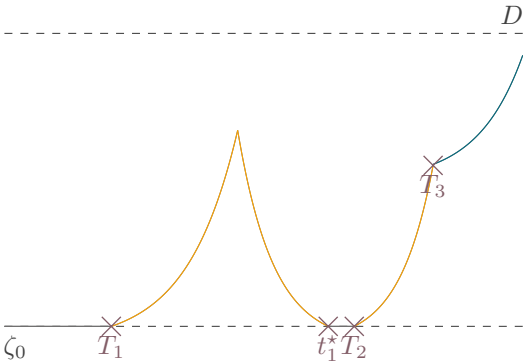
Le flot

Description de la partie déterministe du processus.

$$\Phi^{\ell}(x,t) = (m, k, \phi^{\ell}_{m,k}(\zeta,t), u+t)$$

Caractérisation d'un PDMP

Un PDMP se définit par trois caractéristiques locales.



L'intensité de saut

Description des mécanismes de saut du processus.

- Saut à la frontière (déterministe)

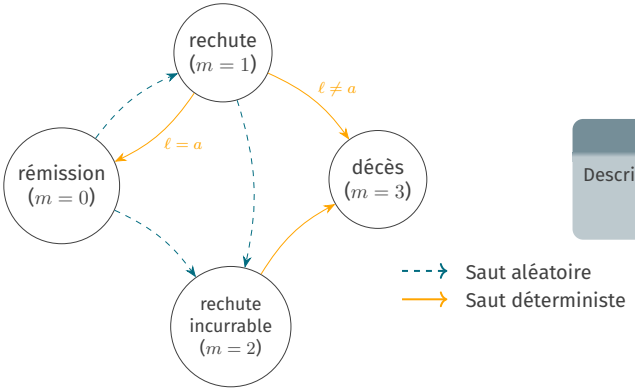
$$t^\star(x) = t_m^{\ell\star}(\zeta) = \inf\{t > 0 : \phi_{m,k}^{\ell}(\zeta, t) \in \{\zeta_0, D\}\}$$

- Saut aléatoire

$$\mathbb{P}(T > t) = e^{-\int_0^t \lambda_m^{\ell}(\Phi^{\ell}(x,s))\,\mathrm{d}s}$$

Caractérisation d'un PDMP

Un PDMP se définit par trois caractéristiques locales.



Le noyau

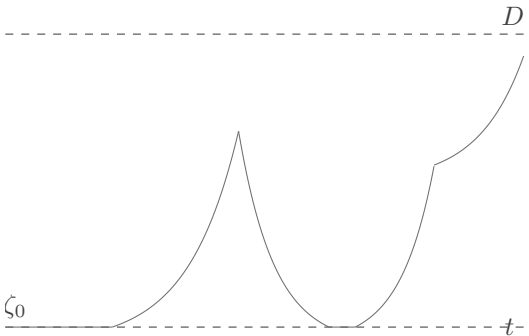
Description de l'état du processus après chaque saut.

$$\mathbb{P}(X' \in A | X = x) = \int_A Q_m^d(\Phi^\ell(x, T), \mathrm{d}x')$$

Sommaire

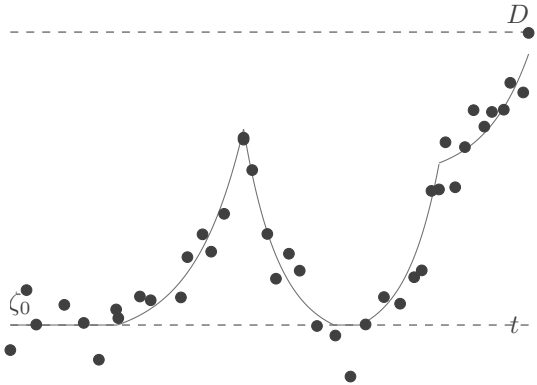
Un processus partiellement observé

L'état de santé du patient n'est pas observé



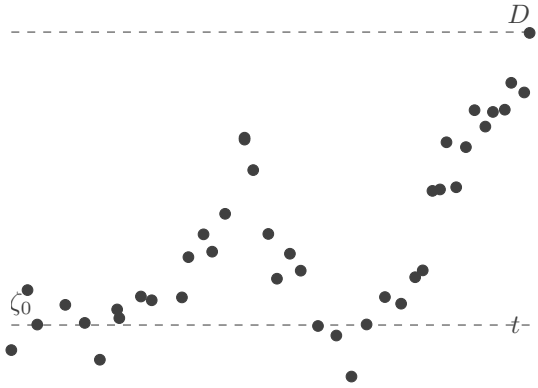
Un processus partiellement observé

L'état de santé du patient n'est pas observé et les mesures sont bruitées.
De plus, les données sont obtenues en temps discret.
De plus, les données sont obtenues en temps discret.



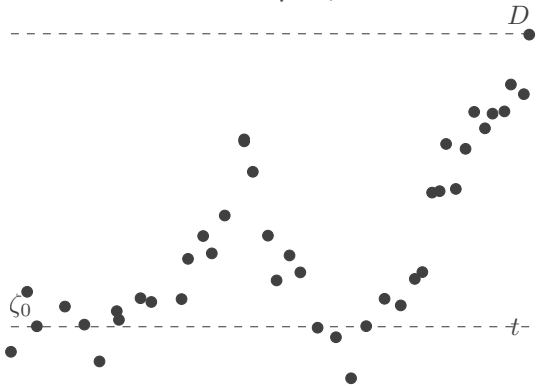
Un processus partiellement observé

L'état de santé du patient n'est pas observé et les mesures sont bruitées.
De plus, les données sont obtenues en temps discret.
De plus, les données sont obtenues en temps discret.



Un processus partiellement observé

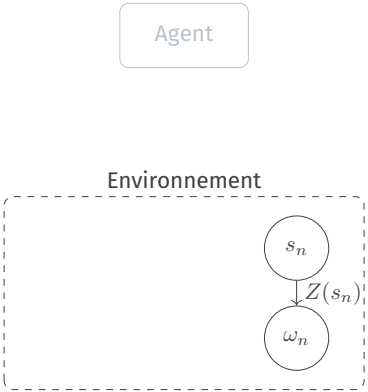
L'état de santé du patient n'est pas observé et les mesures sont bruitées.
De plus, les données sont obtenues en temps discret.
De plus, les données sont obtenues en temps discret.



Il y a des contraintes dans les décisions:

- Une chimiothérapie dure 45 jours au minimum;
- La date du prochain rendez-vous ne peut dépasser la date de suivi;
- Un mort ne reçoit pas de traitement.

Un processus markovien de décision partiellement observé

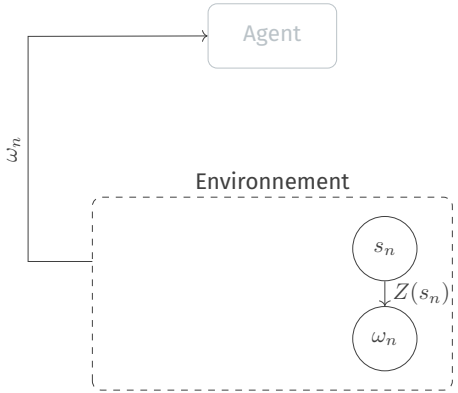


POMDP

Un POMDP se définit par un tuple $(\mathcal{S}, \mathcal{D}, \mathcal{K}, \mathcal{P}, \Omega, \mathcal{Z}, C)$.

- L'état du patient $s = (m, k, \zeta, u, \tau, t)$;
- Les décisions restent inchangées;
- $\mathcal{K}(\omega) \subseteq D$ l'espace des décisions admissibles selon l'observation ω;
- La probabilité de transition $\mathcal{P}(s, d)(s')$;
- Les observations $\omega = (z, F(\zeta), \tau, t)$;
- La fonction d'observations $\mathcal{Z}(s)(\omega)$;
- La fonction de coût $C : \mathcal{D} \times \mathcal{S} \to \mathbb{R}$.

Un processus markovien de décision partiellement observé



POMDP

Un POMDP se définit par un tuple $(\mathcal{S}, \mathcal{D}, \mathcal{K}, \mathcal{P}, \Omega, \mathcal{Z}, C)$.

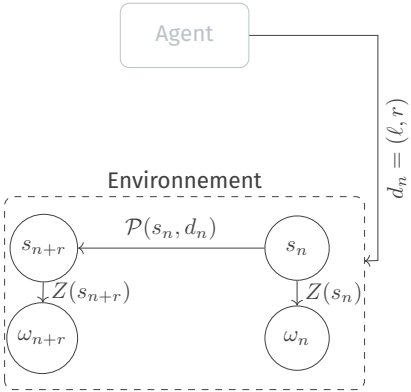
- L'état du patient $s = (m, k, \zeta, u, \tau, t)$;
- Les décisions restent inchangées;
- $\mathcal{K}(\omega) \subseteq D$ l'espace des décisions admissibles selon l'observation ω;
- La probabilité de transition $\mathcal{P}(s, d)(s')$;
- **Les observations** $\omega = (z, F(\zeta), \tau, t)$**;**
- **La fonction d'observations** $\mathcal{Z}(s)(\omega)$**;**
- La fonction de coût $C : \mathcal{D} \times \mathcal{S} \to \mathbb{R}$.

La fonction d'observation est:

$$Z(s_n) = \omega_n = (\mathbb{1}_{m=3}, F(\zeta), \tau, t),$$

avec $F(\zeta) = \zeta e^{\epsilon}$ et où $\epsilon \sim \mathcal{N}(0, 1)$.

Un processus markovien de décision partiellement observé

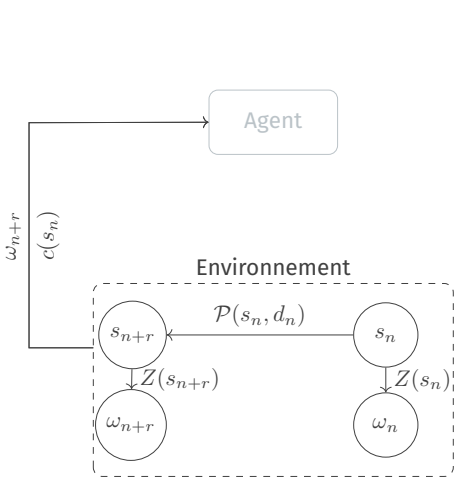


POMDP

Un POMDP se définit par un tuple $(\mathcal{S}, \mathcal{D}, \mathcal{K}, \mathcal{P}, \Omega, \mathcal{Z}, C)$.

- L'état du patient $s = (m, k, \zeta, u, \tau, t)$;
- Les décisions restent inchangées;
- $\mathcal{K}(\omega) \subseteq D$ l'espace des décisions admissibles selon l'observation ω;
- **La probabilité de transition** $\mathcal{P}(s, d)(s')$**;**
- Les observations $\omega = (z, F(\zeta), \tau, t)$;
- La fonction d'observations $\mathcal{Z}(s)(\omega)$;
- La fonction de coût $C : \mathcal{D} \times \mathcal{S} \to \mathbb{R}$.

Un processus markovien de décision partiellement observé

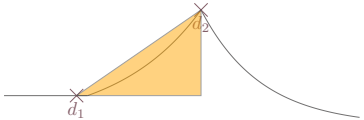


POMDP

Un POMDP se définit par un tuple $(\mathcal{S}, \mathcal{D}, \mathcal{K}, \mathcal{P}, \Omega, \mathcal{Z}, C)$.

- L'état du patient $s = (m, k, \zeta, u, \tau, t)$;
- Les décisions restent inchangées;
- $\mathcal{K}(\omega) \subseteq D$ l'espace des décisions admissibles selon l'observation ω;
- La probabilité de transition $\mathcal{P}(s, d)(s')$;
- Les observations $\omega = (z, F(\zeta), \tau, t)$;
- La fonction d'observations $\mathcal{Z}(s, \omega)$;
- **La fonction de coût** $C : \mathcal{D} \times \mathcal{S} \to \mathbb{R}$.

$$c(s_n) = r \times \frac{1}{2}|\zeta_{n+r} - \zeta_0|(+...)$$



Sommaire

Introduction: apprentissage par renforcement

Objectif:

Identifier une politique $\pi : S \to A$ qui minimise les coûts le long d'une trajectoire

Introduction: apprentissage par renforcement

Objectif:
Identifier une politique $\pi : S \to A$ qui minimise les coûts le long d'une trajectoire

Critère d'optimisation

$$V(\pi, s) = \mathbb{E}_s^{\pi}[\sum_{n=0}^{N-1} c(S_n, d_n, S_{n+1})]$$

Introduction: apprentissage par renforcement

Objectif:
Identifier une politique $\pi : S \to A$ qui minimise les coûts le long d'une trajectoire

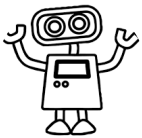
Critère d'optimisation

$$V(\pi, s) = \mathbb{E}_s^{\pi}[\sum_{n=0}^{N-1} c(S_n, d_n, S_{n+1})]$$

Le problème d'optimisation

Soit $\pi^\star$ la politique optimale tel que : $V^\star(s) = \min_{\pi \in \Pi} V(\pi, s)$

L'apprentissage par renforcement

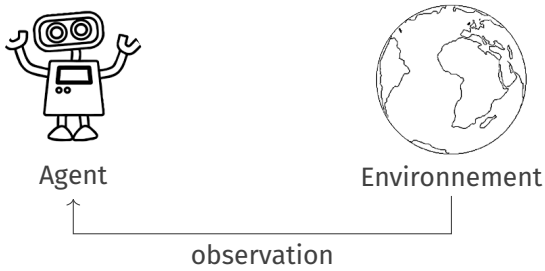


Agent

Environnement

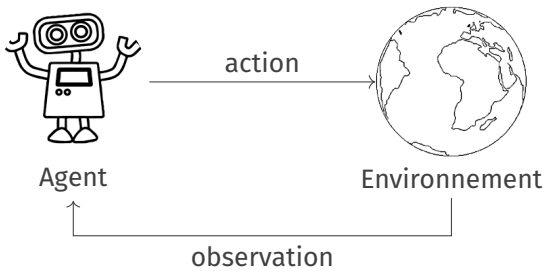
L'objectif est d'apprendre les actions à mener en fonction des expériences passées et des coûts perçus.

L'apprentissage par renforcement



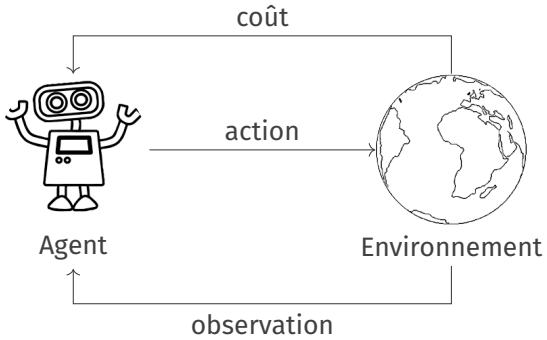
L'objectif est d'apprendre les actions à mener en fonction des expériences passées et des coûts perçus.

L'apprentissage par renforcement



L'objectif est d'apprendre les actions à mener en fonction des expériences passées et des coûts perçus.

L'apprentissage par renforcement



L'objectif est d'apprendre les actions à mener en fonction des expériences passées et des coûts perçus.

L'apprentissage par renforcement pour un POMDP

Le problème d'optimisation

Soit $h_t = (\omega_0, d_0, \omega_1, d_1, \cdots, \omega_t)$ l'historique des observations passées tel que:
$V^\star(h_t) = \min_{\pi \in \Pi} V(\pi, h_t)$

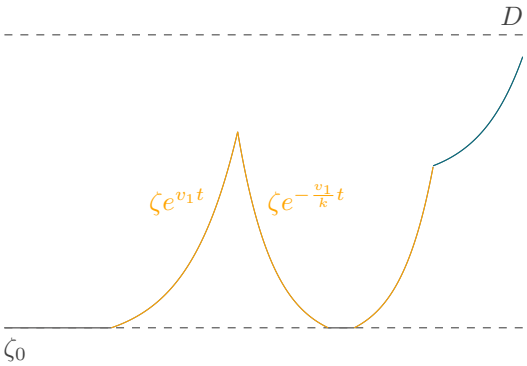
Problème : on a besoin de beaucoup d'intéractions avec l'environnement

Idée : à partir du modèle on simule un environnement pour interagir avec lui.

Sommaire

Un modèle partiellement connu

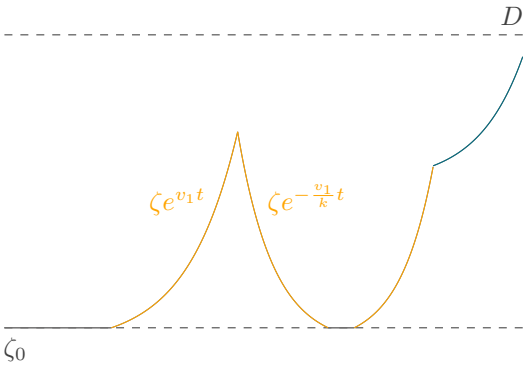
On ne connait pas le paramètre de pente v_1 de la maladie.



Hypothèse:
$v_1 \sim$ log-normale (μ, σ^{-2}).

Un modèle partiellement connu

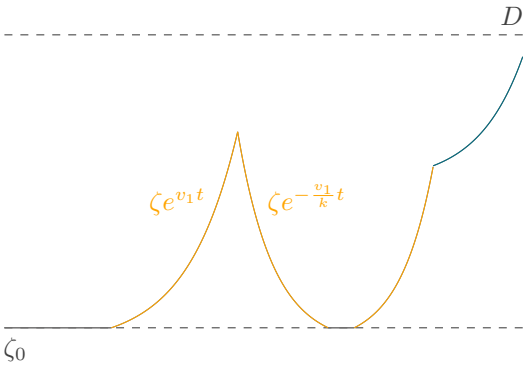
On ne connait pas le paramètre de pente v_1 de la maladie.



Hypothèse:
$v_1 \sim$ log-normale (μ, σ^{-2}).
Inférence bayésienne:
$(\mu, \sigma^{-2}) \sim$ gamma-log-normale$(\alpha, \beta, \kappa, \nu)$.

Un modèle partiellement connu

On ne connait pas le paramètre de pente v_1 de la maladie.

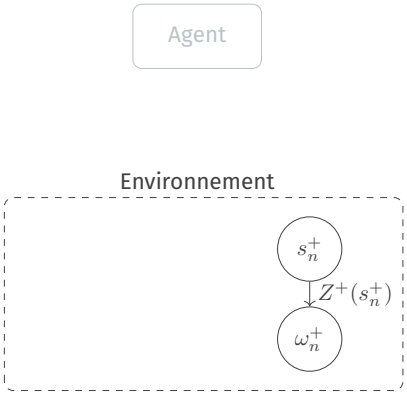


Hypothèse:
$v_1 \sim$ log-normale (μ, σ^{-2}).
Inférence bayésienne:
$(\mu, \sigma^{-2}) \sim$ gamma-log-normale$(\alpha, \beta, \kappa, \nu)$.

Mise à jour des hyperparamètres

- $\alpha_{n+1} = \frac{\beta_n \alpha_n + \log(\hat{v_1})}{\beta_n + 1}$
- $\beta_{n+1} = \beta_n + 1$
- $\kappa_{n+1} = \kappa_n + \frac{1}{2}$
- $\nu_{n+1} = \nu_n + \frac{\beta_n (\log(\hat{v_1} - \alpha_n)^2)}{2(\beta_n + 1)}$

Un BAMDP[2] partiellement observé



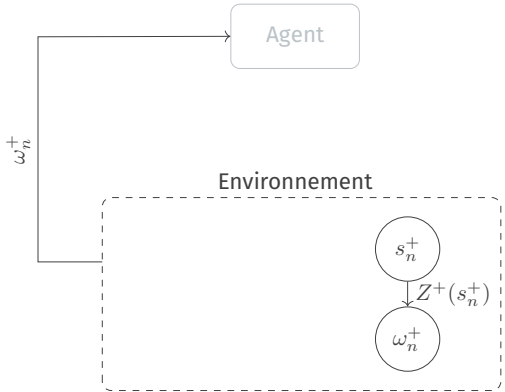
BAMDP PO

Un BAMDP-PO se définit par un tuple $(\mathcal{S}^+, \mathcal{D}, \mathcal{K}, \mathcal{P}^+, \Omega^+, \mathcal{Z}^+, C)$.

- L'hyper-état du patient $s^+ = (m, k, \zeta, u, \tau, t, \alpha, \beta, \kappa, \nu)$;
- Les décisions restent inchangées;
- $\mathcal{K}(\omega) \subseteq D$ l'espace des décisions admissibles selon l'observation ω;
- La probabilité de transition $\mathcal{P}(s^+, d)(s')$;
- Les observations $\omega^+ = (z, F(\zeta), \tau, t, \tilde{\alpha}, \tilde{\beta}, \tilde{\kappa}, \tilde{\nu})$;
- La fonction d'observations $\mathcal{Z}(s^+)(\omega^+)$;
- La fonction de coût $C : \mathcal{D} \times \mathcal{S} \to \mathbb{R}$.

[2] Processus de décision Markovien Bayes adaptatif

Un BAMDP[2] partiellement observé



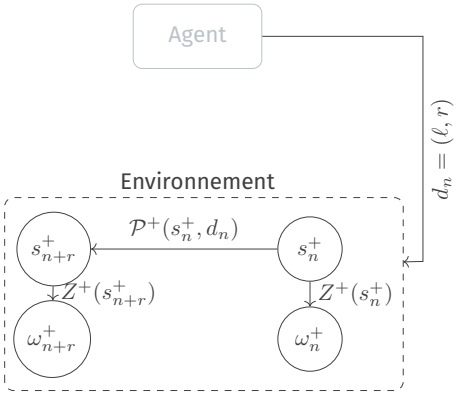
BAMDP PO

Un BAMDP-PO se définit par un tuple $(\mathcal{S}^+, \mathcal{D}, \mathcal{K}, \mathcal{P}^+, \Omega^+, \mathcal{Z}^+, C)$.

- L'hyper-état du patient $s^+ = (m, k, \zeta, u, \tau, t, \alpha, \beta, \kappa, \nu)$;
- Les décisions restent inchangées;
- $\mathcal{K}(\omega) \subseteq D$ l'espace des décisions admissibles selon l'observation ω;
- La probabilité de transition $\mathcal{P}(s^+, d)(s')$;
- Les observations $\omega^+ = (z, F(\zeta), \tau, t, \tilde{\alpha}, \tilde{\beta}, \tilde{\kappa}, \tilde{\nu})$;
- La fonction d'observations $\mathcal{Z}(s^+)(\omega^+)$;
- La fonction de coût $C : \mathcal{D} \times \mathcal{S} \to \mathbb{R}$.

[2] Processus de décision Markovien Bayes adaptatif

Un BAMDP[2] partiellement observé



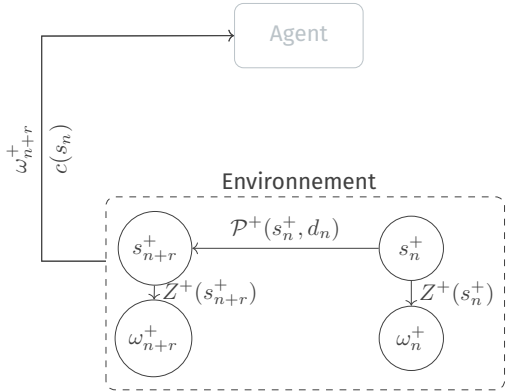
BAMDP PO

Un BAMDP-PO se définit par un tuple $(\mathcal{S}^+, \mathcal{D}, \mathcal{K}, \mathcal{P}^+, \Omega^+, \mathcal{Z}^+, C)$.

- L'hyper-état du patient $s^+ = (m, k, \zeta, u, \tau, t, \alpha, \beta, \kappa, \nu)$;
- Les décisions restent inchangées;
- $\mathcal{K}(\omega) \subseteq D$ l'espace des décisions admissibles selon l'observation ω;
- La probabilité de transition $\mathcal{P}(s^+, d)(s')$;
- Les observations $\omega^+ = (z, F(\zeta), \tau, t, \tilde{\alpha}, \tilde{\beta}, \tilde{\kappa}, \tilde{\nu})$;
- La fonction d'observations $\mathcal{Z}(s^+)(\omega^+)$;
- La fonction de coût $C : \mathcal{D} \times \mathcal{S} \to \mathbb{R}$.

[2] Processus de décision Markovien Bayes adaptatif

Un BAMDP[2] partiellement observé



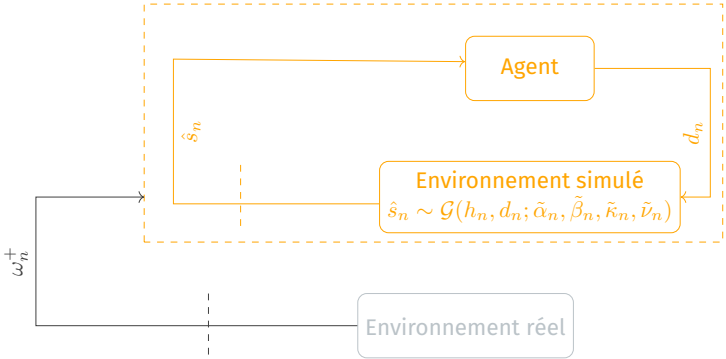
BAMDP PO

Un BAMDP-PO se définit par un tuple $(\mathcal{S}^+, \mathcal{D}, \mathcal{K}, \mathcal{P}^+, \Omega^+, \mathcal{Z}^+, C)$.

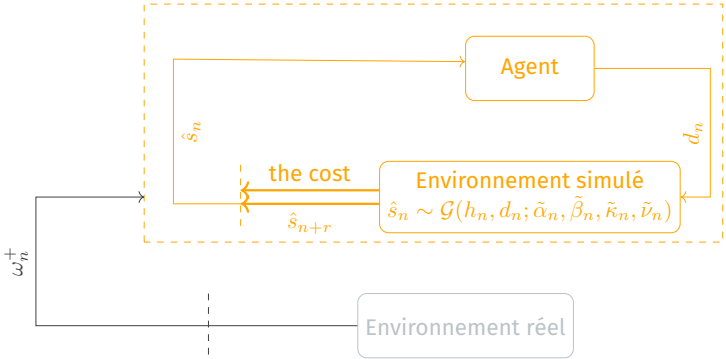
- L'hyper-état du patient $s^+ = (m, k, \zeta, u, \tau, t, \alpha, \beta, \kappa, \nu)$;
- Les décisions restent inchangées;
- $\mathcal{K}(\omega) \subseteq D$ l'espace des décisions admissibles selon l'observation ω;
- La probabilité de transition $\mathcal{P}(s^+, d)(s')$;
- Les observations $\omega^+ = (z, F(\zeta), \tau, t, \tilde{\alpha}, \tilde{\beta}, \tilde{\kappa}, \tilde{\nu})$;
- La fonction d'observations $\mathcal{Z}(s^+)(\omega^+)$;
- La fonction de coût $C : \mathcal{D} \times \mathcal{S} \to \mathbb{R}$.

[2] Processus de décision Markovien Bayes adaptatif

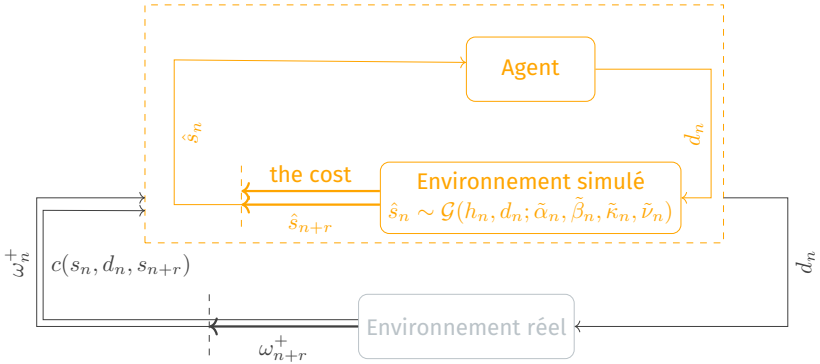
Une suggestion de résolution



Une suggestion de résolution

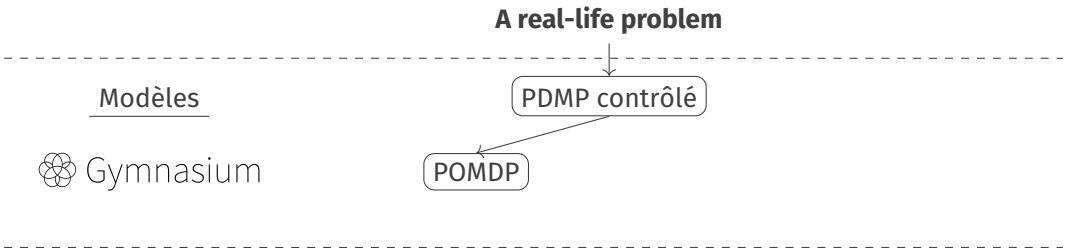


Une suggestion de résolution

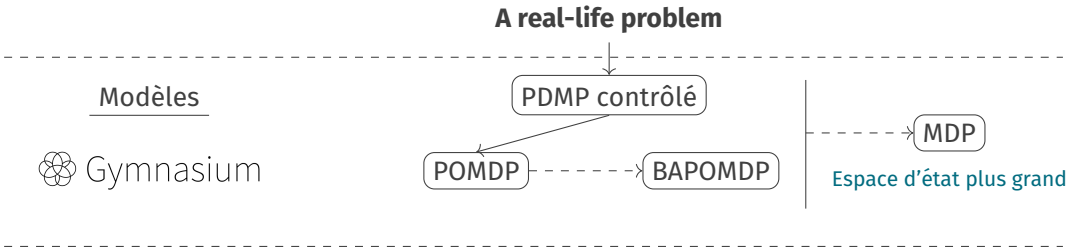


Sommaire

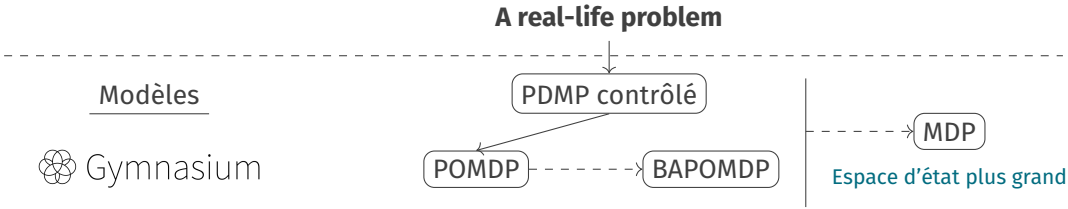
Conclusion et perspectives



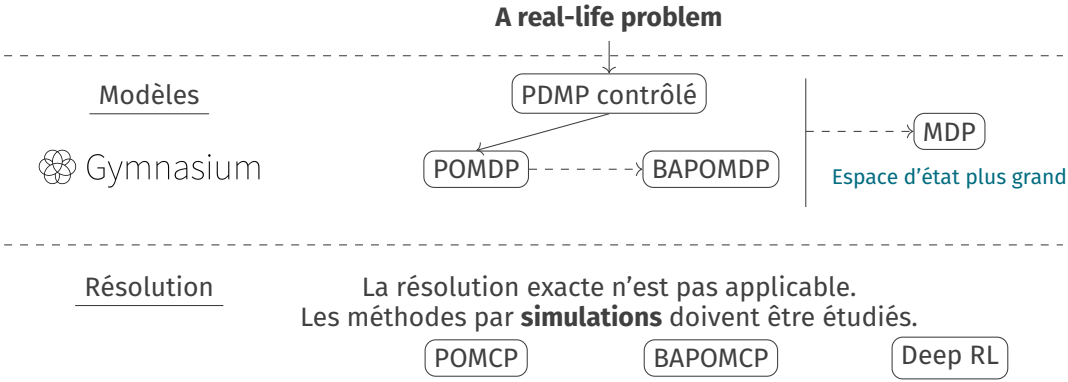
Conclusion et perspectives



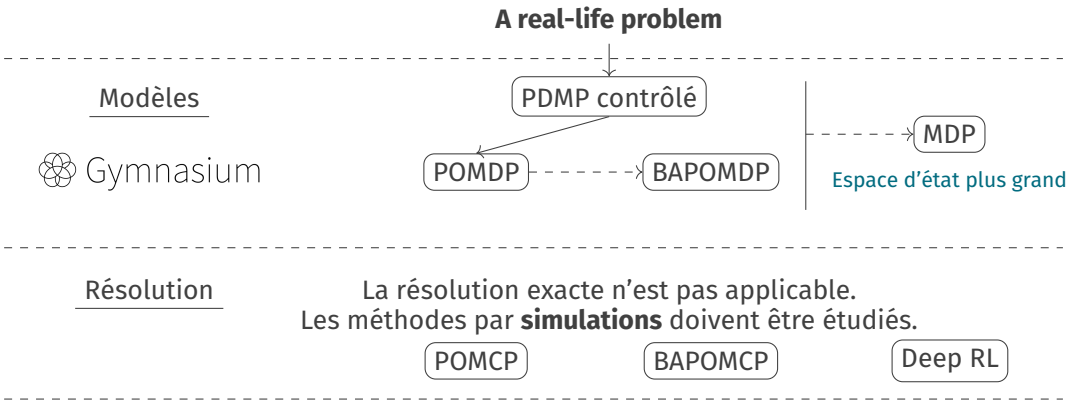
Conclusion et perspectives



Conclusion et perspectives



Conclusion et perspectives



Merci pour votre attention !