

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

¹IMAG, Univ Montpellier, CNRS, Montpellier, France

²John Curtin School of Medical Research, The Australian National University, Canberra, ACT, Australia

³Univ Toulouse, INRAE-MIAT, Toulouse, France

13 Mars 2024



UNIVERSITÉ DE
MONTPELLIER

INRAE

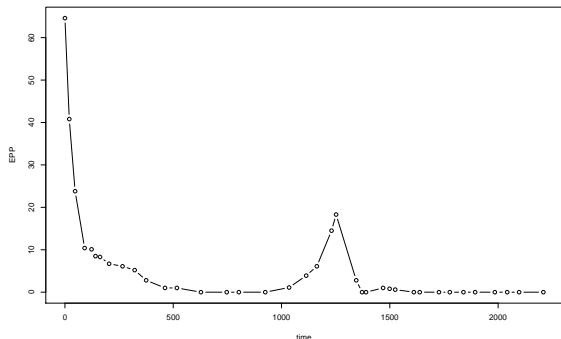
IMAG

INSTITUT MONTPELLIERAIN
ALEXANDER GROTHENDIECK



anr[®]

Le contexte médical

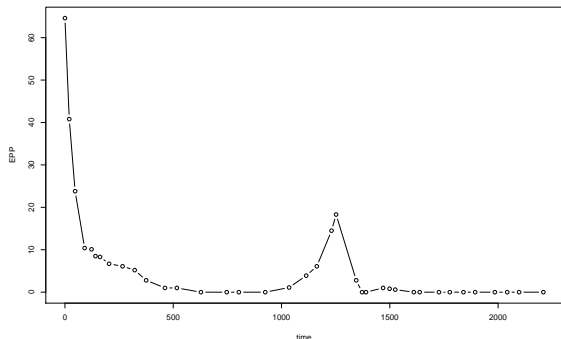


- Des patients ayant eu un **cancer** bénéficient d'un **suivi régulier**;
- La concentration d'**immunoglobuline clonale** est mesurée **dans le temps**;
- Le médecin doit prendre de nouvelles **décisions** à chaque visite.

Figure: Exemple de données d'un patient^a

^aIUCT Oncopole et CRCT, Toulouse, France

Le contexte médical



- 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.

⇒ **Contrôle dynamique stochastique**

Figure: Exemple de données d'un patient^a

^aIUCT Oncopole et CRCT, Toulouse, France

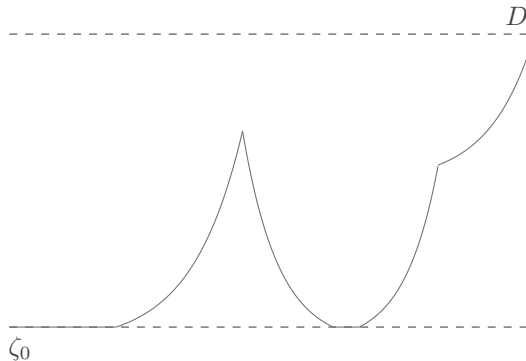
Sommaire

- **Modélisation de la Trajectoire d'un Patient**
- Problème Partiellement Observé et Modèle Inconnu
- Apprentissage par Renforcement
- Conclusion et Perspectives



Le modèle PDMP¹ contrôlé

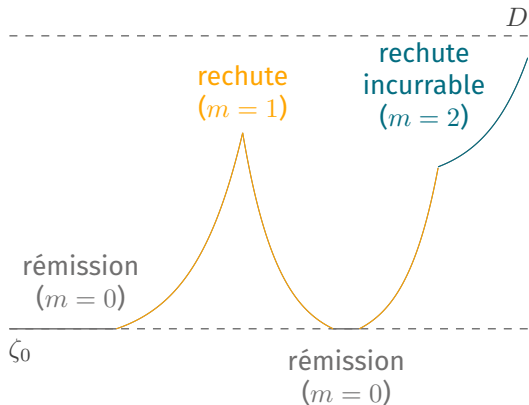
On passe aléatoirement d'un régime déterministe à un autre.



¹Processus Markovien Déterministe par Morceaux

Le modèle PDMP¹ contrôlé

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



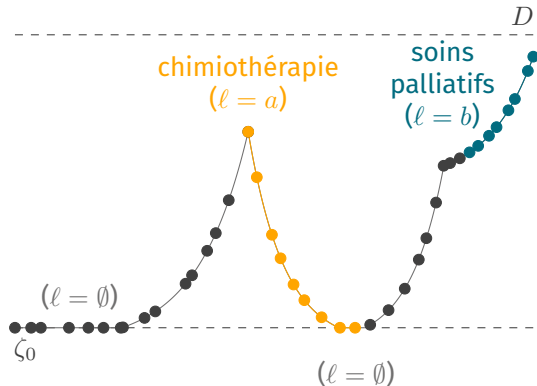
Soit l'état du patient $x = (m, k, \zeta, u)$:

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

¹Processus Markovien Déterministe par Morceaux

Le modèle PDMP¹ 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.

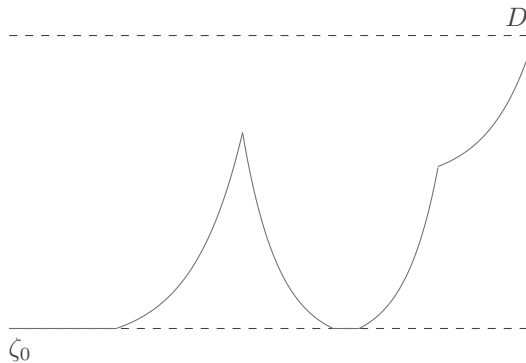
Soit d la décision telle que: $d = (\ell, r)$:

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

¹Processus Markovien Déterministe par Morceaux

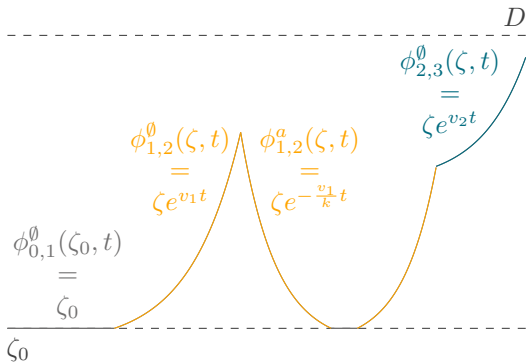
Caractérisation d'un PDMP

Un PDMP se définit par trois caractéristiques locales.



Caractérisation d'un PDMP

Un PDMP se définit par trois caractéristiques locales.



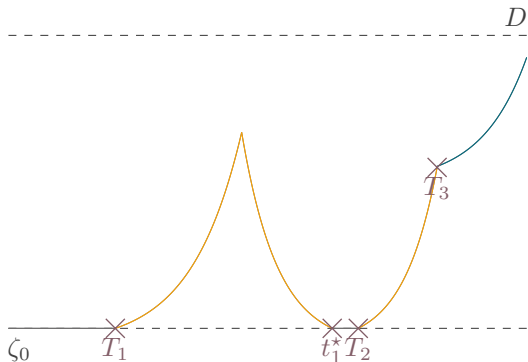
Le flot

Description de la partie déterministe du processus.

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

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)

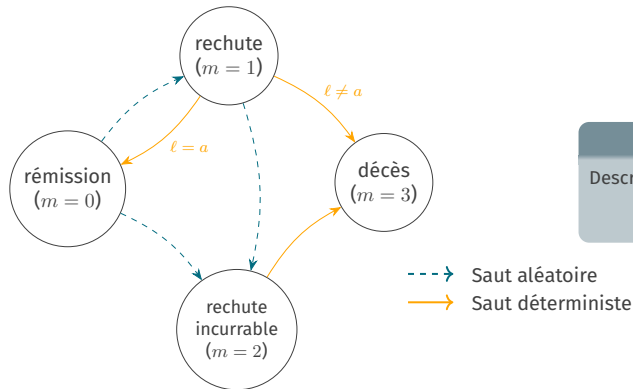
$$t^*(x) = t_m^*(\zeta) = \inf\{t > 0 : \phi_{m,k}^\ell(\zeta, t) \in \{\zeta_0, D\}\}$$

- Saut aléatoire

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

Caractérisation d'un PDMP

Un PDMP se définit par trois caractéristiques locales.



Le noyau

Description de l'état du processus après chaque saut.

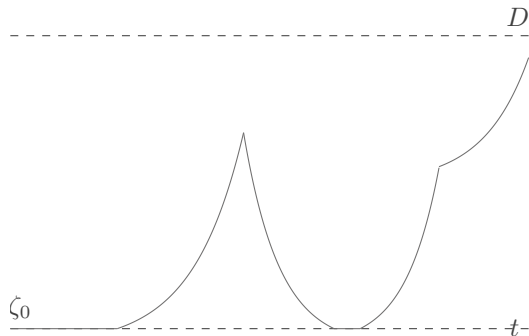
$$\mathbb{P}(X' \in A | X = x) = \int_A Q_m^d(\Phi^\ell(x, T), dx')$$

Sommaire

- ▶ Modélisation de la Trajectoire d'un Patient
- ▶ **Problème Partiellement Observé et Modèle Inconnu**
- ▶ Apprentissage par Renforcement
- ▶ Conclusion et Perspectives

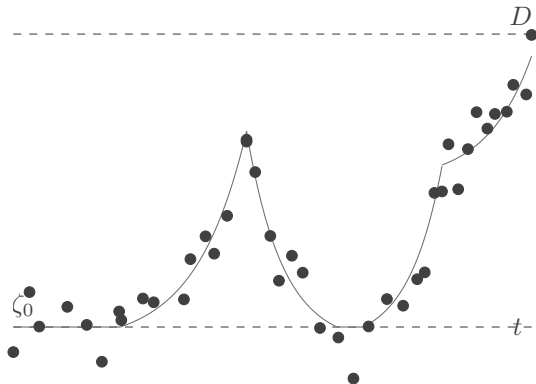


Un processus partiellement observé



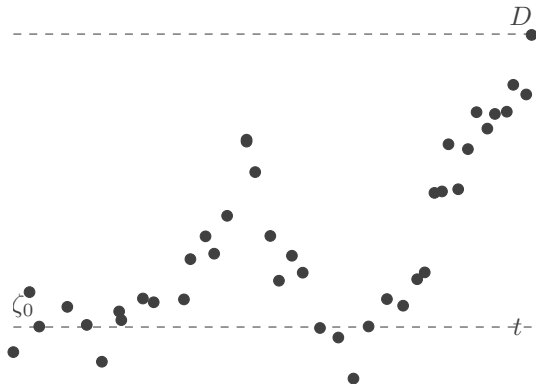
Un processus partiellement observé

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**.

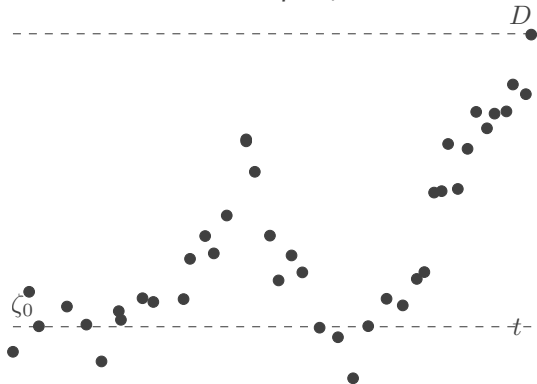


Un processus partiellement observé

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**.

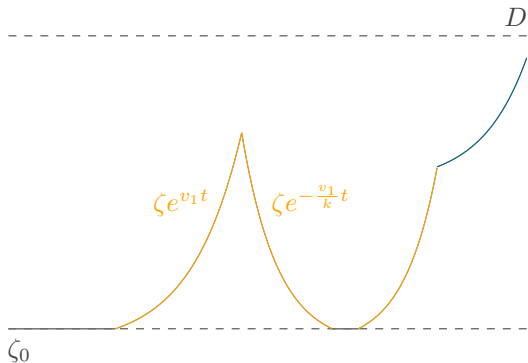


Il y a des **contraintes** dans les décisions:

- 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 modèle partiellement connu

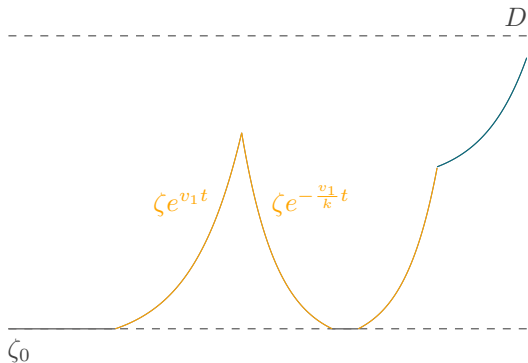
On ne connaît pas le paramètre de pente v_1 de la maladie.



Hypothèse:
 $v_1 \sim \text{log-normale } (\mu, \sigma^{-2}).$

Un modèle partiellement connu

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



Hypothèse:

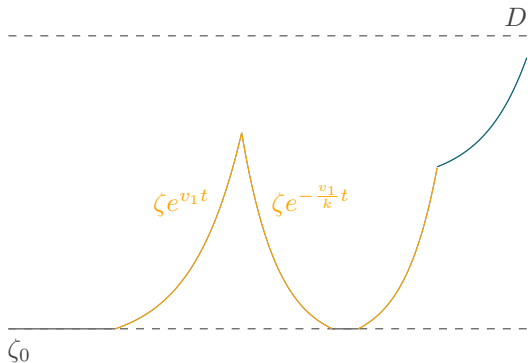
$$v_1 \sim \text{log-normale}(\mu, \sigma^{-2}).$$

Inférence bayésienne:

$$(\mu, \sigma^{-2}) \sim \text{gamma-log-normale}(\alpha, \beta, \kappa, \nu).$$

Un modèle partiellement connu

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



Hypothèse:

$$v_1 \sim \text{log-normale}(\mu, \sigma^{-2}).$$

Inférence bayésienne:

$$(\mu, \sigma^{-2}) \sim \text{gamma-log-normale}(\alpha, \beta, \kappa, \nu).$$

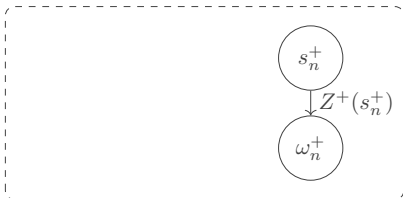
Mise à jour des hyperparamètres

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

Un BAMDP² partiellement observé

Agent

Environnement



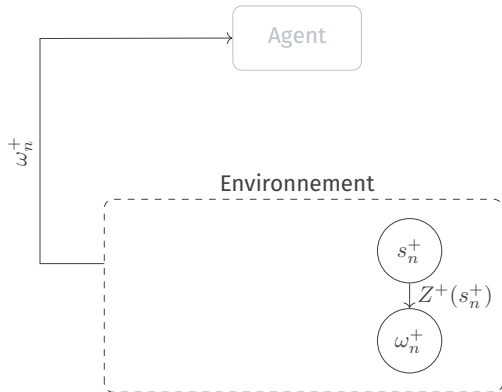
BAMDP PO

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;
- $\mathcal{K}(\omega) \subseteq \mathcal{D}$ 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 $C : \mathcal{D} \times \mathcal{S} \rightarrow \mathbb{R}$.

²Processus de décision Markovien Bayes adaptatif

Un BAMDP² 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;
- $\mathcal{K}(\omega) \subseteq \mathcal{D}$ 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 $C : \mathcal{D} \times \mathcal{S} \rightarrow \mathbb{R}$.

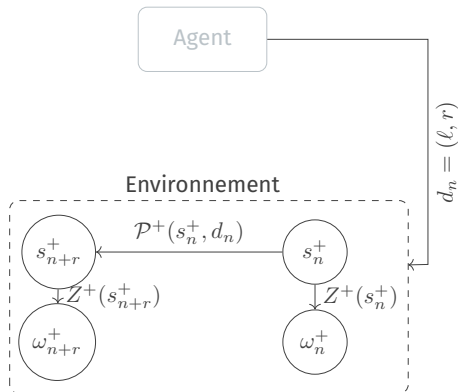
La fonction d'observation est:

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

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

²Processus de décision Markovien Bayes adaptatif

Un BAMDP² partiellement observé



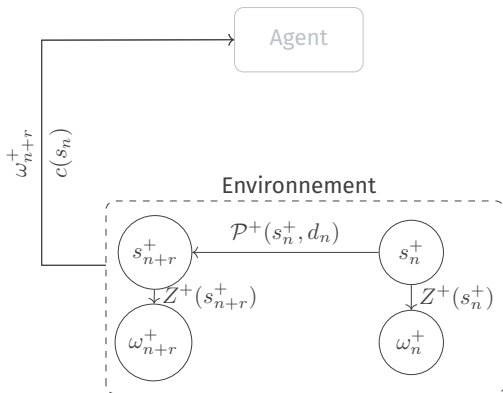
BAMDP PO

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;
- $\mathcal{K}(\omega) \subseteq \mathcal{D}$ 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 $C : \mathcal{D} \times \mathcal{S} \rightarrow \mathbb{R}$.

²Processus de décision Markovien Bayes adaptatif

Un BAMDP² partiellement observé

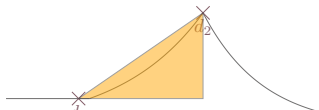


BAMDP PO

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

- L'hyper-état du patient $s^+ = (m, k, \zeta, u, \tau, t, \alpha, \beta, \kappa, \nu)$;
- Les décisions restent inchangées;
- $\mathcal{K}(\omega) \subseteq \mathcal{D}$ 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** $C : \mathcal{D} \times S \rightarrow \mathbb{R}$.

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



Sommaire

- ▶ Modélisation de la Trajectoire d'un Patient
- ▶ Problème Partiellement Observé et Modèle Inconnu
- ▶ **Apprentissage par Renforcement**
- ▶ Conclusion et Perspectives

Introduction à l'apprentissage par renforcement

Objectif:

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

Introduction à l'apprentissage par renforcement

Objectif:

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

Critère d'optimisation

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

Introduction à l'apprentissage par renforcement

Objectif:

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

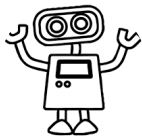
Critère d'optimisation

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

Le problème d'optimisation

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

Introduction à l'apprentissage par renforcement



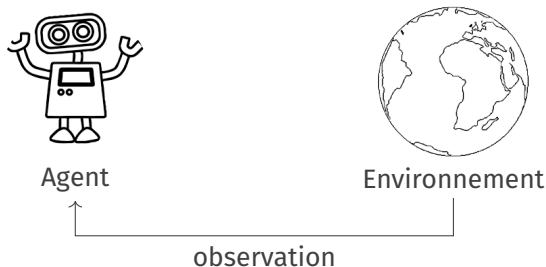
Agent



Environnement

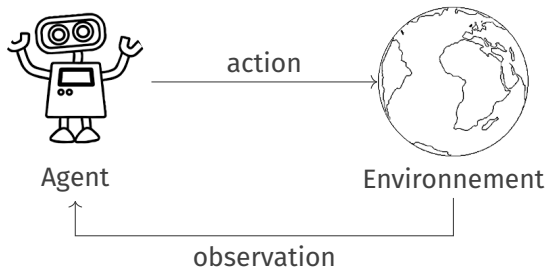
L'objectif est d'**apprendre les actions** à mener en fonction des expériences passées et des coûts perçus.

Introduction à l'apprentissage par renforcement



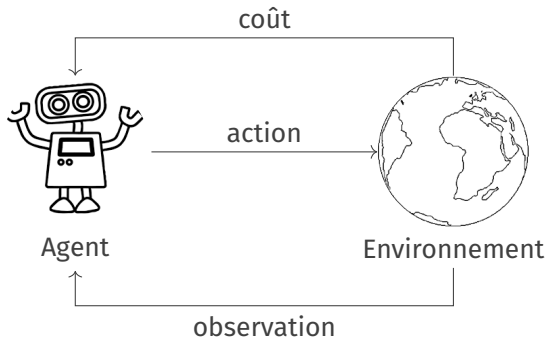
L'objectif est d'**apprendre les actions** à mener en fonction des expériences passées et des coûts perçus.

Introduction à l'apprentissage par renforcement



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

Introduction à l'apprentissage par renforcement



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

Pour un MDP partiellement observé

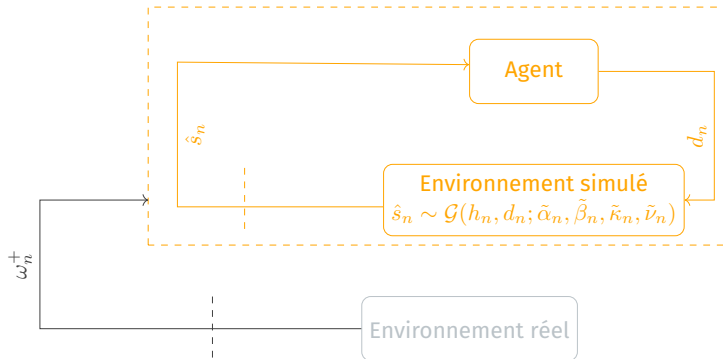
Le problème d'optimisation

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

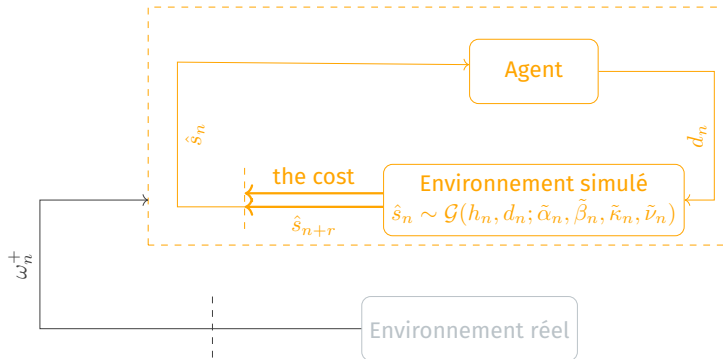
Problème : on a besoin de **beaucoup d'interactions** avec l'environnement

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

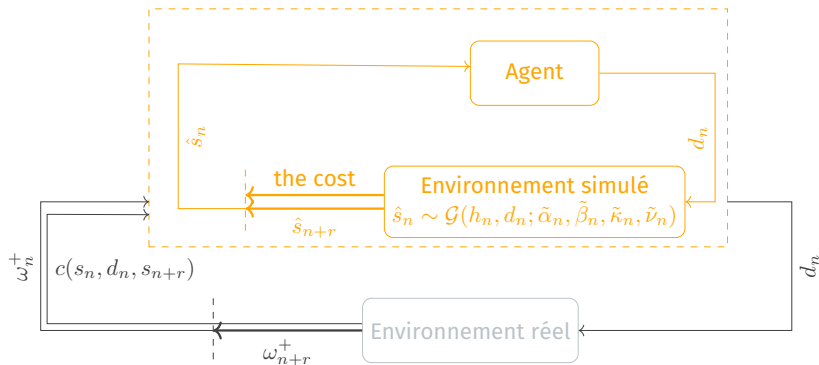
Une suggestion de résolution



Une suggestion de résolution



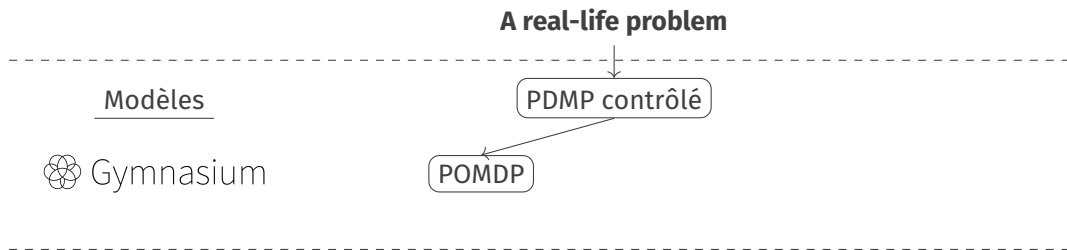
Une suggestion de résolution



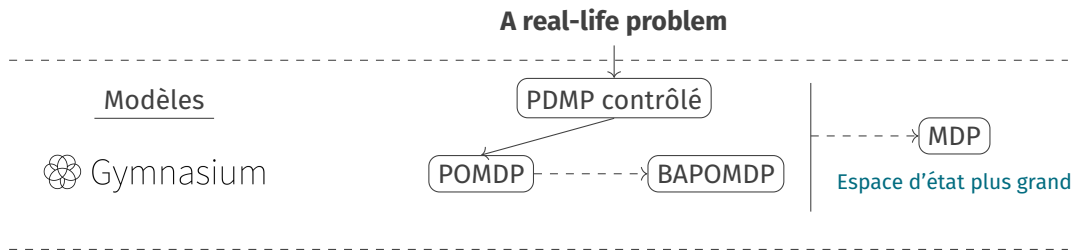
Sommaire

- ▶ Modélisation de la Trajectoire d'un Patient
- ▶ Problème Partiellement Observé et Modèle Inconnu
- ▶ Apprentissage par Renforcement
- ▶ Conclusion et Perspectives

Conclusion et perspectives

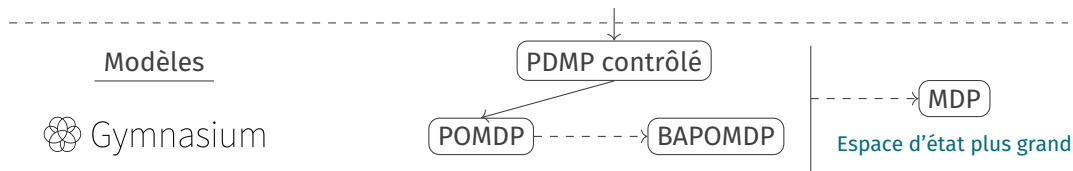


Conclusion et perspectives



Conclusion et perspectives

A real-life problem



Résolution

La résolution exacte n'est pas applicable.
Les méthodes par **simulations** doivent être étudiés.



Meritxell Vinyals

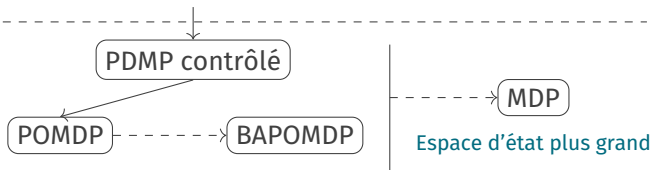
Conclusion et perspectives

A real-life problem

Modèles



Gymnasium



Résolution



Meritxell Vinyals

La résolution exacte n'est pas applicable.
Les méthodes par **simulations** doivent être étudiées.

POMCP

BAPOMCP

Deep RL

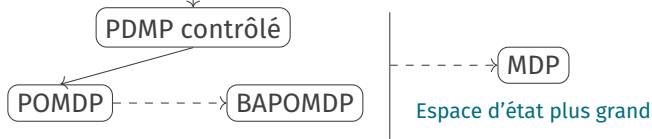
Conclusion et perspectives

A real-life problem

Modèles



Gymnasium



Résolution



Meritxell Vinyals

La résolution exacte n'est pas applicable.
Les méthodes par **simulations** doivent être étudiées.

POMCP

BAPOMCP

Deep RL

Merci pour votre attention !