

Contrôle stochastique pour l'optimisation de traitement médical

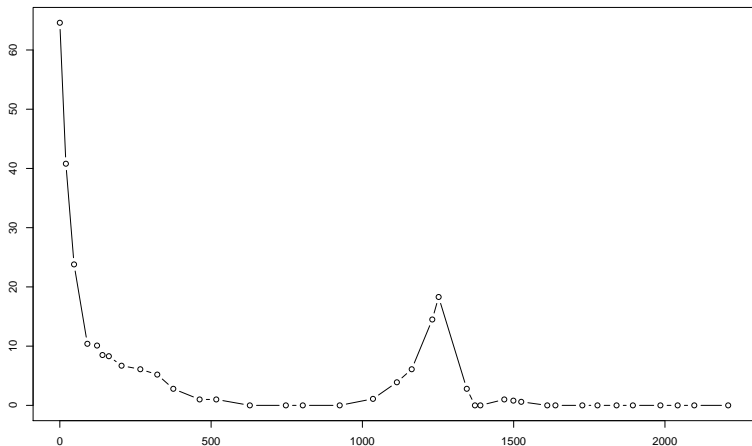
Alice Cleynen, Benoîte de Saporta

Institut Montpelliérain Alexander Grothendieck
CNRS, Univ. Montpellier, France



Motivation : Myélome multiple

Essai clinique : Intergroupe Francophone du Myélome 2009



Mesure de l'immunoglobuline monoclonale en fonction du temps
pour un-e patient-e

Source : Centre de Recherche en Cancérologie de Toulouse

Optimisation de traitement médical

En utilisant les **mesures** de l'immunoglobuline monoclonale **choisir** de façon séquentielle

- ▶ le **traitement** approprié à appliquer jusqu'à la prochaine visite au centre médical
- ▶ la **date** de la prochaine visite au centre médical

pour **améliorer** l'état de santé : maintenir à un faible niveau l'immunoglobuline monoclonale tout en minimisant les **effets secondaires indésirables** et les **contraintes/coûts** de traitement

Difficultés

- ▶ **détection des rechutes** : état de santé global **aléatoire** et **non observable directement** : utilisation de l'immunoglobuline monoclonale comme **marqueur** de la maladie, évolution **continue** du marqueur, mais observations à des dates **peu fréquentes** et les observations peuvent être **bruitées**

Difficultés

- ▶ **détection des rechutes** : état de santé global **aléatoire** et **non observable directement** : utilisation de l'immunoglobuline monoclonale comme **marqueur** de la maladie, évolution **continue** du marqueur, mais observations à des dates **peu fréquentes** et les observations peuvent être **bruitées**
- ▶ **détection du type de rechute / choix du traitement** **plusieurs** types de rechutes et de traitement possibles

Difficultés

- ▶ **détection des rechutes** : état de santé global **aléatoire** et **non observable directement** : utilisation de l'immunoglobuline monoclonale comme **marqueur** de la maladie, évolution **continue** du marqueur, mais observations à des dates **peu fréquentes** et les observations peuvent être **bruitées**
- ▶ **détection du type de rechute / choix du traitement** **plusieurs** types de rechutes et de traitement possibles
- ▶ **choix de la prochaine date de visite / durée du traitement** : compromis non trivial entre
 - ▶ **trop tôt** traitements lourds avec effets secondaires importants
 - ▶ **trop tard** risque accru de décès si la maladie n'est pas traitée à temps

Approche

- ▶ proposer un **modèle** de l'évolution de l'état de santé et du marqueur
- ▶ **formuler** le problème d'optimisation de traitement comme un problème de **contrôle stochastique**
- ▶ proposer une méthode de **numérique** pour construire une politique **explicite** proche de l'optimalité
- ▶ étudier les performances de cette politique sur données **simulées**

Plan de l'exposé

Modélisation du problème de contrôle

Dynamique en temps continu de l'état de santé et du marqueur

Dynamique à temps discret

Construction d'une politique proche de l'optimalité

Résultats numériques

Conclusion et perspectives

Variables d'intérêt

- ▶ **mode m** : état de santé global
 - ▶ $m = 0$: sain/rémission,
 - ▶ $m = 1$: maladie/rechute de type 1
 - ▶ $m = 2$: maladie/rechute de type 2
 - ▶ $m = 3$: mort

- ▶ **marqueur ζ** : immunoglobuline monoclonale $\zeta \in [\zeta_0, D]$
 - ▶ $\zeta = \zeta_0$: valeur nominale dans l'état sain/rémission
 - ▶ $\zeta = D$: seuil de mort

Traitements

La dynamique du marqueur dépend de l'état de santé global et du traitement choisi

Traitements l possibles

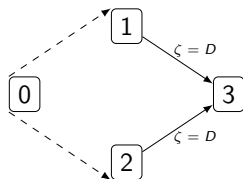
- ▶ $l = \emptyset$: pas de traitement
- ▶ $l = a$: traitement a
 - ▶ efficace sur la maladie 1
 - ▶ ralentit l'évolution de la maladie 2
- ▶ $l = b$: traitement b
 - ▶ efficace sur la maladie 2
 - ▶ ralentit l'évolution de la maladie 1

Evolution de l'état de santé global

Intensité et noyau de saut

L'état de santé global est **constant par morceaux**, il change à des instants **déterministes** (traits pleins) ou **aléatoires** (tirets) dont la loi dépend

- ▶ de la valeur courante du **marqueur** ζ et / ou du **temps** passé dans cet état m (nécessite une variable supplémentaire)
- ▶ et du **traitement** appliqué ℓ



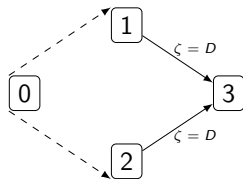
Transitions possibles
sans traitement

Evolution de l'état de santé global

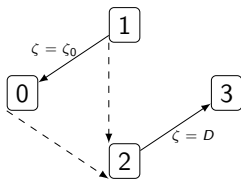
Intensité et noyau de saut

L'état de santé global est **constant par morceaux**, il change à des instants **déterministes** (traits pleins) ou **aléatoires** (tirets) dont la loi dépend

- ▶ de la valeur courante du **marqueur** ζ et / ou du **temps** passé dans cet état m (nécessite une variable supplémentaire)
- ▶ et du **traitement** appliqué ℓ



Transitions possibles
sans traitement



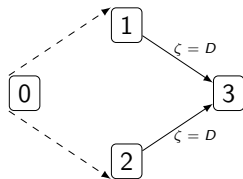
Transitions possibles
sous traitement a

Evolution de l'état de santé global

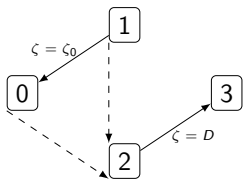
Intensité et noyau de saut

L'état de santé global est **constant par morceaux**, il change à des instants **déterministes** (traits pleins) ou **aléatoires** (tirets) dont la loi dépend

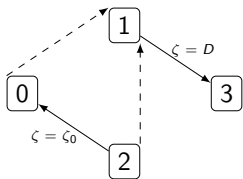
- ▶ de la valeur courante du **marqueur** ζ et / ou du **temps** passé dans cet état m (nécessite une variable supplémentaire)
- ▶ et du **traitement** appliqué ℓ



Transitions possibles
sans traitement



Transitions possibles
sous traitement a



Transitions possibles
sous traitement b

Evolution du marqueur

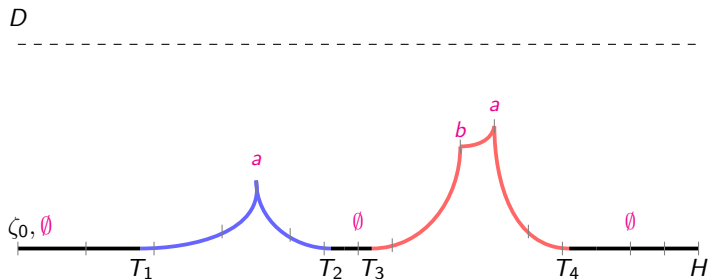
Flott

Conditionnellement à la valeur courante du mode m et du traitement ℓ , l'évolution du marqueur est déterministe

$$\zeta(t+s) = \zeta(t) \exp(v_m^\ell s)$$

	$\ell = \emptyset$	$\ell = a$	$\ell = b$
$m = 0$	$v_0^\emptyset = 0$	$v_0^a = 0$	$v_0^b = 0$
$m = 1$	$v_1^\emptyset > 0$	$v_1^a < 0$	$0 < v_1^b < v_1^\emptyset$
$m = 2$	$v_2^\emptyset > 0$	$0 < v_2^a < v_2^\emptyset$	$v_2^b < 0$
$m = 3$	$v_3 = 0$		

Exemple de trajectoire contrôlée



Problème de contrôle stochastique

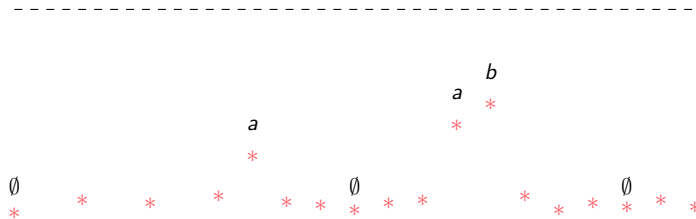
Décider séquentiellement

- ▶ quel **traitement** appliquer
- ▶ pour combien de temps

pour maintenir le marqueur proche de ζ_0 , alors que

- ▶ le mode est caché
- ▶ les dates de saut sont cachées
- ▶ les observations du marqueur sont **bruitées** et **peu fréquentes**

Exemple d'information disponible

 D 

Contrôle en temps continu

Modélisation en temps continu

- ▶ proche de la réalité biologique
- ▶ permet de modéliser le problème avec **peu** de paramètres qui sont tous interprétables
- ▶ cadre théorique et numérique du **contrôle impulsionnel** pour les **processus Markoviens déterministes par morceaux** (PDMP) bien défini dans un cadre **d'observation parfaite à tout instant** [Davis 93],[dSDZ 14] ou sous observation partielle lorsque **les dates de saut sont observées** [dSDZ 14], [BL 17]
 - ▶ choisir les dates des impulsions
 - ▶ choisir les nouvelles positions après impulsion
- ▶ construction théorique et numérique explicite de politiques ϵ -optimales **très difficile** dès que les dates de saut ne sont **pas observées**

Contrôle en temps continu

Modélisation en temps continu

- ▶ proche de la réalité biologique
- ▶ permet de modéliser le problème avec **peu** de paramètres qui sont tous interprétables
- ▶ cadre théorique et numérique du **contrôle impulsionnel** pour les **processus Markoviens déterministes par morceaux** (PDMP) bien défini dans un cadre **d'observation parfaite à tout instant** [Davis 93],[dSDZ 14] ou sous observation partielle lorsque **les dates de saut sont observées** [dSDZ 14], [BL 17]
 - ▶ choisir les dates des impulsions
 - ▶ choisir les nouvelles positions après impulsion
- ▶ construction théorique et numérique explicite de politiques ϵ -optimales **très difficile** dès que les dates de saut ne sont **pas observées**

Simplifier le problème : limiter drastiquement le nombre de possibilités pour les durées de traitement

Plan de l'exposé

Modélisation du problème de contrôle

Dynamique en temps continu de l'état de santé et du marqueur

Dynamique à temps discret

Construction d'une politique proche de l'optimalité

Résultats numériques

Conclusion et perspectives

Délais jusqu'à la prochaine visite

δ délai minimal entre deux observations

- ▶ δ n'est pas petit : 15 jours pour le modèle du myélome
- ▶ choix du délai r jusqu'à la prochaine date de visite **restreint** parmi les multiples de δ : $r \in \{15, 30, 60\}$ jours

Délais jusqu'à la prochaine visite

δ délai minimal entre deux observations

- ▶ δ n'est pas petit : 15 jours pour le modèle du myélome
- ▶ choix du délai r jusqu'à la prochaine date de visite restreint parmi les multiples de δ : $r \in \{15, 30, 60\}$ jours

On n'a besoin que des transitions de t à $t + 15$, $t + 30$ et $t + 60$ du processus continu : besoin uniquement de connaître l'état du processus aux instants discrets multiples de δ

Délais jusqu'à la prochaine visite

δ délai minimal entre deux observations

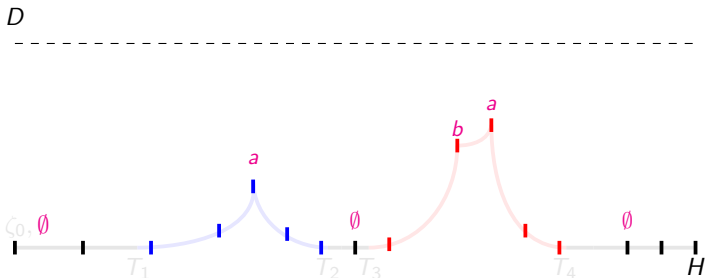
- ▶ δ n'est pas petit : 15 jours pour le modèle du myélome
- ▶ choix du délai r jusqu'à la prochaine date de visite **restreint** parmi les multiples de δ : $r \in \{15, 30, 60\}$ jours

On n'a besoin que des transitions de t à $t + 15$, $t + 30$ et $t + 60$ du processus continu : besoin uniquement de connaître l'état du processus aux instants **discrets** multiples de δ

Il ne s'agit pas d'une **approximation** discrète du processus en temps continu, mais de regarder le vrai processus avec sa dynamique continue uniquement à des dates discrètes

Processus à temps discret

Il ne s'agit pas d'une **approximation** discrète du processus en temps continu, mais de regarder le **vrai processus** avec sa dynamique continue uniquement à des **dates discrètes**. En particulier, le processus peut changer de mode entre deux dates



Processus et observations à temps discret

Initialisation

- ▶ $X_0 = (m(0), \zeta(0)) = (0, \zeta_0)$ état de santé initial, et valeur initiale du marqueur + variables additionnelles
- ▶ $Y_0 = \zeta(0) + \epsilon_0$ observation initiale, où ϵ_0 bruit aléatoire + variables ad.
- ▶ $n \leftarrow 0$ étape courante, $t \leftarrow 0$ date courante

Processus et observations à temps discret

Initialisation

- ▶ $X_0 = (m(0), \zeta(0)) = (0, \zeta_0)$ état de santé initial, et valeur initiale du marqueur + variables additionnelles
- ▶ $Y_0 = \zeta(0) + \epsilon_0$ observation initiale, où ϵ_0 bruit aléatoire + variables ad.
- ▶ $n \leftarrow 0$ étape courante, $t \leftarrow 0$ date courante

Itérations

 Sachant la valeur courante de n , t , X_n et Y_n

- ▶ en utilisant seulement les observations Y_0, \dots, Y_n , choisir la prochaine décision $A_n = (\ell, r)$ ℓ = traitement, r = délai jusqu'à la prochaine étape
- ▶ tirer la prochaine valeur cachée : $X_{n+1} = (m(t+r), \zeta(t+r))$ suivant la dynamique continue, conditionnellement à $X_n = (m(t), \zeta(t))$ et au traitement ℓ
- ▶ tirer la prochaine observation : $Y_{n+1} = \zeta(t+r) + \epsilon_{n+1}$
- ▶ $n \leftarrow n+1$, $t \leftarrow t+r$

Problème d'optimisation

Processus Markovien décisionnel (MDP)

- ▶ Une **politique** admissible π est une suite de règles de décision ne prenant en compte que les observations disponibles à chaque instant. On note Π l'ensemble des politiques admissibles
- ▶ L'**horizon** d'optimisation est $H = N\delta$
- ▶ c est la fonction de **coût instantané**, C la fonction de **coût terminal**

$$V(x, y) = \inf_{\pi \in \Pi} \mathbb{E}_{(x, y)}^{\pi} \left[\sum_{n=0}^{N-1} c(X_n, Y_n, A_n) + C(X_N, Y_N) \right]$$

On cherche une **approximation calculable** de la **fonction valeur** V et une politique π^* **calculable** proche du minimum

Plan de l'exposé

Modélisation du problème de contrôle

Construction d'une politique proche de l'optimalité

Espace des croyances

Discrétisations

Résultats numériques

Conclusion et perspectives

Problème 1 : observations partielles

Le processus contient une partie **cachée** X_n et une partie **observée** Y_n : problème de contrôle partiellement observé POMPD

Problème 1 : observations partielles

Le processus contient une partie **cachée** X_n et une partie **observée** Y_n : problème de contrôle partiellement observé POMPD

Solution classique : se ramener à un problème complètement observé sur un espace plus grand contenant l'**espace des croyances** en utilisant le processus des croyances, ou **filtre** [BR 11], [CdS 18]

$$\Theta_n(\cdot) = \mathbb{P}(X_n \in \cdot \mid Y_0, Y_1, \dots, Y_n)$$

C'est un processus à valeurs dans un **espace de lois probabilités** qui mesure la croyance que l'on a en les valeurs possibles du processus sachant les observations.

Formule de récurrence explicite mais inutilisable numériquement qui relie Θ_{n+1} à Θ_n et Y_{n+1}

Equivalence avec le problème totalement observé

$$V(x, y) = V'(\delta_x, y) = \inf_{\pi \in \Pi} \mathbb{E}_{(\delta_x, y)}^{\pi} \left[\sum_{n=0}^{N-1} c'(\Theta_n, Y_n, A_n) + C'(\theta_N, Y_N) \right]$$

avec

- ▶ $c'(\theta, y, a) = \int c(x, y, a) \theta(dx)$
- ▶ $C'(\theta, y) = \int C(x, y) \theta(dx)$

Programmation dynamique

V' peut se calculer récursivement par programmation dynamique :
on pose

$$V'_N(\theta, y) = C'(\theta, y)$$

$$V'_n(\theta, y) = \inf_a \{c'(\theta, y, a) + R'V'_{n+1}(\theta, y, a)\}$$

alors $V'_0 = V'$ et une politique optimale est donnée par les arginf à chaque étape

R' est le noyau de transition contrôlé de la suite (Θ_n, Y_n) défini par

$$R'f(\theta, y', a) = \mathbb{E}[f(\Theta_{n+1}, Y_{n+1}) | (\Theta_n, Y_n) = (\theta, y), A_n = a]$$

Plan de l'exposé

Modélisation du problème de contrôle

Construction d'une politique proche de l'optimalité

Espace des croyances

Discrétisations

Résultats numériques

Conclusion et perspectives

Problème 2 : dimension infinie

Θ_n vit dans un espace de dimension infinie, on ne sait pas intégrer analytiquement ni numériquement contre le noyau R'

- ▶ on ne sait pas résoudre la programmation dynamique
- ▶ on ne sait pas simuler le filtre

Problème 2 : dimension infinie

Θ_n vit dans un espace de dimension infinie, on ne sait pas intégrer analytiquement ni numériquement contre le noyau R'

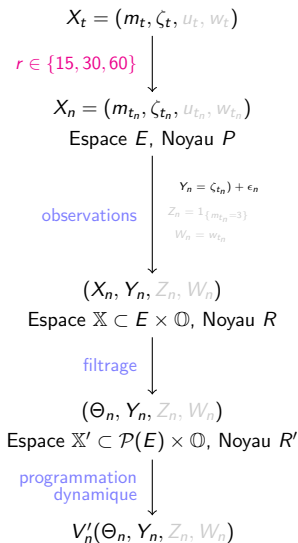
- ▶ on ne sait pas résoudre la programmation dynamique
- ▶ on ne sait pas simuler le filtre

Solution : discrétisation en deux temps

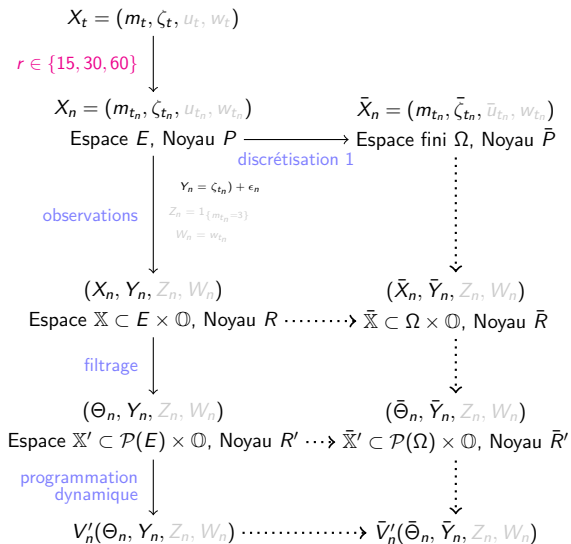
- ▶ discrétiser l'espace d'état du marqueur pour obtenir un filtre approché $\bar{\Theta}$
 - ▶ $\bar{\Theta}_n$ est à **support fini** : $\bar{\Theta}_n$
 - ▶ la relation de récurrence entre $\bar{\Theta}_n$ et $\bar{\Theta}_{n+1}$ est calculable
 - ▶ $\bar{\Theta}_n$ est simulable
- ▶ discrétiser les probabilités en chaque point du support du filtre

Alors l'intégrale contre R' se ramène à une somme pondérée calculable et si R' est **suffisamment régulier**, on obtiendra ainsi une bonne approximation de V

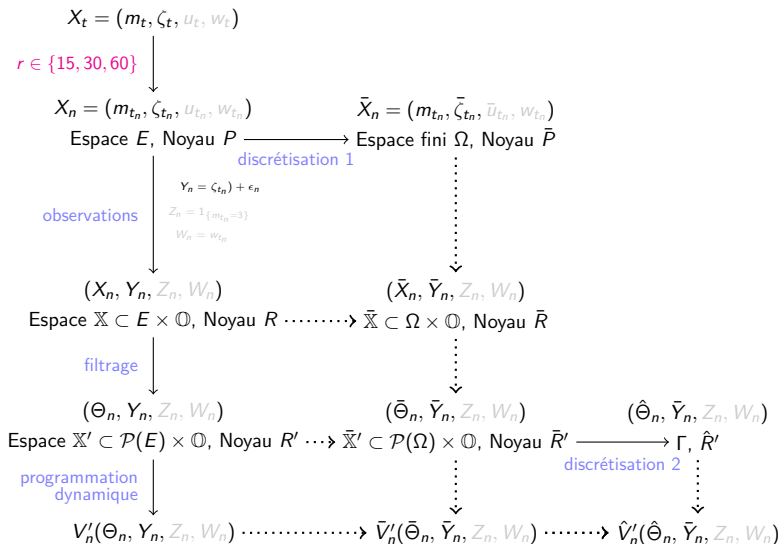
Récapitulatif de la démarche



Récapitulatif de la démarche



Récapitulatif de la démarche



Problème 3 : régularité des noyaux

A cause des sauts aux frontières lorsque le marqueur atteint ζ_0 ou D , le noyau P n'est pas régulier sur tout l'espace E , mais seulement localement Lipschitz sur une partition de E

Solution : prendre des précautions dans la création de la grille Ω pour qu'un point et sont projeté soient toujours dans une zone de l'espace où P est régulier : placer les points symétriquement par rapport à certaines valeurs seuils.

Approximation de la fonction valeur

Théorème

Sous des hypothèse de **régularité des paramètres** et de compatibilité entre **grille** et frontières

$$\begin{aligned} |V'_N(\bar{\theta}, \bar{y}) - \bar{V}'_N(\bar{\theta}, \bar{y})| &= 0 \\ |V'_n(\bar{\theta}, \bar{y}) - \bar{V}'_n(\bar{\theta}, \bar{y})| &\leq C_{V'_n} \max \mathcal{D}_j, \quad 0 \leq n < N, \\ |\hat{V}'_N(\hat{\theta}, \bar{y}) - \bar{V}'_N(\hat{\theta}, \bar{y})| &= 0, \\ |\hat{V}'_n(\hat{\theta}, \bar{y}) - \bar{V}'_n(\hat{\theta}, \bar{y})| &\leq C_{\bar{V}'_n} \max \bar{\mathcal{D}}_j, \end{aligned}$$

où $C_{V'_n}$ et $C_{\bar{V}'_n}$ ne dépendent que de n , N , δ et des paramètres de régularité et \mathcal{D}_j est le diamètre de la j -ème cellule de la première grille, $\bar{\mathcal{D}}_j$ celui de la j -ème cellule de la deuxième grille

Mise en oeuvre de la stratégie proposée

Initialisation

- ▶ Filtre initialisé à $\bar{\theta} \leftarrow \delta_{(0, \zeta_0)}$
- ▶ observation initiale y
- ▶ instant courant $t \leftarrow 0$

Mise en oeuvre de la stratégie proposée

Initialisation

- ▶ Filtre initialisé à $\bar{\theta} \leftarrow \delta_{(0, \zeta_0)}$
- ▶ observation initiale y
- ▶ instant courant $t \leftarrow 0$

Tant que l'horizon ou la mort n'est pas atteint

- ▶ projeter le filtre courant $\bar{\theta}$ sur la grille Γ pour obtenir $\hat{\theta}$
- ▶ choisir l'action $a^* = (\ell, r)$ donnée par la programmation dynamique pour $(\hat{\theta}, y)$
- ▶ donner le traitement ℓ jusqu'à l'instant $t + r$
- ▶ recueillir la nouvelle observation y à la date $t + r$
- ▶ mettre à jour le filtre avec l'opérateur discrétisé à partir de $\bar{\theta}$ et y
- ▶ $t \leftarrow t + r$

Plan de l'exposé

Modélisation du problème de contrôle

Construction d'une politique proche de l'optimalité

Résultats numériques

- Construction des grilles

- Performance de la politique

Conclusion et perspectives

Problème 4 : les décisions influencent la dynamique

Comme les décisions influencent significativement la dynamique, on ne peut pas explorer toutes les trajectoires possibles par simulation et le recours aux simulations est ainsi très limité pour construire les grilles

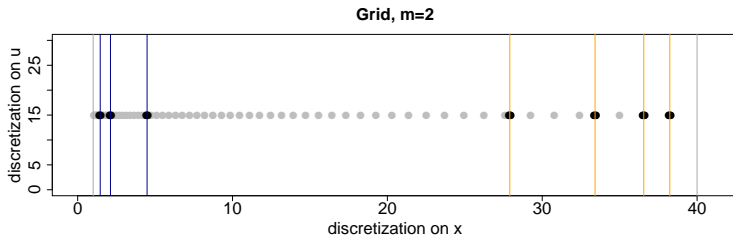
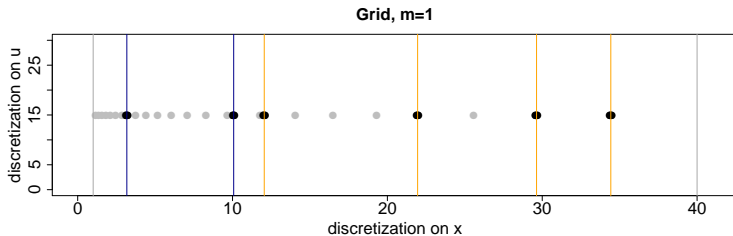
Problème 4 : les décisions influencent la dynamique

Comme les décisions influencent significativement la dynamique, on ne peut pas explorer toutes les trajectoires possibles par simulation et le recours aux simulations est ainsi très limité pour construire les grilles

Solution : rentabiliser la rigidité du processus

Dans un PDMP, la seule source d'aléa provient des sauts : tant qu'on n'a pas sauté pour la première fois, le processus reste en ζ_0 . Le processus ne va pas visiter toutes les zones de l'espace d'états, utiliser cette information a priori au mieux et enrichir les grilles avec des simulations sous la politique "optimale" proposée

Première discrétisation



Deuxième discrétisation

- ▶ Construire une première grille en dimension $|\Omega|$: pour chaque point ω dans Ω on prend la loi de probabilité qui charge ω avec proba 0.95, et le reste des points selon une loi de Dirichlet, estimation du noyau \hat{R} par simulations Monte Carlo
- ▶ Calculer la politique optimale pour cette grille
- ▶ Itérations
 - ▶ simuler $nsim$ trajectoires contrôlées avec la politique optimale de la grille courante
 - ▶ pour chaque trajectoire et chaque instant, calculer la **distance** entre le filtre estimé et sa projection sur la grille courante et rajouter le filtre estimé dans la prochaine grille si cette distance dépasse un certain seuil
 - ▶ Estimer \hat{R} par Monte Carlo sur la nouvelle grille et relancer la programmation dynamique pour mettre à jour la politique optimale

Plan de l'exposé

Modélisation du problème de contrôle

Construction d'une politique proche de l'optimalité

Résultats numériques

Construction des grilles

Performance de la politique

Conclusion et perspectives

Politiques concurrentes

- ▶ **Gold Standard** – inatteignable – décisions basées sur l'observation du mode à chaque instant de mesure
 - ▶ attribuer le bon traitement
 - ▶ visites à pas fixe
- ▶ **Filtrage** – une seule discrétisation nécessaire – utiliser le filtre pour estimer le mode
 - ▶ attribuer le traitement adapté au mode le plus probable
 - ▶ visites à pas fixe
- ▶ **Standard** – protocole hospitalier de référence – basé sur des seuils s_{rel} pour la rechute et s_{rem} pour la rémission.
 - ▶ Tant que $y \leq s_{rel}$, $\ell = \emptyset$, $r = 60$ jours
 - ▶ Si $y > s_{rel}$, $\ell = b$ (correspondant au type de rechute le plus fréquent 2) et $r = 15$ jours
 - ▶ Si à la visite suivante y a diminué, traitement b est maintenu avec visites tous les 15 jours jusqu'à ce que s_{rem} soit atteint
 - ▶ Sinon, $\ell = a$ avec des visites tous les 15 jours

Fonctions de coût

Choix de $c(x, a)$ de la forme

$$c(x, a) = \mathbb{E}[\tilde{c}(X_0, A_0, X_1) | X_0 = x, A_0 = a]$$

avec

$$\tilde{c}(x, a, x') = C_V + \kappa|\zeta' - \zeta_0|r + \beta r \mathbb{1}_{\{m=0, \ell \neq \emptyset\}}.$$

si $m \neq 3$, où

- ▶ $x = (m, \zeta)$, $a = (\ell, r)$, $x' = (m', \zeta')$
- ▶ C_V : coût fixe par visite – charge émotionnelle + frais médicaux
- ▶ $\beta > 0$: pénalité d'application d'un traitement sans maladie – effets secondaires
- ▶ $\kappa|\zeta' - \zeta_0|r$: approximation du temps passé dans la maladie et de la sévérité de la maladie
- ▶ M : coût de la mort (payé une seule fois)

Comparaison des performances

	Dates visites	coût réel (écart-type)	coût filtré (e-t)
méthode de l'exposé	Choix optimisé	136.23 (3.91)	134.74 (0.82)
	15 jours	213.92 (1.66)	215.16 (0.75)
	60 jours	145.37 (4.94)	140.58 (0.99)
Filtre	15 jours	209.96 (2.38)	210.2 (0.72)
	60 jours	169.39 (6.76)	170.56 (2.15)
Gold Standard	15 jours	161.51 (0.04)	
	60 jours	52.31 (0.82)	
Standard		438.92 (20.42)	

Plan de l'exposé

Modélisation du problème de contrôle

Construction d'une politique proche de l'optimalité

Résultats numériques

Conclusion et perspectives

Récapitulatif

Méthode proposée

- ▶ premier résultat de constructif de stratégie "optimale" pour un problème de contrôle impulsif de PDMP avec temps de saut non observés

Récapitulatif

Méthode proposée

- ▶ premier résultat de constructif de stratégie "optimale" pour un problème de contrôle impulsif de PDMP avec temps de saut non observés
- ▶ garanties théoriques sur l'approximation de la fonction valeur

Récapitulatif

Méthode proposée

- ▶ premier résultat de constructif de stratégie "optimale" pour un problème de contrôle impulsif de PDMP avec temps de saut non observés
- ▶ garanties théoriques sur l'approximation de la fonction valeur
- ▶ bonnes performances numériques

Récapitulatif

Méthode proposée

- ▶ premier résultat de constructif de stratégie "optimale" pour un problème de contrôle impulsif de PDMP avec temps de saut non observés
- ▶ garanties théoriques sur l'approximation de la fonction valeur
- ▶ bonnes performances numériques
- ▶ numériquement intensif, et très problème-dépendant

Récapitulatif

Méthode proposée

- ▶ premier résultat de constructif de stratégie "optimale" pour un problème de contrôle impulsif de PDMP avec temps de saut non observés
- ▶ garanties théoriques sur l'approximation de la fonction valeur
- ▶ bonnes performances numériques
- ▶ numériquement intensif, et très problème-dépendant
- ▶ généralisable dans une certaine mesure :
 - ▶ pas trop de modes / de variables, rester en faible dimension
 - ▶ pas trop de sauts possibles entre deux observations
 - ▶ dynamique déterministe entre les sauts générique (avec un minimum de régularité)
 - ▶ loi des temps de sauts générique (avec un minimum de régularité), ajout d'autres sauts aux frontières possible

Perspectives : ANR HSMM-INCA

Etape clé : [estimer/simuler/discrétiser le filtre](#)

- ▶ exploration en cours d'autres méthodes basées sur les simulations : Monte Carlo Tree Search, Filtre à particules
stage d'Aymar Thierry d'Argenlieu

Perspectives : ANR HSMM-INCA

Etape clé : [estimer/simuler/discrétiser le filtre](#)

- ▶ exploration en cours d'autres méthodes basées sur les simulations : Monte Carlo Tree Search, Filtre à particules
[stage d'Aymar Thierry d'Argenlieu](#)

Vers des modèles plus réalistes

- ▶ durée minimale de traitement un fois un traitement entamé
- ▶ adapter les paramètres au nombre de rechutes : résistance au traitement

Apprendre les paramètres et optimiser le traitement simultanément

- ▶ cadre de l'apprentissage par renforcement [thèse d'Orlane Le Quelenec avec Alice Cleynen et Régis Sabbadin](#)

Références

[BL 17] N. Bäuerle, D. Lange *Markov decision processes with applications to finances*

[BR 11] N. Bäuerle, U. Rieder *Optimal control of partially observed PDMPs*

[CdS 18] A. Cleynen, B. de Saporta *Change-point detection for PDMPs*

[CdS 22] A. Cleynen, B. de Saporta *Sequential decision making for a class of hidden Markov processes, application to medical treatment optimisation*

<https://github.com/acleynen/PDMP-control>

[Davis 93] M. Davis, *Markov models and optimization*

[dSDZ 14] B. de Saporta, F. Dufour, H. Zhang *Numerical methods for simulation and optimization of PDMPs: application to reliability*