

Définition et estimation d'une méta-cible pour la prédiction d'épisodes hypotensifs en unité de soin aigu par monitoring continu

Antoine Chambaz^{1*}, M. Cherifa^{3*}, R. Pirracchio^{2*}, M. Resche-Rigon^{3*}

¹ MAP5, Université de Paris

² UCSF & HEGP, Université Paris Descartes

³ ECSTRA, Hôpital Saint-Louis, APHP, Université de Paris

* ACTERRÉA

Journées Statistique et Santé, CNAM



A. Chambaz (MAP5)



ACTERRÉA



Mission Interdisciplinarité
défi MASTODONS

11/10/19

1 / 20

Introduction

Modélisation

Application

Le projet ACTERRÉA

- Chez un patient en réanimation, tous les organes sont susceptibles à tout moment de cesser de fonctionner normalement
- Les objectifs d'ACTERRÉA :
 1. prédire de tels dysfonctionnements
 2. suggérer une intervention thérapeutique
- Ce sur quoi ACTERRÉA peut prendre appui :
 - ▶ données MIMIC-III
 - ▶ flux de données patient en temps réel
 - ▶ méthodes d'apprentissage ciblé “machine learning” et statistique en temps réel
 - ▶ modèles déterministes d'organes
 - ▶ un réseau d'experts !

Les données MIMIC-III (1/3)

- MIMIC-III : “*Medical Information Mart for Intensive Care III*”, base de données relationnelle
- ~ 40,000 patients passés par une *réanimation* entre 2001 et 2012 au Beth Israel Deaconess Medical Center de Boston, Massachusett

Les données MIMIC-III (2/3)

- complexité croissante
- $\geq 58,000$ passages, monitoring et enregistrement de
 - ▶ signes vitaux (série temporelle, à l'échelle de la minute au mieux)
typiquement, rythmes cardiaque et respiratoire, quantité d'oxygène dans le sang (SpO2), pression artérielle systolique, moyenne et diastolique, glycémie
 - ▶ administration de médicaments
 - ▶ diagnostics
 - ▶ "waveforms" (courbes en quasi temps réel)
typiquement, électrocardiographie (ECG, de un à cinq courbes), photopléthysmogramme (PPG, variations de la circulation sanguine), pression artérielle (ABP), respiration
 - ▶ rapports biologiques effectués en laboratoire
 - ▶ notes prises par l'équipe de soignants
 - ▶ images diverses
 - Évidemment, données anonymisées

Les données MIMIC-III (3/3)

- Accès ouvert grâce à la collaboration du LCP (MIT), de Philips Health Care et du Beth Israel Deaconess Medical Center (*cf* références)
- Deux pré-requis :
 1. signer une convention stipulant les modalités d'utilisation des données
en particulier, les chercheurs doivent rendre leurs codes accessibles
 2. suivre une formation concernant la protection des informations dans le cadre d'études impliquant des femmes et des hommes
- Garantie de transparence, d'une meilleure coopération, de *reproductibilité*

Prédiction d'épisodes hypotensifs en unités de soin aigu

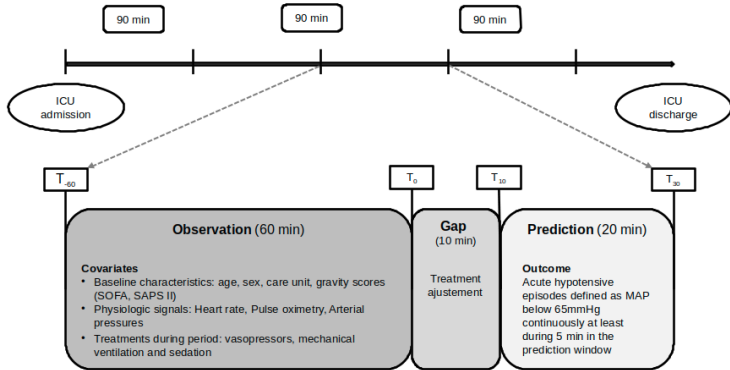
- Épisode hypotensif :
 - ▶ chute de la pression artérielle moyenne sous 65 mmHg pendant au moins 5 minutes
 - ▶ événement fréquent en unités de soin aigu
 - ▶ événement fortement associé à des événements délétères ultérieurs
 - ▶ événement qui nécessite une intervention rapide
- Cas d'étude pertinent pour la mise en œuvre de protocoles de prédiction
- Objectif :
 - ▶ élaborer une procédure pour anticiper les épisodes hypotensifs en unité de soin aigu... à partir...
 - ▶ des *données cliniques*, des *numerics*, et des *informations relatives aux traitements administrés*

Introduction

Modélisation

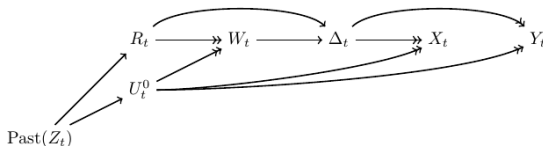
Application

Décomposition en blocs



- Objectif : anticiper les épisodes hypotensifs avec au moins 10 minutes d'avance

Modélisation (1/3)



Pour chaque période $1 \leq t \leq t_{\max}$, $Z_t \triangleq (U_t^0, R_t, W_t, \Delta_t, X_t, Y_t)$

- ▶ $U_t^0 \triangleq (W_t^0, X_t^0, Y_t^0)$: données complètes (possiblement contrefactuelles)
- ▶ R_t : indicateur de présence en unité de soin aigu
- ▶ W_t : résumé des données passées & données cliniques
- ▶ Δ_t : indicateur de disponibilité des *numerics*
- ▶ X_t : résumé des *numerics*
- ▶ Y_t : occurrence d'un épisode hypotensif

Modélisation (2/3)

- Π_0 la loi jointe de $(Z_1, \dots, Z_{t_{\max}})$

- ▶ hypothèse fondamentale :

il existe $c > 0$ tel que $\inf_{1 \leq t \leq t_{\max}} \Pi_0(\Delta_t = 1 | W_t, R_t = 1) \geq c > 0$

- Pour $\tau \in \{1, \dots, t_{\max}\}$ arbitrairement fixé :

- ▶ $O_\tau \triangleq (R_\tau, W_\tau, \Delta_\tau, X_\tau, Y_\tau) \subset Z_\tau$, observation générique à la période τ

- ▶ sous Π_0 , $O_\tau \sim P_{0\tau}$

- ▶ sous $\Pi_0(\text{do}(\Delta_\tau = 1))$, $O_\tau \sim P'_{0\tau}$

- ▶ objectif statistique attaché à τ :

prédire Y_τ à partir de (W_τ, X_τ) pour O_τ échantillonné sous $P'_{0\tau}$ conditionnellement à $R_\tau = 1$, grâce à O^1, \dots, O^n indépendantes de loi $P_{0\tau}$

- ▶ cible spécifique à τ :

$$f_{0\tau}^\ell \triangleq \arg \min_{f \in \mathcal{F}} E_{P'_{0\tau}} [\ell(f)(W_\tau, X_\tau, Y_\tau)]$$

ℓ , fonction de perte adaptée; \mathcal{F} , ensemble des fonctions de $\mathcal{X} \times \mathcal{W}$ dans $[0, 1]$

Modélisation (3/3)

- Une collection de t_{\max} cibles statistiques

$$f_{01}^{\ell}, \dots, f_{0\tau}^{\ell}, \dots, f_{0t_{\max}}^{\ell} \quad (1)$$

- ▶ multiplicité des cibles (et leur caractère assez artificiel)
 - ▶ si estimation individuelle, nécessité de prendre la dépendance en compte
 - ▶ $\Pi_0(\Delta_t = 0) \gg 0$ pour de nombreux $1 \leq t \leq t_{\max}$
 - ▶ grand intérêt à tirer profit des données à *travers* les périodes
- Une solution : fusionner les cibles multiples (1) en une unique **méta-cible**...

étant donné un *choix* de $\omega \in (\mathbb{R}_+)^{t_{\max}}$ tel $\sum_{t=1}^{t_{\max}} \omega_t = 1$,

$$f_0^{\ell\omega} \triangleq \arg \min_{f \in \mathcal{F}} \left\{ \sum_{t=1}^{t_{\max}} \omega_t \times P_{0t}(R_t = 1) \times E_{P'_{0t}} [\ell(f)(W_t, X_t, Y_t)] \right\}$$

Analyse statistique

- Interprétation : soit

- ▶ T indépendante de $Z \sim \Pi_0$ et telle que $Pr(T = t) = \omega_t$ pour tout $1 \leq t \leq t_{\max}$
- ▶ $\tilde{\omega}'_t(W, X) \triangleq P'_0(T = t | X_T = X, W_T = W)$ pour tout $1 \leq t \leq t_{\max}$

alors

$$f_0^{\ell\omega}(W, X) = \sum_{t=1}^{t_{\max}} \tilde{\omega}'_t(W, X) f_{0t}^{\ell}(W, X)$$

- Identification : soit, pour tout $1 \leq t \leq t_{\max}$,

- ▶ ϕ_{0t} la densité de la loi de W_t sachant $\Delta_t = 1$ % la loi de W_t sachant $R_t = 1$

alors

$$\arg \min_{f \in \mathcal{F}} \left\{ \sum_{t=1}^{t_{\max}} \omega_t \times P_{0t}(R_t = 1) \times E_{P_{0t}} \left[\phi_{0t}(W_t)^{-1} \ell(f)(W_t, X_t, Y_t) \middle| \Delta_t = 1 \right] \right\} = f_0^{\ell\omega}$$

Introduction

Modélisation

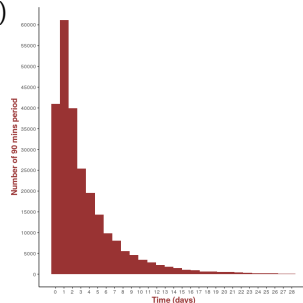
Application

Estimation “machine learning” de traits pertinents

- Pondération : pour tout $1 \leq t \leq t_{\max}$,
 - ▶ le facteur $\omega_t \times P_{0t}(R_t = 1) \times \phi_{0t}(W_t)^{-1}$ est un **poids** attaché à O_t
 - ▶ $\phi_{0t}(W_t)^{-1} = \Pi_0(\Delta_t = 1 | R_t = 1) / \Pi_0(\Delta_\tau = 1 | W_\tau, R_\tau = 1)$
- Machine learning à la super learning (agrégation) :
 - ▶ Ψ_1, \dots, Ψ_K algorithmes concurrents
 - ▶ pour tout $\alpha \in (\mathbb{R}_+)^K$, $\sum_{k=1}^K \alpha_k = 1$, $\Psi_\alpha \triangleq \sum_{k=1}^K \alpha_k \Psi_k$
 - ▶ principe de validation croisée pour déterminer un $\hat{\alpha}^L$ L -optimal et son $\hat{\Psi}_{\hat{\alpha}^L}^L$ correspondant
- Concrètement, ici :
 - ▶ package SuperLearner, algorithmes `SL.nnet`, `SL.glm`, `SL.glmnet`, `SL.rpart`, `SL.bayesglm`, `SL.xgboost`, `SL.gam` ($K = 7$)
 - ▶ validation croisée à $V = 5$ volets

Données

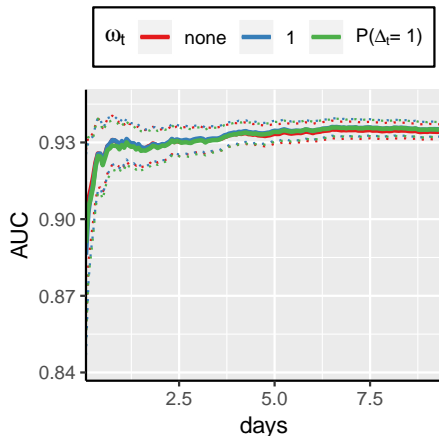
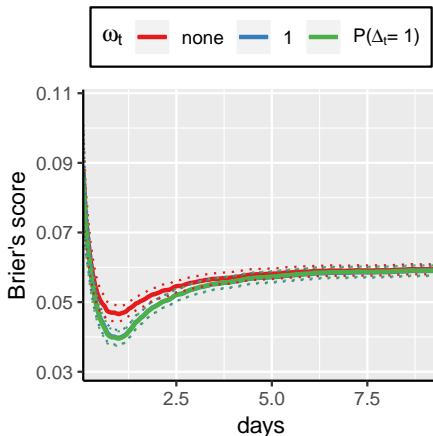
- Environ 4,750 patients et 200,000 périodes
- Conservation des périodes de 1 à 225 (environ 9 jours)
- W_t : c("gender", "age", "periode", "sapsii", "sofa", "care_unit", "ventilation", "sedation", "amine", "event_cumule", "event_5hours", "event_24hours", "event_one_half")
- X_t : c("hr", "spo2", "abp_mean", "abp_sys", "abp_dias")
- 70% pour l'apprentissage, 30% pour la validation



Mise en œuvre (1/2)

- Poids : pour tout $1 \leq t \leq t_{\max}$,
 - ▶ $\omega_t \propto 1$ ou $\omega_t \propto \hat{P}_{0t}(\Delta_t = 1)$
 - ▶ $\hat{P}_{0t}(\Delta_t = 1)$ et $\hat{P}_{0t}(R_t = 1)$ déterminés directement sur les données
 - ▶ $\widehat{\phi}_{0t}$ estimé par super learning directement sur les données, selon le mêmes modalités que $f_0^{\ell\omega}$
- Perte L : perte des moindres carrés
- Étude de simulation :
 - ▶ abandon des blocs pour lesquels $\Delta_t^{(i)} = 0$, sur les données d'apprentissage et de validation
 - ▶ fabrication d'une loi conditionnelle synthétique de Δ_t sachant W_t et $R_t = 1$
 - ▶ censure de $X_t^{(i)}$ selon les $\Delta_t^{(i)}$ générés, sur les données d'apprentissage uniquement

Mise en œuvre (2/2)



- ▶ score de Brier : moyenne des $(Y_t^i - \hat{f}_0^{\ell\omega}(X_t^{(i)}, W_t^{(i)}))$ sur les (i, t) disponibles et tels que $t \leq \tau$, pour tout $1 \leq \tau \leq 225$
- ▶ AUC : sous-échantillonnage d'une unique période par patient

Conclusion

- Un exercice intéressant :
 - ▶ délicat, instructif
 - ▶ questions épineuses (i) du choix des algorithmes et de leur paramétrage, (ii) du choix de la perte L , (iii) du protocole de validation par étude de simulation
- Perspectives :
 - ▶ enrichissement de la collection d'algorithmes concurrents
 - ▶ peaufinage de la perte L fondée sur l'AUC (attention à la pondération !)
 - ▶ introduction d'un hyper-paramètre $\beta \in [0, 1]$, substitution de poids^β à poids et sélection du β optimal par validation croisée externe
 - ▶ enrichissement de W_t et, surtout, de X_t

Références

- Cherifa M, Blet A, Chambaz A, Gayat É, Pirracchio R, and Resche-Rigon M. Prediction of an acute hypotensive episode during an ICU hospitalization with a super learner machine learning algorithm, *Anesthesia & Analgesia* (2019)
- Johnson AEW, Pollard TJ, Shen L, Lehman L, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, and Mark RG, *MIMIC-III, a freely accessible critical care database*, *Scientific Data* (2016)
- Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE, *PhysioBank, PhysioToolkit, and PhysioNet : Components of a New Research Resource for Complex Physiologic Signals*, *Circulation* 101(23) :e215-e220 (2000)
- Pirracchio R, Cohen J, Lee C, Malenica I, Cannesson M, Cohen M, Chambaz A, Resche-Rigon M, and Hubbard A. Big data and targeted machine learning in action to assist medical decision in the ICU : the past, the present and the future. *Anaesthesia, Critical Care & Pain Medicine* (2018)