

Modélisation statistique de la mortalité maternelle et néonatale pour l'aide à la planification et à la gestion des services de santé en Afrique Sub-Saharienne

Cheikh Ndour

Sous la direction de : S. Mercier & A. Diop
le co-encadrement de : S. Dossou Gbété & A. Dumont

LMAP (UPPA - France) – LERSTAD (UGB - Sénégal)

19 Mai 2014

1 Problématique et Objectifs

Sommaire

- 1 Problématique et Objectifs
- 2 Profil et Classifieur associé à un profil

Sommaire

- 1 Problématique et Objectifs
- 2 Profil et Classifieur associé à un profil
- 3 Extraction des profils fréquents

Sommaire

- 1 Problématique et Objectifs
- 2 Profil et Classifieur associé à un profil
- 3 Extraction des profils fréquents
- 4 Elagage des profils redondants

Sommaire

- 1 Problématique et Objectifs
- 2 Profil et Classifieur associé à un profil
- 3 Extraction des profils fréquents
- 4 Elagage des profils redondants
- 5 Sélection de profils et formulation d'un classifieur : cas d'un échantillon iid

Sommaire

- 1 Problématique et Objectifs
- 2 Profil et Classifieur associé à un profil
- 3 Extraction des profils fréquents
- 4 Elagage des profils redondants
- 5 Sélection de profils et formulation d'un classifieur : cas d'un échantillon iid
- 6 Sélection de profils et formulation d'un classifieur : cas de données groupées

Sommaire

- 1 Problématique et Objectifs
- 2 Profil et Classifieur associé à un profil
- 3 Extraction des profils fréquents
- 4 Elagage des profils redondants
- 5 Sélection de profils et formulation d'un classifieur : cas d'un échantillon iid
- 6 Sélection de profils et formulation d'un classifieur : cas de données groupées
- 7 Application aux données de la mortalité maternelle au Sénégal et au Mali

Problématique

- ① Dans ce travail, nous abordons un problème d'apprentissage supervisé avec une variable réponse binaire dont la modalité d'intérêt présente une très faible occurrence. Du point de vue de la statistique, on parle alors de **données déséquilibrées**.
- ② Nous considérons deux protocoles d'échantillonnage :
 - (a) les données sont **indépendantes et identiquement distribuées**
 - (b) les données présentent une structure **d'hétérogénéité**.
 - ① On sélectionne de façon aléatoire un ensemble de groupe à partir d'une population de groupes donnée.
 - ② Dans chaque groupe, on tire aléatoirement un ensemble d'observations.

Etat de l'art

Dans la littérature, il est bien documenté que les méthodes de classement standards telles que

- la régression logistique
- les arbres binaires de classement
- l'analyse discriminante
- etc..

ne sont pas appropriées pour l'analyse de données déséquilibrées.

⇒ Puisqu'elles produisent des **classifieurs faibles** (classifieurs avec de faibles performances).

Etat de l'art

① Les méthodes préconisant un ré-échantillonnage des données :

- ① Sur-échantillonner les données observées par rapport à la classe rare
- ② Sous-échantillonner les données observées par rapport à la classe majoritaire

Limites : Forte instabilité par rapport à toute modification de l'ensemble d'apprentissage



McCarthy K, Zabar B, Weiss G (2005), *Does cost-sensitive learning beat sampling for classifying rare classes?*, *Proceedings of the 1st international workshop on utility-based data mining*, ACM Press, New York, pp 69 - 77

② Les méthodes intervenant au niveau du processus d'apprentissage :

- ① Tenir compte des coûts de mauvais classement
- ② Combiner les classifieurs faibles
- ③ Associer le ré-échantillonnage avec les techniques d'agrégation

Limites : Classifieur sous forme d'une boîte noire



Breiman, L (1996), *Bagging Predictors*, *Mach. Learn.*, Kluwer Academic Publishers, Hingham, MA, USA, pp 123-140

Etat de l'art

Pour surmonter la difficulté qui résulte de la faible occurrence de la modalité d'intérêt, nous proposons une approche basée sur les **règles d'association**. Cette approche est plus connue sous le nom de **classification associative**. Parmi les approches pionnières on peut citer :

- 1 **Algorithme CBA** (Classification Based on Association)
- 2 **Algorithme CMAR** (Classification Based on Multiples Association Rules)

Principe : Extraire d'abord l'ensemble des règles d'association "valides", puis de supprimer les règles redondantes et les moins intéressantes de cet ensemble.

Limites : Ces algorithmes produisent un vaste ensemble de profils qui est très difficile à manipuler pour l'aide à la prise de décision.



Bing Liu, Wynne Hsu, and Yiming Ma. (1998), *Integrating classification and association rule mining*. In *Proceedings KDD'98*, pp 80–86. AAAI Press, 1998

Objectifs

Notre objectif est donc de proposer une méthode d'apprentissage statistique qui

- 1 Est capable de produire un classifieur efficace et qui prend en compte les avantages des algorithmes CBA et CMAR.
- 2 Permet d'améliorer les stratégies de sélection d'un ensemble optimal de profils développées dans les algorithmes CBA et CMAR en utilisant une approche statistique.
- 3 Permet de produire un classifieur qui est moins dépendant de l'ensemble d'apprentissage, qui est facile à manipuler et qui peut être visualisé pour servir d'aide à la prise de décision.

Principe de la méthode

- 1 Explorer un ensemble d'apprentissage dans le but de générer tous les profils fréquents et significativement corrélés avec la variable réponse.
- 2 Elaguer les profils redondants en utilisant des mesures de performance calculées à partir des données d'apprentissage.
- 3 Extraire un petit ensemble de profils pertinents $\{U_k\}_{k=1:m}$ à partir de la collection précédente en utilisant un ensemble de validation.

Etablir la règle de classement comme une fonction des profils pertinents $\{U_k\}_{k=1:m}$.

Profil

Soit $X = (X_j)_{j=1:p}$ une suite finie de p variables aléatoires où chaque X_j est une variable non numérique à q_j modalités.

Définition : Profil

On appelle profil toute intersection d' événements élémentaires (item)

$$U = \bigcap_{j \in J} \{X_j = m_{h(j)}^{X_j}\} \quad (1)$$

où $J \subseteq 1 : p$

Pour simplifier l'expression, on note

$$U = \left(m_{h(j)}^{X_j} \right)_{j \in J}$$

- Un profil peut être vu comme la réalisation conjointe de $|J|$ variables $(X_j)_{j \in J}$.
- La taille (ou longueur) du profil détermine sa complexité.
- Plus le nombre de variables conjointement réalisées augmente, plus la taille du profil est grande.

Profils emboîtés & Profils disjoints

Définition

Soient $U = \left(m_{h(l)}^{X_l} \right)_{l \in L}$ et $U' = \left(m_{h(j)}^{X_j} \right)_{j \in J}$ deux profils. On dit que U' est emboîté dans U et on note $U' \prec U$ si les conditions suivantes sont vérifiées.

a) $L \subset J$

b) $\forall l \in L, \forall h(l) \in \{1 : q_l\} \quad \exists ! j \in J, \exists ! k(j) \in \{1 : q_j\}$ tels que $m_{h(l)}^{X_l} = m_{k(j)}^{X_j}$

Définition

Soient $U = \left(m_{h(l)}^{X_l} \right)_{l \in L}$ et $U' = \left(m_{h(j)}^{X_j} \right)_{j \in J}$ deux profils. On dit que les profils U et U' sont disjoints lorsque $L \cap J = \emptyset$.

Classifieur associé à un Profil

On peut associer à chaque profil $U = \left(m_{h(U)}^{X_j} \right)_{j \in J}$ un classifieur binaire $\phi(U, X)$ définie par :

$$\phi(U, X) = \begin{cases} 1 & \text{si } \prod_{j \in J} \mathbb{1} \left(m_{h(U)}^{X_j} \right) (X_j) = 1 \\ 0 & \text{sinon} \end{cases} \quad (2)$$

En présence de données déséquilibrées $\phi(U, X)$ est en général un classifieur faible.

Notre objectif est de combiner un nombre réduit de classifieurs de ce type pour construire un classifieur performant $\Phi(X)$ défini par

$$\Phi(X) = \mathbb{1} \left(\sum_{k=1}^m \phi(U_k, X) > 0 \right) \quad (3)$$

Indicateurs de performance de base

La fréquence d'un profil $U = \left(m_{h(j)}^{X_j} \right)_{j \in J}$ est mesurée à partir de sa couverture, de son support et de sa confiance.

Définition

- ① On appelle couverture du profil U la probabilité $\Pr \{ \phi(U, X) = 1 \}$.
- ② On appelle support du profil U la probabilité $\Pr \{ \phi(U, X) = 1, Y = 1 \}$.
- ③ On appelle confiance du profil U la probabilité conditionnelle $\Pr \{ Y = 1 \mid \phi(U, X) = 1 \}$.

Du point de vue de la statistique, la confiance représente la valeur prédictive positive du classifieur $\phi(U, X)$.

Indicateurs de performance de base

Puisque on associe à chaque profil U un classifieur $\phi(U, X)$, nous proposons d'utiliser des indicateurs de performance classiques tels que :

Indicateurs de performance

- ① L'erreur de classement : $Err(U, Y) = \Pr\{\phi(U, X) \neq Y\}$
- ② La sensibilité : $TVP(U, Y) = \frac{\Pr(\phi(U, X) = 1, Y = 1)}{\Pr(Y = 1)}$
- ③ La spécificité : $TVN(U, Y) = \frac{\Pr(\phi(U, X) = 0, Y = 0)}{\Pr(Y = 0)}$
- ④ La valeur prédictive positive : $VPP(U, Y) = \frac{\Pr(\phi(X, U) = 1, Y = 1)}{\Pr(\phi(X, U) = 1)}$

pour identifier les profils les plus pertinents pour construire un classifieur efficace.

Indicateurs de performance de base

Les indicateurs de performance peuvent être exprimés en fonction de la valeur prédictive positive, de la couverture et de la loi a priori de la variable réponse.

$$Err(U, Y) = \Pr(Y = 1) + \Pr(\phi(U, X) = 1)(1 - 2VPP(U, Y))$$

$$TVP(U, Y) = VPP(U, Y) \frac{\Pr(\phi(U, X) = 1)}{\Pr(Y = 1)}$$

$$TVN(U, Y) = 1 - (1 - VPP(U, Y)) \frac{\Pr(\phi(U, X) = 1)}{1 - \Pr(Y = 1)}$$

Par conséquent, en contrôlant la valeur prédictive positive et la couverture on contrôle l'évolution des indicateurs de performance.

Profils fréquents

Le classifieur associé à un profil est une implication de la forme

$$[\phi(X, U) = 1] \rightarrow [Y = 1] \quad (4)$$

Ce qui est une règle d'association particulière communément appelée règle d'association binaire.

Elle est fréquente dès lors qu'on exige que

$$\Pr(\phi(X, U) = 1, Y = 1) > s_0 \quad \text{et} \quad \Pr(Y = 1 | \phi(X, U) = 1) > c_0$$

De telles règles d'association peuvent être produites en utilisant les algorithmes d'extraction des règles d'association.

Profils fréquents

Depuis plus de deux décennies (1994), plusieurs algorithmes d'extraction des profils fréquents ont vu le jour. Parmi eux on peut citer **Algorithme APRIORI** qui est programmé dans l'environnement de programmation **R** sous le packages **arules**.

L'algorithme **APRIORI** est l'algorithme d'exploration des règles d'association le plus connu, utilisant s_0 , c_0 et l_0 pour l'extraction des règles les plus fréquentes, mais aussi pour contrôler systématiquement la croissance exponentielle des règles candidates.



Agrawal, R and Srikant, R (1994), *Fast Algorithms for Mining Association Rules in Large Databases, Proceedings of the 20th International Conference on Very Large Data Bases, VLDB '94, Morgan Kaufmann Publishers Inc., pp 487 - 499*

Profils redondants

Définition : Profils redondants

Soient $U_1 = \left(m_{h(l)}^{X_l} \right)_{l \in L}$ et $U_2 = \left(m_{h(l)}^{X_j} \right)_{j \in J}$ deux profils emboîtés. On dit que le profil U_t , $t \in \{1, 2\}$, est redondant par rapport au profil $U_{t'}$, $t' \in \{1, 2\}$ et $t \neq t'$, si les indicateurs de performance de $\phi(U_{t'}, X)$ sont meilleurs (plus forts) que les indicateurs de performance de $\phi(U_t, X)$.

La stratégie d'élagage des profils redondants est basée sur trois résultats principaux que nous présentons sous forme de propositions.

Elagage des profils redondants

Proposition 1

Soient $U = \left(m_{h(l)}^{X_l} \right)_{l \in L}$ et $U' = \left(m_{h(l')}^{X_{l'}} \right)_{l' \in J}$ deux profils tels que $U' \prec U$.

Si $\Pr \{ \phi(U, X) = 1 \} = \Pr \{ \phi(U', X) = 1 \}$ alors

- ① $\Pr \{ \phi(U, X) = 1, Y = 1 \} = \Pr \{ \phi(U', X) = 1, Y = 1 \}$
- ② $\Pr \{ \phi(U, X) = 0, Y = 0 \} = \Pr \{ \phi(U', X) = 0, Y = 0 \}$

En conclusion, le classifieur $\phi(U, X)$ associé au profil U a les mêmes performances que le classifieur $\phi(U', X)$ associé au profil U' .

Elagage des profils redondants

Proposition 2

Soient $U = \left(m_{h(l)}^{X_l} \right)_{l \in L}$ et $U' = \left(m_{h(U)}^{X_j} \right)_{j \in J}$ deux profils tels que $U' \prec U$.

Si $\Pr \{ \phi(U, X) = 1, Y = 1 \} = \Pr \{ \phi(U', X) = 1, Y = 1 \}$ alors

- 1 $VPP(U, Y) \leq VPP(U', Y)$
- 2 $TVN(U, Y) \leq TVN(U', Y)$
- 3 $Err(U, Y) \geq Err(U', Y)$

Il résulte de cette proposition que le classifieur $\phi(U', X)$ associé au profil U' est plus performant que le classifieur $\phi(U, X)$ associé au profil U .

Elagage des profils redondants

Proposition 3

Soient $U = \left(m_{h(l)}^{X_l} \right)_{l \in L}$ et $U' = \left(m_{h(U)}^{X_j} \right)_{j \in J}$ deux profils tels que $U' \prec U$.

Si $\Pr \{ \phi(U, X) = 0, Y = 0 \} = \Pr \{ \phi(U', X) = 0, Y = 0 \}$ alors

- ① $VPP(U', Y) \leq VPP(U, Y)$
- ② $TVP(U', Y) \leq TVP(U, Y)$
- ③ $Err(U', Y) \geq Err(U, Y)$

En conclusion, le classifieur $\phi(U, X)$ associé au profil U est plus performant que le classifieur $\phi(U', X)$ associé au profil U' .

Test d'hypothèse

Pour supprimer les profils redondants on effectue un test d'hypothèse sur :

- ① **l'égalité des couvertures**
- ② **l'égalité des supports**
- ③ **l'égalité des spécificités**

Ces trois tests d'hypothèse sont tous basés sur le même principe.

Test d'hypothèse sur les couvertures

Soient $U = (m_{h(l)}^{X_l})_{l \in L}$ et $U' = (m_{h(l')}^{X_{l'}})_{l' \in J}$ deux profils tels que $U' \prec U$.

On note $\mathcal{D} = \{(y_i, x_i); i = 1 : n\}$ dont (y_i, x_i) est une réalisation de la variable aléatoire (Y_i, X_i) . Dans le cas où la suite (Y_i, X_i) , $i = 1 : n$ est iid, les variables aléatoires

$$Z(X_i) = \phi(U, X_i) - \phi(U', X_i), \quad i = 1 : n \quad (5)$$

constituent une suite indépendante de variables aléatoires de Bernoulli de même paramètre

$$\theta = \Pr(\phi(U, X) = 1) - \Pr(\phi(U', X) = 1) \quad (6)$$

L'hypothèse nulle d'égalité des couvertures peut se traduire par l'énoncé $\theta = 0$.

Test d'hypothèse sur les couvertures

On considère alors la statistique de test

$$\varphi(\mathcal{D}) = \left[\sum_{i=1}^n Z(X_i) > 0 \right] + (1 - \gamma) \left[\sum_{i=1}^n Z(X_i) = 0 \right] \quad (7)$$

avec $\gamma \in]0, 1[$ et la règle de décision suivante :

- si $\varphi(\mathcal{D}) = 1$, rejeter l'égalité des couvertures avec la probabilité 1
- sinon rejeter l'égalité des couvertures au hasard avec une probabilité γ

On obtient ainsi un test stochastique d'hypothèse de niveau γ et de puissance

$$\pi(n) = 1 - (1 - \gamma)(1 - \theta)^n \quad (8)$$

Réduction de l'ensemble des profils fréquents et non-redondants

Nous avons remarqué plus tôt que :

- 1 la sensibilité et la spécificité sont des fonctions croissantes de la valeur prédictive positive
- 2 l'erreur de classement est une fonction décroissante de la valeur prédictive positive

Pour cette raison, nous avons basé la sélection des profils sur un test de comparaison des valeurs prédictives positives lorsque deux profils sont emboîtés.

Réduction de l'ensemble des profils : cas des données iid

Considérons le système complet d'événements ci-dessous :

$$\begin{aligned}
 E_1 &= \{Y = 1, \phi(U, X) = 1, \phi(U', X) = 1\} & E_2 &= \{Y = 1, \phi(U, X) = 1, \phi(U', X) = 0\} \\
 E_3 &= \{Y = 1, \phi(U, X) = 0, \phi(U', X) = 0\} & E_4 &= \{Y = 0, \phi(U, X) = 1, \phi(U', X) = 1\} \\
 E_5 &= \{Y = 0, \phi(U, X) = 1, \phi(U', X) = 0\} & E_6 &= \{Y = 0, \phi(U, X) = 0, \phi(U', X) = 0\}
 \end{aligned}$$

dont les probabilités respectives sont :

$$\theta = \begin{pmatrix} p_1 \\ \vdots \\ p_6 \end{pmatrix}$$

La variable aléatoire $Z = (I_{E_1}, \dots, I_{E_6})$ suit une distribution de Bernoulli généralisée de paramètre θ et de matrice de variance covariance

$$\Lambda(\theta) = \text{diag}(\theta) - \theta^T \theta$$

Réduction de l'ensemble des profils : cas des données iid

On a

$$VPP(U, Y) = \frac{p_1 + p_2}{p_1 + p_2 + p_4 + p_5} \qquad VPP(U', Y) = \frac{p_1}{p_1 + p_4}$$

On note par

$$g(\theta) = \log \left(\frac{VPP(U, Y)}{VPP(U', Y)} \right) = \log \left(\frac{(p_1 + p_4)(p_1 + p_2)}{p_1(p_1 + p_2 + p_4 + p_5)} \right)$$

On a

$$\nabla g(\theta) = \begin{pmatrix} \frac{1}{p_1 + p_4} + \frac{1}{p_1 + p_2} - \frac{1}{p_1} - \frac{1}{p_1 + p_2 + p_4 + p_5} \\ \frac{1}{p_1 + p_2} - \frac{1}{p_1 + p_2 + p_4 + p_5} \\ 0 \\ \frac{1}{p_1 + p_4} - \frac{1}{p_1 + p_2 + p_4 + p_5} \\ -\frac{1}{p_1 + p_2 + p_4 + p_5} \\ 0 \end{pmatrix}$$

Réduction de l'ensemble des profils : cas des données iid

Soit $(Z_i)_{i=1:n}$ une suite de variables aléatoires indépendantes et identiquement distribuées suivant une distribution de Bernoulli généralisée. Si on considère

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n Z_i$$

l'estimateur empirique de θ .

Il résulte de la Delta-Method que

$$\frac{\sqrt{n} \left(g(\hat{\theta}_n) - g(\theta) \right)}{\sqrt{^T \nabla g(\hat{\theta}_n) \Lambda(\hat{\theta}_n) \nabla g(\hat{\theta}_n)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

Réduction de l'ensemble des profils : cas des données iid

Sous l'hypothèse nulle $H_0 : VPP(U, Y) = VPP(U', Y)$, on a $g(\theta) = 0$. Par conséquent, on peut établir une stratégie de sélection des profils pertinents à partir du test suivant.

- 1 Supprimer le profil U' si

$$g(\hat{\theta}_n) > -q_{1-\alpha/2} \sqrt{\frac{\tau \nabla g(\hat{\theta}_n) \Lambda(\hat{\theta}_n) \nabla g(\hat{\theta}_n)}{n}} \quad (9)$$

- 2 Supprimer le profil U si

$$g(\hat{\theta}_n) \leq -q_{1-\alpha/2} \sqrt{\frac{\tau \nabla g(\hat{\theta}_n) \Lambda(\hat{\theta}_n) \nabla g(\hat{\theta}_n)}{n}} \quad (10)$$

où $q_{1-\alpha/2}$ est le quantile d'ordre $1 - \alpha/2$ de la loi normale centrée réduite.

Comparaison du classifieur avec des méthodes alternatives

Population d'étude : Les données contiennent des sujets :

- âgés de plus de 16 ans
- ayant un revenu brut ajusté supérieur à 1
- un volume horaire de travail positif

Au total, elles contiennent 45222 sujets hormis les données manquantes.

Variables : Les données contiennent 14 covariables dont 5 sont continues et 8 sont nominales dont une variable réponse binaire indexant la classe du revenu annuel d'un sujet (24.78% d'individus ayant un revenu annuel supérieur à \$50K).



Ron Kohavi (1996), *Scaling Up the Accuracy of Naive-Bayes Classifiers : a Decision-Tree Hybrid*, *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, pp 202-207



Fayyad, Usama M. and Irani, Keki B (1993), *Multi-Interval Discretization of Continuous-Valued Attributes for Classification Learning*, *IJCAI*, pp 1022-1029

Comparaison du classifieur avec des méthodes alternatives

Pour évaluer la procédure d'apprentissage des règles d'association binaire, nous avons effectué plusieurs expériences en sous-échantillonnant les données. Pour obtenir un ensemble de données déséquilibrées, on procède de la manière suivante :

- Sélectionner tous les individus de la classe prévalente. Soit n le nombre d'individus de la classe majoritaire.
- Se donner un nombre réel $\alpha \in]0, 1[$. On sélectionne $n' = n\alpha/(1 - \alpha)$ d'individus de la classe rare.

On obtient ainsi, un échantillon de $n + n'$ observations avec une proportion α de la classe rare.

Comparaison du classifieur avec des méthodes alternatives

Distributions "0" - "1"	Sensibilité					
	ARM	CTREE	Naive Bayes	SMOTE	Boosting	Random Forests
0.993 - 0.007	0,700	0,593	0,814	0,547	0,756	0,698
0.985 - 0.015	0,671	0,793	0,799	0,696	0,766	0,799
0.970 - 0.030	0,644	0,812	0,842	0,716	0,823	0,791
0.930 - 0.070	0,754	0,797	0,850	0,783	0,842	0,804
0.850 - 0.150	0,791	0,813	0,832	0,805	0,836	0,780

Distributions "0" - "1"	Spécificité					
	ARM	CTREE	Naive Bayes	SMOTE	Boosting	Random Forests
0.993 - 0.007	0,779	0,878	0,719	0,770	0,793	0,851
0.985 - 0.015	0,821	0,777	0,761	0,821	0,814	0,794
0.970 - 0.030	0,807	0,766	0,750	0,755	0,801	0,780
0.930 - 0.070	0,799	0,805	0,760	0,772	0,804	0,794
0.850 - 0.150	0,774	0,818	0,785	0,792	0,828	0,833

Distributions "0" - "1"	Erreur de classement					
	ARM	CTREE	Naive Bayes	SMOTE	Boosting	Random Forests
0.993 - 0.007	0,222	0,124	0,279	0,231	0,208	0,150
0.985 - 0.015	0,181	0,222	0,238	0,181	0,187	0,205
0.970 - 0.030	0,198	0,232	0,247	0,247	0,199	0,220
0.930 - 0.070	0,204	0,196	0,233	0,227	0,193	0,204
0.850 - 0.150	0,223	0,183	0,207	0,207	0,171	0,176

Comparaison du classifieur avec des méthodes alternatives

Distributions "0" - "1"	Aire en dessous de la courbe ROC (AUC)					
	ARM	CTREE	Naive Bayes	SMOTE	Boosting	Random Forests
0.993 - 0.007	0,740	0,736	0,766	0,659	0,774	0,774
0.985 - 0.015	0,746	0,785	0,780	0,758	0,790	0,796
0.970 - 0.030	0,726	0,789	0,796	0,736	0,812	0,786
0.930 - 0.070	0,776	0,801	0,805	0,778	0,823	0,799
0.850 - 0.150	0,782	0,815	0,808	0,798	0,832	0,806

Distributions "0" - "1"	Score de Pierce					
	ARM	CTREE	Naive Bayes	SMOTE	Boosting	Random Forests
0.993 - 0.007	0,479	0,471	0,533	0,317	0,549	0,549
0.985 - 0.015	0,492	0,570	0,560	0,517	0,580	0,593
0.970 - 0.030	0,451	0,578	0,592	0,471	0,624	0,571
0.930 - 0.070	0,553	0,602	0,610	0,555	0,646	0,598
0.850 - 0.150	0,565	0,631	0,617	0,597	0,664	0,613

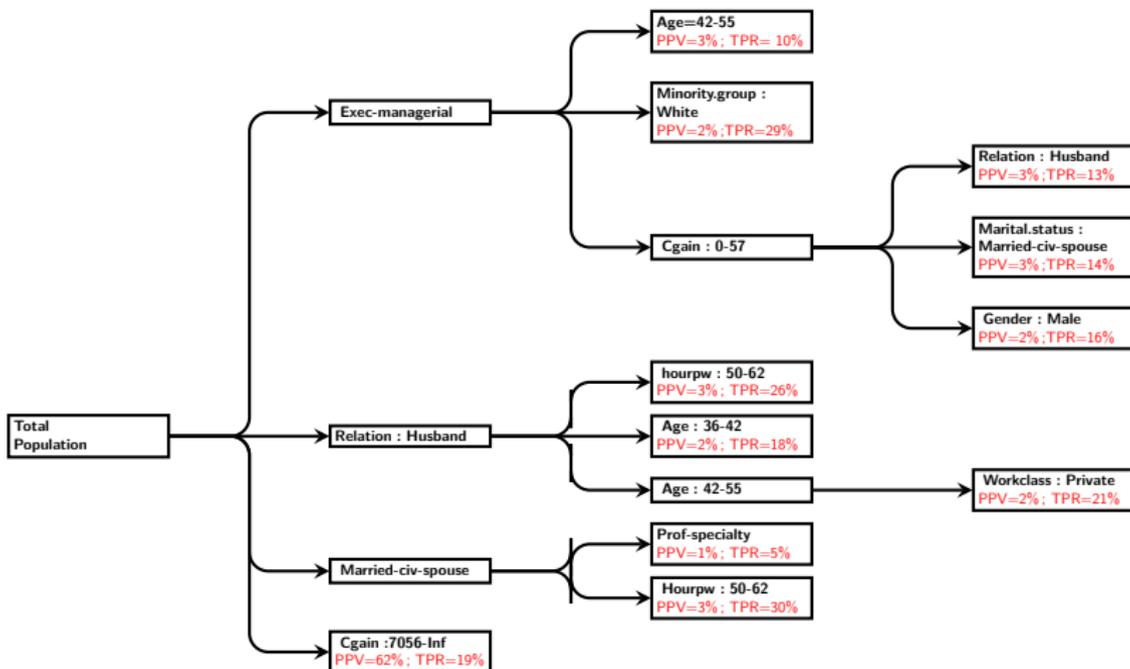
Variabilité des indicateurs de performance

Echantillon : $\alpha = 0.7\%$

Paramètres d'apprentissage : $s_0 = 0.1\%$; $c_0 = 2.8\%$; $l_0 = 4$

	nb-freq	nb-optim	sensibilité	spécificité	erreur	auc
1	80	12	0.71	0.77	0.24	0.74
2	101	14	0.78	0.73	0.25	0.76
3	80	11	0.74	0.78	0.23	0.76
4	47	13	0.70	0.82	0.20	0.76
5	93	7	0.57	0.83	0.23	0.70
6	46	14	0.73	0.83	0.20	0.78
7	113	14	0.72	0.73	0.27	0.73
8	71	12	0.55	0.83	0.23	0.69
9	94	13	0.74	0.79	0.22	0.77
10	51	8	0.54	0.85	0.22	0.70
11	102	4	0.46	0.91	0.19	0.69
12	53	6	0.49	0.90	0.20	0.70
moy	65	11	0.65 (± 0.007)	0.82 (± 0.006)	0.22 (± 0.006)	0.73 (± 0.007)

Présentation du classifieur



Réduction de l'ensemble des profils : cas des données groupées

Dans la suite, il est question d'adapter le test de comparaison des valeurs prédictives positives pour deux profils emboîtés sous l'hypothèse d'hétérogénéité des données.

Nous supposons que les données soumises au traitement statistique sont partitionnées en H sous-ensembles correspondant à des sous-populations sélectionnées de façon aléatoire.

Réduction de l'ensemble des profils : cas données groupées

Modèle hiérarchique

- On considère un premier niveau d'échantillonnage sur l'ensemble $Prob(Y, X)$ des lois de probabilité sur $\Omega = \{0, 1\} \times \prod_{j=1}^p Dom(X_j)$.
- On échantillonne dans $Prob(Y, X)$ suivant une loi de probabilité μ et il en résulte H réalisations indépendantes $[Y, X]_h \in Prob(Y, X)$ de la loi de probabilité μ .

$$\mathcal{D} = \bigcup_{h=1}^H \mathcal{D}_h$$

où

- $\mathcal{D}_h = \{ (y_{i[h]}, , x_{i[h]}) , i[h] = 1 : n_h \}$
- les $(y_{i[h]}, , x_{i[h]})$ sont des réalisations indépendantes de la loi $[Y, X]_h$

Réduction de l'ensemble des profils : cas données groupées

Pour un profil U fixé, on suppose que :

H1 $[Y|\theta_h^U, [Y, X]_h] = \text{Bernoulli}(\theta_h^U)$ où $\theta_h^U = \Pr(Y = 1|U(X), [Y, X]_h)$

H2 θ_h^U suit une loi Bêta de paramètres α_U et β_U .

Les variables θ_h^U ; $h = 1, \dots, H$ sont indépendantes et identiquement distribuées.

Réduction de l'ensemble des profils : cas données groupées

Connaissant la loi a priori de θ_h^U , on peut écrire

$$\mathbb{E}(\theta_h^U | \alpha_U, \beta_U) = \frac{\alpha_U}{\alpha_U + \beta_U}$$

et

$$\text{Var}(\theta_h^U | \alpha_U, \beta_U) = \frac{\alpha_U}{\alpha_U + \beta_U} \left(1 - \frac{\alpha_U}{\alpha_U + \beta_U}\right) \frac{1}{(\alpha_U + \beta_U + 1)}$$

L'application $(\alpha_U, \beta_U) \longrightarrow \left(\begin{array}{l} \pi_U = \frac{\alpha_U}{\alpha_U + \beta_U} \\ \gamma_U = \frac{1}{\alpha_U + \beta_U + 1} \end{array} \right)$ étant injective, on peut envisager de reparamétriser la famille des lois Bêta par la moyenne π_U et le paramètre de dispersion γ_U .

Modèle hiérarchique

Dans la suite, nous considérons le modèle suivant :

$$\left\{ \begin{array}{l} Y | \theta_h^U, [Y, X]_h \sim \prod_{k=1}^m \left\{ (\theta_h^U)^{\mathbb{1}_{\{Y=1\}}(Y)} \delta_{\{1, [Y, X]_h\}}(U(X), [Y, X]_k) \right. \\ \left. (1 - \theta_h^U)^{(1 - \mathbb{1}_{\{Y=1\}}(Y)) \delta_{\{1, [Y, X]_h\}}(U(X), [Y, X]_k)} \right\} \\ \theta_h^U | \pi_U, \gamma_U \sim \text{Beta}(\pi_U, \gamma_U) \end{array} \right.$$

Idee : déterminer la loi a posteriori de θ_h^U pour pouvoir évaluer la $VPP(U, Y, h)$.

$$VPP(U, Y, h) = \int_0^1 [Y = 1 | U(X), [Y, X]_h] [\theta_h^U | [Y, X]_h, \pi_U, \gamma_U] d\theta_h^U \quad (11)$$

Puisque la loi a priori des paramètres π_U et γ_U n'a pas été spécifiée, nous proposons une approche bayésienne empirique pour l'estimation a posteriori de $VPP(U, Y, h)$.

Principe de l'estimation bayésienne empirique

En supposant que les paramètres π_U et γ_U sont inconnus, nous les estimerons à partir de la distribution marginale de toutes les données.

$$[(S_{hU})_{h=1:m} | \pi_U, \gamma_U] = \prod_{h=1}^m \left\{ \int_0^1 [S_{hU} | \theta_h^U] [\theta_h^U | \pi_U, \gamma_U] d\theta_h^U \right\} \quad (12)$$

où

$$\begin{aligned} [S_{hU} | \theta_h^U] &= \text{Binomiale}(n_{hU}, \theta_h^U) \\ [\theta_h^U | \pi_U, \gamma_U] &= \text{Beta}\left(\pi_U \left(\frac{1}{\gamma_U} - 1\right), (1 - \pi_U) \left(\frac{1}{\gamma_U} - 1\right)\right) \end{aligned}$$

Ces estimations sont habituellement obtenues par la méthode du maximum de vraisemblance (MLE) ou la méthode des moments (MOM) à partir de la distribution marginale des données.

On obtient la distribution a posteriori estimée : $[\theta_h^U | S_{hU}, \hat{\pi}_U, \hat{\gamma}_U]$

Réduction de l'ensemble des profils : cas données groupées

Indicateur de performance de base de la procédure

La valeur prédictive positive a posteriori du profil U pour un cluster h (VPP)

$$VPP(U, Y, h) = \frac{S_{hU} + \hat{\pi}_U \left(\frac{1}{\hat{\gamma}_U} - 1 \right)}{n_{hU} + \left(\frac{1}{\hat{\gamma}_U} - 1 \right)} \quad (13)$$

On peut écrire $VPP(U, Y, h)$ sous la forme d'une combinaison linéaire convexe de $\frac{S_{hU}}{n_{hU}}$ et de $\hat{\pi}_U$.

$$VPP(U, Y, h) = \frac{S_{hU}}{n_{hU}} \left(1 - \frac{(1 - \hat{\gamma}_U)/\hat{\gamma}_U}{n_{hU} + (1 - \hat{\gamma}_U)/\hat{\gamma}_U} \right) + \hat{\pi}_U \left(\frac{(1 - \hat{\gamma}_U)/\hat{\gamma}_U}{n_{hU} + (1 - \hat{\gamma}_U)/\hat{\gamma}_U} \right)$$

Réduction de l'ensemble des profils : cas données groupées

Si on considère $\phi_h(U, X)$ le classifieur généré par le profil U pour le cluster h et par $\mathcal{D} = \{(y_i, x_i); i = 1 : n\}$ l'ensemble des observations.

On peut interpréter la sensibilité du classifieur $\phi_h(U, X)$ comme une fonctionnelle de la loi a posteriori de $\phi_h(U, X)$ conditionnellement aux données \mathcal{D} et à $Y = 1$. Si on a U' un profil emboîté dans U alors

$$\frac{\Pr\{\phi_h(U', X) = 1 \mid Y = 1, \mathcal{D}\}}{\Pr\{\phi_h(U, X) = 1 \mid Y = 1, \mathcal{D}\}} = \left[\frac{\Pr\{\phi_h(U', X) = 1 \mid \mathcal{D}\}}{\Pr\{\phi_h(U, X) = 1 \mid \mathcal{D}\}} \right] \left[\frac{VPP(U', Y, h)}{VPP(U, Y, h)} \right]$$

D'où l'interprétation du quotient $\frac{VPP(U', Y, h)}{VPP(U, Y, h)}$ comme un facteur de Bayes.

Réduction de l'ensemble des profils : cas données groupées

On compare deux profils emboîtés en interprétant le quotient de leurs valeurs prédictives positives a posteriori comme un facteur de Bayes.

$$BF(U', U) = \frac{VPP(U', Y, h)}{VPP(U, Y, h)}$$

Facteur de Bayes	Interprétation
1-3	on ne peut pas soutenir que le profil U' est un meilleur classifieur que U
3-10	on peut soutenir que U' est un meilleur classifieur que U
10-100	On peut fortement soutenir que U' est un meilleur classifieur que U
≥ 100	il n'y pas de doute que U' est un meilleur classifieur que U



Robert E. Kass et Adrian E. Raftery (1995), *Bayes Factors*, *Journal of the American Statistical Association*, Vol. 90, No. 430, pp. 773-795

Contexte de l'étude

Le ratio de mortalité maternelle est l'indicateur de la santé public pour lequel la disparité entre pays industrialisés et pays en voie de développement est la plus grande. Il est environ 100 fois plus élevé que dans les pays développés. Des études menées en Afrique de l'Ouest ont révélé des taux de mortalité maternelle très élevés et variables d'un hôpital à un autre, au sein d'un même pays, mais aussi d'un pays à un autre.

Cadre d'étude

- Ce travail est réalisé dans le cadre d'un essai contrôlé randomisé (ECR) : essai QUARITE.
 - l'hôpital est l'unité de randomisation et d'intervention
 - l'unité d'analyse est la patiente admise pour un accouchement

Objectifs de l'étude

Objectifs

- Identifier les profils caractéristiques des patientes décédées chez les femmes enceintes fréquentant les hôpitaux de référence au Mali et au Sénégal.
- Mettre en place une règle de classement
 - qui est performante
 - facile à comprendre par les professionnels de soins de santé
 - peut aider à la prise en charge des patients accouchant dans les établissements de santé.



Présentation des données d'étude

Hôpital de référence

Les hôpitaux qui ont participé à la collecte des données ont été tirés au hasard parmi ceux de leur pays disposant :

- Un bloc opératoire fonctionnel
- Pratiquant au moins 800 accouchements par an
- Un consentement signé par le chef de service de la maternité et le directeur de l'établissement

Au total 46 hôpitaux dont 24 au Sénégal et 22 au Mali

Présentation des données d'étude

Population d'étude

La population d'étude est l'ensemble des femmes enceintes qui sont prises en charge pour un accouchement dans les 46 hôpitaux de référence au cours de la phase pré-intervention.

• Sont incluses

- les femmes admises pour un accouchement
- les parturientes dirigées secondairement vers un des hôpitaux concernés par l'étude

• Sont exclues

- les femmes admises après un accouchement à domicile
- les femmes prises en charge dans une autre structure

Au total 89518 patientes incluses dont 617 décès ($6.9^0/00$)

Analyse des données

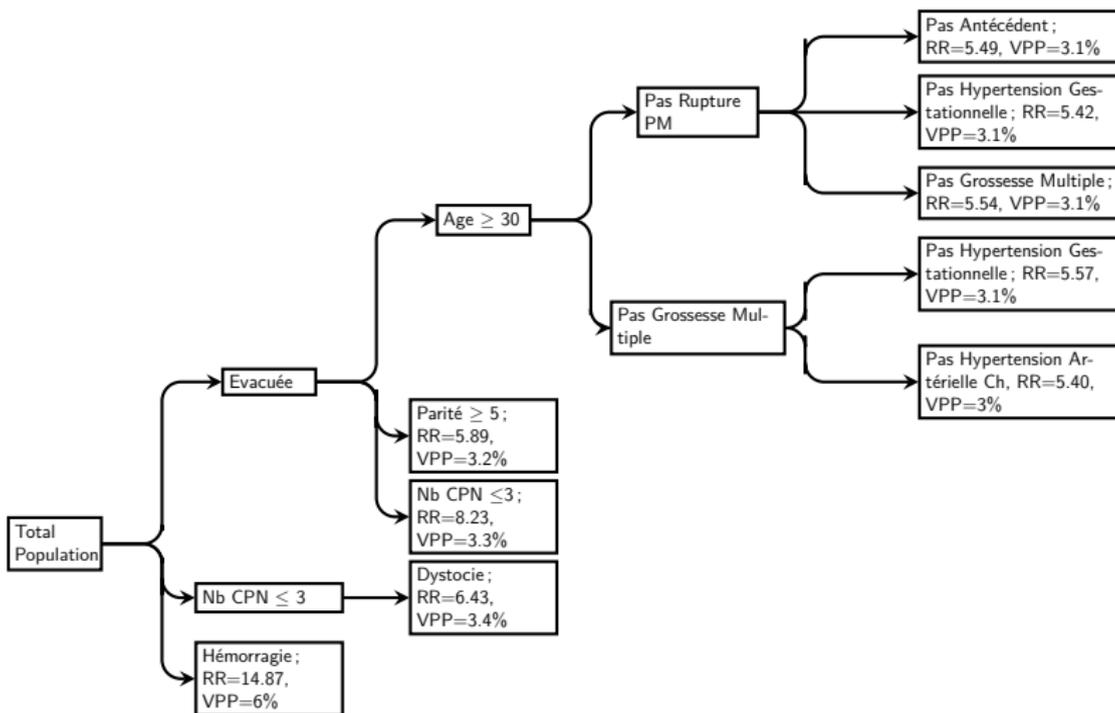
Echantillonnage :

- On sélectionne un sous-échantillon \mathcal{E}_1 de 40 hôpitaux tirés au hasard parmi les 46 hôpitaux de référence.
- On subdivise l'échantillon \mathcal{E}_1 en trois sous-ensembles : Apprentissage, Validation et Test, de telle sorte que chaque sous-ensemble contient les 40 hôpitaux avec les mêmes pourcentages de décès observés dans \mathcal{E}_1 .
- Les 6 hôpitaux restant constituent un échantillon \mathcal{E}_2 . Ces hôpitaux sont considérés comme de nouveaux hôpitaux.

Paramètres d'apprentissage :

- Pour générer l'ensemble des profils fréquents et significativement corrélés avec la variable réponse : $s_0 = 0.002$, $c_0 = 0.03$ et $l_0 = 5$
- On obtient un ensemble de 1522 profils fréquents.

Présentation du classifieur



Présentation des résultats

Hôpital participant à l'estimation de π et γ

Hopital	taille U_{λ}^h	sensib	spécif	erreur	auc
01	9	1.00	0.98	0.02	0.99
03	9	0.92	0.78	0.22	0.85
04	9	0.92	0.54	0.45	0.73
05	9	-	0.97	0.03	-
06	9	0.00	0.92	0.08	0.54
07	9	-	0.97	0.03	-
08	9	0.67	0.88	0.12	0.78
09	9	0.60	0.79	0.21	0.69
11	9	1.00	0.67	0.32	0.84
14	9	0.71	0.59	0.41	0.65
⋮	⋮	⋮	⋮	⋮	⋮
moyenne		0.824	0.756	0.241	0.786

Présentation des résultats

Hôpital n'ayant pas participé à l'estimation de π et γ

- On construit un classifieur général dont l'élaboration est basé sur $\hat{\pi}_U$
- On évalue ses performance sur \mathcal{E}_2

Hopital	taille \mathcal{U}_λ^2	sensib	spécif	erreur	auc
02	9	0.91	0.88	0.12	0.89
10	9	0.87	0.70	0.29	0.79
12	9	0.60	0.88	0.13	0.74
13	9	1.00	0.88	0.12	0.94
21	9	1.00	0.95	0.05	0.97
22	9	0.73	0.78	0.23	0.75
moyenne		0.852	0.845	0.156	0.846

Conclusion

La méthode d'apprentissage statistique que nous avons développée permet :

- 1 de limiter l'impact du déséquilibre des données en produisant un classifieur dont les performances sont sensiblement égales à des méthodes telles que le Boosting et le Random forest.
- 2 de produire un ensemble parcimonieux de profils qui est très facile à interpréter par un personnel soignant même non qualifié.
- 3 la structure d'arbre pourra être affichée dans les salles d'accouchement afin de permettre au personnel de détecter dès à l'admission certains signaux d'alarme.
- 4 En envoyant aux urgences 8 cas de décès sur 10, elle participe à la réduction du taux de mortalité maternelle en Afrique.

Perspective

Dans des travaux à venir, nous allons étendre notre modèle en introduisant une matrice de covariables \mathbf{M} de dimension $n \times q$, où $n = \sum_{h=1}^m n_h$ est le nombre d'observations et q le nombre de caractéristiques observables. Le nombre de covariables est supposé être strictement supérieur à un ($q > 1$).

$$\left\{ \begin{array}{l} Y|_{\theta_h^U, [Y, X]_h} \sim \prod_{k=1}^m \left\{ (\theta_h^U)^{\mathbb{1}_{\{Y=1\}}(y)} \delta_{\{1, [Y, X]_h\}}(U(X), [Y, X]_k) \right. \\ \qquad \qquad \qquad \left. (1 - \theta_h^U)^{(1 - \mathbb{1}_{\{Y=1\}}(y))} \delta_{\{1, [Y, X]_h\}}(U(X), [Y, X]_k) \right\} \\ \\ \theta_h^U | \pi_U, \gamma_U \sim \text{Beta}(\pi_U, \gamma_U) \\ \\ \text{logit}(\pi_U) = \mathbf{W}^t \beta_U \end{array} \right.$$

où $\mathbf{W} = (U(X), \mathbf{M}_i)_{n \times (q+1)}$ est un vecteur de dimension $(q+1)$ et β_U est le vecteur des coefficients de régression associés au profil $U(X)$.

MERCI DE VOTRE ATTENTION

Réduction de l'ensemble des profils : cas des données i.i.d

Soit $(Z_i)_{i=1:n}$ une suite de variables aléatoires indépendantes et identiquement distribuées suivant une distribution de Bernoulli généralisée. Si on considère

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n Z_i$$

l'estimateur empirique de θ , le théorème central limite permet de dire que

$$\sqrt{n} (\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Lambda(\theta))$$

A l'aide du théorème de la méthode Delta généralisé on a :

$$\sqrt{n} (g(\hat{\theta}_n) - g(\theta)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, {}^T \nabla g(\theta) \Lambda(\theta) \nabla g(\theta))$$

où

$${}^T \nabla g(\theta) \Lambda(\theta) \nabla g(\theta) = \sum_{i=1}^6 p_i \nabla_i^2 - \left(\sum_{i=1}^6 p_i \nabla_i \right)^2 > 0$$

Réduction de l'ensemble des profils : cas des données i.i.d

En utilisant le fait que $\hat{\theta}_n \xrightarrow{p.s} \theta$ et la continuité des applications

$$\theta \longrightarrow \nabla g(\theta) \quad \text{et} \quad \theta \longrightarrow \Lambda(\theta)$$

on obtient

$${}^T \nabla g(\hat{\theta}_n) \Lambda(\hat{\theta}_n) \nabla g(\hat{\theta}_n) \xrightarrow{p.s} {}^T \nabla g(\theta) \Lambda(\theta) \nabla g(\theta)$$

A l'aide du théorème de Slutsky, on a

$$\frac{\sqrt{n} \left(g(\hat{\theta}_n) - g(\theta) \right)}{\sqrt{{}^T \nabla g(\hat{\theta}_n) \Lambda(\hat{\theta}_n) \nabla g(\hat{\theta}_n)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

Estimation par maximum de vraisemblance

Pour simplifier les notations, on pose $\tau_U = \frac{1}{\gamma_U} - 1$. La log-vraisemblance des paramètres π_U et τ_U est définie par :

$$L(\pi_U, \tau_U) = \sum_{h=1}^m \left\{ \log \binom{s_h}{n_h} - \sum_{j=0}^{n_h-1} \log(\tau_U + j) + \sum_{k=0}^{s_h-1} \log(\pi_U \tau_U + k) + \sum_{l=0}^{n_h-s_h-1} \log((1 - \pi_U)\tau_U + l) \right\} \quad (14)$$

Problème : Le problème d'optimisation

$$\underset{(\pi_U, \tau_U) \in]0,1[\times]0,+\infty[}{\operatorname{argmax}} \quad L(\pi_U, \tau_U)$$

n'admet pas une solution analytique. On propose d'utiliser l'algorithme MM pour trouver une solution itérative.

Estimation par maximum de vraisemblance

Le principe de l'algorithme consiste à remplacer le problème d'optimisation numérique de la vraisemblance L compliquée à implémenter par celui de l'optimisation d'une fonction auxiliaire g dont l'optimum correspond à un optimum local de L . L'algorithme procède en trois étapes :

- 1 Trouver une fonction auxiliaire g telle que pour (π', τ') fixé,

$$L(\pi, \tau) \geq g(\pi, \tau | \pi', \tau') \quad (15)$$

$$L(\pi, \tau) = g(\pi, \tau | \pi, \tau) \quad \forall (\pi, \tau) \quad (16)$$

- 2 Trouver un couple $(\hat{\pi}, \hat{\tau})$ qui maximise la fonction $g(\pi, \tau | \pi', \tau')$.

$$(\hat{\pi}, \hat{\tau}) = \underset{(\pi, \tau)}{\operatorname{argmax}} g(\pi, \tau | \pi', \tau')$$

- 3 On répète l'étape (2) jusqu'à ce que le couple $(\hat{\pi}, \hat{\tau})$ vérifie un critère d'arrêt.

Estimation par maximum de vraisemblance

Proposition

Soient $L(\pi_U, \tau_U)$ la log-vraisemblance du couple des paramètres (π_U, τ_U) et (π'_U, τ'_U) une valeur connue des paramètres (π_U, τ_U) . La fonction auxiliaire $g(\pi_U, \tau_U | \pi'_U, \tau'_U)$ définie par

$$\begin{aligned}g(\pi_U, \tau_U | \pi'_U, \tau'_U) &= A(\pi'_U, \tau'_U)[\log(\pi_U) + \log(\tau_U)] \\ &+ B(\pi'_U, \tau'_U)[\log(1 - \pi_U) + \log(\tau_U)] \\ &- (\tau_U - \tau'_U)C(\tau'_U) + D(\pi'_U, \tau'_U)\end{aligned}$$

vérifie les conditions (15) et (16).

Il en résulte que

$$\hat{\pi}_U = \frac{A(\pi'_U, \tau'_U)}{A(\pi'_U, \tau'_U) + B(\pi'_U, \tau'_U)} \quad \hat{\tau}_U = \frac{A(\pi'_U, \tau'_U) + B(\pi'_U, \tau'_U)}{C(\tau'_U)}$$

Estimation par maximum de vraisemblance

Algorithme : MM (Minimisation-Maximisation)

- Entrées : $\mathcal{D} = \{(s_h, n_h); h = 1 : m\}$; valeurs initiales : (π^0, τ^0) et *maxiter*.
- Sortie : le couple $(\widehat{\pi}, \widehat{\tau})$
- *cond.arret* = vrai ; $t = 0$ et $(\pi^t, \tau^t) \leftarrow (\pi^0, \tau^0)$

Tant que *cond.arret* est vrai **faire** :

$$t \leftarrow t + 1; \quad (\pi^{(t-1)}, \tau^{(t-1)}) \leftarrow (\pi^{(t)}, \tau^{(t)})$$

- $$\pi^{(t)} \leftarrow \frac{A(\pi^{(t-1)}, \tau^{(t-1)})}{A(\pi^{(t-1)}, \tau^{(t-1)}) + B(\pi^{(t-1)}, \tau^{(t-1)})}$$

- $$\tau^{(t)} \leftarrow \frac{A(\pi^{(t-1)}, \tau^{(t-1)}) + B(\pi^{(t-1)}, \tau^{(t-1)})}{C(\tau^{(t-1)})}$$

$$\textit{cond.arret} = \left\{ \left((\pi^{(t)} - \pi^{(t-1)})^2 + (\tau^{(t)} - \tau^{(t-1)})^2 + 1 \neq 1 \right), (t < \textit{maxiter}) \right\}$$

fin tant que

Valeurs estimées des RMSE de $\pi_U = 0.007$ et $\gamma_U = 0.005$

B=250

	n=20.000, m=10			n=50.000, m=50			n=100.000, m=100		
	MOMK	MOMG	EMV	MOMK	MOMG	EMV	MOMK	MOMG	EMV
RMSE ($\hat{\pi}$)	0.00193	0.00193	0.00193	0.00105	0.00106	0.00105	0.00075	0.00075	0.00075
RMSE ($\hat{\gamma}$)	0.00407	0.00337	0.00315	0.00238	0.00204	0.00152	0.00155	0.00147	0.00121

B=250

	n=200.000, m=150			n=300.000, m=200			n=400.000, m=250		
	MOMK	MOMG	EMV	MOMK	MOMG	EMV	MOMK	MOMG	EMV
RMSE ($\hat{\pi}$)	0.00055	0.00055	0.00055	0.00048	0.00048	0.00049	0.00041	0.00041	0.00041
RMSE ($\hat{\gamma}$)	0.00126	0.00114	0.00087	0.00095	0.00088	0.00067	0.00091	0.00091	0.00064

Valeurs estimées des RMSE de $\pi = 0.007$ et $\gamma = 0.05$

B=250

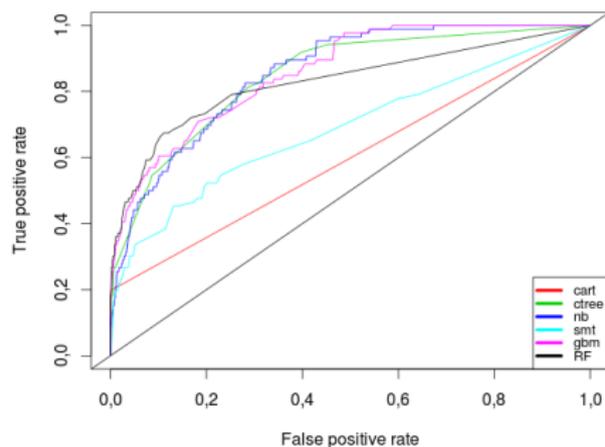
	n=20.000, m=10			n=50.000, m=50			n=100.000, m=100		
	MOMK	MOMG	EMV	MOMK	MOMG	EMV	MOMK	MOMG	EMV
RMSE ($\hat{\pi}$)	0.00557	0.00560	0.00542	0.00248	0.00248	0.00244	0.00185	0.00186	0.00184
RMSE ($\hat{\gamma}$)	0.03650	0.03622	0.04707	0.02239	0.02210	0.02024	0.01952	0.01936	0.01538

B=250

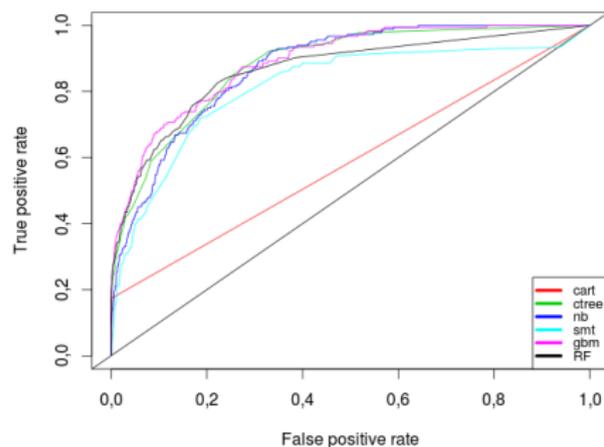
	n=200.000, m=150			n=300.000, m=200			n=400.000, m=250		
	MOMK	MOMG	EMV	MOMK	MOMG	EMV	MOMK	MOMG	EMV
RMSE ($\hat{\pi}$)	0.00147	0.00147	0.00147	0.00129	0.00130	0.00129	0.00122	0.00122	0.00122
RMSE ($\hat{\gamma}$)	0.01686	0.01640	0.01134	0.01700	0.01540	0.01077	0.01403	0.01385	0.00971

Courbes ROC des méthodes alternatives

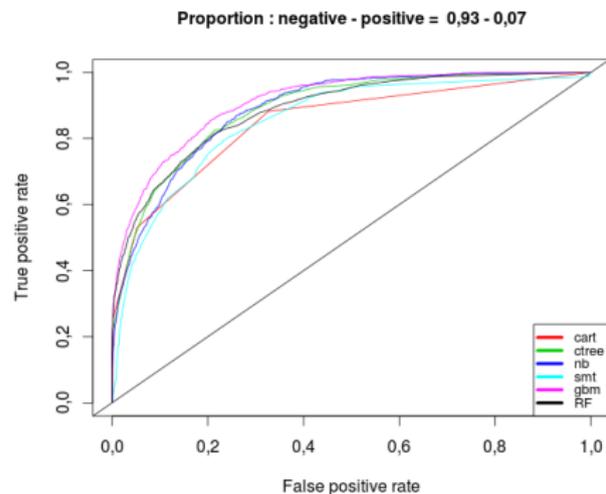
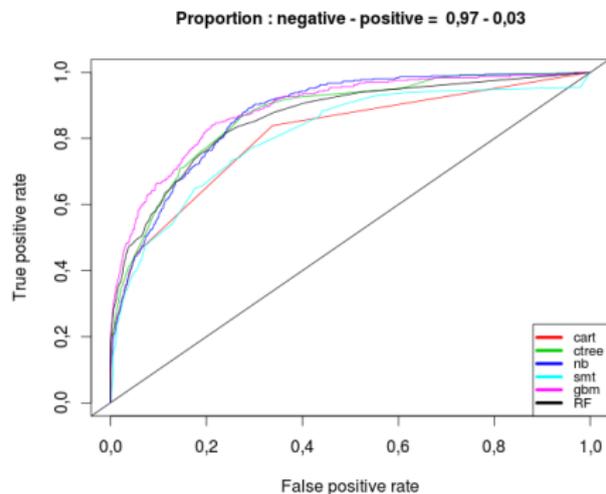
Proportion : negative - positive = 0,993 - 0,007



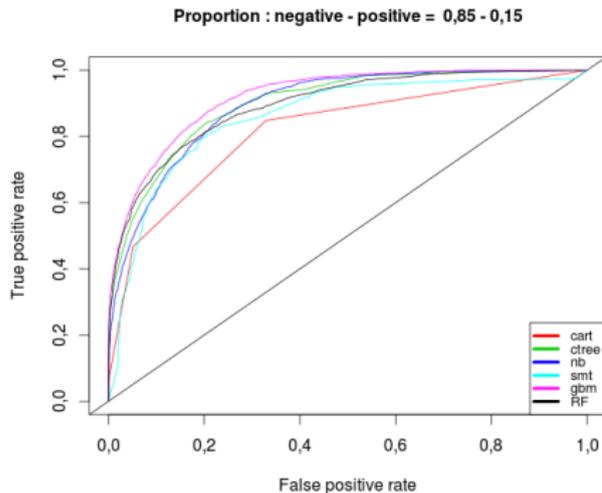
Proportion : negative - positive = 0,985 - 0,015



Courbes ROC des méthodes alternatives



Courbes ROC des méthodes alternatives



On peut constater à partir des graphes ci-dessus que lorsque la proportion d'observations positives devient de plus en plus grande, les courbes ROC se rapprochent de plus en plus.