

Modélisation de processus multi-états, risques concurrents et évènements récurrents

Ségolen Geffray

17 Octobre 2008

M2 Modélisation en Pharmacologie clinique et Epidémiologie

- 1 Modélisation de processus multi-états
 - Etats, transitions, modèles classiques
 - Modèle markovien ou semi-markovien ?
 - Modèle homogène ou non-homogène ?
- 2 Risques concurrents
 - Risques dépendants ou indépendants ?
 - Problème d'identifiabilité
 - Risques concurrents dépendants en présence de censure indépendante
- 3 Evènements récurrents
 - Processus des évènements récurrents, instants d'arrivée, instants inter-arrivées
 - Approche conditionnelle, non-conditionnelle, "frailty"
 - Quelques modèles extraits de la littérature
- 4 Evènements récurrents en présence de décès avec peu de récurrences
 - Résultats pour une étude réelle

1^{ère} partie

Modélisation de processus multi-états

Quelques définitions

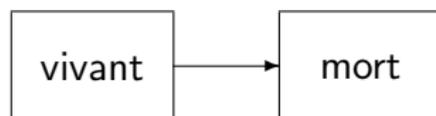
- **Processus multi-états X** : processus stochastique qui représente l'état occupé par un individu au cours du temps. L'espace des états est un ensemble discret (dénombrable). Lorsque $X(t) = i$, cela signifie que l'individu se trouve dans l'état i à l'instant t .
- **Etat** : condition physique (sain, malade, mort,...). Un état est **absorbant** si aucune sortie de cet état n'est possible (ex : mort).
- **Transition** : changement d'état = évènement (déclaration d'une maladie, survenue de complication).
- Le modèle statistique spécifie quelles sont les transitions possibles et parmi celles-ci quelles sont les forces de passage entre états au moyen de **probabilités de transition** et **fonctions de risque (ou hasard) instantané**.

Modèle de survie (à 1 cause de mort)

Pour chaque patient, on observe un évènement et un seul : le décès

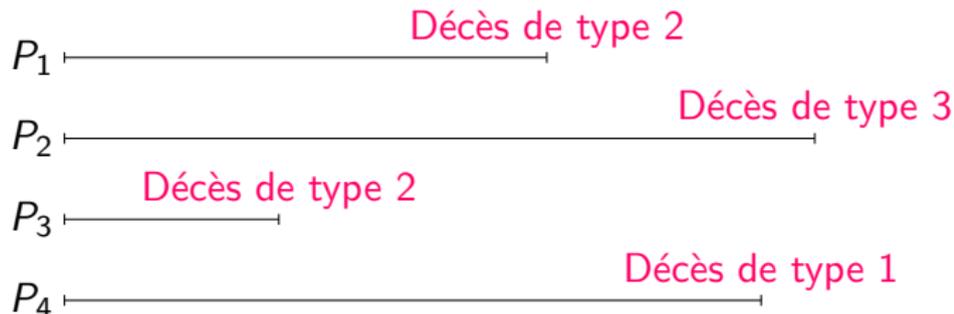


Modèle à 2 états :

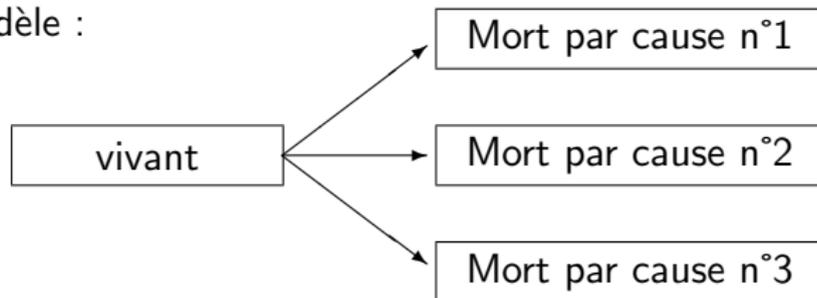


Modèle des risques concurrents

Plusieurs types d'évènements différents sont susceptibles de survenir mais on observe un seul évènement par patient.



Modèle :



Cas particulier : 2 risques concurrents

- Lorsqu'il y a seulement deux risques concurrents, on retrouve un schéma de censure aléatoire droite. Les 2 risques "se censurent l'un l'autre".
- X : v.a. positive représentant une durée de loi F donnée pour $t \geq 0$ par $F(t) = \mathbb{P}[X \leq t]$.
NB : de manière équivalente, on peut considérer la fonction de survie : $S(t) = \bar{F}(t) = 1 - F(t) = \mathbb{P}[X > t]$
- X est censurée aléatoirement à droite par une v.a. positive C de loi G donnée pour $t \geq 0$ par $G(t) = \mathbb{P}[C \leq t]$.
- On observe seulement :

$$T = X \wedge C,$$

$$\delta = I(X \leq C).$$

Rappel : estimateur de Kaplan-Meier

Lorsque X et C sont **indépendantes**, sur la base d'un échantillon i.i.d., on peut estimer F (et G) de façon non-paramétrique par l'estimateur de Kaplan-Meier. Soit $(T_i, \delta_i)_{i=1, \dots, n}$ un échantillon distribué comme (T, δ) .

$$1 - \hat{F}_n(t) = \prod_{i=1}^n \left(1 - \frac{I(T_i \leq t, \delta_i = 1)}{\sum_{j=1}^n I(T_j \geq T_i)} \right),$$

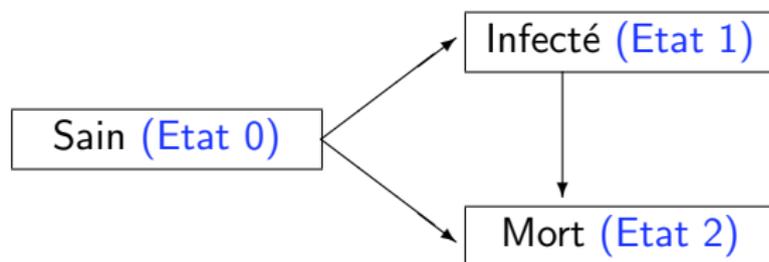
$$1 - \hat{G}_n(t) = \prod_{i=1}^n \left(1 - \frac{I(T_i \leq t, \delta_i = 0)}{\sum_{j=1}^n I(T_j \geq T_i)} \right).$$

Lorsque le poids de la censure est élevé, l'estimateur de KM présente un problème de biais et d'explosion de variance quand t augmente.

Modèle maladie incurable-mortalité

(illness-death, disability-death)

exemple : infection à Herpès virus

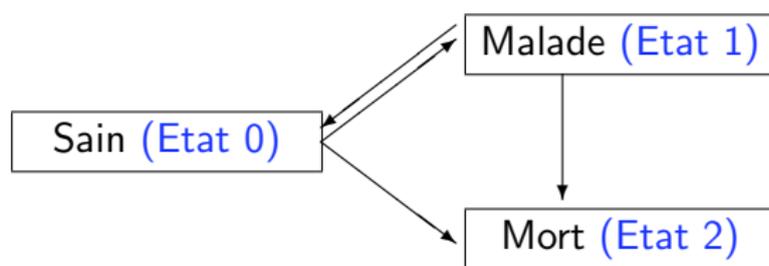


Interprétons les données suivantes (temps t exprimé en mois) :

- $X_1(t) = 0$ pour $0 \leq t < 500$, $X_1(t) = 1$ pour $500 \leq t < 750$ et $X_1(t) = 2$ pour $t \geq 750$
- $X_2(t) = 0$ pour $0 \leq t < 200$, $X_2(t) = 1$ pour $200 \leq t < 900$ et $X_2(t) = 2$ pour $t \geq 900$
- $X_3(t) = 0$ pour $0 \leq t < 800$, et $X_3(t) = 2$ pour $t \geq 800$

Modèle maladie curable-mortalité

exemple : grippe

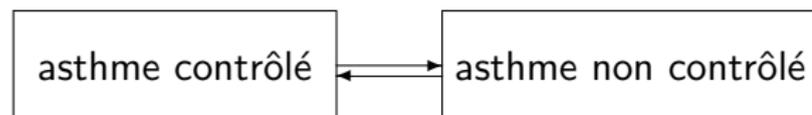


Interprétons les données suivantes (temps t exprimé en mois) :

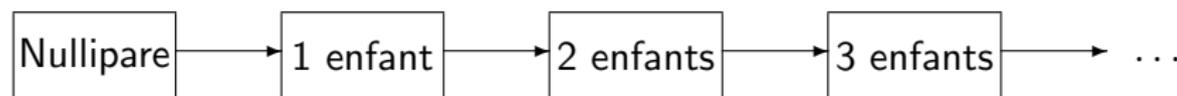
- $X_1(t) = 0$ pour $0 \leq t < 120$, $X_1(t) = 1$ pour $120 \leq t < 121$,
 $X_1(t) = 0$ pour $121 \leq t < 240$, $X_1(t) = 1$ pour $500 \leq t < 502$,
 $X_1(t) = 0$ pour $502 \leq t < 850$ et $X_1(t) = 2$ pour $t \geq 750$
- $X_2(t) = 0$ pour $0 \leq t < 800$, $X_2(t) = 1$ pour $800 \leq t < 801$ et
 $X_2(t) = 2$ pour $t \geq 801$
- $X_3(t) = 0$ pour $0 \leq t < 700$, et $X_3(t) = 2$ pour $t \geq 700$

Modèles d'évènements récurrents

Le modèle dépend de ce que l'on veut montrer!!!! Considérons le cas de patients asthmatiques sous traitement pour lesquels on étudie le contrôle de l'asthme.



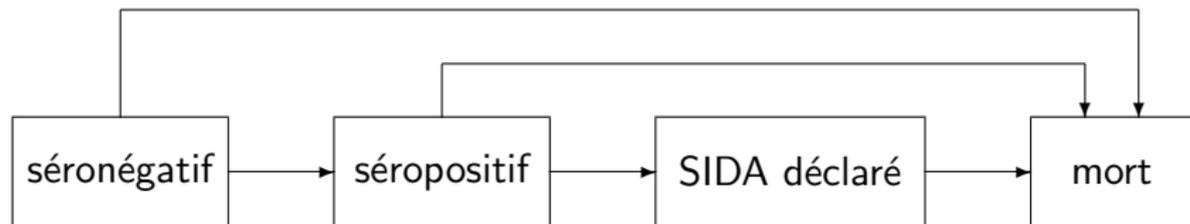
Considérons le cas de l'étude de la fertilité des femmes en âge de procréer.



Il n'est pas pertinent de mettre le décès dans ces modèles.

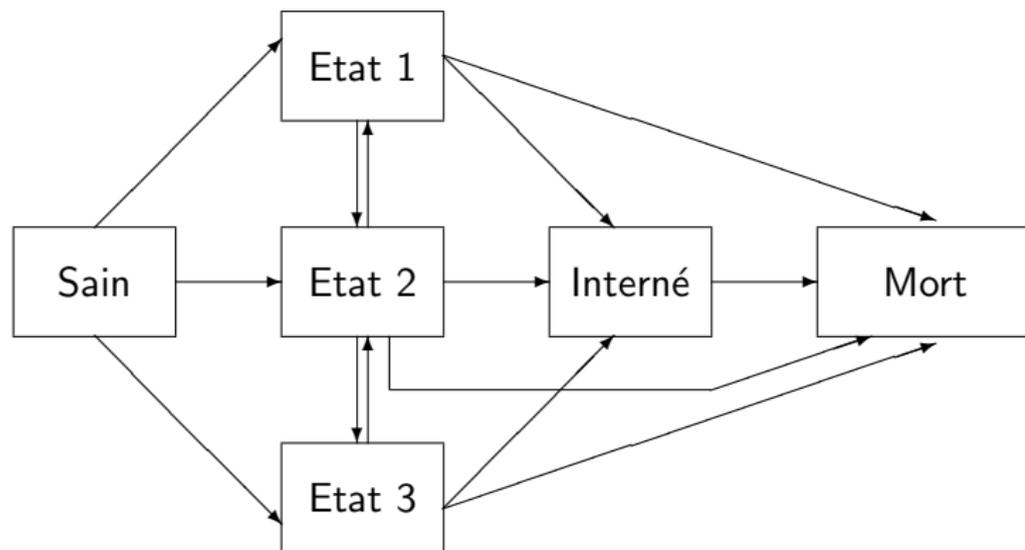
Modèle pour l'infection à VIH

Considérons une étude de l'épidémie de SIDA dans un pays donné. L'individu traverse des états successifs selon une évolution irréversible.



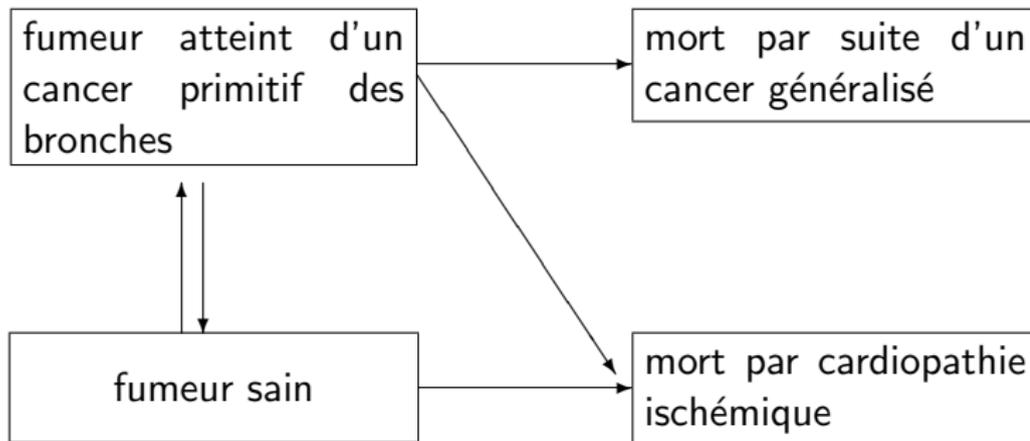
Modèle de démence sénile

Considérons une étude de l'évolution des malades d'Alzheimer. C'est une maladie à états transitoires avec éventuellement des périodes d'amélioration avant que l'état du patient nécessite éventuellement le placement dans un institut spécialisé.



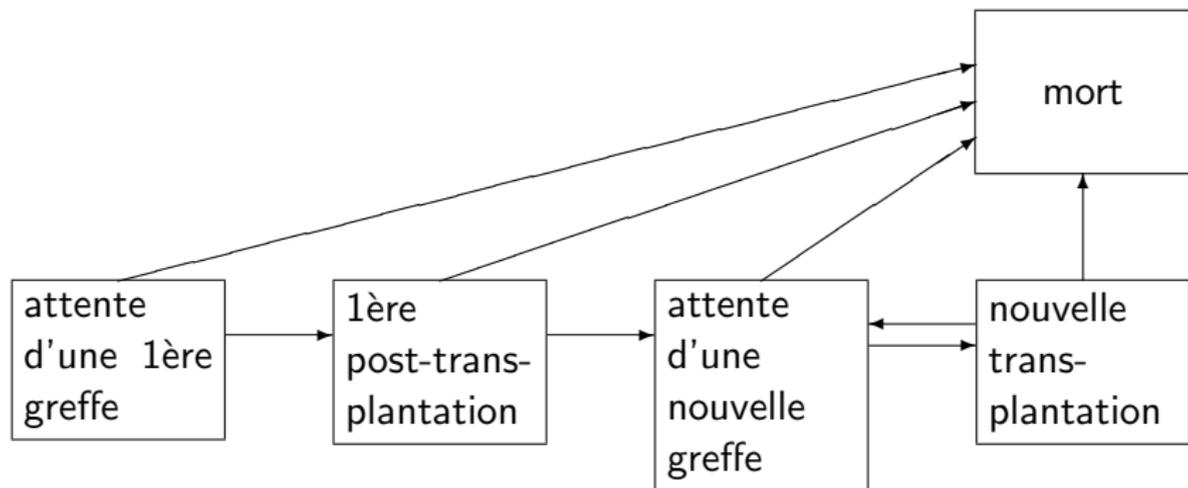
Modèle de pathologie due au tabac

Considérons une étude portant sur la population des fumeurs.



Modèle de transplantation rénale

Considérons le cas de patients sous dialyse dont l'état de santé nécessite une greffe rénale. Les médecins savent que l'évolution médicale d'un patient est différente selon qu'il a subi ou non une 1ère greffe rénale.



Quelques définitions supplémentaires

- Formellement, étudier un processus multi-états (à temps continu) revient à étudier une famille de variables aléatoires $\{X(t) : t \geq 0\}$ où les valeurs prises par $X(t)$ sont des entiers codant pour les différents états possibles.
- Pour cela, on étudie les probabilités multivariées pour tout entier k et pour toute suite d'instants ordonnés notés $t_1 < t_2 < \dots < t_k < t_{k+1}$:

$$\mathbb{P}\left[X(t_{k+1}) = j | X(t_1), \dots, X(t_k)\right]$$

- Parfois, il est pertinent de prendre en compte tout le passé des individus. Parfois, il est plus pertinent de ne pas surcharger l'analyse avec le passé.

- **Hypothèse de Markov** : le passé est entièrement résumé par la seule connaissance de l'état présent. On en déduit :

$$\mathbb{P}\left[X(t_{k+1}) = j | X(t_1), \dots, X(t_k)\right] = \mathbb{P}\left[X(t_{k+1}) = j | X(t_k)\right]$$

- On s'intéresse alors à la probabilité de transition de l'état i vers l'état j (avec $s \leq t$) :

$$P_{i,j}(s, t) = \mathbb{P}\left[X(t) = j | X(s) = i\right].$$

- On s'intéresse également à la fonction de risque instantané de la transition du processus X de l'état i vers l'état j (=force de transition de l'état i vers l'état j) :

$$\alpha_{i,j}(t) = \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{P}\left[X(t+h) = j | X(t) = i\right].$$

Modèle de Markov (suite)

- **Modèle de Markov homogène** : les fonctions de risque instantané des différentes transitions sont constantes ($\alpha_{i,j}(t) = \alpha_{i,j}$) et les temps de séjour suivent une loi exponentielle (sans mémoire) : restrictif!
- **Modèle de Markov inhomogène** : les taux de risque instantané dépendent de l'âge (ou du temps calendaire) t . Ce modèle est bien adapté au cas où un évènement donné a une importance particulière dans l'histoire de l'individu (ex : début d'un traitement ou déclaration d'une maladie qui amènent à considérer des proba de transition qui dépendent de l'âge).

Hypothèse semi-markovienne

- **Hypothèse semi-markovienne** : le passé est parfaitement résumé par la connaissance de l'état présent **et** de la durée depuis laquelle l'individu est dans l'état présent.
- **Modèle semi-markovien homogène** : les fonctions de risque instantané des différentes transitions du processus ne dépendent que de d la durée écoulée depuis l'entrée dans l'état présent. Ce modèle est bien adapté au cas où le patient peut être stabilisé dans un état donné.
- **Modèle semi-markovien inhomogène** : les fonctions de hasard instantané des différentes transitions du processus dépendent de l'âge (ou du temps calendaire) t et de d la durée écoulée depuis l'entrée dans l'état présent. Ce modèle est peu utilisé car il est complexe.

Modèle semi-markovien homogène (SMH)

- Le modèle SMH peut se reformuler ainsi : les observations consistent en $(Y_k, J_k)_{k \geq 0}$ où $0 = Y_0, Y_1, Y_2, \dots$ sont les temps de séjour consécutifs (ou durées inter-arrivées) dans les états J_0, J_1, J_2, \dots avec $J_{k-1} \neq J_k$ pour tout k .



- L'hyp SMH se formule alors ainsi :

$$\begin{aligned} & \mathbb{P} \left[J_k = j, Y_k \leq t \mid J_0, Y_1, J_1, J_2, Y_2, \dots, J_{k-1}, Y_{k-1} \right] \\ &= \mathbb{P} \left[J_k = j, Y_k \leq t \mid J_{k-1} \right] \end{aligned}$$

i.e. les temps de séjour sont indépendants et leurs distributions dépendent uniquement des états contigus

Modèle semi-markovien homogène (suite)

- On s'intéresse alors à la probabilité de transition de l'état i vers l'état j (indépendante de k =numéro de la transition) pour $s \leq t$:

$$P_{i,j}(t) = \mathbb{P}\left[J_k = j, Y_k \leq t | J_{k-1} = i\right].$$

- On s'intéresse également à la fonction de risque (hasard) instantané du temps de séjour de la transition de l'état i vers l'état j :

$$\lambda_{i,j}(t) = \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{P}\left[t \leq Y_k < t + h | Y_k \geq t, J_{k-1} = i, J_k = j\right]$$

- On s'intéresse aussi à la fonction de risque (hasard) instantané du processus SMH de la transition de l'état i vers l'état j :

$$\alpha_{i,j}(t) = \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{P}\left[t \leq Y_k < t + h, J_k = j | Y_k \geq t, J_{k-1} = i\right]$$

Modèle semi-markovien homogène (suite2)

Compte tenu de l'hypothèse SMH, on obtient :

$$\alpha_{i,j}(t) = \frac{\mathbb{P}[J_k = j | J_{k-1} = i]}{\mathbb{P}[Y_k \geq t | J_{k-1} = i]} \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{P}[t \leq Y_k < t + h | J_k = j, J_{k-1} = i]$$

ainsi $\alpha_{i,j}(t)$ est d'autant plus grande que

- $\mathbb{P}[J_k = j | J_{k-1} = i]$ = proba de transition de i vers j est grande
- $\mathbb{P}[Y_k \geq t | J_{k-1} = i]$ = fonction de survie dans i est faible i.e. le temps déjà passé dans i est grand
- la densité $f_{i,j}(t)$ est grande :
$$f_{i,j}(t) = \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{P}[t \leq Y_k < t + h | J_k = j, J_{k-1} = i]$$

Illustration du modèle de Markov homogène

Question importante : dans quel état est le patient ?

Nb : ne convient pas pour décrire la fertilité d'une femme car il y a au moins 9 mois entre 2 accouchements.

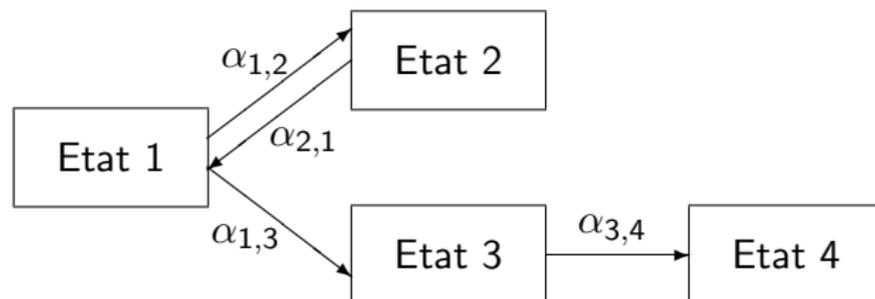


Illustration du modèle de Markov inhomogène

Questions importantes : dans quel état est le patient ? quel âge a-t-il ?

Nb : ne convient tjs pas pour décrire la fertilité d'une femme car il y a au moins 9 mois entre 2 accouchements.

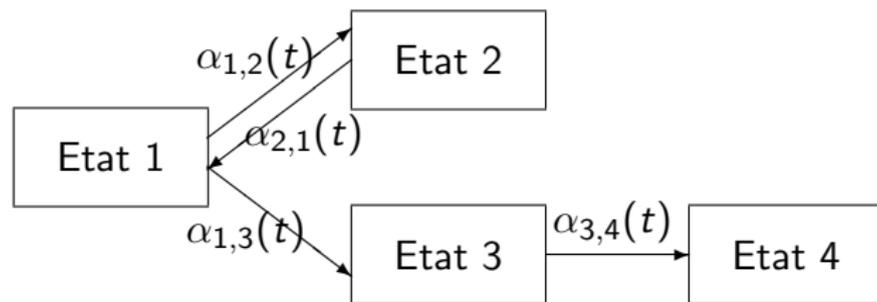


Illustration du modèle semi-markovien homogène

Questions importantes : dans quel état est le patient ? depuis quand est-t-il comme ça ?

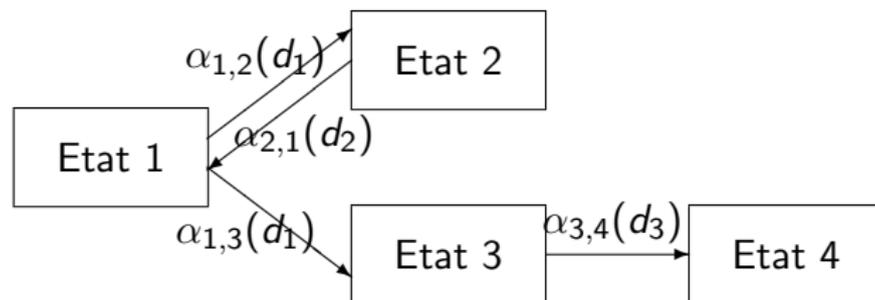
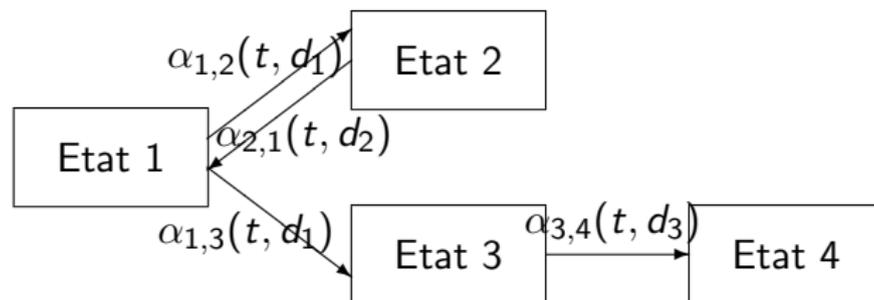


Illustration du modèle semi-markovien inhomogène

Questions importantes : dans quel état est le patient ? depuis quand est-t-il comme ça ? quel âge a-t-il ?



2^{ème} partie

Risques concurrents

Définition

- \mathcal{J} : nombre total de risques ($\mathcal{J} \geq 2$).
- X_j pour $j \in \{1, \dots, \mathcal{J}\}$: durée jusqu'à l'apparition d'un évènement du $j^{\text{ème}}$ type en l'absence des autres risques.
- On observe seulement :

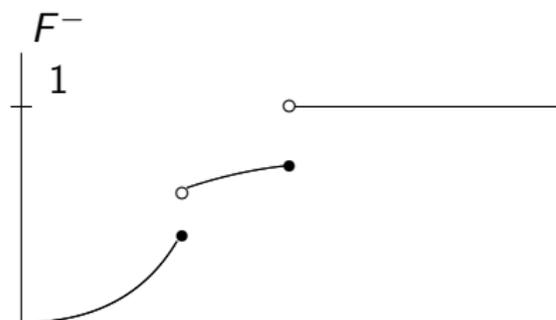
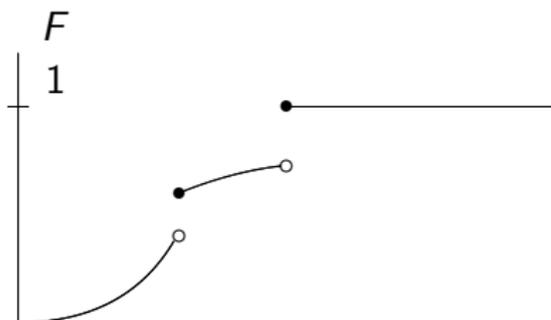
$$X = \min(X_1, \dots, X_{\mathcal{J}}),$$

$$\mathcal{C} = j \text{ lorsque } \min(X_1, \dots, X_{\mathcal{J}}) = X_j.$$

i.e. $\mathcal{C} = j$ lorsque l'évènement est dû au $j^{\text{ème}}$ risque parmi les \mathcal{J} causes possibles.

Rappel

Une fonction de répartition F est croissante, nulle en $t = 0$, vaut 1 en $+\infty$, est continue à droite i.e. est continue sauf éventuellement en un nombre dénombrable de points et on a $F(a) = \lim_{x \rightarrow a^+} F(x)$. La modification continue à gauche de F notée F^- est croissante, nulle en $t = 0$, vaut 1 en $+\infty$ mais est continue à gauche.



Différentes fonctions de répartition

- Fonction de répartition (ou d'incidence) globale :

$$F(t) = \mathbb{P}[X \leq t]$$

- Fonction de répartition (incidence) nette associée à la $j^{\text{ème}}$ cause :

$$F_j(t) = \mathbb{P}[X_j \leq t].$$

- Fonction de répartition (ou incidence) spécifique (ou brute) associée à la $j^{\text{ème}}$ cause :

$$F^{(j)}(t) = \mathbb{P}[X \leq t, \mathcal{C} = j].$$

- On a toujours $F = \sum_{j=1}^J F^{(j)}$. Si les risques concurrents sont **indépendants**, alors $1 - F = \prod_{j=1}^J (1 - F_j)$.
- NB : $F^{(j)}(\infty) < 1$ et $\mathbb{P}[X > t, \mathcal{C} = j] \neq 1 - \mathbb{P}[X \leq t, \mathcal{C} = j]$.

Différentes fonctions de hasard instantané

- Fonction de hasard (risque) instantané global :

$$\lambda(t) = \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{P} \left[t \leq X < t + h \mid X \geq t \right].$$

- Fonction de hasard instantané nette associée à la $j^{\text{ème}}$ cause :

$$\lambda_j(t) = \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{P} \left[t \leq X_j < t + h \mid X_j \geq t \right].$$

- Fonction de hasard instantané spécifique (ou brute) associée à la $j^{\text{ème}}$ cause :

$$\lambda^{(j)}(t) = \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{P} \left[t \leq X < t + h, \mathcal{C} = j \mid X \geq t \right].$$

- On a toujours

$$\lambda = \sum_{j=1}^{\partial} \lambda^{(j)}.$$

Différentes fonctions de hasard cumulées

- Fonction de risque cumulée globale :

$$\Lambda(t) = \int_0^t \lambda(u) du = \int_0^t \frac{dF(u)}{1 - F^-(u)}.$$

- Fonction de risque cumulée nette associée à la $j^{\text{ème}}$ cause :

$$\Lambda_j(t) = \int_0^t \lambda_j(u) du = \int_0^t \frac{dF_j(u)}{1 - F_j^-(u)}.$$

- Fonction de risque cumulée spécifique (ou brute) associée à la $j^{\text{ème}}$ cause :

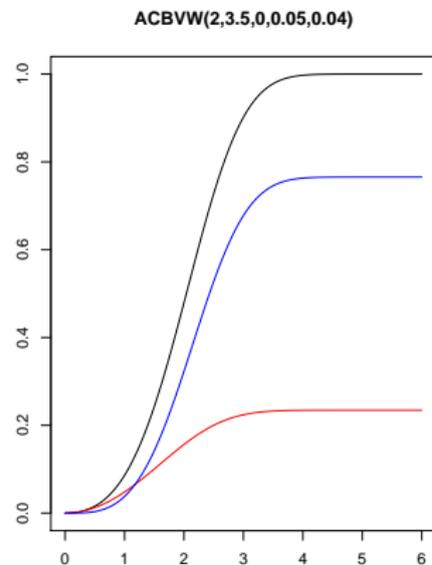
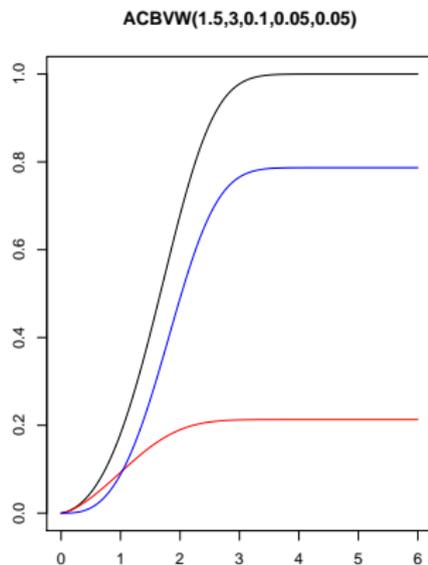
$$\Lambda^{(j)}(t) = \int_0^t \lambda^{(j)}(u) du = \int_0^t \frac{dF^{(j)}(u)}{1 - F^-(u)}.$$

- On a toujours $\Lambda = \sum_{j=1}^{\mathcal{J}} \Lambda^{(j)}$. En général, $\Lambda^{(j)} \neq \Lambda_j$ mais si les risques concurrents sont **indépendants**, alors $\Lambda^{(j)} = \Lambda_j$.

Le problème d'identifiabilité

- Problème : on ne peut pas tester l'indépendance des différents risques!!!! Il faut donc effectuer ou non une hypothèse d'indépendance...
- Constatons que l'indépendance ne se voit pas à l'oeil nu!!!!
- 1er schéma : tracé de fonctions de répartition spécifiques pour 2 risques dépendants
- 2ème schéma : tracé de fonctions de répartition spécifiques pour 2 risques indépendants

Tracé de fonctions de répartition spécifiques



Estimation pour des risques concurrents dépendants

- Si les risques concurrents sont **dépendants**, les quantités d'intérêt sont les fonctions $F^{(j)}$. On peut les estimer pour $t \geq 0$ par

$$F_n^{(j)}(t) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq t, \mathcal{C}_i = j)$$

et on a :

$$F_n(t) = \sum_{j=1}^{\mathcal{J}} F_n^{(j)}(t) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq t).$$

- NB : on ne sait pas estimer les fonctions F_j sauf en cas d'hypothèses restrictives.

Estimation pour des risques concurrents indépendants

- Si les risques concurrents sont **indépendants**, les quantités d'intérêt sont les fonctions F_j . On peut les estimer pour $t \geq 0$ par

$$\widehat{F}_{n,j}(t) = 1 - \prod_{i=1}^n \left(1 - \frac{I(X_i \leq t, \mathcal{C}_i = j)}{\sum_{j=1}^n I(X_j \geq X_i)} \right)$$

- On peut encore estimer les fonctions $F^{(j)}$ pour $t \geq 0$ par

$$F_n^{(j)}(t) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq t, \mathcal{C}_i = j).$$

- On a :

$$F_n(t) = 1 - \prod_{j=1}^{\mathcal{J}} \left(1 - \widehat{F}_{n,j}(t) \right) = \sum_{j=1}^{\mathcal{J}} F_n^{(j)}(t) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq t).$$

Risques concurrents dépendants en présence de censure indépendante

- X_j pour $j \in \{1, \dots, J\}$: durée jusqu'à l'apparition d'un évènement du $j^{\text{ème}}$ type. On observe seulement :

$$X = \min(X_1, \dots, X_J),$$

$$C = j \text{ lorsque } \min(X_1, \dots, X_J) = X_j.$$

Les variables X_j ne sont pas supposées indépendantes.

- X (de loi F) est censurée aléatoirement à droite par C de fonction de répartition G , **indépendante** de X et de C .
- On observe seulement :

$$T = \min(X, C),$$

$$J = C I(X \leq C).$$

Estimation

- Echantillon i.i.d. : $(T_i, J_i)_{i=1, \dots, n}$.
- Soit H la loi de T définie par $H(t) = \mathbb{P}[T \leq t]$ et soit $H^{(1,j)}$ définie par $H^{(1,j)}(t) = \mathbb{P}[T \leq t, J = j]$.
- Fonctions de répartition empiriques associées à H et $H^{(1,j)}$:

$$H_n(t) = \frac{1}{n} \sum_{i=1}^n I(T_i \leq t),$$

$$H_n^{(1,j)}(t) = \frac{1}{n} \sum_{i=1}^n I(T_i \leq t, J_i = j).$$

- Estimateur de Kaplan-Meier de F ($F(t) = \mathbb{P}[X \leq t]$) :

$$\hat{F}_n(t) = 1 - \prod_{i=1}^n \left(1 - \frac{I(T_i \leq t, J_i \neq 0)}{\sum_{j=1}^n I(T_j \geq T_i)} \right).$$

Estimateur de Aalen-Johansen

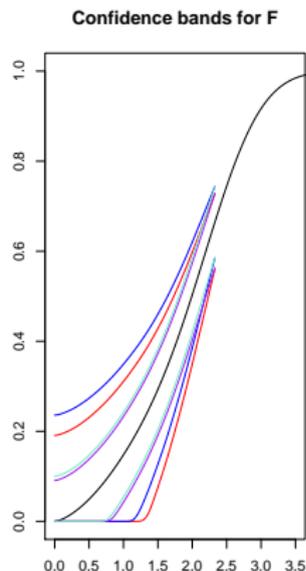
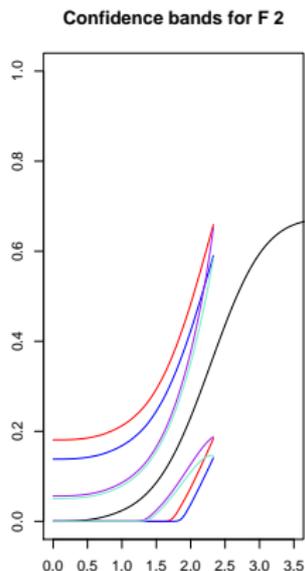
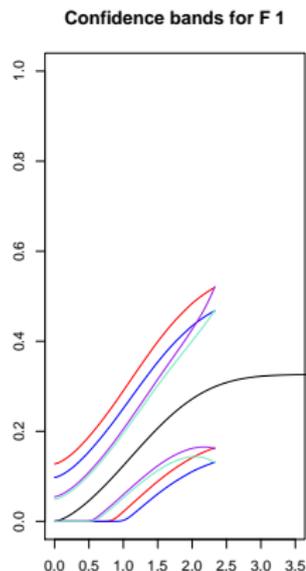
- **Hypothèse classique et nécessaire** : Pour $j \in \{1, \dots, \mathcal{J}\}$, les fonctions $F^{(j)}$ ont des points de discontinuité différents et donc il n'y a pas d'ex aequo entre deux événements différents observés chacun chez deux individus différents.
- Estimateur de Aalen-Johansen de $F^{(j)}(t) = \mathbb{P}[X \leq t, \mathcal{C} = j]$:

$$\widehat{F}_n^{(j)}(t) = \frac{1}{n} \sum_{i=1}^n \frac{1 - \widehat{F}_n^-(T_i)}{1 - H_n^-(T_i)} I(T_i \leq t, \mathcal{C}_i = j) = \int_0^t \frac{1 - \widehat{F}_n^-}{1 - H_n^-} dH_n^{(1,j)}$$

- NB : l'estimateur de Aalen-Johansen de $F^{(j)}(t)$ est obtenu par "plug-in" à partir de la relation suivante pour $t \geq 0$:

$$F^{(j)}(t) = \int_0^t \frac{1 - F^-}{1 - H^-} dH^{(1,j)}.$$

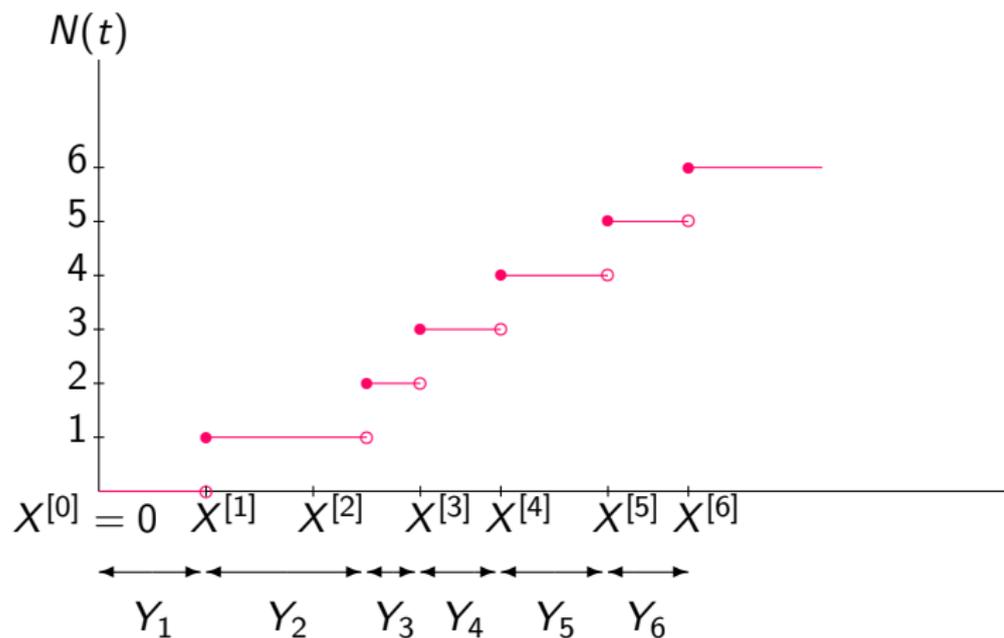
Bandes de confiance pour les fonctions d'incidence en présence de censure



3^{ème} partie

Evènements récurrents

Quantités d'intérêt



Variables d'intérêt

- $N(t)$ = nombre d'évènements récurrents se produisant dans l'intervalle de temps $[0, t]$
- $X^{[k]}$ = instant survenue du k ème évènement récurrent
- $Y^{[k]} = X^{[k]} - X^{[k-1]}$ = k ème durée inter-arrivée
- \mathbf{Z} = vecteur de covariables indépendantes du temps ou bien $\mathbf{Z}(t)$ vecteur de covariables externes dépendantes du temps
- D = instant de décès (dépendant de $N(\cdot)$, des $X^{[k]}$ et des $Y^{[k]}$) lorsqu'il est pris en compte
- C = instant de censure (définitive) **indépendante** de $N(\cdot)$, des $X^{[k]}$, $Y^{[k]}$, de D et éventuellement de \mathbf{Z}

- $N_C(t) = N(\min(t, C))$
- $T_X^{[k]} = \min(X^{[k]}, C)$ pour $k \leq N_C(\infty)$
- $T_Y^{[k]} = T_X^{[k]} - T_X^{[k-1]}$ pour $k \leq N_C(\infty)$
- $J^{[k]} = I(X^{[k]} \leq C)$ pour $k \leq N_C(\infty)$
- et éventuellement $T = \min(D, C)$ avec $\delta = I(D \leq C)$
- et éventuellement **Z**

Différentes approches

- **approche conditionnelle** pour $N(\cdot)$: consiste à modéliser la fonction d'intensité du processus $N(\cdot)$ ou celle des durées $X^{[k]}$ ou $Y^{[k]}$. Dans tous les cas, la fonction d'intensité est définie **conditionnellement** à l'histoire médicale du patient recueillie jusqu'à l'instant noté t^- précédant immédiatement l'instant t . L'intensité du processus $N(\cdot)$ est donnée par :

$$\begin{aligned}\lambda(t) &= \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{P}[N(t+h) - N(t) > 0 | \{N(s) : s < t\}] \\ &= \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{E}[N(t+h) - N(t) | \{N(s) : s < t\}] \\ &= \mathbb{E}[dN(t) | \{N(s) : s < t\}]\end{aligned}$$

ou en présence de covariables :

$$\lambda(t|\mathbf{Z}) = \mathbb{E}[dN(t) | \{N(s), \mathbf{Z}(s) : s < t\}]$$

Différentes approches (suite)

- **approche marginale** pour $N(\cdot)$: consiste à modéliser la fonction moyenne cumulée (FCM) définie par

$$M(t) = \mathbb{E}[N(t)]$$

ou sa dérivée notée $\mu(t) = M'(t)$ qui s'écrit aussi

$$\begin{aligned}\mu(t) = \mathbb{E}[dN(t)] &= \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{E}[N(t+h) - N(t)] \\ &= \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{P}[N(t+h) - N(t) > 0]\end{aligned}$$

souvent appelée fonction de taux d'occurrence (FTO). En présence de covariables \mathbf{Z} , on modélise

$$M(t|\mathbf{Z}) = \mathbb{E}[N(t)|\mathbf{Z}]$$

ou

$$\mu(t|\mathbf{Z}) = \mathbb{E}[dN(t)|\mathbf{Z}].$$

Différentes approches (fin)

- **approche basée sur les durées** : on peut modéliser les fonctions de répartition (en général, en l'absence de covariables) ou les intensités (en général, en présence de covariables). La fonction de répartition de $X^{[k]}$ est donnée par :

$$F^{[k]}(t) = \mathbb{P}[X^{[k]} \leq t]$$

L'intensité de $X^{[k]}$ est donnée par :

$$\lambda^{[k]}(t) = \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{P}[t \leq X^{[k]} \leq t + h | X^{[k]} \geq t]$$

ou en présence de covariables :

$$\lambda^{[k]}(t|\mathbf{Z}) = \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{P}[t \leq X^{[k]} \leq t + h | X^{[k]} \geq t, \{\mathbf{Z}(s) : s < t\}]$$

La fonction de répartition de $Y^{[k]}$ est donnée par :

$$\tilde{F}^{[k]}(t) = \mathbb{P}[Y^{[k]} \leq t]$$

L'intensité de $Y^{[k]}$ est donnée par :

$$\tilde{\lambda}^{[k]}(t) = \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{P}[t \leq Y^{[k]} \leq t + h | Y^{[k]} \geq t]$$

ou en présence de covariables :

$$\tilde{\lambda}^{[k]}(t|\mathbf{Z}) = \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{P}[t \leq Y^{[k]} \leq t + h | Y^{[k]} \geq t, \{\mathbf{Z}(s) : s < t\}]$$

Quelques recommandations

- Modéliser $N(t)$ est intéressant lorsqu'on s'attend à beaucoup de récurrences (crises d'asthme, crises d'épilepsie).
- Modéliser les $Y^{[k]}$ est intéressant lorsqu'on s'attend à peu de récurrences (infarctus).
- Modéliser les $X^{[k]}$ est délicat car si on ne prend pas compte l'ordre existant entre les instants de survenue successifs, il faut s'attendre à une perte d'efficacité considérable. Or, aucun modèle vraiment adapté à cela n'a encore été proposé.
- L'usage de l'intensité apparaît plus approprié pour des prédictions individuelles .
- L'usage de la FCM est plus pertinent pour des comparaisons de populations auquel cas la FCM fournit une interprétation plus directe pour l'identification des effets du traitement et pour la comparaison de facteurs de risque.

Quelques modèles sans décès

- approche marginale pour $N(\cdot)$:
 - sans covariables : estimateur non-paramétrique pour $M(\cdot)$: Lawless et Nadeau (1995), Wang et Chiang (2002)
 - avec covariables : modèle semi-paramétrique
 $M(t|\mathbf{Z}) = M_0(t)g(\beta' \cdot \mathbf{Z}(t))$ avec g connue et M_0 et β à estimer : Lawless et Nadeau (1995) avec l'hypothèse que $N(\cdot)$ est un processus de Poisson inhomogène
- approche conditionnelle pour $N(\cdot)$: avec covariables
 - modèle de type Cox : $\lambda(t|\mathbf{Z}) = \lambda_0(t) \exp(\beta' \cdot \mathbf{Z}(t))$: Andersen et Gill (1982)
 - modèle stratifié de Prentice, Williams et Peterson (1981) par rapport à une variable notée $s = \varphi(t, \{N(s), \mathbf{Z}(s)\})$:
 $\lambda(t|\mathbf{Z}) = \lambda_0^s(t) \exp(\beta'_s \cdot \mathbf{Z}(t))$ ou
 $\lambda(t|\mathbf{Z}) = \lambda_0^s(t - t^{[N(t)]}) \exp(\beta'_s \cdot \mathbf{Z}(t))$
- modèle avec covariable pour l'intensité de $X^{[k]}$ de Wei, Lin et Weissfeld (1989) : $\lambda^{[k]}(t|\mathbf{Z}) = \lambda_0^{[k]}(t) \exp(\beta' \cdot \mathbf{Z})$ ou
 $\lambda^{[k]}(t|\mathbf{Z}) = \lambda_0^{[k]}(t) \exp(\beta'_k \cdot \mathbf{Z})$

Quelques modèles avec décès

- approche marginale pour $N(\cdot)$:
 - sans covariables : estimateur non-paramétrique pour $M(\cdot)$: Lawless et Nadeau (1997), Ghosh et Lin (2000)
 - avec covariables : modèle semi-paramétrique $\mu(t|\mathbf{Z}) = \mu_0(t) \exp(\beta' \cdot \mathbf{Z}(t))$ avec μ_0 et β à estimer : Ghosh et Lin (2002)
- modèle (sans covariable) pour l'intensité de $X^{[k]}$ de Li et Lagakos (1997) : estimateur non-paramétrique
- estimateur non-paramétrique de la fonction de répartition des $Y^{[k]}$: Geffray (2006)
- modèle joint pour $(N(\cdot), D)$: approche "frailty"

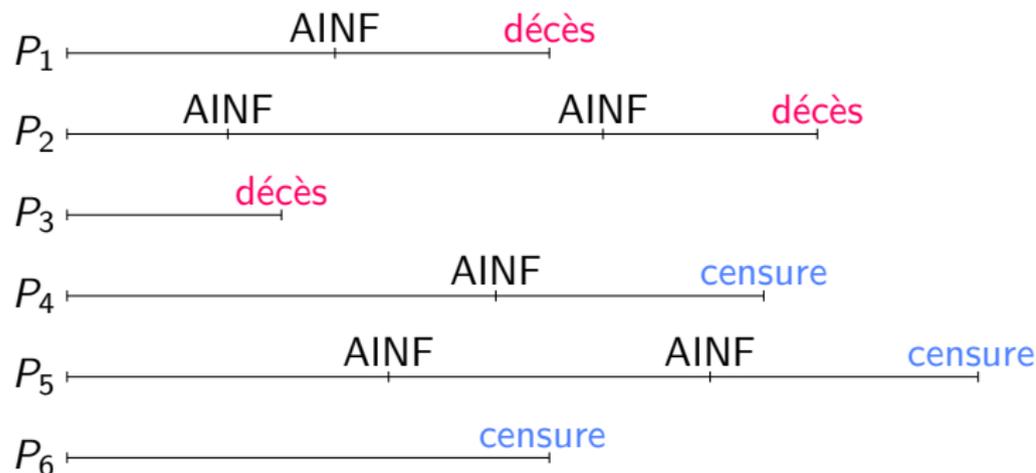
4^{ème} partie

Evènements récurrents (peu de récurrences)
terminés par un décès en présence de censure
indépendante

Motivation : l'étude CAPRIE

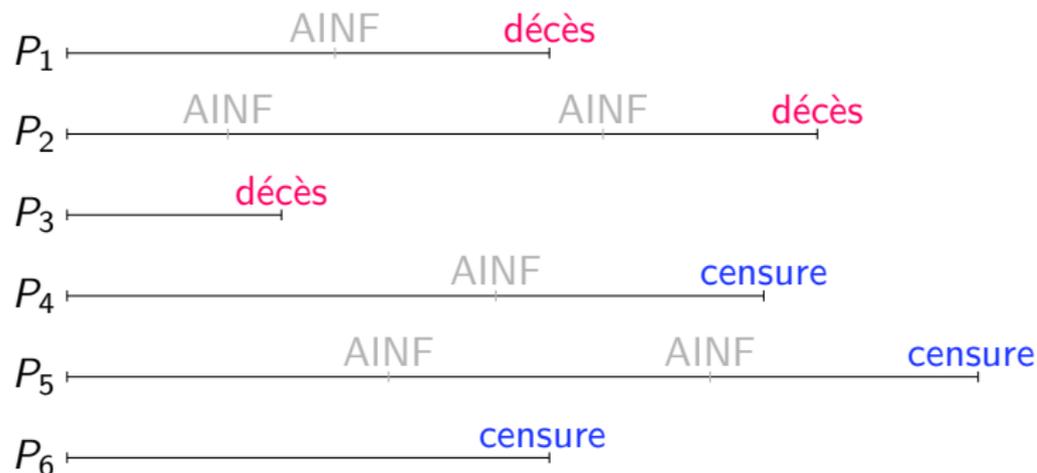
- CAPRIE : Clopidogrel versus Aspirin in Patients at Risk of Ischaemic Events.
- 20000 patients souffrant d'athérosclérose randomisés avec clopidogrel ou aspirine, durée de suivi allant de 1 à 3 ans.
- tous les accidents ischémiques non-fatals sont enregistrés puis le décès ou la censure
- maxi 8 évènements
- 90% d'observations censurées et 3% de décès en premier évènement, 70% d'observations censurées et 17% de décès parmi les patients restants à chaque occurrence ultérieure.
- Censure identique dans les deux groupes.

Exemple de données recueillies



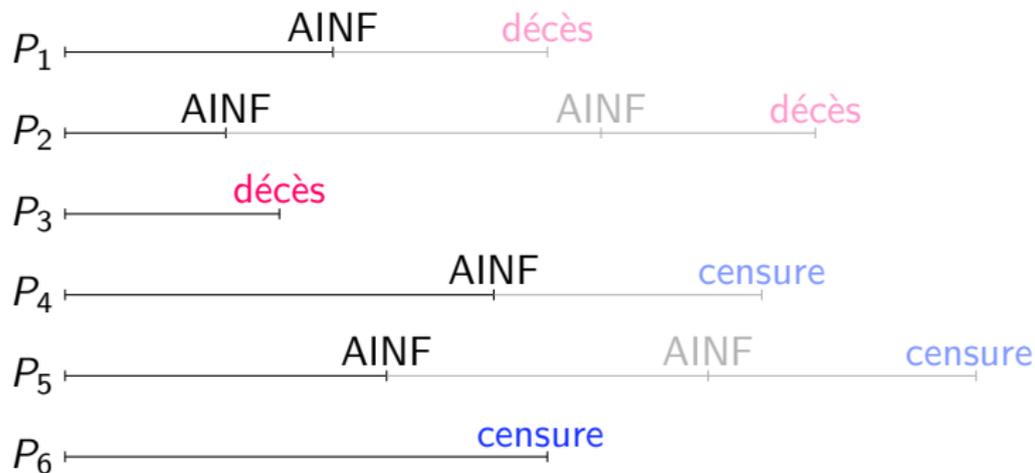
AINF = Accident Ischémique Non-Fatal

Une analyse classique : Kaplan-Meier pour la survie



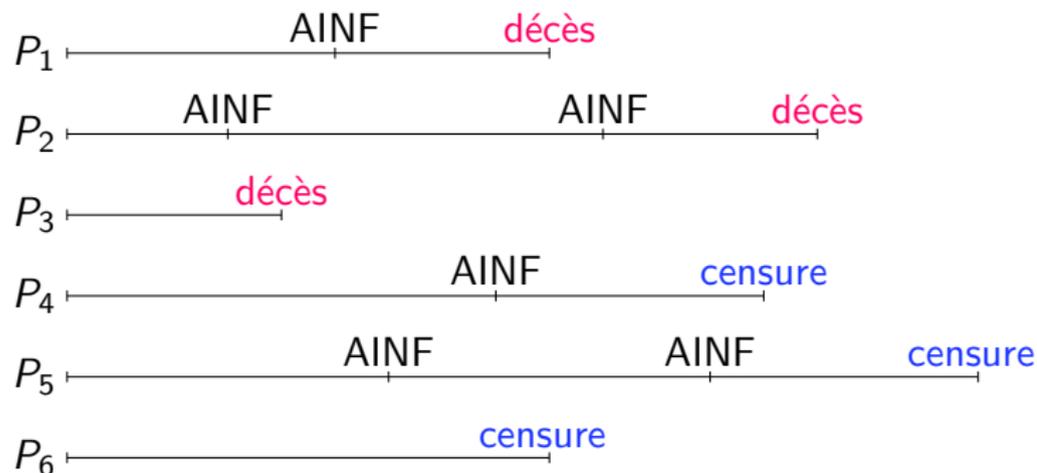
AINF = Accident Ischémique Non-Fatal

Autre quantité intéressante : durée jusqu'au premier évènement (*primary endpoint*)



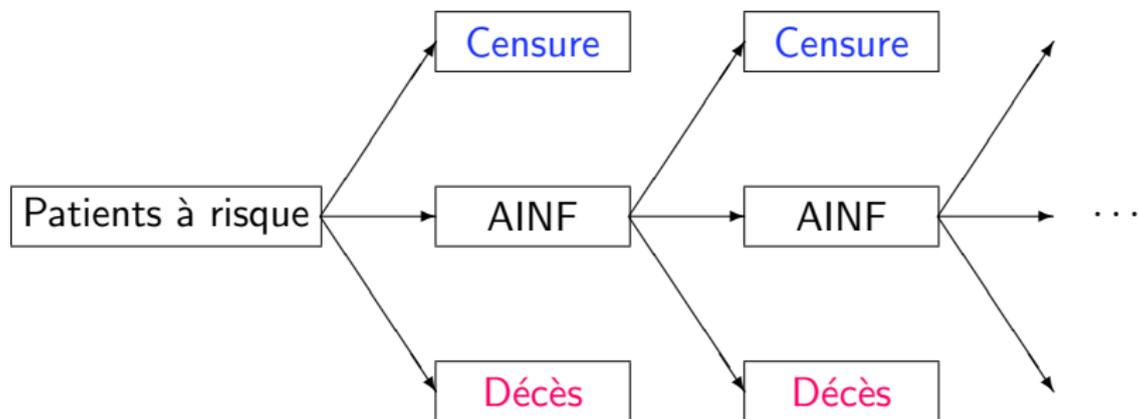
AINF = Accident Ischémique Non-Fatal

Rappel : exemple de données recueillies



AINF = Accident Ischémique Non-Fatal

Modèle



- N : nombre d'AINF subis par un patient.
- D : instant de décès du patient, de f.r. F_D .
- $X^{[k]}$: instant de survenue du $k^{\text{ème}}$ AI, pour $k = 1, \dots, N + 1$.
Ces v.a. satisfont $0 = X^{[0]} < X^{[1]} < \dots < X^{[N+1]}$.
- $\mathcal{C}^{[k]} = \begin{cases} 1 & \text{si le } k^{\text{ème}} \text{ AI n'est pas fatal,} \\ 2 & \text{si le } k^{\text{ème}} \text{ AI est fatal,} \end{cases}$
pour $k = 1, \dots, N + 1$.
- $Y^{[k]} = X^{[k]} - X^{[k-1]}$, pour $k = 1, \dots, N + 1$, durée entre le $(k - 1)^{\text{ème}}$ et le $k^{\text{ème}}$ AI. Ces v.a. sont corrélées.
- C : instant de censure, de f.r. G , **indépendante** des $X^{[k]}$ (donc aussi des $Y^{[k]}$) et des $\mathcal{C}^{[k]}$.

- Avec le premier accident ischémique :

$$F^{[1]}(t) = \mathbb{P} [Y^{[1]} \leq t],$$

$$F^{[1(1)]}(t) = \mathbb{P} [Y^{[1]} \leq t, \mathcal{C}^{[1]} = 1],$$

$$F^{[1(2)]}(t) = \mathbb{P} [Y^{[1]} \leq t, \mathcal{C}^{[1]} = 2].$$

- Avec les deux premiers accidents ischémiques (loi jointe) :

$$F^{[1(1),2]}(t_1, t_2) = \mathbb{P} [Y^{[1]} \leq t_1, \mathcal{C}^{[1]} = 1, Y^{[2]} \leq t_2],$$

$$F^{[1(1),2(1)]}(t_1, t_2) = \mathbb{P} [Y^{[1]} \leq t_1, \mathcal{C}^{[1]} = 1, Y^{[2]} \leq t_2, \mathcal{C}^{[2]} = 1],$$

$$F^{[1(1),2(2)]}(t_1, t_2) = \mathbb{P} [Y^{[1]} \leq t_1, \mathcal{C}^{[1]} = 1, Y^{[2]} \leq t_2, \mathcal{C}^{[2]} = 2].$$

Quantités d'intérêt (suite)

- Avec les deux premiers accidents ischémiques (loi conditionnelle " \leq ") :

$$F^{[2/1]}(t_2/t_1) = \mathbb{P} [Y^{[2]} \leq t_2 \mid Y^{[1]} \leq t_1, C^{[1]} = 1],$$

$$F^{[2(1)/1]}(t_2/t_1) = \mathbb{P} [Y^{[2]} \leq t_2, C^{[2]} = 1 \mid Y^{[1]} \leq t_1, C^{[1]} = 1],$$

$$F^{[2(2)/1]}(t_2/t_1) = \mathbb{P} [Y^{[2]} \leq t_2, C^{[2]} = 2 \mid Y^{[1]} \leq t_1, C^{[1]} = 1].$$

- Avec les deux premiers accidents ischémiques (loi conditionnelle " $=$ ") :

$$F^{[2/1]}(t_2/t_1) = \mathbb{P} [Y^{[2]} \leq t_2 \mid Y^{[1]} = t_1, C^{[1]} = 1],$$

$$F^{[2(1)/1]}(t_2/t_1) = \mathbb{P} [Y^{[2]} \leq t_2, C^{[2]} = 1 \mid Y^{[1]} = t_1, C^{[1]} = 1],$$

$$F^{[2(2)/1]}(t_2/t_1) = \mathbb{P} [Y^{[2]} \leq t_2, C^{[2]} = 2 \mid Y^{[1]} = t_1, C^{[1]} = 1].$$

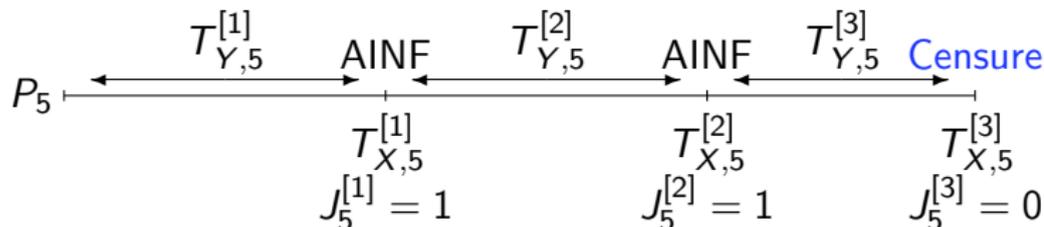
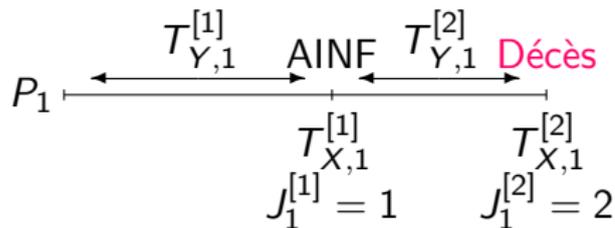
- etc...
- Les v.a. $Y^{[1]}$, $C^{[1]}$, $Y^{[2]}$ et $C^{[2]}$ ne sont pas observables à cause de la censure.

Variables observables

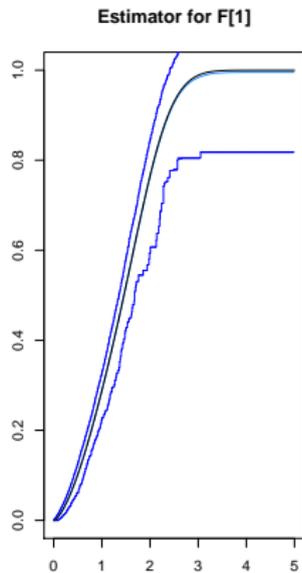
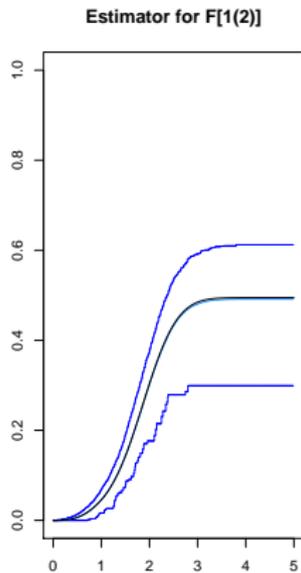
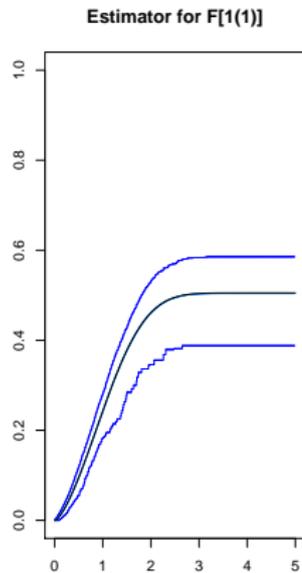
- K : nombre d'évènements observés ($K \leq N + 1$).
- $T_X^{[k]} = \min(X^{[k]}, C)$: instant de survenue du $k^{\text{ème}}$ évènement, pour $k = 1, \dots, K$. Ces v.a. satisfont $T_X^{[1]} < \dots < T_X^{[K]}$.
- $T_Y^{[k]} = T_X^{[k]} - T_X^{[k-1]}$: durée entre le $(k - 1)^{\text{ème}}$ et le $k^{\text{ème}}$ évènement, pour $k = 1, \dots, K$. Ces v.a. sont corrélées.
- Pour $k = 1, \dots, K$,
 $J^{[k]} = \begin{cases} 0 & \text{si le } k^{\text{ème}} \text{ évènement est une censure,} \\ 1 & \text{si le } k^{\text{ème}} \text{ évènement est un AINF,} \\ 2 & \text{si le } k^{\text{ème}} \text{ évènement est un décès.} \end{cases}$

Visualisation des observations

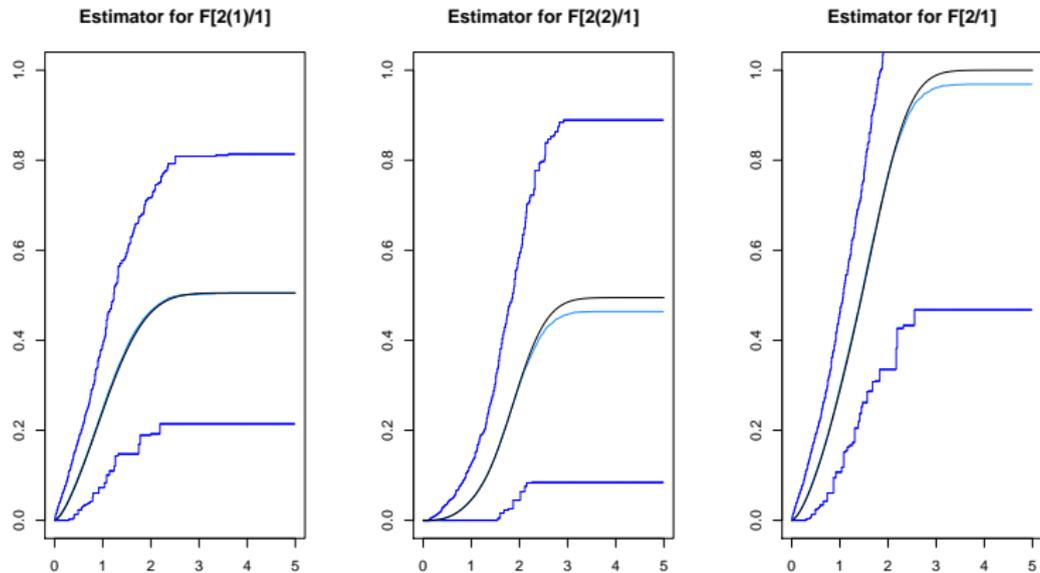
Echantillon i.i.d. pour $i = 1, \dots, n$: $\left(T_{Y,i}^{[k]}, J_i^{[k]} \right)_{k=1, \dots, K_i}$.



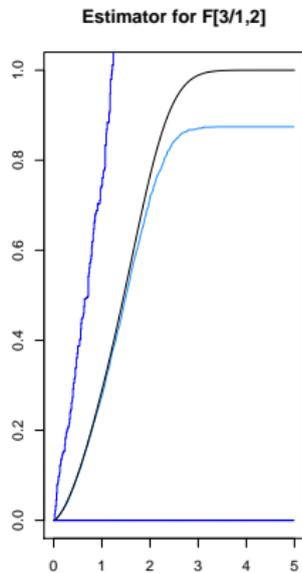
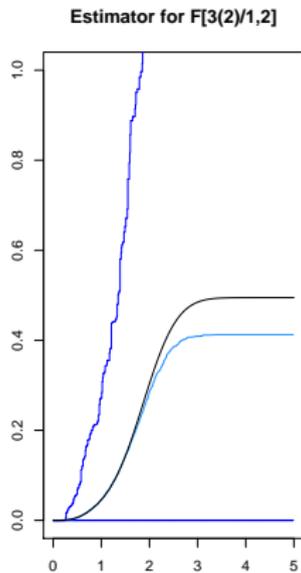
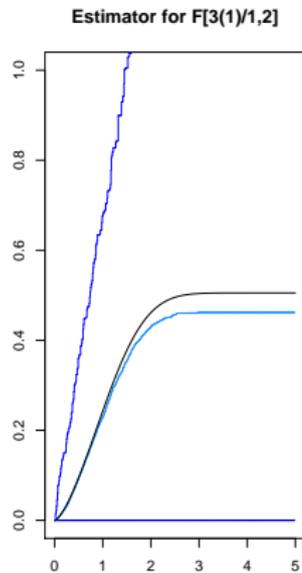
Simulations : première durée



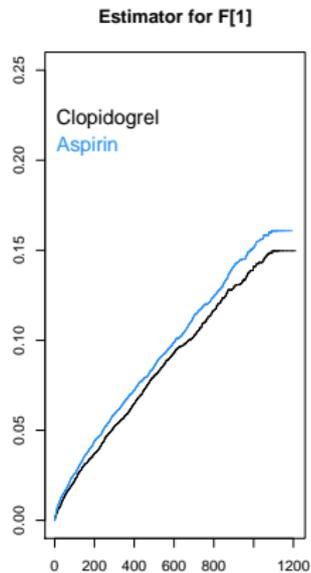
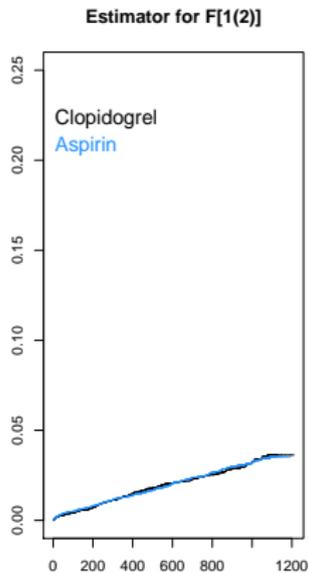
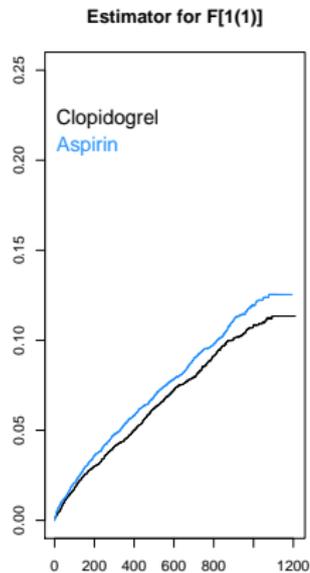
Simulations : seconde durée



Simulations : troisième durée

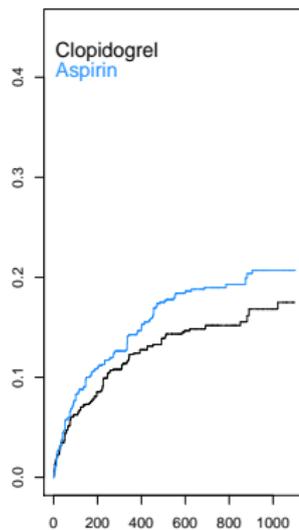


CAPRIE : première durée

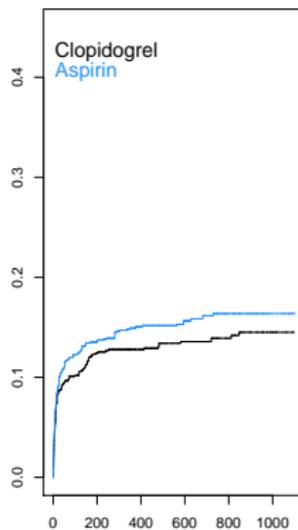


CAPRIE : seconde durée

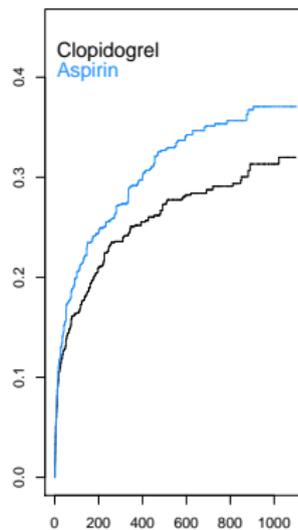
Estimator for $F[2(1)/1]$



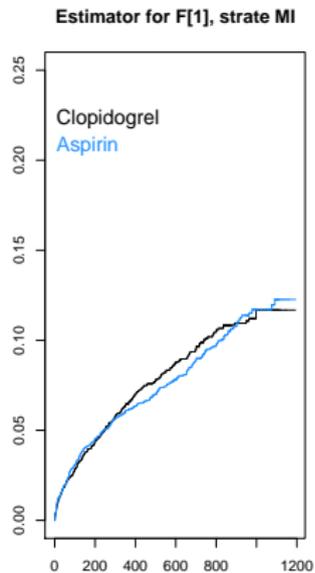
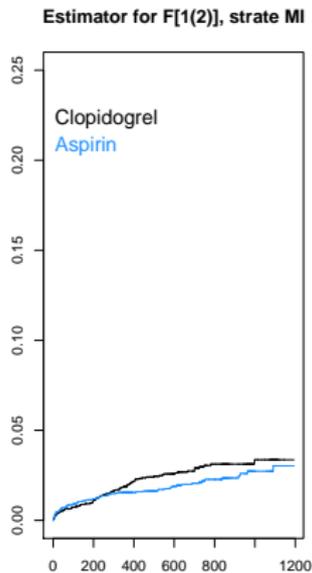
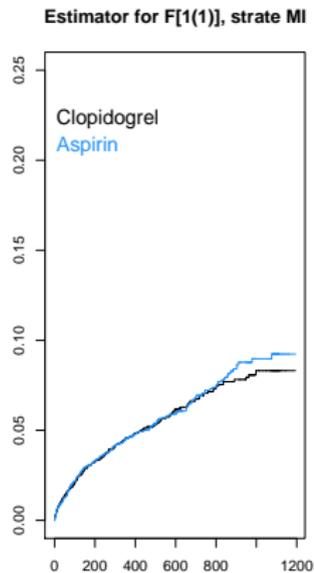
Estimator for $F[2(2)/1]$



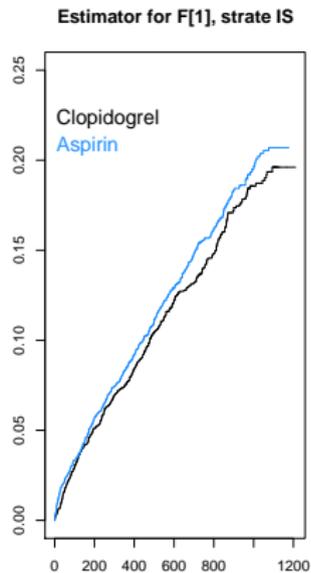
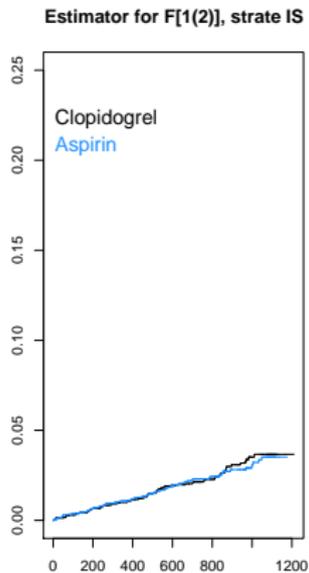
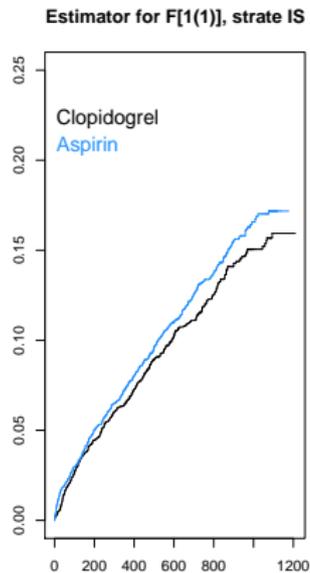
Estimator for $F[2/1]$



CAPRIE : première durée, strate MI

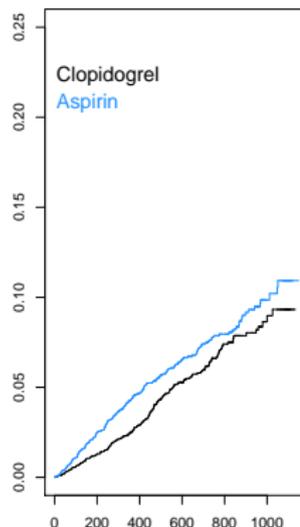


CAPRIE : première durée, strate IS

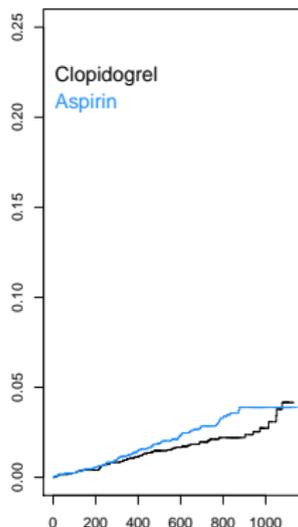


CAPRIE : première durée, strate PAD

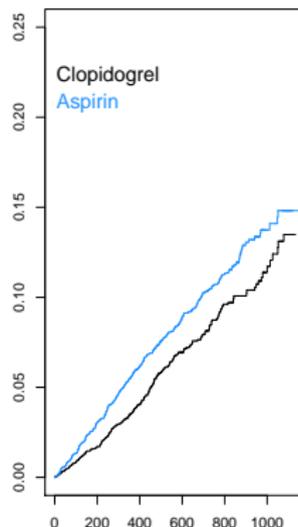
Estimator for $F[1(1)]$, strate PAD



Estimator for $F[1(2)]$, strate PAD

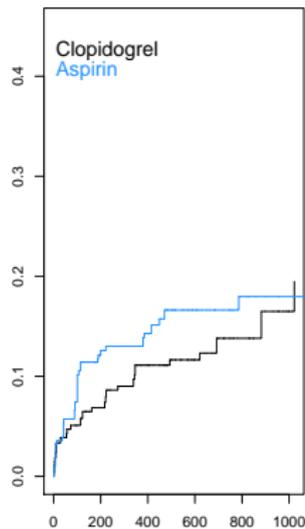


Estimator for $F[1]$, strate PAD

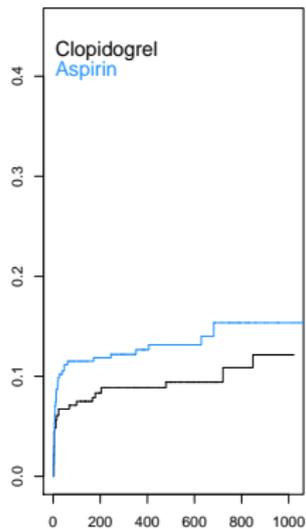


CAPRIE : seconde durée, strate MI

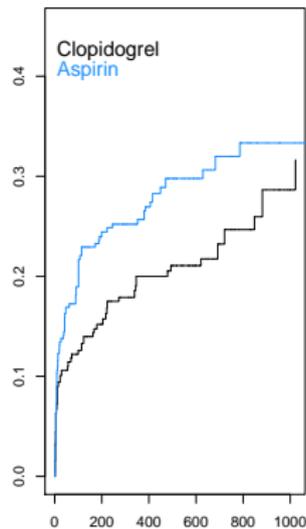
Estimator for F[2(1)/1], strate MI



Estimator for F[2(2)/1], strate MI

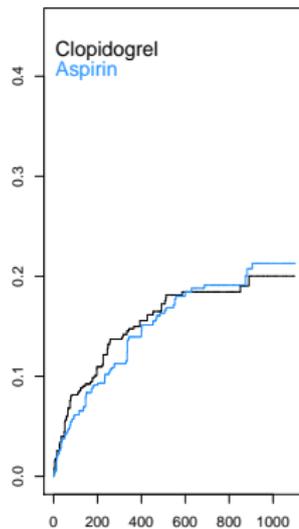


Estimator for F[2/1], strate MI

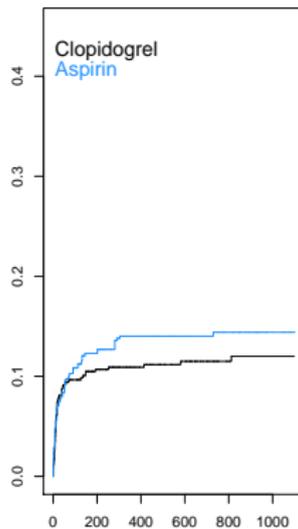


CAPRIE : seconde durée, strate IS

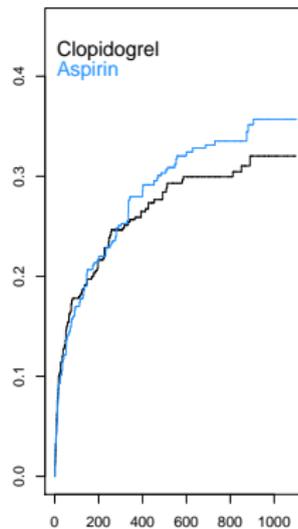
Estimator for F[2(1)/1], strate IS



Estimator for F[2(2)/1], strate IS

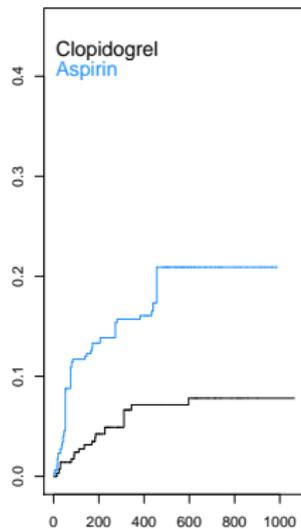


Estimator for F[2/1], strate IS

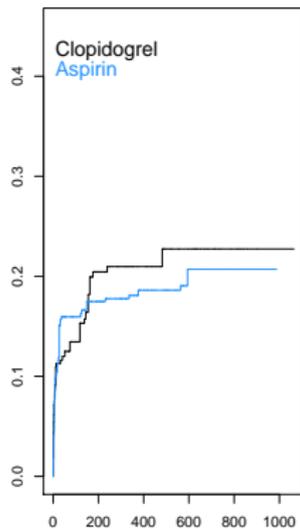


CAPRIE : seconde durée, strate PAD

Estimator for $F[2(1)/1]$, strate PAD



Estimator for $F[2(2)/1]$, strate PAD



Estimator for $F[2/1]$, strate PAD

