



**HAL**  
open science

# Approches statistiques asymptotique de la théorie de la décision et application quant à l'utilisation de l'incertain d'un modèle

Lauzetto Durand

► **To cite this version:**

Lauzetto Durand. Approches statistiques asymptotique de la théorie de la décision et application quant à l'utilisation de l'incertain d'un modèle. *Méthodologie [stat.ME]*. 2013. dumas-00858955

**HAL Id: dumas-00858955**

**<https://dumas.ccsd.cnrs.fr/dumas-00858955v1>**

Submitted on 6 Dec 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université de Strasbourg

UFR Mathématique-Informatique

Année Universitaire 2012-2013

Master Statistique et Applications

**Approches statistiques asymptotiques de la théorie de la  
décision et application quant à l'utilisation de l'incertain  
d'un modèle.**

Présenté par : Durand Lauzetto

Date de la soutenance : 29 Août 2013

Maître de stage : M. Dominik Pfluger

Entreprise d'accueil : Novendi SAS

Adresse : 15 rue des Frères Lumière.  
Mulhouse 68200



# Remerciements

Avant de débiter ce mémoire je tiens à remercier M.Dominik H. Pfluger pour m'avoir proposé un sujet si intéressant et accueillis au sein de son entreprise.

Je remercie l'ensemble des élèves de ma promotion pour la bonne ambiance apportée en cours ces deux dernières années et je remercie nos professeurs pour certaines remarques qui ne prennent tout leur sens que plus tard.

Je remercie celui qui se reconnaîtra.

Je remercie le hasard.

Aux premiers, au bref et non-oublié compagnon de route et à la futur copie non égale.

# Notations

i.i.d indépendant et identiquement distribué (pour un échantillon).

$\mathbf{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$  un caractère en gras représente un vecteur.

$\mathbf{u}' = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}' = (u_1, \dots, u_n)$  une apostrophe à droite d'un vecteur représente la transposée.

$M_{n,p}$ : ensemble des matrices à n lignes et p colonnes.

$\mathbb{R}$  : ensemble des réels.

$\mathbb{R}^+$ : ensemble des réels positifs.

$\pi(\theta|x)$  : loi à posteriori sur  $\theta$ .

$\pi(\theta)$  : loi à priori sur  $\theta$ .

$\in$  : symbole d'appartenance.

$\mathbb{P}$  : probabilité

$\mathcal{B}(E)$  : ensemble des boréliens de E.

E: espérance.

$\hat{\beta}$  : un chapeau représente un estimateur.

$\text{Card}\{E\}$  : nombre d'éléments que l'ensemble E possède.

## Table des matières

Remerciements .....	3
Introduction.....	6
Analyse de survie.....	7
1.1 Fonctions de survie et de hasard .....	7
1.2 Censure et troncature.....	9
1.3 Différentes approches.....	10
Modèles semi-paramétriques .....	13
2.1 Modèle de Cox .....	13
2.2 Modèle au hasard constant par morceaux.....	19
Statistique Bayésienne .....	20
3.1 Présentation générale.....	20
3.3 Non-paramétrique.....	24
3.4 Algorithmes de simulations .....	25
Applications et discussion .....	27
4.1 Exemple.....	27
4.2 Calcul du nombre de sujets .....	27
References.....	33

# Introduction

Des problèmes statistiques dans la conduite d'essais cliniques sont fréquents. Le choix de la taille de l'échantillon, l'interprétation des résultats, la validité d'un modèle, l'étendue des outils statistiques et des règles d'arrêt peuvent compliquer les analyses statistiques.

La mise en marché d'un médicament passe par plusieurs études différentes (ou essais cliniques) avec chacune différents objectifs. Dans notre cas nous nous intéressons à une étude de phase II avec différents objectifs classés par ordre d'importance. Les études de phase II sont considérées comme des intermédiaires sensibles dans le sens où l'on ne possède généralement pas assez de sujets pour conclure d'une manière définitive, mais peuvent directement stopper le développement d'un nouveau médicament.

Nous reprenons la méthodologie statistique pour répondre à ces différents objectifs, en particulier deux critères de décision ainsi qu'une approche Bayésienne sont proposés. Des simulations nous ont permis de faire un contrôle de cohérence de ces critères. Par la suite une revue de la littérature nous a permis d'expliquer théoriquement cette approche, ainsi que d'adapter nos outils statistiques à l'approche Bayésienne pour proposer une modélisation du phénomène observé.

Dans la partie 1 nous rappelons la base de l'analyse de survie, dans la partie 2 nous présenterons deux modèles semi-paramétriques, dans la partie 3 nous présenterons quelques principes de l'analyse bayésienne et dans la partie 4 nous expliquerons l'approche recommandée.

# Partie 1

## Analyse de survie

Dans cette partie nous allons présenter et introduire les définitions et notations de base de l'analyse de survie. On peut trouver différentes notations dans les articles et livres pour représenter nos prédicteurs, notre variable de survie et l'indicateur de censure. Nous utiliserons les notations jugées comme les plus adaptées.

### 1.1 Fonctions de survie et de hasard

**Définition 11.1** Soit  $T$  une variable aléatoire positive et continue représentant la « durée de vie » d'un individu de notre population. Sa **fonction de répartition** est

$$F(t) = \mathbb{P}(T \leq t) = \int_0^t f(u) du .$$

où  $f$  est la densité de probabilité de  $T$ . La fonction de répartition représente la probabilité de décès de l'individu avant l'instant  $t$  (i.e entre 0 et  $t$ ).

Remarques : (i) Le vocabulaire "durée de vie", "décès de l'individu" est une convention de l'analyse de survie bien que la théorie s'applique dans divers domaines où l'événement d'intérêt peut représenter par exemple la durée de latence d'un patient séropositif, la durée de chômage pour les licenciés, etc. (ii) On s'est placé dans le cas où  $T$  admet une densité par rapport à la mesure de Lebesgue sur  $\mathbb{R}^+$  ; toutes les fonctions définies dans cette partie sont définies sur  $[0, +\infty[$ . (iii) De même, sauf si précisé on considèrera qu'un individu ne peut "mourir" qu'une seule fois.

**Définition 11.2** La **fonction de survie** est définie par

$$\begin{aligned} S(t) &= \mathbf{1} - F(t) \\ &= \mathbb{P}(T > t) . \end{aligned}$$

et représente la probabilité de survie au-delà de l'instant  $t$ .



Remarquons que  $S(t)$  est une fonction monotone décroissante ainsi que  $S(0) = 1$  et  $\lim_{t \rightarrow \infty} S(t) = 0$ .

**Définition 11.3** La *fonction de hasard* est définie par

$$h(t) = \lim_{\Delta t \rightarrow 0^+} \frac{\mathbb{P}(t \leq T < t + \Delta t | T > t)}{\Delta t} = \frac{f(t)}{S(t)}.$$

$h(t)\Delta t$  représente la probabilité approchée qu'un individu décède dans  $[t, t+\Delta t[$  en sachant qu'il a survécu jusqu'en  $t$ .

**Définition 11.4** La *fonction de hasard cumulé* est donnée par

$$H(t) = \int_0^t h(u)du = -\log(S(t)).$$

Remarque :  $h(t)$  est aussi appelée la **fonction de hasard instantané**, **fonction de risque**, ou **fonction de risque instantané**.

**Proposition 11.1.** Les fonctions  $S(t)$ ,  $f(t)$ ,  $F(t)$ ,  $H(t)$  et  $h(t)$  sont en bijection les unes des autres. Par conséquent la distribution de probabilité de  $T$  est entièrement définie par une de ces fonctions.

En effet : 
$$f(t) = \frac{d}{dt} F(t) = -\frac{d}{dt} S(t)$$

Et donc 
$$h(t) = -\left(\frac{d}{dt} S(t)\right) \frac{1}{S(t)} = -\frac{d}{dt} \log S(t)$$

En intégrant des deux côtés puis en passant à l'exponentielle nous avons :

$$S(t) = \exp\left(-\int_0^t h(u)du\right) = e^{-H(t)}$$

Et donc  $f(t) = h(t) \exp\left(-\int_0^t h(u)du\right)$ . Comme on a  $S(t) = e^{-H(t)}$  et  $\lim_{t \rightarrow \infty} S(t) = 0$ , on en déduit  $\lim_{t \rightarrow \infty} H(t) = \infty$ .

## 1.2 Censure et troncature

**Définition 12.1** Une observation est **censurée** lorsque la durée exacte est inconnue mais que l'on dispose d'une information partielle sous la forme d'un minimum ou d'un maximum ou les deux. Si l'on observe la **variable de censure**  $C$ , et non  $T$ , et que l'on sait que  $T > C$  (respectivement  $T < C$ ,  $C_1 < T < C_2$ ), on dit que  $T$  est **censurée à droite** (respectivement à **gauche**, par **intervalle**).

**Définition 12.2** Un jeu de données est sujet à la **troncature** lorsque le caractère incomplet du jeu de données est dû à un processus de sélection systématique inhérent à la planification expérimentale. Lorsqu'en raison de la planification expérimentale on recueille des observations seulement sur les individus pour lesquels l'évènement d'intérêt se produit au-delà d'un seuil aléatoire ou fixe (respectivement avant, au-delà d'un premier seuil mais avant un deuxième seuil), il s'agit de données **tronquées à gauche** (respectivement à **droite**, par **intervalle**).

La troncature se distingue de la censure dans le sens où la troncature élimine une partie de nos observations de notre échantillon, tandis que pour la censure l'observation est présente dans l'échantillon mais incomplète.

On parle de censure **non-informative** ou censure **aléatoire** lorsque la censure est indépendante de l'évènement étudié.

Pour la suite nous nous placerons dans le cas de censure aléatoire droite, sans troncature.

**Définition 12.3** Prenons un échantillon de durées de survie avec censure aléatoire droite, dans ce cas nous observons soit  $T_i$  de manière exacte, soit  $C_i$  en sachant que  $T_i > C_i$ . Dans tous les cas nous observons  $Y_i = \min(T_i, C_i)$  et

l'**indicateur de censure**  $\delta_i = \begin{cases} 1 & \text{si } T_i \leq C_i \\ 0 & \text{si } T_i > C_i \end{cases}$ .

### 1.3 Différentes approches

Plusieurs modélisations sont possibles avec chacune avantages et inconvénients, présentons brièvement sans exhaustivité des avantages et inconvénients.

Pour l'approche **paramétrique** on fait l'hypothèse que la loi de probabilité réelle de la population appartient à une classe particulière de lois, qui dépendent d'un certain nombre fini de paramètres. L'avantage de cette approche est la relative facilité d'analyse en termes d'estimation des paramètres, de construction de tests et d'intervalle de confiance. L'inconvénient est une possible inadéquation entre le phénomène observé et le modèle choisi.

Pour l'approche **non-paramétrique** on fait un minimum d'hypothèses sur la densité réelle des observations. L'avantage est la capacité à s'ajuster à quasiment tous les jeux de données. L'inconvénient est la taille de l'échantillon nécessaire ainsi que l'estimation des paramètres qui est plus compliquée.

L'approche **semi-paramétrique** est une sorte de compromis entre les deux, le paramètre à estimer a une composante appartenant à un espace de dimension finie et une composante appartenant à un espace de dimension infinie. On est souvent seulement intéresser à estimé seulement le paramètre de dimension finie.

Exemple 1: Modèle paramétrique exponentiel.

Supposons qu'on ait un échantillon (i.i.d) de taille  $n$  de durée de survie,  $T$  est notre variable d'intérêt et on suppose que  $T$  suit une loi exponentielle de paramètre  $\lambda$ . Avec les notations introduites précédemment nos données sont représentées par  $D = (n, \mathbf{y}, \boldsymbol{\delta})$  où  $\mathbf{y} = (y_1, y_2, \dots, y_n)'$ ,  $\boldsymbol{\delta} = (\delta_1, \delta_2, \dots, \delta_n)'$ .

En cas d'observation non censurée on observe  $y_i = t_i$  et  $\delta_i = 1$ , la contribution à la vraisemblance est  $f(t_i|\lambda) = f(y_i|\lambda) = \lambda * \exp(-\lambda y_i)$ , en cas de censure droite on observe  $y_i = c_i$  en sachant que  $t_i > y_i$ , l'information apportée par l'observation  $i$  à la vraisemblance est  $\mathbb{P}(t_i > y_i|\lambda) = S(y_i|\lambda) = \exp(-\lambda y_i)$ . La vraisemblance s'écrit :

$$\begin{aligned} L(\lambda|D) &= \prod_{i=1}^n f(y_i|\lambda)^{\delta_i} * S(y_i|\lambda)^{1-\delta_i} = \prod_{i=1}^n \lambda \exp(-\lambda y_i)^{\delta_i} * \exp(-\lambda y_i)^{1-\delta_i} \\ &= \lambda^n \exp(-\lambda \sum_{i=1}^n y_i) \end{aligned}$$

où  $d = \sum_{i=1}^n \delta_i$ .

Si on a des prédicteurs nos données sont  $D = (n, \mathbf{y}, \boldsymbol{\delta}, X)$ , avec  $X \in M_{n,p}$ , notons  $\mathbf{x}_i$  le vecteur de taille  $p \times 1$  représentant les prédicteurs de l'individu  $i$ ,  $\boldsymbol{\beta}$  le vecteur  $p \times 1$  des coefficients de régression. On peut introduire nos prédicteurs par le paramètre  $\lambda$  en posant  $\lambda_i = r(\mathbf{x}_i' \boldsymbol{\beta})$  avec  $r(\cdot)$  une fonction choisie.

$$L(\lambda|D) = \prod_{i=1}^n f(y_i|\lambda)^{\delta_i} * S(y_i|\lambda)^{1-\delta_i} = \prod_{i=1}^n r(x_i) \exp(-r(\mathbf{x}_i)y_i)^{\delta_i} * \exp(-\lambda y_i)^{1-\delta_i}$$

Remarque : La loi exponentielle possède la propriété d'être « sans mémoire »

C'est à dire que  $\forall (s, t) \in \mathbb{R}^2$ ,  $\mathbb{P}(T > s + t | T > t) = \mathbb{P}(T > s)$ , (i.e on fait l'hypothèse qu'un individu a le même risque instantané de mourir à chaque instant).

Exemple 2: Modèle paramétrique de Weibull avec deux paramètres.

La densité de la loi de Weibull dans sa forme la plus générale est donnée par:

$$f(y|\alpha, \gamma, \mu) = \begin{cases} \frac{\alpha}{\gamma} * (\frac{t-\mu}{\gamma})^{\alpha-1} \exp\left(-(\frac{t-\mu}{\gamma})^\alpha\right), & y > 0, \alpha > 0, \gamma > 0, \mu \geq 0 \\ 0 & \text{sinon} \end{cases}$$

avec  $\alpha$  paramètre de forme,  $\gamma$  paramètre d'échelle et  $\mu$  paramètre de position.

En posant  $\mu = 0$  on a la densité de la loi de Weibull à deux

paramètres.  $f(y|\alpha, \gamma, \mu) = \begin{cases} \frac{\alpha}{\gamma} * (\frac{t}{\gamma})^{\alpha-1} \exp\left(-(\frac{t}{\gamma})^\alpha\right), & y > 0, \alpha > 0, \gamma > 0, \mu \geq 0 \\ 0 & \text{sinon} \end{cases}$

avec  $\alpha$  paramètre de forme,  $\gamma$  paramètre d'échelle. Sa fonction de répartition est donc  $F(y|\alpha, \gamma) = 1 - \exp[-(\frac{t}{\gamma})^\alpha]$ , c'est la forme utilisé dans R ou SAS. On trouve différentes écritures dans littérature de ce modèle suivant les applications en posant  $\lambda = 1/\gamma$ ,  $\lambda = \gamma^\alpha$ ,  $\lambda = (1/\gamma)^\alpha$  ou encore  $\lambda = \log[(\frac{1}{\gamma})^\alpha]$  (le lecteur pourra trouver plus de détails dans le livre de Murthy *et al.*[13]).

En prenant la dernière notation on a :

$$f(y|\alpha, \lambda) = \begin{cases} \alpha y^{\alpha-1} \exp(\lambda - \exp(\lambda) y^\alpha), & y > 0, \alpha > 0, \lambda \in \mathbb{R} \\ 0 & \text{sinon} \end{cases}$$

Et donc  $S(y|\alpha, \lambda) = \exp(-\exp(\lambda) y^\alpha)$

La vraisemblance s'écrit :

$$L(\lambda|D) = \prod_{i=1}^n f(y_i|\alpha, \lambda)^{\delta_i} * S(y_i|\alpha, \lambda)^{1-\delta_i} = \prod_{i=1}^n \dots = \prod_{i=1}^n \alpha$$

$$= \alpha^d \exp \left\{ d * \lambda + \sum_{i=1}^n (\delta_i(\alpha - 1) \log(y_i) - \exp(\lambda)y_i^\alpha) \right\}$$

On peut introduire une régression par le paramètre  $\lambda$  en posant  $\lambda_i = \mathbf{x}_i' \boldsymbol{\beta}$ .

Remarque: Pour un modèle de Weibull on fait l'hypothèse que le hasard varie d'une manière monotone, en posant  $\alpha = 1$  on se ramène au modèle exponentiel, la loi de Weibull peut donc être vue comme une généralisation de la loi exponentielle.

## Partie 2

# Modèles semi-paramétriques

## 2.1 Modèle de Cox

Introduit par Cox (1972) puis communément appelé en son nom, on trouve aussi l'appellation modèle de hasards proportionnels, modèle à risques proportionnels ou encore modèle à risques proportionnels de Cox. On garde les notations introduites au chapitre précédent.

**Définition 21.1** *On dit que  $T$  satisfait un modèle à **risques proportionnels** lorsque l'on suppose que la fonction de risque est définie par :*

$$\forall t \geq 0, \forall \mathbf{x} \quad h(t|\mathbf{x}) = h_0(t) * \exp\{G(\mathbf{x}, \boldsymbol{\beta})\},$$

*le plus souvent utilisé en supposant un prédicteur linéaire:*

$$\forall t \geq 0, \forall \mathbf{x} \quad h(t|\mathbf{x}) = h_0(t) * \exp(\mathbf{x}'\boldsymbol{\beta})$$

où  $h_0(t)$  est une fonction représentant la fonction de hasard lorsque  $\mathbf{x} = 0$ .

Cette écriture implique l'**hypothèse des risques proportionnels** c'est-à-dire que le rapport des hasards pour deux individus est constant dans le temps et dépend uniquement de la différence entre leur prédicteur. En effet pour deux individus  $i$  et  $j$  on a :

$$\frac{h_0(t) * \exp(\mathbf{x}_i' \boldsymbol{\beta})}{h_0(t) * \exp(\mathbf{x}_j' \boldsymbol{\beta})} = \frac{\exp(\mathbf{x}_i' \boldsymbol{\beta})}{\exp(\mathbf{x}_j' \boldsymbol{\beta})} = \exp((\mathbf{x}_i' - \mathbf{x}_j') \boldsymbol{\beta})$$

qui ne dépend pas du temps. Cette hypothèse se vérifie pour chaque variable explicative de plusieurs manières soit graphiquement ou avec des tests statistiques. On peut par exemple tracer l'estimateur de Kaplan-Meier sur l'échelle  $\log(-\log())$ , en effet avec l'hypothèse des risques proportionnels:

$$\log(-\log[S_i(t)]) = \log[H_0(t)] - \mathbf{x}_i' \boldsymbol{\beta}$$

Et donc sur cette échelle les courbe pour les différents niveaux d'un prédicteur devrait être "environ" parallèle.

(voir [14] pour plus d'exemples de tests sur l'hypothèse des risques proportionnels.)

Remarque: Dans un cadre plus général on peut supposer que les variables explicatives ou que le vecteur des coefficients dépend du temps, par exemple pour modéliser le cas où un médicament perd de son effet dans le temps. Le lecteur pourra se référer au livre de Therneau et Grambsch (2000) pour plus de détails sur le modèle de Cox. Ici on se restreint au cas où les variables explicatives et le vecteur des coefficients est fixe dans le temps.

**Définition 21.2** *Le rapport des hasards (ou Hazard Ratio en anglais) est défini mathématiquement par son nom (implicitement dit le rapport des fonctions de hasard pour deux populations différentes, ou deux sous parties d'une même population). On notera HR.*

Remarque: Le terme rapport des hasards est souvent employé en même temps qu'on fait l'hypothèse des risques proportionnels, et est donc supposé constant dans le temps c'est aussi dans ce cadre que nous sommes et donc nous le définissons à l'intérieur de la partie sur le modèle de Cox bien qu'on aurait pu le définir dans le chapitre 1.

**Propriété 21.1** *En supposant le modèle de Cox vrai on a la relation*

$$S(t) = \exp\left(-\int_0^t h(u)du\right) = \exp\left(-\exp(\mathbf{x}'\boldsymbol{\beta}) \int_0^t h_0(u)du\right) = \exp\left[-\int_0^t h_0(u)du\right]^{\exp(\mathbf{x}'\boldsymbol{\beta})}.$$

Exemple: Sans perte de généralité on considère un simple modèle de Cox avec une variable qualitative de type 0 ou 1 pour évaluer l'efficacité de deux traitements (A et B) le rapport des hasards s'écrit  $\frac{h_A(t|\mathbf{x})}{h_B(t|\mathbf{x})} = \frac{h_0(t)*\exp(\boldsymbol{\beta})}{h_0(t)*\exp(0)} = \exp(\boldsymbol{\beta})$ , d'où le passage fréquent du rapport des hasards à son logarithme.

Le rapport des hasards permet d'interpréter l'effet d'un médicament en termes de risque relatif. Si on suppose que l'effet d'un des deux traitements sur la population est connu la relation entre rapport des hasards et durée de survie est  $S_A(t|D) = S_B(t|D)^{HR}$  ce qui permet de répondre à certaines questions en terme de durée de survie de type « Si on a la proportion  $p_1$  de patients qui survivent jusqu'au temps t avec le traitement A, quelle sera la proportion  $p_2$  de patients qui survivront jusqu'au temps t avec le traitement B ? », et d'interpréter le rapport des hasards en terme de *différence clinique*, c'est-à-dire un écart qu'on considère

avoir une importance clinique, par exemple 20% en plus de bénéfice. Si  $\beta = 0$ , équivalent à  $HR=1$ , le traitement n'a pas d'effet.

Remarque : En fonction de la façon de coder notre variable qualitative un rapport des hasards supérieur à 1 signifie que le nouveau ou l'ancien traitement est meilleur.

### Estimation des paramètres du modèle de Cox:

L'estimation des paramètres se fait par maximum de vraisemblance, la vraisemblance du modèle de Cox est:

$$\begin{aligned} L(\beta|D) &= \prod_{i=1}^n f(y_i)^{\delta_i} * S(y_i)^{1-\delta_i} = \prod_{i=1}^n [h(y_i)S(y_i)]^{\delta_i} * S(y_i)^{1-\delta_i} \\ &= \prod_{i=1}^n [h_0(y_i)\exp(x'\beta)S(y_i)]^{\delta_i} * S(y_i)^{1-\delta_i}. \end{aligned}$$

En passant à la log-vraisemblance:

$$l(\beta|D) = \sum_{i=1}^n \delta_i \log(h_0(y_i)) + \delta_i x' \beta + \exp(x' \beta) * \log(S_0(y_i))$$

On ne peut pas maximiser cette vraisemblance sans spécifier  $h_0(t)$ .

Supposons qu'on ait un jeu de donnée avec  $n$  sujets, en ordonnant les temps de survie  $y_{(1)} < y_{(2)} < \dots < y_{(n)}$ . Notons  $R_{(j)}$  l'ensemble des individus à risque au temps  $y_j$  (i.e l'ensemble des individus toujours en vie au temps  $y_j$ ). Cox (1972) a proposé une vraisemblance conditionnelle ("dans le sens conditionnelle à l'ensemble  $R_{(j)}$ ") pour estimer  $\beta$  en laissant  $h_0(.)$  non spécifié :

$$L(\beta|D) = \prod_{i=1}^n \left[ \frac{\exp(x_j' \beta)}{\sum_{k \in R_{(i)}} \exp(x_k' \beta)} \right]^{\delta_i}$$

On peut voir dans Breuils, C. (2003) que cette vraisemblance peut aussi être considérée comme une vraisemblance profilée et un exemple d'heuristique pour le calcul de la vraisemblance conditionnelle que nous modifions pour expliquer



cette vraisemblance, supposons que nous sommes en  $t_i$  et qu'un seul individu puisse mourir en  $[t_i ; t_i + \Delta t[$  nous raisonnons conditionnellement à  $R_{(i)}$  :

$$\begin{aligned} & \mathbb{P}(\text{l'individu } i \text{ soit mort en } [t_i ; t_i + \Delta t[ \mid \text{une mort en } [t_i ; t_i + \Delta t[) \\ &= \frac{\mathbb{P}(\text{une mort en } [t_i ; t_i + \Delta t[ \cap i \text{ meure en } [t_i ; t_i + \Delta t[)}{\mathbb{P}(\text{une mort en } [t_i ; t_i + \Delta t[)} \\ &= \frac{\mathbb{P}(i \text{ meure en } [t_i ; t_i + \Delta t[)}{\mathbb{P}(\text{une mort en } [t_i ; t_i + \Delta t[)} \approx \frac{h_i(t)\Delta t}{\sum_{j \in R_{(i)}} h_j(t)\Delta t} = \frac{\exp(\mathbf{x}'_i \boldsymbol{\beta})}{\sum_{j \in R_{(i)}} \exp(\mathbf{x}'_j \boldsymbol{\beta})} \end{aligned}$$

et Cox (1975) a montré que cette vraisemblance pouvait être vue comme une vraisemblance partielle, plusieurs auteurs ont déduit ses propriétés asymptotiques, on peut utiliser asymptotiquement cette vraisemblance de façon classique pour faire des tests tel que le test de Wald, le test du score ou le rapport des vraisemblances.

Remarque : Cette vraisemblance ne dépend que du rang auxquels les individus meurent et non pas de la valeur de ces durées.

En cas d'égalité on a deux façons différentes d'écrire la vraisemblance et on ne sait pas laquelle choisir, en notant  $I$  le nombre total de morts ( $I \leq n$ ),  $\mathbf{D}_{(i)}$  l'ensemble des individus morts au temps  $t_i$ , et  $\phi_i = \exp(\mathbf{x}'_i \boldsymbol{\beta})$  ; Breslow (1972) propose une approximation :

$$L(\boldsymbol{\beta} | D) = \prod_{i=1}^I \frac{\prod_{j \in \mathbf{D}_{(i)}} \phi_j}{\left( \sum_{j \in R_{(i)}} \phi_j \right)^{\text{card}\{\mathbf{D}_{(i)}\}}}$$

Efron (1977) propose une meilleure approximation :

$$L(\boldsymbol{\beta} | D) = \prod_{i=1}^I \frac{\prod_{j \in \mathbf{D}_{(i)}} \phi_j}{\prod_{k=1}^{\text{card}\{\mathbf{D}_{(i)}\}} \left[ \sum_{j \in R_{(i)}} \phi_j - \frac{k-1}{\text{card}\{\mathbf{D}_{(i)}\}} \sum_{l \in \mathbf{D}_{(i)}} \phi_l \right]}$$

La vraisemblance exacte est :

$$L(\boldsymbol{\beta} | D) = \prod_{i=1}^I \frac{\prod_{j \in \mathbf{D}_{(i)}} \phi_j}{\sum_{q \in Q_{(i)}} \phi_q}$$

Avec  $Q_{(i)}$  l'ensemble de tous les  $\text{card}\{\mathbf{D}_{(i)}\}$ -uples inclus dans  $R_{(i)}$ .

Exemple : Considérons un simple jeu de données où 5 individus sont à risque au temps  $t_i$  ; supposons que bien que notre variable aléatoire soit continue, les deux premiers individus sont notés comme mort en même temps dû à des raisons pratiques inhérentes à la récolte des données.

En utilisant la vraisemblance partielle de Cox le terme de la vraisemblance pour le temps  $t_i$  peut s'écrire

$$\left(\frac{\phi_1}{\phi_1 + \phi_2 + \phi_3 + \phi_4 + \phi_5}\right) \left(\frac{\phi_2}{\phi_2 + \phi_3 + \phi_4 + \phi_5}\right)$$

ou

$$\left(\frac{\phi_2}{\phi_1 + \phi_2 + \phi_3 + \phi_4 + \phi_5}\right) \left(\frac{\phi_1}{\phi_1 + \phi_3 + \phi_4 + \phi_5}\right)$$

on ne sait pas quel dénominateur de droite choisir.

L'approximation de Breslow donne  $\frac{\phi_2\phi_1}{(\phi_1+\phi_2+\phi_3+\phi_4+\phi_5)^2}$

L'approximation d'Efron donne  $\left(\frac{\phi_2\phi_1}{\phi_1+\phi_2+\phi_3+\phi_4+\phi_5}\right) \left(\frac{1}{0.5\phi_1+0.5\phi_2+\phi_3+\phi_4+\phi_5}\right)$ . Si on avait trois temps de morts égaux l'approximation serait :

$\left(\frac{\phi_1\phi_2\phi_3}{\phi_1+\phi_2+\phi_3+\phi_4+\phi_5}\right) \left(\frac{1}{2/3\phi_1+2/3\phi_2+2/3\phi_3+\phi_4+\phi_5}\right) \left(\frac{1}{1/3\phi_1+1/3\phi_2+1/3\phi_3+\phi_4+\phi_5}\right)$  une explication est que chaque sujet est certain d'être dans le premier dénominateur, chaque sujet à deux chances sur trois d'être dans le deuxième dénominateur et une chance sur trois d'être dans le troisième dénominateur.

Le terme de la vraisemblance exact est  $\left(\frac{\phi_2\phi_1}{\phi_1\phi_2+\phi_1\phi_3+\phi_1\phi_4+\dots+\phi_4\phi_5}\right)$ , si on a  $k$  égalités pour  $n$  sujets à risque on a  $\binom{n}{k}$  termes au dénominateur ce qui peut être gourmand en temps de calcul dans de rare cas.

Remarques : (i) Le modèle de Cox est habituellement vu comme un modèle semi-paramétrique, bien qu'on puisse aussi le traiter de manière non-paramétrique ou paramétrique (avec des splines par exemple voir [10], ou avec une loi exponentielle, weibull ou gompertz [2]).

(ii) Dans le cas d'un modèle mal spécifié (s'il manque une variable explicative par exemple) l'estimateur du maximum de vraisemblance partielle de  $\beta$  est biaisé.

Modèle de Cox stratifié :

Dans le cas où l'hypothèse d'indépendance entre les individus est violée ou si l'hypothèse des risques proportionnels est violée on peut stratifier le modèle.

On suppose que les individus dans des strates différentes ont différentes fonctions de hasard de référence mais une valeur commune pour le vecteur des coefficients  $\beta$ . On divise notre échantillon en  $k$  strates de tailles  $n_1, \dots, n_k$ ; si l'individu  $i$  appartenant à la strate  $j$  a fonction de hasard

$$h(t|\mathbf{x}) = h_j(t) * \exp(\mathbf{x}_i' \beta)$$

Exemple: Supposons qu'on a  $p$  variables explicatives où l'hypothèse des risques proportionnels tient et  $m$  variables explicatives où l'hypothèses des risques proportionnels ne tient pas. Une solution est de transformer judicieusement les variables continues en variables qualitatives et de créer autant de strates que de combinaisons possibles. La vraisemblance est :

$$L(\beta|D) = \prod_{j=1}^k \prod_{i=1}^{n_j} \left[ \frac{\exp(\mathbf{x}_j' \beta)}{\sum_{l \in R(i)} \exp(\mathbf{x}_l' \beta)} \right]^{\delta_i}$$

avec  $R(i)$  l'ensemble des individus à risque appartenant à la même strate que l'individu  $i$ .

## 2.2 Modèle au hasard constant par morceaux

**Définition 22.1** *On peut modéliser  $h_0(t)$  en construisant un modèle **au hasard constant par morceaux** lorsqu'on suppose que le hasard est constant sur des intervalles spécifiés.*

En construisant une partition de l'axe du temps  $0 < s_1 < \dots < s_J$ , avec  $\forall i = 1, \dots, n : s_j > y_i$ . On a donc  $J$  intervalles :  $]0, s_1], \dots, ]s_{J-1}, s_J]$ .

En posant le hasard :

$$\forall t \geq 0, \forall \mathbf{x} \quad h(t|\mathbf{x}) = h_j(t) * \exp(\mathbf{x}'\boldsymbol{\beta})$$

Remarque: Le modèle n'est pas paramétrique dans le sens où la forme de la fonction de hasard n'est pas définie à l'avance mais est définie par les données d'une manière non-paramétrique (en forme d'escalier).. Le modèle n'est pas non-paramétrique dans le sens où le nombre d'intervalles (et donc de paramètres à estimer) est définie par l'utilisateur et ne tend pas vers l'infini, la seule contrainte étant qu'il y a au moins un mort dans chaque intervalle.

Un choix judicieux d'intervalles est de prendre de petits intervalles où la fonction de survie varie beaucoup et des intervalles plus larges où la fonction de survie varie peu.

## Partie 3

# Statistique Bayésienne

Dans cette partie nous allons présenter quelques éléments de la statistique bayésienne. Dans la première partie nous avons vu trois approches permettant de modéliser un phénomène complexe par un modèle probabiliste, maintenant nous considérons deux approches statistiques d'un point de vue plus général qui est l'approche *fréquentiste* ou *classique* et l'approche *bayésienne*. Bien que d'un point de vue pragmatique on puisse ne pas faire de distinction entre les deux approches, pour une meilleure explication et compréhension nous faisons une séparation entre les deux. Nous ferons aussi une séparation entre *inférence* et *décision*. Il serait impossible de mettre assez de sources ici mais le lecteur voulant s'intéresser au sujet pourra se référer à [8],[17], [19], et des auteurs tels que Berger, Lindley, Savage, etc pour une réelle compréhension et non une illusion de compréhension.

### 3.1 Présentation générale

La première division entre fréquentiste et bayésien vient de l'interprétation du théorème de Bayes, en prenant le cas continu nous avons :

$$g(x|y) = \frac{f(x|y)g(y)}{\int f(x|y)g(y)dy}$$

Dans le cadre bayésien l'incertitude sur le paramètre  $\theta$  d'un modèle est décrite par une loi de probabilité  $\pi(\theta)$ , appelée loi à *priori*; l'inférence est alors fondée sur une loi à *postérieure* de  $\theta$  conditionnellement aux données,

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d\theta}$$

On a actualisé l'information sur  $\theta$  grâce aux données. Le paramètre  $\theta$  n'est pas considéré inconnu mais aléatoire.

On a  $\pi(\theta|x) \propto L(x|\theta)\pi(\theta)$ ; le terme  $\int f(x|\theta)\pi(\theta)d\theta$  est constant.

Remarques : (i) Le fait qu'on écrive  $f(x|\theta)$ , au lieu de  $f(x, \theta)$  met l'accent sur le caractère conditionnel du modèle. Avec les exemples 1 et 2 du chapitre 1 on était donc déjà dans un cadre Bayésien.

(ii) Cette différence entre paramètre aléatoire et paramètre inconnu peut expliquer le terme *fréquentiste* simplement dans la notion d'intervalle de confiance et d'intervalle de crédibilité. En effet dans l'approche classique le

paramètre n'étant pas une variable aléatoire mais un réel la probabilité que la valeur du paramètre soit dans l'intervalle de confiance est 0 ou 1; au long terme c'est-à-dire avec un grand nombre d'échantillons de notre population on a la propriété qu'approximativement  $(1 - \alpha) * 100\%$  des intervalles de confiance contiennent la vraie valeur du paramètre contrairement à l'approche bayésienne où on considère que pour l'expérience donnée le paramètre se trouve dans l'intervalle de crédibilité avec probabilité  $1 - \alpha$ .

L'approche Bayésienne est basée sur des axiomes. On y trouve plusieurs principes.

Le **principe de vraisemblance** affirme que l'information apportée par une observation  $x$  sur  $\theta$  est entièrement contenu dans la fonction de vraisemblance. Si  $x_1$  et  $x_2$  sont deux observations avec vraisemblance et qu'il existe une constante  $c$  tel que  $L_1(\theta|x) = cL_2(\theta|x)$ , alors  $x_1$  et  $x_2$  doivent apporter la même information (au sens large) sur  $\theta$  et conduire à la même inférence.

Le **principe de conditionnement** affirme que si deux expériences sur  $\theta$  sont utilisables et si on choisit une de ces expériences avec probabilité  $p$ , l'inférence sur  $\theta$  ne doit dépendre que de l'expérience choisie.

Le **principe d'exhaustivité** : deux observations  $x$  et  $y$  donnant la même valeur d'une statistique exhaustive  $T$ , c'est-à-dire telles que  $T(x) = T(y)$  doivent conduire à la même inférence sur  $\theta$ .

Le **principe des règles d'arrêt** : si une suite d'expériences,  $\varepsilon_1, \varepsilon_2, \dots$ , admet une règle d'arrêt,  $t$ , qui indique quand doivent s'arrêter les expériences, l'inférence sur  $\theta$  ne doit dépendre de  $t$  qu'à travers l'échantillon résultant.

Remarque : (i) Le principe de conditionnement et celui d'exhaustivité sont des principes qu'on retrouve dans les deux approches. En revanche le principe de vraisemblance est déduit du principe de conditionnement et d'exhaustivité. Certains de ces principes divisent fréquentistes et bayésiens (et bayésiens parmi bayésiens), notamment en **analyse séquentielle**, certains bayésiens modifient leur information a priori pour s'adapter à une règle d'arrêt.

Supposons qu'on a une distribution  $f(x|\theta)$  sur nos données et une loi à priori  $\pi(\theta)$ .

On définit : (i) la **distribution jointe** de  $(\theta, x)$  :  $\varphi(\theta, x) = f(x|\theta)\pi(\theta)$  ;

(ii) la **distribution marginale** de  $x$ ,  $m(x) = \int \varphi(\theta, x)d\theta = \int f(\theta, x)\pi(\theta)d\theta$

(iii) la **distribution de prédiction**  $p(x_{nouv}|x) = \int f(x_{nouv}|\theta)\pi(\theta|x)d\theta$ , peut être utilisée si on veut prédire une issue future du phénomène aléatoire.

**Définition 31.1** Une famille de loi  $\mathcal{F}$  est dite *conjuguée* si, pour toute loi a priori dans  $\mathcal{F}$ , la loi a posteriori appartient elle aussi à  $\mathcal{F}$ .

Exemple : Dans le cas où notre variable aléatoire  $X \sim \mathcal{N}(\mu, \sigma^2)$  et  $\theta = \mu$  ( $\sigma^2$  connu).

$$\pi(\mu) \sim \mathcal{N}\left(\mu_0, \frac{\sigma_0^2}{n_0}\right)$$

Pour un échantillon à une réalisation le principe d'exhaustivité nous donne:

$$\pi(\mu|x) \propto \mathcal{N}\left(\frac{n_0 \mu_0 + n x}{n_0 + n}, \frac{\sigma^2}{n_0 + n}\right)$$

(On a juste mis à jour les paramètres de notre loi, c'est pratique pour mettre à jour de l'information).

Décision :

Dans une approche décisionnelle on distingue  $\mathcal{X}$  l'espace des observations  $x$ ,  $\Theta$  l'espace des paramètres et  $\mathcal{D}$  l'espace des décisions possibles.

**Définition 32.1** Une fonction de **coût** est une fonction  $L$  de  $\Theta \times \mathcal{D}$  dans  $[0, +\infty[$ .

On cherche à prendre la décision qui minimise la fonction de coût. On considère le coût moyen a posteriori.

$$p(\pi, d|x) = \int_{\Theta} L(\theta, d) \pi(\theta|x) d\theta$$

Remarque : En statistique classique on intègre par sur l'espace des observations  $\mathcal{X}$  :  $R(\theta, d) = \int_{\mathcal{X}} L(\theta, dx) f(x|\theta) dx$

## 3.2 Moyenne de modèles

En pratique dans l'approche classique lors de la construction d'un modèle on fait souvent de nombreux tests pour inclure certaines variables explicatives ou non, souvent à la suite des autres, on doit aussi faire certains choix, et on ne tient pas compte de toutes les erreurs que l'on a pu faire lors du choix du modèle. On a des critères mesurant comment le modèle s'ajuste aux données avec une pénalité sur le nombre de paramètres pour favoriser les modèles simples, ce qui nous permet de choisir le "meilleur modèle". Un des avantages de l'approche bayésienne qui ne semble pas être présent en statistique classique est de tenir compte de toutes ces erreurs, avec la même philosophie que pour le paramètre de notre modèle on va essayer de tenir compte de l'incertain de notre modèle. Cette approche est décrite dans [11]:

Soit  $\mathcal{M}$  l'espace des modèles, on suppose  $\mathcal{M} = \{M_1, \dots, M_k\}$

Soit  $\Delta$  un paramètre d'intérêt (ex: une future observation, un ensemble de coefficients de régression, l'utilité d'une suite d'actions).

La distribution à postériori de  $\Delta$  est :  $\pi(\Delta|D) = \sum_{i=1}^k \pi(\Delta|D, M_i) p(M_i|D)$

où  $\pi(\Delta|D, M_i)$  est la distribution à postériori de  $\Delta$  sous le modèle  $M_i$ , et  $p(M_i|D)$  est la distribution à postériori du modèle.

$$p(M_i|D) = \frac{p(D|M_i) * p(M_i)}{\sum_{i=1}^k \pi(D|M_i) p(M_i)}$$

L'implémentation peut-être difficile dû à la difficulté de calcul de  $p(M_i|D)$  et à un très grand nombre de modèle (et donc de termes) qui peuvent être considéré dans  $\pi(\Delta|D)$ .

Une solution pour réduire le grand nombre de termes est de ne considérer que les modèles avec une probabilité à postériori dépassant un certain seuil : on choisit  $C$  tel que :

$\mathcal{A}' = \{M_i: \frac{\max_l \{p(M_l|D)\}}{p(M_i|D)} \leq C\}$  et les modèles n'appartenant pas à  $\mathcal{A}'$  sont exclus. Un deuxième critère est d'exclure tout les modèles qui ont de meilleurs sous-modèles, i.e tout les modèles appartenant à  $\mathcal{B} = \{M_i: \exists M_l \in \mathcal{M}, M_l \subset M_i, \frac{p(M_l|D)}{p(M_i|D)} > 1\}$ .



### 3.3 Non-paramétrique

Nous avons vu dans la partie 2 plusieurs modèles semi-paramétriques, en revanche dans un cadre Bayésien on doit tout modéliser et donc il n'y a pas de semi-paramétrique.

Reprenons les méthodes que nous avons vues précédemment pour éliminer un paramètre de nuisance d'un point de vue général. Supposons que nos paramètres d'intérêts sont notés  $\beta$ , et  $h$  nos paramètres qui n'ont pas un intérêt premier. La façon classique de faire est de trouver  $\beta$  et  $h$  qui maximisent la vraisemblance. Si  $h$  appartient à un espace de dimension infinie cette maximisation peut présenter des difficultés et on est obligé d'utiliser une approche non-paramétrique alors que notre but est simplement de faire une inférence sur  $\beta$ .

En notant  $L(\beta, h|D)$  la vraisemblance de nos données. La vraisemblance complète est  $L(\beta, h|D)$ . Si on peut écrire la vraisemblance partielle comme un produit de deux vraisemblance  $L(\beta, h|D) = L_1(\beta|D) * L_2(h|D)$  alors on maximise seulement  $L_1(\beta|D)$ .

Dans un cadre Bayésien, il est plus intéressant d'utiliser une vraisemblance marginale,  $h$  est un paramètre dépendant du temps et nous devons donc le modéliser par un processus aléatoire.

#### Définition 33.1 Processus aléatoire :

Soit  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace probabilisé,  $T$  un sous-ensemble de  $\mathbb{R}^+$  et  $(E, \varepsilon)$  un espace mesurable. La famille  $Z = (Z_{(t)}, t \in \mathbb{R})$  de variables aléatoires sur  $(\Omega, \mathcal{F}, \mathbb{P})$  avec valeurs dans  $(E, \varepsilon)$  est appelé un **processus stochastique** (ou **aléatoire**).

On appelle  $T$  l'ensemble des paramètres. Pour chaque  $\omega \in \Omega$  la fonction de  $T$  dans  $E$  défini par  $t \rightarrow Z_t(\omega)$  est appelée **trajectoire** (ou **chemin**) du processus. On se restreint au cas où  $E = \mathbb{R}^+$  et  $\varepsilon = \mathcal{B}(\mathbb{R}^+)$ .

Concrètement un processus aléatoire sert à décrire l'évolution d'une variable aléatoire dans le temps.

#### Définition 33.2 Processus aléatoire Gamma

Notons  $\mathcal{G}(\alpha, \lambda)$  la distribution gamma avec paramètre  $\alpha > 0$  et  $\lambda > 0$ . Soit  $\alpha(t)$ ,  $t \geq 0$  une fonction continue à gauche et croissante tel que  $\alpha(0) = 0$ , et soit  $Z(t), t \geq 0$  un processus aléatoire avec les propriétés :

- (i)  $Z(0) = 0$ ,

(ii)  $Z(t)$  a des sauts indépendants en des intervalles disjoints,

(iii)  $t > s, Z(t) - Z(s) \sim \mathcal{G}(c(\alpha(t) - \alpha(s)), c)$

Alors le processus  $\{Z(t), t \geq 0\}$  est un processus aléatoire gamma, noté  $Z(t) \sim \mathcal{G}(c\alpha(t), c)$ .

Remarquons que  $\alpha(t)$  est la moyenne du processus et  $c$  est un poids, ou paramètre de confiance de la moyenne. La trajectoire du processus est approximativement croissante. L'utilisation d'un processus gamma pour représenter  $H(t)$  est cohérente avec les propriétés de la partie 1.

Kalbfleisch (1978) a montré que la vraisemblance partielle de Cox est un cas limite d'une loi à posteriori marginale de  $\beta$  pour des variables explicatives indépendantes du temps. Les principaux travaux sur l'utilisation de la vraisemblance partielle du modèle de Cox dans un cadre bayésien sont expliqués par Sinha (2003). Et donc en supposant  $\beta$  et  $h$  indépendants on peut utiliser la vraisemblance partielle du modèle de Cox pour conduire une inférence sur  $\beta$ .

### 3.4 Algorithmes de simulations

En pratique la loi à posteriori n'a pas toujours une forme analytique et on simule  $\pi(\theta|x)$  à une constante près. Robert [17] développe très bien le sujet Ici nous présentons deux résultats fondamentaux.

**Proposition 34.1** Soit  $X$  une v.a réelle de fonction de répartition  $F(x) = \mathbb{P}(X \leq x)$ , dont l'inverse est défini par  $F^{-1}(u) = \inf\{x : F(x) \geq u\}$ . Si  $U$  est une variable aléatoire de loi uniforme sur  $[0, 1]$ , alors  $F^{-1}(U)$  a même loi que  $X$ .

Preuve : pour tout  $x \in \mathbb{R}$ , on a  $\mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x) = \mathbb{P}(X \leq x)$ .

**Lemme 35.1: L'algorithme d'acceptation-rejet :**

Soit  $f$  notre densité cible  $M$  une constante et  $g$  une densité instrumentale telle que  $f(x) \leq Mg(x)$  sur l'ensemble de définition de  $f$ .

L'algorithme :

(i) simuler  $X \sim g, U \sim U_{[0,1]}$  ;

(ii) accepter  $Y = X$  si  $U \leq f(x)/Mg(x)$  ;

(iii) retourner à 1 sinon. ;

génère une variable  $Y$  distribué selon  $f$ .

Conséquence :

La conséquence principale de ce lemme est de fournir une méthode générique pour simuler de n'importe qu'elle densité qui est connu à un facteur multiplicatif près (la constante de normalisation n'a pas besoin d'être connu). Ce qui est utile dans le cadre bayésien puisqu'on a  $\pi(\theta|x) \propto L(x|\theta)\pi(\theta)$ .

## Partie 4

# Applications et discussion

### 4.1 Applications

Les nombreux exemples que nous avons vus peuvent être directement appliqués lors de l'analyse principale du jeu de données pour répondre aux différentes questions de l'étude. Dans le cas de la moyenne de modèle le package `bic.surv` du logiciel R est recommandé plutôt que SAS. Dans le cas où l'hypothèse de proportionnalité du modèle de Cox ne tient pas nous avons vu que stratifier par variables explicatives pouvait être une solution ; d'autres solutions sont possibles tel que stratifier l'axe du temps ou changer de modèle (un modèle additif par exemple ou un modèle plus général).

### 4.2 Calcul du nombre de sujets

On souhaite comparer deux traitements, un connu contre un nouveau. Une des étapes principale de la statistique est de passer des données au modèle de la façon la plus rigoureuse possible afin de tirer les conclusions les plus précises et les plus justes possibles quant au phénomène observé. Nous nous plaçons une étape avant c'est-à-dire au moment de désigner une stratégie afin d'aboutir à l'objectif principal de l'essai clinique. La fin principale de l'essai clinique est d'évaluer l'effet du nouveau traitement comparé à l'ancien, puis lors de l'arrêt de l'étude de prendre une décision justifiée (ou de justifier une décision) quand à poursuivre ou non l'essai clinique dans une autre phase. Le moyen est de déterminer le nombre de sujets nécessaire pour atteindre un niveau d'évidence suffisant. On se place dans le cadre le plus simple au point de vue de l'interprétation où on suppose que les données vont suivre un modèle de Cox avec un simple effet des médicaments sans interactions avec d'autres variables explicatives.

Soit  $\mathcal{D}$  notre espace de décisions.

Une approche souvent utilisée en statistique est de réduire une prise de décision à un test d'hypothèse où l'on cherche à contrôler l'erreur de type I et type II.

	$H_0$ vraie	$H_0$ fausse
$H_0$ non-rejeté	Correct : $1 - \alpha$	Erreur de type II : $\beta$
$H_0$ rejeté	Erreur type I : $\alpha$	correct : $1 - \beta$

L'erreur de type I est souvent fixée à une valeur conventionnelle (souvent 0.05), et l'erreur de type II doit être contrôlée au mieux (en recherchant un test uniformément plus puissant) et calculé au moins en un point.

Remarques : (i) Une « erreur » statistique peut se produire si on ne contrôle pas l'erreur de type II et qu'on pose pour  $H_0$  une hypothèse peu probable, dans ce cas  $H_1$  sera (trop) souvent décidée à tort.

(ii) Certaines décisions peuvent nécessiter plusieurs tests, ou une procédure basée sur une suite de tests, c'est le cas par exemple lors du choix d'un modèle.

La difficulté consiste à réduire notre espace de décisions à deux éléments puis à traduire formellement l'arbitraire du choix d'une décision en test d'hypothèse sur un paramètre de notre modèle.

Plusieurs méthodes :

Considérons  $\theta$  notre paramètre d'intérêt, le logarithme du rapport des hasards.

Test d'hypothèse ponctuel : L'approche habituelle pour ce genre de problème est d'utiliser un test statistique de type log-rank pour tester l'hypothèse ponctuel  $H_0: \theta = 0$  où  $\theta$  est le logarithme du rapport des hasards, en pratique l'usage est de

spécifier une hypothèse alternative  $H_1: \theta = \theta_A$  ponctuelle est de s'assurer que sous  $H_1$  on atteint une puissance assez élevée.  $\theta_A$  peut prendre plusieurs valeurs afin de contrôler le risque de deuxième espèce :

- (i) La plus petite différence clinique qui en vaut la peine
- (ii) Une différence clinique que les chercheurs jugent comme valant la peine d'être détecté
- (iii) la différence clinique la plus probable d'apparaître.

### Plage d'équivalence :

Spiegelhalter et al [23] discutent de ce sujet, pour savoir quand arrêter un essai clinique en analyse séquentielle et d'utiliser une approche bayésienne avec différents à priori. Dans certaines circonstances les traitements peuvent être inégaux en termes de coût (toxicité du médicament par rapport à son effet, coût financier, etc...), dans ce cas le médicament le plus coûteux devra atteindre une certaine valeur  $\theta_I$  ( $\theta_I$  peut être  $\neq 0$ ) avant d'être simplement considéré; une autre valeur peut être défini comme celle pour laquelle le traitement puisse indiquer une « différence clinique supérieure »  $\theta_S$  ( $\theta_S$  peut-être  $\neq 0$ ). Si  $\theta \leq \theta_I$  le nouveau traitement est jugé inférieur, si  $\theta \geq \theta_S$  le nouveau traitement est jugé supérieur, et si  $\theta_S \leq \theta \leq \theta_I$  nous ne savons pas faire un choix entre les deux traitements.

Le point (i) du paragraphe précédent correspond à poser  $\theta_A = \theta_S$ , le point (ii) laisse le choix au chercheur qui peut prendre  $\theta_A$  aussi grand qu'il veut pour réduire la taille de l'échantillon, le point (iii) permet de calculer une taille d'échantillon sur ce qui est le plus probable de se passer.

Si  $\theta$  est le log-hazard ratio en pratique on utilise l'approximation suivante. On suppose  $\theta = 0$  et l'hypothèse des risques proportionnels, la statistique du log-rank donne  $4 * \frac{S_m}{m} = \hat{\theta}$  et  $\hat{\theta} \sim \mathcal{N}\left(\theta, \frac{4}{m}\right)$  voir [21],[22],  $m$  étant le nombre d'évènements et  $S_m$  la statistique du log-rank. Cette approximation permet de se ramener à un cadre paramétrique mais peut aussi être utilisé en analyse séquentielle. Cet exemple montre que dans beaucoup de cas il est plus simple de chercher à calculer le nombre d'évènements requis en premier.

### Décision basée sur deux critères :

C'est l'approche expliquée par Neuenschwander [14]. Soit  $\theta$  le rapport des hasards, soit  $\mathcal{D} = \{d_0, d_1\}$  notre espace de décision réduit.

ici  $d_0 = \text{"l'essai clinique est un échec"}$  ;  $d_1 = \text{"l'essai clinique est un succès"}$ .

On a deux *critères* de décision : (c1) :  $\hat{\theta} \leq \theta_c$

$$(c2) : \int_{-\infty}^{\theta_l} \pi(\theta|x) d\theta \geq 1-\alpha$$

Ici  $\theta_l$  représente la valeur de comparaison pour laquelle on peut dire que le nouveau traitement est inférieur, i.e si  $\theta < \theta_l$  le nouveau traitement n'est pas inférieur.

$1-\alpha$  représente le *niveau d'évidence* que le traitement n'est pas inférieur.  $\theta_c$  représente la valeur pour laquelle on ne sait pas décider si on souhaite continuer le traitement ou pas. Si  $\theta = \theta_c$  on est indifférent quant à continuer le traitement ou non (ce qui revient à faire un pile ou face), si  $\theta < \theta_c$  on souhaite continuer le traitement et si  $\theta > \theta_c$  on souhaite arrêter le traitement.

Notre *règle de décision* est :

$$(c1) \text{ et } (c2) \Rightarrow d_1$$

$$\text{non}((c1) \text{ et } (c2)) \Leftrightarrow \text{non}((c1)) \text{ ou } \text{non}((c2)) \Rightarrow d_0$$

Supposons la vraie valeur de  $\theta$  comme  $\theta = \theta_c$ , pour être indifférent quant à continuer le succès ou non on souhaite ajuster la taille de l'échantillon pour décider  $d_1$  une fois sur deux.

Avec une loi à priori impropre, l'approche fréquentiste et l'approche classique donnent les mêmes résultats écrits de deux manières différentes, en supposant la variance connue :

$$\hat{\theta} \sim \mathcal{N}\left(\theta_c, \frac{4}{m}\right) \text{ approche fréquentiste} \quad (1)$$

$$\theta|\hat{\theta} \sim \mathcal{N}\left(\hat{\theta}, \frac{4}{m}\right) \text{ approche bayésienne} \quad (2)$$

(1) donne qu'on a une chance sur deux d'avoir (c1), il faut donc ajuster (c2) pour que (c1)  $\Rightarrow$  (c2). En prenant  $m$  la plus petite valeur tel que (c1)  $\Rightarrow$  (c2), on a aussi (c2)  $\Rightarrow$  (c1).

Le nombre d'évènements requis est donc :

$$m = 4 * \frac{z_{\alpha}^2}{(\theta_1 - \theta_c)^2} \quad (3)$$

Avec  $z_{\alpha}$  le quantile de la loi gaussienne centrée réduite au niveau  $\alpha$ . Pour le critère 2 il est important de prendre une valeur de  $\alpha$  supérieure à 0.5 pour mesurer *un niveau d'évidence* plutôt qu'un *niveau de « non-évidence »*. Dans le cas contraire nos données et notre analyse aurait moins de poids quant à la non-infériorité du traitement que de lancer une pièce et de procéder à un pile-ou face. (3) prévient ce problème grâce à la symétrie de la fonction quantile. Un des avantages de cette approche est que le deuxième critère est transparent et plus facilement compréhensible pour des non-statisticiens qu'un calcul de puissance lors d'un test d'hypothèse.

Une autre solution cohérente serait de tester  $H_0: \theta > \theta_c$  vs  $H_1: \theta < \theta_c$ , ou de calculer un intervalle de confiance qui ne contienne pas  $\theta_c$ , mais dans certains cas cela demanderait trop de sujets.

Une fois qu'on a décidé le nombre d'évènements pour atteindre un bon niveau de conclusion, on cherche à calculer le nombre de sujets.

Plusieurs approches sont possibles Schoenfeld [18] propose une formule basée sur la période de recrutement  $a$ , le taux de recrutement  $r$ , la durée entre le dernier recrutement et la fin de l'étude  $f$ , et la courbe de survie pour l'ancien traitement  $S_A$ . Cette formule permet de déterminer approximativement le nombre de patients qui vont mourir pendant l'étude; en supposant un taux de censure global ou par période il est alors facile de passer du nombre d'évènements requis au nombre de sujet requis.

Remarque : Nous avons présenté ce dernier paragraphe comme un problème direct mais nous aurions très bien pu le présenter comme un problème inverse.

Une fois qu'on a trouvé le nombre d'évènements requis et un nombre de sujet permettant d'obtenir le nombre d'évènement requis en un temps raisonnable il est facile d'utiliser (1) ou (2) pour calculer une probabilité conditionnelle de décider  $d_1$  ou  $d_0$ .



### Exemple :

En prenant  $\theta_I = 1$ ,  $\theta_C = 0.75$ ,  $\alpha = 0.1$ , notons  $\mathcal{H}$  l'ensemble de nos hypothèses, avec  $m = 70$ .

$$\mathbb{P}(d_1|\mathcal{H}, \theta = 0.75) \approx 0.41 \quad \mathbb{P}(d_1|\mathcal{H}, \theta = 0.70) \approx 0.49 \quad \mathbb{P}(d_1|\mathcal{H}, \theta = 0.65) \approx 0.57$$

$$\mathbb{P}(d_1|\mathcal{H}, \theta = 0.55) \approx 0.72 \quad \mathbb{P}(d_1|\mathcal{H}, \theta = 0.85) \approx 0.25 \quad \mathbb{P}(d_1|\mathcal{H}, \theta = 1) \approx 0.1$$

## Conclusion

Ainsi, j'ai effectué mon stage de fin d'étude du Master de Statistique de Strasbourg. Les questions initiales ont été explorées par simulations puis répondues et expliquées par la théorie. Des pistes de réflexions ainsi que des outils statistiques ont été proposés. Lors de ce stage j'ai pu mettre en perspective et mieux comprendre les connaissances théoriques acquises pendant ma formation universitaire grâce à une réelle réflexion et une recherche bibliographique, j'ai pu voir différents intérêts de passer par un cadre formaliste pour résoudre un problème et enfin ce stage a été une bonne introduction au domaine biostatistique et aux essais cliniques dans l'industrie pharmaceutique. J'aborde avec confiance la suite.

# References

- [1] Adcock, C. J. (1997). Sample size determination: a review. *Journal of the Royal Statistical Society: Series D (The Statistician)*, **46**(2), 261-283.
- [2] Bender, R., Augustin, T., & Blettner, M. (2005). Generating survival times to simulate Cox proportional hazards models. *Statistics in medicine*, **24**(11), 1713-1723.
- [3] Bjørnstad, J. F. (1996). On the generalization of the likelihood function and the likelihood principle. *Journal of the American Statistical Association*, **91**(434), 791-806.
- [4] Breuils, C. (2003). *Analyse de durées de vie: analyse séquentielle du modèle des risques proportionnels et tests d'homogénéité* (Doctoral dissertation, Université de Technologie de Compiègne).
- [5] Breslow, N. E. (1972). Discussion of Professor Cox's paper. *Journal of the Royal Statistical Society, Series B*, **34**, 216-217.
- [6] Cox, D. R. (1972). Regression models and life tables(with Discussion). *Journal of the Royal Statistical Society. Series B (Methodological)*, **34** 187-220.
- [7] Cox D.R. (1975). Partial Likelihood. *Biometrika*, **62**, 269-276.
- [8] Cox, D., & Mayo, D. G. (2010). II: Objectivity and Conditionality in Frequentist Inference. *Error and inference: Recent exchanges on experimental reasoning, reliability, and the objectivity and rationality of science*, 276.
- [9] Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *Journal of the American statistical Association*, **72**(359), 557-565.
- [10] Grambsch, T. M., & Therneau, T. M. (2000). *Modeling survival data: extending the Cox model*. ff. New York Berlin Heidelberg: Springer, 127.
- [11] Ibrahim J.G., Chen M.-H., & Sinha, D. (2001). *Bayesian Survival Analysis*. New York: Springer-Verlag.
- [12] Kalbfleisch, J. D. (1978). Non-parametric Bayesian analysis of survival time data. *Journal of the Royal Statistical Society. Series B (Methodological)* **40**, 214-221.
- [13] Murthy, D. P., Xie, M., & Jiang, R. (2004). *Weibull models* (Vol. 505). John Wiley & Sons.
- [14] Neuenschwander, B., Rouyrre, N., Hollaender, N., Zuber, E., & Branson, M. (2011). A proof of concept phase II non-inferiority criterion. *Statistics in medicine*, **30**(13), 1618-1627.
- [15] Persson, I. (2002). *Essays on the assumption of proportional hazards in Cox regression*. Uppsala University.
- [16] Robert, C. P., Casella, G., & Robert, C. P. (1999). *Monte Carlo statistical methods* (Vol. 58). New York: Springer.

- [17] Robert, C. P. (2005). *Le choix bayésien: Principes et pratique*. Springer.
- [18] Schoenfeld, D. A. (1983). Sample-size formula for the proportional-hazards regression model. *Biometrics*, **39**, 499-503.
- [19] Senn, S. (2011). You may believe you are a Bayesian but you are probably wrong. *Rationality, Markets and Morals*, **2**(42).
- [20] Sinha D., Ibrahim J.G, & Chen M.-H. (2003). A Bayesian Justification of Cox's Partial Likelihood. *Biometrika*, **90**, 629-641.
- [21] Spiegelhalter, D. J., Freedman, L. S., & Parmar, M. K. (1994). Bayesian approaches to randomized trials(with Discussion). *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, **157**, 357-416.
- [22] Tsiatis, A. A. (1981). The asymptotic joint distribution of the efficient scores test for the proportional hazards model calculated over time. *Biometrika*, **68**(1), 311-315.