Analyse de survie

Résumé du cours

Laurent Gardes

Table des matières

Chapitre 1

Introduction

Dans ce cours, toutes les variables aléatoires considérées sont définies sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.
L'analyse de survie a pour objectif l'étude de la loi d'une durée de survie modélisée par une variable aléatoire positive T. Une durée de survie est le temps écoulé avant l'apparition d'un événement d'intérêt. La loi de T est d'un point de vu formel la mesure image de $\mathbb{P}$ par la fonction mesurable T. Autrement dit, la loi $\mathbb{P}_T$ de T est définie pour tout $A \in \mathcal{B}(\mathbb{R}^+)$ par

$$\mathbb{P}_T(A) = \mathbb{P}(T^{-1}(A)) = \mathbb{P}(\{\omega \in \Omega;\ \ T(\omega) \in A\}) = \mathbb{P}(T \in A).$$

Dans toute la suite de ce cours, on supposera que $\mathbb{P}_T(\{0\}) = \mathbb{P}(T = 0) = 0$.

1.1 Fonction de survie

La mesure de probabilité $\mathbb{P}_T$ peut être caractérisée plus simplement par la fonction de répartition F définie pour tout $t > 0$ par $F(t) := \mathbb{P}_T([t, \infty[) = \mathbb{P}(T \leq t)$ ou de manière équivalente par la fonction de survie. La fonction de survie S d'une variable aléatoire $T : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$ est définie pour tout $t \geq 0$ par

$$\boxed{S(t) := 1 - F(t) = \mathbb{P}_T(]t, \infty[) = \mathbb{P}(T > t).}$$

Certains auteurs notent par $\overline{F}$ la fonction de survie.

1.2 Fonction de hasard

La loi de T est également caractérisée par la fonction de hasard dont la définition est donnée ci-après. S'il existe une fonction $\vartheta(\cdot) : [0,\infty[\to [0,\infty[$ telle que la limite

$$\boxed{\begin{array}{rl} \lambda(t) & = \lim_{h\to 0} \vartheta(h) \frac{S(t-h)-S(t)}{S(t-h)} = \lim_{h\to 0} \vartheta(h) \frac{\mathbb{P}(t-h<T\leq t)}{\mathbb{P}(T>t-h)} \\ & = \lim_{h\to 0} \vartheta(h) \mathbb{P}(t-h<T\leq t \mid T>t-h), \end{array}}$$

existe, la fonction $\lambda(\cdot)$ est appelée fonction de hasard de T.

Loi absolument continu En notant f la densité et en prenant $\vartheta(h) = 1/h$,

$$\boxed{\lambda(t) := \frac{f(t)}{S(t)}.}$$

Pour tout $t > 0$,

$$\boxed{\ln S(t) = -\int_0^t \lambda(x)dx.}$$

Loi discrète Dans le cas où T est une variable aléatoire discrète prenant ses valeurs dans l'ensemble $\{t_1, t_2, \cdots\}$, en prenant $\vartheta(h) = 1$, la fonction de hasard est déterminée par les valeurs

$$\boxed{\lambda(t_j) = \mathbb{P}(T = t_j \mid T \geq t_j) = \frac{\mathbb{P}(T=t_j)}{\mathbb{P}(T\geq t_j)}, \quad j \in \mathbb{N}\setminus\{0\}.}$$

Les liens existants entre les probabilités $\mathbb{P}(T = t_j)$ et les $\lambda(t_j)$ sont donnés ci-dessous. On note $\{t_{(1)} < t_{(2)} < \cdots\}$ les valeurs rangées par ordre croissant et on pose $p_j := \mathbb{P}(T = t_{(j)})$ pour tout $j \in \mathbb{N}\setminus\{0\}$.

$$\boxed{\lambda(t_{(j)}) =: \lambda_j = p_j \Big/ \sum_{k\geq j} p_k \in [0,1].}$$

De plus, $\lambda_1 = p_1$ et pour tout $j \geq 2$,

$$\boxed{p_j = \lambda_j \prod_{k=1}^{j-1}(1-\lambda_k).}$$

Enfin,

$$\boxed{S(t) = \mathbb{P}(T > t) = \prod_{k:t_{(k)} \leq t} (1 - \lambda_k).}$$

1.3 Fonction de hasard cumulé

On peut également caractériser la loi d'une durée de survie par sa fonction de hasard cumulé. Si la loi de T est absolument continue, la fonction de hasard cumulé $\Lambda(\cdot)$ de T est définie pour tout $t > 0$ par

$$\boxed{\Lambda(t) := \int_0^t \lambda(x)dx.}$$

Dans ce cas, $\Lambda'(t) = \lambda(t)$.
Si T est une variable aléatoire discrète prenant ses valeurs dans l'ensemble $\{t_1, t_2, \cdots\}$, la fonction de hasard cumulé est donnée par

$$\boxed{\Lambda(t) := \sum_{i:t_i \leq t} \lambda(t_i).}$$

Chapitre 2

Estimation de la loi d'une durée de survie

L'objectif de ce chapitre est d'estimer la fonction de survie d'une durée de survie T. Pour ce faire, on utilisera un ensemble de n individus. Nous nous placerons pour commencer sous le modèle idéal où l'on observe pour ces n individus les réalisations $t_1, \cdots, t_n$ de n réplications indépendantes $T_1, \cdots, T_n$ de T. Nous nous concentrerons ensuite sur le cas plus intéressant (mais plus difficile) où les durées de survie ne sont pas complètement observées.

2.1 Cas d'un échantillon i.i.d.

2.1.1 Loi empirique

Lorsqu'on dispose des observations des variables aléatoires indépendantes $T_1, \cdots, T_n$ de même loi $\mathbb{P}_T$, l'estimateur usuel de la loi $\mathbb{P}_T$ est la distribution empirique

$$\boxed{\widehat{\mathbb{P}}_{n,T} = \frac{1}{n}\sum_{i=1}^{n} \delta_{T_i},}$$

où δ_a est la mesure de Dirac au point $a \in \mathbb{R}$. Noter que $\widehat{\mathbb{P}}_{n,T}$ est une mesure de probabilité aléatoire. Si $\mathbf{t} = (t_1, \cdots, t_n)$ est la valeur observée du vecteur aléatoire $\mathbf{T} = (T_1, \cdots, T_n)$, la mesure empirique observée est définie pour tout $A \in \mathcal{B}(\mathbb{R}^+)$ par

$$\widehat{\mathbb{P}}_{n,T}(A \mid \mathbf{t}) = \frac{1}{n}\sum_{i=1}^{n} \delta_{t_i}(A).$$

La mesure empirique de l'ensemble A sachant les observations est donc simplement la proportion d'observations appartenant à A. On en déduit l'expression de la fonction de survie empirique qui est l'estimateur classique de $S(x)$ pour tout $x > 0$.

$$\boxed{\widehat{S}_n(x) := \widehat{\mathbb{P}}_{n,T}(]t,\infty[) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{]t,\infty[}(T_i).}$$

La valeur observée de la fonction de survie empirique est donc la valeur

$$\widehat{\mathbb{P}}_{n,T}(]t,\infty[\mid \mathbf{t}) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{]t,\infty[}(t_i) \in [0,1].$$

La mesure empirique peut également être obtenue par maximisation d'une vraisemblance non paramétrique.

2.1.2 Maximum de vraisemblance non paramétrique

Dans l'idéal, il s'agit de maximiser sur l'ensemble $\mathcal{Q}_+$ est l'ensemble des lois de probabilité sur $\mathbb{R}^+$ la fonction définie pour tout $h > 0$ par

$$\boxed{L_h(\mathbb{P}_T \mid \mathbf{t}) = \prod_{i=1}^{n} \mathbb{P}_T(]t_i - h, t_i]).}$$

La fonction $L_h(\cdot \mid \mathbf{t})$ est la vraisemblance non paramétrique de $\mathbb{P}_T$ sachant les observations $\mathbf{t}$. Cette maximisation est malheureusement impossible. On restreint donc la recherche du maximum à un sous ensemble $\mathcal{P}$ de l'ensemble des lois de probabilité défini de la façon suivante. On note m le nombre de valeurs distinctes dans l'ensemble des valeurs observées $\{t_1, \cdots, t_n\}$ et $t_{(1)} < \cdots < t_{(m)}$ l'ensemble des valeurs distinctes rangées par ordre croissant. Le sous ensemble $\mathcal{P}$ est donné par

$$\mathcal{P} = \left\{ \tilde{\mathbb{P}} : \mathcal{A} \to [0,1];\ \ \tilde{\mathbb{P}} = \sum_{j=1}^{m} p_j \delta_{t_{(j)}},\ p_1, \cdots, p_m > 0,\ \sum_{i=1}^{m} p_i = 1 \right\}.$$

En notant $\sharp E$ le cardinal d'un ensemble E, posons $n_j = \sharp\{i \mid t_i = t_{(j)}\}$ pour tout $j = 1, \cdots, m$. Pour tout $h < \min(t_{(j)} - t_{(j-1)};\ j = 1, \cdots, m-1)$ et pour tout $\tilde{\mathbb{P}} \in \mathcal{P}$,

$$\boxed{L_h(\tilde{\mathbb{P}} \mid \mathbf{t}) = L_h(\mathbf{p} \mid \mathbf{t}) = \prod_{j=1}^{m} p_j^{n_j},}$$

où $\mathbf{p} = (p_1, \cdots, p_m)$ avec la contrainte $p_1 + \cdots + p_m = 1$. On peut résoudre directement ce problème d'optimisation mais il est plus simple à résoudre en effectuant un changement de variable. Posons pour tout $j = 1, \cdots, m$,

$$\lambda_j = p_j \Big/ \sum_{k=j}^{m} p_k \in [0,1] \text{ et } N_j := \sum_{k=j}^{m} n_k.$$

En posant $\lambda = (\lambda_1, \cdots, \lambda_{m-1})$,

$$\boxed{L_h(\tilde{\mathbb{P}} \mid \mathbf{t}) = L_h(\lambda \mid \mathbf{t}) = \prod_{j=1}^{m-1} \lambda_j^{n_j} (1-\lambda_j)^{N_{j+1}}.}$$

L'estimateur du maximum de vraisemblance de λ est

$$\left(\widehat{\lambda}_{n,1}, \cdots, \widehat{\lambda}_{n,m-1}\right) = \underset{\lambda \in [0,1]^{m-1}}{\arg\max}\ L_h(\lambda \mid \mathbf{t})$$

avec pour $k = 1, \cdots, m-1$,

$$\boxed{\widehat{\lambda}_{n,k} = \frac{n_k}{N_k},}$$

L'estimateur du maximum de vraisemblance de $\mathbf{p} = (p_1, \cdots, p_m)$ est donné pour tout $j = 1, \cdots, m-1$ par

$$\widehat{p}_{n,j} = \widehat{\lambda}_{n,j} \prod_{k=1}^{j-1} (1 - \widehat{\lambda}_{n,k}) = \frac{n_j}{n}.$$

L'estimateur empirique coincide donc avec l'estimateur NPML de la loi de T.

2.2 Cas de données incomplètes

Dans le cas idéal, on dispose des observations des durées de vie pour la totalité des n individus de l'étude. En pratique, il arrive souvent que certaines observations soient manquantes. Il y a principalement deux types de données manquantes : les données manquantes par censure et les données manquantes par troncature. Nous allons dans un premier temps définir les notions de censure et de troncature. Nous nous intéresserons par la suite à l'estimation d'une fonction de survie en présence de données manquantes par censure (à droite).

2.2.1 Censure et troncature

Soit T une variable aléatoire positive modélisant la durée avant l'apparition du phénomène d'intérêt. Donnons pour commencer les définitions de la censure et de la troncature.

Troncature

La variable aléatoire T est dite tronquée (non aléatoirement) sur un sous ensemble $A \subset \mathbb{R}$ si on observe uniquement les valeurs de T dans l'ensemble A. Ce type de données manquantes ne sera pas considéré dans la suite de ce cours.

Censure

Une variable aléatoire T est dite censurée si observe le couple aléatoire $(T^\star, \Delta)$ avec $T^\star = T$ si $\Delta = 1$ et $T^\star \neq T$ si $\Delta = 0$.

Il existe plusieurs types de censure.

1) **Censure à droite de type I** – Soit $c \in \mathbb{R}$. Le couple aléatoire $(T^\star, \Delta)$ est donné par $T^\star = \min(T, c)$ et $\Delta = \mathbb{I}_{]-\infty,c]}(T)$.
2) **Censure aléatoire à droite** – Le couple aléatoire $(T^\star, \Delta)$ est donné par $T^\star = \min(T, C)$ et $\Delta = \mathbb{I}_{T \leq C}$ où C est une variable aléatoire. Si la variable C est indépendante de T on dira que la censure est non informative.

Dans toute la suite de ce cours, on se placera dans le cas d'une censure aléatoire à droite non informative.

2.2.2 Estimateur de la loi d'une durée de survie censurée

Soient $(T_1^\star, \Delta_1), \cdots, (T_n^\star, \Delta_n)$ des couples de variables aléatoires indépendants et de même loi que le couple $(T^\star, \Delta)$ avec $T^\star = \min(T, C)$ et $\Delta = \mathbb{I}_{T \leq C}$. Les variables aléatoires positives T et C sont indépendantes. La variable T est la durée de survie qui nous intéresse et C est la censure aléatoire. On souhaite estimer la loi $\mathbb{P}_T$ de T à partir des observations $(\mathbf{t}^\star, \mathbf{d}) = \{(t_1^\star, d_1), \cdots, (t_n^\star, d_n)\}$. La vraisemblance non paramétrique (observée) est donnée pour tout $h > 0$ par

$$L_h(\mathbb{P}_T \mid (\mathbf{t}^\star, \mathbf{d})) = \prod_{i=1}^{n} \begin{array}{l} [\mathbb{P}(\{t_i^\star - h < T \ \leq t_i^\star\} \cap \{T \leq C\})]^{d_i} \\ \times [\mathbb{P}(\{t_i^\star - h < C \ \leq t_i^\star\} \cap \{T > C\})]^{1-d_i}. \end{array}$$

Comme dans le paragraphe précédent, il n'est pas envisageable de maximiser cette fonction sur toutes les lois possibles pour le couple (T, C). En notant toujours $t^\star_{(1)} < \cdots < t^\star_{(m)}$ les valeurs distinctes de l'ensemble $\{t^\star_1, \cdots, t^\star_n\}$ rangées par ordre croissant, on va chercher le maximum de la vraisemblance non paramétrique sur le sous-ensemble des lois de probabilité

$$\mathcal{P} := \left\{ \tilde{\mathbb{P}} = \tilde{\mathbb{P}}_T \otimes \tilde{\mathbb{P}}_C : \mathcal{A} \otimes \mathcal{A} \to [0,1];\ \ \tilde{\mathbb{P}} = \sum_{j=1}^{m+1} p_j \delta_{t^\star_{(j)}} \otimes \sum_{j=1}^{m} q_j \delta_{t^\star_{(j)}} \right\},$$

où $t^\star_{(m+1)} = \infty$, $p_1, \cdots, p_{m+1} > 0$, $q_1, \cdots, q_m > 0$ et

$$\sum_{i=1}^{m+1} p_i = \sum_{i=1}^{m} q_i = 1.$$

Posons pour tout $j = 1, \cdots, m+1$,

$$O_j := O_j(\mathbf{t}^\star, \mathbf{d}) = \sum_{i : t^\star_i = t^\star_{(j)}} d_i \ \text{ et } \ n_j = n_j(\mathbf{t}^\star) = \sharp\{i \mid t^\star_i = t^\star_{(j)}\}.$$

Pour tout $h < \min(t^\star_{(j)} - t^\star_{(j-1)};\ j = 1, \cdots, m)$ (avec $t^\star_{(0)} = 0$) et pour tout $\tilde{\mathbb{P}}_T \otimes \tilde{\mathbb{P}}_C \in \mathcal{P}$,

$$\boxed{L_h(\tilde{\mathbb{P}}_T \mid (\mathbf{t}^\star, \mathbf{d})) = L_h(\mathbf{p} \mid (\mathbf{t}^\star, \mathbf{d})) = K \prod_{j=1}^{m} p_j^{O_j} \left(\sum_{k=j+1}^{m+1} p_k \right)^{n_j - O_j},}$$

où K est une constante ne dépendant pas des p_i et $\mathbf{p} = (p_1, \cdots, p_{m+1})$ avec $p_1 + \cdots + p_{m+1} = 1$. Comme précédemment, on va effectuer un changement de variable pour éviter d'avoir à maximiser sous contrainte. Pour tout $j = 1, \cdots, m+1$, on pose

$$\lambda_j = p_j \Big/ \sum_{k=j}^{m+1} p_k \ .$$

En posant $\lambda = (\lambda_1, \cdots, \lambda_m)$, on a pour tout $\tilde{\mathbb{P}}_T \otimes \tilde{\mathbb{P}}_C \in \mathcal{P}$,

$$\boxed{L_h(\tilde{\mathbb{P}}_T \mid (\mathbf{t}^\star, \mathbf{d})) = L_h(\lambda \mid (\mathbf{t}^\star, \mathbf{d})) = K \prod_{j=1}^{m} \lambda_j^{O_j} (1 - \lambda_j)^{N_j - O_j},}$$

L'estimateur NPML de λ est

$$\boxed{\left(\widehat{\lambda}_{n,1},\cdots,\widehat{\lambda}_{n,m}\right)=\underset{\lambda\in[0,1]^m}{\arg\max}\prod_{j=1}^{m}\lambda_j^{O_j}(1-\lambda_j)^{N_j-O_j}=\left(\frac{O_1}{N_1},\ldots,\frac{O_m}{N_m}\right).}$$

La valeur observée de l'estimateur NPML de la loi $\mathbb{P}_T$ est :

$$\widehat{\mathbb{P}}_{n,T}^{(ML)}(\cdot\mid(\mathbf{t}^\star,\mathbf{d}))=\sum_{j=1}^{m+1}\widehat{p}_{n,j}\delta_{t^\star_{(j)}},\tag{2.1}$$

avec $\boxed{\widehat{p}_{n,1}=\widehat{\lambda}_{n,1}}$ et pour $j=2,\cdots,m+1$

$$\boxed{\widehat{p}_{n,j}=\widehat{\lambda}_{n,j}\prod_{k=1}^{j-1}(1-\widehat{\lambda}_{n,k}).}$$

2.2.3 Estimateur de Kaplan-Meier

L'estimateur NPLM de la fonction de survie S de la durée de survie T est appelé l'estimateur de Kaplan-Meier. Il est donné pour tout $x>0$ par

$$\boxed{\widehat{S}_n^{(KM)}(x\mid(\mathbf{t}^\star,\mathbf{d}))=\prod_{j:t^\star_{(j)}\leq x}\left(1-\frac{O_j}{N_j}\right).}$$

Propriétés asymptotiques de l'estimateur de Kaplan-Meier

Un estimateur de la variance (asymptotique) de l'estimateur $\widehat{S}_n^{(KM)}(x)$ de Kaplan-Meier est l'estimateur de Greenwood

$$\boxed{\mathcal{V}_n^{(G)}(x):=\left[\widehat{S}_n^{(KM)}(x)\right]^2\sum_{j:T^\star_{(j)}\leq x}\frac{O_j(\mathbf{T}^\star,\mathbf{\Delta})}{N_j(\mathbf{T}^\star)(N_j(\mathbf{T}^\star)-O_j(\mathbf{T}^\star,\mathbf{\Delta}))}}$$

On peut montrer que $\mathcal{V}_n^{(G)}(x)$ est un estimateur consistant. De plus, sous certaines conditions

$$\boxed{\left[\mathcal{V}_n^{(G)}(x)\right]^{-1/2}\left(\widehat{S}_n^{(KM)}(x)-S(x)\right)\xrightarrow{d}\mathcal{N}(0,1).}$$

En posant $z_{1-\alpha/2}$ le quantile d'ordre $1-\alpha/2$ d'une loi $\mathcal{N}(0,1)$, on montre que

l'intervalle aléatoire $\mathrm{CI}_\alpha^{(KM)}(x)$ donné par

$$\left[\widehat{S}_n^{(KM)}(x) - z_{1-\alpha/2}\{\mathcal{V}_n^{(G)}(x)\}^{1/2}; \widehat{S}_n^{(KM)}(x) + z_{1-\alpha/2}\{\mathcal{V}_n^{(G)}(x)\}^{1/2} \right]$$

est tel que $\mathbb{P}[\mathrm{CI}_\alpha^{(KM)}(x) \ni S(x)] \to 1-\alpha$ lorsque $n \to \infty$ et ceci pour tout $x \in]0, t^\star_{(m)}[$.

2.2.4 Estimateur de Nelson-Aalen du risque cumulé

L'estimateur de Nelson-Aalen de la fonction de hasard cumulé de T est donnépour tout $x > 0$

$$\boxed{\widehat{A}_n(x \mid (\mathbf{t}^\star, \mathbf{d})) = \sum_{i: t^\star_{(i)} \leq x} \frac{O_i}{N_i}.}$$

Dans le cas où la loi de T est absolument continue, on sait que $\ln(S(t)) = -\Lambda(t)$. On en déduit donc un autre estimateur de la fonction de survie donné par

$$\boxed{\widehat{S}_n^{(NA)}(t) = \exp\left(-\widehat{A}_n(t)\right).}$$

Sous certaines conditions,

$$\left[\mathcal{V}_n^{(NA)}(x)\right]^{-1/2} \left(\widehat{A}_n(x) - \Lambda(x)\right) \xrightarrow{d} \mathcal{N}(0,1),$$

avec,

$$\mathcal{V}_n^{(NA)}(x) := \sum_{j: T^\star_{(j)} \leq x} \frac{O_j(\mathbf{T}^\star, \boldsymbol{\Delta})[N_j(\mathbf{T}^\star) - O_j(\mathbf{T}^\star, \boldsymbol{\Delta})]}{[N_j(\mathbf{T}^\star)]^3}.$$

Chapitre 3

Test de comparaison de fonctions de survie

Dans tout ce chapitre, nous nous plaçons sur un modèle de censure aléatoire et non informative à droite. L'objectif de ce chapitre est de comparer deux (ou plusieurs) fonctions de survie.

3.1 Test du log-rank : comparaison de 2 groupes

On se place dans le cadre suivant. Soient deux groupes (groupe 1 et groupe 2) d'individus. On supposera que les $n(1)$ individus du groupe 1 sont indépendants (différents) des $n(2)$ individus du groupe 2. On notera $n = n(1)+n(2)$ le nombre d'individus lorsqu'on fusionne les deux groupes. Dans le groupe $g \in \{1, 2\}$, on observe les réalisations des $n(g)$ vecteurs aléatoires

$$(\mathbf{T}^{(g),\star}, \mathbf{\Delta}^{(g)}) := \left\{ (T_1^{(g),\star}, \Delta_1^{(g)}), \cdots, (T_{n(g)}^{(g),\star}, \Delta_{n(g)}^{(g)}) \right\},$$

que l'on suppose indépendants et de même loi qu'un couple $(T^{(g),\star}, \Delta^{(g)})$ avec $T^{(g),\star} = \min(T^{(g)}, C^{(g)})$ et $\Delta^{(g)} = \mathbb{I}_{T^{(g)} \leq C^{(g)}}$, les variables aléatoires positives $T^{(g)}$ et $C^{(g)}$ étant supposées indépendantes. Comme précédemment, les observations dans le groupe $g \in \{1, 2\}$ seront notées

$$(\mathbf{t}^{(g),\star}, \mathbf{d}^{(g)}) = \left\{ (t_1^{(g),\star}, d_1^{(g)}), \cdots, (t_{n(g)}^{(g),\star}, d_{n(g)}^{(g)}) \right\}.$$

Les valeurs $\{t_{(1)}^{(g),\star} < \cdots < t_{(m(g))}^{(g),\star}\}$ sont les valeurs distinctes et ordonnées de l'ensemble $\{t_1^{(g),\star}, \cdots, t_{n(g)}^{(g),\star}\}$. L'objectif du test du log-rank est de tester l'égalité en loi des deux groupes. Les hypothèses nulle et alternative du test sont

$$H_0 \ : \ S^{(1)}(\cdot) = S^{(2)}(\cdot) \ \text{ contre } \ H_1 \ : \ S^{(1)}(\cdot) \neq S^{(2)}(\cdot),$$

où pour $g \in \{1,2\}$, $S^{(g)}(\cdot) = \mathbb{P}(T^{(g)} > \cdot)$. Sous l'hypothèse H_0, l'échantillon $(\mathbf{T}^\star, \mathbf{\Delta})$ de taille n est i.i.d. de loi commune la loi du couple aléatoire (T, Δ). On note comme d'habitude $(\mathbf{t}^\star, \mathbf{d}) = \{(t_1^\star, d_1), \cdots, (t_n^\star, d_n)\}$, les observations correspondantes.

Construction du test On note $\{t_{(1)}^\star < \cdots < t_{(m)}^\star\}$ les valeurs distinctes et ordonnées de $\{t_1^\star, \cdots, t_n^\star\}$. On considère également le vecteur $\Lambda^{(g)} := (\lambda_1^{(g)}, \cdots, \lambda_m^{(g)})$ avec pour $k \in \{1, \cdots, m\}$, $\lambda_k^{(g)} = \lambda^{(g)}(t_k^\star)$, avec $\lambda^{(g)}(\cdot)$ la fonction de hasard de la variable aléatoire $T^{(g)}$. On note $O_j := O_j^{(1)} + O_j^{(2)}$ le nombre d'événements observés dans la réunion des deux groupes à l'instant $t_{(j)}^\star$. De la même façon, $N_j = N_j^{(1)} + N_j^{(2)}$ est le nombre d'individus à risque dans la réunion des deux groupes à l'instant $t_{(j)}^\star$. On rappelle que pour $g \in \{1,2\}$, la valeur observée de l'estimateur du maximum de vraisemblance de $\mathbf{\Lambda}^{(g)}$ est

$$\widehat{\mathbf{\Lambda}}_n^{(g)} := \left(\widehat{\lambda}_{n,1}^{(g)}, \cdots, \widehat{\lambda}_{n,m}^{(g)}\right) = \left(\frac{O_1^{(g)}}{N_1^{(g)}}, \ldots, \frac{O_m^{(g)}}{N_m^{(g)}}\right).$$

Sous l'hypothèse nulle, on a $\Lambda^{(1)} = \Lambda^{(2)} =: \Lambda$ et, après fusion des deux groupes

$$\widehat{\mathbf{\Lambda}}_n := \left(\widehat{\lambda}_{n,1}, \cdots, \widehat{\lambda}_{n,m}\right) = \left(\frac{O_1}{N_1}, \ldots, \frac{O_m}{N_m}\right).$$

est la valeur observée de l'estimateur du maximum de vraisemblance de Λ. Ainsi, sous H_0, on s'attend à avoir

$$\sum_{k=1}^{m} W_{k,n}\left(\widehat{\lambda}_{n,k}^{(1)} - \widehat{\lambda}_{n,k}\right) \approx 0,$$

où $W_{1,n}, \cdots, W_{m,n}$ sont les observations de variables aléatoires positives qui pondèrent les écarts en fonction de l'instant d'observation. Pour le test classique du log-rank, les poids sont donnés pour tout $k \in \{1, \cdots, m\}$ par $W_{k,n} = N_k^{(1)}$. Le test du log-rank est donc basé sur la variable aléatoire $K(\mathbf{T}^\star, \mathbf{\Delta})$ dont la

valeur observée est

$$K := \sum_{k=1}^{m} \left(O_k^{(1)} - N_k^{(1)} \frac{O_k}{N_k} \right).$$

Si on n'est plus sous H_0, le comportement de K peut être de 2 types.

i) Si on a $\widehat{\lambda}_{n,k}^{(1)} < \widehat{\lambda}_{n,k}^{(2)}$ pour tout k (ou bien $\widehat{\lambda}_{n,k}^{(1)} > \widehat{\lambda}_{n,k}^{(2)}$ pour tout k) (autrement dit si les courbes des estimateurs de Kaplan-Meier ne se croisent pas), alors, sous H_1 on s'attend à observer pour K une valeur très différente de 0. Cette variable aléatoire permet donc dans ce cas de conclure sur la plausibilité de H_0.

ii) Dans le cas où les valeurs de $\widehat{\lambda}_{n,k}^{(1)}$ et $\widehat{\lambda}_{n,k}^{(2)}$ se croisent, on pourra observer une valeur de K proche de 0 sous H_1. Le test sera dans ce cas moins puissant puisque le risque d'accepter H_0 à tort est plus important (ou de manière équivalente, l'acceptation de H_1 à raison sera moins fréquente).

Pour décider si la valeur observée de la variable aléatoire $K(\mathbf{T}^\star, \mathbf{\Delta})$ est suffisamment éloignée de 0 pour conclure à la non validité de H_0, il faut connaître le comportement attendu sous H_0 ou de manière plus formelle la loi asymptotique de $K(\mathbf{T}^\star, \mathbf{\Delta})$ sous H_0. On introduit la statistique $\mathrm{LR}(\mathbf{T}^\star, \mathbf{\Delta})$ dont la valeur observée est

$$\boxed{\mathrm{LR}(\mathbf{t}^\star, \mathbf{d}) := \frac{1}{\mathcal{V}_n^{(LR)}} \left[\sum_{k=1}^{m} \left(O_k^{(1)} - N_k^{(1)} \frac{O_k}{N_k} \right) \right]^2,}$$

avec

$$\mathcal{V}_n^{(LR)} := \sum_{k=1}^{m} \frac{N_k^{(1)} N_k^{(2)}}{N_k} \frac{O_k}{N_k} \left(1 - \frac{O_k}{N_k} \right).$$

On dispose du résultat suivant.

> Sous certaines hypothèses, notamment si les lois des couples $(T^{(1)}, C^{(1)})$ et $(T^{(2)}, C^{(2)})$ sont indépendantes et s'il existe $\zeta \in]0,1[$ tel que $n(1)/n \to \zeta$ lorsque $n \to \infty$ alors, sous l'hypothèse nulle H_0, la loi de la statistique $\mathrm{LR}(\mathbf{T}^\star, \mathbf{\Delta})$ converge vers une loi du khi-deux à 1 degré de liberté.

On peut à présent mettre en place la stratégie du test du log-rank. On compare la valeur observée $\mathrm{LR}(\mathbf{t}^\star, \mathbf{d})$ de la statistique de test au quantile d'ordre $1-\alpha$ d'une loi du khi-deux à 1 degré de liberté, α étant l'erreur de première espèce.

3.2 Test du log-rank pour plusieurs groupes

Nous nous contentons de donner la statistique de test pour la comparaison de trois groupes mais elle se généralise très facilement pour la comparaison de plus de trois groupes. Les hypothèses nulles et alternatives sont donc

$$H_0 \ : \ S^{(1)}(\cdot) = S^{(2)}(\cdot) = S^{(3)}(\cdot) \ \text{ contre } \ H_1 \ : \ \exists i \neq j \text{ t.q. } S^{(i)}(\cdot) \neq S^{(j)}(\cdot).$$

On supposera toujours que les individus des trois groupes sont différents (groupes indépendants). On utilise les mêmes notations que dans le paragraphe précédent. De plus, on pose $\mathbf{K} := (K^{(1)}, K^{(2)})^\top$ avec, pour $g \in \{1, 2\}$,

$$\boxed{K^{(g)} := K^{(g)}(\mathbf{t}^\star, \mathbf{d}) := \sum_{k=1}^{m} \left(O_k^{(g)} - N_k^{(g)} \frac{O_k}{N_k} \right),}$$

où l'on rappelle que,

$$N_k = \sum_{g=1}^{3} N_k^{(g)} \text{ et } O_k = \sum_{g=1}^{3} O_k^{(g)}.$$

Pour toute matrice V symétrique et inversible de dimension 2×2, on note $\|\cdot\|_V$ la norme de $\mathbb{R}^2$ définie pour tout $x \in \mathbb{R}^2$ par $\|x\|_V = x^\top V x$. Sous l'hypothèse nulle H_0, on s'attend à ce que $K^{(1)}(\mathbf{t}^\star, \mathbf{d}) \approx K^{(2)}(\mathbf{t}^\star, \mathbf{d}) \approx 0$ et donc à $\|\mathbf{K}\|_V \approx 0$. Il s'agit à présent de trouver la « bonne » matrice V qui nous permettra de connaître la loi asymptotique de cette norme sous H_0. Avant d'énoncer le résultat, introduisons les notations suivantes. Comme nous l'avons vu dans le paragraphe précédent, pour $g \in \{1, 2\}$, la valeur observée de l'estimateur de la variance asymptotique de $K^{(g)}(\mathbf{T}^\star, \boldsymbol{\Delta})$ est

$$\mathcal{V}_n^{(g,LR)} := \sum_{k=1}^{m} \frac{N_k^{(g)} N_k^{(-g)}}{N_k} \frac{O_k}{N_k} \left(1 - \frac{O_k}{N_k} \right),$$

où $N_k^{(-g)} = N_k - N_k^{(g)}$. La valeur observée de l'estimateur de la covariance asymptotique entre les variables aléatoires $K^{(1)}(\mathbf{T}^\star, \boldsymbol{\Delta})$ et $K^{(2)}(\mathbf{T}^\star, \boldsymbol{\Delta})$ est

$$\mathcal{C}_n^{(LR)} := -\sum_{k=1}^{m} \frac{N_k^{(1)} N_k^{(2)}}{N_k} \frac{O_k)}{N_k} \left(1 - \frac{O_k}{N_k} \right).$$

On introduit la statistique $\mathrm{LR}(\mathbf{T}^\star, \mathbf{\Delta})$ dont la valeur observée est donnée par

$$\boxed{\mathrm{LR}(\mathbf{t}^\star, \mathbf{d}) := \left(K^{(1)}, K^{(2)}\right) \begin{pmatrix} \mathcal{V}_n^{(1,LR)} & \mathcal{C}_n^{(LR)} \\ \mathcal{C}_n^{(LR)} & \mathcal{V}_n^{(2,LR)} \end{pmatrix}^{-1} \begin{pmatrix} K^{(1)} \\ K^{(2)} \end{pmatrix}.}$$

La loi asymptotique de cette statistique de test est donnée dans le résultat suivant.

Sous certaines hypothèses, notamment si les lois des couples $(T^{(g)}, C^{(g)})$, $g \in \{1,2,3\}$ sont indépendantes et s'il existe $(\zeta(1), \zeta(2), \zeta(3)) \in]0,1[^3$ avec $\zeta(1) + \zeta(2) + \zeta(3) = 1$ et tel que pour tout $g \in \{1,2,3\}$, $n(g)/n \to \zeta(g)$ lorsque $n \to \infty$ alors, sous l'hypothèse nulle H_0, la loi de la statistique $\mathrm{LR}(\mathbf{T}^\star, \mathbf{\Delta})$ converge vers une loi du khi-deux à deux degrés de liberté.

Chapitre 4

Estimation de la durée de survie en présence de covariables

On observe maintenant les réalisations de n copies indépendantes du vecteur aléatoire $(T^\star, \Delta, X)$ où $T^\star = \min(T, C)$, $\Delta = \mathbb{I}_{T \leq C}$ avec T et C indépendantes conditionnellement à X i.e., pour tout $y > 0$ et $z > 0$,

$$\mathbb{P}(\{T \leq y\} \cap \{C \leq z\} \mid X) = \mathbb{P}(T \leq y \mid X)\mathbb{P}(C \leq z \mid X), \text{ p.s.}$$

La principale différence avec ce qui a été fait précédemment est que l'on souhaite à présent estimer la loi de T sachant que $X = x$ pour tout $x \in \mathcal{S}$ en utilisant les observations $(\mathbf{t}^\star, \mathbf{d}, \mathbf{x}) := \{(t_i^\star, d_i, x_i),\ i = 1, \cdots, n\}$.

4.1 Maximum de vraisemblance non paramétrique

Notons $\mathbb{P}_{T|X}$ la loi conditionnelle de T sachant X. La vraisemblance de ce modèle est donnée pour $h > 0$ par

$$\boxed{L_h(\mathbb{P}_{T|X} \mid (\mathbf{t}^\star, \mathbf{d}, \mathbf{x})) = \prod_{i=1}^{n} \mathbb{P}\left(\{t_i^\star - h < T^\star \ \leq t_i^\star\} \cap \{\Delta = d_i\} | \, X = x_i\right) f_X(x_i),}$$

avec $f_X(\cdot)$ la densité de X (éventuellement par rapport à une mesure de comptage). On note $t^\star_{(1)} < \cdots < t^\star_{(m)}$ les valeurs distinctes et ordonnées de $\mathbf{t}^\star$. On va

chercher le maximum de la vraisemblance non paramétrique sur le sous-ensemble des lois de probabilité

$$\mathcal{P} := \Bigg\{ \tilde{\mathbb{P}}(\cdot \mid X) = \tilde{\mathbb{P}}_{T|X} \otimes \tilde{\mathbb{P}}_{C|X} : \mathcal{A} \otimes \mathcal{A} \to [0,1] \\ \text{avec } \tilde{\mathbb{P}}(\cdot \mid X) = \sum_{j=1}^{m+1} p_j(X)\delta_{t^\star_{(j)}} \otimes \sum_{j=1}^{m} q_j(X)\delta_{t^\star_{(j)}} \Bigg\},$$

où pour tout $x \in \mathcal{S}$, $p_1(x), \cdots, p_{m+1}(x), q_1(x), \cdots, q_m(x) > 0$ avec la contrainte

$$\sum_{i=1}^{m+1} p_i(x) = \sum_{i=1}^{m} q_i(x) = 1.$$

Ainsi, pour tout $\tilde{\mathbb{P}}(\cdot \mid X) \in \mathcal{P}$, la vraisemblance non paramétrique s'écrit

$$\boxed{L_h(\tilde{\mathbb{P}}_{T|X} \mid (\mathbf{t}^\star, \mathbf{d}, \mathbf{x})) = K \prod_{j=1}^{m} \prod_{i:t_i^\star = t^\star_{(j)}} [\, p_j(x_i)]^{d_i} \left[\sum_{k=j+1}^{m+1} p_k(x_i) \right]^{1-d_i},}$$

où K est une quantité ne dépendant pas des $p_j(x_i)$. Comme nous l'avons fait pour obtenir l'estimateur de Kaplan-Meier, nous effectuons à présent le changement de variable

$$\lambda_j(x_i) = p_j(x_i) \Big/ \sum_{k=j}^{m+1} p_k(x_i)\ ,$$

pour tout $j = 1, \cdots, m+1$ et $i = 1, \cdots, n$. On obtient alors l'expression suivante pour la vraisemblance (qui ne dépend plus de h).

$$\boxed{L_h(\tilde{\mathbb{P}}_{T|X} \mid (\mathbf{t}^\star, \mathbf{d}, \mathbf{x})) = K \prod_{j=1}^{m} \left\{ \prod_{i:t_i^\star = t^\star_{(j)}} \left[\frac{\lambda_j(x_i)}{1-\lambda_j(x_i)} \right]^{d_i} \times \prod_{i:t_i^\star \geq t^\star_{(j)}} [1 - \lambda_j(x_i)] \right\}.}$$

Il est inutile d'essayer de maximiser cette vraisemblance pour obtenir un estimateur des fonctions $\lambda_j(\cdot)$ pour $j = 1, \cdots, m$. Tout d'abord, la vraisemblance ne dépend que des quantités $\lambda_j(x_i)$ pour tous les couples (i,j) tels que $t_i^\star \geq t^\star_{(j)}$. Il sera donc impossible de pouvoir estimer par la méthode NPML la valeur de $\lambda_j(x)$ pour tout $x \in \mathcal{S}$. Enfin, on peut remarquer facilement que si $t_i^\star = t^\star_{(j)}$ alors l'estimateur NPML de $\lambda_j(x_i)$ est $d_i \in \{0,1\}$ et si $t_i^\star > t^\star_{(j)}$ alors l'estimateur

NPML de $\lambda_j(x_i)$ est égal à 0.

4.2 Modèles paramétriques

Une solution pour estimer les fonctions de hasard sur l'ensemble du support $\mathcal{S} \subset \mathbb{R}^p$ de X est d'imposer un modèle paramétrique à cette fonction. Autrement dit, nous supposerons qu'il existe un vecteur de paramètres $\theta \in \Theta \subset \mathbb{R}^q$, $q \in \mathbb{N}\backslash\{0\}$ et une fonction $g(\cdot; x, \theta)$ (positive et dont l'expression est connue) tels que $\lambda(t \mid x) = g(t; x, \theta)$, où $\lambda(t \mid x)$ est la fonction de hasard de la loi conditionnelle de T sachant $X = x$ à l'instant t. Pour estimer la fonction de hasard (et donc la loi conditionnelle de T sachant $X = x$) pour tout $x \in \mathcal{S}$, il suffit donc d'estimer le paramètre $\theta \in \Theta$. Pour ce faire, la méthode du maximum de vraisemblance est ici bien adaptée. Nous considérons dans la suite deux cas pour la loi du couple (T, C).

4.2.1 Cas discret

Nous nous plaçons sous l'hypothèse que la loi conditionnelle du couple (T, C) sachant X est discrète à support fini. Nous avons montré en (4.1) que, si les lois de C et de X ne dépendent pas du paramètre θ, la vraisemblance s'écrit

$$L(\theta \mid (\mathbf{t}^\star, \mathbf{d}, \mathbf{x})) = K \prod_{j=1}^{m} \left\{ \prod_{i:t_i^\star = t_{(j)}^\star} \left[\frac{g(t_{(j)}^\star; x_i, \theta)}{1 - g(t_{(j)}^\star; x_i, \theta)} \right]^{d_i} \times \prod_{i:t_i^\star \geq t_{(j)}^\star} [1 - g(t_{(j)}^\star; x_i, \theta)] \right\},$$

où la constante K ne dépend pas de θ et où $g(t; x, \theta) \in [0, 1]$ pour tout $(t, x, \theta) \in]0, \infty[\times\mathcal{S} \times \Theta$. L'estimateur du maximum de vraisemblance (s'il existe) est donc la variable aléatoire donnée par

$$\widehat{\theta}_n := \arg\max_{\theta \in \Theta} L\left(\theta \mid (\mathbf{t}^\star, \mathbf{d}, \mathbf{x})\right).$$

Ainsi, pour tout $j \in \{1, \cdots, m\}$ et $x \in \mathcal{S}$, on estime la fonction de hasard à l'instant $t_{(j)}^\star$ par l'estimateur du maximum de vraisemblance $\widehat{\lambda}_{n,j}(x) = g(t_{(j)}^\star; x, \widehat{\theta}_n)$ et donc la fonction de survie par

$$\widehat{S}_n(t \mid x) := \prod_{j:t_{(j)}^\star \leq t} \left(1 - \widehat{\lambda}_{n,j}(x)\right).$$

4.2.2 Cas absolument continu

Dans le cas où la loi du vecteur aléatoire (T, C) est absolument continue, il faut estimer la fonction de hasard

$$\lambda(t \mid x) := \frac{f(t \mid x)}{S(t \mid x)},$$

pour tout $t > 0$ et $x \in \mathcal{S}$ et où $f(\cdot \mid x)$ et $S(\cdot \mid x)$ sont respectivement la densité et la fonction de survie de la loi de T conditionnellement à $X = x$. Noter que dans ce cas, avec une probabilité égale à 1, les observations $t_1^\star, \cdots, t_n^\star$ sont distinctes. On montre que la vraisemblance du modèle s'écrit

$$\boxed{K \prod_{i=1}^{n} \left[\lambda(t_i^\star \mid x_i)\right]^{d_i} \exp\left(-\int_0^{t_i^\star} \lambda(z \mid x_i) dz\right),}$$

où K est une constante ne dépendant que des lois de C et X. Ainsi, sous le modèle paramétrique $\lambda(t \mid x) = g(t; x, \theta)$ où $g(\cdot; x, \theta)$ est une fonction positive, l'estimateur du maximum de vraisemblance de θ est (s'il existe)

$$\widehat{\theta}_n := \arg\max_{\theta \in \Theta} \prod_{i=1}^{n} \left[g(t_i^\star; x_i, \theta)\right]^{d_i} \exp\left(-\int_0^{t_i^\star} g(z; x_i, \theta) dz\right).$$

L'estimateur du maximum de vraisemblance de la fonction de survie conditionnelle de T sachant $X = x$ est alors

$$\widehat{S}_n(t \mid x) = \exp\left(-\int_0^{t} g(z; x_i, \widehat{\theta}_n) dz\right).$$

4.2.3 Comportement asymptotique et tests d'hypothèses

On profite des bonnes propriétés du maximum de vraisemblance pour établir la loi asymptotique de $\widehat{\theta}_n$ et effectuer des tests d'hypothèses. Sous les hypothèses de régularité requises, on a

$$\left[\mathrm{I}_n(\theta)\right]^{1/2} \left(\widehat{\theta}_n - \theta\right) \xrightarrow{d} \mathcal{N}\left(0, I_q\right),$$

où I_q est la matrice identité de dimension $q \times q$ et $\mathrm{I}_n(\theta)$ est la matrice d'information de Fischer (de dimension $q \times q$).

Test de Wald Il est possible de tester l'hypothèse nulle $H_0 : \theta_k = \theta_0 \in \mathbb{R}$ où θ_k est la k-ième composante du vecteur θ contre l'alternative $H_1 : \theta_k \neq \theta_0$ (ou

$H_1 : \theta_k > \theta_0$ ou $H_1 : \theta_k < \theta_0$) en utilisant la statistique de Wald définie par

$$W_{k,n} := \frac{\widehat{\theta}_{n,k} - \theta_0}{\widehat{\sigma}_{n,k}},$$

où $\widehat{\sigma}^2_{n,k} = [\mathrm{I}_n(\widehat{\theta}_n)]_{k,k}$ est un estimateur consistant de la variance (asymptotique) de $\widehat{\theta}_{n,k}$. Sous l'hypothèse H_0, la statistique $W_{k,n}$ converge en loi vers une loi $\mathcal{N}(0,1)$ ce qui nous permet de mettre en place le test de Wald.

Test du rapport de vraisemblance De manière plus générale, on peut également proposer un test de l'hypothèse nulle $H_0 : \theta \in \Theta_0$ contre l'alternative $H_1 : \theta \notin \Theta_0$ où $\Theta_0 \subset \Theta$. On note $m < q$ le nombre de paramètres contraints dans l'ensemble Θ_0. La statistique de test de H_0 contre H_1 est la statistique du rapport de vraisemblance donnée par

$$\mathrm{RV}_n = -2 \log \frac{L\left(\widehat{\theta}_{n,0} \mid (\mathbf{T}^\star, \boldsymbol{\Delta}, X)\right)}{L\left(\widehat{\theta}_n \mid (\mathbf{T}^\star, \boldsymbol{\Delta}, X)\right)},$$

avec

$$\widehat{\theta}_{n,0} := \arg\max_{\theta \in \Theta_0} L\left(\theta \mid (\mathbf{T}^\star, \boldsymbol{\Delta}, X)\right).$$

On sait que sous H_0, la loi de la statistique RV_n converge vers une loi du khi-deux à m degrés de liberté.

4.2.4 Quelques exemples de modèles paramétriques

Loi discrète Modèle logistique :

$$\boxed{\lambda(t \mid x) := g(t; x, \beta) = \frac{\exp(\beta^\top x)}{1 + \exp(\beta^\top x)},}$$

où $\beta \in \mathbb{R}^p$ est le vecteur des paramètres.

Loi absolument continue Modèle exponentiel :

$$\boxed{\lambda(t \mid x) := g(t; x, \beta) = \exp(\beta^\top x),}$$

où $\beta \in \mathbb{R}^p$ est le vecteur des paramètres. Il correspond au cas où la fonction de hasard (risque instantané) est constante au cours du temps.

4.3 Modèles semi-paramétrique

4.3.1 Modèles à risque proportionnel

Une fonction de survie conditionnelle $S(\cdot \mid X)$ est dite à risques proportionnels s'il existe une fonction de survie $S_0(\cdot)$ et une fonction positive $\rho(\cdot \mid \theta)$ avec $\theta \in \Theta \in \mathbb{R}^q$ telle que pour tout $x \in \mathcal{S}$,

$$\boxed{S(t \mid x) = [S_0(t)]^{\rho(x;\theta)}.}$$

Dans le cas où la loi conditionnelle de T sachant $X = x$ est absolument continue, l'égalitéci-dessus est équivalente à la condition sur la fonction de hasard :

$$\boxed{\lambda(t \mid x) = \lambda_0(t) \times \rho(x;\theta).}$$

Dans le cas où la loi conditionnelle de T sachant X est discrète, elle est équivalente à dire que pour tout t dans le support de T

$$\boxed{1 - \lambda(t \mid x) = (1 - \lambda_0(t))^{\rho(x;\theta)}.}$$

Loi absolument continue – Le modèle de Cox

Le choix le plus courant pour la fonction paramétrique est $g(\cdot \mid \beta) = \exp(\beta^\top \cdot)$, avec $\beta \in \mathbb{R}^p$. Ceci nous conduit au modèle introduit par Cox en 1972. La fonction de hasard conditionnelle de T sachant $X = x$ est donnée par

$$\boxed{\lambda(t \mid x) = \lambda_0(t) \exp\left(\beta^\top x\right)}.$$

Estimation du modèle de Cox La vraisemblance est proportionnelle à

$$\boxed{\prod_{i=1}^{n} \left[\lambda_0(t_i^\star) \exp(\beta^\top x_i)\right]^{d_i} \exp\left(-\exp(\beta^\top x_i)\Lambda_0(t_i^\star)\right).}$$

où $\Lambda_0(\cdot)$ est la fonction de hasard cumulé. Nous allons chercher le maximum sur l'ensemble des fonctions $\lambda_0(\cdot)$ telles que la fonction de hasard cumulé est

$$\Lambda_0(t) = \sum_{j:t_{(j)}^\star \leq t} \lambda_{0,j}.$$

On a en particulier que $\lambda_{0,j} = \lambda_0(t_{(j)}^\star)$. On note $x_{(i)}$ et $d_{(i)}$ les observations de X et Δ associées à l'"observation $t_{(i)}^\star$. En posant $\ell_0 := \{\lambda_{0,1}, \cdots, \lambda_{0,n}\}$, la

vraisemblance s'écrit

$$\boxed{\overline{L}(\ell_0, \beta \mid (\mathbf{t}^\star, \mathbf{d}, \mathbf{x})) := \prod_{i=1}^{n} \left[\lambda_{0,i} \exp(\beta^\top x_{(i)})\right]^{d_{(i)}} \prod_{j=i}^{n} \exp\left\{-\lambda_{0,i} \exp(\beta^\top x_{(j)})\right\}.}$$

Posons

$$r_j(\beta) := r_j(\beta, \mathbf{t}^\star, \mathbf{x}) := \sum_{i: t_i^\star \geq t_{(j)}^\star} \exp(\beta^\top x_i) = \sum_{i=j}^{n} \exp(\beta^\top x_{(i)}).$$

L'estimateur est donné par

$$\left(\widehat{\ell}_{0,n}, \widehat{\beta}_n\right) := \underset{(\ell_0, \beta)}{\arg\max}\, \overline{L}(\ell_0, \beta \mid (\mathbf{t}^\star, \mathbf{d}, \mathbf{x})).$$

avec

$$\boxed{\widehat{\beta}_n = \underset{\beta \in \mathbb{R}^p}{\arg\max} \prod_{j=1}^{n} \left[\frac{\exp(\beta^\top x_{(j)})}{r_j(\beta)}\right]^{d_{(j)}}}$$

et $\widehat{\ell}_{0,n} = (\widehat{\lambda}_{0,j},\ j = 1, \cdots, n)$ avec $\boxed{\widehat{\lambda}_{0,j} = d_{(j)}/r_j(\widehat{\beta}_n)}$. La valeur observée de l'estimateur de la fonction de survie conditionnelle est

$$\widehat{S}_n(t \mid x) = \exp\left[\ -\exp\left(\widehat{\beta}_n^\top x\right) \sum_{i: t_{(i)}^\star \leq t} \widehat{\lambda}_{0,i}\right].$$

Notons pour finir que l'on peut mettre en place les tests d'hypothèses de Wald et du rapport de vraisemblance sur le paramètre β.

Loi discrète – Présence d'ex æquo

Il existe différentes solution pour traiter le cas des ex æquo.

Modèle logistique On va utiliser la vraisemblance du cas d'une loi discrète à support de cardinal fini. On note comme d'habitude $t_{(1)}^\star < \cdots < t_{(m)}^\star$ avec $m < n$ les observations distinctes et ordonnées. L'équivalent du modèle de Cox est le modèle logistique donné par

$$\boxed{\frac{\lambda(t \mid x)}{1 - \lambda(t \mid x)} = \beta_0(t) \exp(\beta^\top x),}$$

où $\beta \in \mathbb{R}^p$ et $\beta_0(\cdot)$ est une fonction positive inconnue. On pose $\mathbf{b}_0 := \{\beta_{0,j} := \beta_0(t^\star_{(j)}),\ j = 1, \cdots, m\}$. La vraisemblance s'écrit

$$L(\mathbf{b}_0, \beta \mid (\mathbf{t}^\star, \mathbf{d}, \mathbf{x})) = K \prod_{j=1}^{m} \left\{ \prod_{i:t_i^\star = t^\star_{(j)}} \left[\ \beta_{0,j} \exp(\beta^\top x_i)\right]^{d_i} \right.$$
$$\left. \times \prod_{i:t_i^\star \geq t^\star_{(j)}} \left[1 + \beta_{0,j} \exp(\beta^\top x_i)\right]^{-1} \right\},$$

où la constante K ne dépend pas de $\mathbf{b}_0$ et β. L'estimateur du maximum de vraisemblance est donné par

$$\left(\widehat{\mathbf{b}}_{0,n}, \widehat{\beta}_n\right) := \underset{(\mathbf{b}_0, \beta)}{\arg\max}\, L\left(\mathbf{b}_0, \beta \mid (\mathbf{t}^\star, \mathbf{t}, \mathbf{x})\right).$$

Approximation de Breslow Elle se base sur le fait que si $\beta_{0,j} \exp(\beta^\top x_i) \approx 0$ alors $[1 + \beta_{0,j} \exp(\beta^\top x_i)]^{-1} \approx \exp(-\beta_{0,j} \exp(\beta^\top x_i))$.
Il en découle que $L(\mathbf{b}_0, \beta \mid (\mathbf{t}^\star, \mathbf{d}, \mathbf{x})) \approx \widetilde{L}(\mathbf{b}_0, \beta \mid (\mathbf{t}^\star, \mathbf{d}, \mathbf{x}))$ avec

$$\widetilde{L}(\mathbf{b}_0, \beta \mid (\mathbf{t}^\star, \mathbf{d}, \mathbf{x})) = K \prod_{j=1}^{m} \left\{ \prod_{i:t_i^\star = t^\star_{(j)}} \left[\ \beta_{0,j} \exp(\beta^\top x_i)\right]^{d_i} \right.$$
$$\left. \times \prod_{i:t_i^\star \geq t^\star_{(j)}} \exp\left[-\beta_{0,j} \exp(\beta^\top x_i)\right] \right\},$$

Posons O_j le nombre d'événements observés à l'instant $t^\star_{(j)}$ et

$$r_j(\beta) = r_j(\beta, \mathbf{t}^\star, \mathbf{x}) := \sum_{i:t_i^\star \geq t^\star_{(j)}} \exp(\beta^\top x_i).$$

Les paramètres $(\mathbf{b}_0, \beta)$ sont estimés par

$$\left(\widetilde{\mathbf{b}}_{0,n}, \widetilde{\beta}_n\right) := \underset{(\mathbf{b}_0, \beta)}{\arg\max}\, \widetilde{L}\left(\mathbf{b}_0, \beta \mid (\mathbf{t}^\star, \mathbf{t}, \mathbf{x})\right).$$

avec

$$\boxed{\widetilde{\beta}_n = \underset{\beta \in \mathbb{R}^p}{\arg\max} \prod_{j=1}^{m} \prod_{i:t_i^\star = t^\star_{(j)}} \left[\ \frac{\exp(\beta^\top x_i)}{r_j(\beta)}\right]^{d_i}.}$$

et $\widetilde{\mathbf{b}}_{0,n} = (\widetilde{\beta}_{0,j},\ j = 1, \cdots, m)$ avec $\widetilde{\beta}_{0,j} = O_j / r_j(\widetilde{\beta}_n)$. La fonction de survie est estimée par

$$\tilde{S}_n(t \mid x) = \exp\left[\ - \sum_{i:t^\star_{(i)} \leq t} \frac{\widetilde{\beta}_{0,i} \exp(\widetilde{\beta}_n^\top x)}{1 + \widetilde{\beta}_{0,i} \exp(\widetilde{\beta}_n^\top x)} \right].$$

Remarquons pour finir que les estimateurs $\widehat{\beta}_n$ et $\widetilde{\beta}_n$ sont identiques dans le cas où il n'y a aucun ex æquo.

Modèle « Complementary log-log »

La fonction de hasard conditionnelle de T sachant $X = x$ suit un modèle « complementary log log » s'il existe une fonction $\beta_0(\cdot)$ et $\beta \in \mathbb{R}^p$ tels que pour tout $x \in \mathcal{S}$ et t dans le support de T,

$$\boxed{\lambda(t \mid x) = 1 - \exp\left[\ - \exp\left(\beta_0(t) + \beta^\top x\right)\right].}$$

Pour l'estimation de ce modèle, on supposera que pour tout $t \in [t^\star_{(i)}, t^\star_{(i+1)}[$, la fonction $\beta_0(t)$ est constante et égale à $\beta_{0,i}$. On estime le paramètre β et les coefficients $\beta_{0,i}$ en utilisant la fonction de vraisemblance (4.1).

4.3.2 Analyse des résidus du modèle de Cox

Pour le modèle de Cox, la validité des estimateurs repose sur l'hypothèse forte que la fonction de hasard de la loi conditionnelle de T sachant $X = x$ est de la forme $\lambda(t \mid x) = \lambda_0(t) \exp(\beta^\top x)$. Nous présentons ci-après deux méthodes graphiques permettant de vérifier cette hypothèse et qui sont utilisables lorsqu'il n'y a pas d'ex aequo.

Loi exponentielle censurée Cette méthode est basée sur le résultat suivant.

> Si Z est une variable aléatoire positive absolument continue de fonction de survie $S(\cdot)$ strictement décroissante. La variable aléatoire $\Lambda(Z)$, où $\Lambda(\cdot)$ est la fonction de hasard cumulé, suit une loi exponentielle de paramètre 1.

En notant $\Lambda(\cdot \mid x)$ la fonction de hasard cumulé de la loi conditionnelle de T sachant $X = x$, on peut voir les couples $\{(\Lambda(t_i^\star \mid x_i), d_i)\,,\ i = 1, \cdots, n\}$, comme les réalisations d'un couple $(E^\star, \Delta)$ avec $E^\star = \min(E_1, E_2)$, E_1 étant une variable aléatoire de loi exponentielle de paramètre 1 et E_2 une variable aléatoire positive indépendante de E_1. Ainsi, si on applique l'estimateur de Nelson-Aalen

aux couples $\{(\Lambda(t_i^\star \mid x_i), d_i)\,,\ i = 1, \cdots, n\}$, on doit estimer la fonction de hasard cumulé d'une loi exponentielle. La fonction $\Lambda(\cdot \mid x)$ est inconnue mais, si le modèle de Cox est valide, on est censé l'estimer correctement par

$$\widehat{A}_n(t \mid x) := \exp(\widehat{\beta}_n^\top x) \sum_{k:t_{(k)}^\star \leq t} \widehat{\lambda}_{0,k}.$$

Ceci nous conduit à calculer l'estimateur de Nelson-Aalen en utilisant les points

$$\left\{ \left(\widehat{A}_n(t_i^\star \mid x_i), d_i \right) ;\ i = 1, \cdots, n \right\},$$

et à le comparer avec la première bissectrice. Des points éloignés de cette droite suggèreront que le modèle de Cox n'est vraisemblablement pas valide.

Résidus de martingale Le second graphique permettant de vérifier la validité du modèle de Cox est basé sur le résultat suivant.

> Soit la variable aléatoire $T^\star := \min(T, C)$ où T et C sont des variables aléatoires positives et indépendantes. En notant $\Lambda(\cdot)$ la fonction de hasard cumulé de T on a $\mathbb{E}[\Lambda(T^\star)] = \mathbb{P}(T \leq C)$.

La définition des résidus de martingale découle de ce résultat en remarquant que pour tout $i = 1, \cdots, n$, $\mathbb{E}\left[\,\Delta_i - \Lambda(T_i^\star \mid X_i) | \, X_i\right] = 0$. En remplaçant la fonction de hasard cumulé par son estimateur sous le modèle de Cox, on définit les résidus de martingale par

$$\boxed{\mathrm{MR}_i := d_i - \exp\left(\widehat{\beta}_n^\top x_i\right) \sum_{j:t_{(j)}^\star \leq t_i^\star} \widehat{\lambda}_{0,j}.}$$

On représente sur une graphique les points $\{(t_i^\star, \mathrm{MR}_i);\ i = 1, \cdots, n\}$. Si le modèle de Cox est valable, les points doivent être situés aux alentours de la droite horizontale d'équation $y = 0$.