

Le modèle de Cox fait partie de la famille des modèles semi-paramétriques. A partir de là, il comporte une composante complètement non-paramétrique (de type fonctionnelle) et une composante purement paramétrique (de forme paramétrique exponentielle). C'est le modèle d'analyse de données de survie le plus utilisé dans le cadre de l'étude d'impact de covariables sur la durée de vie.

### I - Le cadre des modèles à hazards proportionnels -

On introduit dans ces modèles un effet multiplicatif des covariables sur le taux de hasard, d'où le nom de "hazards proportionnels".

Ces modèles comportent un taux de hasard de base qui donne la forme générale des taux de hasard, commun à tous les individus de la population. Ils se caractérisent ensuite par une relation du type :  $\forall t > 0$ ,

$$\boxed{\lambda(t|z) = \lambda_0(t) h(\beta, z)} \quad , \text{ avec } z \text{ un vecteur de } \left\{ \begin{array}{l} \text{covariables} \\ \text{caractéristiques} \\ \text{facteurs de risque} \end{array} \right.$$

On note que la fonction  $h$ , d'ajustement du risque, est indépendante du temps. C'est une fonction positive qui permet de tenir compte des caractéristiques spécifiques aux individus et d'ajuster ainsi le taux de hasard.

Pour simplifier, on suppose en général que  $h(\beta, z)$  s'écrit  $h(\beta^T z)$ , ce qui revient à dire que l'impact des covariables se résume à un réel, donné par la combinaison



linéaire des paramètres avec les covariables.

Ce type de modèle est dit à risques proportionnels car si l'on prend deux individus  $i$  et  $j$  avec covariables  $z_i$  et  $z_j$ , le rapport de leurs fonctions de hasard est  $\perp$  du temps (car les hasards de base se simplifient!):

$$\frac{\lambda(t|z_i)}{\lambda(t|z_j)} = \frac{\lambda_0(t) h(\beta^T z_i)}{\lambda_0(t) h(\beta^T z_j)} = \frac{h(\beta^T z_i)}{h(\beta^T z_j)}$$

Ainsi  $\lambda(t|z_i) = \underbrace{\frac{h(\beta^T z_i)}{h(\beta^T z_j)}}_{\in \mathbb{R}} \lambda(t|z_j) \Rightarrow$  les 2 risques sont proportionnels, et le facteur de proportionnalité ne dépend pas de  $t$ .

→ Cela signifie qu'il faudra vérifier cette hypothèse par une bonne utilisation de cette classe de modèle!

→ Si la fonction  $\lambda_0$  et/ou  $h$  ont une forme inconnue, on parle donc de modèle semi-paramétrique.

## II - Le modèle de Cox

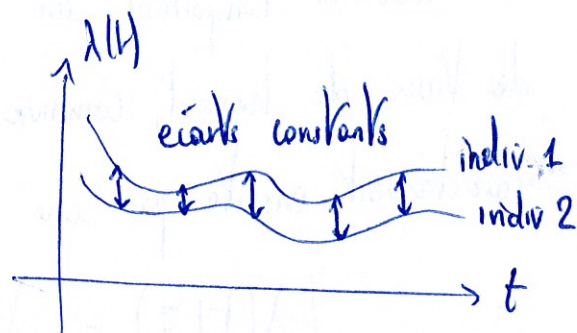
### ① - Définition

Dans le modèle de Cox, la fonction  $h$  est la fonction exponentielle. En pratique, cette fonction se révèle extrêmement pratique car elle est positive, et  $h(0) = 1$ .

On obtient ainsi le modèle

$$\lambda(t|z) = \lambda_0(t) e^{\beta^T z} \quad \text{avec} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} \quad \text{et} \quad z = \begin{pmatrix} z_1 \\ \vdots \\ z_p \end{pmatrix}.$$

$\beta$  est donc le vecteur des coefficients de régression,





## ② - Vraisemblance partielle de Cox.

Ne connaissant pas la forme de la fonction  $\lambda_0(t)$ , le principe est d'estimer uniquement le coefficient  $\beta$  en considérant cette dernière comme un paramètre de nuisance.  $\Rightarrow$  on ne cherche pas à estimer  $\lambda_0(t)$  !

$\rightarrow$  Intuition: dans les intervalles où aucun événement ne survient, on n'a aucune information pour évoluer les  $\beta$ . (on peut considérer que  $\lambda_0$  soit nulle dans ces intervalles).  
 $\Rightarrow$  On travaille alors conditionnellement à l'ensemble des instants où un décès survient.

$\rightarrow$  Notation: On reprend les notations déjà introduites dans les chapitres précédents:

- $D$  est le nombre de décès observés parmi les  $n$  individus à l'étude;
- $T_1 < T_2 < \dots < T_D$  sont des temps d'événements distincts;
- $(1), (2), \dots, (D)$  les indices des individus décédés en  $T_1, T_2, \dots, T_D$ ;
- $Z_i$ : la valeur des covariables de l'individu  $i$ ;
- $R(T_i)$  est l'exposition juste avant  $T_i$ : ensemble des individus à risque en  $T_i$ .

### a) Cas d'événements distincts

Ici, il n'y a qu'un seul décès à chaque temps d'événement. On s'intéresse à la probabilité qu'il y ait un événement en  $T_i$  (+ précisément dans  $[T_i, T_i + dt]$ ). Cette probabilité instantanée est donnée par le taux de hazard: (on a survécu jusqu'en  $T_i$ )

$$\sum_{j \in R(T_i)} \lambda_0(T_i) e^{\beta^T Z_j} : \text{proba d'au moins 1 événement en } T_i.$$
  
(Union d'événements: indiv. 1 ou indiv. 2 ou indiv. 3...  $\Rightarrow$  somme des probas)

La proba pur que ce soit l'individu  $i$  qui subisse l'événement en  $T_i$  est donc: (parmi tous les individus)

$$\frac{\lambda_0(T_i) e^{\beta^T Z_i}}{\sum_{j \in R(T_i)} \lambda_0(T_i) e^{\beta^T Z_j}} = \frac{e^{\beta^T Z_i}}{\sum_j e^{\beta^T Z_j}}$$



→ Cette probabilité ne dépend que du paramètre  $\beta$ .

→ Cette probabilité est une contribution individuelle à la vraisemblance globale: comme il y a contribution à chaque temps de décès, la vraisemblance partielle de Cox est définie comme le produit sur les temps de décès:

$$L^{\text{Cox}}(\beta; T_1, \dots, T_D) = \prod_{i=1}^D \frac{e^{\beta^T Z_{(i)}}}{\sum_{j \in R(T_i)} e^{\beta^T Z_j}}$$

→ Remarques:

\* Cette vraisemblance ne dépend pas de  $\lambda_0(t)$ : on peut donc estimer  $\beta$  sans connaître  $\lambda_0(t)$ , le hazard de base.

\* Cette vraisemblance est partielle au sens où elle n'est pas définie comme dans la statistique classique. En revanche, elle possède des propriétés similaires et peut donc être utilisée asymptotiquement pour l'estimation des tests des coef.  $\beta$ .

### b) Cas d'événements simultanés

Dans la réalité, il arrive qu'il y ait des décès "en même temps". (par exemple si l'on collecte les données par semaines...).

Lorsqu'il y a plusieurs événements, on admet tout de même que les événements se produisent les uns à la suite des autres. Mais on ne connaît pas l'ordre...!

→ Exemple de 2 individus  $s_1$  et  $s_2$  qui décèdent en  $T_i$ : la contrib. à la vraisemblance est:

$$\frac{e^{\beta^T Z_1}}{\sum_{j \in R(T_i)} e^{\beta^T Z_j}} \times \frac{e^{\beta^T Z_2}}{\sum_{j \in R(T_i) \setminus s_1} e^{\beta^T Z_j}} + \frac{e^{\beta^T Z_2}}{\sum_{j \in R(T_i)} e^{\beta^T Z_j}} \times \frac{e^{\beta^T Z_1}}{\sum_{j \in R(T_i) \setminus s_2} e^{\beta^T Z_j}} = \begin{cases} 1^{\text{er}} \text{ terme: cas où } s_1 \text{ décède en premier} \\ 2^{\text{e}} \text{ terme: cas où c'est } s_2! \end{cases}$$



→ On peut généraliser ceci au cas où  $k > 2$  décès interviennent "en même temps". Cependant, le calcul explicite deviendrait beaucoup trop long... (3)

⇒ On utilise l'approximation de Breslow!

→ Idee de l'approximation: la contribution des  $d_i$  événements survenus en  $T_i$  est le produit des probabilités  $p_j$  pour les individus décédés en  $T_i$ .

Ainsi 
$$\sum_{j \in R(T_i)} e^{\beta^T Z_j} \approx \sum_{j \in R(T_i) \setminus R} e^{\beta^T Z_j}$$

D'où la vraisemblance partielle de Breslow qui vaut, pour une contribution individuelle:

$$L_B(T_i) = \prod_{\substack{j: \text{uniques} \\ \text{décédés en } T_i}} p_j = \frac{e^{\beta^T \sum_{j: \text{uniques} \\ \text{décédés en } T_i} Z_j}}{\sum_{R \in R(T_i)} e^{\beta^T Z_R}}$$

et donc la vraisemblance totale de Breslow: 
$$L^{\text{Breslow}} = \prod_{i=1}^D L_B(T_i)$$

⇒ Temps de calcul rapide  
Si le nb d'événements simultanés est raisonnable, alors l'approximation est précise.

### ③ - Estimation

a) Estimation des coefficients de régression  $\beta$ :

En maximisant la vraisemblance partielle, on obtient une estimation du vecteur  $\beta$ .  
On considère pour ce faire la log-vraisemblance:

$$\ell(\beta) = \ln(L^{\text{Cox}}(\beta)) = \sum_{i=1}^D \left[ \beta^T Z_{(i)} - \ln \left( \sum_{j \in R(T_i)} e^{\beta^T Z_j} \right) \right];$$



et le score  $U(\beta)$  (vecteur de taille  $p$ ), les dérivées premières de  $\mathcal{L}(\beta)$ :

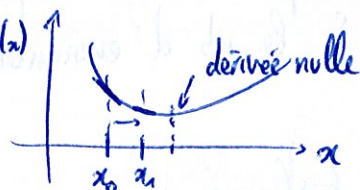
$$\begin{aligned}
 U(\beta) &= \frac{\partial \mathcal{L}(\beta)}{\partial \beta} = \left( \frac{\partial \mathcal{L}(\beta)}{\partial \beta_1}, \dots, \frac{\partial \mathcal{L}(\beta)}{\partial \beta_p} \right) \\
 &= \sum_{i=1}^D \left[ Z_{(i)} - \frac{\sum_{j \in R(i)} Z_j e^{\beta^T Z_j}}{\sum_{j \in R(i)} e^{\beta^T Z_j}} \right] \\
 &= \left( \underbrace{\sum_{i=1}^D \left[ Z_{(i),1} - \frac{\sum_{j \in R(i)} Z_{j,1} e^{\beta^T Z_j}}{\sum_{j \in R(i)} e^{\beta^T Z_j}} \right]}_{1^{\text{er}} \text{ composante pour } \beta_1}, \dots, \underbrace{\sum_{i=1}^D \left[ Z_{(i),p} - \frac{\sum_{j \in R(i)} Z_{j,p} e^{\beta^T Z_j}}{\sum_{j \in R(i)} e^{\beta^T Z_j}} \right]}_{p^{\text{ième}} \text{ composante pour } \beta_p} \right)
 \end{aligned}$$

← dérivée de  $\ln(u) = \frac{u'}{u}$

→ Ainsi, l'estimateur de Cox  $\hat{\beta}$  des coefficients de régression annule le score:

$$U(\hat{\beta}) = 0.$$

Comme il n'y a pas de solution explicite à la résolution du système d'équations, on utilise d'habitude l'algorithme de Newton-Raphson. (vient de la formule de Taylor-Lagrange où  $f \approx L'$ , à l'ordre 1).



→ Hérité de la théorie asymptotique de la maximisation de la vraisemblance, on retrouve un estimateur consistant de la matrice de variance-covariance de  $\beta$ , cet estimateur se calcule à partir de l'inverse de la matrice d'information de Fisher,

$$\widehat{\text{Var}}(\hat{\beta}) = (I(\hat{\beta}))^{-1} \text{ avec } [I(\beta)]_{i,j} = - \frac{\partial^2 \mathcal{L}(\beta)}{\partial \beta_i \partial \beta_j}$$

↑  
terme  $(i,j)$  de la matrice.



## b) Estimation du risque cumulé de base ( $\Lambda_0$ ):

Une fois estimés les coefficients de régression  $\beta$ , on peut utiliser l'estimateur de Breslow du risque cumulé de base (extension de l'estimateur de Nelson-Aalen):

$$\boxed{\hat{\Lambda}_0(t) = \sum_{i: T_i \leq t} \frac{d_i}{\sum_{j \in R(t)} e^{\hat{\beta}^T z_j}}}, \text{ avec } d_i \text{ le nb de décès en } T_i.$$

→ Si  $\hat{\beta} = 0$ , alors on retrouve l'estimateur de Nelson-Aalen!

→ De cette expression, on déduit un estimateur de la fonction de survie pour un individu de covariables  $Z$ : comme  $S(t|Z) = e^{-\int_0^t \lambda(u|Z) du}$ , on a:

$$\boxed{\hat{S}(t|Z) = e^{-\int_0^t \hat{\lambda}(u|Z) du} = e^{-\hat{\Lambda}_0(t) e^{\hat{\beta}^T Z}}}$$

## c) Tests sur les coefficients de régression $\beta$ :

Comme dans le cas de la régression classique, on souhaite tester la significativité des coefficients de régression. Autrement dit, à partir des propriétés asymptotiques de  $\hat{\beta}$ , on souhaiterait tester l'hypothèse:

$$\boxed{H_0: \beta = \beta_0.}$$

Pour tester la significativité de  $\beta$ , on prendra  $\beta_0 = 0$ .

On introduit d'abord les 3 statistiques couramment utilisées pour réaliser un test sur le vecteur  $\beta$ .



### a) La statistique du rapport de vraisemblance:

Comme son <sup>nom</sup> l'indique, elle mesure la distance entre  $\ln L^{\text{lox}}(\hat{\beta})$  et  $\ln L^{\text{lox}}(\beta_0)$ . La différence de deux logarithmes étant un rapport, on l'appelle "rapport de vraisembl.":

$$\text{Likelihood} \left| \begin{array}{l} X_{\text{LRT}}^2 \\ \text{ratio Test} \end{array} \right. = 2 \left[ \ln L^{\text{lox}}(\hat{\beta}) - \ln L^{\text{lox}}(\beta_0) \right] \underset{\text{sous } H_0}{\sim} \chi_p^2.$$

Asymptotiquement, cette statistique suit une loi du  $\chi^2$  (Chi-2) à  $p$  degrés de liberté, où  $p$  est le nombre de composantes du vecteur  $\beta$ .

Rq: On obtient ce résultat en loi en faisant un développement de Taylor à l'ordre 2 autour de  $\hat{\beta}$ .

### b) La statistique de Wald:

Elle mesure l'écart entre  $\hat{\beta}$  et  $\beta_0$ . Elle est définie par:

$$\left| \begin{array}{l} X_{\text{Wald}}^2 \end{array} \right. = (\hat{\beta} - \beta_0)^T I(\hat{\beta}) (\hat{\beta} - \beta_0) \underset{\text{sous } H_0}{\sim} \chi_p^2.$$

On rappelle que si  $\hat{\beta}$  est un estimateur du maximum de vraisemblance, la variance de l'estimateur est donnée par l'inverse de la matrice d'information de Fisher. Ainsi, on retrouve la formule plus connue dans le cadre unidimensionnel: si  $\beta \in \mathbb{R}$ ,

$$X_W^2 = \frac{(\hat{\beta} - \beta_0)^2}{\text{Var}(\hat{\beta})} \sim \chi_1^2.$$

Rq: Ce résultat est assez intuitif dans la mesure où la statistique elle-même est une gaussienne normalisée au carré, ce qui conduit à une loi du Chi-2.



### c) La statistique du score:

Elle s'intéresse à la mesure de la pente de la tangente en  $\beta_0$  : ↖ score

Sous  $H_0$ ,  $\beta = \beta_0$ . On rappelle que le score s'annule en  $\hat{\beta}$ , donc on mesure le score en  $\beta_0$  plutôt que  $\hat{\beta}$  (sachant que sous  $H_0$ ,  $\hat{\beta}$  devrait être proche de  $\beta_0$ ) :

$$\boxed{\underset{\text{score}}{X_S^2} = U(\beta_0)^T (I(\beta_0))^{-1} U(\beta_0) \underset{\text{sous } H_0}{\sim} \chi_p^2.}$$

→ Les 3 statistiques vues ci-dessus suivent la même loi !

→ Pour conduire le test, on remplacera les v.a. par leurs réalisations, ce qui permettra de calculer la statistique de manière effective. Puis on comparera cette valeur au quantile d'intérêt (selon le niveau du test) de la loi du  $\chi^2$  à  $p$  degrés de liberté. Si elle excède ce quantile, on rejettera l'hypothèse nulle  $H_0$ .

→ Rq :

- $X_{LRT}^2$  ne nécessite pas de calculer des dérivées secondes.
- $X_S^2$  ne nécessite pas l'estimation de  $\hat{\beta}$ .
- $X_W^2$  requiert ces 2 calculs.

### d) Tests partiels.

Ces tests sont des cas particuliers des tests précédents où l'on s'intéresse à une coordonnée particulière du vecteur  $\beta$  (ou plusieurs coordonnées).

Par exemple, on peut vouloir tester l'ajout d'une nouvelle covariable  $Z_p$  dans un modèle à  $(p-1)$  variables. On aimerait savoir si cet ajout apporte (plus d'information sur la distribution des durées de vie, comparé au modèle plus simple   
 suffisamment à  $(p-1)$  variables.



Autrement dit, on teste l'hypothèse:  $H_0: \beta = \beta_0$ , où  $\begin{cases} \beta = (\beta_1, \dots, \beta_p) \\ \beta_0 = (\beta_1, \dots, \beta_{p-1}, 0) \end{cases}$ .

En fait on teste donc  $\boxed{H_0: \beta_p = 0}$ .

Dans ce cas, on a les résultats suivants:

- $X_{LRT}^2 = 2 [\ln L^{\text{cox}}(\hat{\beta}) - \ln L^{\text{cox}}(\hat{\beta}_0)] \underset{\text{sous } H_0}{\sim} \chi_1^2$  car 1 paramètre à tester!
- $X_W^2 = (\hat{\beta} - \hat{\beta}_0)^T I(\hat{\beta}) (\hat{\beta} - \hat{\beta}_0) \underset{\text{sous } H_0}{\sim} \chi_1^2$ .
- $X_S^2 = U(\hat{\beta}_0)^T (I(\hat{\beta}_0))^{-1} U(\hat{\beta}_0) \underset{\text{sous } H_0}{\sim} \chi_1^2$ .

avec  $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_p)$  et  $\hat{\beta}_0 = (\hat{\beta}_1, \dots, \hat{\beta}_{p-1}, 0)$ .

### ⑤ - Interprétation des coefficients de régression.

Dans le modèle de Cox, le risque relatif entre 2 individus  $i$  et  $j$  à l'instant  $t$  est donné par:

$$RR(t) = \frac{\lambda(t|z_i)}{\lambda(t|z_j)} = \frac{\cancel{\lambda(t)} e^{\beta^T z_i}}{\cancel{\lambda(t)} e^{\beta^T z_j}} = e^{\beta^T (z_i - z_j)}.$$

$\Rightarrow$  Il est constant au cours du temps!

Ainsi, dans un modèle simplifié avec une seule covariable  $Z$ ,  $RR(t)$  vaut:

- pour  $Z$  binaire codée 0 et 1:  $RR = e^{\beta}$
  - pour  $Z$  binaire codée  $a$  et  $b$ :  $RR = e^{\beta(b-a)}$
  - pour  $Z$  continue:  $e^{\beta}$  est le RR pour une augmentation d'une unité de  $Z$ .
- (quelque soit la valeur de la covariable: il s'agit d'une hyp. de log-linéarité).

Bilan: 2 hypothèses à vérifier pour utiliser Cox:  $\begin{cases} \rightarrow \text{risques proportionnels} \\ \rightarrow \text{log-linéarité} \end{cases}$



## ⑥ - Test d'adéquation du modèle.

⑥

### a) Risques proportionnels.

Pour valider l'hyp selon laquelle les risques sont proportionnels et pouvoir utiliser le modèle de Cox, il existe plusieurs méthodes.

- Considérer une covariable dépendante du temps: on introduit alors un terme d'interaction entre le temps et la covariable, puis on teste la significativité de l'effet du temps. Par exemple, pour une covariable  $Z$ , on considère la variable  $\beta Z + \gamma Z \ln(t)$ , ce qui induit:  $\lambda(t|Z) = \lambda_0(t) e^{\beta Z + \gamma Z \ln t} = \lambda_0(t) (e^{\beta Z} t^{\gamma Z})$ ,  
et  $RR(t) = \frac{\lambda(t|Z_i)}{\lambda(t|Z_j)} = e^{\beta(Z_i - Z_j)} t^{\gamma(Z_i - Z_j)}$

Pour une covariable binaire, on obtient donc:  $RR(t) = e^{\beta} t^{\gamma}$ .

Puis on teste si  $\gamma$  est statistiquement  $\neq$  de 0:

→ s'il l'est, l'hypothèse de proportionnalité n'est pas respectée!

→ sinon c'est OK.

Dans le premier cas, une solution peut être de stratifier suivant les modalités de la covariable, ou d'utiliser des coefficients de régression dépendant du temps. Avec la première solution, on perd en puissance (car on diminue  $n$  en formant des sous-groupes); avec la 2<sup>e</sup> solution, on n'a plus un modèle à risques proportionnels!

- Utiliser des validations graphiques: trop subjectif mais donne une première idée. Par exemple, pour une covariable avec 2 modalités, on peut estimer la fonction de survie dans les 2 groupes (hommes et femmes par exemple).



Puis on trace les courbes :

$$\ln(-\ln \hat{S}(t)) = \ln\left(\int_0^t \hat{\lambda}(u|Z) du\right) = \ln\left(\hat{\Lambda}_0(t) e^{\hat{\beta}^T Z}\right) \\ = \ln(\hat{\Lambda}_0(t)) + \hat{\beta}^T Z.$$

⇒ Il suffit alors ensuite de vérifier que les courbes obtenues pour chaque modalité présentent un écart constant au cours du temps.

On peut aussi utiliser les résidus de Schoenfeld.

- Utiliser les résidus de Cox-Snell : ils permettent de tester la validité globale du modèle. On sait que la fonction de survie de la v.a.  $T$  vérifie :

$$S(t) = P(T > t) = e^{-\Lambda(t|Z)}, \text{ avec ici } \Lambda(t|Z) = \Lambda_0(t) e^{\beta^T Z}.$$

En considérant la fonction de survie de la v.a.  $V = \Lambda(T|Z)$ , on a :

$$P(\Lambda(T|Z) > y) = P(V > y) = P(T > \Lambda^{-1}(y|Z)) = e^{-y}.$$

⇒ On en conclut que  $V = \Lambda(T|Z) \sim \text{Exp}(1)!$

⇒ Le modèle est adéquat si le risque cumulé de la v.a.  $V$  est proche de la droite  $y=x$ , car cette dernière correspond au risque cumulé d'une loi  $\text{Exp}(1)$ .

→ On procède donc comme suit :

1. Estimation de  $\Lambda(\cdot|Z)$  par l'estimateur semi-paramétrique  $\hat{\Lambda}(\cdot|Z) = \hat{\Lambda}_0(\cdot) e^{\hat{\beta}^T Z}$
2. Pour chaque  $T_i$  observé ( $i=1, \dots, n$ ), on associe la variable  $v_i$  (résidus de Cox-Snell) :  
$$v_i = \hat{\Lambda}(T_i|Z) = \hat{\Lambda}_0(T_i) e^{\hat{\beta}^T Z}$$
3. Estimation du risque cumulé des variables  $v_i$  par l'estimateur de Nelson-Aalen, noté  $\hat{\Lambda}_r$ .
4. Tracé des fonctions  $y = \hat{\Lambda}_r(x)$  et  $y=x$  sur un graphique. Si le modèle est correct, les courbes sont proches de se superposer.



## b) Hypothèse de log-linéarité:

(7)

Sous cette hypothèse, le logarithme du risque est une fonction linéaire de  $Z$ ,

$$\ln \lambda(t|Z) - \ln \lambda_0(t) = \beta^T Z.$$

→ Cette hypothèse implique que RR est constant pour une  $\uparrow$  d'une unité, quelque soit la valeur de la covariable.

→ Rq: - pour une variable continue, c'est une hypothèse qui peut se révéler particulièrement contraignante... Ex: si l'âge est très explicatif d'une maladie touchant essentiellement les personnes âgées, prendre un an d'âge aurait le même effet à 30 ans qu'à 70 ans...  
- pour une variable catégorielle, on peut prendre un codage dichotomique par relâche la contrainte (par ex. une variable à 3 modalités peut être recodée par 2 variables binaires).

## III - Extensions

### ① - Covariables dépendantes du temps.

On peut prendre en compte des variables explicatives dépendantes du temps dans le modèle de Cox. Il faut que  $Z(t)$  soit connue au temps  $t$  ( $Z(t)$  prédictible).

L'approche statistique (estimation, test, ...) reste la même, avec tout de même:

→ la connaissance nécessaire de la valeur des covariables par chaque temps d'événement en effet pour calculer  $L^{Cox}(\beta) = \prod_{i=1}^D \frac{e^{\beta^T Z_{(i)}(T_i)}}{\sum_{j \in R(T_i)} e^{\beta^T Z_j(T_i)}}$  - parfois difficile dans certaines applications...

→ Interprétation ardue!

→ L'effet de la covariable ne change pas au cours du temps, c'est la variable qui change.

→ L'utilisation de certaines covariables dépendantes du temps permet de tester l'hyp. de risque proportionnel.



## ②. Modèle de Cox stratifié.

Si une covariable qualitative ne vérifie pas l'hypothèse de hazards proportionnels, on peut considérer un modèle de Cox stratifié.

→ Exemple: soit une covariable  $Y$  codée 0 ou 1, comme le sexe.

Dans ce modèle, le risque de base est différent dans les 2 strates, mais les autres covariables  $Z$  agissent identiquement sur les 2 hazards, i.e.:

$$\begin{cases} \lambda(t|Z, Y=0) = \lambda_0(H) e^{\beta^T Z} \\ \lambda(t|Z, Y=1) = \lambda_1(H) e^{\beta^T Z} \end{cases}$$

→ la vraisemblance partielle se calcule dans chacune des strates, et la vraisemblance totale est le produit des vraisemblances de chaque strate.

→ On peut tester l'hypothèse selon laquelle les covariables  $Z$  agissent de la même manière dans chaque strate en considérant: le même pour toutes les strates.

$$X^2_{LRT} = 2 \left[ \left( \sum_{j=1}^s \ln L^{\text{Cox}}(\hat{\beta}_j) \right) - \ln L^{\text{Cox}}(\hat{\beta}) \right] \underset{\text{sous } H_0}{\sim} \chi^2_{sp-p}$$

avec  $s$  le nombre de strates. les  $\beta$  sont  $\neq$  de chacune des strates.

## ③. Modèles "Frailty" (de fragilité).

On suppose dans le modèle de Cox que la population est homogène (on tient déjà compte d'une forme d'hétérogénéité, "observable", via les covariables).

Parfois, certaines covariables importantes ne sont pas observables... (ou inconnues!).

⇒ On considère alors les modèles fragilité.



Ces modèles permettent de prendre en compte l'hétérogénéité inobservable. (8)

Considérons une nouvelle covariable  $Z_0$  non-observée.

La fonction de risque s'exprime alors

$$\lambda(t|Z, Z_0) = \lambda_0(t) e^{\beta_0 Z_0 + \beta^T Z} = \lambda_0(t) e^{\beta_0 Z_0} e^{\beta^T Z}.$$

On note maintenant  $w = e^{\beta_0 Z_0}$  la v.a. réelle positive, appelée "fragilité".

$$\lambda(t|Z, w) = \lambda_0(t) w e^{\beta^T Z}.$$

$\uparrow$  v.a.  $\Rightarrow$  modèle mélange!

Alors la fonction de survie conditionnelle s'écrit:

$$S(t|Z, w) = \exp\left(-\int_0^t \lambda_0(s) w e^{\beta^T Z} ds\right) = e^{-w e^{\beta^T Z} \Lambda_0(t)}.$$

Puisque  $w$  est aléatoire, on intègre sur la loi de  $w$  pour avoir la survie non-conditionnelle:

$$S(t|Z) = \int_0^{+\infty} e^{-w e^{\beta^T Z} \Lambda_0(t)} f_w(w) dw, \text{ avec } f_w \text{ la densité de } w.$$

→ Utilisation: le + souvent, ces modèles sont utilisés pour prendre en compte une dépendance entre les temps d'événement de certains individus. En effet, cette dépendance peut provenir d'une caractéristique commune non-observée à l'intérieur d'un sous-groupe. (même famille, même hôpital, ...).

$\Rightarrow$  la v.a.  $w$  de fragilité est alors commune aux individus d'un sous-groupe, et varie pour un autre sous-groupe.



→ Formellement, on considère  $X_{ij} = \min(T_{ij}, C_{ij})$ , avec  $\begin{cases} i \text{ le sous-groupe } (i=1, \dots, 6) \\ j \text{ l'individu} \end{cases}$ .

On a aussi:  $\lambda_{ij}(H) = \lambda_{ij}(H|Z_j, w_i) = \lambda_0(H) w_i e^{\beta^T Z_j}$   
 $\underbrace{\hspace{10em}}_{\text{groupe } i}.$

où  $w_i$  est la fragilité du groupe  $i$ .

- Les  $w_i$  sont iid. En général, on suppose  $E[w_i] = 1$  et  $Var(w_i) = \sigma$ .

→  $\sigma$  permet de mesurer l'hétérogénéité entre les groupes.

$E(w_i) = 1 \Rightarrow$

$w_i > 1$	$\equiv$	le risque du groupe i est $>$ en moyenne au risque de base.
$w_i < 1$	$\equiv$	" " " " " "

- Dans ce modèle, les observations sont iid, conditionnellement aux  $w_i$ .

→ On prend souvent en pratique la loi Gamma comme v.a. de fragilité; car elle simplifie les calculs.

→ En toute généralité, l'algorithme EM est utilisé pour estimer les paramètres du modèle.