Simulation de covariables dépendantes du temps via une distribution de Weibull

AARAB Ayoub BURGAT PAUL LAABSI Zakaria

Faculté des Sciences
Université de Montpellier

Année 2020-2021

Sommaire

Présentation du projet

Modèle de Cox standard
- Modèle de Cox
- Hypothèses du modèle
- Génération d'une durée de survie invariante au temps

Extension du modèle de Cox
- Modèle de Cox avec covariables générées via un modèle linéaire mixte
- Génération des durées de survie avec dépendance du temps

Sommaire

Présentation du projet

Modèle de Cox standard
- Modèle de Cox
- Hypothèses du modèle
- Génération d'une durée de survie invariante au temps

Extension du modèle de Cox
- Modèle de Cox avec covariables générées via un modèle linéaire mixte
- Génération des durées de survie avec dépendance du temps

L'Analyse de la survie

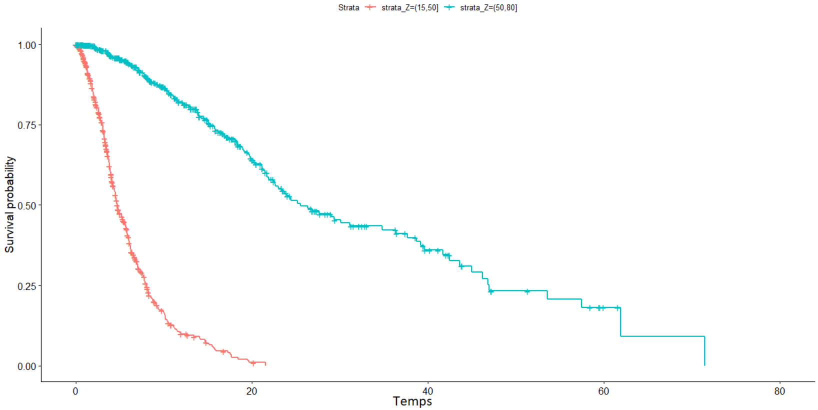

FIGURE – Courbes de survie : $\forall t \geq 0, \quad S(t) = \mathbb{P}(X > t)$

Sommaire

Présentation du projet

Modèle de Cox standard
- Modèle de Cox
- Hypothèses du modèle
- Génération d'une durée de survie invariante au temps

Extension du modèle de Cox
- Modèle de Cox avec covariables générées via un modèle linéaire mixte
- Génération des durées de survie avec dépendance du temps

Modèle de Cox

Le modèle de Cox donne l'expression suivante pour la fonction de risque instantané de décès.

$$h\,(t \mid \mathbf{Z}) = h_0(t)\exp\left(\sum_{k=1}^{n}\beta_k Z_k\right) = h_0(t)\exp\left(\mathbf{Z}'\beta\right) \tag{1}$$

avec :

$$\mathbf{Z} = \left(Z_1, ..., Z_p\right)' \text{ le vecteur des covariables}$$

$$\beta = \left(\beta_1, ..., \beta_p\right)' \text{ le vecteur des constantes}$$

β_k est le paramètre pour la k-ème covariable Z_k. La fonction h caractérise la loi de la durée de survie :

$$S(t) = \exp\left(-\int_0^t h(u)\,du\right),\ t \geqslant 0$$

Hypothèses du modèle

Hypothèse des risques proportionnels : Le rapport des risques relatifs RR entre deux individus i et j est indépendant du temps t.

$$\frac{h\left(t \mid \mathbf{Z}^{(i)}\right)}{h\left(t \mid \mathbf{Z}^{(j)}\right)} = \frac{h_0(t)\exp\left(\mathbf{Z}^{(i)\prime}\beta\right)}{h_0(t)\exp\left(\mathbf{Z}^{(j)\prime}\beta\right)} = \exp\left(\left(\mathbf{Z}^{(i)} - \mathbf{Z}^{(j)}\right)^{\prime}\beta\right)$$

Hypothèse de log-linéarité : Le modèle de Cox standard est un modèle log-linéaire :

$$\log(h(t \mid \mathbf{Z})) = \log(h_0(t)) + \mathbf{Z}'\beta \quad \text{est une fonction linéaire des } Z_i$$

Distribution de Weibull

Densité de probabilité de la loi de Weibull, pour $x > 0$:

$$f(x;k,\lambda) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1}\exp\left(-\left(\frac{x}{\lambda}\right)\right)^{k}$$

- $k > 0$: paramètre de forme
- $\lambda > 0$: paramètre d'échelle de la distribution

Fonction de répartition :

$$F(x;k,\lambda) = 1 - \exp\left(-\left(\frac{x}{\lambda}\right)^{k}\right)$$

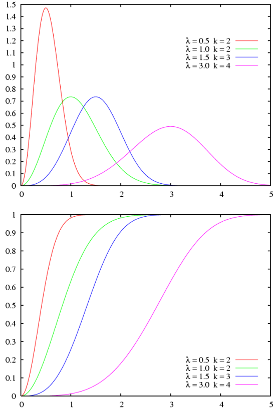

FIGURE – Densité de probabilité et fonction de répartition

Génération de données vérifiant les hypothèses du modèle de Cox :

	Transformée inverse	**Modèle log-linéaire**
Durée de survie T	$\mathrm{T} = \left(-\frac{\log(V)}{\lambda \exp(\mathbf{Z}'\beta)}\right)^{\frac{1}{\alpha}}$ où $V \sim U(0,1), \lambda, \alpha > 0$	$\mathrm{T} = \exp(\mu + \gamma \mathbf{Z} + \sigma W)$ où $\mu, \gamma, \sigma \in \mathbf{R}$ et $W \sim F_W$ suit la loi des extrêmes.
Distribution de probabilité	Weibull $(\lambda(\mathbf{Z}), \alpha)$ avec $\lambda(\mathbf{Z}) = \lambda \exp(\mathbf{Z}'\beta)$	Weibull $\left(\exp(\mu + \gamma Z), \frac{1}{\sigma}\right)$

TABLE – Deux méthodes de génération de durée de survie avec indépendance du temps

Validation des hypothèses sur un jeu de données simulées

Soient $Z \sim Unif(15, 80)$, $\mu = 10$, $\gamma = 1$ et $\sigma = 0.5$.
Hypothèse nulle : la covariable Z est indépendante du temps.

	chisq	df	p-value
Z	1.07	1	0.3
GLOBAL	1.07	1	0.3

TABLE – Test d'hypothèse de la dépendance au temps d'une covariable

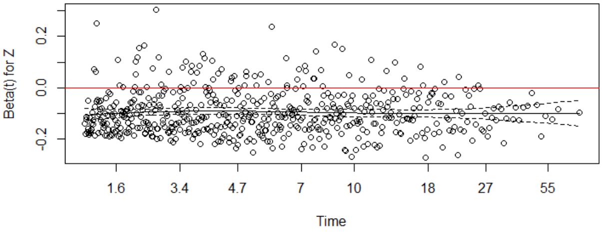

FIGURE – Test des résidus de Schoenfeld
Droite de régression $r_{ij} = at + \epsilon_i$

Sommaire

Présentation du projet

Modèle de Cox standard
- Modèle de Cox
- Hypothèses du modèle
- Génération d'une durée de survie invariante au temps

Extension du modèle de Cox
- Modèle de Cox avec covariables générées via un modèle linéaire mixte
- Génération des durées de survie avec dépendance du temps

Nouvelle approche lorsque les covariables varient avec le temps

(inspirée de Julius S.Ngwa et al, 2019)

Echantillon d'individus de taille n avec covariables indépendantes et dépendantes du temps :

$$\left\{ \left(T_i,\ \mathbf{Y}_i(t),\ \mathbf{Z}_i,\ \Delta_i \right) \mid\ 0 \leq t \leq T_i \quad \forall\, i = 1, 2, .., n \right\}$$

- T_i : le temps d'évènement pour l'individu i
- Δ_i : l'indicatrice d'évènement de l'individu i
- $\mathbf{Z}_i$: le vecteur des covariables indépendantes du temps de l'individu i
- $\mathbf{Y}_i(t)$: le vecteur des covariables dépendantes du temps de l'individu i avec m_i mesures au cours du temps

Nouvelle écriture avec covariables dépendantes du temps

Nouvelle fonction de risque instantané de décès :

$$\begin{aligned} h(t \mid \mathbf{Z}_i, \mathbf{Y}_i(t)) &= h_0(t) \exp\left(\mathbf{Z}_i'\beta + \gamma \mathbf{1} \cdot \mathbf{Y}_i(t)\right) \\ &= h_0(t) \exp\left(\mathbf{Z}_i'\beta + \gamma \sum_{j=1}^{m_i} Y_{ij}(t)\right) \end{aligned}$$

- $h_0(t)$: la fonction de risque de base
- β : le vecteur des coefficients pour des covariables invariantes dans le temps.
- $\gamma \in \mathbf{R}$: Le paramètre (scalaire) qui relie les covariables dépendantes du temps au risque.

Modèle linéaire mixte

Modèle linéaire à effets fixes	Modèle linéaire mixte
$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \varepsilon_{ij}$	$Y_{ij} = \beta_0 + \beta_1 X_{ij} + b_{i,0} + b_{i,1} X_{ij} + \varepsilon_{ij}$

TABLE – Comparaison des deux modèles linéaires

- Y_{ij} est la réponse de l'individu i au temps d'évènement j avec $i \in \{1, \dots, n\}$ et $j \in \{1, \dots, m_i\}$
- X_{ij} est la variable variant dans le temps de l'individu i.
- β_0 et β_1 les coefficients identiques pour tous les individus (effets fixes)
- $\beta_{i,0}$ et $\beta_{i,1}$ les coefficients propres à l'individu i
- $b_i = (\beta_i - \beta)$ les effets aléatoires : les déviations par rapport aux coefficients β du groupe d'individus

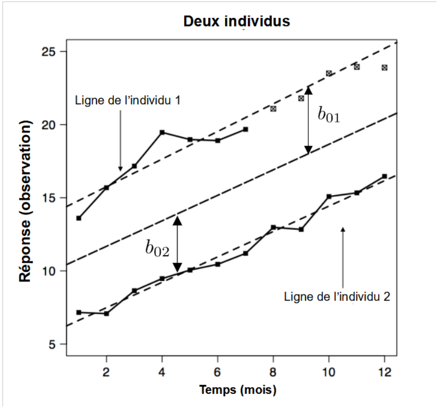

FIGURE – Trajectoires longitudinales pour deux individus. D'après Belle & al.

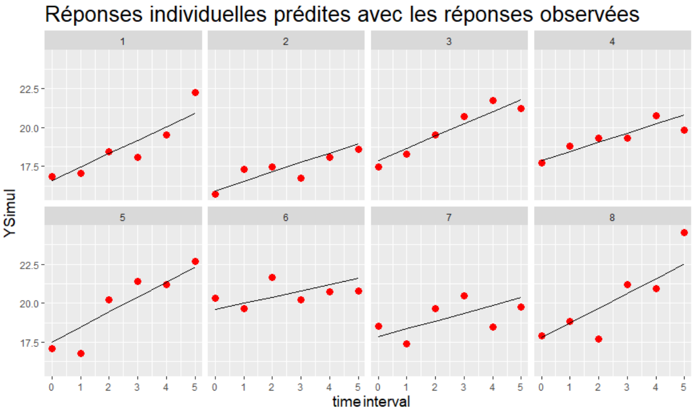

FIGURE – Mesures longitudinales générées via une simulation pour les 8 premiers individus

Expression de la durée de survie

Génération de la durée de survie T via la transformée inverse et la fonction de Lambert W :

$$T = \frac{1}{\gamma \cdot \left(\beta_{i,1} \times \frac{1}{\alpha}\right)} \times W\left(\gamma \cdot \left(\beta_{i,1} \times \frac{1}{\alpha}\right) \times \left(\frac{-\log(Q)}{\lambda \exp(Z'\beta + \gamma(\beta_{i,0}))}\right)^{\frac{1}{\alpha}}\right)$$

Fonction W de Lambert : fonction réciproque de $f(w) = we^w$ telle que

$$z = f(w) = we^w \iff w = W(z) \quad \text{avec } z, w \in \mathbf{C}$$

Extrait commenté code R

1. ```
   RMVN=rmvnorm(N, mean=beta, sigma=S,
   method = "chol")
   ```
2. ```
   RMVNb=rmvnorm(N, mean= matrix(0,1,2),
   sigma=S, method = "chol")
   ```
3. ```
   Z=matrix(0,M,length(dim(S)))
   Z[,1] <- 1
   Z[,2] <- c(time)
   Y <- matrix(0,N,M)
   R <- diag(rnorm(M,1.000,0.00001), M)
   V=Z%*%S%*%t(Z) + R
   ```

1. Génération des effets aléatoires mixtes β_i :
$$\beta_i = \begin{pmatrix} \beta_{i,0} \\ \beta_{i,1} \end{pmatrix} \sim N(\beta, \Sigma)$$
2. Et effets aléatoires b_i
$$b_i = \begin{pmatrix} \beta_{i,1} - \beta_1 \\ \beta_{i,2} - \beta_2 \end{pmatrix} \sim N(0_p, \Sigma)$$
3. Création des $Y_{ij}(t_{ij})$ et de la variance V_i avec :
$$V_i = Z_i \Sigma Z_i' + R_i$$
   ```

Estimation des coefficients

	Paramètres fixes	Matrice de variance-covariance des effets aléatoires
Observés	$\beta = \begin{pmatrix} 16.7010 \\ 0.6601 \end{pmatrix}$	$\Sigma = \begin{pmatrix} 5.4149 & -0.3214 \\ -0.3214 & 0.0513 \end{pmatrix}$
Estimés	$\widehat{\beta} = \begin{pmatrix} 15.6085 \\ 0.6759 \end{pmatrix}$	$\widehat{\Sigma} = \begin{pmatrix} 4.6235 & -0.2911 \\ -0.2911 & 0.0584 \end{pmatrix}$

TABLE – Paramètres statistiques observés et leurs estimations via les packages R `nlme` et `lme4`

Bilan

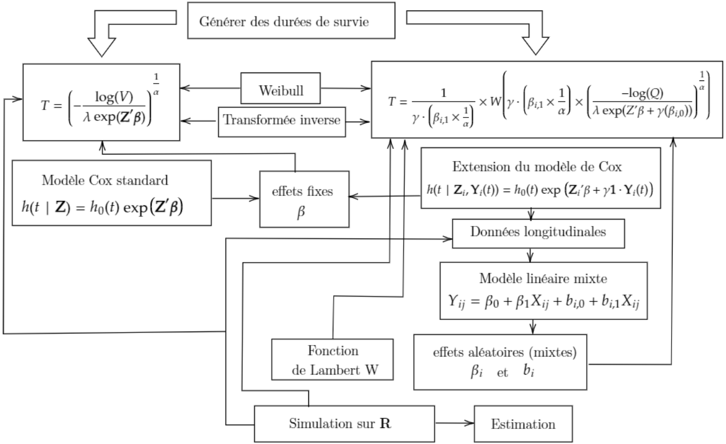

FIGURE – Organigramme du projet

Merci de votre attention !