

TD 3: MODÈLE À HASARD PROPORTIONNEL.

Modèle de durée M1 DUAS– Semestre 2
P.-O. Goffard

1. On considère un modèle de hasard proportionnel pour lequel le risque de base est donnée par une loi de Weibull. La fonction de hasard de T conditionnellement au vecteur de covariables Z est donnée par

$$h(t) := h(t|z; \alpha, \beta) = \alpha t^{\alpha-1} e^{z\beta}.$$

Soit un échantillon de données censurées à droite

$$(x_i, \delta_i) = (t_i \wedge c_i, \mathbb{I}_{t_i \leq c_i}), \quad i = 1, \dots, n.$$

- (a) Ecrire la log vraisemblance du modèle.

Solution: La vraisemblance s'écrit

$$\mathcal{L}(\mathcal{D}; \theta) = \prod_{i=1}^n (\alpha x_i^{\alpha-1} e^{t_i z_i \beta})^{\delta_i} \exp(e^{t_i z_i \beta} x_i^\alpha).$$

La log vraisemblance est donnée par

$$l(\mathcal{D}; \theta) = \ln(\alpha) \sum_{i=1}^n \delta_i + (\alpha - 1) \sum_{i=1}^n \delta_i \ln(x_i) + \sum_{i=1}^n \delta_i t_i z_i \beta - \sum_{i=1}^n e^{t_i z_i \beta} x_i^\alpha.$$

- (b) Ecrire l'expression des dérivées premières de la log vraisemblance.

Solution: Les dérivées premières sont données par

$$\frac{\partial}{\partial \alpha} l(\mathcal{D}; \theta) = \frac{1}{\alpha} \sum_{i=1}^n \delta_i + \sum_{i=1}^n \delta_i \ln(x_i) - \sum_{i=1}^n e^{t_i z_i \beta} \ln(x_i) x_i^\alpha.$$

et

$$\frac{\partial}{\partial \beta} l(\mathcal{D}; \theta) = \sum_{i=1}^n \delta_i z_i - \sum_{i=1}^n z_i e^{t_i z_i \beta} x_i^\alpha.$$

- (c) Ecrire un code R qui simule des données depuis un modèle à hasard proportionnel de Weibull et ajuste le modèle aux données.

1. Ecrire une fonction qui prend en entrée α , β et la matrice des covariables.
2. Produire un dataframe contenant 1,000 observations
 - pour chaque individu, on dispose de deux covariables $Z_1 \sim \text{Bin}(1, p = 1/2)$ et $Z_2 \sim \text{Unif}([0, 1])$.
 - Les observations non censurées t_i sont distribués suivant un modèle de Weibull à hasard proportionnel tel que $\alpha = 1/2$ et $\beta = (-1/2 \ 1/2)$

- Les observations censurées à droite x_i proviennent d'une variable de censure non informative c_i suivant le même modèle que les observations c'est à dire un modèle de Weibull à hasard proportionnel tel que $\alpha = 1/2$ et $\beta = (-1/2 \ 1/2)$.
3. Ecrire une fonction pour ajuster le modèle aux données et appliquer cette fonction à votre dataframe.
 4. Vous pouvez comparer le résultat avec celui rendu par la fonction `survreg`¹ de la librairie `survival`.

¹Voir <http://dwo11.de/rexrepos/posts/survivalParametric.html> pour un tuto