TD 3: MODÈLE À HASARD PROPORTIONEL.

Modèle de duréee 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 conditionellement au vecteur de covariables Z est donnée par

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

Soit un échantillon de données censurées à droite

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

(a) Ecrire la log vraisemblance du modèle.

Solution: La vraisemblance s'écrit

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

La log vraisemblance est donnée par

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

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

Solution: Les dérivées premières sont données par

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

et

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

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

1. Ecrire une fonction qui prend en entrée α, β et la matrice des covariables et qui retourne un échantillon suivant le modèle à hasard proportionnel de Weibull. On pourra considérer deux covariables $Z_1 \sim \mathsf{Bin}(1,p)$ et $Z_2 \sim \mathsf{Unif}([0,1])$.
2. Produire un dataframe contenant $1,000$ observations
   - pour chaque individu, on dispose de la valeur des deux covariables $Z_1 \sim \mathsf{Bin}(1,p=1/2)$ et $Z_2 \sim \mathsf{Unif}([0,1])$.
   - Les observations non censurées t_i sont distribués suivant un modèle de Weibull à hasard proportionel 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 proportionel tel que $\alpha = 1/2$ et $\beta = (-1/2\ 1/2)$.
- L'indication $\delta_i = \mathbb{I}_{t_i \leq c_i}$ sur la censure de l'observation i

3. Ecrire une fonction pour ajuster le modèle aux données et appliquer cette fonction à votre dataframe. Il s'agit de maximiser la vraisemblance obntenu à la question a).
4. Vous pouvez comparer le résultat avec celui rendu par la fonction `survreg`[1] de la librairie survival.
5. Vous pouvez auusi regarder le résultat lorsque vous utiliser la fonction `coxph` de la librairie survival.

[1] Voir `http://dwoll.de/rexrepos/posts/survivalParametric.html` pour un tuto