Devoir 2
STT3260 - modèles de survie
Échéance: vendredi 3 octobre.

Tâche de lecture:

- Finissez la lecture du Chapitre 2, sections 2.1 à 2.5 [Klein & Moeschberger].
- vendredi 3 octobre:
  Chapitre 2: Section 2.6 [Klein & Moeschberger]. Chapitre 3, Sections 3.1 à 3.5 [Klein & Moeschberger].

Bonne lecture!

1.- [**50 points**] Le tableau suivant présente une étude clinique sur la leucémie aiguë. Dans cet essai, l'événement à considérer est le délai entre le traitement et la rechute de symptômes de la leucémie. La ligne "Nombre de rechutes" représente le nombre de personnes qui ont rechuté à la date "Temps" correspondant.

Temps	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Nombre de rechutes	1	22	3	12	8	17	2	11	8	12	2	5	4	15	8	23	5	11	4	1	9

(a) Calculez la vraisemblance maximale pour les modèles log-normale et gamma.

(b) Considérez le modèle gamma généralisée avec densité $f(t) = \frac{\alpha\lambda^\beta}{\Gamma(\beta)} t^{\alpha\beta-1}\exp\{-\lambda t^\alpha\}$. Supposez que T suit une loi gamma généralisée.

i. Montrez que $Y = \log T$ peut s'écrire

$$Y = -\frac{\log\lambda - \log\beta}{\alpha} + \frac{1}{\alpha\sqrt{\beta}}W,$$

où W suit une loi log-gamma avec densité

$$g(w) = Q\frac{(Q^{-2})^{Q^{-2}}}{\Gamma(Q^{-2})}\exp\left\{\frac{Qw - \exp\{Qw\}}{Q^2}\right\}$$

où $Q = \beta^{-1/2}$. C'est-à-dire le logarithme d'une variable gamma généralisée suit un modèle de position-échelle avec une variable log-gamma comme variable d'erreur.

ii. Let $\mu = -\frac{\log\lambda - \log\beta}{\alpha}$, $\sigma = \frac{1}{\alpha\sqrt{\beta}}$ et $Q = \beta^{-1/2}$. Trouvez la densité de T en fonction de (μ, σ, Q).

iii. Calculez la vraisemblance maximale pour le modèle gamma généralisée pour les données de rechutes.
Aide : Utilizez la commande `flexsurvreg` de **R** pour trouver le maximum (après avoir installé la bibliothèque `flexsurv` et avoir fait `library(flexsurv)`) La paramétrization utilisée par `flexsruvreg` pour la densité gamma généralisée est celui de la partie précédante avec (μ, σ, Q).

(c) Tracez le graphique de la fonction de survie empirique (basé sur les données)
Aide : la fonction de répartition empirique est $\hat{F}(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{t_i \leq t\}$, et $\hat{S}(t) = 1 - \hat{F}(t)$, où n est la taille de l'échantillon $\{t_1, \ldots, t_n\}$, et $\mathbf{1}\{t_i \leq t\} = 1$ ssi $t_i \leq t$, et il est zéro sinon.

(d) Sur ce même graphique, tracez les fonctions de survie de la log-normale et la loi gamma avec le paramètres que vous avez trouvé plus tôt. Commentez : Est-ce que le modèle log-normale semble approprié pour ces données ? Est-ce que le modèle gamma semble approprié pour ces données ?

(e) On peut montrer que lorsque $Q \to 0$, la loi gamma généralisée converge (en loi) à une loi log-normale. Comparez l'ajustement de la loi gamma généralisée à celui de la loi log-normale. En particulier, tracez sur le même graphique précédante la fonction de survie de la gamma généralisée avec le paramètres que vous avez trouvé plus tôt. Lequel modèle semble plus approprié pour les données de rechute. Expliquez.

(f) Faites un test d'hypothèses pour tester si la loi log-normale est adéquate en partant de la loi gamma généralisée.

2.- **[50 points]**

Considérez le modèle suivant, où les temps de décès T_i des individus $i = 1, \ldots, n$ suivent une loi exponentielle de paramètre $\lambda_i = e^{\beta x_i}$, où x_i est une valeur supposée non-aléatoire et connue pour chaque patient (une variable explicative). Les temps T_i sont indépendants d'un individu à un autre.

(a) Écrire la vraisemblance du modèle.

(b) Quelle équation doit satisfaire l'estimateur du maximum de vraisemblance de β ?

(c) Quelle est la variance asymptotique de l'estimateur du maximum de vraisemblance $\hat{\beta}$?

On veut tester l'hypothèse nulle que $\beta = 1$ contre l'alternative $\beta \neq 1$.

(d) Décrivez la procédure de test pour un niveau α quelconque (statistique de test et seuil de rejet) pour le test de Wald (basé sur la normalité asymptotique de $\hat{\beta}$.)

(e) Décrivez la procédure de test pour un niveau α quelconque (statistique de test et seuil de rejet) pour le test du rapport de vraisemblance (utilisez la loi asymptotique du rapport de vraisemblance).