

Survie

Feuille de TP : généralités

Exercice 1

Considérons n individus et leurs temps de survie respectifs $\tilde{T}_1, \dots, \tilde{T}_n$. Considérons que les données sont aléatoirement censurées à droite. On observe donc (T_i, Δ_i) , $i = 1 \dots, n$ avec $T_i = \min(\tilde{T}_i, C_i)$, où C_i la censure et $\Delta_i = \mathbb{1}_{\tilde{T}_i \leq C_i}$, l'indicateur de censure. On se propose d'estimer S la fonction de survie de \tilde{T} à partir des observations (T_i, Δ_i) , $i = 1 \dots, n$. Soient

$$\hat{S}_1(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{T_i > t} \text{ et } \hat{S}_2(t) = \frac{1}{\sum_{i=1}^n \Delta_i} \sum_{i=1}^n \mathbb{1}_{T_i > t, \Delta_i = 1}$$

1. Rappeler la définition de $S(t)$.
2. Interpréter $\hat{S}_1(t)$.
3. Vers quelle fonction $\hat{S}_1(t)$ converge-t'elle ? Conclusion ?
4. Interpréter $\hat{S}_2(t)$.
5. Vers quelle fonction $\hat{S}_2(t)$ converge-t'elle ? Conclusion ?

Exercice 2 (*Simulations de données censurée sous R*)

1. Simuler 100 temps d'intérêt et 100 temps de censure suivants tous les deux une loi uniforme de paramètres $[0, 5]$.
2. Créer le vecteur des observations (T_i, Δ_i) .
3. Tracer la fonction de répartition des "vrais" temps d'intérêt et comparer avec les estimateurs "naïfs" :

$$\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{T_i > t}, \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{T_i > t, \Delta_i = 1} \text{ et } \frac{1}{\sum_{i=1}^n \Delta_i} \sum_{i=1}^n \mathbb{1}_{T_i > t, \Delta_i = 1}.$$

Calculer également les moyennes de chaque estimateur et comparer à la "vraie" moyenne.

4. Simuler de nouveau en changeant la taille d'échantillon et commenter. Simuler de nouveau en changeant la proportion de données censurées.
5. Simuler de nouveau en changeant les lois des temps d'intérêt et de la censure. On pourra étudier des lois Exponentiels et Weibull.

Exercice 3 (*Introduction à R- Étude préliminaire*)

On s'intéresse au délai d'apparition d'une infection localisée au point d'insertion du cathéter chez des patients souffrant de maladie rénale et ayant un équipement de dialyse portable. Le cathéter peut être enlevé pour des raisons autres que l'infection, ce qui conduit à des observations censurées. Les données se trouvent sous R après avoir chargé la librairie `survival`, en tapant `kidney`. Le jeu de données contient les informations suivantes :

Patient : numéro du patient

Time : délai d'apparition d'une infection

Status : 0 pour observation censurée, 1 sinon

Age : âge

Sex : 1 pour homme 2 pour femme

Disease : type de la maladie (0=GN, 1=AN, 2=PKD, 3=Other).

1. Compter le nombre de données censurées dans les deux groupes définis par le sexe.
2. Récupérer les âges respectifs des hommes et des femmes dans deux tableaux. Faire une représentation graphique de l'âge en fonction du sexe.
3. Calculer la moyenne et la variance de l'âge dans les deux groupes définis par le sexe.
4. Comparer la moyenne d'âge entre les hommes et les femmes.
5. Comparer la répartition du type d'atteinte rénale en fonction du sexe.
6. L'âge moyen est-il le même quel que soit le type d'atteinte rénale ?