– TP 1 : Modèle exponentiel –

L'objectif de ce premier TP est d'illustrer les résultats établis en cours autour du modèle exponentiel en présence de censure. Le programme R commenté sera sauvegardé et transmis par mail.

Exercice 1 On utilise les notations du cours. Soit X une variable aléatoire représentant une durée de vie (durée s'écoulant jusqu'à la survenue d'un événement d'intérêt), de loi exponentielle de paramètre $\lambda = 1$.

1) *Simulation d'un échantillon.* Simuler à l'aide de la fonction `rexp()` un échantillon de taille $n = 50$ de couples (X_i, C_i), $i = 1, \cdots, n$, avec X_i indépendant de C_i, de lois exponentielles de paramètres respectifs $\lambda = 1$ et $\mu = 0.5$.

Définir un vecteur $T = \min(X, C)$ et un vecteur $\delta = \mathbb{I}(X \leq C)$. On pourra utiliser `as.numeric(X==T)`.

Déterminer (calcul à la main) la "proportion" théorique de censure c'est-à-dire la probabilité $\mathbb{P}(T > C)$.

Calculer la proportion observée de censure sur l'échantillon simulé c'est-à-dire

$$1 - \frac{1}{n}\sum_{i=1}^{n} \delta_i.$$

Répéter plusieurs fois ce calcul pour plusieurs réalisations de l'échantillon et observer la fluctuation de la proportion de censure.

2) Calculer une estimation ponctuelle du paramètre λ pour l'échantillon simulé. Répéter plusieurs fois ce calcul pour plusieurs réalisations de l'échantillon. Observer la fluctuation de l'échantillonnage.

3) *Illustration de la convergence p.s de $\hat{\lambda}_n$ vers λ.*

Reprendre le mécanisme de simulation de l'échantillon censuré de la question **1)** et faire varier la taille n de l'échantillon.

Pour n allant de 10 à 5000, calculer les estimateurs $\hat{\lambda}_n$ et les représenter graphiquement en fonction de n. On représentera aussi sur le même graphique la vraie valeur du paramètre $\lambda = 1$. Observer la convergence de $\hat{\lambda}_n$.

4) *Intervalle de confiance de λ pour un niveau de confiance de 95% pour l'échantillon complètement observé.*

Reprendre la simulation de la question **1)**. Pour n allant de 10 à 5000, calculer les bornes de l'intervalle de confiance dans le cas non-censuré, pour chaque valeur de n :

$$\hat{\lambda}_n\left(1-\frac{1.96}{\sqrt{n}}\right)\leq\lambda\leq\hat{\lambda}_n\left(1+\frac{1.96}{\sqrt{n}}\right)$$

5) *Intervalle de confiance de λ pour un niveau de confiance de 95% pour l'échantillon censuré.*

Même question que **4)** avec les formules de bornes de l'intervalle de confiance dans le cas censuré.

6) *Consistance de l'estimateur plug-in de la fonction de survie en présence de censure.*
Pour un échantillon de taille n, calculer les valeurs de l'estimateur plug-in $\hat{S}_n(t)=\exp(-\hat{\lambda}_n t)$ avec t un vecteur de pas 0.01 allant de 0 à 3. On pourra utiliser la fonction `seq(0,3,0.01)`.

Tracer la courbe représentative de $\hat{S}_n(t)$ en utilisant la fonction `plot()`.

Superposer la courbe théorique $S(t)=\exp(-t)$ en utilisant la fonction `curve()`.

Ajouter les bandes de confiance vues en cours.

Reprendre la question pour différentes tailles d'échantillons $n=50,100,200,1000$.

7) *Illustration de la convergence en loi de l'estimateur $\hat{\lambda}_n$.*
Pour un échantillon de taille $n=100$, on considère les variables :

$$\hat{Z}_n=\sqrt{\sum_{i=1}^{n}\delta_i}\left(\frac{\hat{\lambda}_n-\lambda}{\hat{\lambda}_n}\right)\quad\text{et } Z_n=\sqrt{n\mathbb{P}(\delta_1=1)}\left(\frac{\hat{\lambda}_n-\lambda}{\lambda}\right)$$

On a vu en cours la convergence en loi de :

$$Z_n\rightsquigarrow\mathcal{N}(0,1)\quad\text{et } \hat{Z}_n\rightsquigarrow\mathcal{N}(0,1)$$

Simuler un échantillon de 1000 réalisations de la variable $\hat{Z}_n$ et de la variable Z_n pour $n=100$.

Calculer la fonction de répartition empirique de $\hat{Z}_n$ et de Z_n à partir des 1000-échantillons. On utilisera la fonction `ecdf()`.

Tracer les fonctions de répartition empirique de $\hat{Z}_n$ et de Z_n entre -2.5 et 2.5 et superposer la courbe de la fonction de répartition de la loi normale standard $\mathcal{N}(0,1)$. Observer l'adéquation des fonctions de répartitions empiriques avec la fonction de répartition $\mathcal{N}(0,1)$.