M1 SSD - UE Tests - TP1

Dans ce TP, on s'intéresse à la problématique de l'obsolescence des appareils électroniques. Certains fabricants sont soupconnés d'obsolescence programmée : les appareils électroniques auraient une durée de vie programmée dans un certain intervalle de temps (par exemple entre 5 et 6 ans de vie). Au contraire, en l'absence d'obsolescence programmée, on s'attend à avoir des durées de vie aléatoires, prenant des valeurs dans $\mathbb{R}^+$. Une modélisation classique des durées de vie est alors d'utiliser la loi de Weibull. Cet exercice a pour but d'étudier cette loi, de proposer un estimateur des paramètres ainsi qu'un intervalle de confiance et un test statistique sur ce paramètre.

1. On admet que la durée de vie est représentée par une variable aléatoire X suivant une loi de Weibull de paramètres θ, a et c positifs. Cette loi, notée $W(\theta, a, c)$ a pour fonction de répartition

$$F(x) = (1 - \exp(-\frac{(x-a)^c}{\theta}))I\{x > a\},$$

   et donc pour densité la fonction

$$f(x) = \frac{c}{\theta}(x-a)^{c-1}\exp(-\frac{(x-a)^c}{\theta})I\{x > a\}.$$

   1. Montrer que la variable aléatoire $Y = \frac{2}{\theta}(X-a)^c$ suit la loi du χ^2_2.
   2. Que représente le paramètre a?
   3. Simuler $n = 100$ variables aléatoires $X_1, \dots X_n$ sous la loi de Weibull de paramètres $a = 0$, $c = 1/2$ et $\theta = 1$ avec la fonction `rweibull` (les paramètres sont `shape=c`, `scale = theta^(1/c)`, vous expliquerez pourquoi). Calculer les variables $Y_1, \dots, Y_n$ correspondantes. Proposer des graphiques pour illustrer leurs distributions.

2. Dans la suite on suppose connus les paramètres a et c de la loi de $W(\theta, a, c)$, le paramètre θ étant inconnu. On dispose d'un échantillon $(X_1, \dots, X_n)$ des durées de vie observées sur n appareils électroniques, X_i étant des réalisations i.i.d. de la variable aléatoire X.

   1. Calculer la vraisemblance du modèle.
   2. Montrer que l'estimateur du maximum de vraisemblance de θ est:

$$\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - a)^c.$$

   3. Cet estimateur est-il sans biais? convergent?
   4. On veut illustrer les différentes propriétés de l'estimateur par des simulations. Quel protocole de simulations permet d'illustrer le biais ? la convergence ? l'efficacité ? Mettre en place ces protocoles de simulation.

3. On cherche à construire un intervalle de confiance pour θ.

   1. Quelle est la loi exacte de l'estimateur $\hat{\theta}_n$? Construire un intervalle de confiance pour θ de niveau exact 90% (donner la formule).
   2. Donner une loi approchée de l'estimateur $\hat{\theta}_n$. Construire l'intervalle de confiance associé, qui est un intervalle de niveau asymptotique 90% (donner la formule).

3. Quel protocole de simulations permet d'illustrer le niveau de confiance de ces intervalles ? Quel protocole permet d'illustrer l'influence de la taille de l'échantillon n ? Mettre en place ces illustrations.

4. On veut maintenant mettre en place un test statistique de l'hypothèse simple $H_0 : \ \theta = \theta_0$ contre l'hypothèse alternative simple $H_1 : \ \theta = \theta_1$, vérifiant $\theta_1 > \theta_0 > 0$.

   1. Le lemme de Neyman-Pearson conduit à une région critique de la forme

   $$R = \{(X_1, \dots, X_n) : \sum_{i=1}^{n} (X_i - a)^c \geq k\},$$

   où $k > 0$ est une constante. Soit $0 < \alpha < 1$ un niveau de signification fixé et soit $\chi^2_{2n}(\gamma)$ le quantile d'ordre γ de la loi χ^2_{2n}. Déterminer en fonction de θ_0 et de $\chi^2_{2n}(\gamma)$, la région critique d'un test de niveau α.

   2. Exprimer en fonction de θ_0, θ_1 et de $\chi^2_{2n}(\gamma)$ et à l'aide de la fonction de répartition F de la loi χ^2_{2n}, le risque de seconde espèce et la puissance de ce test.

   3. Mettre en place sous R ce test pour un échantillon de taille $n = 100$ avec les paramètres $a = 0$, $c = 1/2$ et $\theta = \theta_0 = 1$. Illustrer le niveau du risque de première espèce de ce test. Illustrer numériquement la puissance du test en faisant varier α pour un θ_1 fixé, puis en faisant varier θ_1 pour α fixé.

5. On considère maintenant le test statistique de l'hypothèse $H_0 : \ \theta \leq 1$ contre $H_1 : \theta > 1$.

   1. Proposer un test uniformément le plus puissant pour un niveau de signification $0 < \alpha < 1$ fixé.

   2. Proposer un test asymptotique de niveau α de l'hypothèse $H_0 : \ \theta \leq 1$ contre $H_1 : \ \theta > 1$.

   3. Proposer une illustration numérique du niveau des deux tests en faisant varier α.