TP2 - ESTIMATION PAR INTERVALLES

Soit $X_1, \ldots, X_n$ un échantillon aléatoire issu d'une loi exponentielle de paramètre $\lambda > 0$ inconnu. On s'intéresse ici au paramètre $\mu = 1/\lambda$, qui est la moyenne des X_i.

L'objectif de ce TP est de construire plusieurs intervalles de confiance, de niveau de confiance $(1-\alpha)$, pour $\alpha \in (0,1)$ fixé, puis d'investiguer s'ils ont les pourcentages de couvertures attendus, et enfin de les comparer.

1. Ecrire l'intervalle de confiance asymptotique pour μ basé sur le théorème de la limite centrale, sans utiliser l'expression de la variance. On notera I_1 cet intervalle.

2. Ecrire l'intervalle de confiance asymptotique pour μ basé sur le théorème de la limite centrale, en utilisant le fait que la variance d'une $\mathcal{E}$xp est égale à μ^2. On notera I_2 cet intervalle.

3. On cherche une transformation monotone permettant de stabiliser la variance, en déterminant une fonction φ telle que l'on ait :

$$\sqrt{n}\left(\varphi(\bar{X}) - \varphi(\mu)\right) \xrightarrow{\mathcal{L}} \varphi'(\mu)\sigma(\mu)Z \ ,$$

   où $Z \sim \mathcal{N}(0,1)$ et avec $\varphi'(\mu)\sigma(\mu) = \text{cte}$. Expliciter cette fonction φ et en déduire un nouvel intervalle de confiance, que l'on notera I_3.

4. Calculer la fonction de répartition de la variable définie par

$$Y_n = \max_{i=1\ldots n} X_i/\mu \ ,$$

   et en déduire un intervalle de confiance pour μ à $100(1-\alpha)\%$. On notera I_4 cet intervalle.

5. En se souvenant du lien entre loi exponentielle et loi gamma, ainsi que des propriétés de la loi gamma[1], vérifier que $Z_n = n\bar{X}/\mu$ suit une loi $\gamma(n,1)$, puis en déduire un intervalle de confiance que l'on notera I_5.

[1] Si X et Y sont indépendantes, de loi respective $\gamma(a_X, b)$ et $\gamma(a_Y, b)$, alors $X+Y \sim \gamma(a_X+a_Y, b)$; si $X \sim \gamma(a,b)$, alors $cX \sim \gamma(a, b/c)$, pour tout $c > 0$.

Nous voici donc en possession de 5 intervalles de confiance.

6. Écrire une fonction R permettant de calculer les 5 intervalles de confiance à $100(1-\alpha)\%$ obtenus plus haut ; on spécifiera comme paramètre d'entrée l'échantillon $x_1, \ldots, x_n$ ainsi que la valeur de α, et on récupèrera en sortie une matrice de taille 5x2 contenant en ligne k le vecteur (borne inférieure ; borne supérieure) de l'intervalle I_k.

7. Écrire une fonction R permettant de calculer les pourcentages de couverture empiriques des 5 intervalles de confiance précédents. Cette fonction admettra comme paramètres d'entrée : `n` = taille des échantillons qui seront générés ; `nrep` = nombre d'échantillons à générer, μ = moyenne de la loi exponentielle simulée, et α. Elle aura comme paramètres de sortie une liste à deux éléments, contenant `couv` = vecteur des pourcentages de couverture empiriques des 5 intervalles de confiance ; et `longueur` = vecteur des longueurs médianes de chaque intervalle.

8. Effectuer un graphe qui contiendra en abscisse la taille de l'échantillon, et en ordonnée le pourcentage de couverture empirique associé à chacun des 5 intervalles de confiance, ainsi que la valeur théorique attendue. Commenter.

9. Effectuer un graphe qui contiendra en abscisse la taille de l'échantillon, et en ordonnée la longueur médiane de chacun des 5 intervalles de confiance. Commenter.