Feuille de travaux dirigés 6 : intervalles de confiance

Exercice 1 (transformation par la fonction de répartition)**:**
Soit $X : \Omega \to \mathbb{R}$ une variable aléatoire de fonction de répartition F continue sur $\mathbb{R}$.

1. Montrer que la variable aléatoire $U = F(X)$ suit une loi uniforme sur $[0, 1]$.
   indication : introduire $F^{\leftarrow}(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}$. Montrer que $F(F^{\leftarrow}(u)) = u$ pour $0 < u < 1$ en utilisant la continuité à droite de F. Monter ensuite que pour $0 < u < 1$, $F(x) < u \iff x < F^{\leftarrow}(u)$.Conclure.
2. Montrer que ce n'est plus le cas si l'on ne suppose pas que X est continue.

Exercice 2 (Loi uniforme)**:**
Soit $(X_i)_{i=1,\ldots n}$ des variables aléatoire i.i.d. selon une loi uniforme sur $[0, \theta]$ ($\theta > 0$), et $X = (X_1, \ldots X_n)$.

1. On pose $M_n(X) = \max_{i=1}^n X_i$. Donner la loi de M_n.
2. En déduire une fonction pivotale $\varphi(M_n(X), \theta)$ telle que la loi de $Z = \varphi(M_n(X), \theta)$ ne dépende pas de θ. On utilisera la question précédente et le résultat de l'exercice 1.
3. En déduire un intervalle de confiance pour le paramètre θ de type $I = [M_n(X), R(X)]$ de niveau de confiance $1 - \alpha$.
4. Pour $\theta_0 > 0$, on considère le test d'hypothèse : $H_0 : \theta \geq \theta_0$. En utilisant la dualité tests-intervalles de confiance, et en étudiant le sens de variation de la fonction $\theta \mapsto \mathbb{P}_\theta(M_n(X) < c)$ (c fixé), construire un test de l'hypothèse H_0 (contre $H_1 = H_0^c$) de niveau α.
5. Application numérique : $n = 10$, et $M_n(X) = 5$. Quel intervalle de confiance de niveau 95% obtenez-vous ? Rejetez-vous l'hypothèse $H_0 : \theta \geq 6$?

Exercice 3 (Moyenne et variance empirique, Cochran)**:**
Soit $X \sim \mathcal{N}(0, \mathbf{I}_n)$. Soient E_1 et E_2 deux sous-espaces orthogonaux de $\mathbb{R}^n$ de dimension respective r et s et p_1 et p_2 les projecteurs orthogonaux, associés.

1. Soient $(h_1, \ldots, h_r)$ et $(h_{r+1}, \ldots, h_{r+s})$ deux bases orthonormées respectivement de E_1 et E_2. On note $H_1 = [\mathbf{h_1} \ldots \mathbf{h_r}]$ et $H_2 = [\mathbf{h_{r+1}} \ldots \mathbf{h_{r+s}}]$ les matrices correspondantes dans la base canonique.

   (a) Vérifiez que les matrices P_1 et P_2 de p_1 et p_2 s'écrivent respectivement $P_1 = H_1 H_1^\top$ et $H_2 H_2^\top$.

   (b) Montrer que les vecteurs $V_1 = H_1^\top X$ et $V_2 = H_2^\top X$ sont indépendants et calculer leur loi.

   (c) Montrer que les projetés $P_1 X$ et $P_2 X$ son indépendants et donner leurs lois.

   (d) Montrer que $\|P_1 X\|^2 \sim \chi^2(r)$, $\|P_2 X\|^2 \sim \chi^2(s)$ et que ces deux variables sont indépendantes.

   Les deux points c) et d) ci-dessus sont connus sous le nom de « théorème de Cochran », très utilisé pour la modélisation par régression linéaire.

2. On pose $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ et $S_n = \sum_{i=1}^n (X_i - \bar{X}_n)^2$. On considère le projecteur orthogonal p_1 sur le vecteur constant $\mathbb{1}_n = (1, \dots, 1)$ et p_2 le projecteur orthogonal sur le supplémentaire orthogonal de $\mathrm{vect}(\mathbb{1}_\mathrm{n})$ dans $\mathbb{R}^n$.
   (a) Exprimer $V_1 = H_1^\top X$ en fonction de $\mathbb{1}_n$ et de $\bar{X}_n$.
   (b) Exprimer $\|P_2 X\|^2$ en fonction de S_n.
   (c) En déduire que $\frac{\sqrt{n}\bar{X}_n}{\sqrt{S_n/(n-1)}}$ suit une loi de Student à $(n-1)$ degrés de liberté.
3. Soit maintenant $Y = (Y_1, \dots, Y_n)$ où $Y_i \overset{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2)$. On pose $\bar{Y}_n = \frac{1}{n}\sum Y_i$ et $\widehat{\sigma^2}_n = \frac{1}{n}\sum_{i=1}^n (Y_i - \bar{Y}_n)^2$. Déduire de ce qui précède que $\widehat{\sigma^2}_n$ et $\bar{Y}_n$ sont indépendantes et que $Z = \frac{\sqrt{n}(\bar{Y}_n - \mu)}{\sqrt{\frac{n}{n-1}\widehat{\sigma^2}_n}}$ suit une loi de Student à $n-1$ degrés de libertés. En déduire un intervalle de confiance pour μ de niveau $(1-\alpha)$, fondé sur l'observation Y.

Exercice 4 (Intervalle de confiance pour la variance)**:**
Soient X et V deux variables gaussiennes réelles, indépendantes, de variances respectives σ_X^2 et σ_V^2. La variable X est un signal utile et V est un bruit de mesure. L'observation est donnée par $Y = X + V + \mu$ où μ est un biais systématique de l'appareil, inconnu. On reçoit n observations indépendantes. On connaît $\sigma_V^2 = 1$ mais pas σ_X^2 ni μ. On veut construire un intervalle de confiance pour σ_X^2 de niveau $(1-\alpha)$.

En utilisant le résultat de l'exercice 3, questions 1) et 2) construisez une variable aléatoire $Z = \varphi(Y, \sigma_X^2)$ distribuée selon une loi du χ^2 à $(n-1)$ degrés de libertés.

En déduire la construction d'un intervalle de confiance pour σ_X^2, en fonction des quantiles de la loi du $\chi^2(n-1)$.