TD 3 : intervalles de confiance

Les questions marquées d'un astérisque **(*)** *sont facultatives.*

Exercice 0. Questions de cours.

1. Soit $\{\mathbb{P}_\theta, \theta \in \Theta\}$ un modèle statistique et f une fonction $\Theta \to \mathbb{R}$. Soit $\alpha \in [0,1]$. Rappeler la définition d'un intervalle de confiance de niveau $1-\alpha$ de $f(\theta)$.
2. Soit X une variable aléatoire réelle qui admet une densité strictement positive par rapport à la mesure de Lebesgue et $\alpha \in (0,1)$. Le **quantile** d'ordre α de X est le réel q tel que $\mathbb{P}(X \leq q) = \alpha$.
   Pour une définition générale, cherchez « Fonction quantile ».
   En notant z_β le quantile d'ordre β de la loi $\mathcal{N}(0,1)$, en déduire un intervalle de confiance de niveau $1-\alpha$
   - bilatéral (de la forme $[a,b]$)
   - unilatéral à droite (de la forme $[a,+\infty)$)
   - unilatéral à gauche (de la forme $(-\infty,b]$)

   de l'espérance de la loi $\mathcal{N}(\mu,\sigma^2)$ fondé sur une observation $X \sim \mathcal{N}(\mu,\sigma^2)$, en supposant que σ est connu.
3. On admet la conséquence suivante du Théorème de Portmanteau :
   « Soit X une variable aléatoire réelle et $(X_n)_{n\geq 1}$ une suite de variables aléatoires réelles telle que X_n converge vers X en loi. Alors pour tout intervalle $[a,b]$ tel que $\mathbb{P}(X = a \text{ ou } X = b) = 0$,
   $$\mathbb{P}(X_n \in [a,b]) \underset{n\to\infty}{\longrightarrow} \mathbb{P}(X \in [a,b]). \text{ »}$$
   Montrer que si $(Y_n)_{n\geq 1}$ est une suite de variables aléatoires réelles telle que $\sqrt{n}\frac{Y_n-\mu}{\sigma} \to \mathcal{N}(0,1)$ en loi, alors $\mathbb{P}(\mu \in [Y_n - \sigma\frac{z_{1-\alpha/2}}{\sqrt{n}}, Y_n + \sigma\frac{z_{1-\alpha/2}}{\sqrt{n}}]) \underset{n\to\infty}{\longrightarrow} 1-\alpha$ pour tout $\alpha \in (0,1)$.
   On dit que $[Y_n - \sigma\frac{z_{1-\alpha/2}}{\sqrt{n}}, Y_n + \sigma\frac{z_{1-\alpha/2}}{\sqrt{n}}]$ *est un intervalle de confiance* ***asymptotique*** *de niveau* $1-\alpha$ *de* μ.

Exercice 1. Soient $X_1,\dots,X_n$ des v.a. i.i.d. de loi $\mathcal{E}(\lambda)$ avec λ inconnu.

1. Rappeler la formule d'un estimateur $\hat{\lambda}_n$ du maximum de vraisemblance de λ.
2. Dans cette question uniquement, on suppose qu'il existe $\lambda_0 > 0$ connu tel que $\lambda \geq \lambda_0$.
   (a) En utilisant l'inégalité de Bienaymé-Tchebychev, donner un intervalle de confiance de niveau $1-\alpha$ de $1/\lambda$.
   (b) En déduire un intervalle de confiance de niveau $1-\alpha$ de λ.
3. (a) Énoncer le théorème central limite sur la moyenne empirique $\bar{X}_n$.
   (b) Montrer qu'il existe $\sigma(\lambda) > 0$ tel que sous le paramètre λ,
   $$\sqrt{n}(\hat{\lambda}_n - \lambda) \underset{n\to+\infty}{\longrightarrow} \mathcal{N}(0,\sigma(\lambda)^2) \quad \text{en loi}$$
   et donner la formule de $\sigma(\lambda)$.
   (c) En utilisant le théorème de Slutsky, en déduire un intervalle de confiance asymptotique de niveau $1-\alpha$ de λ.
4. **(*)** Refaire la question 3 dans le cas où $X_1,\dots,X_n \sim \mathcal{B}(p)$. Comparer les cas où p proche de $1/2$ et où p proche de zéro.

Exercice 2. On répète de manière indépendante une expérience ayant une probabilité p de succès. On suppose $p \geq p_0$ pour $p_0 > 0$ connu.

1. Combien de répétitions faut-il faire pour être sûr avec probabilité au moins $1 - \alpha$ d'obtenir au moins un succès ?

   Autrement dit, si on note $\mathbb{P}_p$ la loi des expériences lorsque la probabilité de succès est p, donner un entier N ne dépendant pas de p tel que pour tout $p \geq p_0$, $\mathbb{P}_p$(au moins un succès lors de N expériences) $\geq 1 - \alpha$.
2. Rappeler l'énoncé de l'inégalité de Hoeffding.
3. Combien de répétitions faut-il faire pour être sûr avec probabilité au moins $1 - \alpha$ d'obtenir au moins k succès ?

Exercice 3. On sonde n personnes sur k questions où elles peuvent répondre oui ou non. On suppose que les personnes répondent indépendamment les unes des autres (mais les réponses d'une même personne aux questions ne sont, elles, pas indépendantes), et que la probabilité de répondre oui à une question donnée ne dépend pas de la personne. On note p_j la probabilité qu'une personne dise oui à la question j, et $X_{i,j}$ la variable qui vaut 1 si la personne i dit oui à la question j et 0 sinon.

1. Donner un intervalle de confiance de niveau $1 - \alpha$ de p_j pour tout j.
2. En déduire une région de confiance de niveau $1 - \alpha$ pour le vecteur $p = (p_1, \dots, p_k)$.

Exercice 4. (*) Soient $X_1, \dots, X_n$ des v.a. i.i.d. de loi $\mathcal{N}(\mu, \sigma^2)$ pour (μ, σ^2) inconnu. On admet que la matrice de variance-covariance de (X_1, X_1^2) est

$$\begin{pmatrix} \sigma^2 & 2\mu\sigma^2 \\ 2\mu\sigma^2 & 2\sigma^4 + 4\mu^2\sigma^2 \end{pmatrix}$$

1. Énoncer le théorème central limite pour le couple $(\frac{1}{n}\sum_{i=1}^n X_i, \frac{1}{n}\sum_{i=1}^n X_i^2)$.
2. En utilisant la delta-méthode, en déduire la normalité asymptotique de l'estimateur de la variance $\frac{n-1}{n} S_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ où $\bar{X}_n = \frac{1}{n}\sum_{j=1}^n X_j$.
3. En déduire la normalité asymptotique de S_n^2. *Indication pour vérifier vos calculs : la variance de la limite dépend-elle de μ ?*
4. En déduire un intervalle de confiance de σ^2.
5. **(*)** Démontrer la formule de la matrice de variance-covariance.