Exercice 1. Soit θ le pourcentage (inconnu) de fumeurs dans une population. On considère un échantillon de n personnes et on estime θ avec $\hat{\theta}_n$ = proportion de fumeurs dans l'échantillon. Comment choisir n pour garantir que

$$\mathrm{P}(|\hat{\theta}_n - \theta| < 0.005) \geq 0.95 \quad ?$$

Indication : Utiliser une approximation normale.

Exercice 2. Intervalle de confiance et test d'hypothèse.
Durant l'année 2009 (année choisie au hasard), sur $n = 78286$ naissances en Suisse, on a observé $x = 40407$ naissances de garçons. Soit $\theta := \mathrm{P}($ naissance d'un garçon $)$.

(a) Quelle est la loi de la v.a. X, le nombre de naissance de garçons, sous θ ?

(b) Donner un intervalle de confiance à 95% pour θ en se basant sur l'approximation par une loi normale. Donner $I(x)$.

(c) Tester l'hypothèse $H_0 : \theta = \theta_0 = 1/2$ contre l'alternative $H_1 : \theta \neq 1/2$ en utilisant le test

$$\phi(X) = \begin{cases} 1 & \text{si } \frac{1}{2} \notin I(X) \\ 0 & \text{sinon} \end{cases}$$

On définit $\beta_\phi(\theta) = \mathbb{E}_\theta(\phi(X))$. Que vaut et comment interpréter $\beta_\phi(\frac{1}{2})$?

(d) Comparer avec le test binomial exact `binom.test` sur `R`.

Exercice 3. Loi uniforme - intervalle de confiance.
On considère un échantillon aléatoire $X_1, \ldots, X_n$ où les X_i sont distribués uniformément sur $[0, \theta]$.

(a) Donner un intervalle de confiance de niveau $1 - \alpha$ pour θ.
Indication : Considérer la distribution de $X_{(n)}/\theta$ et proposer un intervalle de la forme $[X_{(n)}; c \cdot X_{(n)}]$ où $X_{(n)}$ est le maximum des X_i et $c > 1$.

(b) Que vaut c lorsque $n = 10, 100, 1000$ ou 10000 et $\alpha = 0.05$?

Exercice 4. Borne inférieure non-paramétrique pour la médiane.
Soit $X := (X_1, X_2, \ldots, X_n)$ où les X_i sont i.i.d. avec une fonction de répartition F, continue et strictement croissante. La médiane θ_F de F est définie comme l'unique valeur pour laquelle $F(\theta_F) = \frac{1}{2}$. Soit $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ l'échantillon $(X_1, X_2, \ldots, X_n)$ ordonné et soit $0 < \alpha < 1$.

(a) Calculer $\mathrm{P}_F(X_{(k)} \leq \theta_F) = \mathrm{P}_F($au moins k parmi $X_1, X_2, \ldots, X_n \leq \theta_F)$.

(b) Montrer qu'il existe un $1 \leq k \leq n$ et un $0 \leq \rho \leq 1$ tels que si $L(X) = X_{(k)}$ avec probabilité ρ et $L(X) = X_{(k+1)}$ avec probabilité $1 - \rho$ alors

$$\mathrm{P}_F(L(X) \leq \theta_F) = 1 - \alpha, \quad \forall F$$

Indication : Choisir le plus grand nombre k tel que $\mathrm{P}_F(X_{(k)} \leq \theta_F) \geq 1 - \alpha$.

(c) Calculer explicitement k et ρ dans le cas où $n = 100$ et $\alpha = 0.05$.
Indication : Dans `R`, la fonction `pbinom(m,n,p)` calcule $\sum_{j=0}^{m} \binom{n}{j} p^j (1-p)^{n-j}$. Dans le cas où `m` est un vecteur, la fonction est appliquée à chaque composante de `m`.

À rendre jusqu'au vendredi 26 avril, à 12h.