Exercice 1. (8 points). Modèle de Hardy-Weinberg.
On considère le modèle de Hardy-Weinberg : on essaie d'estimer dans une population la fréquence $\theta \in \Theta = (0,1)$ d'une allèle de type A, en ne pouvant observer que les phénotypes AA, aA et aa. Pour le faire, on dispose de n observations i.i.d. $X_1, \ldots, X_n$ avec

$$\mathrm{P}_\theta(X_1 = \mathrm{AA}) = \theta^2, \quad \mathrm{P}_\theta(X_1 = \mathrm{aA}) = 2\theta(1-\theta) \quad \text{et} \quad \mathrm{P}_\theta(X_1 = \mathrm{aa}) = (1-\theta)^2.$$

On définit encore les variables aléatoires $N_1 := \sum_{i=1}^n \mathbb{I}_{\{\mathrm{AA}\}}(X_i)$ et $N_2 := \sum_{i=1}^n \mathbb{I}_{\{\mathrm{aA}\}}(X_i)$ le nombre d'individus observés de phénotype AA resp. aA.

(a) Montrer que l'estimateur du maximum de vraisemblance pour θ est $\hat{\theta} = \frac{2N_1+N_2}{2n}$.

(b) Montrer que $\hat{\theta}$ est sans biais.

(c) Calculer $\mathrm{Var}(\hat{\theta})$
Indication : pour calculer $\mathbb{E}(N_1N_2)$ développer le produit en termes de fonctions indicatrices.

(d) Calculer l'information de Fisher $I(\mathrm{P}_\theta)$ et montrer le lien entre $I(\mathrm{P}_\theta)$ et $\mathrm{Var}(\hat{\theta})$.

(e) Montrer que l'estimateur $\hat{\theta}$ est asymptotiquement normal, plus précisément :

$$\sqrt{nI(\mathrm{P}_\theta)}(\hat{\theta} - \theta) \stackrel{\mathcal{L}}{\Longrightarrow} \mathcal{N}(0,1) \quad \text{lorsque} \quad n \to \infty.$$

(f) Illustrer le point précédent par simulation.
Pour ce faire, choisir un $\theta \in (0,1)$ et tirer 1000 échantillons de taille n (employer la fonction `rmultinom`), chaque échantillon donnant un estimateur $\hat{\theta}$. Dessiner l'histogramme de ces 1000 valeurs normalissées selon la formule ci-dessus et comparer avec la loi normale standard.

Exercice 2. Estimateur sans biais.
On considère la variable aléatoire X suivant une loi de Poisson tronquée, i.e.

$$\mathrm{P}_\theta(X = k) = \frac{\theta^k}{k!}(\exp(\theta) - 1)^{-1} \quad \text{pour } k = 1, 2, \ldots$$

Montrer que le seul estimateur sans biais pour $(1 - \exp(-\theta))$ est donné par

$$T(k) = \begin{cases} 0 & \text{si } k \text{ est impair} \\ 2 & \text{si } k \text{ est pair} \end{cases}$$

Pourquoi cet estimateur est-il absurde ?

Exercice 3. Information de Fisher et loi normale.
Dans cet exercice, nous supposons que les hypothèses du théorème de Cramer-Rao sont satisfaites.

(a) Supposons que $X_1, X_2, \ldots, X_n$ soit un n-échantillon issu de la variable aléatoire X de densité $g_\theta(x)$ et désignons par $\mathrm{I}_n(\theta)$ son information de Fisher. Vérifier que $\mathrm{I}_n(\theta) = n\mathrm{I}_1(\theta)$.

(b) Soit $X_1, X_2, \ldots, X_n$ un n-échantillon issu de la variable aléatoire X de loi normale $\mathcal{N}(\theta, \sigma^2)$ (on suppose σ connu). Calculer l'information de Fisher de X et en déduire que parmi tous les estimateurs sans biais pour θ, $\bar{X}_n$ est celui de variance minimale.

À rendre jusqu'au vendredi 19 avril, à 12h.