Interrogation n°2

**Question :** Rappeler brièvement la méthodologie de construction d'un intervalle (bilatéral) de confiance pour l'espérance $m$ d'une loi normale $N(m, \sigma^2)$ :

- lorsque l'écart-type $\sigma$ est connu
- lorsque l'écart-type $\sigma$ est inconnu

**Exercice 1 :**

1. On considère un échantillon $(X_1, X_2, .., X_n)$ tiré de façon i.i.d. dans $B(1,p)$.

1.a. Donner les expressions de la vraisemblance et de la log-vraisemblance de cet échantillon.

1.b. En déduire que l'estimateur du maximum de vraisemblance du paramètre $p$ est la proportion empirique $\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

1.c. Calculer l'espérance et la variance de l'estimateur $\hat{p}_n$. En déduire que $\hat{p}_n$ est convergent.

1.d. L'estimateur $\hat{p}_n$ est-il efficace?

2. On suppose dans cette question que $n = 400$ et que la valeur observée de l'estimateur $\hat{p}_n$ est $0,4$. On admet que l'approximation normale de la loi binômiale est valable.

2.a. Déterminer une borne inférieure de confiance à 95% pour le paramètre $p$.

2.b. Déterminer une borne supérieure de confiance à 95% pour le paramètre $p$.

3. On suppose maintenant qu'on dispose d'un autre échantillon $(Y_1, Y_2, .., Y_m)$ tiré de façon i.i.d dans une loi de Bernouilli de paramètre $q$. On suppose en outre que les variables $X_1, X_2, .., X_n, Y_1, Y_2, .., Y_m$ sont indépendantes.

3.a. Déterminer, sous l'hypothèse de validité de l'approximation normale da la loi binômiale, l'expression d'un intervalle bilatéral de confiance à 95% pour la différence $p - q$.

3.b. Application : On effectue deux sondages aléatoires indépendants, l'un auprès de la population masculine et l'autre auprès de la population féminine pour déterminer la proportion de fumeurs $p$ dans l'ensemble de la population masculine et la proportion de fumeuses $q$ dans l'ensemble de la population féminine. Le premier sondage porte sur un échantillon de 300 hommes dont 105 déclarent fumer et le second porte sur un échantillon de 200 femmes dont 48 déclarent fumer. Déterminer un intervalle bilatéral de confiance à 95% pour la différence $p - q$.

**Exercice 2 :**

Considérons le modèle linéaire :

$$y_i = ax_i + b + \varepsilon_i \qquad i = 1, .., N$$

sous les hypothèses standards $E(\varepsilon_i) = 0$; $V(\varepsilon_i) = \sigma^2$ et $cov(\varepsilon_i, \varepsilon_j) = 0$ pour $i \neq j$.

1. Que minimise-t-on lorsqu'on utilise la méthode des moindres carrés? En déduire les équations normales qui caractérisent les estimateurs des MCO $\hat{a}$ et $\hat{b}$ des paramètres $a$ et $b$. Résoudre ces équations et retrouver les expressions classiques des estimateurs $\hat{a}$ et $\hat{b}$.

2. Rappeler les propriétés des estimateurs $\hat{a}$ et $\hat{b}$.

3. Donner (sans démonstration) l'expression d'un estimateur sans biais de $\sigma^2$.

4. Posons, pour $i = 1, .., N$,

$$q_i = y_i - \bar{y}; \quad p_i = x_i - \bar{x}; \quad u_i = \varepsilon_i - \bar{\varepsilon}$$

4.a. Montrer que pour $i = 1, ..N$,

$$q_i = ap_i + u_i$$

4.b. Montrer que $E(u_i) = 0$ pour $i = 1, ..N$

4.c. Montrer que, pour $i = 1, ..N$, $V(u_i) = \sigma^2\left(1 - \frac{1}{N}\right)$ et $cov(u_i, u_j) = -\frac{\sigma^2}{N}$ pour $i \neq j$.