TD 5 : Tests d'hypothèse

Exercice 1. Soit $(X_1, X_2, \cdots, X_n)$ un n-échantillon de loi exponentielle de paramètre $\theta > 0$.

1) Construire un test de H_0 : "$\theta = \theta_0$" contre H_1 : "$\theta > \theta_0$" de niveau α.

2) Construire un test de H_0 : "$\theta = \theta_0$" contre H_1 : "$\theta \neq \theta_0$" de niveau α.

Exercice 2. Si un nouveau-né pèse moins de $2,5$ kg à la naissance, on dira que son poids est faible. Aux États-Unis, il a été observé qu'environ 7% des bébés ont effectivement un poids faible. On peut utiliser cette donnée comme indicateur de malnutrition des parents.

On note p la proportion des nouveaux-nés au Soudan dont le poids est faible, on observe $n = 209$ bébés, 23 d'entre eux ont un poids faible.

1) Donner un bon estimateur du paramètre p.

2) Trouver des intervalles de confiance asymptotiques de niveau 95%, 99% (on utilisera de préférence le TCL, le paramètre p étant trop petit pour exploiter la méthode des variances uniformément bornées).

3) On veut tester H_0 : "$p = 0,07$" contre H_1 : "$p > 0,07$".

- **a)** Que peut-on en conclure au niveau $\alpha = 5\%$? Et au niveau $\alpha = 1\%$?
- **b)** Calculer la p-valeur associée à ce test ?

On utilisera les valeurs suivantes pour les quantiles d'une loi $\mathcal{N}(0,1)$: $q_{0,95} = 1,65$, $q_{0,975} = 1,96$, $q_{0,99} = 2,33$ *et* $q_{0,995} = 2,58$.

Exercice 3. Des plaignants[1] ont poursuivi en justice le Ministère israëlien de la Santé suite à une campagne de vaccination menée sur des enfants et ayant entraîné des dommages fonctionnels irréversibles pour certains d'entre eux. Ce vaccin était connu pour entraîner ce type de dommages en de très rares circonstances. Des études antérieures menées dans d'autres pays ont montré que ce risque était d'un cas sur 310 000 vaccinations. Les plaignants avaient été informés de ce risque et l'avaient accepté. Les doses de vaccin ayant provoqué les dommages objet de la plainte provenaient d'un lot ayant servi à vacciner un groupe de 300 533 enfants. Dans ce groupe, quatre cas de dommages ont été détectés.

1) On modélise l'événement "le vaccin provoque des dommages fonctionnels irréversibles sur l'enfant i" par une variable aléatoire de Bernoulli, X_i, de paramètre p. Calculer la valeur p_0 correspondant aux résultats des études antérieures.

2) Justifier qu'on peut modéliser la loi du nombre N de cas de dommages par une loi de Poisson de paramètre θ. Calculer la valeur θ_0 attendue si le vaccin est conforme aux études antérieures.

1. cf. Murray Aitkin, Evidence and the Posterior Bayes Factor, 17 Math. Scientist 15 (1992)

3) L'hypothèse H_0 : "$p = p_0$" correspond au risque que les plaignants avaient accepté, l'hypothèse alternative étant H_1 : "$p > p_0$". Construire un test de niveau α à partir de la variable N. Accepte-t'on H_0 au seuil de 5% ? Donner la p-valeur de ce test.

Exercice 4. Une agence de voyage souhaite cibler sa clientèle. Elle sait que les coordonnées du lieu de vie d'un client (X, Y) rapportées au lieu de naissance $(0, 0)$ sont une information significative pour connaître le goût de ce client. Elle distingue :

⋆ La population 1 (Hypothèse H_0) dont la loi de répartition a pour densité :

$$p_1 : (x, y) \longmapsto \frac{1}{2\pi} \exp\left(-\frac{x^2 + y^2}{2}\right)$$

⋆ La population 2 (Hypothèse H_1) dont la loi de répartition a pour densité :

$$p_2 : (x, y) \longmapsto \frac{1}{16} 1_{[-2,2]}(x)\, 1_{[-2,2]}(y)$$

L'agence souhaite tester l'hypothèse qu'un nouveau client vivant en (x, y) appartient à la population 1 plutôt qu'à la population 2.

1) Proposer un test de niveau inférieur à $\alpha = 5\%$ et de puissance maximale, construit à partir du rapport de vraisemblance.

2) Donner une statistique de test et caractériser graphiquement la région critique dans $\mathbb{R}^2$.

Exercice 5. Pour des raisons géologiques, on considère habituellement que la taille des « grands » champs de pétrole (plus d'un million de barils) d'une région pétrolifère est convenablement modélisée comme étant la réalisation de n (n de l'ordre de quelques centaines ou quelques milliers) variables aléatoires indépendantes et identiquement distribuées $X_1, \ldots, X_n$ de loi p_θ de densité

$$f_\theta : x \longmapsto \theta x^{-(\theta+1)} \mathbb{1}_{]1,\infty[}(x)$$

où $\theta > 0$ est le paramètre inconnu.

1) Justifier que f_θ est bien une densité de probabilité.

2) Calculer la queue de la fonction de répartition associée, c'est-à-dire $\mathbb{P}_\theta(X_1 > x)$ pour tout $x > 1$. Un géologue considère que θ est un paramètre mesurant la dispersion au sein du bassin pétrolier : comment reliez-vous cette opinion au calcul qui précède ?

3) Déterminer $\widehat{\theta}_n$ l'estimateur du maximum de vraisemblance de θ. Est-il sans biais et consistant ?

4)a) En utilisant la méthode Delta, déterminer une fonction g telle que

$$\sqrt{n}\left(g\left(1/\widehat{\theta}_n\right) - g(1/\theta)\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, 1)$$

b) En déduire un intervalle de confiance pour θ de niveau 92%, faisant apparaître $\widehat{\theta}_n$.

5) Construire le test du rapport de vraisemblance de niveau 4% de l'hypothèse H_0 : "$\theta \geq 1$" contre H_1 : "$\theta < 1$".