UNIVERSITÉ DES LAGUNES Année universitaire 2021-2022

Licence 2 Sciences et technologie **TD de Statistique 2**

Exercice 1

Soit T_1 et T_2 deux estimateurs sans biais et indépendants, d'un parametre θ, de variances respectives σ_1^2 et σ_2^2, deux nombres reels strictement positifs. Soit α un réel appartenant à $[0, 1]$.

1. Montrer que la variable aléatoire $T = \alpha T_1 + (1-\alpha)T_2$ est aussi un estimateur sans biais de θ.
2. Pour quelle valeur de α, la variance de l'estimateur T est-elle minimale ?

Exercice 2

Soient $X_1, ..., X_n$ des variables aléatoires indépendantes suivant une loi avec fonction de densité de probabilité

$$f_\theta(x) = (\theta+1)x^\theta \ \ \forall \ \ 0 \leq x \leq 1 \ \text{ avec } \ \theta > 0.$$

1. Déterminer l'estimateur des moments de θ.
2. Déterminer l'estimateur du maximum de vraisemblance de θ.

Exercice 3

Soit $Y_1, ..., Y_n$ des réalisations indépendantes d'une variable aléatoire distribuée selon une loi dont la fonction de densité de probabilité est la suivante :

$$f(y;\theta) = \theta^{-1}|y| \exp\left(-\frac{y^2}{\theta}\right), \ \ y \in \mathbb{R}, \ \ \theta > 0.$$

1. Déterminer l'estimateur du maximum de vraisemblance $\widehat{\theta}_{MV}$ de θ.
2. Sachant que $\mathbb{E}(Y_i^2) = \theta$ et $\mathbb{E}(Y_i^4) = 2\theta^2$, calculer l'espérance et la variance de $\widehat{\theta}_{MV}$.
3. L'estimateur du maximum de vraisemblance $\widehat{\theta}_{MV}$ est-il sans biais ?, consistant ?
4. Calculer l'information de Fisher au point θ apportée par l'échantillon $(Y_1, ..., Y_n)$.
5. L'estimateur du maximum de vraisemblance $\widehat{\theta}_{MV}$ est-il efficace ?

Exercice 4

Soient $X_1, ..., X_n$ des variables aléatoires indépendantes et identiquement distribuées dont la fonction de densité de probabilité est définie par

$$f_\theta(x) = \begin{cases} \frac{x}{\theta}\exp\left(-\frac{x^2}{2\theta}\right), & \text{si } x > 0 \text{ avec } \theta > 0\,; \\ 0, & \text{sinon.} \end{cases}$$

1. Déterminer l'estimateur des moments de θ.
2. Déterminer l'estimateur du maximum de vraisemblance $\widehat{\theta}_{MV}$ de θ.
3. Montrer que la variable aléatoire $X_i^2/2$ suit une loi exponentielle de paramètre $1/\theta$.
4. $\widehat{\theta}_{MV}$ est-il sans biais ?
5. Existe-t-il un autre estimateur non biaisé de θ dont la variance est strictement plus petite que celle de $\widehat{\theta}_{MV}$?

Exercice 5

Soit X une variable aléatoire dont la fonction de densité de probabilité est définie, pour $\theta > 0$, par

$$f_\theta(x) = \begin{cases} \frac{1}{\theta} x^{-1-1/\theta}, & \text{si } x > 1\,; \\ 0, & \text{sinon.} \end{cases}$$

1. Calculer $\mathbb{E}(X)$ et en déduire un estimateur T_n de θ par la méthode des moments, construit à partir d'un échantillon $(X_1, ..., X_n)$ de X.
2. Déterminer la loi de probabilité de la variable aléatoire $Y = \ln(X)$.
3. Déterminer l'estimateur $\widehat{\theta}_n$ de θ par la méthode du maximum de vraisemblance.
4. Etudier les propriétés de l'estimateur $\widehat{\theta}_n$ (biais, efficacité, consistance).

Exercice 6

Soit X une variable aléatoire suivant une loi uniforme sur l'intervalle $[0, 2a]$. On considère une suite $X_1, ..., X_n$ de n variables aléatoires indépendantes et de même loi que X.

On pose $\overline{X} = \frac{1}{n}(X_1 + ... + X_n)$ et $T = \max(X_1, ..., X_n)$.

1. Calculer l'espérance $\mathbb{E}(X)$ et la variance $V(X)$ de la variable aléatoire X.
2. Montrer que $\overline{X}$ est un estimateur convergent de a.
3. Déterminer la densité de la variable aléatoire T. Calculer l'espérance et la variance de T.
4. En déduire un autre estimateur $\overline{T}$ sans biais de a.
5. Comparer les estimateurs $\overline{X}$ et $\overline{T}$ de a.

Exercice 7

Soit $X_1, ..., X_n$ un échantillon aléatoire provenant d'une distribution géométrique de paramètre p qui s'écrit

$$P(X = x) = (1-p)^{x-1} p, \quad x = 1, 2, 3, ...$$

1. Déterminer l'estimateur des moments de p.
2. Déterminer l'estimateur du maximum de vraisemblance de p.

Exercice 8

Dans une usine de fabrication, il est produit en série des tôles métalliques. La surface X des tôles est une variable dont il est admis qu'elle est normale d'écart-type égal à 4. Après mise en place d'un nouveau processus de fabrication, afin de déterminer une estimation de la moyenne m de X, on prélève un échantillon de 28 tôles. On trouve une moyenne empirique $\overline{X} = 45,25 dm^2$.

1. Construire un intervalle de confiance pour m au niveau 95% en supposant que l'écart-type n'a pas changé au moment de la mise en place du nouveau processus.
2. Pour un même niveau de confiance, on souhaite réduire la largeur de l'intervalle trouvé dans la question précédente en choisissant un échantillon de taille supérieure. En souhaitant obtenir une largeur d'intervalle de $1 dm^2$, quelle doit être la taille du nouvel échantillon ?
3. Suite à la mise en place du nouveau processus, on considère maintenant que l'écart-type de la variable X ne peut pas être supposé invariant. On relève l'écart-type empirique de l'échantillon et on trouve $s = 4 dm^2$. Construire le nouvel intervalle de confiance de m avec le même niveau de confiance et le comparer à celui trouvé dans la première question.

Exercice 9
Un fabricant de voitures fait un test d'efficacité d'un certain modèle. On mesure les litres de consommation d'essence pour 100 kilomètres et on obtients les résultats suivants :
$14,60 \;\; 11,21 \;\; 15,56 \;\; 11,37 \;\; 13,68 \;\; 11,06 \;\; 26,58 \;\; 13,37 \;\; 15,98 \;\; 12,07 \;\; 13,22 \;\; 12,01 \;\; 15,07$.
On suppose que la consommation d'essence suit une loi normale de moyenne m et de variance $\sigma^2 = 16$.
1. Calculer un intervalle de confiance à 95% pour la moyenne.
2. Quel niveau de confiance correspond à un intervalle de longueur $3litres/100km$?
3. Combien d'observations additionnelles sont nécessaires pour avoir un intervalle à 99% de longueur $2litres/100km$?

Exercice 10
On considère un échantillon de 40 paquets de biscuits provenant d'une production de 2000 unités. Le poids moyen obtenu pour cet échantillon est égal à $336g$ et l'écart-type empirique, c'est-à-dire la quantité s, est égal à $0,86g$.
Quelle est l'estimation, par intervalle de confiance, du poids moyen de ces paquets de biscuits, pour l'ensemble de la fabrication, avec un niveau de confiance $0,98$?

Exercice 11
Dix bouteilles d'eau minérale provenant d'une source donnée sont analysées. On relève les taux de nitrates suivants, en mg/l : $3,61 \;\; 3,56 \;\; 3,67 \;\; 3,56 \;\; 3,64 \;\; 3,62 \;\; 3,44 \;\; 3,52 \;\; 3,55 \;\; 3,52$.
Donner un intervalle de confiance à 95% pour l'écart-type du taux de nitrates dans les bouteilles produites (on supposera ce taux gaussien).

Exercice 12
Lors d'une élection opposant 2 candidats, un sondage d'opinion réalisé sur un échantillon de 1000 personnes donne 52% des voix au candidat A et 48% au candidat B.
1. Donner un intervalle de confiance au niveau $0,95$ des intentions de vote pour A.
2. Combien suffirait-il interroger de personnes pour qu'il y ait moins de 5% de chances que B l'emporte, si A a recueilli 52% des intentions de vote dans le sondage ?

Exercice 13
On désire tester si la durée de vie moyenne d'un tube électronique est égale à 1600 heures ou si elle est plutôt inférieure à cette valeur. Les observations sur un échantillon de taille 16 suivent une loi normale avec $\overline{X} = 1590$ heures et $s = 30$ heures.
1. Donner les hypothèses H_0 et H_1.
2. Peut-on rejeter H_0 au seuil de 1% ?
3. Calculer l'erreur de deuxième espèce et la puissance du test au seuil de 1% pour $\mu = 1570$.

Exercice 14
Les teneurs en azote de 30 échantillons d'un même terrain ont été mesurées par 2 méthodes d'analyse différentes : 15 échantillons par la méthode A et 15 autres par la méthode B. Sur la base des résultats suivants et en considérant un modèle normal, peut-on admettre, au seuil de 5%, que les

2 méthodes donnent en moyenne des résultats analogues ?
- Méthode A
3,51 3,01 3,33 3,31 3,54 3,17 3,50 2,72 3,24 3,48 1,97 2,85 2,51 2,93 1,83
- Méthode B
2,34 3,12 3,30 2,15 3,13 2,84 2,97 3,65 3,89 3,23 2,46 2,70 2,44 2,41 2,85

Exercice 15
Le tableau suivant reprend des données concernant le niveau des dépenses mensuelles pour l'achat de produits cosmétiques observé sur un échantillon aléatoire simple de 500 femmes adultes différenciées selon leur statut professionnel.

	Plein temps	Temps partiel	Sans profession
Moins de 10 dollars	30	20	60
De 10 à 25 dollars	55	60	65
Plus de 25 dollars	55	80	75

Sur la base de ce tableau, peut-on dire qu'il y a indépendance entre le niveau de dépenses et le statut professionnel (utilisez un risque de première espèce $\alpha = 5\%$).

Exercice 16
Afin de mieux gérer les demandes de crédits de ses clients, un directeur d'agence bancaire réalise une étude relative à la durée de traitement des dossiers, supposée suivre une distribution normale. Un échantillon de 30 dossiers a donné :

Durée de traitement (en jours)	$[0,10[$	$[10,20[$	$[20,30[$	$[30,40[$	$[40,50[$	$[50,60[$
Effectif	3	6	10	7	3	1

1. Déterminer un estimateur de la moyenne m par la méthode du maximum de vraisemblance. Etudier ses propriétés.
2. Donner une estimation ponctuelle de la moyenne m.
3. Donner une estimation de m par intervalle de confiance au seuil de risque 5%.
4. Au seuil de 5%, tester l'hypothèse $H_0 : m = 30$ contre $H_1 : m < 30$. Que pouvez-vous conclure ?

Exercice 17
La société "Votre santé" est une entreprise de vente par correspondance de produits de beauté dits "naturels". Elle gère un fichier de 350000 clients et propose chaque mois une offre promotionnelle accompagnée d'un cadeau. Le taux de réponse à cette offre est généralement de 15%, la marge moyenne par réponse de 340 fcfa. Mlle Claire, nouvellement en charge de ce fichier, a retenu comme cadeau un abonnement gratuit de six mois au mensuel "Votre beauté Madame". Elle pense que cela pourrait augmenter le taux de réponse à la prochaine offre ; toutefois cette proposition ne serait rentable que si le taux de réponse dépassait les $17,5\%$ (avec la même marge moyenne évidemment). Elle envisage de tester la réalité de ces hypothèses sur un échantillon de clientes. La précision voulue pour son estimation est de l'ordre de 2%.
1. Quelle taille d'échantillon doit-elle choisir afin d'atteindre la précision voulue (avec un niveau de confiance de $0,95$) ?
2. Les résultats d'un sondage sur un échantillon de 1225 clientes vous sont donnés ci-dessous :

	Total	Anciens clients
Nombre d'individus	1225	850
Nombre de réponses	258	193

a. Donner une estimation par intervalle de confiance au niveau $0,95$ du pourcentage p de réponses positives attendues à cette offre.
b. Mlle Claire se propose de procéder au test d'hypothèses suivant $H_0 : p = 17,5\%$ contre $H_1 : p > 17,5\%$. Qu'en concluez-vous ?
c. Mlle Claire pense que les nouveaux clients (inscrits depuis moins de 6 mois) ont un taux de réponse inférieur aux anciens. Confirmer ou infirmer cette hypothèse.