

Examen 29 juin 2017 (9h45 - 11h45)

Les exercices sont indépendants et peuvent être traités dans n'importe quel ordre. Il est possible d'admettre un résultat des questions précédentes pour résoudre les questions suivantes. Tous les résultats doivent être justifiés. La notation prendra en compte la clarté et le soin de la rédaction. Le seul document autorisé est le résumé de cours manuscrit que vous avez préparé. La calculatrice est autorisée. Quelques quantiles utiles sont reportées en fin d'énoncé.

Barème indicatif: exercice 1: 5 points, exercice 2: 8 points, exercice 3: 3 points, exercice 4: 4 points

Exercice 1

Il est habituellement admis que la séquence (publicités et bandes-annonces) précédant le début de la projection d'un film au cinéma dure 15 min en moyenne, mais il vous semble avoir constaté une augmentation. Inspiré par le cours MA101 vous décidez d'appliquer un test statistique. Sur un mois, vous sélectionnez aléatoirement des séances. Les temps constatés (en minutes) sur ces séances sont 14 min, 21 min, 15 min, 20 min, 17 min, 13 min, 18 min.

1. En préambule, vous effectuez sur votre échantillon un test de Shapiro-Wilks de niveau 5% testant l'hypothèse nulle suivant laquelle les données sont issues d'une loi gaussienne. La région de rejet de ce test est de la forme $\{W < q_\alpha\}$ où W est la statistique de test et q_α un seuil défini pour garantir un niveau α . Un logiciel donne une p-value de 0.74 sur vos données.
 - (a) Rappelez la définition de la p-value.
 - (b) Quelle décision prenez-vous, et avec quel risque?
 - (c) Sur un schéma, représenter schématiquement le seuil, la région de rejet, la p-value et le niveau.
2. La durée de la séquence précédant le début d'un film a-t-elle augmenté? Quel test proposez-vous d'utiliser et précisez-le: modèle, hypothèses (H_0) et (H_1), statistique de test, loi sous (H_0), région de rejet.
Quelle décision prenez-vous au niveau $\alpha = 10\%$? Quel est le risque ?
3. En fait, vous aviez enregistré les durées de séquence dans le même cinéma il y a cinq ans. Sur 5 séances, le temps d'attente moyen était de 15 min avec un écart-type corrigé de 4 min. En faisant l'hypothèse d'un modèle d'égalité des variances, proposez un autre test prenant en compte cette nouvelle situation. Quelle est votre décision et son risque? Commenter le résultat.
4. La démarche précédente est-elle modifiée si on ne fait pas l'hypothèse gaussienne, et si oui de quelle façon?

Exercice 2

Dans un grand lac, on a remarqué l'apparition de tâches de décoloration sur certains poissons. Afin d'étudier l'ampleur de ce phénomène, n expériences sont réalisées. Une expérience consiste à capturer de façon indépendante des poissons un par un, jusqu'à l'obtention d'un poisson décoloré. La population piscicole est suffisamment grande pour considérer que l'opération a lieu avec remise.

1. Soit X la variable aléatoire modélisant le nombre de captures nécessaires jusqu'à l'obtention d'un premier poisson décoloré. Montrer que

$$\mathbb{P}(X = x) = p(1 - p)^{x-1}, \quad x \in \mathbb{N}^*$$

où p est la proportion (inconnue) de poissons décolorés.

Note: cette loi est appelée loi géométrique $\mathcal{G}(p)$, et elle est telle que $\mathbb{E}(X) = 1/p$ et $\text{var}(X) = (1 - p)/p^2$.

n expériences $(X_1, \dots, X_n)$ ont été menées indépendamment dans différentes parties du lac supposé peuplé de façon homogène.

2. Proposer un estimateur des moments $\hat{p}$ de p . Est-il consistant?
3. Montrer que $\hat{p}$ est asymptotiquement normal de variance limite $p^2(1 - p)$.
On pourra utiliser la delta-méthode: Soit $\hat{\theta}$ un estimateur consistant de $\theta \in \mathbb{R}$ et asymptotiquement normal $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V)$ et soit h une fonction de classe C^1 de dérivée h' , alors $\sqrt{n}(h(\hat{\theta}) - h(\theta)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, [h'(\theta)]^2 V)$
4. Proposer la construction d'un intervalle de confiance bilatéral de niveau asymptotique 95% de la proportion de poissons atteints.
Application: 30 expériences de capture ont amené à capturer 1500 poissons au total.
5. Construire un test asymptotiquement UPP de $p = p_0$ contre $p = p_1$, où $p_1 > p_0$.
Application: le ministère de la pêche a défini un seuil d'alerte à 3% de la population piscicole, avec un taux de fausse alerte de 1%. L'alerte doit-elle être lancée? Quel est le risque de cette décision?
6. Déterminer l'expression de la courbe de puissance du test précédent. La représenter schématiquement. Le test est-il sans biais? Est-il consistant?

Exercice 3

Soit un n -échantillon i.i.d. $(X_1, \dots, X_n)$ de loi mère de Pareto $X_i \sim \mathcal{P}(a, \theta)$ de densité définie sur $[a; +\infty[$ par

$$f(x) = \theta a^\theta x^{-(\theta+1)}.$$

On rappelle que l'espérance de la loi de Pareto vaut $\theta a/(\theta - 1)$ pour $\theta > 1$ et sa variance $\theta a^2/[(\theta - 1)^2(\theta - 2)]$ pour $\theta > 2$.

Le paramètre θ est supposé connu et est supérieur à 2.

1. Construire un estimateur $\hat{a}$ de a par la méthode des moments. Déterminer son risque.
2. Soit $M_n = \min(X_1, \dots, X_n)$. Calculer la fonction de répartition de M_n . En déduire que M_n suit une loi de Pareto $\mathcal{P}(a, n\theta)$. L'estimateur M_n est-il non biaisé pour estimer a ?
3. Calculer le risque de M_n . $\hat{a}$ est-il admissible?

Exercice 4

On cherche à calculer la constante π par simulation statistique.

1. Dans le plan $2D$, soit $\mathcal{C}$ le carré centré en l'origine du repère et de côté 2, et soit $\mathcal{D}$ le disque unitaire centré à l'origine. Soit $U = (X, Y)$ la loi uniforme sur ce carré, de densité

$$f(x, y) = c \mathbb{I}_{[-1;1]}(x) \times \mathbb{I}_{[-1;1]}(y)$$

Quelle est la valeur de c ?

2. On définit la variable aléatoire Z qui vaut 1 si U appartient à $\mathcal{D}$, 0 sinon. Quelle est la loi de Z ? Préciser la valeur de son paramètre p .
3. En déduire un estimateur de π . Quel est son risque? Est-il consistant?
4. Estimer π par intervalle de confiance de niveau 99%. Combien faut-il de simulations de U pour assurer que cet intervalle de confiance ait une précision de $\pm 10^{-2}$? Commenter ce résultat.
5. Un intervalle de confiance de niveau 99% est calculé avec un échantillon de taille 10000. La vraie valeur de π n'appartient pas à l'intervalle de confiance observé. Y a-t-il un problème dans la simulation?

Quelques quantiles utiles

Quantile d'ordre α de la loi gaussienne centrée réduite:

α	0.90	0.95	0.975	0.99	0.995	0.999	0.9995
quantile	1.28	1.64	1.96	2.33	2.58	3.09	3.29

Quantile d'ordre α de la loi de Student en fonction du nombre de degrés de liberté (ddl):

$\alpha \setminus \text{ddl}$	4	5	6	7	8	9	10	11	12	13	14
0.9	1.53	1.48	1.44	1.41	1.40	1.38	1.37	1.36	1.36	1.35	1.35
0.95	2.13	2.02	1.94	1.89	1.86	1.83	1.81	1.80	1.78	1.77	1.76
0.975	2.78	2.57	2.45	2.36	2.31	2.26	2.23	2.20	2.18	2.16	2.14
0.99	3.75	3.36	3.14	3.00	2.90	2.82	2.76	2.72	2.68	2.65	2.62

Quantile d'ordre α de la loi du Khi-deux en fonction du nombre de degrés de liberté (ddl):

$\alpha \setminus \text{ddl}$	4	5	6	7	8	9	10	11	12	13	14
[1,]	7.78	9.24	10.64	12.02	13.36	14.68	15.99	17.28	18.55	19.81	21.06
0.9	9.49	11.07	12.59	14.07	15.51	16.92	18.31	19.68	21.03	22.36	23.68
0.95	11.14	12.83	14.45	16.01	17.53	19.02	20.48	21.92	23.34	24.74	26.12
0.975	13.28	15.09	16.81	18.48	20.09	21.67	23.21	24.72	26.22	27.69	29.14
0.99	14.86	16.75	18.55	20.28	21.95	23.59	25.19	26.76	28.30	29.82	31.32

Quantile d'ordre 0.95 de la loi de Fisher en fonction du nombre de degrés de liberté au numérateur (num) et dénominateur (dénom):

dénom/num	3	4	5	6	7
3	9.28	9.12	9.01	8.94	8.89
4	6.59	6.39	6.26	6.16	6.09
5	5.41	5.19	5.05	4.95	4.88
6	4.76	4.53	4.39	4.28	4.21
7	4.35	4.12	3.97	3.87	3.79