

Examen 26 juin 2018 (9h00 - 12h00)

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 Partie I: 5 points; exercice 1 Partie II: 4 points; exercice 2: 6 points; exercice 3: 5 points.

Exercice 1.

On considère n variables indépendantes $X_1, \dots, X_n$ suivant la même loi de densité

$$f_\theta(x) = \frac{x}{\theta} \exp \left[-\frac{x^2}{2\theta} \right] I_{\mathbb{R}^+}(x),$$

avec $\theta > 0$ et où $I_{\mathbb{R}^+}$ est la fonction indicatrice sur $\mathbb{R}^+$ ($I_{\mathbb{R}^+}(x) = 1$ si $x \geq 0$, $I_{\mathbb{R}^+}(x) = 0$ sinon).

Cette loi est appelée *loi de Rayleigh* de paramètre θ et notée $\mathcal{R}(\theta)$. On admettra que les premiers moments de cette loi sont donnés par les relations suivantes:

$$\mathbb{E}(X_i) = \sqrt{\frac{\pi}{2}} \sqrt{\theta}, \quad \mathbb{E}(X_i^2) = 2\theta, \quad \mathbb{E}(X_i^3) = \left(3\sqrt{\frac{\pi}{2}} \right) \theta^{3/2}, \quad \mathbb{E}(X_i^4) = 8\theta^2.$$

Les deux parties de l'exercice 1 peuvent être traitées de façon indépendante et doivent être composées sur des **copies différentes**.

Partie I. Estimation

1. Montrer que la log-vraisemblance de l'observation $X_1, \dots, X_n$ admet un unique maximum global, puis déterminer l'estimateur du maximum de vraisemblance de θ noté $\hat{\theta}$.
2. Déterminer le risque quadratique de $\hat{\theta}$. Montrer que $\hat{\theta}$ est un estimateur consistant du paramètre θ .
3. A l'aide de l'expression de l'espérance d'une loi de Rayleigh, donner un estimateur T_n de θ en utilisant la méthode des moments.
Justifier que T_n est un estimateur consistant. Est-il non biaisé?
4. Proposer un estimateur $\hat{\theta}_b$ non biaisé de θ fonction de $\bar{X}$.
Sans faire de calcul, que pensez-vous de l'assertion suivante: $\hat{\theta}_b$ domine T_n parce que son biais est nul.

***** CHANGER DE COPIE *****

Partie II. Test statistique On observe le résultat d'un n -échantillon $X_1, \dots, X_n$ de loi de Rayleigh $\mathcal{R}(\theta)$ définie précédemment.

On souhaite effectuer le test d'hypothèse nulle $H_0 : \theta = \theta_0$ contre l'hypothèse alternative $H_1 : \theta = \theta_1$ où θ_0 et θ_1 sont deux valeurs connues telles que $0 < \theta_1 < \theta_0$.

5. Ecrire l'expression de la vraisemblance que vous avez calculée à la question 1 de la partie I.
6. Rappeler la statistique de test et la forme de la région de rejet du test de Neyman-Pearson de (H_0) contre (H_1)
 Expliquer la propriété remarquable de ce test et préciser sous quelles conditions elle est valide.
7. Montrer que la région de rejet de ce test peut s'écrire sous la forme $\{\sum_i X_i^2 < c_\alpha\}$ où le seuil c_α dépend du niveau α du test.
8. La loi du χ^2 (*Khi-deux*) à d degrés de liberté admet pour fonction génératrice des moments
 $M(t) = \mathbb{E}(e^{tY}) = (1 - 2t)^{-d/2}$ et pour fonction de densité

$$g(x) = \frac{1}{2^{d/2}\Gamma(d/2)} x^{\frac{d}{2}-1} \exp(-x/2) I_{\mathbb{R}^+}(x),$$

où la fonction Gamma est définie par $\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt$ pour $x > 0$.

- (a) Montrer que si X_i suit une loi de Rayleigh de paramètre θ , alors $Y_i = X_i^2/\theta$ suit une loi du χ^2 à 2 degrés de liberté.
 - (b) Montrer que $S_n = \sum_{i=1}^n X_i^2/\theta$ suit la loi du χ^2 à $2n$ degrés de liberté.
 - (c) En déduire l'expression du seuil c_α du test de niveau α défini à la question 6.
9. On observe une réalisation d'un échantillon de loi de Rayleigh de taille $n = 7$, $x_1, \dots, x_7$, avec

$$\sum_{i=1}^7 x_i^2 = 10.5.$$

Effectuer au niveau 5% le test de Neyman-Pearson de $H_0 : \theta = 2$ contre $H_1 : \theta = 1$.

Justifier votre décision et son risque. Calculer la p-value du test.

Répondre directement sur l'énoncé, après avoir indiqué vos nom et prénom

NOM:

Prénom

Exercice 2.

On a déterminé dans le sang le taux x_i d'une certaine protéine sur $n = 15$ sujets atteints d'une même maladie. La moyenne des taux sur les 15 sujets vaut $m = \sum_i x_i/n = 141$, et l'écart-type corrigé $s = \sqrt{\sum_i (x_i - m)^2 / (n - 1)} = 5.8$.

1. La p-value du test de Shapiro-Wilks vaut 0.42. Pouvez-vous proposer une modélisation gaussienne? Préciser le risque de votre décision.
2. Calculer un intervalle de niveau de confiance 95% de l'espérance du taux de protéine de la population dont sont extraits les 15 individus. Construire également un intervalle de confiance de niveau 95% de la variance ce taux.

- 4

Répondre directement sur l'énoncé, après avoir indiqué vos nom et prénom

NOM: Prénom

Exercice 3.

Un gouvernement songe à une profonde réforme mais craint une agitation populaire à quelques mois d'une élection importante. Les renseignements généraux estiment que tant que 15% de la population n'est pas prête à se mobiliser activement (dans la rue ou sur les réseaux sociaux) contre la réforme, cette dernière peut passer.

Un institut de sondage interroge 1000 Français, parmi lesquels 980 s'expriment: 131 sont prêts à agir activement mais les 849 autres, tout en ne soutenant pas nécessairement la réforme, affirment qu'ils se contenteront au mieux de pester dans leur coin. Un conseiller non statisticien du Premier Ministre affirme que la proportion observée étant de $131/980 \simeq 13.4\%$, bien inférieure à 15%, le gouvernement doit se lancer.

- Indiquer la population visée et définir le modèle.
- Quelle taille d'échantillon retenir?
- Que pensez-vous de la réponse du non-statisticien?
- Préciser le paramètre d'intérêt. Sachant que le gouvernement est prudent et qu'il ne tient pas à mettre le feu aux poudres, définir les hypothèses du test (justifier votre choix).

- La statistique de test envisagée est de la forme $T_n = \sqrt{n} \left((\bar{X} - p_0) / \sqrt{p_0(1 - p_0)} \right)$.
Quelle valeur proposer pour p_0 ? Quelle est la loi asymptotique de T_n ? Est-il possible de l'utiliser sur l'échantillon considéré?

- Déterminer la région de rejet de niveau 5%.

- Quelle est la décision, et son risque?

- Calculer la p-value.

- Quelle conclusion stratégique apporter au Premier Ministre?

- Calculer et représenter la puissance de ce test.

Quelques quantiles utiles

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

α	0.90	0.92	0.95	0.975	0.99	0.995	0.999	0.9995
quantile	1.28	1.40	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	15
0.9	1.53	1.48	1.44	1.41	1.40	1.38	1.37	1.36	1.36	1.35	1.35	1.34
0.95	2.13	2.02	1.94	1.89	1.86	1.83	1.81	1.80	1.78	1.77	1.76	1.75
0.975	2.78	2.57	2.45	2.36	2.31	2.26	2.23	2.20	2.18	2.16	2.14	2.13
0.99	3.75	3.36	3.14	3.00	2.90	2.82	2.76	2.72	2.68	2.65	2.62	2.6
0.995	4.60	4.03	3.71	3.50	3.36	3.25	3.17	3.11	3.05	3.01	2.98	2.95
$\alpha \setminus \text{ddl}$	16	17	18	19	20	21	22	23	24	25	26	27
0.9	1.34	1.33	1.33	1.33	1.33	1.32	1.32	1.32	1.32	1.32	1.31	1.31
0.95	1.75	1.74	1.73	1.73	1.72	1.72	1.72	1.71	1.71	1.71	1.71	1.70
0.975	2.12	2.11	2.10	2.09	2.09	2.08	2.07	2.07	2.06	2.06	2.06	2.05
0.99	2.58	2.57	2.55	2.54	2.53	2.52	2.51	2.50	2.49	2.49	2.48	2.47
0.995	2.92	2.90	2.88	2.86	2.85	2.83	2.82	2.81	2.80	2.79	2.78	2.77

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

$\alpha \setminus \text{ddl}$	10	11	12	13	14	15	16	17	18	19	20
0.01	2.56	3.05	3.57	4.11	4.66	5.23	5.81	6.41	7.01	7.63	8.26
0.025	3.25	3.82	4.40	5.01	5.63	6.26	6.91	7.56	8.23	8.91	9.59
0.05	3.94	4.57	5.23	5.89	6.57	7.26	7.96	8.67	9.39	10.12	10.85
0.1	4.87	5.58	6.30	7.04	7.79	8.55	9.31	10.09	10.86	11.65	12.44
0.9	15.99	17.28	18.55	19.81	21.06	22.31	23.54	24.77	25.99	27.20	28.41
0.95	18.31	19.68	21.03	22.36	23.68	25.00	26.30	27.59	28.87	30.14	31.41
0.975	20.48	21.92	23.34	24.74	26.12	27.49	28.85	30.19	31.53	32.85	34.17
0.99	23.21	24.72	26.22	27.69	29.14	30.58	32.00	33.41	34.81	36.19	37.57

Quantile d'ordre 0.975 de la loi de Fisher en fonction du nombre de degrés de liberté au numérateur (num) et dénominateur (dénom):

dénom/num	11	12	13	14	15
11	3.47	3.43	3.39	3.36	3.33
12	3.32	3.28	3.24	3.21	3.18
13	3.20	3.15	3.12	3.08	3.05
14	3.09	3.05	3.01	2.98	2.95
15	3.01	2.96	2.92	2.89	2.86

On rappelle que $q_{\alpha}^{n,m} = 1/q_{1-\alpha}^{m,n}$