

Partiel n°3

Durée: 90 minutes.

Il sera tenu compte du soin apporté à la rédaction.

Attention : toute réponse devra être précisément justifiée. Une réponse oui ou non ne sera pas suffisante pour obtenir les points.

Le barème est indicatif, il vise à vous donner une idée du point relatif de chaque question, puis je multiplierais la note par un certain coefficient en fonction des résultats de tous les étudiants.

Vous avez le droit à votre feuille bleue mais pas à la calculatrice.

1 Estimation dans un modèle de Pareto

Pour $\theta \in]0, +\infty[$, on note $\mathcal{P}(\theta)$ la distribution sur \mathbb{R} ayant pour densité

$$p_\theta(x) = \frac{3\theta^3}{x^4} 1_{x \geq \theta}.$$

Calculs préliminaires

[a] (1pt) Montrer que si $X \sim \mathcal{P}(1)$, $\theta X \sim \mathcal{P}(\theta)$.

[b] (2pt) Si bien définies, calculer l'espérance et la variance d'une loi $\mathcal{P}(\theta)$.

Estimations : On se donne X_1, \dots, X_n i.i.d. de loi $\mathcal{P}(\theta)$, et on cherche à estimer θ . On note $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$.

[c] (0.5pt) Écrire le modèle sous la forme d'un triplet, espace d'observation, tribu, famille de probabilités.

[d] (0.5pt) Le modèle est-il identifiable ?

[e] (2pt) Proposer un estimateur $\hat{\theta}_1$ par la méthode des moments et calculer sa variance.

[f] (1pt) Soit $\hat{\theta}_2 = \frac{2}{3}\bar{X}$, montrer que $\hat{\theta}_2$ est un estimateur fortement consistant.

[g] (3pt) Pour cette question uniquement on suppose que $\theta \in [0, 10]$. En utilisant l'inégalité de Bienaymé-Tchebychev et $\hat{\theta}_2$, trouver un intervalle au niveau de confiance $1 - \alpha$ sur θ . On le notera I_1 .

[h] (3pt) Trouver un intervalle de confiance asymptotique (au niveau $1 - \alpha$) sur θ que l'on notera I_2 .

[i] (1,5pt) Déterminer l'estimateur du maximum de vraisemblance.

[j] (2pt) On note $m_n = \min_{i=1, \dots, n} X_i$. Soit $u > 0$. Montrer que

$$P_\theta \left(\frac{m_n}{\theta} \geq 1 + u \right) = \left(\frac{1}{1 + u} \right)^{3n}.$$

[k] (2pt) Montrer que $3n \left(\frac{m_n}{\theta} - 1 \right) \rightsquigarrow \mathcal{E}(1)$ (loi Exponentielle de paramètre 1).

[l] (1,5pt) En déduire un intervalle de confiance de niveau asymptotique $1 - \alpha$ sur θ , que l'on notera I_3 .

[m] (0.5pt) Lorsque $n \rightarrow +\infty$, comparer les longueurs de I_2 et I_3 , lequel préférez-vous ?

2 On veut étudier la liaison entre les caractères : «être fumeur» (plus de 20 cigarettes par jour, pendant 10 ans) et «avoir un cancer de la gorge», sur une population de 1000 personnes, dont 500 sont atteintes d'un cancer de la gorge. Voici les résultats observés : Tableau observé

Fumeur ? \ Malade ?	cancer	non cancer	tous
	fumeur	342	258
non fumeur	158	242	400

Identifier la situation d'échantillonnage et poser l'hypothèse nulle correspondant à la question informelle : le fait d'avoir un cancer est-il identique pour les fumeurs et les non fumeurs ?

- [a] (1pt) Calculer la statistique du test du χ^2 . On pourra se contenter de donner la réponse sur la forme d'une formule contenant des nombres réels et des opérations élémentaires (addition, puissance,...).
- [b] (1pt) Que conclue-t-on avec les risques de première espèce $\alpha = 0.01$ et $\alpha = 0.05$? On pourra donner la réponse sous la forme d'une condition contenant des formules exprimées avec des nombres réels et des opérations élémentaires (addition, puissance,...). Par contre il faudra expliciter les quantiles.
- [c] (1pt) Pourquoi a-t-on $K_{0.05} < K_{0.01}$ (où K_α est le seuil du test du χ^2) ?
- [d] (0.5pt) Peut-on calculer la puissance du test ? Pourquoi ?

[3] On s'intéresse au problème des algues toxiques qui atteignent certaines plages de France ; après étude on constate que 10% des plages sont atteintes par ce type d'algues et on veut tester l'influence de rejets chimiques nouveaux sur l'apparition de ces algues. Pour cela 500 plages proches de zones de rejet chimiques, sont observées ; on compte alors le nombre de plages atteintes par l'algue nocive : on constate que 100 plages sont atteintes par l'algue. On s'intéresse donc à $\theta > 0$ la proportion de plages atteintes par ce type d'algue quand il y a des rejets chimiques à proximité.

- [a] (2pt) Donner un intervalle de confiance de niveau asymptotique $1 - \alpha$ sur θ .
- [b] (1pt) Grace à l'intervalle de confiance précédent, pouvez-vous en déduire un test permettant de répondre à la question «Les rejets chimiques ont-t-il modifié, de façon significative, avec le risque $\alpha = 0,05$, le nombre de plages atteintes?»

[4] **Chaîne de Markov à deux états.** On considère une chaîne de Markov sur $E = \{1, 2\}$, dont la matrice de transition est $P = ((P(i, j))_{(i, j) \in E^2}) = \begin{pmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix}$.

On suppose que : $\mathbb{P}(X_0 = 1) = \nu(1)$ et $\mathbb{P}(X_0 = 2) = \nu(2) = 1 - \nu(1)$

- [a] (3pt) Calculer $\mathbb{P}(X_{n+1} = 1)$ en fonction de $\nu(1)$.
- [b] (2pt) Calculer P^n .
- [c] (1,5pt) Déterminer sa probabilité stationnaire π .

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of χ^2								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.09
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.21
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68	24.72
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03	26.22
13	4.107	5.892	7.042	9.299	12.340	15.98	19.81	22.36	27.69
14	4.660	6.571	7.790	10.165	13.339	17.12	21.06	23.68	29.14
15	5.229	7.261	8.547	11.037	14.339	18.25	22.31	25.00	30.58
16	5.812	7.962	9.312	11.912	15.338	19.37	23.54	26.30	32.00
17	6.408	8.672	10.085	12.792	16.338	20.49	24.77	27.59	33.41
18	7.015	9.390	10.865	13.675	17.338	21.60	25.99	28.87	34.80
19	7.633	10.117	11.651	14.562	18.338	22.72	27.20	30.14	36.19
20	8.260	10.851	12.443	15.452	19.337	23.83	28.41	31.41	37.57
22	9.542	12.338	14.041	17.240	21.337	26.04	30.81	33.92	40.29
24	10.856	13.848	15.659	19.037	23.337	28.24	33.20	36.42	42.98
26	12.198	15.379	17.292	20.843	25.336	30.43	35.56	38.89	45.64
28	13.565	16.928	18.939	22.657	27.336	32.62	37.92	41.34	48.28
30	14.953	18.493	20.599	24.478	29.336	34.80	40.26	43.77	50.89
40	22.164	26.509	29.051	33.660	39.335	45.62	51.80	55.76	63.69
50	27.707	34.764	37.689	42.942	49.335	56.33	63.17	67.50	76.15
60	37.485	43.188	46.459	52.294	59.335	66.98	74.40	79.08	88.38