

MAPSI – Examen réparti – 05/11/24 – 50 pts

Durée : 1h30

Seuls documents autorisés : Calculatrice, antiscèhes recto-verso
– Barème indicatif –

Exercice 1 (20pts) – ML & MAP végétal

Dans une pépinière, des jardiniers se sont amusés à collecter des données sur un échantillon de 1000 plantes vertes d'une espèce qui les intéresse particulièrement : leur idée était de distinguer 3 stades d'évolution de la plante, sachant que l'arrivée au 3^{ème} stade permet de mettre en vente lesdites plantes.

Temps pour arriver au stade 3 (en semaines)	1	2	3	4	5
Effectif	50	350	100	250	250

Q 1.1 Modélisation

Le statisticien de l'équipe propose de modéliser chaque semaine comme une épreuve de Bernoulli indépendante. On nomme X^* la loi de la durée (en semaine) pour arriver en phase 3, correspondant à la première réussite de l'épreuve.

Q 1.1.1 X^* suit alors une loi géométrique. Donner la formule pour $p(X^* = k)$

Q 1.1.2 Exprimer la log-vraisemblance de l'échantillon en fonction du paramètre. Puis estimer la valeur optimale de ce paramètre au sens du maximum de vraisemblance.

Q 1.2 Un collègue statisticien, également féru de jardinage, fait remarquer que l'expérience a été menée en automne et que les données sont peut-être légèrement biaisées. Il propose d'ajouter un a priori sous la forme d'une loi Gamma de densité :

$$g(x) = \frac{1}{\Gamma(5)} x^4 e^{-x}$$

Q 1.2.1 Exprimer la log-probabilité a posteriori.

Q 1.2.2 Donner la valeur optimale du paramètre au sens du maximum a posteriori.

Q 1.3 En étudiant les valeurs du tableau, quelle suggestion pourriez-vous faire pour améliorer la modélisation ?

Exercice 2 (20 pts) – Des urnes et des lois

Soit deux urnes A et B , remplies de boules de couleur dans lesquelles nous avons effectué 100 tirages avec remise pour le résultat suivant :

Couleur	R	B	V	J
Nb boules	35	10	32	23

Urne A

Couleur	R	B	V	J	N
Nb boules	28	9	25	33	5

Urne B

Q 3.1 Quels sont les états de la chaîne de Markov ?

Q 3.2 Donner la matrice de transition de la chaîne et le graphe associé, ainsi que le vecteur de probabilité initial.

Q 3.3 La chaîne de Markov est-elle irréductible ? Apériodique ?

Q 3.4 La chaîne de Markov converge-t-elle vers une distribution stationnaire unique ? Justifier.

Q 2.1 Supposons que l'on tire successivement (toujours avec remise) une boule Jaune et une Verte dans une même urne ;

Q 2.1.1 Sachant ces données, pouvez-vous décider si l'urne de ces 2 tirages est l'urne A ou l'urne B ?

Q 2.1.2 Un expert nous indique qu'il y a 54% de chance que le tirage vienne de A . Sachant cette donnée supplémentaire, votre décision sur l'urne de tirage change-t-elle ?

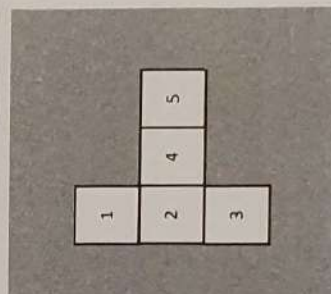
Q 2.2 Le tirage d'une boule Rouge dans l'urne A s'apparente à un processus de Bernoulli. Alors que l'expérience relatée ci-dessus fait apparaître un probabilité de 35%, l'expert nous indique que la probabilité de victoire historique est de $p = 32\%$. Peut-on affirmer, avec une confiance de 95%, que le taux de boules Rouges dans l'urne A a augmenté dans la période de notre tirage ? Soyez précis sur les hypothèses à faire pour pouvoir mener les calculs.

Q 2.3 Nous sommes passés dans la question précédente par un célèbre théorème pour vérifier une hypothèse. Mais nous aurions en fait pu directement calculer la probabilité de tirer n boules rouges en 100 tirages sous l'hypothèse $p = 32\%$. Poser les calculs (sans faire l'application numérique) et expliquer comment tirer une conclusion.

Q 2.4 Peut-on affirmer que les deux urnes sont en fait probablement identiques ? Toujours avec une confiance de 95%.

Il existe des tests pour comparer deux échantillons. Toutefois, pour des raisons numériques, nous proposons de prendre les résultats de l'urne B comme la distribution théorique et nous testerons si l'expérience de l'urne A peut être considérée comme un échantillon issu de cette distribution théorique.

Exercice 3 (15pts) – Modélisation markovienne



Le labyrinthe ci-contre est constitué de 5 pièces. Les cases grisées représentent des murs, les cases blanches des pièces. L'explorateur démarrera toujours sur la case 1. Il peut passer d'une pièce à l'autre seulement si elle sont adjacentes (les mouvements en diagonale ne sont pas possibles), et les mouvements sont réalisés avec des probabilités égales. Par exemple, de la salle 1, l'explorateur peut rester sur place ou aller en salle 2, avec des probabilités égales.

Q 3.5 Donner la définition formelle de la distribution stationnaire et interpréter la solution du problème.

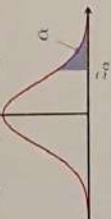
Q 3.6 Rappelez comment trouver la distribution stationnaire en utilisant l'analyse spectrale de la matrice de transition A .

Donner le code de la fonction `d = distribution_stationnaire(A)`, qui calcule cette distribution stationnaire. **N.B.** : Pour la diagonalisation matricielle, utiliser la fonction `eig` du module `np.linalg`.

Q 3.7 Un trésor se trouve dans la salle 5, on se demande en combien de temps, en moyenne, l'explorateur va le trouver. En faisant l'hypothèse que vous disposez déjà de la fonction `tirage_multinomiale(p)` qui retourne l'indice de l'état sélectionné à partir d'une distribution p , donner le code de la méthode `trouver_tresor(A)` qui calcule le nombre d'itérations nécessaire pour UNE quête. Donner la ligne de code qui estime le temps moyen nécessaire sur 1000 quêtes.

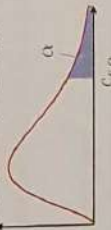
Extraits des tables de la loi normale et du χ^2

Tableau des z_α tels que $P(Z \geq z_\alpha) = \alpha$ avec $Z \sim \mathcal{N}(0, 1)$



z_α	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2297	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0722	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0466	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0352	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233

Tableau des $c_{r,\alpha}$ tels que $P(Z \geq c_{r,\alpha}) = \alpha$ avec $Z \sim \chi^2_r$



$r \backslash \alpha$	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	0.000393	0.00157	0.000982	0.00393	0.0158	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.101	0.211	4.61	5.99	7.38	9.21	10.6
3	0.0717	0.115	0.216	0.352	0.584	6.25	7.81	9.35	11.3	12.8
4	0.207	0.297	0.484	0.711	1.06	7.78	9.49	11.1	13.3	14.9
5	0.412	0.554	0.831	1.15	1.61	9.24	11.1	12.8	15.1	16.7
6	0.676	0.872	1.24	1.64	2.20	10.6	12.6	14.4	16.8	18.5
7	0.989	1.24	1.69	2.17	2.83	12.0	14.1	16.0	18.5	20.3
8	1.34	1.65	2.18	2.73	3.49	13.4	15.5	17.5	20.1	22.0
9	1.73	2.09	2.70	3.33	4.17	14.7	16.9	19.0	21.7	23.6
10	2.16	2.56	3.25	3.94	4.87	16.0	18.3	20.5	23.2	25.2
11	2.60	3.05	3.82	4.57	5.58	17.3	19.7	21.9	24.7	26.8
12	3.07	3.57	4.40	5.23	6.30	18.5	21.0	23.3	26.2	28.3
13	3.57	4.11	5.01	5.89	7.04	19.8	22.4	24.7	27.7	29.8
14	4.07	4.66	5.63	6.57	7.79	21.1	23.7	26.1	29.1	31.3
15	4.60	5.23	6.26	7.26	8.55	22.3	25.0	27.5	30.6	32.8