MAPSI – Examen Réparti – 60 pts

Durée : 2h

Seuls documents autorisés : Calculatrice, antisèche recto-verso,
tables de lois de probabilité
Le barème sur 60 pts (1 pt = 2 minutes)
– Barème indicatif –

Exercice 1 (6pts) – **Probabilités conditionnelles**

Soit le tableau suivant, partiellement rempli, correspondant à la distribution conditionnelle $p(B|A)$.

		A	
		1	2
	1	0.1	$?$
B	2	0.2	$?$
	3	$?$	$?$

Q 1.1 (1.5pt) Compléter le tableau ci-dessus sachant que $P(B = k|A = 2) \propto k$, c'est à dire que $P(B = k|A = 2) = pk$ où p est une constante à déterminer.

Q 1.2 (1.5pt) Déterminer le tableau de la loi jointe $P(A, B)$ sachant que la marginale en A est uniforme. Déterminer la marginale en B.

Q 1.3 (1.5pt) Les deux variables A et B sont-elles indépendantes ?

Q 1.4 (1.5pt) Soit le tableau `pAB` obtenu à la question 1.2 (quelles que soient les valeurs à l'intérieur). Donner les lignes de code permettant de vérifier (ou pas) l'indépendance entre les variables.

Exercice 2 – Attente au distributeur de billets (6.5pts)

En supposant que le traitement d'une demande d'un client au distributeur est réalisé en temps constant (égal à 1 unité de temps) et que la distribution des arrivées des clients au distributeur suit une loi de Poisson de paramètre μ (unités de temps), la distribution de probabilité des durées X pendant lesquelles le distributeur est occupé suit une distribution de Borel de paramètre μ définie par :

$$P(X = n \text{ unités de temps}) = \frac{e^{-\mu n}(\mu n)^{n-1}}{n!}$$

où le paramètre μ appartient à l'intervalle $[0, 1]$. Sur une période de quelques jours, on a observé les durées d'occupation suivantes :

2	3	1	5	2	7

Q 2.1 (2.5pts) Déterminez par maximum de vraisemblance la valeur du paramètre μ.

Q 2.2 (4pts) Un expert de la banque affirme qu'*a priori* la distribution des paramètres μ dans la région où est situé le distributeur suit une loi Beta de paramètres $\alpha = 7$ et $\beta = 21$. On rappelle que la loi Beta est définie de la manière suivante :

$$\text{Beta}(x; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1 - x)^{\beta-1}.$$

Où Γ désigne la loi usuelle dont la formulation est volontairement omise. En utilisant cet apriori, calculez par *maximum a posteriori* la valeur du paramètre μ.

Exercice 3 (11 pts) – **Pour quelques patates de plus**

Nous nous intéressons ici à l'optimisation de la culture de pommes de terres. Pour commencer à travailler, nous avons fait un prélèvement de 100 éléments qui se répartissent comme suit en 4 catégories :

Diamètre moyen (en cm)	3	5	8	12
Nombres de pommes de terres	30	20	30	20

Q 3.1 (2.5pts) L'agriculteur considère que toutes les classes sont grossièrement équi-probable... Qu'en pensez-vous du point de vue statistique –avec un niveau de confiance 0.95– ?

Q 3.2 (2.5pts) L'année précédente, le diamètre moyen des pommes de terre était de 6.4cm sur l'ensemble de la récolte. Depuis l'agriculteur a tenté d'optimiser la production par différents moyens. Peut-on conclure que le diamètre moyen des pommes de terre a augmenté (avec une confiance de 95%, en prenant un écart type sur le diamètre de 3 cm) ?

Q 3.3 L'agriculteur nous fait remarquer qu'il peut y avoir un biais sur le diamètre D selon les champs C d'où viennent les pommes de terre. L'échantillon considéré vient de 3 champs, chaque diamètre de patate correspondant la distribution suivante :

Diamètre moyen (en cm)	3	5	8	12
Champ 1	0.80	0.30	0.20	0.00
Champ 2	0.10	0.30	0.20	0.50
Champ 3	0.10	0.40	0.60	0.50

Q 3.3.1 (1pt) A quoi correspond la distribution du tableau ci-dessus par rapport aux variables aléatoires Champ C et Diamètre D ?

Q 3.3.2 (2.5pts) Est-il possible de calculer le diamètre moyen des patates du champ 1 à partir des informations présentes dans les questions précédentes ? Dans l'affirmative, faites le calcul, sinon, justifier de l'impossibilité de le faire.

Q 3.4 (2.5pts) Chaque pomme de terre engendre des coûts fixe de 5cts (plantation, arrosage, récolte). Elles sont ensuite vendues en moyenne selon leur calibre :

Diamètre moyen (en cm)	3	5	8	12
Prix unitaire	3cts	6cts	10cts	15cts

Est ce que l'un des champs est déficitaire ? Donner les étapes principales analytiquement puis l'application numérique.

Exercice 4 (4pts) – **[CODE] Multiplication matricielle**

Q 4.1 (2.5pts) Donner le code de la fonction `multmat(A,B)` qui prend en argument deux matrices et retourne la matrice résultant du produit matriciel.
Vous travaillerez obligatoirement avec des boucles `for`, après avoir extrait les dimensions des matrices d'entrées mais vous considérerez que les dimensions des matrices sont de dimensions compatibles sans faire de test.

Q 4.2 (1.5pt) Vous donnerez les lignes de script permettant l'invocation de la méthode et la vérification du résultat par rapport à l'opérateur `@`. On supposera A et B pré-existantes.

Exercice 5 (14pts) – **Max de vraisemblance**

Un fabricant d'ampoules *basse consommation* affirme que la durée de vie de ses ampoules est de 2 ans et demi. Un échantillon de 1000 ampoules a été testé et les durées de vie constatées (en dizaines d'années) ont été reportées dans le tableau ci-dessous :

durée (dizaines d'années)	0.05	0.1	0.2	0.25	0.3	0.35
nombre d'ampoules	200	100	300	100	200	100

Q 5.1 (2.5pts) Sachant que la durée de vie d'une ampoule de l'entreprise est modélisée par une loi exponentielle de densité $p(x) = \lambda e^{-\lambda x}$, estimez le paramètre λ par maximum de vraisemblance en vous appuyant sur l'échantillon ci-dessous. Vous détaillerez vos calculs.

Q 5.2 Un expert de l'entreprise intervient alors pour vous expliquer qu'il existe deux gammes de produits : la gamme basique, qui représente approximativement 60% des ventes et la gamme premium, plus résistante, qui représente 40% des ventes. Les deux gammes sont modélisables par des lois exponentielles.

Q 5.2.1 (2pts) Identifier les paramètres initiaux des deux modèles. Afin d'éviter un nouveau calcul de max de vraisemblance, vous pourrez utiliser directement le fait que lorsque X suit une loi exponentielle, $E[X] = 1/\lambda$

Q 5.2.2 (1pt) Donner les formules permettant de calculer la probabilité d'appartenance à un modèle sachant la durée de vie à partir des paramètres initiaux (i.e. la probabilité des variables cachées sachant les observations).
Note : donner la formulation sans faire les applications numériques

Q 5.2.3 (3pts) Rappeler la formalisation du problème d'optimisation pour la mise à jour des paramètres des deux modèles en détaillant le calcul de la log-vraisemblance en fonction des $Q_i(j)$.
Calculer la valeur des nouveaux paramètres optimaux en fonction des $Q_i(j)$.

Q 5.2.4 (1.5pt) Donner l'algorithme général et proposer un critère d'arrêt.

Q 5.3 (4pts) Donner une implémentation en python de l'algorithme ci-dessus. Vous préciserez en une phrase vos hypothèses sur la forme des données en entrée et vous utiliserez obligatoirement au moins deux fonctions (permettant de bien identifier les arguments dont vous avez besoin pour le calcul).

Exercice 6 (10.5 pts) – **Tests divers**

Q 6.1 (1.5 pt) Rappeler l'espérance d'une variable de Bernoulli B puis démontrer que la variance vaut $V[B] = p(1-p)$.

Q 6.2 (0.5 pt) *Espérance et variance d'une loi binomiale*
En remarquant qu'une variable X suivant une loi binomiale peut s'écrire comme la somme de variables X_i indépendantes suivant toutes une même loi de Bernoulli, calculer l'espérance et la variance de X.

Q 6.3 *Compositions de groupes*
Dans une université bien connue, 800 étudiants se sont inscrits à une UE de L1. Ils sont répartis aléatoirement en 200 groupes de travail de 4 étudiants. Sur ces 200 groupes :

- 13 groupes sont composés de garçons uniquement,
- 65 groupes contiennent une unique fille (et donc 3 garçons)
- 72 groupes contiennent exactement 2 filles (et donc 2 garçons),
- 35 groupes contiennent exactement 3 filles (et donc 1 garçon),
- 15 groupes contiennent exactement 4 filles (et donc pas de garçons),

Les enseignants de l'UE se demande si cette répartition est compatible avec une hypothèse de répartitions équilibrées entre filles et garçons (autant de chance pour un membre d'un groupe d'être une fille ou un garçon).

Q 6.3.1 (0.5pt) Quel type de test doit-on donc effectuer ?

Q 6.3.2 (1pt) Exprimer H_0. Soit X la variable représentant le nombre de filles dans un groupe, quelle loi doit-elle suivre sous H_0 ?

Q 6.3.3 (1.5pt) Quels seraient alors les effectifs attendus pour les groupes comprenant de 0 à 4 filles ?
Note : $\binom{4}{i} = C_4^i = [1, 4, 6, 4, 1]$ pour i allant de 0 à 4

Q 6.3.4 (2.5pt) Vérifier que les conditions d'acceptation du test sont vérifiés puis vérifier si H_0 est acceptable pour un seuil de signification $\alpha = 0.05$.

Q 6.4 (3 pts) On se rend compte rapidement en étudiant la répartition des groupes que l'hypothèse globale d'équilibre entre les sexes n'est pas respectée dans cette UE.
Donner les calculs permettant d'estimer la répartition garçons/filles à partir de cet échantillon puis vérifier votre hypothèse H_0 modifiée par rapport à ce nouveau paramètre, toujours au même niveau de confiance. Si vous n'avez pas le temps de faire les calculs, donnez votre intuition (motivée) sur le résultat que nous allons obtenir.