MAPSI

Durée : 2 heures

Seuls documents autorisés : Calculatrices et votre antisèche recto-verso
– Barème indicatif –

Exercice 1 (3 points) – **Qu'est-ce ?**

Dans une petite supérette, on a noté le nombre de clients passant par heure à chacune des 5 caisses :

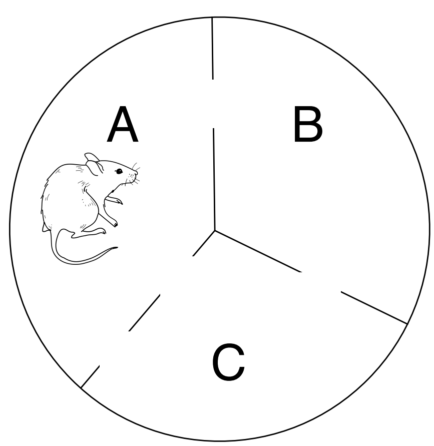
caisse	1	2	3	4	5
nombre de clients	13	18	32	15	22

Q 1.1 Le gérant de la supérette pense que la répartition entre les différentes caisses correspond à la distribution de probabilité suivante, où X représente la variable aléatoire Â« caisse choisie par le client Â».

X	1	2	3	4	5
probabilité $P(X)$	0.1	0.2	0.4	0.2	0.1

À l'aide d'un test d'ajustement de niveau de confiance 95%, peut-on dire que le gérant a raison ? Vous justifierez mathématiquement votre réponse.

Exercice 2 (5.5 points) – **Souris**



— La souris commence toujours dans la pièce A (temps = 1),
— elle ne reste jamais dans la même pièce entre deux pas de temps,
— toutes les portes sont équiprobables à chaque pas de temps.

On souhaite modéliser l'expérience suivante à l'aide d'une chaîne de Markov.

Q 2.1 (1.5 pt) Proposer une modélisation : que signifie les états ? Quels sont les paramètres de la chaîne de Markov ?

Q 2.2 (1.5 pt) Quelle est la probabilité que la souris se trouve dans chaque pièce aux temps 2, 3, 4 ?

Q 2.3 (0.5 pt) Que signifie la distribution stationnaire dans cette expérience ? Est-elle unique ?

Q 2.4 (1.5 pt) Calculer la distribution stationnaire.

Q 2.5 (0.5 pt) Souris schizophrène
Imaginons que la souris ait 2 modes comportementaux distincts, sans que l'on sache quand elle passe de l'un à l'autre. Comment modéliser les déplacements de la souris dans ce cas ? (Proposer un modèle) Est-il possible de prendre en compte la contrainte disant que la souris change de pièce à chaque pas de temps ?

Exercice 3 (4 points) – **Emails**

Un usager souhaite modéliser la distribution de probabilité du nombre d'emails qu'il reçoit chaque jour dans sa boite mail. La distribution la plus appropriée pour cela est la loi de Poisson de paramètre λ définie par $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$, où X représente la variable aléatoire Â≪ nombre d'emails reçus par jour Â≫. Pendant cinq jours, il a noté le nombre d'emails reçus, comme indiqué dans le tableau ci-dessous :

5	3	7	2	8

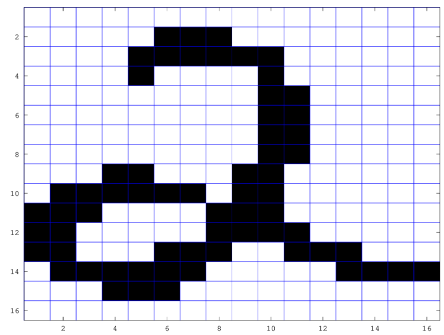
Q 3.1 Les observations sont mutuellemment indépendantes. Estimez par maximum de vraisemblance la valeur du paramètre λ. Vous préciserez bien les formules que vous utiliserez.

Q 3.2 L'usager, féru de statistiques, sait que la loi conjuguée de la loi de Poisson est la loi Gamma, de paramètres α, β, définie par :

$$\Gamma(x; \alpha, \beta) = \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)},$$

où $\Gamma(\alpha)$ est la valeur de la fonction Γ en α. L'usager souhaite maintenant utiliser cette loi comme *a priori*. Le paramètre α de la loi Gamma représente la forme de cette loi et l'usager l'estime à $\alpha = 2$. En revanche, le paramètre β, qui représente son échelle, lui est inconnu. Estimez par maximum *a posteriori* le paramètre λ de la question précédente ainsi que le paramètre β de la loi Γ.

Exercice 4 (5 points) – Reconnaissance de caractères



Afin de reconnaitre des chiffres manuscrits comme ceux vus en TME ce semestre (image binaire 16x16), on propose d'utiliser le modèle suivant :

— Chaque ligne j est décrite par une variable x_j donnant l'indice du premier pixel allumé sur la ligne j. x_j suit une loi géométrique de paramètre p_j : on part de la gauche, chaque pixel est une épreuve de Bernoulli et on attend le premier succès (pixel allumé).

— Afin de stabiliser les calculs, on ajoute une colonne (17) de pixels allumés.

— L'observation de chaque ligne est stockée dans une variable k_j donnant l'indice du premier pixel allumé. Par exemple sur l'illustration précédente : $k_1 = 17, k_2 = 6$...

— Les lignes sont supposées mutuellement indépendantes.

Notations : $\mathbf{x}$ est une image composée de N lignes : $\mathbf{x} = \{x_1, \dots, x_N\}$. On note $X = \{\mathbf{x}^1, \dots, \mathbf{x}^P\}$ un ensemble de P images, avec : $\mathbf{x}^i = \{x_1^i, \dots, x_N^i\}$.

Q 4.1 (0.5 pt) Quelles valeurs peuvent prendre les x_j ?

Q 4.2 (0.5 pt) Donner l'expression de la loi géométrique $p(x_j = k_j)$ en fonction du paramètre p_j et de k_j.
Rappel : dans la loi de probabilité géométrique, la variable x_j prend la valeur k_j si on a eu $k_j - 1$ échecs puis 1 succès (dans cet ordre).

Q 4.3 (0.5 pt) Donner l'expression de la probabilité d'observation d'une image $p(\mathbf{x})$ si la valeur observée pour le pixel x_j est k_j.

Q 4.4 (1 pt) Donner la log vraisemblance $\mathcal{L}$ de l'ensemble des images X, en supposant que toutes les images sont mutuellement indépendantes.

Q 4.5 (1.5 pt) Calculer les paramètres p_j maximisant la vraisemblance. Exprimer p_j en fonction de P et $K_j = \sum_i k_j^i$, où k_j^i représente la valeur observée du $j$ème pixel de la $i$ème image. Proposer une interprétation du résultat.

Q 4.6 (1 pt) Expliquer la procédure complète pour construire et évaluer un classifieur d'image au sens du maximum de vraisemblance basé sur ce modèle.

Q 4.7 (0.5 pt) On dispose maintenant d'un a priori $p(y)$ sur les différents chiffres (qui ne sont pas équiprobables dans la majorité des applications). Donner un critère de classification des images au sens du maximum à posteriori.

Exercice 5 (3 points) – **Indépendance**

Soit trois variables aléatoires binaires X, Y, Z. La distribution de probabilité jointe de ces trois variables est donnée dans le tableau ci-dessous :

$P(X,Y,Z) =$

	z_1		z_2	
	y_1	y_2	y_1	y_2
x_1	$0,03$	$0,07$	$0,1$	$0,1$
x_2	$0,12$	$0,28$	$0,25$	$0,05$

Q 5.1 Calculez la distribution $P(X, Y|Z)$.

Q 5.2 Est-ce que X et indépendant de Y conditionnellement à Z ? Vous justifierez mathématiquement votre réponse.