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 ≪ caisse choisie par le client ≫.

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

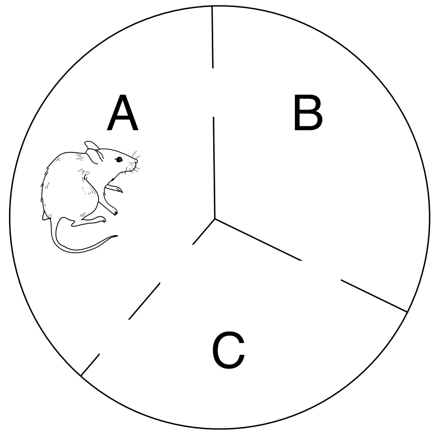
À l'aide d'un test d'ajustement de niveau de confiance 95%, peut-on dire que le gérant a raison ? Vous justifierez mathématiquement votre réponse.

Le nombre d'observations de l'échantillon est $N = 100$. La statistique d'ajustement est donc :

$$d^2 = \frac{(13-10)^2}{10} + \frac{(18-20)^2}{20} + \frac{(32-40)^2}{40} + \frac{(15-20)^2}{20} + \frac{(22-10)^2}{10} = 18,35.$$

Le nombre de degrés de liberté est égal à $5 - 1 = 4$. La loi du χ^2 à 4 degrés de liberté donne, pour le niveau de confiance 95% (risque $\alpha = 5\%$), le seuil $d^2_\alpha = 9,49$. Comme $d > d^2_\alpha$, on peut en déduire que le gérant a tort.

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 ?

– 0.5 pour les états
– 0.5 pour pi et les 0 sur la diagonale
– 0.5 pour A juste
Les états sont les pièces.
$\Pi = [1, 0, 0]$

$$A = \begin{bmatrix} 0 & 1/3 & 2/3 \\ 1/2 & 0 & 1/2 \\ 2/3 & 1/3 & 0 \end{bmatrix}$$

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

– 1 pt si pas de valeur numérique (mais formule juste)
– 1.5 pour les bonnes valeurs

```
p = np.array([1.,0,0])
A = np.array([[0, 1/3., 2/3.],[.5,0,.5], [2/3.,1/3.,0]])
p.dot(A)
Out[25]: array([ 0.        ,  0.33333333,  0.66666667])
p.dot(A).dot(A)
Out[26]: array([ 0.61111111,  0.22222222,  0.16666667])
p.dot(A).dot(A).dot(A)
Out[27]: array([ 0.22222222,  0.25925926,  0.51851852])
```

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

Je n'ai pas encore trouvé de réponse honnête... D'où le peu de point dans le barème
Au bout d'un grand nombre d'itération, la distribution va être stable (= distribution stationnaire)
Interprétation : probabilité de se trouver dans chaque état
Cette distribution est unique car la chaine est irréductible et positive

- une chaîne de Markov est irréductible si tout état est accessible à partir de n'importe quel autre état ;
- un état est récurrent positif si l'espérance du temps de premier retour en cet état, partant de cet état, est finie.

Q 2.4 (1.5 pt) Calculer la distribution stationnaire.

- 1.5 pour ceux qui ont posé et résolu le système (évidemment)
- 1 pt pour ceux qui ont posé le système (avec la normalisation à 1) mais fait une faute dans les calculs
- 0.5 pt pour ceux qui ont posé la version empirique (en multipliant A plein de fois) + 0.5 s'ils ont trouvé la solution numérique

Calcul

$$\mu A = \mu \text{ et } \sum_i \mu_i = 1$$

Ecriture sous forme de système d'équations

$$\begin{aligned} .5\mu_B + 2\mu_C/3 &= \mu_A && (1)\\ \mu_A/3 + \mu_C/3 &= \mu_B && (2)\\ 2\mu_A/3 + .5\mu_B &= \mu_C && (3) \end{aligned}$$

Puis résolution :
eq3 - eq1 $\Rightarrow \mu_A = \mu_C$
eq2 $\Rightarrow 2/3\mu_A = 2/3\mu_C = \mu_B$
Normalisation :

$$2/3\mu_A + \mu_A + \mu_A = 1 \Rightarrow \mu_A = 3/8$$

$$\mu_A = 3/8 \quad \mu_C = 3/8 \quad \mu_B = 1/4$$

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 ?

– On donne les points s'ils parlent de HMM raisonnablement
On modélise avec un HMM, les états sont les modes comportementaux de la souris. La loi d'émission donne la pièce dans laquelle se trouve la souris à chaque étape... Mais on ne peut plus modéliser la contrainte de changement de pièce.
Sinon, il faut modéliser 6 états (3 pièces x 2 états de souris) avec un modèle simple. Dans ce cas, on peut prendre en compte la contrainte.

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 ≪ nombre d'emails reçus par jour ≫. 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.

La vraisemblance est :

$$\mathcal{L} = \prod_{i=1}^{5} P(X = k_i) = \prod_{i=1}^{5} \frac{\lambda^{k_i} e^{-\lambda}}{k_i!},$$

où k_i est le nombre d'emails reçus lors de la $i$ème journée. En passant au log, on obtient :

$$\log \mathcal{L} = \left(\sum_{i=1}^{5} k_i\right) \log(\lambda) - 5\lambda - \sum_{i=1}^{5} \log(k_i!).$$

En dérivant par rapport à λ, on obtient :

$$\frac{\partial \log \mathcal{L}}{\partial \lambda} = \left(\sum_{i=1}^{5} k_i\right) \frac{1}{\lambda} - 5 = 0$$

d'où $\lambda = \frac{\sum_{i=1}^{5} k_i}{5} = 5$.

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.

$$\pi(\lambda, \beta|\mathbf{x}) = \left[\prod_{i=1}^{5} \frac{\lambda^{k_i} e^{-\lambda}}{k_i!}\right] \times \Gamma(\lambda; 2, \beta) = \left[\prod_{i=1}^{5} \frac{\lambda^{k_i} e^{-\lambda}}{k_i!}\right] \times \frac{\lambda e^{-\lambda/\beta}}{\beta^2 \Gamma(2)}.$$

En passant au log, on obtient :

$$\log(\pi(\lambda, \beta|\mathbf{x})) = \left(\sum_{i=1}^{5} k_i\right) \log(\lambda) - 5\lambda - \sum_{i=1}^{5} \log(k_i!) + \log(\lambda) - \frac{\lambda}{\beta} - 2\log(\beta) - \log(\Gamma(2)).$$

En dérivant par rapport à λ et β, on obtient :

$$\frac{\partial \log(\pi(\lambda, \beta|\mathbf{x}))}{\partial \lambda} = \left(\sum_{i=1}^{5} k_i\right) \frac{1}{\lambda} - 5 + \frac{1}{\lambda} - \frac{1}{\beta} = 0 \tag{4}$$

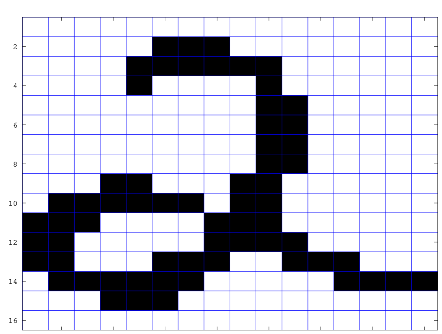
$$\frac{\partial \log(\pi(\lambda, \beta|\mathbf{x}))}{\partial \beta} = \frac{\lambda}{\beta^2} - \frac{2}{\beta} = 0 \tag{5}$$

De (5), on a $\beta = \lambda/2$. En remplaçant β par cette valeur dans l'équation (4), on obtient :

$$\left(\sum_{i=1}^{5} k_i\right) \frac{1}{\lambda} - 5 + \frac{1}{\lambda} - \frac{2}{\lambda} = \left(-1 + \sum_{i=1}^{5} k_i\right) \frac{1}{\lambda} - 5 = 0,$$

d'où $\lambda = \frac{-1+\sum_{i=1}^{5} k_i}{5} = \frac{24}{5} = 4,8$ et $\beta = \lambda/2 = 2,4$.

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 : $\mathbf{x} = \{x_1, \ldots, x_N\}$. On note $X = \{\mathbf{x}^1, \ldots, \mathbf{x}^P\}$ un ensemble de P images, avec : $\mathbf{x}^i = \{x_1^i, \ldots, x_N^i\}$.

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

> 17 valeurs : 1 à 17

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).

> Rappelons que pour la loi géométrique : $p(x_j = k_j) = p_j(1 - p_j)^{k_j - 1}$

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.

> $p(\mathbf{x}) = \prod_{j=1}^N p(x_j = k_j) = \prod_{j=1}^N p_j(1 - p_j)^{k_j - 1}$

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.

> – On ne pénalise pas ceux qui ont fait faux à la question précédente mais qui ont juste à celle-ci
>
> $$logL = \sum_{ij} \log p(x_j^i = k_j^i)$$
>
> $$logL = \sum_{ij} \log(p_j) + (k_j^i - 1)\log(1 - p_j)$$

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.

> $$\frac{\partial \mathcal{L}}{\partial p_j} = \sum_{i=1}^P \frac{1}{p_j} + (k_j^i - 1)\frac{-1}{1 - p_j}$$
>
> $$\frac{\partial \mathcal{L}}{\partial p_j} = 0 \Longleftrightarrow \frac{P}{p_j} = \frac{K_j - P}{1 - p_j} \Longleftrightarrow p_j = P/K_j$$
>
> Le paramètre p_j correspond à $1/n_j$ où n_j est le nombre moyen de pixels à parcourir sur la ligne j pour trouver un pixel allumé.

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.

1. Séparer les données en un ensemble d'apprentissage et un de test
2. Séparer les images d'apprentissage de chaque classe (0, 1,...)
3. Apprendre les N paramètres p_j pour chaque classe
4. Evaluation : prendre l'ensemble de test
5. Calculer les probabilités d'appartenance à chaque classe
6. La plus grande proba donne la classe, on évalue ensuite le pourcentage de bonne classification et/ou la matrice de confusion

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.

On attribut à chaque échantillon la classe

$$\arg\max_y p(y|\mathbf{x}) = \arg\max_y p(\mathbf{x}|y)p(y)/p(\mathbf{x}) = \arg\max_y p(\mathbf{x}|y)p(y)$$

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 $P(X, Y|Z)$.

On commence par calculer $P(Z) = [0.5, 0.5]$ en sommant les nombres sur les quatre cases de gauche et les quatre cases de droite. Ensuite, on divise $P(X, Y, Z)$ par $P(Z)$. On obtient donc :

$P(X, Y, Z) =$

	z_1		z_2	
	y_1	y_2	y_1	y_2
x_1	$0,06$	$0,14$	$0,2$	$0,2$
x_2	$0,24$	$0,56$	$0,5$	$0,1$

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

> Conditionnellement à $Z = z_1$, il y a bien indépendance car la 2ème ligne est proportionnelle à la premiere (ou, de manière équivalente, la 2ème colonne est proportionnelle à la première). En revanche, ce n'est pas le cas pour $Z = z_2$. Donc, X n'est pas indépendant de Y conditionnellement à Z.