Premier examen réparti du module RFIDEC

C. Gonzales

Durée : 2 heures

Seuls documents autorisés :
Les transparents de cours. Calculatrices autorisées.

Exercice 1 (7 pts)

Un péage d'autoroute est constitué de 10 cabines de péage, numérotées de 1 à 10. On a recensé le nombre de clients se présentant à chaque cabine sur une heure donnée de la journée. Voici les effectifs obtenus :

N° cabine	1	2	3	4	5	6	7	8	9	10
Nb clients	24	14	18	20	23	13	23	24	23	18

Q 1.1 À l'aide d'un test d'ajustement, testez si la distribution des clients dans les différentes cabines du péage est uniforme. Utilisez pour cela un niveau de test $\alpha = 0,05$ (niveau de confiance du test = 95%).

Bien évidemment notre test d'hypothèse sera : $H_0 =$ « la répartition des clients est uniforme » v.s. $H_1 =$ « la répartition n'est pas uniforme ». Si l'on avait eu une loi uniforme, chaque chiffre aurait du apparaître exactement 20 fois. Si l'on appelle X_i la variable « effectif » recensé pour la $i$ème cabine, la statistique d'ajustement est donc :

$$D^2 = \sum_{i=1}^{10} \frac{(X_i - 20)^2}{20}.$$

On sait que cette statistique suit une loi χ^2_9, c'est-à-dire une loi du χ^2 à 9 degrés de liberté.

$$\begin{aligned} \alpha = 0,05 &= P(\text{rejeter } H_0 | H_0 \text{ est vraie}) \\ &= P\left(D^2 > d_\alpha \;\middle|\; D^2 \sim \chi^2_9\right). \end{aligned}$$

D'après la table du χ^2, $d_\alpha = 16,9$. Autrement dit, si la valeur d^2 de D^2 observée sur l'échantillon est supérieure à 16,9, alors on conclura que la répartition des clients n'est pas uniforme, sinon on la considérera comme uniforme. Calculons maintenant la valeur de D^2 observée :

$$\begin{aligned} d^2 = \tfrac{1}{20}[&(14-20)^2 + (24-20)^2 + (18-20)^2 + (20-20)^2 + (23-20)^2 + (13-20)^2 + \\ &(23-20)^2 + (18-20)^2 + (24-20)^2 + (23-20)^2] = 7,6. \end{aligned}$$

Donc on peut effectivement estimer que la répartition est uniforme.

Exercice 2 (6 pts)

Soit trois variables aléatoires X, Y, Z, de modalités respectives $\{x_1, x_2\}$, $\{y_1, y_2, y_3\}$ et $\{z_1, z_2\}$. La probabilité jointe de ces trois variables est la suivante :

	x_1			x_2		
	y_1	y_2	y_3	y_1	y_2	y_3
z_1	$0,02$	$0,16$	$0,03$	$0,1$	$0,12$	$0,02$
z_2	$0,18$	$0,04$	$0,07$	$0,1$	$0,08$	$0,08$

Q 2.1 Déterminez si X est indépendante de Y, si X est indépendante de Z et si Y est indépendante de Z. Vous justifierez bien évidemment vos réponses.

D'après la table ci-dessus, on a :

$P(X,Y) =$

	y_1	y_2	y_3
x_1	$0,2$	$0,2$	$0,1$
x_2	$0,2$	$0,2$	$0,1$

$P(Y,Z) =$

	y_1	y_2	y_3
z_1	$0,12$	$0,28$	$0,05$
z_2	$0,28$	$0,12$	$0,15$

$P(X,Z) =$

	x_1	x_2
z_1	$0,21$	$0,24$
z_2	$0,29$	$0,26$

On voit que X est indépendant de Y car les 2 lignes du tableau de gauche sont identiques. En revanche, X n'est pas indépendant de Z car la deuxième ligne du tableau de droite n'est pas égale à la première. De même, Y n'est pas indépendant de Z.

Q 2.2 En déduire un réseau bayésien représentant les indépendances que vous avez déterminées.

Le réseau bayésien est $X \longrightarrow Z \longleftarrow Y$.

Exercice 3 (7 pts)

Dans le cadre des « Plans de prévention des risques d'inondations (PPR-I) », des photos du satellite SPOT5 sont utilisées afin de caractériser des zones susceptibles d'être innondées. Dans le passé, pour une région donnée, il a été déterminé que 60% des parcelles étaient innondables (PI), 10% étaient partiellement innondables (PPI), les autres étant non innondables (NI).
D'après les caractéristiques techniques de SPOT5, les probabilités d'obtenir un certain niveau de gris n sur une image conditionnellement au fait que la zone représentée est de type PI ou PPI sont respectivement $P(n|PI)$ et $P(n|PPI)$. Des études ont montré que $P(n|PI) = \mathcal{N}(\mu_1, \sigma_1^2)$ et $P(n|PPI) = \mathcal{N}(\mu_2, \sigma_2^2)$ sont des lois normales de caractéristiques respectives :

$$\mu_1 = 100 \qquad \sigma_1 = 20 \qquad \mu_2 = 85 \qquad \sigma_2 = 5.$$

Q 3.1 Calculez les valeurs des fonctions de densité $p(n|PI)$ et $p(n|PPI)$ pour $n = 80$.

On sait que la fonction de densité d'une loi normale $\mathcal{N}(\mu, \sigma^2)$ est :

$$f(x) = \frac{1}{\sqrt{2\pi}.\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$$

Par conséquent :

$$f_{PI}(80) = \frac{1}{\sqrt{2\pi} \times 20} \exp\left\{-\frac{1}{2}\left(\frac{80-100}{20}\right)^2\right\} = \frac{1}{20\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\right\} \approx 0,0121$$

$$f_{PPI}(80) = \frac{1}{\sqrt{2\pi} \times 5} \exp\left\{-\frac{1}{2}\left(\frac{80-85}{5}\right)^2\right\} = \frac{1}{5\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\right\} \approx 0,0484$$

Q 3.2 Une nouvelle image envoyée par SPOT5 révèle pour la région mentionnée ci-dessus une zone Z dont le niveau de gris est $n = 80$. Calculez la vraisemblance d'obtenir la zone Z sous l'hypothèse $\theta_1 =$ « Z est de type PI », puis sous l'hypothèse $\theta_2 =$ « Z est de type PPI ». Si l'on prend une décision fondée sur le critère du maximum de vraisemblance, quelle classe $\theta_1 =$ PI ou $\theta_2 =$ PPI serait la plus probable pour la zone Z ?

Ici, on a un échantillon $\mathbf{x}$ qui est constitué de la nouvelle image. Par conséquent, la vraisemblance $L(\mathbf{x}, \theta_1) = p(80|PI) = 0,0121$. De même, $L(\mathbf{x}, \theta_2) = p(80|PPI) = 0,0484$. D'après le maximum de vraisemblance, il est plus probable que la zone Z soit de type PPI.

Q 3.3 Si l'on prend une décision fondée selon le critère du maximum *a posteriori* (MAP), à quelle classe $\theta_1 = \text{PI}$ ou $\theta_2 = \text{PPI}$ doit-on attribuer la zone ?

Avec le maximum de vraisemblance, on doit calculer $\pi(\theta_1|\mathbf{x})$ et $\pi(\theta_1|\mathbf{x})$ définis respectivement par :

$$\pi(\theta_1|\mathbf{x}) = \frac{L(\mathbf{x}, \theta_1)\pi(\theta_1)}{\sum_{\theta \in \Theta} L(\mathbf{x}, \theta)\pi(\theta)} \qquad \pi(\theta_2|\mathbf{x}) = \frac{L(\mathbf{x}, \theta_2)\pi(\theta_2)}{\sum_{\theta \in \Theta} L(\mathbf{x}, \theta)\pi(\theta)}.$$

Par conséquent, $\pi(\theta_1|\mathbf{x}) \propto 0,0121 \times 0,6 = 0,00726$ et $\pi(\theta_2|\mathbf{x}) \propto 0,0484 \times 0,1 = 0,00484$. Donc la zone devrait, selon MAP, être une parcelle inondable (PI).

Q 3.4 En fait, on n'est pas très sûr de la probabilité *a priori* des zones partiellement innondables (PPI) qui, selon les années, varie de 8% à 20% de la région. Est-ce que cela peut changer notre décision selon le critère MAP ?

Si l'on fait varier la probabilité a priori de PPI, cela revient, pour notre calcul de MAP à changer $\pi(\theta_2|\mathbf{x}) \propto 0,0484 \times 0,1$ par $\pi(\theta_2|\mathbf{x}) \propto 0,0484 \times \alpha$ pour $\alpha \in [0,08; 0,2]$. En $0,08$, $\pi(\theta_2|\mathbf{x}) \propto 0,00387$, et donc notre décision ne change pas par rapport à la question précédente. En revanche, lorsque $\alpha = 0,2$, on a $\pi(\theta_2|\mathbf{x}) \propto 0,00968$ et, dans ce cas, la zone Z doit être considérée comme PPI.