Examen de 2ème session du module RFIDEC

Christophe Gonzales

Durée : 2 heures

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

Exercice A (4 points)

On souhaite comparer l'efficacité de deux médicaments censés combattre la même maladie. Le premier médicament est générique et son prix est réduit, le deuxième est un médicament de marque de prix beaucoup plus élevé. La Sécurité Sociale a effectué une enquête sur les guérisons obtenues grâce à chacun de ces médicaments. Le nombre de guérisons et de non guérisons (sur les 250 personnes testées) sont consignés dans le tableau ci-dessous :

	générique	marque
guérisons	44	156
non guérisons	6	44

À un niveau de risque de 5%, peut-on estimer que le taux de guérison dépend du médicament (générique ou marque) ? Justifiez votre réponse mathématiquement.

Le problème revient à déterminer s'il y a indépendance entre les guérisons et le type de médicament. On commence par calculer les effectifs marginaux :

	générique	marque	total
guérisons	44	156	200
non guérisons	6	44	50
total	50	200	250

S'il y avait réellement indépendance, on obtiendrait donc le tableau suivant :

	générique	marque	total
guérisons	40	160	200
non guérisons	10	40	50
total	50	200	250

On peut maintenant calculer la statistique du test d'indépendance :
$\chi_{(2-1)\times(2-1)} = \frac{(44-40)^2}{40} + \frac{(156-160)^2}{160} + \frac{(6-10)^2}{10} + \frac{(44-40)^2}{40} = 2.5$
À un niveau de risque de 5%, la valeur critique du χ_1 est égale à 3.84. Par conséquent, on peut en déduire qu'il y a bien indépendance entre le taux de guérison et le type de médicament.

Exercice B (6 points)

Soit X une variable aléatoire définie sur l'ensemble des nombres entiers positifs. On sait que X suit la loi géométrique de paramètre $p \in [0, 1]$ si $P(X = n) = (1-p)^{n-1}p$. On a observé 7 réalisations (obtenues indépendamment les unes des autres) d'une variable X suivant la loi géométrique :

2	3	8	3	4	7	8

.

Q B.1 Estimez par maximum de vraisemblance la valeur du paramètre $\theta = p$ de la loi.

La vraisemblance est égale à $L(\mathbf{x}, \theta) = \prod_{i=1}^{7} P(x_i|\theta)$, où θ est notre estimation de p. En remplaçant les $P(x_i|\theta)$ par leur valeurs, on obtient que $L(\mathbf{x}, \theta) = (1-\theta)^{28}\theta^7$. Par conséquent, la log-vraisemblance est égale à $\ln L(\mathbf{x}, \theta) = 28\ln(1-\theta) + 7\ln\theta$. Donc $\frac{\partial \ln L(\mathbf{x},\theta)}{\partial\theta} = \frac{-28}{1-\theta} + \frac{7}{\theta} = \frac{7-35\theta}{p(1-\theta)}$. La dérivée de la log-vraisemblance est donc positive jusqu'à $\theta = 0,2$ et négative ensuite. Le maximum de vraisemblance est donc atteint pour $p = 0,2$.

Q B.2 Avant le tirage de l'échantillon, nous avions une connaissance a priori sur le paramètre θ : ce dernier suivait a priori une loi Beta de paramètres $a = \alpha$ et $b = 2$. Sachant que l'estimation du paramètre $\theta = p$ par maximum a posteriori est égale à $0,16$, quelle était la valeur de α ?

On doit calculer $\text{Argmax}_\theta \pi(\theta|\mathbf{x}) = \text{Argmax}_\theta \pi(\mathbf{x}|\theta)\pi(\theta) = \text{Argmax}_\theta (1-\theta)^{27+\alpha}\theta^8$. En passant au log, on obtient $\text{Argmax}_\theta (27+\alpha)\ln(1-\theta) + 8\ln\theta$. La dérivée par rapport à θ s'annule lorsque $(35+\alpha)\theta = 8$. Autrement dit, pour que l'estimation par maximum a posteriori de p soit égale à $0,16$, il faut que $(35+\alpha)\times 0.16 = 8$, soit $\alpha = 15$.