Premier examen réparti du module RFIDEC

C. Gonzales

Durée : 2 heures

Seuls documents autorisés :
Une page format A4. Calculatrices autorisées.

Exercice 1 (5 pts)

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 1.1 À partir de cet échantillon, faites une estimation ponctuelle de la moyenne et de la variance de la population « ensemble des ampoules du fabricant ». Vous détaillerez les calculs permettant d'obtenir ces estimations.

La moyenne μ est estimée par la moyenne de l'échantillon. Donc $\mu = 2$ années (ou 0.2 dizaines d'années). La variance estimée est la variance corrigée de l'échantillon. La variance de l'échantillon est $S^2 = 1$. Donc la variance corrigée est égale à $\sigma^2 = 1 * 1000/999 \approx 1.001$.

Q 1.2 On sait que la durée de vie des ampoules de l'entreprise peut être modélisée par une loi exponentielle de densité $p(x) = \lambda e^{-\lambda x}$. Une telle loi a pour espérance $1/\lambda$ et pour variance $1/\lambda^2$. Le fabricant se demande s'il ne trompe pas ses clients en affirmant que la durée de vie moyenne des ampoules est de 2 ans et demi. À l'aide d'un test d'hypothèse de niveau de confiance 95%, peut-on déduire qu'il abuse ses clients ? Vous justifierez votre réponse.

Soit X la variable aléatoire « durée de vie d'une ampoule » et soit $\overline{X}$ la durée de vie moyenne d'un échantillon. Le test d'hypothèse est ici :

$$H_0 : \mu = 2.5 \qquad \text{v.s. } H_1 : \mu < 2.5$$

Comme, ici, l'échantillon est de grande taille (1000 individus), le théorème central limite nous assure que $\overline{X}$ suit approximativement une loi normale $\mathcal{N}(\mu, \sigma^2)$. Ici, σ^2 est la variance de la population et peut donc être supposée égale à 10.01 d'après la question 1.

$$P(\overline{X} < c|\mu = 2.5) = P\left(\frac{\overline{X} - 2.5}{\sqrt{1.001/1000}} < \frac{c - 2.5}{\sqrt{1.001/1000}}|\mu = 2.5\right) = 0.05$$

implique que $\frac{c-2.5}{\sqrt{1.001/1000}} = -1,645$, d'où $c \approx 2.48355$. Comme la moyenne de l'échantillon est de 2, on peut rejeter H_0 et en déduire que le fabricant abuse ses clients.

Exercice 2 (5 pts)

On considère ici le même fabricant d'ampoules que celui de l'exercice 1. 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.

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

La vraisemblance $L(\mathbf{x}, \lambda)$ est égale à $\prod_{i=1}^{1000} p(x_i|\lambda)$, où x_i correspond aux réalisations du tableau ci-dessus. Autrement dit, $L(\mathbf{x}, \lambda) = \prod_{i=1}^{1000} \lambda e^{-\lambda x_i} = \lambda^{1000} e^{-\lambda \sum_{i=1}^{1000} x_i} =$

$\lambda^{1000}e^{-200\lambda}$. Par conséquent, $\frac{\partial L}{\partial \lambda} = 1000\lambda^{999}e^{-200\lambda} - 200\lambda^{1000}e^{-200\lambda}$ et, donc, $\frac{\partial L}{\partial \lambda} = 0 \Longleftrightarrow 1000 - 200\lambda = 0$. Par conséquent, $\lambda = 5$.

Exercice 3 (5 pts)

Soit deux urnes A et B emplies de boules rouges et jaunes. L'urne A contient 20 boules rouges et 30 boules jaunes ; l'urne B contient 10 boules rouges et 10 boules jaunes. On sélectionne au hasard une des deux urnes (tirage équiprobable) et, dans cette urne, on tire une boule au hasard. La boule est rouge. Déterminez la probabilité que la boule provienne de l'urne A.

Soit U_1 l'événement « la boule tirée provient de l'urne A », U_2 l'événement « la boule tirée provient de l'urne B » et R l'événement « la boule tirée est rouge ». On souhaite donc calculer $P(U_1|R)$. D'après la formule de Bayes, on a :

$$P(U_1|R) = \frac{P(R|U_1)P(U_1)}{P(R)} = \frac{P(R|U_1)P(U_1)}{P(R|U_1)P(U_1) + P(R|U_2)P(U_2)}.$$

Or $P(U_1) = P(U_2) = 1/2$; $P(R|U_1) = 20/50$; $P(R|U_2) = 10/20$. Donc :

$$P(U_1|R) = \frac{\frac{20}{50} \times \frac{1}{2}}{\frac{1}{2} \times \frac{20}{50} + \frac{1}{2} \times \frac{10}{20}} = \frac{4}{9}.$$

Exercice 4 (5 pts)

Soit trois variables aléatoires X, Y, Z, de modalités respectives $\{x_1, x_2\}$, $\{y_1, y_2\}$ et $\{z_1, z_2\}$. La probabilité jointe de ces trois variables est fournie dans le tableau suivant :

	x_1		x_2	
	y_1	y_2	y_1	y_2
z_1	$0,032$	$0,06$	$0,056$	$0,576$
z_2	$0,048$	$0,06$	$0,024$	$0,144$

Q 4.1 Déterminez si X est indépendante de Y. Vous justifierez bien évidemment votre réponse.

D'après la table ci-dessus, on a :

$P(X, Y) =$

	x_1	x_2
y_1	$0,08$	$0,08$
y_2	$0,12$	$0,72$

Étant donné que la deuxième ligne n'est pas un multiple de la première, X ne peut être indépendant de Y.

Q 4.2 Déterminez si Z est indépendante de Y conditionnellement à X. Vous justifierez votre réponse.

Pour savoir si Z et indépendante de Y conditionnellement à X, il faut vérifier si $P(Z, Y|X) = P(Z|X)P(Y|X)$. Or $P(Z, Y|X) = P(X, Y, Z)/P(X)$. On commence donc par calculer $P(X)$:

$$P(Z) = \sum_{X,Y} P(X, Y, Z) = \begin{array}{|c|}\hline 0,2 \\ \hline 0,8 \\ \hline\end{array}$$

Par conséquent :

$$P(Z, Y|X) =$$

	x_1		x_2	
	y_1	y_2	y_1	y_2
z_1	0, 16	0, 3	0, 07	0, 72
z_2	0, 24	0, 3	0, 03	0, 18

Si Z est indépendant de Y conditionnellement à X, alors les deux colonnes de gauche (resp. de droite) doivent être proportionnelles l'une par rapport à l'autre. Ce n'est pas le cas ici. Donc Y n'est pas indépendant de Z conditionnellement à X.

Q 4.3 En déduire un réseau bayésien représentant les dépendances et indépendances que vous avez déterminées.

Le réseau bayésien est n'importe quel DAG complet.