Premier examen réparti du module RFIDEC

C. Gonzales

Durée : 2 heures

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

Exercice 1 (7 pts)

La loi exponentielle est une loi continue dont la fonction de densité est : $f(x) = \lambda e^{-\lambda x}$ pour tout $x > 0$. Elle sert, entre autres, pour caractériser la durée de vie des composants électroniques. Le tableau suivant recense les durées de vie (en années) observées pour un échantillon de 10 composants électroniques :

2	7	3	4	1	2	6	5	1	9

Q 1.1 On suppose que la distribution des durées de vie est effectivement une loi exponentielle. Estimez par maximum de vraisemblance la valeur de λ.

La vraisemblance est égale à $L(\mathbf{x}, \theta) = \pi(\mathbf{x}|\theta) = \prod_{i=1}^{10} p(x_i|\theta) = \prod_{i=1}^{10} \left[\theta e^{-\theta x_i}\right]$. Par conséquent, la log-vraisemblance est égale à $\ln L(\mathbf{x}, \theta) = \sum_{i=1}^{10} \ln\left(\theta e^{-\theta x_i}\right) = 10 \ln \theta - \theta \sum_{i=1}^{10} x_i$. Donc :

$$\frac{\partial \ln L(\mathbf{x}, \theta)}{\partial \theta} = \frac{10}{\theta} - \sum_{i=1}^{10} x_i = 0 \Longrightarrow \theta_{MV} = \frac{10}{\sum_{i=1}^{10} x_i}.$$

D'après le tableau ci-dessus, $\sum_{i=1}^{10} x_i = 40$. Donc $\lambda_{MV} = \frac{1}{4}$.

Q 1.2 Après discussion avec un expert en électronique, on a un *a priori* sous la forme d'une loi Gamma de densité $g(x) = \frac{1}{\Gamma(5)} x^4 e^{-x}$. Estimez par maximum a posteriori la valeur de λ.

Le maximum a posteriori est défini par :

$$\theta_{MAP} = \text{Argmax}_\theta \pi(\mathbf{x}|\theta)\pi(\theta) = \text{Argmax}_\theta \left(\prod_{i=1}^{10} \theta e^{-\theta x_i}\right) \frac{\theta^4 e^{-\theta}}{\Gamma(5)}.$$

En passant au log, on doit donc optimiser $\sum_{i=1}^{10} [\ln \theta - \theta x_i] + 4 \ln \theta - \theta - \ln \Gamma(5)$. En dérivant par rapport à θ, et en fixant cette dérivée à 0, on obtient :

$$\frac{10}{\theta} - \left(\sum_{i=1}^{10} x_i\right) + \frac{4}{\theta} - 1 = 0.$$

Autrement dit, on doit résoudre $14/\theta - (\sum_{i=1}^{10} x_i + 1) = 0$. Donc $\theta_{MAP} = 14/(\sum_{i=1}^{10} x_i + 1) = 14/41 \approx 0,34$.

Exercice 2 (6 pts)

Soit trois variables aléatoires X, Y, Z, de modalités respectives $\{x_1, x_2, x_3\}$, $\{y_1, y_2\}$ et $\{z_1, z_2\}$. La probabilité jointe de ces trois variables est fournie dans le tableau suivant :

	z_1			z_2		
	x_1	x_2	x_3	x_1	x_2	x_3
y_1	$0,012$	$0,048$	$0,06$	$0,126$	$0,042$	$0,252$
y_2	$0,018$	$0,072$	$0,09$	$0,084$	$0,028$	$0,168$

Q 2.1 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	x_3
y_1	$0,138$	$0,09$	$0,312$
y_2	$0,102$	$0,1$	$0,258$

$P(X) =$

$0,24$
$0,19$
$0,57$

$P(Y) =$

$0,54$
$0,46$

On en déduit donc que :

$P(X)P(Y) =$

	x_1	x_2	x_3
y_1	$0,1296$	$0,1026$	$0,3078$
y_2	$0,1104$	$0,0874$	$0,2622$

S'il y avait eu indépendance entre X et Y, on aurait eu $P(X,Y) = P(X)P(Y)$. Donc X et Y ne sont pas des variables indépendantes.

Q 2.2 Déterminez si X est indépendante de Y conditionnellement à Z. Vous justifierez votre réponse.

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

$$P(Z) = \sum_{X,Y} P(X,Y,Z) = \begin{array}{|c|}\hline 0,3 \\ \hline 0,7 \\ \hline\end{array}$$

Par conséquent :

$P(X,Y|Z) =$

	z_1			z_2		
	x_1	x_2	x_3	x_1	x_2	x_3
y_1	$0,04$	$0,16$	$0,2$	$0,18$	$0,06$	$0,36$
y_2	$0,06$	$0,24$	$0,3$	$0,12$	$0,04$	$0,24$

$P(X|Z) =$

	x_1	x_2	x_3
z_1	$0,1$	$0,4$	$0,5$
z_2	$0,3$	$0,1$	$0,6$

$P(Y|Z) =$

	y_1	y_2
z_1	$0,4$	$0,6$
z_2	$0,6$	$0,4$

On peut effectivement vérifier que $P(X,Y|Z) = P(Y|Z)P(X|Z)$. Donc X et Y sont indépendantes conditionnellement à Z.

Q 2.3 En déduire un réseau bayésien représentant les indépendances que vous avez déterminées.

> Le réseau bayésien est $X \longleftarrow Z \longrightarrow Y$ ou bien $X \longleftarrow Z \longleftarrow Y$ ou encore $X \longrightarrow Z \longrightarrow Y$. Seul le réseau $X \longrightarrow Z \longleftarrow Y$ ne représente pas les indépendances ci-dessus.

Exercice 3 (7 pts)

Un grand quotidien souhaite estimer la proportion de français opposés au projet de loi du gouvernement sur les retraites. Il commande donc une étude auprès d'un organisme de sondage. Celui-ci demande à des personnes dans la rue s'ils sont opposés ou non au projet. Le tableau ci-dessous recense 10 réponses : 0 signifie que l'on est « favorable » au projet et 1 signifie que l'on y est « opposé ».

1	1	0	1	1	1	0	0	1	1

Q 3.1 Donnez une estimation ponctuelle de la proportion de la population opposée à la loi sur les retraites.

> L'estimation ponctuelle revient à calculer $\overline{p} = 7/10$.

Q 3.2 En utilisant un intervalle de confiance de niveau de confiance 95%, estimez la proportion de la population opposée à la loi sur les retraites.

> Soit p la proportion de la population opposée à la réforme. On sait que $\frac{\overline{P}-p}{\sqrt{\frac{p(1-p)}{n}}} \sim \mathcal{N}(0,1)$. Donc, d'après la table de la loi normale :
>
> $$\text{Prob}\left(-z_{\alpha/2} \leq \frac{\overline{P}-p}{\sqrt{\frac{p(1-p)}{n}}} \leq z_{\alpha/2}\right) = 95\% \Longrightarrow z_{\alpha/2} = 1,96.$$
>
> Par conséquent, l'intervalle de confiance que l'on cherche est :
>
> $$\left[\overline{p} - 1,96\sqrt{\frac{p(1-p)}{n}}; \overline{p} + 1,96\sqrt{\frac{p(1-p)}{n}}\right],$$
>
> ce que l'on approche par :
>
> $$\left[\overline{p} - 1,96\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}; \overline{p} + 1,96\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}\right].$$
>
> Or, ici, $n = 10$ et $\overline{p} = 0,7$. Donc l'intervalle de confiance que l'on cherche est : $[0,416; 0,984]$.

Q 3.3 Le gouvernement annonce que moins de 50% des gens sont opposés à la loi. En effectuant un test d'hypothèse de niveau de confiance 95%, pouvez-vous confirmer cette affirmation au vu de l'échantillon que vous observez ?

Nous allons confronter les hypothèses $H_0 =$ « la proportion d'opposants est égale à 0,5 » contre $H_1 =$ « la proportion d'opposants est supérieure ou égale à 50% ».
Sous l'hypothèse H_0, on sait que $\frac{\overline{P}-0{,}5}{\sqrt{\frac{0{,}5\times 0{,}5}{10}}} \sim \mathcal{N}(0,1)$. Par conséquent :

$$
\begin{aligned}
0,05 &= \text{Prob}\left(\overline{P} < c \,\middle|\, \frac{\overline{P}-0,5}{\sqrt{\frac{0{,}5\times 0{,}5}{10}}} \sim \mathcal{N}(0,1)\right) \\
&= \text{Prob}\left(\frac{\overline{P}-0{,}5}{\sqrt{\frac{0{,}5\times 0{,}5}{10}}} < \frac{c-0{,}5}{\sqrt{\frac{0{,}5\times 0{,}5}{10}}} \,\middle|\, \frac{\overline{P}-0{,}5}{\sqrt{\frac{0{,}5\times 0{,}5}{10}}} \sim \mathcal{N}(0,1)\right)
\end{aligned}
$$

Donc, d'après la table de la loi normale, $\frac{c-0{,}5}{\sqrt{\frac{0{,}5\times 0{,}5}{10}}} = 1,645$, ou encore $c \approx 0,76$. Notre règle de décision est donc de rejeter H_0 si et seulement si $\overline{p} > 0,76$. Ici, l'échantillon que nous avons prélevé ayant un $\overline{p} = 0,7$, on ne peut rejeter H_0.