Correction, Statistiques et Probabilités, L2 MASS, Partiel
24/04/2013, Durée : deux heures

Exercice 1. Couple d'entiers
1. La somme des probabilités jointes vaut 1, c.a.d.

$$\sum_{i=1}^{\infty}\sum_{j=0}^{\infty} P(X=i,Y=j) = e\frac{a}{1-a} = 1,$$

on en déduit que $a = \frac{1}{e+1}$.
2. $P(X=i) = \sum_{j=0}^{\infty} P(X=i,Y=j) = a^i e = (\frac{1}{e+1})^i e = (\frac{1}{e+1})^{i-1}\frac{e}{e+1}$.
$P(Y=j) = \sum_{i=1}^{\infty} P(X=i,Y=j) = \frac{1}{j!}\frac{a}{1-a} = \frac{1}{j!}e^{-1}$.
3. $X \sim \mathcal{G}(\frac{e}{e+1})$, $Y \sim \mathcal{P}(1)$.
4. Oui.
5. $E(S) = E(X) - E(Y) = \frac{e+1}{e} - 1 = \frac{1}{e}$.
$Var(S) = Var(X) + Var(Y) = \frac{1}{e+1}/(\frac{e}{e+1})^2 + 1 = \frac{e^2+e+1}{e^2}$.

Exercice 2. Chaîne de Markov
2. M$=\begin{bmatrix} 1/2 & 0 & 1/3 \\ 1/6 & 1 & 1/6 \\ 1/3 & 0 & 1/2 \end{bmatrix}$.
3. L'état récurrent est e_2, les états transitoires sont e_1 et e_3.
4. $\mu = (0,1,0)^T$.
5. Oui, car e_2 est un état absorbant.
6. Puisque $X_1 = MX_0 = (1/3, 1/6, 1/2)^T$, $P(X_1 = a) = 1/2$.
Puisque $X_2 = MX_1 = (1/3, 11/36, 13/36)^T$, $P(X_2 = a) = 13/36$.
$P(X_1 = a|X_2 = a) = \frac{P(X_2=a|X_1=a)P(X_1=a)}{P(X_2=a)} = \frac{1/2\times 1/2}{13/36} = \frac{9}{13}$.

Exercice 3. Régression
1. Par la linéarité des lois gaussiennes, on a $Y_i \sim \mathcal{N}(ax_i, 4)$, d'où vient

$$L(a, y_1, \dots, y_n) = \prod_{i=1}^{n} f_{Y_i}(y_i) = \prod_{i=1}^{n} f_{\mathcal{N}(ax_i,4)}(y_i).$$

2. La log-vraisemblance est $H(a) = n\ln\frac{1}{2\sqrt{2\pi}} - \sum_{i=1}^{n}\frac{(y_i-ax_i)^2}{8}$. En prenant la dérivée $H'(a) = 1/4\sum_{i=1}^{n} x_i(y_i - ax_i) = 0$, on obtient $\hat{a} = \frac{\sum x_i y_i}{\sum x_i^2}$. On vérifie que $H''(a) = -\sum x_i^2/4 \le 0$.

3. $E(\hat{a}) = a$, $Var(\hat{a}) = 4/\sum x_i^2$.

4. Le deuxième jeu de données est meilleur, car la variance de $\hat{a}$ décroît en fonction de $\sum x_i^2$.

5. D'après le lemme Neyman et Pearson, la région de rejet du test le plus puissant a la forme $L_0/L_1 < K$.

On a $L_0 = (\frac{1}{2\sqrt{2\pi}})^n \exp(-\frac{1}{8}\sum Y_i^2)$, et $L_1 = (\frac{1}{2\sqrt{2\pi}})^n \exp(-\frac{1}{8}\sum(Y_i - x_i)^2)$. Ainsi on a $T = \sum x_i Y_i > 1/2(\sum x_i^2 - 8\ln K) := C$.

6. Sous l'hypothèse H_0, la statistique T suit la loi gaussienne d'espérance 0 et de variance 400.

On a donc $P(T/20 > 1.64) = 5\%$, $W = (32, 8, \infty)$.