LSTAT 2040 - TP 1

Rappels sur les vecteurs aléatoires et introduction aux modèles paramétriques

Exercice 1

Soit $X \sim \mathrm{N}_3(\mu, \Sigma)$ où

$$\Sigma = \begin{pmatrix} 4 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 3 \end{pmatrix}.$$

Pour quelles valeurs de a et de b la variable aléatoire $X_3 + aX_1 + bX_2$ est-elle indépendante de (X_1, X_2) ?

Exercice 2

Soit (X_1, X_2) un vecteur aléatoire dont la fonction de densité conjointe est donnée par

$$f(x_1, x_2) = 1 + \alpha(2x_1 - 1)(2x_2 - 1), \quad 0 < x_1, x_2 < 1, \quad -1 \leq \alpha \leq 1$$

et $f(x_1, x_2) = 0$ ailleurs.

(a) Vérifier que cette fonction est effectivement une fonction de densité.
(b) Calculer $\mathrm{E}[X_1 | X_2 = x_2]$ et $\mathrm{Var}[X_1 | X_2 = x_2]$.
(c) Pour cette famille de distribution, montrer que X_1 est indépendante de X_2 si et seulement si $\mathrm{Corr}[X_1, X_2] = 0$.

Exercice 3

Soient X_1 et X_2 deux variables aléatoires exponentielles de moyennes $1/\theta_1$ et $1/\theta_2$ respectivement, supposées indépendantes. Quelle est la distribution de $T = \max(X_1, X_2)$?

Exercice 4

On considère le vecteur aléatoire $X \sim \mathrm{N}_2(\mu, \Sigma)$, où

$$\mu = (2, 2)^t, \qquad \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Notons $A = (1, 1)$ et $B = (1, -1)$. Les variables aléatoires AX et BX sont-elles indépendantes ?

Exercice 5

On considère le vecteur aléatoire $T = (X, Y, Z)$ de densité

$$f_T(x, y, z) = k(x + yz), \quad x, y, z \in [0, 1]$$

et $f_T(x, y, z) = 0$ ailleurs.

(a) Déterminer la valeur de k pour que f_T soit une fonction de densité.
(b) Déterminer la distribution marginale de X.
(c) Déterminer la covariance entre les variables aléatoires Y et Z.
(d) Déterminer la distribution de $(Y, Z | X)$.

Exercice 6

Soit $X_1, \ldots, X_n$ un échantillon iid issu d'une loi normale $\mathrm{N}(\mu, \sigma^2)$. On définit $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ et $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$.

(a) Montrer que $\overline{X}_2$ et S_2^2 sont indépendants.

(b) Pour $n \geq 2$, montrer que

$$\begin{pmatrix} \overline{X}_n \\ X_1 - \overline{X}_n \\ \vdots \\ X_n - \overline{X}_n \end{pmatrix} = B \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix},$$

où B est une matrice de dimension $(n+1) \times n$ satisfaisant

$$BB^t = \begin{pmatrix} \frac{1}{n} & 0 & \ldots & 0 \\ 0 & & & \\ \vdots & & A & \\ 0 & & & \end{pmatrix},$$

où A est une certaine matrice de dimension $n \times n$ qu'il ne faut pas déterminer.

(c) En déduire, pour $n \geq 2$, que $\overline{X}_n$ et S_n^2 sont indépendants.

Exercice 7

Soit le modèle paramétrique formé par une mixture de deux normales :

$$f(x;\theta) = p\phi(x;\mu_1,\sigma_1^2) + (1-p)\phi(x;\mu_2,\sigma_2^2), \qquad \theta = (p, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2) \in (0,1) \times \mathbb{R}^2 \times (0,\infty)^2,$$

où $\phi(\cdot;\mu,\sigma^2)$ est la fonction de densité d'une normale $\mathrm{N}(\mu, \sigma^2)$. Ce modèle est-il identifiable ?

Exercice 8

Vérifier si les distributions suivantes appartiennent à une famille exponentielle :

(a) $\mathrm{N}(1, \sigma^2)$, $\quad \sigma > 0$.
(b) $\mathrm{N}(\mu, \mu)$, $\quad \mu > 0$.
(c) $\mathrm{Geo}(\pi)$, $\quad 0 < \pi \leq 1$.
(d) $\mathrm{Unif}[0, \theta]$, $\quad \theta > 0$.
(e) $\mathrm{Rayleigh}(\sigma)$, $\quad \sigma > 0$.
(f) $\mathrm{Weibull}(\lambda, k)$, $\quad \lambda, k > 0$.

Exercice 9

Soit $X_{11}, \ldots, X_{1n_1}$ et $X_{21}, \ldots, X_{2n_2}$ deux échantillons iid de distribution $\mathrm{Exp}(\lambda_i)$, $i \in \{1,2\}$, respectivement. Soit $\theta = (\lambda_1, \lambda_2)$. Montrer que la distribution (jointe) du vecteur aléatoire

$$X = (X_{11}, \ldots, X_{1n_1}, X_{21}, \ldots, X_{2n_2})$$

appartient à une famille exponentielle de dimension k et déterminer k.

Exercice 10

Soit $\mathcal{P}$ la classe des distributions multinomiales avec paramètres n connu et $\pi = (\pi_1, \ldots, \pi_m) \in (0,1)^m$ tel que $\sum_{j=1}^m \pi_j = 1$ inconnu. Montrer que $\mathcal{P}$ est une famille exponentielle de dimension k et déterminer k.