MAT7381 Exercices
Chapitre 2 - Loi normale

Rappelons la convention adoptée dans ce cours : $\boldsymbol{e}$ ou $\boldsymbol{e}_n$ désigne un vecteur de n « 1 » ; $\mathbf{J}$ ou $\mathbf{J}_n$ désigne une matrice $n \times n$ de « 1 ». Généralement $\mathbf{C}$ ou $\mathbf{C}_n$ désigne la matrice $\mathbf{I}$-$\mathbf{J}/n$.

2.1 Soit $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim \mathcal{N}(\mathbf{0} \,;\, \mathbf{I}_2)$. On sait que $(x_1 - x_2)$ et $(x_1 + x_2)$ sont alors indépendantes. Soit $z_1 = 2x_1 + 3x_2$. Déterminer une fonction linéaire $z_2 = a_1x_1 + a_2x_2$ telles que z_1 et z_2 sont indépendantes.

2.2 Soit $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim \mathcal{N}(\mathbf{0} \,;\, \boldsymbol{\Sigma})$ où $\boldsymbol{\Sigma} = \begin{bmatrix} 41 & -12 \\ -12 & 34 \end{bmatrix}$.

a) Soit $z_1 = x_1 + x_2$. Déterminer une fonction linéaire $z_2 = a_1x_1 + a_2x_2$ telles que z_1 et z_2 sont indépendantes.

b) Déterminer la matrice de covariance du vecteur aléatoire $\boldsymbol{y} = \begin{bmatrix} x_1 \\ x_2 \\ x_1 - x_2 \end{bmatrix}$ et montrer qu'elle n'est pas définie positive.

c) Montrer que les vecteurs $\boldsymbol{\ell}_1 = \begin{bmatrix} 0,6 \\ 0,8 \end{bmatrix}$ et $\boldsymbol{\ell}_2 = \begin{bmatrix} 0,8 \\ -0,6 \end{bmatrix}$ sont des vecteurs propres de $\boldsymbol{\Sigma}$ et que les valeurs propres correspondantes sont $\lambda_1 = 25$ et $\lambda_2 = 50$.

d) Soit $\mathbf{L} = [\boldsymbol{\ell}_1 \,;\, \boldsymbol{\ell}_2]$. Montrer que $\boldsymbol{z} = \mathbf{L}'\boldsymbol{x}$ est $\mathcal{N}(\mathbf{0} \,;\, \mathbf{D})$, où $\mathbf{D} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \begin{bmatrix} 25 & 0 \\ 0 & 50 \end{bmatrix}$.

2.3 Supposons que $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim \mathcal{N}(\mathbf{0} \,;\, \boldsymbol{\Sigma})$ où $\boldsymbol{\Sigma} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$. Déterminer une matrice $\mathbf{A}_{2\times 2}$ telle que $\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \mathbf{A}\boldsymbol{x}$ soit $\mathcal{N}(\mathbf{0} \,;\, \mathbf{I}_2)$.

2.4 Soit y_1 ; ... ; y_n des variables aléatoires indépendantes, y_i de loi $\mathcal{N}(\mu_i \,;\, \sigma^2)$. Montrez que

$$E\left[\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}\right] = \sigma^2 + \frac{\sum_{i=1}^{n}(\mu_i - \bar{\mu})^2}{n-1}\text{, où } \bar{\mu} = \frac{1}{n}\sum_{i=1}^{n}\mu_i\ .$$

[*Écrivez* $\sum_{i=1}^{n}(y_i - \bar{y})^2$ *sous sa forme matricielle* $\mathbf{y}'\mathbf{C}\mathbf{y}$ *et appliquez le théorème* 2.1.3]

2.5 Soit $\bar{y}_1, \ldots, \bar{y}_k$ les moyennes de k échantillons de tailles n_1, ..., n_k indépendants tirés de populations normales de moyennes μ_1, ... , μ_k et de variances σ_i^2.

a) Supposons que les moyennes sont égales : $\mu_1 = \ldots = \mu_k = \mu$.

1. Montrez que $\Sigma a_i \bar{y}_i$ est sans biais pour μ si et seulement si $\Sigma a_i = 1$.

2. Montrez que sous la contrainte $\Sigma a_i = 1$, la variance de $\Sigma a_i \bar{y}_i$ est minimisée pour $a_i = \dfrac{n_i/\sigma_i^2}{\sum_{j=1}^{k} n_j/\sigma_j^2}$. [*Considérez* $1 = \Sigma a_i = \sum_{i=1}^{k} \dfrac{a_i\sigma_i}{\sqrt{n_i}}\dfrac{\sqrt{n_i}}{\sigma_i}$, *et utilisez l'inégalité de Cauchy-Schwartz; ou encore, minimisez sous contrainte à l'aide du multiplicateur de Lagrange*].

b) Déterminez la matrice de covariance $\boldsymbol{\Sigma}$ du vecteur $\bar{\boldsymbol{y}} = \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_k \end{bmatrix}$.

c) Soit les constantes a_1 , ... , a_k et b_1 , ... , b_k. Déterminez une condition nécessaire et suffisante pour que

$\mathrm{Cov}(\Sigma a_i\, \bar{y}_i\, ; \Sigma b_i\, \bar{y}_i\,) = 0$. Exprimez la condition en termes de $\mathbf{\Sigma}$ et des vecteurs $\boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{bmatrix}$ et $\boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}$.

2.6 Soit y_1 ;... ; y_n n variables aléatoires normales indépendantes de variances $\sigma_1^2, \ldots, \sigma_n^2$.

a) Déterminez une condition nécessaire et suffisante pour qu'une combinaison linéaire $\Sigma a_i y_i$ soit indépendante de la moyenne $\bar{y}$.

b) Supposez que les variances sont égales: $\sigma_1^2 = \ldots = \sigma_n^2 = \sigma^2$. Montrez que pour tout i, la variable $y_i - \bar{y}$ est indépendante de $\bar{y}$, et concluez que $\bar{y}$ et $S^2 = \sum_{i=1}^n (y_i - \bar{y})^2/(n-1)$ sont indépendantes.

2.7 Soit y_1, y_2, et y_3 des variables aléatoires indépendantes, chacune de moyenne 0 et de variance σ^2. Soit $z_1 = y_1$, $z_2 = y_1 + y_2$, et $z_3 = y_1 + y_2 + y_3$. Déterminez la matrice de covariance de $\boldsymbol{z} = (z_1\, ; z_2\, ; z_3)'$.

2.8 Considérer un vecteur aléatoire $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$ de loi normale multivariée de moyenne $\boldsymbol{\mu} = \begin{bmatrix} 2 \\ 4 \\ 5 \\ 3 \end{bmatrix}$ et de matrice de covariance

$$\mathbf{\Sigma} = \begin{bmatrix} 11.69 & 6.45 & 8.32 & -2.32 \\ 6.45 & 26.00 & 13.06 & 8.45 \\ 8.32 & 13.06 & 29.36 & 13.43 \\ -2.32 & 8.45 & 13.43 & 17.61 \end{bmatrix}.$$

Les commandes R qui suivent chacune des questions vous proposent une façon de procéder. On devine assez bien le sens de chacune. Voici comment saisir les données.

```
Sigma<-matrix(c(11.69,6.45,8.32,-2.32,6.45,26,13.06,8.45,8.32,13.06,29.36,13.43,-2.32,8.45,13.43,17.61),4,4)
mu<-c(2,4,5,3)
```

a) Déterminer la distribution conditionnelle de x_1 étant donné $\begin{bmatrix} x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 3 \end{bmatrix}$.

```
Sigma11<-Sigma[1,1]
Sigma12<-Sigma[1,2:4]
Sigma22<-Sigma[2:4,2:4]
Sigma22Inverse<-solve(Sigma22)                                   L'inverse de Σ22
```

b) Déterminer la distribution conditionnelle de $\boldsymbol{y}_1 = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ étant donné $\begin{bmatrix} x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$

c) Déterminer la distribution marginale de $\boldsymbol{z} = \begin{bmatrix} 1 & 2 & 1 & 6 \\ 3 & 4 & 2 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$.

```
L<-matrix(c(1,2,1,6,3,4,2,4),4,2)
V<-t(L)%*%Sigma%*%L
```

d) Déterminer la distribution conditionnelle de $\underline{x}_1$ étant donné $2x_2+3x_3 = 3$.

e) Définir une transformation $\boldsymbol{y} = \mathbf{L}'\boldsymbol{x}$ telle que les composantes de $\boldsymbol{y}$ soient indépendantes ($\mathbf{L}$ est 4×4).

f) Déterminer la probabilité que la moyenne des composantes de $\boldsymbol{x}$ soit inférieure ou égale à 14. [On peut obtenir les probabilités cumulatives $P(X \le x)$ d'une variable X de loi normale avec R à l'aide de la commande `> pnorm(x,m,sigma)` où `m` est la moyenne et `sigma` est l'écart-type]

2.9 Soit $\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \sim \mathcal{N}_3(\boldsymbol{\mu}_y\,;\,\boldsymbol{\Sigma}_y)$, où $\boldsymbol{\Sigma}_y = \frac{1}{6}\begin{bmatrix} 11 & -1 & -4 \\ -1 & 11 & -4 \\ -4 & -4 & 14 \end{bmatrix}$ et $\boldsymbol{\mu}_y = \begin{bmatrix} 5 \\ 5 \\ 5 \end{bmatrix}$. Les valeurs propres de $\boldsymbol{\Sigma}$ sont $\lambda_1 = 1$, $\lambda_2 = 2$, et $\lambda_3 = 3$ avec vecteurs propres normalisés correspondants $\boldsymbol{p}_1 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $\boldsymbol{p}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$ et $\boldsymbol{p}_3 = \frac{1}{\sqrt{6}}\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$.

a) Trouver une fonction linéaire $\boldsymbol{z} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \mathbf{L}'\boldsymbol{y}$ de moyenne $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$. Choisir $\mathbf{L}$ de telle sorte que la matrice de covariance de $\boldsymbol{z}$ soit non singulière.[*Il est possible de répondre sans faire le moindre calcul, mais vous devez justifier votre choix*]

b) Trouver deux fonctions linéaires $z_1 = \boldsymbol{a}'\boldsymbol{y}$ et $z_2 = \boldsymbol{b}'\boldsymbol{y}$ telles que z_1 et z_2 sont indépendantes.[*Ici aussi, vous pouvez vous en tirer sans aucun calcul, à condition de justifier votre choix*]

c) Montrez que pour une certaine valeur de k, la statistique $F = \frac{k(y_1 - y_2)^2}{(y_1 + y_2 - 2y_3)^2}$ suit une loi de Fisher et déterminer la valeur de k. [*Vous pouvez simplifier les calculs en faisant intervenir les vecteurs propres dans votre discussion*].

2.10 Soit $\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \sim \mathcal{N}_3(\boldsymbol{\mu}_y\,;\,\boldsymbol{\Sigma}_y)$, où $\boldsymbol{\Sigma}_y = \begin{bmatrix} 5 & -2 \\ -2 & 8 \end{bmatrix}$ et $\boldsymbol{\mu}_y = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$. Les valeurs propres de $\boldsymbol{\Sigma}$ sont $\lambda_1 = 4$ et $\lambda_2 = 9$, avec vecteurs propres normalisés correspondants $\boldsymbol{p}_1 = \frac{1}{\sqrt{5}}\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ et $\boldsymbol{p}_2 = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ -2 \end{bmatrix}$. Trouver une fonction linéaire $\boldsymbol{z} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \mathbf{A}\boldsymbol{y}$ telle que $\mathrm{Var}(\boldsymbol{z}) = \mathbf{I}_2$.

2.11 Soit $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim \mathcal{N}(\mathbf{0}\,;\,\sigma^2\mathbf{I}_2)$, $\sigma^2 > 0$ fixé. Identifiez les lois des variables aléatoires suivantes :

a) $2x_1$ b) $x_1 + x_2$ c) $\frac{x_1^2}{x_2^2}$ d) $\frac{(x_1 - x_2)^2}{(x_1 + x_2)^2}$ e) $\frac{x_1}{|x_2|}$

2.12 Supposons que $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim \mathcal{N}(\mathbf{0}\,;\,\boldsymbol{\Sigma})$ où $\boldsymbol{\Sigma} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$. Trouvez la matrice $\mathbf{A}_{2\times 2}$ telle que $\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \mathbf{A}\boldsymbol{x}$ soit $\mathcal{N}(\mathbf{0}\,;\,\mathbf{I}_2)$ et en déduire une fonction de $\boldsymbol{x}_1$ et $\boldsymbol{x}_2$ qui soit de loi χ^2_2.

2.13 Soit $y_1, \ldots, y_n$ des variables aléatoires indépendantes, chacune de loi $\mathcal{N}(\mu\,;\,\sigma^2)$. [*Écrivez ces variables sous forme matricielle et vérifiez que* $\sum_{i=1}^n (y_i - \bar{y})^2 = \boldsymbol{y}'\mathbf{C}\boldsymbol{y}$]

a) Montrez que $\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sigma^2} \sim \chi^2_{n-1}$.

b) Montrez que $\frac{n\bar{y}^2}{\sigma^2} \sim \chi^2_1$ non centrale, avec paramètre de non centralité $n\mu^2/\sigma^2$.

c) Montrez que $\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sigma^2}$ et $\frac{n\bar{y}^2}{\sigma^2}$ sont indépendantes.

d) Montrez que $\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sigma^2}$ et $\bar{y}$ sont indépendantes.

2.14 Soit $X_1, X_2, \ldots, X_m$ et $Y_1, Y_2, \ldots, Y_n$ deux échantillons normaux indépendants de moyennes μ_1 et μ_2, respectivement et de même variance σ^2. Montrer que la « statistique t » habituelle, $T = \frac{\bar{Y} - \bar{X}}{S\sqrt{\frac{1}{m} + \frac{1}{n}}}$, où $S = \sqrt{\frac{\sum_{i=1}^m (X_i - \bar{X})^2 + \sum_{j=1}^n (Y_j - \bar{Y})^2}{m + n - 2}}$, $\bar{X} = \frac{\sum_{i=1}^m X_i}{m}$ et $\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n}$, suit une loi de Student à m+n-2 degrés de liberté lorsque $\mu_1 = \mu_2$.

2.15 Soit $y_1, \ldots, y_n$ des variables aléatoires indépendantes de même moyenne μ et de variances $\mathrm{Var}(y_i) = \sigma_i^2$ et soit $\bar{y} = \sum_{i=1}^n y_i / n \sum_{i=1}^n y_i / n$.

a) Déterminez $\mathrm{Var}(\bar{y})$

b) Déterminez une constante K telle que $K\sum_{i=1}^n (y_i - \bar{y})^2$ est un estimateur sans biais de $\mathrm{Var}(\bar{y})$.

2.16 Soit $\boldsymbol{y} \sim \mathcal{N}_n(\mu\boldsymbol{e}\,;\, \boldsymbol{\Sigma})$, où μ est un scalaire, $\boldsymbol{e} = (1, \ldots, 1)'$, et $\boldsymbol{\Sigma}$ est tel que $\mathrm{Var}(y_i) = \sigma^2\ \forall\ i$, $\mathrm{cov}(y_i, y_j) = \rho\sigma^2\ \forall\ i \neq j$, ρ étant le coefficient de corrélation commun entre toute paire de variables.

a) Montrez que $\dfrac{\sum_{i=1}^n (y_i - \bar{y})^2}{(1-\rho)\sigma^2} \sim \chi^2_{n-1}$ centrale [*Vous pouvez trouver une expression pour* $\boldsymbol{\Sigma}$ *en fonction du vecteur* $\boldsymbol{e}$, *de l'identité* $\mathbf{I}$ *et des scalaires* σ^2 *et* ρ].

b) Montrez que $\rho \geq -1/(n-1)$ [*Utilisez le fait que* $\mathrm{Var}(\Sigma_i\, y_i) \geq 0$].

2.17 Soit b le revenu *moyen* de tous les ménages d'une ville tirée au hasard d'une très grande (pratiquement infinie) population de villes. Supposons que $b \sim \mathcal{N}(\beta\,;\, \sigma_b^2)$. Soit y le revenu d'un ménage choisi au hasard dans une ville *donnée*, et supposons que, conditionnellement, $y \sim \mathcal{N}(b\,;\, \sigma^2)$, où b est le revenu moyen des ménages de la ville sélectionnée (Vous supposerez que le nombre de ménages dans chaque ville est pratiquement infini). Vous tirez une ville dans la population, puis r ménages dans la ville sélectionnée. Soit $y_1, y_2, \ldots, y_r$ les revenus de ces r ménages et $\bar{y}$ leur moyenne.

a) Montrer que $\mathrm{E}(y_1) = \mathrm{E}(\bar{y}) = \beta$

b) Montrer que $\mathrm{Var}(y_1) = \sigma_b^2 + \sigma^2$

c) Montrer que $\mathrm{Cov}(y_j\,;\, y_k) = \sigma_b^2$ pour $j \neq k$.

d) Montrer que $\mathrm{Var}(\bar{y}) = \sigma^2/r + \sigma_b^2$

e) Montrer que $\mathrm{E}\left(\dfrac{\sum_{j=1}^r (y_j - \bar{y})^2}{r-1}\right) = \sigma^2$ et que $\dfrac{\sum_{j=1}^r (y_j - \bar{y})^2}{\sigma^2} \sim \chi^2_{r-1}$ [*Écrire sous forme matricielle la moyenne* $\boldsymbol{\mu}$ *et la matrice de covariance* $\boldsymbol{\Sigma}$ *du vecteur* $\boldsymbol{y} = [y_1; y_2\,; \ldots; y_r]'$].

f) Montrer que $\sum_{j=1}^r (y_j - \bar{y})^2$ et $\bar{y}$ sont indépendantes.

[*Suggestion : Une formulation équivalente à celle de l'énoncé ci-dessus est celle-ci* : $y_i = b + \varepsilon_i$, *où b est une variable aléatoire normale de moyenne* β *et de matrice de covariance* $\sigma_b^2\mathrm{I}$, *indépendantes de* $\varepsilon_i \sim \mathcal{N}(0\,;\, \sigma^2)$].

2.18 Soit y le pouls d'une personne tirée au hasard et soumis à un exercice d'une durée x. Nous supposerons que la relation entre y et x est celle d'une régression linéaire simple, c'est-à-dire, $y = a + \beta x + \varepsilon$, où $\varepsilon \sim \mathcal{N}(0\,;\, \sigma^2)$. On suppose que β est un paramètre fixe alors que a est une valeur qui varie d'une personne à l'autre. Admettons que $a \sim \mathcal{N}(\alpha\,;\, \sigma_a^2)$, indépendamment de ε. Supposons que k mesures $y_1, y_2, \ldots, y_k$ sont prises sur un même individu après des séances d'exercices de durées $x_1, x_2, \ldots, x_k$ fixées d'avance. On a donc le modèle suivant : $\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix} = a\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \beta\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_k \end{bmatrix}$, ou $\boldsymbol{y} = a\boldsymbol{e} + \beta\boldsymbol{x} + \boldsymbol{\varepsilon}$. Donc $\boldsymbol{y}$ est de moyenne $\boldsymbol{\mu} = \alpha\boldsymbol{e} + \beta\boldsymbol{x}$ et de matrice de covariance $\boldsymbol{\Sigma} = \sigma^2\mathbf{I} + \sigma_a^2\mathbf{J}$

a) Soit $\bar{y} = \dfrac{1}{k}\sum y_i = \dfrac{\boldsymbol{e}'\boldsymbol{y}}{k}$. Montrer que $\mathrm{E}(\bar{y}) = \alpha + \beta\bar{x}$ et que $\mathrm{Var}(\bar{y}) = \sigma_a^2 + \dfrac{\sigma^2}{k}$.

b) Soit $\mathbf{E}$ une matrice $k\times(k-1)$ dont les colonnes sont orthogonales à $\boldsymbol{e}$ et $\boldsymbol{z} = \begin{bmatrix} z_1 \\ \boldsymbol{z}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{e}'/k \\ \mathbf{E}' \end{bmatrix}\boldsymbol{y}$. Montrer que $\mathrm{E}(\boldsymbol{z}) = \begin{bmatrix} \alpha + \beta\bar{x} \\ \beta\mathbf{E}'\boldsymbol{x} \end{bmatrix}$ et que $\mathrm{Var}(\boldsymbol{z}) = \begin{bmatrix} \sigma_a^2 + \sigma^2/k & \mathbf{0} \\ \mathbf{0} & \sigma^2\mathbf{I}_{k-1} \end{bmatrix}$.[*Si on choisit* $\mathbf{E}$ *de telle sorte que ses colonnes soient orthogonales* (*ce qu'on a toujours le loisir de faire*) *alors* $\mathbf{EE}'$ *est la matrice* $\mathbf{C} = \mathbf{I} - (1/n)\boldsymbol{ee}'$].

c) Montrer que $\hat{\beta} = \dfrac{\boldsymbol{x}'\mathbf{C}\boldsymbol{y}}{\boldsymbol{x}'\mathbf{C}\boldsymbol{x}} = \dfrac{\sum(x_i - \bar{x})y_i}{\sum(x_i - \bar{x})^2}$ est un estimateur sans biais de β et que la variance de $\hat{\beta}$ est $\dfrac{\sigma^2}{\boldsymbol{x}'\mathbf{C}\boldsymbol{x}}$.

d) Montrer que $\hat{\sigma}^2 = \dfrac{\sum(y_i - \bar{y})^2 - \dfrac{[\sum(x_i - \bar{x})y_i]^2}{\sum(x_i - \bar{x})^2}}{k-2}$ est un estimateur sans biais de σ^2 et que $\dfrac{(k-2)\hat{\sigma}^2}{\sigma^2}$ est de loi χ^2 centrale.[*Vérifier que* $\hat{\sigma}^2 = \dfrac{\boldsymbol{y}'\mathbf{C}\boldsymbol{y} - \dfrac{\boldsymbol{y}'\mathbf{C}\boldsymbol{x}\boldsymbol{x}'\mathbf{C}\boldsymbol{y}}{\boldsymbol{x}'\mathbf{C}\boldsymbol{x}}}{k-2}$.]

e) Montrer que $\hat{\alpha} = \dfrac{\bar{y}\sum x_i^2 - \bar{x}\sum x_i y_i}{\sum(x_i - \bar{x})^2}$ est un estimateur sans biais de α.

2.19 Soit $\boldsymbol{y} \sim \mathcal{N}_n(\boldsymbol{\mu} ; \mathbf{I}_n)$ où $\boldsymbol{\mu} \in \mathcal{C}(\mathbf{X})$ et $\mathbf{X}$ est une matrice n x q de rang plein.

a) Montrez que $\boldsymbol{y}'[\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\boldsymbol{y} \sim \chi^2_{n-q}$ centrale.

b) Montrez que $\boldsymbol{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y} \sim \chi^2_q$ non centrale avec paramètre de non centralité $\boldsymbol{\mu}'\boldsymbol{\mu}$.

2.20 Soit $\boldsymbol{y} = [y_1; \ldots; y_{n-1} ; y_n]' \sim \mathcal{N}(\mu\boldsymbol{e}_n ; \sigma^2\mathbf{I}_n)$.

a) Déterminer la distribution conjointe du *vecteur* $\boldsymbol{z} = [\boldsymbol{z}_1' ; \bar{y}]'$, où $\boldsymbol{z}_1 = [y_1; \ldots; y_{n-1}]'$, et $\bar{y}$ est la moyenne des n observations.

b) Montrer que la distribution *conditionnelle de* $\boldsymbol{z}_1$ étant donné $\bar{y}$ est normale de moyenne $\bar{y}\boldsymbol{e}_{n-1}$ et de matrice de covariance $\sigma^2\left(\mathbf{I}_{n-1} - \dfrac{1}{n}\mathbf{J}_{n-1}\right)$.

c) Montrer que $Q = \dfrac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sigma^2} = \dfrac{(\boldsymbol{z}_1 - \bar{y}\boldsymbol{e})'(\mathbf{I}_{n-1} + \mathbf{J}_{n-1})(\boldsymbol{z}_1 - \bar{y}\boldsymbol{e})}{\sigma^2}$.

d) Montrer que la distribution *conditionnelle* de $Q = \dfrac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sigma^2}$ étant donné $\bar{y}$ est χ^2_{n-1}. En déduire que c'est également la distribution *non* conditionnelle et que Q est indépendante de $\bar{y}$.

2.21 Soit $\boldsymbol{z} = \begin{bmatrix} \boldsymbol{z}_1 \\ \boldsymbol{z}_2 \end{bmatrix}$ un vecteur de loi normale ($\boldsymbol{z}_1$ et $\boldsymbol{z}_2$ de dimensions n_1 et n_2) de moyenne $\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}$ et de matrice de covariance $\boldsymbol{\Sigma} = \sigma^2\begin{bmatrix} \mathbf{I}_{n_1} & 0 \\ 0 & \mathbf{I}_{n_2} \end{bmatrix}$. Soit $\boldsymbol{f} = \boldsymbol{f}(\boldsymbol{z}_1)$ un vecteur $n\times 1$ ($n = n_1+n_2$) dont les composantes sont fonctions de $\boldsymbol{z}_1$ seulement, $\mathbf{P}$ une matrice $n\times n_2$ telle que $\mathbf{P}'\mathbf{P} = \mathbf{I}_{n_2}$.

a) Montrer que $Q_1 = \dfrac{(\boldsymbol{z}_2\text{-}\boldsymbol{\mu}_2)'\boldsymbol{P}'\boldsymbol{f}\boldsymbol{f}'\boldsymbol{P}(\boldsymbol{z}_2\text{-}\boldsymbol{\mu}_2)}{\boldsymbol{f}'\boldsymbol{P}\boldsymbol{P}'\boldsymbol{f}\sigma^2} \sim \chi^2_1$

b) Montrer que $Q_2 = (\boldsymbol{z}_2\text{-}\boldsymbol{\mu}_2)'\left(\mathrm{I} - \dfrac{\boldsymbol{P}'\boldsymbol{f}\boldsymbol{f}'\boldsymbol{P}}{\boldsymbol{f}'\boldsymbol{P}\boldsymbol{P}'\boldsymbol{f}}\right)(\boldsymbol{z}_2\text{-}\boldsymbol{\mu}_2)/\sigma^2 \sim \chi^2_{n_2-1}$ et que Q_1 et Q_2 sont indépendantes.

2.22 Soit $\boldsymbol{x} \sim \mathcal{N}_p(\boldsymbol{\mu} ; \boldsymbol{\Sigma})$, et $\boldsymbol{a}$ un vecteur fixe.

a) Montrer que $\dfrac{\boldsymbol{a}'(\boldsymbol{x}-\boldsymbol{\mu})}{\sqrt{\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}}} \sim \mathcal{N}(0 ; 1)$.

b) Soit $\boldsymbol{a}$ un vecteur aléatoire indépendant de $\boldsymbol{x}$. Montrer que $\dfrac{\boldsymbol{a}'(\boldsymbol{x}-\boldsymbol{\mu})}{\sqrt{\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}}} \sim \mathcal{N}(0 ; 1)$.

c) Soit $\boldsymbol{x} \sim \mathcal{N}_4(0 ; \mathbf{I})$. Montrer que $\dfrac{x_1 e^{x_3} + x_2 \log|x_4|}{\sqrt{e^{2x_3} + (\log|x_4|)^2}} \sim \mathcal{N}(0 ; 1)$.

2.23 Soit $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ n vecteurs aléatoires indépendants, $\boldsymbol{x}_i = \begin{bmatrix} u_i \\ v_i \end{bmatrix} \sim \mathcal{N}_2(\boldsymbol{\mu}_i ; \boldsymbol{\Sigma})$, $i = 1, \ldots, n$, où $\boldsymbol{\mu}_i = \begin{bmatrix} \mu_i \\ \mu_i + \delta \end{bmatrix}$ et $\boldsymbol{\Sigma}$ =

$\begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix}$ (les u_i et les v_i sont de même variance). Soit $\bar{u} = \frac{1}{n}\sum_{i=1}^{n} u_i$; $\bar{v} = \frac{1}{n}\sum_{i=1}^{n} v_i$; $d_i = v_i - u_i$; $m_i = (u_i+v_i)/2$; $\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i$; $\bar{m} = \frac{1}{n}\sum_{i=1}^{n} m_i$.

a) Montrer que $Q_1 = \frac{n(\bar{v}-\bar{u})^2}{2(1-\rho)\sigma^2} \sim \chi^2_{\nu_1}(\lambda_1)$, déterminer ν_1 et λ_1 et énoncer une condition sous laquelle $\lambda_1 = 0$.

b) Montrer que $Q_2 = \frac{2\sum_{i=1}^{n}(m_i-\bar{m})^2}{(1+\rho)\sigma^2} \sim \chi^2_{\nu_2}(\lambda_2)$, déterminer ν_2 et λ_2 et énoncer une condition sous laquelle $\lambda_2 = 0$.

c) Montrer que $Q_3 = \frac{\sum_{i=1}^{n}(d_i-\bar{d})^2}{2(1-\rho)\sigma^2} \sim \chi^2_{\nu_3}(\lambda_3)$, déterminer ν_3 et λ_3 et énoncer une condition sous laquelle $\lambda_3 = 0$.

2.24 Soit $\boldsymbol{y}' = [y_1 ; y_2 ; \ldots; y_q]$ un vecteur de loi $\mathcal{N}_q(\mathbf{0} ; \mathbf{I})$. Déterminer l'espérance et la variance de $(y_1-y_2)^2 + (y_2-y_3)^2 + \ldots + (y_{q-1}-y_q)^2$. [*Déterminer la matrice* **A** *telle que* $\boldsymbol{y}'\mathbf{A} = [y_1-y_2 ; y_2-y_3; \ldots ; y_{q-1}-y_q]$]

2.25 Soit $y_1, y_2, \ldots, y_n$ n variables aléatoires indépendantes de loi $\mathcal{N}(0 ; \sigma^2)$. Montrer que $Q = \frac{1}{2(n-1)}\sum_{j=1}^{n-1}(y_{j+1}-y_j)^2$ est une estimateur sans biais de σ^2.

2.26 Soit $\boldsymbol{y} \sim \mathcal{N}_q(\boldsymbol{\mu} ; \boldsymbol{\Sigma})$ et **A** une matrice symétrique. Montrer en suivant les étapes ci-dessous, que

$$\mathrm{Var}(\boldsymbol{y}'\mathbf{A}\boldsymbol{y}) = 2\, tr(\mathbf{A\Sigma A\Sigma}) + 4\boldsymbol{\mu}'\mathbf{A\Sigma A}\boldsymbol{\mu}.$$

a) Si $\boldsymbol{\mu} = \mathbf{0}$, et $\boldsymbol{\Sigma} = \mathbf{I}$, alors $\mathrm{Var}(\boldsymbol{y}'\mathbf{A}\boldsymbol{y}) = 2\, tr(\mathbf{A}^2)$.

b) Si $\boldsymbol{\mu} = \mathbf{0}$, $\mathrm{Var}(\boldsymbol{y}'\mathbf{A}\boldsymbol{y}) = 2\, tr(\mathbf{A\Sigma A\Sigma})$.

c) Si $\boldsymbol{\mu} = \mathbf{0}$ et $\boldsymbol{b}$ un vecteur fixe, alors $\mathrm{Cov}(\boldsymbol{y}'\mathbf{A}\boldsymbol{y} ; \boldsymbol{b}'\boldsymbol{y}) = 0$.

d) Démontrer l'énoncé à partir des résultats précédents.

[*Utiliser le fait que si* $X \sim \mathcal{N}(0 ; 1)$, $E(X^3) = 0$ *et* $E(X^4) = 3$.]

2.27 Soit $y \sim \mathcal{N}_n(\mathbf{0} ; \mathbf{I})$ et **A** une matrice $n\times q$ de rang q, $q < n$. Montrer que la distribution conditionnelle de $\boldsymbol{y}'\boldsymbol{y}$ étant donnée $\mathbf{A}'\boldsymbol{y} = \mathbf{0}$ est khi-deux à n-q degrés de liberté. [*Vous pouvez, sans perte de généralité, supposer que les colonnes de* **A** *sont orthogonales. Pourquoi* ?]

2.28 Soit $\boldsymbol{y} \sim \mathcal{N}_n(\mathbf{0} ; \boldsymbol{\Sigma})$ et **A** une matrice symétrique. Montrer que la fonction génératrice des moments de $\boldsymbol{y}'\mathbf{A}\boldsymbol{y}$ est $|\mathbf{I}-2t\mathbf{A\Sigma}|^{-½}$.