TD1 : Vecteurs gaussiens, modèle gaussien

Exercice 1. Soit $\Sigma_a = \begin{pmatrix} 1 & a \\ a & 1 \end{pmatrix}$ où a est un réel.

1. Donner une condition nécessaire et suffisante sur le réel a pour que Σ soit la matrice de variance-covariance d'un vecteur gaussien.
2. On suppose de plus que ce vecteur gaussien est centré. Donner l'expression analytique de sa densité de probabilité.
3. On note $(X_1, \ldots, X_n)$ un n-échantillon de loi $\mathcal{N}(0, \Sigma_a)$. Proposer un estimateur de a basé sur la formule des moments et déterminer sa loi.
4. Déterminer l'estimateur du maximum de vraisemblance de a, et déterminer les propriétés asymptotiques de cet estimateur lorsque $n \to \infty$.

Exercice 2. Soit $X \in \mathbb{R}^3$ un vecteur gaussien centré de matrice de covariance :

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

1. X possède-t-il une densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^3$? Si oui donner son expression.
2. Trouver un opérateur linéaire $A : \mathbb{R}^3 \to \mathbb{R}^3$ tel que les composantes du vecteur $A \cdot X$ soient des variables indépendantes.
3. Déterminer la loi de $X_1 + 2X_2 - X_3$ où $X = (X_1, X_2, X_3)$.

Exercice 3. Soient U_1, U_2, U_3 des variables aléatoires indépendantes de loi $\mathcal{N}(0, 1)$.

1. Donner la loi du vecteur
$$Y = \begin{pmatrix} U_1 - 2U_2 \\ U_1 + U_2 + U_3 \\ U_2 - U_3 \end{pmatrix}$$
2. Montrer que les variables $(U_1 + U_2 + U_3)$ et $U_2 - U_3$ sont indépendantes.
3. Montrer que $\frac{(U_1+U_2+U_3)^2}{3} + \frac{(U_2-U_3)^2}{2}$ suit la loi $\chi(2)$.

Exercice 4. Dans l'atmosphère, le taux d'un gaz nocif, pour un volume donné, suit une loi normale d'espérance μ et de variance σ^2 inconnues. On effectue n prélèvements conduisant aux valeurs $x_1, x_2, \cdots, x_n$.

1. Sur un échantillon de taille $n = 10$, on observe que $\overline{x}_{10} = 50$ et ${s'}^2_{10} = 100$ où ${s'}^2_{10} = \sum_{i=1}^{10}(x_i - \overline{x}_{10})^2/9$. Quel est l'intervalle de confiance de niveau de confiance 95% du taux moyen μ de gaz dans l'atmosphère ?
2. Quel serait cet intervalle si la variance σ^2 était connue (on la prendra égale à 100) ?
3. On dispose maintenant d'un échantillon 8 fois plus grand conduisant aux résultats suivants : $\overline{x}_{81} = 48$ et ${s'}^2_{81} = 90$. Quel est alors l'intervalle de confiance de niveau de confiance 90% pour μ ?

Exercice 5. On dispose de dix prises de sang recueillies dans les mêmes conditions sur un même sujet. On obtient pour chacune un dosage du cholestérol (en grammes) :

245	248	250	247	249	247	247	246	246	248

Chaque mesure peut être considérée comme une réalisation particulière d'une variable X "taux de cholestérol" suivant une loi normale d'espérance μ et de variance σ^2. On supposera les mesures indépendantes.

1. Déterminer un intervalle de confiance pour μ de niveau de confiance 98%.
2. On admet que la variance de X n'est liée qu'à celle de la méthode de dosage, supposée parfaitement connue ($\sigma^2 = 1,5$). Donner alors un intervalle de confiance pour μ de niveau de confiance 98%.

Exercice 6. On considère un instrument de mesure des longueurs. On suppose que l'erreur de mesure peut être modélisée par une loi normale centrée autour de la vraie longueur, d'écart type 1mm.

1. Calculer la probabilité que l'erreur faite quand on utilise une fois l'instrument soit inférieure à 0.5 mm.
2. Afin d'affiner notre mesure, on suppose que l'instrument est utilisé 9 fois de façon indépendante pour mesurer une longueur donnée. Calculer la probabilité que la moyenne de ces 9 mesures soit au plus distante de 0.5 mm de la vraie longueur.
3. Un autre instrument a été utilisé dix fois de façon indépendante pour mesurer une longueur donnée. On a obtenu les mesures suivantes

19.2	20.1	19.7	20.1	20.2	19.7	19.9	20.1	20.2	20.3

   En supposant que ces mesures sont gaussiennes, calculer un intervalle de confiance de niveau de confiance 98% de la longueur mesurée en deux façons : d'abord en supposant que l'écart type de l'erreur vaut 1mm ensuite en supposant l'écart type inconnu.

Exercice 7. On essaie de déterminer le résultat d'une élection de deux candidats, A et B. Pour ce faire, on a accès aux estimations de trois instituts de sondages, qui prévoient que le candidat A recevra une proportion $p_1 = 46\%$, $p_2 = 51\%$ ou $p_3 = 44\%$ du vote. On déterminera un estimateur de p et on calculera la probabilité que le candidat A gagne l'élection dans les cas suivants :

1. On suppose tout d'abord, au vu de leurs résultats historiques, que les trois instituts de sondages fournissent des résultats distribués selon des gaussiennes centrées autour de la vraie proportion que recevra le candidat A, et de variance $\sigma^2 = 0.05^2$.
2. On suppose maintenant que l'institut 3 est de qualité plus faible que ses compétiteurs, et donne un résultat bruité par une gaussienne de variance $\sigma^2 = 0.1^2$.
3. On ajoute maintenant un "*house effect*" : les instituts ont un biais en faveur du candidat A, ajoutant à son score en moyenne $m_1 = 3\%$, $m_2 = -2\%$ et $m_3 = -6\%$.
4. On considère enfin un effet de "*herding*" : les instituts ont tendance à observer les résultats de leurs concurrent pour améliorer leur propre estimation. On supposera que p_3 est indépendant de p_1 et p_2, mais que p_1 et p_2 ont une corrélation de $0,5$. Décrivez précisément la loi du vecteur (p_1, p_2, p_3) sous la loi $\mathbb{P}_p$ dans ces conditions.