M1 Maths 2014-2015 Université Lille 1
Probabilités M407 Mylene.Maida@math.univ-lille1.fr

Devoir à la maison n°3
Autour des vecteurs gaussiens
À rendre le jeudi 20 novembre 2014

Exercice 1. Définition, caractérisations et premières propriétés des vecteurs gaussiens

Soient $\mathbf{m} := (m_1, \dots, m_d)^t$ un vecteur de $\mathbb{R}^d$ et $\mathbf{C} := (C_{jk})_{j,k=1,\dots,d}$ une matrice de taille $d \times d$ symétrique positive. Un **vecteur gaussien de moyenne m et de matrice de covariance C** est un vecteur aléatoire $\mathbf{X}$ à valeurs dans $\mathbb{R}^d$, dont la fonction caractéristique est donnée par

$$\varphi_{\mathbf{X}}(t_1, \dots, t_d) := \mathbb{E}(e^{i\langle \mathbf{t}, \mathbf{X}\rangle}) = \exp\left(i\sum_{j=1}^d m_j t_j - \frac{1}{2}\sum_{j,k=1}^d C_{jk} t_j t_k\right), \quad \forall \mathbf{t} = (t_1, \dots, t_d) \in \mathbb{R}^d.$$

On note alors $\mathbf{X} \sim \mathcal{N}(\mathbf{m}, \mathbf{C})$. Si $\mathbf{m}$ est le vecteur nul, on dit que le vecteur est gaussien centré ; si $\mathbf{C}$ est la matrice identité de taille $d \times d$, on dit que le vecteur est gaussien réduit.

1. Si $\mathbf{X} \sim \mathcal{N}(\mathbf{m}, \mathbf{C})$, vérifier que son espérance est le vecteur $\mathbf{m}$ et que sa matrice de covariance est $\mathbf{C}$.
2. Soit C une matrice symétrique positive. Montrer qu'il existe une matrice A symétrique positive telle que $A^2 = C$. La matrice A est-elle unique ?
3. Soit $\mathbf{Y} = (Y^1, \dots, Y^d)$ un vecteur aléatoire dont les coordonnées sont des variables aléatoires réelles indépendantes, toutes de loi $\mathcal{N}(0, 1)$.
   (a) Quelle est la loi de $\mathbf{Y}$?
   (b) Soit $\mathbf{R}$ une matrice de taille $d \times d$. Quelle est la loi du vecteur aléatoire $\mathbf{RY}$?
4. Soit $\mathbf{C}$ une matrice symétrique positive et $\mathbf{m}$ un vecteur de taille compatible. Déduire des questions précédentes qu'il existe un vecteur aléatoire de loi $\mathcal{N}(\mathbf{0}, \mathbf{C})$, puis qu'il existe un vecteur aléatoire de loi $\mathcal{N}(\mathbf{m}, \mathbf{C})$.
5. Montrer qu'un vecteur aléatoire $\mathbf{X}$ est un vecteur gaussien centré si et seulement si toute combinaison linéaire de ses coordonnées est une variable aléatoire gaussienne centrée.
6. Supposons que $\mathbf{C}$ soit de taille $d \times d$ symétrique définie positive (en particulier elle est inversible) et $\mathbf{m}$ un vecteur de $\mathbb{R}^d$.
   (a) Soit $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$. Montrer que la densité de la loi de $\mathbf{X}$ est donnée par, $\forall \mathbf{x} \in \mathbb{R}^d$,
   $$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}\sqrt{\det(\mathbf{C})}} \exp\left(-\frac{1}{2}\langle \mathbf{x}, \mathbf{C}^{-1}\mathbf{x}\rangle\right).$$
   (b) Quelle est la densité de la loi de $Z \sim \mathcal{N}(m, C)$?
7. Que se passe-t-il si $\mathbf{C}$ n'est pas inversible ?

Exercice 2. Théorème de Cochran et application en statistiques

Partie A. Préliminaires sur les lois du khi-deux et de Student.
Pour $p \in \mathbb{N}^*$, soient $p+1$ variables aléatoires $N_1, \dots, N_{p+1}$ de loi $\mathcal{N}(0, 1)$ indépendantes. La loi de la variable aléatoire $Z := \sum_{i=1}^p N_i^2$ est appelée loi du khi-deux à p degrés de liberté et notée χ_p^2. La loi de la variable aléatoire $T := \frac{\sqrt{p}N_{p+1}}{\sqrt{Z}}$ est appelée loi de Student à p degrés de liberté et notée T_p.

1. Montrer que la densité f_p de la loi χ_p^2 est donnée par
$$f_p(x) = \frac{1}{2^{p/2}\Gamma(p/2)} x^{p/2-1} e^{-x/2} \mathbf{1}_{\mathbb{R}_+^*}(x), \quad \forall x \in \mathbb{R}.$$
2. Quelle est la densité de la loi T_p ?

Partie B. Théorème de Cochran.
Pour tout $d \in \mathbb{N}^*$, on note $(\mathbf{e}_1, \ldots, \mathbf{e}_d)$ la base canonique de $\mathbb{R}^d$. Pour $p \leq d$, on note $\mathbf{I}_p$ la matrice (dans la base canonique) de la projection sur $(\mathbf{e}_1, \ldots, \mathbf{e}_p)$ (si $p = d$, $\mathbf{I}_d$ est la matrice identité de taille d). Pour tout F sous-espace vectoriel de $\mathbb{R}^d$, $\mathbf{P}_F$ désignera indifféremment l'application projection sur F et la matrice de cette application dans la base canonique. On notera aussi $F^\perp$ l'orthogonal de F pour le produit scalaire euclidien.

Le but de cette deuxième partie est de montrer le théorème de Cochran : Si $\mathbf{X}$ est un vecteur gaussien centré réduit de $\mathbb{R}^d$ et F un sous-espace vectoriel de $\mathbb{R}^d$ de dimension p alors les vecteurs aléatoires $\mathbf{P}_F\mathbf{X}$ et $\mathbf{P}_{F^\perp}\mathbf{X}$ sont indépendants, de lois respectives $\mathbf{P}_F\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{P}_F)$ et $\mathbf{P}_{F^\perp}\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{P}_{F^\perp})$. De plus, les variables aléatoires $\|\mathbf{P}_F\mathbf{X}\|^2$ et $\|\mathbf{P}_{F^\perp}\mathbf{X}\|^2$ sont indépendantes, de lois respectives $\|\mathbf{P}_F\mathbf{X}\|^2 \sim \chi^2_p$ et $\|\mathbf{P}_{F^\perp}\mathbf{X}\|^2 \sim \chi^2_{d-p}$.

1. Soit F un sous-espace vectoriel de $\mathbb{R}^d$ de dimension p. Soit $(\mathbf{u}_1, \ldots, \mathbf{u}_p)$ une base orthonormée de F, $(\mathbf{u}_{p+1}, \ldots, \mathbf{u}_d)$ une base orthonormée de $F^\perp$, et $\mathbf{u} = (\mathbf{u}_1, \ldots, \mathbf{u}_d)$. On note $\mathbf{U}$ la matrice de passage de la base canonique à $\mathbf{u}$.
   (a) Que peut-on dire de la matrice $\mathbf{U}$?
   (b) Exprimer $\mathbf{P}_F$ et $\mathbf{P}_{F^\perp}$ en fonction des matrices $\mathbf{U}$, $\mathbf{I}_d$ et $\mathbf{I}_p$.
2. Soit $\mathbf{X}$ un vecteur gaussien centré réduit. Quelle est la loi de $\mathbf{Y} = \mathbf{U}\mathbf{X}$?
3. Montrer que $\mathbf{I}_p\mathbf{Y}$ et $(\mathbf{I}_d - \mathbf{I}_p)\mathbf{Y}$ sont indépendants et déterminer leur loi.
4. Déterminer la loi jointe de $(\|\mathbf{I}_p\mathbf{Y}\|^2, \|(\mathbf{I}_d - \mathbf{I}_p)\mathbf{Y}\|^2)$.
5. En déduire le théorème de Cochran.

Partie C. Application en statistiques.
Soit $(X_1, \ldots, X_n)$ des variables aléatoires réelles. On note $\overline{X_n} := \frac{1}{n}\sum_{i=1}^n X_i$ leur moyenne empirique et $S_n^2 := \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X_n})^2$ leur variance empirique.

1. On suppose que $(X_1, \ldots, X_n)$ sont des variables aléatoires réelles indépendantes toutes de loi $\mathcal{N}(0,1)$.
   (a) En appliquant le théorème de Cochran pour un sous-espace F bien choisi, montrer que $\overline{X_n}$ et S_n^2 sont indépendantes et déterminer leur loi.
   (b) En déduire la loi de $\sqrt{n}\frac{\overline{X_n}}{S_n}$.
2. Reprendre la question **C**.1. en supposant maintenant que $(X_1, \ldots, X_n)$ sont des variables aléatoires réelles indépendantes toutes de loi $\mathcal{N}(\mu, \sigma^2)$, avec $\mu, \sigma \in \mathbb{R}$.

Exercice 3. Conditionnement gaussien
Cet exercice peut être traité soit en s'aidant de l'exercice 2, soit complètement indépendamment.

1. Soit $(X_1, \ldots, X_m, Y_1, \ldots, Y_n)$ un vecteur gaussien centré. Montrer que les vecteurs $(X_1, \ldots, X_m)$ et $(Y_1, \ldots, Y_n)$ sont indépendants si et seulement si, pour tout couple (i,j) avec $i \in \{1, \ldots, m\}$ et $j \in \{1, \ldots, n\}$, $\mathrm{cov}(\mathrm{X_i}, \mathrm{Y_j}) = 0$.
2. Soit $(Y_1, \ldots, Y_n, X)$ un vecteur gaussien centré. Soit $\hat{X}$ la projection orthogonale de X sur l'espace vectoriel engendré par $Y_1, \ldots, Y_n$. Montrer que $X - \hat{X}$ est indépendant de $Y_1, \ldots, Y_n$ et en déduire que
$$\mathbb{E}_{\{Y_1, \ldots, Y_n\}}[X] = \hat{X}.$$
3. Soit (X, Y, Z) un vecteur gaussien centré de matrice de covariance $\begin{pmatrix} 1 & 0 & -1 \\ 0 & 5 & 3 \\ -1 & 3 & 4 \end{pmatrix}$. Calculer $\mathbb{E}_{\{X,Z\}}(Y)$.
4. Soit (X, Y) un vecteur gaussien centré de matrice de covariance $\begin{pmatrix} 4/3 & -1 \\ -1 & 1 \end{pmatrix}$.
   (a) Calculer $\mathbb{E}_{X-Y}[X]$.
   (b) Pour $t \in \mathbb{R}$, calculer $\mathbb{E}_Y(e^{itX})$ et en déduire la loi de X conditionnellement à $Y = y$.