## 4. Voyage en pays gaussien

*Objectifs : Se familiariser avec les vecteurs gaussiens, leurs propriétés et les calculs associés. Les exercices 4.1 à 4.3 sont à faire pendant le TD, le 4.4 est à chercher de votre côté.*

**Exercice 4.1** (Pour commencer)**.** Soit $Z = (X, Y)$ un vecteur aléatoire sur $\mathbb{R}^2$ ayant une densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^2$ donnée par

$$f(x, y) = C \exp(-x^2 + xy - y^2/2)\,,$$

où $C$ est une constante de normalisation que l'on ne demande pas de calculer.

1. Montrer que $Z$ est un vecteur gaussien dont on précisera l'espérance et la matrice de covariance. *On pourra commencer par chercher $\mu \in \mathbb{R}^2$ et $M \in \mathbb{R}^{2\times 2}$ symétrique telle que $-x^2 + xy - y^2/2 = \frac{1}{2}(z-\mu)^T M(z-\mu)$ pour $z = (x, y)$.* *Solution. La densité s'écrit sous la forme $C\exp(-\frac{1}{2}(x\ y)^T \Sigma^{-1}(x\ y))$ avec $\Sigma^{-1} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$, donc $Z$ est un vecteur gaussien de moyenne nulle et de covariance $\Sigma = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} \mathrm{Var}(X) & \mathrm{cov}(X,Y) \\ \mathrm{cov}(X,Y) & \mathrm{Var}(Y) \end{pmatrix}$.*
2. Quelle est la loi marginale de $X$ ? de $Y$ ? de $2X - Y + 5$ ? *Solution. On lit directement grâce à la linéarité que ce sont des gaussiennes (unidimensionnelles) avec $X \sim \mathcal{N}(0,1), Y \sim \mathcal{N}(0,2)$. Pour $2X - Y + 5$, l'espérance vaut 5, et on utilise la formule de transfo linéaire, ou on calcule directement la variance : $\mathrm{Var}(2X - Y + 5) = 4\mathrm{Var}(X) + \mathrm{Var}(Y) - 2 \times 2\mathrm{cov}(X,Y) = 4 + 2 - 4 = 2$. D'où $2X - Y + 5 \sim \mathcal{N}(5, 2)$.*
3. Montrer que $X$ et $Y - X$ sont indépendants. *Solution. On a $\mathrm{cov}(X, Y - X) = 1 - 1 = 0$ : on conclut à l'indépendance.*
4. Soit $B = \frac{\sqrt{5}}{5}\begin{pmatrix} 3 & -1 \\ -1 & 2 \end{pmatrix}$. Vérifier que $B^2 = \Sigma^{-1}$. Quelle est la loi du vecteur $BZ$ ? *Solution. On utilise la linéarité, on a que $BZ \sim \mathcal{N}(0, B\Sigma B^T) = \mathcal{N}(0, I_2)$ car $\Sigma = B^{-1}(B^T)^{-1}$ ($B$ est symétrique). Faire le dessin avec l'isotropie, avec le cas 1D et le cas multiD. Leur demander ce que sont les directions de l'ellipsoide. (LG: à expliciter pour les chargés de TD)*

**Exercice 4.2** (Un contre-exemple)**.** Soit $X$ variable réelle gaussienne $\mathcal{N}(0,1)$, et $S$ une variable aléatoire de Rademacher, i.e. telle que $\mathbb{P}(S = 1) = \mathbb{P}(S = -1) = 1/2$. On suppose $S$ et $X$ indépendantes. On définit le vecteur aléatoire $Z = (X, SX)$.

1. Montrer que les coordonnées de $Z$ sont des gaussiennes $\mathcal{N}(0,1)$. *On pourra montrer que la fonction de répartition de $SX$ est encore celle d'une gaussienne standard.* *Solution. On a que $SX$ est encore une gaussienne standard, via la fonction de répartition. Pour tout réel $t$, par la formule des probabilités totales,*

$$\begin{aligned}\mathbb{P}(SX \le t) =& \mathbb{P}(SX \le t \,|\, S = -1)\mathbb{P}(S = -1) + \mathbb{P}(SX \le t \,|\, S = 1)\mathbb{P}(S = 1) \\ =& \frac{1}{2}\mathbb{P}(X \ge -t) + \frac{1}{2}\mathbb{P}(X \le t) \\ =& \frac{1}{2}\mathbb{P}(X \le t) + \frac{1}{2}\mathbb{P}(X \le t) = \mathbb{P}(X \le t),\end{aligned}$$

*où l'on a utilisé le fait que la loi d'une gaussienne standard est symétrique en 0 (i.e. $X = -X$ en loi). $SX$ et $X$ ont mêmes fonction de répartition, donc même loi : $SX \sim \mathcal{N}(0,1)$.*

2. Montrer que pourtant, $Z$ n'est pas un vecteur gaussien. *Solution. Pour montrer que $Z$ n'est pas un vecteur gaussien, il y a plusieurs façons de faire. Solution 1. On remarque*

*que* $\mathrm{Cov}(X, SX) = \mathbb{E}[SX^2] - 0 = 0$*. Par l'absurde, si* $Z$ *est un vecteur gaussien, alors* $X$ *et* $SX$ *sont indépendantes. Or, par la formule des probabilités totales,*

$$\begin{aligned}&\mathbb{P}(SX \geq 10 \,|\, X \geq 10)\\ &\qquad = \frac{1}{2}\mathbb{P}(SX \geq 10 \,|\, X \geq 10, S = 1) + \frac{1}{2}\mathbb{P}(SX \geq 10 \,|\, X \geq 10, S = -1) = \frac{1}{2},\end{aligned}$$

*qui n'est vraiment pas égal à* $\mathbb{P}(SX \geq 10)$ *où* $SX \sim \mathcal{N}(0, 1)$*.* *Solution 2. Regardons la somme des deux coordonnées,* $X + SX$*. On a que*

$$\begin{aligned}\mathbb{P}(SX + X = 0) =& \mathbb{P}(SX + X = 0 \,|\, S = -1)\mathbb{P}(S = -1) + \mathbb{P}(SX + X = 0 \,|\, S = 1)\mathbb{P}(S = 1)\\ =& \frac{1}{2} + \frac{1}{2}\mathbb{P}(X = 0) = \frac{1}{2}.\end{aligned}$$

*Si* $X + SX$ *est gaussienne, alors elle n'admet pas de densité par rapport à Lebesgue puisque elle a un atome en* $0$*. Mais dans ce cas elle est nulle ps, contradiction avec* $\mathbb{P}(SX + X = 0) = 1/2 \neq 1$*. Conclusion (importante) : des gaussiennes ne forment pas forcément un vecteur gaussien. Bonus : Mais elles en forment toujours un quand elles sont indépendantes. On pourra juste démontrer qu'une somme de deux gaussiennes indépendantes forme une gaussienne.*

**Exercice 4.3** (Indépendance entre moyenne et variance empiriques dans le modèle gaussien)**.** On considère un échantillon $X = (X_1, \ldots, X_n)$ i.i.d. de loi $\mathcal{N}(\mu, \sigma^2)$ avec $\mu$ et $\sigma^2$ inconnus.

1. On s'intéresse à la variance empirique de l'échantillon définie par $S_n := \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2$. En utilisant le théorème de Cochran, déterminer la loi de $S_n$ et montrer que $S_n$ et $\overline{X}$ sont indépendants. On pourra introduire le vecteur $e$ de $\mathbb{R}^n$ dont toutes les coordonnées valent 1. *Solution. On voit tout ça comme une norme d'un projecteur. Regardons le projecteur orthogonal sur* $E_1 = \mathrm{Vect}(e) = \mathbb{R}e$*.* $e/\sqrt{n}$ *est une base orthonormée triviale de* $E_1$ *et pour* $x \in \mathbb{R}^d$*. On a* $\Pi_1 X = \langle X, e/\sqrt{n}\rangle e/\sqrt{n} = \overline{X}e$*. Prenons* $E_2$ *l'orthogonal de* $E_1$ *dans* $\mathbb{R}^n$*, on a* $\Pi_2 = I_n - \Pi_1$ *et donc* $\|\Pi_2(X - \mu e)\|^2 = \|X - \mu e - \overline{(X - \mu)}e\|^2 = \|X - \overline{X}e\|^2$ *et* $S_n = \frac{1}{n}\|\Pi_2(X - \mu e)\|^2$*. Cochran nous dit que* $S_n$ *est donc indépendant de* $\overline{X}$ *et que* $S_n \sim \frac{\sigma^2}{n}\chi^2(n-1)$*. On a d'ailleurs une moyenne pas tout à fait égale à* $\sigma^2$*, et une variance qui tend sans surprise vers* $0$*.*
2. On s'intéresse à la variance empirique débiaisée, définie par $S'_n := \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$. Quelle est la loi de $\sqrt{n}\frac{\overline{X} - \mu}{\sqrt{S'_n}}$ ? *Solution. application directe du cours, après avoir sorti le* $\sigma$ *en haut et en bas : loi de Student à* $n-1$ *degrés de liberté.*

**Exercice 4.4** (Test des variances dans le modèle gaussien)**.** Nous considérons deux échantillons gaussiens indépendants. Le premier, $X^{(1)}$, est de taille $n_1$ et i.i.d. de loi $\mathcal{N}(\mu_1, \sigma_1^2)$. Le second, $X^{(2)}$, est de taille $n_2$ et i.i.d. de loi $\mathcal{N}(\mu_2, \sigma_2^2)$. Tous les paramètres sont supposés inconnus. On note $S'_1$ (resp. $S'_2$) la variance empirique débiaisée pour l'échantillon $X^{(1)}$ (resp. $X^{(2)}$). On veut tester

$$\mathcal{H}_0 : \sigma_1^2 = \sigma_2^2 \quad \text{contre} \quad \mathcal{H}_1 : \sigma_1^2 \neq \sigma_2^2\,.$$

1. Sous l'hypothèse nulle, quelle est la loi de $\frac{S'_1}{S'_2}$ ? *Solution. application directe du cours, on a deux variables indépendantes avec* $\frac{S'_1}{S'_2} = \frac{\sigma_1^2\chi^2(n_1-1)/(n_1-1)}{\sigma_1^2\chi^2(n_2-1)/(n_2-1)} = \mathcal{F}(n_1 - 1, n_2 - 1)$ *(égalité en loi).*
2. En déduire un test de niveau $\alpha \in ]0, 1[$ répondant au problème. *Solution. Au vu de la question précédente, on va se baser sur la statistique de test* $T = \frac{S'_1}{S'_2}$*. Vu la forme de* $H_1$*,* $T$ *sous l'hypothèse alternative* $H_1$ *sera soit plus grande soit plus petite, typiquement. On va donc prendre ici une région de rejet bilatère de la forme* $\mathcal{R} = \{T < c_1\} \cup \{T > c_2\}$*,*

*et on calibre $c_1, c_2 > 0$ de sorte que $\mathbb{P}_0(\mathcal{R}) = \alpha$. Si l'on répartit cette probabilité équitablement entre les évènements $\{T < c_1\}$ et $\{T > c_2\}$, on prend $c_1 = f_{\alpha/2}^{n_1-1,n_2,2}$ et $c_2 = f_{1-\alpha/2}^{n_1-1,n_2,2}$, avec $f_\beta^{d_1,d_2}$ le quantile d'ordre $\beta$ d'une loi de Fisher $\mathcal{F}(d_1, d_2)$.*

3. Exprimer sa puissance en fonction de $\sigma_1^2, \sigma_2^2$ et une certaine fonction de répartition.
*Solution.* *$H_1$ étant composite, la puissance dépend des paramètres $\sigma_1^2 \neq \sigma_2^2$ sous $H_1$. Sous $H_1$, étant donné $\sigma_1^2 \neq \sigma_2^2$, on a $\frac{S_1'}{S_2'} = \frac{\sigma_1^2\chi^2(n_1-1)/(n_1-1)}{\sigma_2^2\chi^2(n_2-1)/(n_2-1)} = \frac{\sigma_1^2}{\sigma_2^2}\mathcal{F}(n_1-1, n_2-1)$ (égalité en loi). En notant $X$ une variable de loi $\mathcal{F}(n_1-1, n_2-1)$ et $F$ sa fonction de répartition, la puissance s'écrit*

$$\begin{aligned}\mathbb{P}_{\sigma_1^2,\sigma_2^2}(\mathcal{R}) &= \mathbb{P}\left(\frac{\sigma_1^2}{\sigma_2^2}X < c_1\right) + \mathbb{P}\left(\frac{\sigma_1^2}{\sigma_2^2}X > c_2\right) \\ &= F\left(f_{\alpha/2}^{n_1-1,n_2,2}\frac{\sigma_2^2}{\sigma_1^2}\right) + 1 - F\left(f_{1-\alpha/2}^{n_1-1,n_2,2}\frac{\sigma_2^2}{\sigma_1^2}\right).\end{aligned}$$

*(on retrouve que cette puissance vaut $\alpha$ si $\sigma_1^2 = \sigma_2^2$).*