

4. Vecteurs gaussiens

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 $-x^2 + xy - y^2/2 = \frac{1}{2}(X - \mu)^T M (X - \mu)$ pour $X = (x, y)$. [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 $\Sigma = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} \text{Var}(X) & \text{cov}(X, Y) \\ \text{cov}(X, Y) & \text{Var}(Y) \end{pmatrix}$.*
2. Quelle est la loi marginale de X ? de Y ? de $2X - Y$? [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 $2X - Y$ on utilise la formule de transfo linéaire, ou on calcule directement la variance : $\text{Var}(2X - Y) = 4\text{Var}(X) + \text{Var}(Y) - 4\text{cov}(X, Y) = 4 + 2 - 4 = 2$.
3. Montrer que X et $Y - X$ sont indépendants. [Solution.](#) On a $\text{cov}(X, Y - X) = 1 - 1 = 0$: 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'ellipsoïde.

Exercice 4.2 (Un contre-exemple). Soit X variable réelle gaussienne standard, 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 $Z = (X, SX)$.

Montrer que les coordonnées de Z sont des gaussiennes $\mathcal{N}(0, 1)$, mais que Z n'est pas un vecteur gaussien. *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,*

$$\begin{aligned} \mathbb{P}(SX \leq t) &= \mathbb{P}(SX \leq t \mid S = -1)\mathbb{P}(S = -1) + \mathbb{P}(SX \leq t \mid S = 1)\mathbb{P}(S = 1) \\ &= \frac{1}{2}\mathbb{P}(X \geq -t) + \frac{1}{2}\mathbb{P}(X \leq t) \\ &= \frac{1}{2}\mathbb{P}(X \leq t) + \frac{1}{2}\mathbb{P}(X \leq t) = \mathbb{P}(X \leq t), \end{aligned}$$

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 : $SX \sim \mathcal{N}(0, 1)$. Regardons la somme des deux coordonnées, $X + SX$. On a que

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

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 : 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 $X = (X_1, \dots, X_n)$ i.i.d. de loi $\mathcal{N}(\mu, \sigma^2)$ avec μ et σ^2 inconnus.

- On s'intéresse à la variance empirique de l'échantillon définie par $S_n := \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$. En utilisant le théorème de Cochran, déterminer la loi de S_n et montrer que S_n et \bar{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 $E_1 = \text{Vect}(e) = \mathbb{R}e$. e/\sqrt{n} est une base orthonormée triviale de E_1 et pour $x \in \mathbb{R}^d$. On a $\Pi_1 X = \langle X, e/\sqrt{n} \rangle e/\sqrt{n} = \bar{X}e$. Prenons E_2 l'orthogonal de E_1 dans \mathbb{R}^n , on a $\Pi_2 = I_n - \Pi_1$ et donc $\|\Pi_2(X - \mu e)\|^2 = \|X - \mu - (\bar{X} - \mu)_n e\|^2 = \|X - \bar{X}e\|^2$ et $S_n = \frac{1}{n} \|\Pi_2(X - \mu e)\|^2$. Cochran nous dit que S_n est donc indépendants de \bar{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 σ^2 , et une variance qui tend sans surprise vers 0.
- On s'intéresse à la variance empirique débiaisée, définie par $S'_n := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$. Quelle est la loi de $\sqrt{n} \frac{\bar{X} - \mu}{\sqrt{S'_n}}$? *Solution.* application directe du cours, après avoir sorti le σ 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 $\mathcal{N}(\mu_2, \sigma_2^2)$. On note S'_1 (resp. S'_2) la variance empirique débiaisée pour l'échantillon $X^{(1)}$ (resp. $X^{(2)}$). On veut tester

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

- 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 $\frac{S'_1}{S'_2} = \frac{\sigma_1^2 \chi^2(n_1)/n_1}{\sigma_2^2 \chi^2(n_2)/n_2} = \mathcal{F}(n_1, n_2)$ (égalité en loi).
- En déduire un test de niveau $\alpha \in]0, 1[$ répondant au problème.
- Exprimer sa puissance en fonction de σ_1^2, σ_2^2 et une certaine fonction de répartition.