M1 – Statistique – 2020/2021
vincent.brault@univ-grenoble-alpes.fr

TD n°2 : Vecteurs gaussiens et introduction à la statistique

1 Exercice en rapport avec le cours

Démonstrations

1. Montrer que les moments d'ordre k d'une variable aléatoire X suivant une loi gaussienne centrée réduite sont :

$$\forall k \in \mathbb{N},\ \mathbb{E}\left[X^k\right] = \begin{cases} \frac{(2m)!}{2^m m!} & \text{si } k = 2m, \\ 0 & \text{sinon.} \end{cases}$$

2. Étant donnée $\mathbf{X} = (X_1, \ldots, X_d)$ un vecteur gaussien d'espérance m et de matrice de variance covariance Σ montrer qu'alors, pour toute matrice $A \in \mathcal{M}_{k\times d}(\mathbb{R})$ et pour tout vecteur $b \in \mathbb{R}^k$, le vecteur $A\mathbf{X} + b$ est gaussien de loi $\mathcal{N}(Am + b, A\Sigma A^T)$.

Contre-exemple

Soit Z une variable gaussienne centrée réduite et X une variable de Bernoulli de paramètre $p \in]0;1[$ sur $\{-1, 1\}$ indépendante de Z.

1. Montrer que la variable XZ est gaussienne.

2. Montrer que le vecteur $(Z, XZ)^T$ n'est pas gaussien.

3. Calculer la covariance de Z et XZ.

4. Montrer que Z et XZ ne sont pas indépendantes.

2 Application

Cet exercice est tiré du livre **?**. Étant donné $(X, Y, Z)^T$ un vecteur gaussien de moyenne $m = (1, -1, 1)^T$ et de matrice de variance covariance

$$K = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}.$$

Nous posons :

$$\begin{aligned} U &= -X + Y + Z, \\ V &= X - Y + Z, \\ W &= X + Y - Z. \end{aligned}$$

1. Quelles sont les lois de U, V et W ?

2. Déterminer l'espérance et la matrice de variance covariance du vecteur $(U, V, W)^T$. En déduire que les variables U, V et W sont indépendantes.

3. Utiliser ce qui précède pour écrire un algorithme de simulation de la loi $\mathcal{N}(m, K)$ si le logiciel ne permet de simuler que des gaussiennes univariées $\mathcal{N}(0, 1)$.

4. Quelle est la loi de $(U^2 + V^2 + W^2)/4$?

3 Introduction à la statistique

3.1 Espace des paramètres

Pour les lois suivantes, donner un (ou plusieurs) espace de paramètre Θ de telle sorte que les lois correspondent à des modèles identifiables :

Nom	Abréviation	Support	Densité $f(x)$ ou $\mathbb{P}(X = x)$
Bernoulli	$\mathcal{B}(p)$	$\{0, 1\}$	$p^x(1-p)^{1-x}$
Bêta	$\mathcal{B}e(a, b)$	$]0, 1[$	$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$
Binomiale	$\mathcal{B}in(n; p)$	$\{0, 1, \ldots, n\}$	$\binom{n}{x} p^x(1-p)^{n-x}$
Binomiale négative	$\mathcal{NB}in(r; p)$	$\mathbb{N}$	$\binom{x+n-1}{x} p^n(1-p)^x$
Cauchy	$\mathcal{C}au(\mu, a)$	$\mathbb{R}$	$\frac{\pi}{1+\left(\frac{x-\mu}{a}\right)^2}$
Exponentielle	$\mathcal{E}(\lambda)$	$\mathbb{R}^+$	$\lambda e^{-\lambda x}$
Gamma	$\mathcal{G}a(\lambda)$	$\mathbb{R}^\star_+$	$\frac{\beta^\alpha}{\Gamma\alpha} x^{\alpha-1} e^{-\beta x}$
Gaussienne ou normale	$\mathcal{N}(\mu, \sigma^2)$	$\mathbb{R}$	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$
Géométrique	$\mathcal{G}eo(p)$	$\mathbb{N}^\star$	$(1-p)^{x-1}p$
Hypergéométrique	$\mathcal{H}(n, p, A)$	$\{0, 1, \ldots, n\}$	$\frac{\binom{pA}{x}\binom{(1-p)A}{n-x}}{\binom{A}{x}}$
Inverse gamma	$\mathcal{IG}a(\lambda)$	$\mathbb{R}^\star_+$	$\frac{\beta^\alpha}{\Gamma\alpha} x^{-\alpha-1} e^{\frac{-\beta}{x}}$
Laplace	$\text{Laplace}(\mu, b)$	$\mathbb{R}$	$\frac{1}{2b} e^{-\frac{\lvert x-\mu \rvert}{b}}$
Pareto	$\mathcal{P}a(a, k)$	$[a, +\infty[$	$\frac{ka^k}{x^{k+1}}$
Poisson	$\mathcal{P}(\lambda)$	$\mathbb{N}$	$e^{-\lambda}\frac{\lambda^x}{x!}$
Uniforme	$\mathcal{U}(\{0, 1, \ldots, n\})$	$\{0, 1, \ldots, n\}$	$\frac{1}{n+1}$
Uniforme	$\mathcal{U}([a, b])$	$[a, b]$	$\frac{1}{b-a}$

3.2 Modèle linéaire

Dans cet exercice, nous étudions des modèles avec un bruit gaussien mais une moyenne évoluant avec le temps. Pour cela, nous supposons qu'il existe une suite de réels non nuls $(t_n)_{n \in \mathbb{N}^\star}$ connue, un paramètre $a^\star$ inconnu et un paramètre $\sigma^\star > 0$ inconnu tels que pour tout i de $\{1, \ldots, n\}$, nous ayons :

$$Y_i = a^\star t_i + \varepsilon_i \text{ avec } \varepsilon_i \overset{iid}{\sim} \mathcal{N}\left(0, \sigma^{\star 2}\right). \tag{1}$$

Donner les paramètres à estimer dans le modèle (2). Donner la densité en fonction du paramètre connu t et montrer que le modèle est identifiable si $t \neq 0$.