I-1.2 Les outils mathématiques

Cette partie rassemble les outils utilisés dans le chapitre et dans la feuille d'exercices. Les éléments de théorie de la mesure sont développés dans l'annexe du polycopié.

Tribu produit, mesure produit

Soient $(\mathsf{X}, \mathcal{X})$ et $(\mathsf{X}', \mathcal{X}')$ deux espaces mesurables. La *tribu produit* $\mathcal{X} \otimes \mathcal{X}'$ sur $\mathsf{X} \times \mathsf{X}'$ est la plus petite tribu contenant les pavés mesurables $A \times A'$, $A \in \mathcal{X}$, $A' \in \mathcal{X}'$ (voir **??**). Pour deux probabilités $\mathbb{Q}$ sur $\mathcal{X}$ et $\mathbb{Q}'$ sur $\mathcal{X}'$, la *mesure produit* $\mathbb{Q} \otimes \mathbb{Q}'$ (voir **??**) est définie par

$$\mathbb{Q} \otimes \mathbb{Q}'(A \times A') := \mathbb{Q}(A)\, \mathbb{Q}'(A') \ .$$

On note $\mathbb{Q}^{\otimes n}$ le produit de n copies de $\mathbb{Q}$.

Produit de deux modèles statistiques

Soient $(\mathsf{X}, \mathcal{X}, \mathcal{C})$ et $(\mathsf{X}', \mathcal{X}', \mathcal{C}')$ deux modèles statistiques. Un élément z de l'espace des observations $\mathsf{Z} := \mathsf{X} \times \mathsf{X}'$ est un couple $z := (x, x')$ avec $x \in \mathsf{X}$ et $x' \in \mathsf{X}'$. Nous munissons cet espace de la tribu produit $\mathcal{Z} := \mathcal{X} \otimes \mathcal{X}'$.

Définition I-1.7 (Produit de modèles statistiques). *Soient* $(\mathsf{X}, \mathcal{X}, \mathcal{C})$ *et* $(\mathsf{X}', \mathcal{X}', \mathcal{C}')$ *deux modèles statistiques. On appelle* produit de ces deux modèles *et on note* $(\mathsf{X}, \mathcal{X}, \mathcal{C}) \otimes (\mathsf{X}', \mathcal{X}', \mathcal{C}')$ *le modèle statistique* $(\mathsf{X} \times \mathsf{X}',\ \mathcal{X} \otimes \mathcal{X}',\ \mathcal{C} \otimes \mathcal{C}')$ *où :*

$$\mathcal{C} \otimes \mathcal{C}' := \{\mathbb{P} = \mathbb{Q} \otimes \mathbb{Q}' \,,\ \mathbb{Q} \in \mathcal{C},\ \mathbb{Q}' \in \mathcal{C}'\} \ .$$

Soient $X : \mathsf{Z} \to \mathsf{X}$ et $X' : \mathsf{Z} \to \mathsf{X}'$ les statistiques définies pour tout $z = (x, x') \in \mathsf{Z}$ par $X(x, x') = x$ et $X'(x, x') = x'$. Pour tout $\mathbb{P} = \mathbb{Q} \otimes \mathbb{Q}' \in \mathcal{C} \otimes \mathcal{C}'$ et $(A, A') \in \mathcal{X} \times \mathcal{X}'$, nous avons

$$\begin{aligned} \mathbb{P}^{X,X'}(A \times A') &= \mathbb{P}(X \in A, X' \in A') \\ &= \mathbb{Q} \otimes \mathbb{Q}'(A \times A') = \mathbb{Q}(A)\mathbb{Q}'(A') = \mathbb{P}(X \in A)\mathbb{P}(X' \in A') \ . \end{aligned}$$

Les statistiques X et X' sont donc indépendantes. Les modèles statistiques induits par X et X' sont respectivement $(\mathsf{X}, \mathcal{X}, \mathcal{C})$ et $(\mathsf{X}', \mathcal{X}', \mathcal{C}')$. Considérer des produits de modèles statistiques correspond, dans la pratique, à étudier des systèmes d'observations indépendantes.

Remarquons que le n-échantillon $(\mathsf{X}, \mathcal{X}, \mathcal{C})^n$ n'est pas (en général) égal au produit des expériences $(\mathsf{X}, \mathcal{X}, \mathcal{C})$. Pour $n = 2$ par exemple,

$$(\mathsf{X}, \mathcal{X}, \mathcal{C}) \otimes (\mathsf{X}, \mathcal{X}, \mathcal{C}) = (\mathsf{X}^2, \mathcal{X}^{\otimes 2}, \{\mathbb{Q} \otimes \mathbb{Q}' \,,\ \mathbb{Q} \in \mathcal{C}, \mathbb{Q}' \in \mathcal{C}\})$$

alors que

$$(\mathsf{X}, \mathcal{X}, \mathcal{C})^2 = (\mathsf{X}^2, \mathcal{X}^{\otimes 2}, \left\{\mathbb{Q}^{\otimes 2} \,,\ \mathbb{Q} \in \mathcal{C}\right\}) \ .$$

Intuition. *Dans le produit général, les deux coordonnées peuvent suivre des lois* différentes *de la famille ; dans le n-échantillon, c'est* la même *loi qui est répétée. Le n-échantillon est donc une sous-famille (une « diagonale ») du modèle produit.*

Transfert, fonction muette, changement de variables

Pour identifier la loi image $\mathbb{P}^T$, on utilise la *méthode de la fonction muette* : pour toute fonction h mesurable positive sur T,

$$\int_{\mathsf{T}} h(t)\, \mathbb{P}^T(\mathrm{d}t) = \int_{\mathsf{Z}} h \circ T(z)\, \mathbb{P}(\mathrm{d}z) = \mathbb{E}[h(T)] \ .$$

Si l'on trouve une mesure ν et une fonction $q \geq 0$ telles que $\mathbb{E}[h(T)] = \int h\, q\, \mathrm{d}\nu$ pour toute fonction muette h, alors $\mathbb{P}^T = q \cdot \nu$. Quand la statistique est une transformation régulière d'une variable à densité, cette méthode conduit à la formule du changement de variables, que l'on énonce d'abord en dimension 1.

Changement de variable en dimension 1.

Proposition I-1.2.1: Changement de variable, dimension 1

Soient I un intervalle ouvert de $\mathbb{R}$ et X une variable aléatoire réelle de loi $p \cdot \mathrm{Leb}$ avec $\int_I p \,\mathrm{dLeb} = 1$. Soit $\phi : I \to \mathbb{R}$ une application de classe C^1 telle que $\phi'(x) \neq 0$ pour tout $x \in I$. Alors ϕ est strictement monotone, $J := \phi(I)$ est un intervalle ouvert, $\phi^{-1} : J \to I$ est de classe C^1, et la loi de $Y := \phi(X)$ est $q \cdot \mathrm{Leb}$ avec

$$q(y) = \mathbb{1}_J(y)\; p\big(\phi^{-1}(y)\big)\,\big|(\phi^{-1})'(y)\big| = \mathbb{1}_J(y)\;\frac{p\big(\phi^{-1}(y)\big)}{\big|\phi'\big(\phi^{-1}(y)\big)\big|}\,.$$

Démonstration

Soit h mesurable positive sur $\mathbb{R}$. Par le théorème de transfert, $\mathbb{E}[h(\phi(X))] = \int_I h(\phi(x))\,p(x)\,\mathrm{d}x$. La fonction ϕ étant un C^1-difféomorphisme de I sur J, la formule de changement de variable dans l'intégrale de Lebesgue ($x = \phi^{-1}(y)$, $\mathrm{d}x = |(\phi^{-1})'(y)|\,\mathrm{d}y$) donne $\mathbb{E}[h(Y)] = \int_J h(y)\,p(\phi^{-1}(y))\,|(\phi^{-1})'(y)|\,\mathrm{d}y$ pour toute fonction muette h, ce qui identifie la densité de Y. L'égalité $(\phi^{-1})'(y) = 1/\phi'(\phi^{-1}(y))$ vient de la dérivation de $\phi \circ \phi^{-1} = \mathrm{id}_J$. □

Exemple I-1.2.1: Translation et échelle

Si X a pour densité p sur $\mathbb{R}$ et si $\phi(x) = a + b\,x$ avec $b \neq 0$, alors $\phi^{-1}(y) = (y-a)/b$ et $Y = a + bX$ a pour densité

$$q(y) = \frac{1}{|b|}\,p\Big(\frac{y-a}{b}\Big)\,.$$

C'est le calcul de l'Exercice 2 (PC1), coordonnée par coordonnée : pour $X \sim \mathcal{N}(0,1)$ on retrouve la densité de $\mathcal{N}(a, b^2)$.

Exemple I-1.2.2: Une transformation non injective

Si ϕ n'est pas injective, on découpe I en morceaux sur lesquels elle l'est et l'on somme les contributions. Pour $X \sim \mathcal{N}(0,1)$ et $\phi(x) = x^2$, les deux branches $x = \pm\sqrt{y}$ donnent, pour $y > 0$,

$$q(y) = 2 \times \frac{1}{2\sqrt{y}}\,\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-y/2} = \frac{1}{\sqrt{2\pi y}}\,\mathrm{e}^{-y/2}\,,$$

qui est la densité de la loi $\chi^2(1) = \mathrm{Gamma}(1/2, 1/2)$ (Section I-1.2).

Changement de variables en dimension k. On reprend les notations du rappel de l'Exercice 2 (PC1).

Proposition I-1.2.2: Changement de variables, dimension k

Soient un ouvert $\mathcal{O}$ de $\mathbb{R}^k$ et $p \cdot \mathrm{Leb}^{\otimes k}$ une loi sur $\mathbb{R}^k$ telle que $\int_{\mathcal{O}} p \,\mathrm{dLeb}^{\otimes k} = 1$. Soit $\phi : \mathcal{O} \to \mathbb{R}^k$ une application continûment différentiable, injective sur $\mathcal{O}$ et dont le jacobien ne s'annule pas sur $\mathcal{O}$ (de façon équivalente, ϕ est un C^1-difféomorphisme de $\mathcal{O}$ sur $\phi(\mathcal{O})$, c'est-à-dire une bijection de $\mathcal{O}$ sur $\phi(\mathcal{O})$ telle que ϕ et ϕ^{-1} sont de classe C^1). Si $X \sim p \cdot \mathrm{Leb}^{\otimes k}$, alors la loi de $\phi(X)$ est $q \cdot \mathrm{Leb}^{\otimes k}$, où la densité $u \mapsto q(u)$ est donnée par

$$q(u) = \mathbb{1}_{\phi(\mathcal{O})}(u)\; p\big(\phi^{-1}(u)\big)\,\big|\,\mathrm{Det}\,\big(J_{\phi^{-1}}(u)\big)\big|\,,$$

J_ψ désignant, pour toute application différentiable ψ, sa matrice jacobienne (matrice des dérivées partielles). Lorsque ϕ est un C^1-difféomorphisme, la relation $J_\phi(\phi^{-1}(u))\,J_{\phi^{-1}}(u) = \mathrm{I}$

entraîne

$$\mathrm{Det}\left(J_{\phi^{-1}}(u)\right) = \frac{1}{\mathrm{Det}\left(J_{\phi}(\phi^{-1}(u))\right)} .$$

La dimension 1 est le cas $k = 1$, où $J_\phi = \phi'$. La preuve est la même : théorème de transfert, puis formule du changement de variables dans les intégrales multiples, $\int_{\mathcal{O}} h(\phi(x))\, p(x)\, \mathrm{d}x = \int_{\phi(\mathcal{O})} h(u)\, p(\phi^{-1}(u))\, |\mathrm{Det}\, J_{\phi^{-1}}(u)|\, \mathrm{d}u$, pour toute fonction muette h.

Exemple I-1.2.3: Transformation affine de $\mathbb{R}^k$

Soit $\phi(x) = Ax + b$ avec A une matrice $k \times k$ inversible et $b \in \mathbb{R}^k$. Alors $\phi^{-1}(u) = A^{-1}(u - b)$, $J_{\phi^{-1}} = A^{-1}$ et

$$q(u) = \frac{1}{|\mathrm{Det}\, A|}\, p\big(A^{-1}(u - b)\big) .$$

Pour A diagonale, $A = \mathrm{diag}(b_1, \ldots, b_k)$, on retrouve le produit des formules de translation et d'échelle de l'Exercice 2 (PC1) ; pour $X \sim \mathcal{N}(0, \mathrm{I}_k)$, c'est ce calcul qui donne la densité de $\mathcal{N}(b, AA^\top)$ (Section I-1.2).

Lois usuelles utilisées cette semaine

- **Bernoulli** $\mathcal{B}\mathrm{er}(\theta)$ sur $\{0, 1\}$, $\mathbb{P}(\{1\}) = \theta$; **Binomiale** $\mathrm{Bin}(n, \theta)$ sur $\{0, \ldots, n\}$, $\mathbb{P}(\{k\}) = \binom{n}{k}\theta^k(1-\theta)^{n-k}$.
- **Gaussienne** $\mathcal{N}(\mu, \sigma^2)$ sur $\mathbb{R}$, de densité $(2\pi\sigma^2)^{-1/2} \exp(-(x-\mu)^2/(2\sigma^2))$.
- **Exponentielle** de paramètre $\lambda > 0$, de densité $\lambda \mathrm{e}^{-\lambda x} \mathbb{1}_{\mathbb{R}_+}(x)$, de moyenne λ^{-1}. *Absence de mémoire* : pour tous $a, b > 0$, $\mathbb{P}_\lambda\left(X \geq a + b \,|\, X \geq a\right) = \mathbb{P}_\lambda(X \geq b)$.
- **Gamma**. Pour $p > 0$, la loi *Gamma réduite* $\mathrm{Gamma}(p)$ a pour densité $f_p(x) = \frac{1}{\Gamma(p)} \mathrm{e}^{-x} x^{p-1}$, $x > 0$. Pour $\lambda > 0$, $\mathrm{Gamma}(p, \lambda)$ est la loi de Z/λ où $Z \sim \mathrm{Gamma}(p)$, de densité $f_{p,\lambda}(x) = \frac{\lambda^p}{\Gamma(p)} \mathrm{e}^{-\lambda x} x^{p-1}$, $x > 0$. La loi exponentielle de paramètre λ est $\mathrm{Gamma}(1, \lambda)$. *Additivité* : si $X_1, \ldots, X_n$ sont indépendantes de lois $\mathrm{Gamma}(p_i, \lambda)$, alors $\sum_{i=1}^n X_i \sim \mathrm{Gamma}(\sum_{i=1}^n p_i, \lambda)$.
- **Poisson** de paramètre $\lambda > 0$, sur $\mathbb{N}$, de densité $p_\lambda(k) = \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!}$ par rapport à la mesure de comptage. Moyenne et variance λ ; la somme de deux Poisson indépendantes de paramètres λ_1, λ_2 suit une loi de Poisson de paramètre $\lambda_1 + \lambda_2$ (voir l'Exercice 5, PC1).

Vecteurs gaussiens

Le modèle gaussien $\mathcal{N}(\mu \mathbf{1}_n, \sigma^2 \mathrm{I}_n)$ des exemples des tailles et de l'échantillon gaussien, comme le modèle de régression linéaire gaussienne de l'Exercice 4 (PC1), reposent sur la notion de vecteur gaussien. Nous en rappelons la définition et les propriétés utilisées dans ce chapitre, en commençant par le cas réel.

Lois gaussiennes sur $\mathbb{R}$.

Définition I-1.8 (Loi gaussienne réduite). *Une variable aléatoire réelle X est dite* gaussienne centrée réduite*, ce que l'on note $X \sim \mathcal{N}(0, 1)$, si sa loi admet pour densité par rapport à la mesure de Lebesgue sur $\mathbb{R}$*

$$g(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) , \qquad x \in \mathbb{R} .$$

Sa fonction caractéristique vaut

$$\varphi(t) = \mathbb{E}\left[\mathrm{e}^{\mathrm{i}tX}\right] = \exp\left(-\frac{t^2}{2}\right) , \qquad t \in \mathbb{R} .$$

On l'obtient par exemple en remarquant que φ est dérivable (dérivation sous le signe intégral), puis, par une intégration par parties utilisant $xg(x) = -g'(x)$, que $\varphi'(t) = -t\,\varphi(t)$ avec $\varphi(0) = 1$. Comme X admet des moments de tous ordres et que $\varphi^{(k)}(0) = \mathrm{i}^k\,\mathbb{E}[X^k]$, le développement en série entière $\varphi(t) = \sum_{k\geq 0}(-1)^k t^{2k}/(2^k\,k!)$ fournit les moments de la loi gaussienne réduite : les moments d'ordre impair sont nuls et, pour tout $k \in \mathbb{N}$,

$$\mathbb{E}\left[X^{2k}\right] = \frac{(2k)!}{2^k\,k!} = 1 \times 3 \times 5 \times \cdots \times (2k-1)\,.$$

En particulier $\mathbb{E}[X] = 0$, $\mathbb{E}[X^2] = 1$ et $\mathbb{E}[X^4] = 3$.

Définition I-1.9 (Loi gaussienne $\mathcal{N}(\mu, \sigma^2)$). *Soient $\mu \in \mathbb{R}$ et $\sigma^2 \geq 0$. Une variable aléatoire réelle X suit la loi* gaussienne *(ou* normale*) $\mathcal{N}(\mu, \sigma^2)$ si elle a même loi que $\sigma X_r + \mu$, où $X_r \sim \mathcal{N}(0,1)$ et $\sigma = \sqrt{\sigma^2} \geq 0$. On a alors $\mathbb{E}[X] = \mu$ et $\mathrm{Var}(X) = \sigma^2$.*

La convention adoptée dans ce cours est d'autoriser $\sigma^2 = 0$: la loi $\mathcal{N}(\mu, 0)$ est la masse de Dirac δ_μ, et l'on parle de loi gaussienne *dégénérée*. La densité, elle, n'existe que pour $\sigma^2 > 0$: dans ce cas, la formule de changement de variables de la Section I-1.2, appliquée à $x \mapsto \sigma x + \mu$, montre que $\mathcal{N}(\mu, \sigma^2)$ admet pour densité par rapport à Leb

$$g_{\mu,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\,, \qquad x \in \mathbb{R}\,,$$

alors que pour $\sigma^2 = 0$, la loi δ_μ n'est pas absolument continue par rapport à Leb (Section I-1.2). Dans tous les cas ($\sigma^2 \geq 0$), la fonction caractéristique de $\mathcal{N}(\mu, \sigma^2)$ vaut

$$\varphi_{\mu,\sigma^2}(t) = \mathbb{E}\left[\mathrm{e}^{\mathrm{i}t(\sigma X_r + \mu)}\right] = \mathrm{e}^{\mathrm{i}\mu t}\,\varphi(\sigma t) = \exp\left(\mathrm{i}\mu t - \frac{\sigma^2 t^2}{2}\right)\,, \qquad t \in \mathbb{R}\,.$$

Comme la fonction caractéristique caractérise la loi, on en déduit la stabilité de la famille gaussienne par transformation affine et par somme de variables indépendantes.

Lemme I-1.2.1: Stabilité de la famille gaussienne

(i) Si $X \sim \mathcal{N}(\mu, \sigma^2)$ et $a, b \in \mathbb{R}$, alors $aX + b \sim \mathcal{N}(a\mu + b,\ a^2\sigma^2)$.

(ii) Si $X_1, \ldots, X_n$ sont indépendantes, de lois respectives $\mathcal{N}(\mu_1, \sigma_1^2), \ldots, \mathcal{N}(\mu_n, \sigma_n^2)$, alors $X_1 + \cdots + X_n \sim \mathcal{N}(\mu_1 + \cdots + \mu_n,\ \sigma_1^2 + \cdots + \sigma_n^2)$.

Démonstration

Pour (i), $\mathbb{E}[\mathrm{e}^{\mathrm{i}t(aX+b)}] = \mathrm{e}^{\mathrm{i}tb}\,\varphi_{\mu,\sigma^2}(at) = \exp\big(\mathrm{i}(a\mu+b)t - a^2\sigma^2t^2/2\big)$, qui est la fonction caractéristique de $\mathcal{N}(a\mu+b, a^2\sigma^2)$. Pour (ii), par indépendance, $\mathbb{E}[\mathrm{e}^{\mathrm{i}t(X_1+\cdots+X_n)}] = \prod_{j=1}^n \varphi_{\mu_j,\sigma_j^2}(t) = \exp\big(\mathrm{i}t\sum_{j=1}^n \mu_j - t^2\sum_{j=1}^n \sigma_j^2/2\big)$. Dans les deux cas, on conclut par le fait que la fonction caractéristique caractérise la loi. □

Vecteurs gaussiens.

Définition I-1.10 (Vecteur gaussien). *Un vecteur aléatoire $X = (X_1, \ldots, X_n)^\top$ à valeurs dans $\mathbb{R}^n$ est dit* gaussien *si, pour tout $t = (t_1, \ldots, t_n)^\top \in \mathbb{R}^n$, la variable aléatoire réelle $t^\top X = \sum_{j=1}^n t_j X_j$ est gaussienne, éventuellement dégénérée.*

Soit X un vecteur gaussien. Chaque coordonnée $X_i = e_i^\top X$ est gaussienne, donc de carré intégrable ; on peut ainsi définir l'*espérance* $\mu := \mathbb{E}[X] = (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_n])^\top \in \mathbb{R}^n$ et la *matrice de covariance* $\Sigma := \big(\mathrm{Cov}(X_i, X_j)\big)_{1\leq i,j\leq n} = \mathbb{E}\big[(X-\mu)(X-\mu)^\top\big]$ de X. Pour tout $t \in \mathbb{R}^n$, par linéarité de l'espérance,

$$\mathbb{E}\big[t^\top X\big] = t^\top\mu\,, \qquad \mathrm{Var}\left(t^\top X\right) = \sum_{i,j=1}^n t_i t_j\,\mathrm{Cov}(X_i, X_j) = t^\top\Sigma\,t\,,$$

de sorte que $t^\top X \sim \mathcal{N}(t^\top \mu,\ t^\top \Sigma\, t)$. La matrice Σ est symétrique et *semi-définie positive*, puisque $t^\top \Sigma t = \mathrm{Var}(t^\top X) \geq 0$; elle n'est pas nécessairement inversible. On écrit $X \sim \mathcal{N}(\mu, \Sigma)$ pour dire que X est un vecteur gaussien d'espérance μ et de matrice de covariance Σ ; la Proposition I-1.2.3 ci-dessous montre que ces deux paramètres déterminent la loi de X.

D'après le Lemme I-1.2.1 (ii), un vecteur dont les coordonnées sont des variables gaussiennes *indépendantes* est un vecteur gaussien ; en particulier, si $Z_1, \ldots, Z_n$ sont indépendantes de loi $\mathcal{N}(0,1)$, alors $Z = (Z_1, \ldots, Z_n)^\top \sim \mathcal{N}(0, \mathrm{I}_n)$. Sans l'hypothèse d'indépendance, un vecteur dont chaque coordonnée est gaussienne n'est pas nécessairement gaussien (voir la remarque qui suit la Proposition I-1.2.6).

Proposition I-1.2.3: Fonction caractéristique d'un vecteur gaussien

Soient $\mu \in \mathbb{R}^n$ et Σ une matrice symétrique semi-définie positive de taille $n \times n$. Un vecteur aléatoire X à valeurs dans $\mathbb{R}^n$ suit la loi $\mathcal{N}(\mu, \Sigma)$ si et seulement si, pour tout $t \in \mathbb{R}^n$,

$$\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\, t^\top X}\right] = \exp\left(\mathrm{i}\, t^\top \mu - \frac{1}{2}\, t^\top \Sigma\, t\right) .$$

En particulier, la loi d'un vecteur gaussien est entièrement déterminée par son espérance et sa matrice de covariance.

Démonstration

Si $X \sim \mathcal{N}(\mu, \Sigma)$, alors $t^\top X \sim \mathcal{N}(t^\top \mu,\ t^\top \Sigma t)$, et il suffit d'évaluer en 1 la fonction caractéristique de cette loi réelle : $\mathbb{E}[\mathrm{e}^{\mathrm{i}\, t^\top X}] = \varphi_{t^\top \mu,\, t^\top \Sigma t}(1)$. Réciproquement, supposons la formule vérifiée. Pour $t \in \mathbb{R}^n$ et $u \in \mathbb{R}$, en l'appliquant au vecteur ut,

$$\mathbb{E}\left[\mathrm{e}^{\mathrm{i}u\,(t^\top X)}\right] = \exp\left(\mathrm{i}u\, t^\top \mu - \frac{u^2}{2}\, t^\top \Sigma t\right) ,$$

qui est la fonction caractéristique de $\mathcal{N}(t^\top \mu,\ t^\top \Sigma t)$: ainsi $t^\top X \sim \mathcal{N}(t^\top \mu,\ t^\top \Sigma t)$ pour tout t, et X est un vecteur gaussien. En prenant $t = e_i$, on obtient $\mathbb{E}[X_i] = \mu_i$; en notant Σ' la matrice de covariance de X, on a $t^\top \Sigma' t = \mathrm{Var}(t^\top X) = t^\top \Sigma t$ pour tout t, d'où $\Sigma' = \Sigma$ (appliquer l'égalité à $t = e_i$ puis à $t = e_i + e_j$, les deux matrices étant symétriques). Enfin, la fonction caractéristique caractérisant la loi, deux vecteurs gaussiens de mêmes espérance et matrice de covariance ont même loi. □

Proposition I-1.2.4: Transformation affine d'un vecteur gaussien

Soient $X \sim \mathcal{N}(\mu, \Sigma)$ un vecteur gaussien de $\mathbb{R}^n$, A une matrice réelle de taille $m \times n$ et $b \in \mathbb{R}^m$. Alors

$$AX + b \sim \mathcal{N}\big(A\mu + b,\ A\Sigma A^\top\big) .$$

Démonstration

Posons $Y = AX + b$. Pour tout $s \in \mathbb{R}^m$, $s^\top Y = (A^\top s)^\top X + s^\top b$ est une transformation affine de la variable gaussienne $(A^\top s)^\top X$, donc est gaussienne d'après le Lemme I-1.2.1 (i) : Y est un vecteur gaussien. Par linéarité de l'espérance, $\mathbb{E}[Y] = A\mu + b$, et

$$\mathbb{E}\big[(Y - \mathbb{E}[Y])(Y - \mathbb{E}[Y])^\top\big] = \mathbb{E}\big[A(X-\mu)(X-\mu)^\top A^\top\big] = A\,\Sigma\,A^\top .$$

□

On prendra garde à l'ordre des facteurs : la matrice de covariance de $AX + b$ est $A\Sigma A^\top$, de taille $m \times m$, et non $A^\top \Sigma A$.

Proposition I-1.2.5: Représentation d'un vecteur gaussien

Soient $\mu \in \mathbb{R}^n$ et Σ une matrice symétrique semi-définie positive de taille $n \times n$, de rang $k \geq 1$. Il existe une matrice A de taille $n \times k$ telle que $AA^\top = \Sigma$, et pour toute telle matrice :

(i) si $Z \sim \mathcal{N}(0, \mathrm{I}_k)$, alors $AZ + \mu \sim \mathcal{N}(\mu, \Sigma)$;

(ii) si $X \sim \mathcal{N}(\mu, \Sigma)$, il existe un vecteur aléatoire $Z \sim \mathcal{N}(0, \mathrm{I}_k)$ tel que $X = AZ + \mu$ presque sûrement.

En particulier, la loi $\mathcal{N}(\mu, \Sigma)$ existe pour tout couple (μ, Σ) (le cas $\Sigma = 0$ correspondant à la masse de Dirac en μ).

Démonstration

Existence de A. La matrice Σ étant symétrique semi-définie positive, le théorème spectral fournit une base orthonormée $(u_1, \dots, u_n)$ de vecteurs propres, de valeurs propres $\lambda_1 \geq \cdots \geq \lambda_k > 0 = \lambda_{k+1} = \cdots = \lambda_n$; la matrice $A = [\sqrt{\lambda_1}\, u_1, \dots, \sqrt{\lambda_k}\, u_k]$ vérifie $AA^\top = \sum_{j=1}^k \lambda_j u_j u_j^\top = \Sigma$.
(i) C'est la Proposition I-1.2.4 : $AZ + \mu \sim \mathcal{N}(A0 + \mu,\ A\,\mathrm{I}_k A^\top) = \mathcal{N}(\mu, \Sigma)$.
(ii) Soit A de taille $n \times k$ telle que $AA^\top = \Sigma$. Pour $y \in \mathbb{R}^n$, $AA^\top y = 0$ implique $\|A^\top y\|^2 = y^\top AA^\top y = 0$: les matrices Σ et $A^\top$ ont même noyau, donc même rang k, et A est de rang k, c'est-à-dire injective. De même, $A^\top Ax = 0$ implique $\|Ax\|^2 = 0$ puis $x = 0$: la matrice $A^\top A$, de taille $k \times k$, est inversible. Posons $A^\# := (A^\top A)^{-1} A^\top$, de sorte que $A^\# A = \mathrm{I}_k$, et $Z := A^\#(X - \mu)$. D'après la Proposition I-1.2.4,

$$Z \sim \mathcal{N}\big(0,\ A^\# \Sigma (A^\#)^\top\big) = \mathcal{N}\big(0,\ (A^\# A)(A^\# A)^\top\big) = \mathcal{N}(0, \mathrm{I}_k)\ .$$

Il reste à voir que $AZ + \mu = X$ presque sûrement. La matrice $P := AA^\# = A(A^\top A)^{-1} A^\top$ est symétrique et vérifie $P^2 = P$: c'est la matrice de la projection orthogonale sur l'image de A, et $AZ + \mu = P(X - \mu) + \mu$. Or $X - \mu$ appartient presque sûrement à l'image de A. En effet, l'image de $\Sigma = AA^\top$ est contenue dans celle de A et ces deux sous-espaces sont de dimension k, donc $\mathrm{Im}(A) = \mathrm{Im}(\Sigma) = (\mathrm{Ker}\,\Sigma)^\perp$, la matrice Σ étant symétrique. Si $v \in \mathrm{Ker}\,\Sigma$, la variable gaussienne $v^\top(X - \mu)$ est centrée, de variance $v^\top \Sigma v = 0$: elle est nulle presque sûrement. En appliquant ceci aux vecteurs d'une base (finie) de $\mathrm{Ker}\,\Sigma$, on obtient que $X - \mu \in (\mathrm{Ker}\,\Sigma)^\perp = \mathrm{Im}(A)$ presque sûrement, d'où $P(X - \mu) = X - \mu$ et $X = AZ + \mu$ presque sûrement. □

Proposition I-1.2.6: Indépendance et décorrélation

Soient $X \sim \mathcal{N}(\mu, \Sigma)$ un vecteur gaussien de $\mathbb{R}^n$ et $n = n_1 + \cdots + n_m$ une décomposition de n en entiers strictement positifs. On découpe $X = (X_{(1)}^\top, \dots, X_{(m)}^\top)^\top$ en sous-vecteurs de tailles $n_1, \dots, n_m$, de même $\mu = (\mu_{(1)}^\top, \dots, \mu_{(m)}^\top)^\top$, et $\Sigma = (\Sigma_{ij})_{1 \leq i,j \leq m}$ par blocs, où $\Sigma_{ij} = \mathbb{E}\big[(X_{(i)} - \mu_{(i)})(X_{(j)} - \mu_{(j)})^\top\big]$ est de taille $n_i \times n_j$. Alors les sous-vecteurs $X_{(1)}, \dots, X_{(m)}$ sont (mutuellement) indépendants si et seulement si $\Sigma_{ij} = 0$ pour tous $i \neq j$.

Démonstration

Si les sous-vecteurs sont indépendants, alors pour $i \neq j$, $\Sigma_{ij} = \mathbb{E}\big[X_{(i)} - \mu_{(i)}\big]\, \mathbb{E}\big[X_{(j)} - \mu_{(j)}\big]^\top = 0$. Réciproquement, supposons $\Sigma_{ij} = 0$ pour tous $i \neq j$. Pour $t = (t_{(1)}^\top, \dots, t_{(m)}^\top)^\top \in \mathbb{R}^n$ découpé de la même façon, $t^\top \mu = \sum_{i=1}^m t_{(i)}^\top \mu_{(i)}$ et $t^\top \Sigma t = \sum_{i=1}^m t_{(i)}^\top \Sigma_{ii} t_{(i)}$, de sorte que, d'après la Proposition I-1.2.3,

$$\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\, t^\top X}\right] = \prod_{i=1}^m \exp\left(\mathrm{i}\, t_{(i)}^\top \mu_{(i)} - \frac{1}{2}\, t_{(i)}^\top \Sigma_{ii} t_{(i)}\right) = \prod_{i=1}^m \mathbb{E}\left[\mathrm{e}^{\mathrm{i}\, t_{(i)}^\top X_{(i)}}\right]\ ,$$

la dernière égalité venant de ce que $X_{(i)} \sim \mathcal{N}(\mu_{(i)}, \Sigma_{ii})$ (Proposition I-1.2.4, appliquée à la matrice qui extrait les coordonnées du bloc i). La fonction caractéristique de X est donc celle de la loi produit $\mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_m$, où $\mathbb{P}_i$ désigne la loi de $X_{(i)}$; comme la fonction caractéristique caractérise la loi sur $\mathbb{R}^n$, la loi de X est cette loi produit, ce qui est l'indépendance des $X_{(i)}$. □

Remarque. Dans un vecteur gaussien, la décorrélation de deux sous-vecteurs équivaut donc à leur indépendance. Cette équivalence est propre aux vecteurs gaussiens et ne s'étend pas à des variables gaussiennes qui ne formeraient pas un vecteur gaussien. Soient par exemple $X \sim \mathcal{N}(0,1)$ et ε une variable aléatoire indépendante de X telle que $\mathbb{P}(\varepsilon = 1) = \mathbb{P}(\varepsilon = -1) = 1/2$, et posons $Y = \varepsilon X$. Alors $Y \sim \mathcal{N}(0,1)$, car la loi de X est symétrique, et $\mathrm{Cov}(X,Y) = \mathbb{E}[\varepsilon]\,\mathbb{E}[X^2] = 0$, mais X et Y ne sont pas indépendantes puisque $|Y| = |X|$. Le vecteur (X,Y) n'est pas gaussien : $X+Y = (1+\varepsilon)X$ est nulle avec probabilité $1/2$ sans être presque sûrement constante, donc n'est ni dégénérée ni à densité.

Corollaire I-1.2.1: Transformations linéaires indépendantes

Soient $\sigma > 0$, $Z \sim \mathcal{N}(0, \sigma^2 \mathrm{I}_n)$, et A_1, A_2 deux matrices réelles de tailles respectives $m_1 \times n$ et $m_2 \times n$ telles que $A_1 A_2^\top = 0$. Alors le vecteur $(A_1 Z, A_2 Z)$ de $\mathbb{R}^{m_1+m_2}$ est gaussien et les vecteurs aléatoires $A_1 Z$ et $A_2 Z$ sont indépendants. En particulier, si P_1 et P_2 sont deux matrices symétriques, par exemple deux projecteurs orthogonaux, telles que $P_1 P_2 = 0$, alors $P_1 Z$ et $P_2 Z$ sont indépendants.

Démonstration

Le vecteur $(A_1 Z, A_2 Z)$ s'écrit AZ avec $A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}$ de taille $(m_1+m_2) \times n$: il est gaussien d'après la Proposition I-1.2.4, de matrice de covariance

$$\sigma^2 A A^\top = \sigma^2 \begin{pmatrix} A_1 A_1^\top & A_1 A_2^\top \\ A_2 A_1^\top & A_2 A_2^\top \end{pmatrix} = \begin{pmatrix} \sigma^2 A_1 A_1^\top & 0 \\ 0 & \sigma^2 A_2 A_2^\top \end{pmatrix} .$$

Les blocs croisés sont nuls, et la Proposition I-1.2.6 conclut. Pour des matrices symétriques, $P_1 P_2^\top = P_1 P_2$. □

On rencontre aussi ce corollaire sous la forme suivante, qui lui est équivalente : si B_1 et B_2 sont deux matrices ayant n lignes et telles que $B_1^\top B_2 = 0$, alors $B_1^\top Z$ et $B_2^\top Z$ sont indépendants (prendre $A_i = B_i^\top$).

Proposition I-1.2.7: Densité d'un vecteur gaussien

Soient $\mu \in \mathbb{R}^n$ et Σ une matrice symétrique *définie positive* de taille $n \times n$. La loi $\mathcal{N}(\mu, \Sigma)$ admet pour densité par rapport à la mesure de Lebesgue $\mathrm{Leb}^{\otimes n}$ sur $\mathbb{R}^n$

$$g_{\mu,\Sigma}(x) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right) , \qquad x \in \mathbb{R}^n .$$

Démonstration

Traitons d'abord le cas $\mu = 0$, $\Sigma = \mathrm{I}_n$. Si $Z \sim \mathcal{N}(0, \mathrm{I}_n)$, la Proposition I-1.2.6, appliquée avec des blocs de taille 1, montre que les coordonnées $Z_1, \ldots, Z_n$ sont indépendantes, chacune de loi $\mathcal{N}(0,1)$: la loi de Z est la loi produit $\mathcal{N}(0,1)^{\otimes n}$, de densité $\prod_{i=1}^n g(z_i) = (2\pi)^{-n/2} \exp(-\|z\|^2/2)$ par rapport à $\mathrm{Leb}^{\otimes n}$ (densité produit, comme en (I-1.1)).
Dans le cas général, Σ est de rang n : la Proposition I-1.2.5 fournit une matrice A de taille $n \times n$,

inversible, telle que $AA^\top = \Sigma$, et $X \sim \mathcal{N}(\mu, \Sigma)$ a même loi que $AZ + \mu$ avec $Z \sim \mathcal{N}(0, \mathrm{I}_n)$. L'application $\phi : z \mapsto Az + \mu$ est un C^1-difféomorphisme de $\mathbb{R}^n$ sur lui-même, d'inverse $\phi^{-1}(x) = A^{-1}(x - \mu)$, dont la matrice jacobienne est constante, égale à A^{-1}. La formule de changement de variables de la Section I-1.2 donne pour densité de X

$$q(x) = \frac{1}{|\det A|} (2\pi)^{-n/2} \exp\left(-\frac{1}{2} \left\| A^{-1}(x - \mu) \right\|^2 \right) ,$$

et l'on conclut avec $\|A^{-1}(x-\mu)\|^2 = (x-\mu)^\top (A^{-1})^\top A^{-1}(x-\mu) = (x-\mu)^\top (AA^\top)^{-1}(x-\mu) = (x-\mu)^\top \Sigma^{-1}(x-\mu)$ et $(\det A)^2 = \det(AA^\top) = \det \Sigma$. □

Lorsque Σ est de rang $k < n$, la démonstration de la Proposition I-1.2.5 montre que $X - \mu$ prend presque sûrement ses valeurs dans le sous-espace $\mathrm{Im}(\Sigma)$, de dimension k, qui est de mesure de Lebesgue nulle dans $\mathbb{R}^n$: la loi $\mathcal{N}(\mu, \Sigma)$ n'est alors pas absolument continue par rapport à $\mathrm{Leb}^{\otimes n}$ et n'admet pas de densité (loi *dégénérée*). C'est le cas des lois $\mathcal{N}(P_F \mu, \sigma^2 P_F)$ du théorème de Cochran (Section I-1.1.9). Pour $\Sigma = \sigma^2 \mathrm{I}_n$, on retrouve la densité produit de l'exemple de l'échantillon gaussien de la Section I-1.1.5.

Illustration / animation : Les ellipses de niveau de la densité gaussienne suivent les axes propres de la covariance

On illustre ici, pour Γ choisie par σ_1, σ_2, ρ, la carte de densité de $\mathcal{N}(0, \Gamma)$ sur $\mathbb{R}^2$ et ses ellipses de niveau, orientées selon les axes principaux de Γ ; la covariance empirique du nuage $X = AZ$ s'affiche une fois demandée.

link

Fonction Gamma, lois Gamma, du χ^2 et de Student

La fonction Gamma. Pour $\alpha > 0$, l'intégrale

$$\Gamma(\alpha) := \int_0^{+\infty} y^{\alpha-1} \mathrm{e}^{-y} \, \mathrm{d}y$$

converge et définit un réel strictement positif : au voisinage de 0, l'intégrande est équivalent à $y^{\alpha-1}$, intégrable car $\alpha > 0$, et à l'infini la décroissance exponentielle l'emporte. On a $\Gamma(1) = \int_0^{+\infty} \mathrm{e}^{-y} \, \mathrm{d}y = 1$ et, pour $\alpha > 1$, une intégration par parties donne

$$\Gamma(\alpha) = \Big[-y^{\alpha-1} \mathrm{e}^{-y} \Big]_0^{+\infty} + (\alpha - 1) \int_0^{+\infty} y^{\alpha-2} \mathrm{e}^{-y} \, \mathrm{d}y = (\alpha - 1)\, \Gamma(\alpha - 1) \, .$$

Par récurrence, pour tout entier $n \geq 1$, $\Gamma(n) = (n-1)(n-2) \cdots 1 \cdot \Gamma(1) = (n-1)!$: la fonction Gamma prolonge la factorielle. On a aussi $\Gamma(1/2) = \sqrt{\pi}$: le changement de variable $y = x^2/2$ donne $\Gamma(1/2) = \sqrt{2} \int_0^{+\infty} \mathrm{e}^{-x^2/2} \, \mathrm{d}x = \sqrt{2} \cdot \sqrt{2\pi}/2 = \sqrt{\pi}$, la dernière intégrale se calculant à partir de la densité gaussienne réduite. Par construction, $\Gamma(p)$ est la constante de normalisation de la densité f_p de la loi Gamma réduite $\mathrm{Gamma}(p)$ définie en Section I-1.2.

Moments et fonction caractéristique des lois Gamma. Soit $Z \sim \mathrm{Gamma}(p)$, avec $p > 0$. Pour tout réel $r > -p$,

$$\mathbb{E}\left[Z^r \right] = \frac{1}{\Gamma(p)} \int_0^{+\infty} x^{r+p-1} \mathrm{e}^{-x} \, \mathrm{d}x = \frac{\Gamma(p + r)}{\Gamma(p)} \, ,$$

et $\mathbb{E}[Z^r] = +\infty$ si $r \leq -p$. En particulier $\mathbb{E}[Z] = p$, $\mathbb{E}[Z^2] = p(p+1)$ et $\mathrm{Var}(Z) = p$. Pour $X = Z/\lambda \sim \mathrm{Gamma}(p, \lambda)$, on en déduit $\mathbb{E}[X^r] = \lambda^{-r}\, \Gamma(p + r)/\Gamma(p)$, en particulier $\mathbb{E}[X] = p/\lambda$,

$\mathrm{Var}(X) = p/\lambda^2$ et, pour $p > 1$, $\mathbb{E}[1/X] = \lambda/(p-1)$. Ce sont ces formules, avec des exposants r négatifs, qui donnent l'espérance et l'erreur quadratique de l'estimateur $\hat\theta_{n,1} = 1/\bar X_n$ du modèle exponentiel (Section I-1.4) : sous $\mathbb{P}_\theta$, $\sum_{i=1}^n X_i \sim \mathrm{Gamma}(n,\theta)$ et $\mathbb{E}_\theta[\hat\theta_{n,1}] = n\theta\,\Gamma(n-1)/\Gamma(n) = \frac{n}{n-1}\theta$.

La fonction caractéristique de la loi $\mathrm{Gamma}(p,\lambda)$ vaut

$$\phi_{p,\lambda}(t) = \mathbb{E}\left[\mathrm{e}^{\mathrm{i}tX}\right] = \left(1 - \frac{\mathrm{i}t}{\lambda}\right)^{-p}, \qquad t \in \mathbb{R}\,,$$

où la puissance est définie par la détermination principale du logarithme, le nombre complexe $1 - \mathrm{i}t/\lambda$ étant de partie réelle strictement positive. On l'obtient en calculant d'abord, pour $s < \lambda$ réel, $\mathbb{E}[\mathrm{e}^{sX}] = \frac{\lambda^p}{\Gamma(p)} \int_0^{+\infty} x^{p-1}\mathrm{e}^{-(\lambda-s)x}\,\mathrm{d}x = (1-s/\lambda)^{-p}$ (changement de variable $u = (\lambda-s)x$), puis en observant que les deux membres sont des fonctions holomorphes de s sur le demi-plan $\{\mathrm{Re}(s) < \lambda\}$ qui coïncident sur un segment réel : elles coïncident sur tout le demi-plan, en particulier en $s = \mathrm{i}t$. Cette expression démontre la propriété d'additivité énoncée en Section I-1.2 : si $X_1, \dots, X_n$ sont indépendantes de lois $\mathrm{Gamma}(p_i,\lambda)$, la fonction caractéristique de leur somme est $\prod_{i=1}^n (1 - \mathrm{i}t/\lambda)^{-p_i} = (1 - \mathrm{i}t/\lambda)^{-\sum_i p_i}$, qui est celle de $\mathrm{Gamma}(\sum_i p_i, \lambda)$.

Loi du χ^2.

Définition I-1.11 (Loi du χ^2 à ν degrés de liberté). *Soient $\nu \in \mathbb{N}^*$ et $X_1, \dots, X_\nu$ des variables aléatoires indépendantes de loi $\mathcal{N}(0,1)$. La loi de la variable aléatoire*

$$U = \sum_{i=1}^{\nu} X_i^2$$

*est appelée loi du χ^2 (*khi-deux*) centrée à ν degrés de liberté, et notée $\chi^2(\nu)$.*

Proposition I-1.2.8: La loi du χ^2 est une loi Gamma

(i) Si $X \sim \mathcal{N}(0,1)$, alors $X^2 \sim \mathrm{Gamma}(1/2, 1/2)$.

(ii) Pour tout $\nu \in \mathbb{N}^*$, $\chi^2(\nu) = \mathrm{Gamma}(\nu/2, 1/2)$: la loi $\chi^2(\nu)$ admet pour densité par rapport à Leb

$$f_\nu(x) = \frac{1}{2^{\nu/2}\,\Gamma(\nu/2)}\, x^{\nu/2-1}\mathrm{e}^{-x/2}\, \mathbb{1}_{\mathbb{R}_+^*}(x)\,,$$

et si $U \sim \chi^2(\nu)$, alors $\mathbb{E}[U] = \nu$ et $\mathrm{Var}(U) = 2\nu$.

Démonstration

(i) Utilisons la méthode de la fonction muette (Section I-1.2). Pour h mesurable positive sur $\mathbb{R}$, par parité de la densité gaussienne puis par le changement de variable $u = x^2$ sur $\mathbb{R}_+^*$ ($x = \sqrt{u}$, $\mathrm{d}x = \mathrm{d}u/(2\sqrt{u})$),

$$\mathbb{E}\big[h(X^2)\big] = 2\int_0^{+\infty} h(x^2)\,\frac{\mathrm{e}^{-x^2/2}}{\sqrt{2\pi}}\,\mathrm{d}x = \int_0^{+\infty} h(u)\,\frac{1}{\sqrt{2\pi}}\,u^{-1/2}\mathrm{e}^{-u/2}\,\mathrm{d}u\,.$$

Comme $\Gamma(1/2) = \sqrt{\pi}$, on a $1/\sqrt{2\pi} = (1/2)^{1/2}/\Gamma(1/2)$, et la fonction $u \mapsto \frac{(1/2)^{1/2}}{\Gamma(1/2)}\,u^{1/2-1}\mathrm{e}^{-u/2}$ est la densité $f_{1/2,1/2}$ de la loi $\mathrm{Gamma}(1/2,1/2)$.
(ii) Les variables $X_1^2, \dots, X_\nu^2$ sont indépendantes, de loi $\mathrm{Gamma}(1/2,1/2)$; par additivité des lois Gamma de même paramètre $\lambda = 1/2$, leur somme suit la loi $\mathrm{Gamma}(\nu/2,1/2)$, dont la densité $f_{\nu/2,1/2}$ est celle annoncée. Enfin $\mathbb{E}[U] = \sum_{i=1}^\nu \mathbb{E}[X_i^2] = \nu$ et, par indépendance, $\mathrm{Var}(U) = \nu\,\mathrm{Var}(X_1^2) = \nu\big(\mathbb{E}[X_1^4] - \mathbb{E}[X_1^2]^2\big) = 2\nu$, d'après les moments de la loi gaussienne réduite. □

Proposition I-1.2.9: Formes quadratiques de vecteurs gaussiens

(i) Soit $X \sim \mathcal{N}(\mu, \Sigma)$ un vecteur gaussien de $\mathbb{R}^n$ dont la matrice de covariance Σ est définie positive. Alors $(X-\mu)^\top \Sigma^{-1}(X-\mu) \sim \chi^2(n)$.

(ii) Soient $\sigma > 0$, $Z \sim \mathcal{N}(0, \sigma^2 \mathrm{I}_n)$ et Π la matrice d'une projection orthogonale de $\mathbb{R}^n$ de rang $k \geq 1$. Alors $\|\Pi Z\|^2/\sigma^2 \sim \chi^2(k)$.

Démonstration

(i) D'après la Proposition I-1.2.5, X a même loi que $AZ+\mu$ où A est inversible, $AA^\top = \Sigma$ et $Z \sim \mathcal{N}(0, \mathrm{I}_n)$, dont les coordonnées sont indépendantes de loi $\mathcal{N}(0,1)$ (Proposition I-1.2.6). Alors $(X-\mu)^\top \Sigma^{-1}(X-\mu)$ a même loi que $Z^\top A^\top (AA^\top)^{-1} AZ = Z^\top Z = \sum_{i=1}^n Z_i^2 \sim \chi^2(n)$. *(ii)* Soit $H = [h_1, \ldots, h_k]$ la matrice de taille $n \times k$ dont les colonnes forment une base orthonormée de l'image de Π : on a $H^\top H = \mathrm{I}_k$ et $\Pi = HH^\top$, les deux membres étant symétriques, égaux à l'identité sur $\mathrm{Im}(\Pi)$ et nuls sur son orthogonal. Comme Π est symétrique et idempotente, $\|\Pi Z\|^2 = Z^\top \Pi^\top \Pi Z = Z^\top \Pi Z = Z^\top HH^\top Z = \|H^\top Z\|^2$. La Proposition I-1.2.4 donne $H^\top Z \sim \mathcal{N}(0, \sigma^2 H^\top H) = \mathcal{N}(0, \sigma^2 \mathrm{I}_k)$: le vecteur $H^\top Z/\sigma$ a des coordonnées indépendantes de loi $\mathcal{N}(0,1)$, et $\|\Pi Z\|^2/\sigma^2 = \|H^\top Z/\sigma\|^2 \sim \chi^2(k)$. □

Remarque. Si dans (ii) le vecteur Z suit la loi $\mathcal{N}(\mu, \sigma^2 \mathrm{I}_n)$ avec $\mu \neq 0$, le même calcul donne $\|\Pi Z\|^2/\sigma^2 = \|E\|^2$ avec $E \sim \mathcal{N}(H^\top \mu/\sigma, \mathrm{I}_k)$. La loi de $\|E\|^2$ ne dépend de $m = H^\top \mu/\sigma$ que par $\|m\|^2 = \|\Pi\mu\|^2/\sigma^2$, car la loi $\mathcal{N}(0, \mathrm{I}_k)$ est invariante par les transformations orthogonales (Proposition I-1.2.4) : on l'appelle loi du χ^2 à k degrés de liberté *non centrée*, de paramètre de non-centralité $\|\Pi\mu\|^2/\sigma^2$. Elle n'est pas utilisée dans ce chapitre.

Loi de Student.

Définition I-1.12 (Loi de Student à r degrés de liberté). *Soient $r \in \mathbb{N}^*$, $X \sim \mathcal{N}(0,1)$ et $Y \sim \chi^2(r)$ deux variables aléatoires* indépendantes. *La loi de la variable aléatoire*

$$T = \frac{X}{\sqrt{Y/r}}$$

est appelée loi de Student à r degrés de liberté, et notée $\mathrm{t}(r)$.

La variable Y est strictement positive presque sûrement, de sorte que T est bien définie. On montre, par la formule de changement de variables appliquée au couple (X, Y), que la loi $\mathrm{t}(r)$ admet pour densité

$$f_r(t) = \frac{\Gamma\left(\frac{r+1}{2}\right)}{\Gamma\left(\frac{r}{2}\right)\sqrt{r\pi}} \left(1 + \frac{t^2}{r}\right)^{-\frac{r+1}{2}}, \qquad t \in \mathbb{R}\,.$$

Cette densité est symétrique et décroît polynomialement : la loi de Student a des queues plus lourdes que la loi gaussienne. Pour $r = 1$, on retrouve la loi de Cauchy, de densité $1/(\pi(1+t^2))$, qui n'a pas d'espérance ; lorsque $r \to +\infty$, la loi $\mathrm{t}(r)$ converge vers $\mathcal{N}(0,1)$, car Y/r converge vers 1 en probabilité d'après la loi des grands nombres. Ces compléments sont démontrés dans l'annexe du polycopié.

Domination, densités, théorème de Radon-Nikodym

Le théorème de Radon-Nikodym (**??**) montre que le modèle $(\mathsf{Z}, \mathcal{Z}, \mathcal{C})$ est dominé par μ si et seulement si toute loi $\mathbb{P} \in \mathcal{C}$ est absolument continue par rapport à μ (voir **??**). Les mesures de référence usuelles sont la mesure de Lebesgue Leb (lois « à densité » classiques) et les mesures de comptage (lois discrètes). La mesure de domination n'est pas unique, mais ce choix n'affecte pas les quantités utilisées en statistique (rapports de densités, vraisemblance) : deux densités relatives à deux mesures dominantes diffèrent d'un facteur multiplicatif qui ne dépend pas de θ (Section I-1.4, où l'on trouve aussi des exemples de modèles non dominés).

Compléments de théorie de la mesure

Les notions rappelées ici interviennent dès la définition d'un modèle statistique : la tribu borélienne $\mathcal{B}(\mathbb{R}^n)$ des exemples des tailles et de l'échantillon gaussien, et les mesures à densité $p \cdot \mu$ de la définition d'un modèle dominé (Définition I-1.1). Elles sont développées dans l'annexe du polycopié (**??**, **??** et **??**) ; nous nous contentons ici d'en rappeler les énoncés.

Tribu engendrée, tribu borélienne. Soit E un ensemble. L'ensemble $\mathcal{P}(\mathsf{E})$ de toutes les parties de E est une tribu, et l'intersection d'une famille quelconque de tribus sur E est encore une tribu. Étant donnée une classe $\mathcal{C}$ de parties de E, on peut donc parler de la plus petite tribu contenant $\mathcal{C}$.

Définition I-1.13 (Tribu engendrée, tribu borélienne). *Soit $\mathcal{C} \subset \mathcal{P}(\mathsf{E})$. La* tribu engendrée *par $\mathcal{C}$ est l'intersection de toutes les tribus sur E contenant $\mathcal{C}$; on la note $\sigma(\mathcal{C})$.*

Soient E un espace topologique et $\mathcal{O}$ la classe de ses ouverts. La tribu $\sigma(\mathcal{O})$ s'appelle la tribu borélienne *de E et se note $\mathcal{B}(\mathsf{E})$.*

La tribu borélienne est aussi engendrée par les fermés. Dans ce cours, l'espace topologique est presque toujours $\mathbb{R}^d$ muni de sa topologie usuelle (celle de la distance euclidienne) ; la tribu $\mathcal{B}(\mathbb{R}^d)$ est alors aussi engendrée par les boules, par les pavés, et même par les pavés à coordonnées rationnelles (cette dernière famille ayant l'avantage d'être dénombrable). Elle coïncide avec la tribu produit $\mathcal{B}(\mathbb{R})^{\otimes d}$ de la Section I-1.2, ce qui justifie que le n-échantillon d'un modèle sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ait pour espace des observations $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. Enfin, toute application continue de $\mathbb{R}^d$ dans $\mathbb{R}^k$ est mesurable pour les tribus boréliennes : les statistiques usuelles (moyenne empirique, maximum, médiane empirique) sont donc bien des applications mesurables.

Mesures à densité, absolue continuité.

Définition I-1.14 (Mesure à densité, absolue continuité, domination). *Soient $(\mathsf{E}, \mathcal{E})$ un espace mesurable et μ une mesure sur $\mathcal{E}$.*

(i) Pour toute fonction mesurable positive f, l'application $\nu : B \mapsto \int_B f \, \mathrm{d}\mu$ est une mesure sur $\mathcal{E}$, appelée mesure de densité f par rapport à μ *et notée $\nu = f \cdot \mu$. Pour toute fonction mesurable positive h, on a alors $\int h \, \mathrm{d}\nu = \int hf \, \mathrm{d}\mu$.*

(ii) Une mesure ν sur $\mathcal{E}$ est dite absolument continue par rapport à μ, *ce que l'on note $\nu \ll \mu$, si pour tout $B \in \mathcal{E}$, $\mu(B) = 0$ implique $\nu(B) = 0$. On dit aussi que μ* domine *ν.*

Une mesure à densité $f \cdot \mu$ est absolument continue par rapport à μ : l'intégrale de f sur un ensemble μ-négligeable est nulle. Le théorème de Radon-Nikodym affirme la réciproque, sous l'hypothèse que μ est σ-finie, c'est-à-dire que E est réunion dénombrable d'ensembles de μ-mesure finie : c'est le cas de la mesure de Lebesgue sur $\mathbb{R}^d$, des mesures de comptage sur les ensembles dénombrables et de toutes les mesures finies.

Théorème I-1.2.1: Théorème de Radon-Nikodym

Soient μ une mesure σ-finie sur un espace mesurable $(\mathsf{E}, \mathcal{E})$ et ν une mesure absolument continue par rapport à μ. Il existe une fonction mesurable positive f, unique à un ensemble μ-négligeable près, telle que $\nu = f \cdot \mu$. Cette fonction est appelée *densité* (ou *dérivée de Radon-Nikodym*) de ν par rapport à μ et se note $f = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$. De plus,

(i) f est μ-presque partout finie si et seulement si ν est σ-finie ;

(ii) f est μ-intégrable si et seulement si ν est une mesure finie.

Ce théorème est admis dans ce cours ; il est énoncé et commenté dans l'annexe du polycopié (**??**). Appliqué à une loi de probabilité $\mathbb{P}$, qui est une mesure finie, il dit que $\mathbb{P}$ admet une densité par rapport à μ si et seulement si $\mathbb{P} \ll \mu$, la densité étant alors μ-intégrable et d'intégrale 1. C'est

ce qui justifie l'équivalence rappelée en Section I-1.2 : le modèle $(\mathsf{Z}, \mathcal{Z}, \mathcal{C})$ est dominé par la mesure σ-finie μ si et seulement si toute loi $\mathbb{P} \in \mathcal{C}$ est absolument continue par rapport à μ. Ainsi, la loi $\mathcal{N}(m, \sigma^2)$ avec $\sigma^2 > 0$ et la mesure de Lebesgue Leb sont absolument continues l'une par rapport à l'autre, car la densité gaussienne est strictement positive ; en revanche, la masse de Dirac δ_a n'est pas absolument continue par rapport à Leb, puisque $\mathrm{Leb}(\{a\}) = 0$ alors que $\delta_a(\{a\}) = 1$: elle n'admet pas de densité par rapport à Leb. Nous retrouverons cette situation avec les lois gaussiennes dégénérées de la Section I-1.2.