Les fichounettes de Tonton Rémi

Conditionnement Gaussien

par Rémi Peyre

1 L'énoncé à retenir

Rappelons d'abord la notion de *loi conditionnelle*. Si on travaille sur un espace de Lusin probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et que $\mathcal{B}$ est une sous-tribu de $\mathcal{A}$, la loi conditionnelle sachant $\mathcal{B}$, notée $\mathbb{P}(\bullet|\mathcal{B})$, est une *variable aléatoire à valeur dans l'espace des lois sur* $(\Omega, \mathcal{A})$, qui est $\mathcal{B}$-mesurable, et telle que pour tout évènement $\mathcal{B}$-mesurable B, pour tout évènement $A \in \mathcal{A}$,

$$\mathbb{P}(A \cap B) = \int_B \mathbb{P}(A|\mathcal{B})\, d\mathbb{P}(\omega). \tag{1}$$

En particulier, l'espérance conditionnelle d'une variable aléatoire f n'est rien d'autre que l'espérance de f sous la loi conditionnelle.

Les vecteurs gaussiens se comportent particulièrement bien lorsqu'on les conditionne. L'énoncé le plus simple à retenir est le suivant, à partir duquel on pourra facilement retrouver les formules de la § 2 :

Théorème 1. *Soit (X, Y) un vecteur gaussien (pas forcément centré) de $\mathbb{R}^{n+m}$, avec $X \in \mathbb{R}^n$ et $Y \in \mathbb{R}^m$. Alors il existe une constante γ et des matrices $A \in \mathbb{R}^{n\times m}$ et $B \in \mathbb{R}^{m\times m}$ telles que*

$$\mathit{Loi}(Y|X = x) = \gamma + xA + \mathcal{N}(B). \tag{2}$$

Autrement dit, quand on conditionne un vecteur gaussien par rapport à certaines de ses composantes, la loi de ses autres composantes est :

1. *Gaussienne ;*
2. *Avec une variance qui ne dépend pas de la valeur des composantes de conditionnement ;*
3. *Avec une espérance qui dépend affinement de la valeur des composantes de conditionnement.*

En outre, quand le vecteur (X, Y) est centré, on peut prendre $\gamma = 0$ dans (2).

Remarque 2. Je ne démontrerai ce théorème que dans le cas où le vecteur gaussien X est non dégénéré, càd. quand X a pour support $\mathbb{R}^n$ tout entier, ce qui se traduit par « $\mathrm{rg}(\mathrm{Var}\, X) = n$ ». Le cas dégénéré n'est guère plus compliqué, vu qu'on peut se ramener au cas non dégénéré par un changement de base sur les coordonnées de X, mais il est moins élégant, car les formules sympathiques de la § 2 (qui utilisent l'inverse de la matrice $\mathrm{Var}(X)$) ne peuvent plus être utilisées telles quelles, et le choix de γ et A n'est plus unique.

La preuve du théorème 1 réside en fait dans les calculs de la section suivante. Ce qu'on va montrer est qu'on peut effectivement trouver des matrices A et B telles que, si on *pose* que X a la loi qu'il a et que la loi jointe de (X, Y) est ensuite *définie* par (2), alors le vecteur (X, Y) est bien gaussien avec les paramètres qu'on souhaite.

2 Formules explicites

2.1 Cas où (X, Y) est centré

Dans ce paragraphe, on suppose que (X, Y) est centré, de sorte qu'on a $(X, Y) \sim \mathcal{N}\big(\mathrm{Var}(X, Y)\big)$.

On décomposera par blocs la matrice de covariance de (X, Y) comme

$$\mathrm{Var}(X, Y) = \begin{pmatrix} V & C \\ C^\mathsf{T} & W \end{pmatrix}. \tag{3}$$

X étant supposé non dégénéré, la matrice V est alors inversible.

Comme nous l'avons dit à la fin de la § 1, nous allons considérer un vecteur gaussien X de loi $\mathcal{N}(V)$ et *définir* la loi jointe de (X, Y) par (2) avec $\gamma = 0$. Cela revient à dire que (X, Y) s'écrit sous la forme $(X, XA + Z)$ où Z est un vecteur gaussien centré de variance B, *indépendant de X*. Dans ce cas, il est clair d'après les théorèmes sur les combinaisons linéaires et les sommes indépendantes des variables gaussiennes que (X, Y) est gaussien centré ; et on a par indépendance de X et Z :

$$\mathrm{Var}(Y) = \mathrm{Var}(XA + Z) = \mathrm{Var}(XA) + \mathrm{Var}(Z) = A^\mathsf{T}\,\mathrm{Var}(X)A + \mathrm{Var}(Z) = A^\mathsf{T} V A + B \tag{4}$$

et

$$\mathrm{Cov}(X, Y) = \mathrm{Cov}(X, XA + Z) = \mathrm{Cov}(X, XA) = \mathrm{Var}(X)A = VA. \tag{5}$$

On voit ainsi que prendre $A = V^{-1}C$ permet de trouver la valeur attendue pour $\mathrm{Cov}(X, Y)$ dans (5), puis que prendre $B = W - C^\mathsf{T} V^{-1} C$ permet de trouver la valeur attendue de $\mathrm{Var}(Y)$ dans (4) — en utilisant que V est symétrique.

Il reste juste à vérifier que la matrice B est bien une matrice de covariance admissible, càd. qu'elle est symétrique positive : la symétrie est évidente ; et la positivité découle de ce que pour tout vecteur $u \in \mathbb{R}^m$,

$$\begin{aligned} uBu^\mathsf{T} = u(W - C^\mathsf{T} V^{-1} C)u^\mathsf{T} &= (-uC^\mathsf{T} V^{-1}, u) \begin{pmatrix} V & C \\ C^\mathsf{T} & W \end{pmatrix} (-uC^\mathsf{T} V^{-1}, u)^\mathsf{T} \\ &= \mathrm{Var}\big((X, Y) \times (-uC^\mathsf{T} V^{-1}, u)^\mathsf{T}\big) \geqslant 0. \end{aligned} \tag{6}$$

2.2 Cas général

Dans le cas général où le vecteur gaussien n'est pas forcément centré, notons (ν, μ) l'espérance de (X, Y) (avec $\nu \in \mathbb{R}^n$ et $\mu \in \mathbb{R}^m$), de sorte que $(X, Y) \sim (\nu, \mu) + \mathcal{N}\big(\mathrm{Var}(X, Y)\big)$. Notons $\tilde{X} := X - \nu$, resp. $\tilde{Y} := Y - \mu$; alors $(\tilde{X}, \tilde{Y})$ est gaussien centré et rentre donc dans le cadre du paragraphe précédent, avec la même matrice de covariance que (X, Y) — matrice qu'on décompose à nouveau selon (3). Puisque conditionner par rapport à $\{X = x\}$ est équivalent à conditionner par rapport à $\{\tilde{X} = x - \nu\}$, on en déduit la loi conditionnelle de $\tilde{Y}$ sous $\{X = x\}$, et donc la loi conditionnelle de $Y = \mu + \tilde{Y}$:

$$\mathit{Loi}(Y | X = x) = \mu + (x - \nu)V^{-1}C + \mathcal{N}(W - C^\mathsf{T} V^{-1} C). \tag{7}$$