TP2 – Droites de régression

Exercice 1 : estimation de D_{YX} par maximum de vraisemblance

Lancez le script `donnees.m`, qui affiche n points $P_i = (x_i, y_i)$ du plan censés être alignés. En utilisant l'équation paramétrique $y = ax + b$ d'une droite et en notant $r_{a,b}(P_i) = y_i - ax_i - b$ le résidu de cette équation pour le point P_i, il est légitime de modéliser ces résidus par une loi normale centrée :

$$f_{a,b}(P_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{r_{a,b}(P_i)^2}{2\,\sigma^2}\right\} \tag{1}$$

La *droite de régression de* Y *en* X d'un tel nuage de points, notée D_{YX}, est la droite d'équation paramétrique $y = a^*x + b^*$, où a^* et b^* sont les valeurs des paramètres a et b qui maximisent la log-vraisemblance :

$$(a^*, b^*) = \underset{(a,b)\in\mathbb{R}^2}{\arg\max}\left\{\ln\left[\prod_{i=1}^{n} f_{a,b}(P_i)\right]\right\} = \underset{(a,b)\in\mathbb{R}^2}{\arg\min}\left\{n\,\ln\sigma + \frac{1}{2\,\sigma^2}\sum_{i=1}^{n} r_{a,b}(P_i)^2\right\} \tag{2}$$

Si l'on suppose l'écart-type du bruit σ fixé, alors le problème se simplifie :

$$(a^*, b^*) = \underset{(a,b)\in\mathbb{R}^2}{\arg\min}\left\{\sum_{i=1}^{n} r_{a,b}(P_i)^2\right\} = \underset{(a,b)\in\mathbb{R}^2}{\arg\min}\left\{\sum_{i=1}^{n}(y_i - ax_i - b)^2\right\} \tag{3}$$

La résolution de (3) par tirages aléatoires n'est pas aussi simple qu'il y paraît, car les inconnues a et b ne sont pas bornées et a ne suit pas une loi uniforme. Néanmoins, il est facile de montrer que D_{YX} contient le centre de gravité G. On peut donc calculer les coordonnées (x_G, y_G) de G, puis centrer les données. Comme l'équation de D_{YX} devient $y' = a^*x'$ après changement d'origine, le problème (3) se simplifie encore :

$$a^* = \underset{a\in\mathbb{R}}{\arg\min}\left\{\sum_{i=1}^{n}(y'_i - ax'_i)^2\right\} = \tan\left[\underset{\psi\in\,]-\frac{\pi}{2},\frac{\pi}{2}[}{\arg\min}\left\{\sum_{i=1}^{n}(y'_i - \tan\psi\,x'_i)^2\right\}\right] \tag{4}$$

Dans (4), la deuxième égalité vient de ce que le paramètre a d'une droite est égal à la tangente de son angle polaire ψ. La résolution de (4) peut être effectuée par tirages aléatoires de ψ, selon une loi uniforme sur $]-\frac{\pi}{2}, \frac{\pi}{2}[$.

Complétez le script `exercice_1.m` de façon à estimer ψ^*, puis les paramètres a^* et b^*, en suivant ce procédé.

Exercice 2 : estimation de D_{YX} par résolution d'un système linéaire

Le critère à minimiser dans (2) s'écrit $\mathcal{F}(\sigma, a, b) = n\,\ln\sigma + \frac{1}{2\,\sigma^2}\sum_{i=1}^{n} r_{a,b}(P_i)^2$. Le problème (2) peut donc également être considéré comme un problème d'optimisation différentiable. En notant $\mathcal{G}(a, b) = \sum_{i=1}^{n} r_{a,b}(P_i)^2$:

$$\nabla\mathcal{F}(\sigma, a, b) = 0 \iff \left\{\begin{array}{l}\nabla_\sigma\mathcal{F}(\sigma, a, b) = 0\\ \nabla_{a,b}\mathcal{F}(\sigma, a, b) = 0\end{array}\right. \iff \left\{\begin{array}{l}\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} r_{a,b}(P_i)^2\\ \nabla\mathcal{G}(a, b) = 0\end{array}\right. \tag{5}$$

La première de ces équations était prévisible, puisque c'est la définition même de la variance. Quant à la deuxième équation, elle correspond à l'optimalité du critère à minimiser dans (3). Or, ce critère s'écrit aussi :

$$\mathcal{G}(a, b) = \|AX - B\|^2 \quad \text{, avec } A = \begin{bmatrix} x_1 & \dots & x_n\\ 1 & \dots & 1\end{bmatrix}^\top,\ X = \begin{bmatrix} a\\ b\end{bmatrix} \text{ et } B = \begin{bmatrix} y_1 & \dots & y_n\end{bmatrix}^\top \tag{6}$$

Minimiser $\mathcal{G}(a, b)$ revient donc à chercher une solution approchée du système linéaire $AX = B$, au sens des *moindres carrés ordinaires*. Le problème se résout en écrivant les *équations normales* $A^\top AX = A^\top B$, dont la solution s'écrit $X^* = (A^\top A)^{-1}A^\top B = A^+B$, où $A^+ = (A^\top A)^{-1}A^\top$ est la *matrice pseudo-inverse* de A (`pinv`).

Complétez le script `exercice_2.m` de manière à comparer cette méthode d'estimation de D_{YX} avec celle de l'exercice 1. Observez l'évolution des résultats lorsque la valeur de `nb_tirages` diminue. Testez également différentes valeurs du nombre $n \geqslant 2$ de données.

Exercice 3 : estimation de $D_\perp$ par maximum de vraisemblance

Une droite D du plan peut également être définie par son *équation cartésienne normalisée* $x\cos\theta + y\sin\theta = \rho$:

- θ est l'angle polaire du vecteur $\vec{v}$ orthogonal à D, de norme 1, tel que $\theta \in]0, \pi]$ (cf. figure 1).
- À partir d'un point $P = (x, y)$ quelconque de D, on peut calculer le second paramètre $\rho = x\cos\theta + y\sin\theta$.

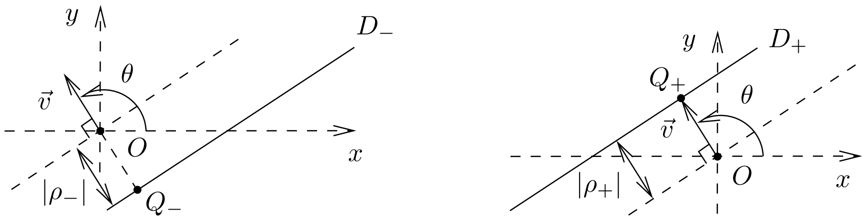

Figure 1 – Les droites D_- et D_+ correspondent aux paramètres $(\theta, \rho_-) = (130, -1)$ et $(\theta, \rho_+) = (130, 1)$.

Si l'on désigne par $r_{\theta,\rho}(P_i) = x_i \cos\theta + y_i \sin\theta - \rho$ le résidu au point P_i de l'équation cartésienne normalisée, il semble à nouveau légitime de modéliser ces résidus par une loi normale centrée :

$$f_{\theta,\rho}(P_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{r_{\theta,\rho}(P_i)^2}{2\,\sigma^2}\right\} \tag{7}$$

La *droite de régression en distance orthogonale* de ce nuage de points, notée $D_\perp$, est la droite d'équation $x\cos\theta^* + y\sin\theta^* = \rho^*$, où θ^* et ρ^* sont les valeurs des paramètres θ et ρ qui maximisent la log-vraisemblance :

$$(\theta^*, \rho^*) = \underset{(\theta,\rho)\in\,]0,\pi]\times\mathbb{R}}{\arg\max} \left\{\ln\left[\prod_{i=1}^{n} f_{\theta,\rho}(P_i)\right]\right\} = \underset{(\theta,\rho)\in\,]0,\pi]\times\mathbb{R}}{\arg\min} \left\{n\,\ln\sigma + \frac{1}{2\,\sigma^2}\sum_{i=1}^{n} r_{\theta,\rho}(P_i)^2\right\} \tag{8}$$

En supposant σ fixé, et sachant que la droite de régression $D_\perp$ contient elle aussi le centre de gravité G, la résolution du problème (8) est en tout point analogue à celle du problème (2). Par analogie avec (4) :

$$\theta^* = \underset{\theta\in\,]0,\pi]}{\arg\min} \left\{\sum_{i=1}^{n} (x'_i \cos\theta + y'_i \sin\theta)^2\right\} \tag{9}$$

Complétez le script `exercice_3.m` de manière à estimer les paramètres θ^* et ρ^* en suivant ce procédé.

Exercice 4 : estimation de $D_\perp$ par résolution d'un système linéaire

Le critère $\mathcal{I}(\theta) = \sum_{i=1}^{n} (x'_i \cos\theta + y'_i \sin\theta)^2$ à minimiser dans (9) s'appelle l'*inertie*. Il s'écrit également :

$$\mathcal{I}(\theta) = \|CY\|^2 \quad \text{, avec } C = \begin{bmatrix} x'_1 & \dots & x'_n \\ y'_1 & \dots & y'_n \end{bmatrix}^\top \text{ et } Y = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \tag{10}$$

Or, la solution approchée du système linéaire $CY = O$, au sens des moindres carrés ordinaires, vaut $C^+O = O$. Pour éviter cette solution, on impose la contrainte $\|Y\| = 1$ (résolution approchée au sens des *moindres carrés totaux*). Ce nouveau problème se résout en introduisant le *lagrangien* $\mathcal{L}(Y, \lambda) = \|CY\|^2 + \lambda\,(1 - \|Y\|^2)$, où λ est un *multiplicateur de Lagrange*. La condition d'optimalité de $\mathcal{L}$ s'écrit :

$$\nabla\mathcal{L}(Y, \lambda) = 0 \quad\Longleftrightarrow\quad \left\{\begin{array}{l} \nabla_Y \mathcal{L}(Y, \lambda) = 0 \\ \nabla_\lambda \mathcal{L}(Y, \lambda) = 0 \end{array}\right. \quad\Longleftrightarrow\quad \left\{\begin{array}{l} C^\top C Y = \lambda Y \\ \|Y\| = 1 \end{array}\right. \tag{11}$$

Sachant que $C^\top C$ est symétrique réelle, cette matrice admet une base orthonormée de vecteurs propres associés à des valeurs propres $\geqslant 0$, car $C^\top C$ est *semi-définie positive*. Le minimiseur de $\mathcal{I}(\theta)$ de norme 1, noté Y^*, est donc le vecteur propre associé à la plus petite valeur propre de $C^\top C$. En effet, pour un vecteur propre Y_p de norme 1, associé à la valeur propre λ_p : $\|CY_p\|^2 = Y_p^\top C^\top C Y_p = \lambda_p Y_p^\top Y_p = \lambda_p$.

Complétez le script `exercice_4.m` de manière à comparer cette méthode d'estimation de $D_\perp$ avec celle de l'exercice 3. Observez l'évolution des résultats pour différentes valeurs de `nb_tirages` et de n.

Enfin, affichez sur une même figure les droites de régression D_{YX} et $D_\perp$: que constatez-vous ?