

I-2.2 Les outils mathématiques

Vous retrouverez ci-dessous les éléments mathématiques ou probabilistes nécessaires pour comprendre cette partie du cours : deux inégalités, la loi faible des grands nombres et quelques faits sur les fonctions convexes, qui interviennent dans la propriété de convergence des moments empiriques, dans l'étude de la limite de la log-vraisemblance et dans la positivité de la divergence de Kullback-Leibler. La fin de cette partie rappelle où trouver, au chapitre 1, les résultats qui sont utilisés sans être redémontrés, et donne la définition d'un intervalle de confiance utilisée dans l'[Exercice 1 \(PC2\)](#).

Inégalités de Markov et de Bienaymé-Tchebychev

Ces deux inégalités bornent la probabilité qu'une variable aléatoire s'écarte de son espérance, à l'aide de ses moments.

Proposition I-2.2.1: Inégalités de Markov et de Bienaymé-Tchebychev

- (i) (Markov) Soit Y une variable aléatoire réelle positive. Pour tout $\delta > 0$, $\mathbb{P}(Y \geq \delta) \leq \mathbb{E}[Y]/\delta$.
- (ii) (Bienaymé-Tchebychev) Soit Z une variable aléatoire réelle telle que $\mathbb{E}[Z^2] < \infty$, et notons $\mu = \mathbb{E}[Z]$. Pour tout $\delta > 0$,

$$\mathbb{P}(|Z - \mu| \geq \delta) \leq \frac{\text{Var}(Z)}{\delta^2}.$$

Démonstration

(i) On a $\delta \mathbb{1}_{\{Y \geq \delta\}} \leq Y$ en tout point ; en prenant l'espérance, $\delta \mathbb{P}(Y \geq \delta) \leq \mathbb{E}[Y]$. (ii) On applique (i) à la variable positive $Y = (Z - \mu)^2$ et au seuil δ^2 : $\mathbb{P}(|Z - \mu| \geq \delta) = \mathbb{P}((Z - \mu)^2 \geq \delta^2) \leq \mathbb{E}[(Z - \mu)^2]/\delta^2$. $\square$

Appliquée à une moyenne empirique $\bar{Y}_n = n^{-1} \sum_{i=1}^n Y_i$ de variables i.i.d. de variance σ^2 , dont la variance vaut σ^2/n , l'inégalité de Bienaymé-Tchebychev donne $\mathbb{P}(|\bar{Y}_n - \mathbb{E}[Y_1]| \geq \delta) \leq \sigma^2/(n\delta^2)$: c'est la forme utilisée dans la [Proposition I-2.1.1](#). La borne n'est pas précise, elle peut même dépasser 1, mais elle a l'avantage d'être explicite et de ne demander aucune hypothèse sur la loi, hormis l'existence d'un moment d'ordre 2.

Convergence en probabilité et loi faible des grands nombres

La convergence en probabilité est le mode de convergence dans lequel s'expriment les résultats de ce chapitre ; elle sera reprise et comparée aux autres modes de convergence dans la suite du cours.

Définition I-2.2.1: Convergence en probabilité

Soient $(Y_n)_{n \in \mathbb{N}}$ une suite de vecteurs aléatoires à valeurs dans $\mathbb{R}^k$ et $y \in \mathbb{R}^k$. On dit que Y_n converge en probabilité vers y , et l'on note $Y_n \xrightarrow{\mathbb{P}-\text{prob}} y$, si, pour tout $\varepsilon > 0$, $\lim_{n \rightarrow \infty} \mathbb{P}(\|Y_n - y\| \geq \varepsilon) = 0$. Lorsque la probabilité est $\mathbb{P}_\theta$, on écrit $Y_n \xrightarrow{\mathbb{P}_\theta-\text{prob}} y$.

Théorème I-2.2.1: Loi faible des grands nombres

Soit $(Y_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles indépendantes, de même loi et intégrables. Alors

$$\frac{1}{n} \sum_{i=1}^n Y_i \xrightarrow{\mathbb{P}\text{-prob}} \mathbb{E}[Y_1] .$$

Démonstration

Lorsque $\mathbb{E}[Y_1^2] < \infty$, c'est une conséquence directe de l'inégalité de Bienaymé-Tchebychev : la moyenne empirique a pour espérance $\mathbb{E}[Y_1]$ et, par indépendance, pour variance $\text{Var}(Y_1)/n$, de sorte que $\mathbb{P}(|n^{-1} \sum_{i=1}^n Y_i - \mathbb{E}[Y_1]| \geq \varepsilon) \leq \text{Var}(Y_1)/(n\varepsilon^2)$, qui tend vers 0. Le cas d'une variable seulement intégrable demande un argument de troncature, ou le théorème de continuité de Lévy ; il est démontré dans l'annexe du polycopié, au chapitre sur les convergences (théorème « loi faible des grands nombres pour des variables i.i.d. intégrables »). $\square$

Le lemme suivant permet de transporter une convergence en probabilité par une fonction continue ; il sert à passer de la convergence de $e(\hat{\theta}_n)$ à celle de $\hat{\theta}_n$.

Lemme I-2.2.1: Convergence en probabilité et fonction continue

Soient $D \subseteq \mathbb{R}^k$, $h : D \rightarrow \mathbb{R}^m$ une fonction, $y \in D$ un point où h est continue, et (Y_n) une suite de vecteurs aléatoires presque sûrement à valeurs dans D . Si $Y_n \xrightarrow{\mathbb{P}\text{-prob}} y$, alors $h(Y_n) \xrightarrow{\mathbb{P}\text{-prob}} h(y)$.

Démonstration

Soit $\varepsilon > 0$. Par continuité de h en y , il existe $\eta > 0$ tel que, pour tout $z \in D$ vérifiant $\|z - y\| < \eta$, on ait $\|h(z) - h(y)\| < \varepsilon$. Ainsi, à un événement négligeable près, $\{\|h(Y_n) - h(y)\| \geq \varepsilon\} \subseteq \{\|Y_n - y\| \geq \eta\}$, et $\mathbb{P}(\|h(Y_n) - h(y)\| \geq \varepsilon) \leq \mathbb{P}(\|Y_n - y\| \geq \eta)$, qui tend vers 0. $\square$

Fonctions convexes et inégalité de Jensen

L'inégalité de Jensen compare l'image de l'espérance et l'espérance de l'image par une fonction convexe ; c'est l'ingrédient de la positivité de la divergence de Kullback-Leibler. Elle repose sur l'existence, en tout point du graphe d'une fonction convexe, d'une droite d'appui.

Lemme I-2.2.2: Droite d'appui d'une fonction convexe

Soient I un intervalle ouvert de $\mathbb{R}$ et $\phi : I \rightarrow \mathbb{R}$ une fonction convexe. Pour tout $t \in I$, il existe une constante $c_t \in \mathbb{R}$ telle que

$$\phi(t) + c_t(y - t) \leq \phi(y) , \quad \text{pour tout } y \in I .$$

Si ϕ est strictement convexe, l'inégalité est stricte pour tout $y \neq t$. Si ϕ est dérivable en t , on peut prendre $c_t = \phi'(t)$.

Démonstration

Pour $s < t < u$ dans I , écrivons $t = \lambda s + (1 - \lambda)u$ avec $\lambda = (u - t)/(u - s) \in]0, 1[$. La convexité donne $\phi(t) \leq \lambda\phi(s) + (1 - \lambda)\phi(u)$, ce qui se réécrit, après réarrangement,

$$\frac{\phi(t) - \phi(s)}{t - s} \leq \frac{\phi(u) - \phi(t)}{u - t} .$$

Toute pente de corde issue de t vers la gauche est donc inférieure à toute pente de corde issue de t vers la droite ; le même argument montre que la pente à droite $u \mapsto (\phi(u) - \phi(t))/(u - t)$ est croissante sur $]t, +\infty[$ et que la pente à gauche $s \mapsto (\phi(t) - \phi(s))/(t - s)$ est croissante sur $] -\infty, t[$. Posons $c_t := \inf \{(\phi(u) - \phi(t))/(u - t) : u \in I, u > t\}$, qui est un réel, puisque cet ensemble est non vide, I étant ouvert, et minoré par n'importe quelle pente à gauche. Pour $u > t$, la définition de c_t donne $\phi(u) \geq \phi(t) + c_t(u - t)$; pour $s < t$, la pente $(\phi(t) - \phi(s))/(t - s)$ minore toutes les pentes à droite, donc leur borne inférieure c_t , ce qui donne $\phi(s) \geq \phi(t) + c_t(s - t)$. Si ϕ est dérivable en t , la borne inférieure c_t de la fonction croissante $u \mapsto (\phi(u) - \phi(t))/(u - t)$ est sa limite en t , c'est-à-dire $\phi'(t)$. Enfin, si ϕ est strictement convexe, les inégalités entre pentes sont strictes : pour $u > t$, la pente en u est strictement supérieure à la pente en tout point de $]t, u[$, elle-même supérieure ou égale à c_t ; pour $s < t$, la pente en s est strictement inférieure à la pente en tout point de $]s, t[$, elle-même inférieure ou égale à c_t . Les inégalités sont donc strictes pour $y \neq t$. $\square$

Théorème I-2.2.2: Inégalité de Jensen

Soient I un intervalle ouvert de $\mathbb{R}$, $\phi : I \rightarrow \mathbb{R}$ une fonction convexe et Y une variable aléatoire réelle telle que $\mathbb{P}(Y \in I) = 1$ et $\mathbb{E}[|Y|] < \infty$. Alors $\mathbb{E}[Y] \in I$, l'espérance $\mathbb{E}[\phi(Y)]$ est bien définie dans $] -\infty, +\infty[$, et

$$\phi(\mathbb{E}[Y]) \leq \mathbb{E}[\phi(Y)] .$$

Si la fonction ϕ est strictement convexe, l'inégalité est stricte dès que la variable aléatoire Y n'est pas presque sûrement constante, c'est-à-dire dès que $\mathbb{P}(Y \neq \mathbb{E}[Y]) > 0$.

Démonstration

Notons $t := \mathbb{E}[Y]$. Si a est l'extrémité gauche de I , avec $a > -\infty$, la variable $Y - a$ est strictement positive presque sûrement, donc $\mathbb{E}[Y - a] > 0$ (une variable aléatoire positive d'espérance nulle est nulle presque sûrement) : $t > a$. De même pour l'extrémité droite, et $t \in I$. Le [Lemme I-2.2.2](#) fournit une constante c telle que $\phi(t) + c(y - t) \leq \phi(y)$ pour tout $y \in I$, donc

$$\phi(t) + c(Y - t) \leq \phi(Y) \quad \text{presque sûrement.}$$

Le membre de gauche est une variable aléatoire intégrable : la partie négative de $\phi(Y)$ est donc intégrable, ce qui donne un sens à $\mathbb{E}[\phi(Y)] \in] -\infty, +\infty[$, et, l'espérance étant croissante, $\phi(t) = \mathbb{E}[\phi(t) + c(Y - t)] \leq \mathbb{E}[\phi(Y)]$. Si ϕ est strictement convexe, l'inégalité du lemme est stricte sur l'événement $\{Y \neq t\}$: la variable aléatoire positive $\phi(Y) - \phi(t) - c(Y - t)$ est strictement positive sur cet événement, qui est de probabilité strictement positive, donc son espérance $\mathbb{E}[\phi(Y)] - \phi(t)$ est strictement positive. $\square$

Le lemme suivant justifie, dans les exemples du chapitre, le passage d'un point critique de la log-vraisemblance à son maximum global.

Lemme I-2.2.3: Maximum d'une fonction strictement concave

Soient I un intervalle ouvert de $\mathbb{R}$ et $\psi : I \rightarrow \mathbb{R}$ une fonction strictement concave, c'est-à-dire telle que $-\psi$ soit strictement convexe. Alors ψ admet au plus un point de maximum sur I . Si ψ est dérivable en un point t avec $\psi'(t) = 0$, alors t est l'unique point de maximum global de ψ . Enfin, si ψ est deux fois dérivable avec $\psi'' < 0$ sur I , elle est strictement concave.

Démonstration

Si $t_1 \neq t_2$ étaient deux points de maximum, la stricte concavité donnerait $\psi\left(\frac{t_1+t_2}{2}\right) > \frac{1}{2}\psi(t_1) + \frac{1}{2}\psi(t_2) = \max \psi$, ce qui est absurde. Si ψ est dérivable en t avec $\psi'(t) = 0$, le [Lemme I-2.2.2](#)

appliqué à $-\psi$, avec $c_t = -\psi'(t) = 0$, donne $-\psi(y) > -\psi(t)$ pour tout $y \neq t$, c'est-à-dire $\psi(y) < \psi(t)$. Enfin, si $\psi'' < 0$ sur I , alors ψ' est strictement décroissante ; pour $s < u$ dans I , $\lambda \in]0, 1[$ et $t = \lambda s + (1-\lambda)u$, l'inégalité des accroissements finis donne $\psi(t) - \psi(s) > \psi'(t)(t-s)$ et $\psi(u) - \psi(t) < \psi'(t)(u-t)$; comme $\lambda(t-s) = (1-\lambda)(u-t)$, en multipliant la première inégalité par λ et la seconde par $1 - \lambda$ puis en soustrayant, on obtient $\psi(t) > \lambda\psi(s) + (1-\lambda)\psi(u)$. $\square$

Résultats du chapitre 1 utilisés dans ce chapitre

Les résultats suivants sont utilisés sans être redémontrés ; ils sont énoncés et démontrés dans les outils mathématiques ou les compléments du chapitre 1.

- *Densité d'un n -échantillon*. Si le modèle d'une observation est dominé par μ avec les densités q_θ , le n -échantillon est dominé par $\mu^{\otimes n}$, avec la densité produit $\prod_{i=1}^n q_\theta(x_i)$ (Section I-1.1.7, équation (I-1.1)) : c'est la fonction dont la vraisemblance est l'évaluation aux observations.
- *Modèle dominé, densités et théorème de Radon-Nikodym* (Section I-1.2 et Section I-1.2) : la mesure dominante n'est pas unique, et deux densités relatives à deux mesures dominantes diffèrent d'un facteur qui ne dépend pas du paramètre (Section I-1.4), ce qui rend l'estimateur du maximum de vraisemblance indépendant de ce choix.
- *Statistiques d'ordre et médiane empirique* (Section I-1.1.5), utilisées pour les modèles uniforme et de Laplace et pour le Z -estimateur de la médiane.
- *Théorème de transfert, fonction muette et changement de variables* (Section I-1.2), pour le calcul des moments exacts et l'écriture de la divergence de Kullback-Leibler comme une espérance.
- *Lois exponentielle et Gamma* : moments $\mathbb{E}_\theta[X_1^k] = k!/\theta^k$ de la loi exponentielle (Section I-1.4), lois Gamma, leurs moments et leur additivité (Section I-1.2), biais et erreur quadratique moyenne de l'estimateur des moments $\hat{\theta}_{n,1}$ (Section I-1.4).
- *Vecteurs gaussiens, loi du χ^2 , théorèmes de Cochran et de Gosset* (Section I-1.2 et Section I-1.1.9), pour la loi des estimateurs du modèle linéaire gaussien de l'Exercice 1 (PC2), et régression linéaire gaussienne à deux coefficients (Section I-1.4).

Intervalle de confiance : définition minimale

Un estimateur ponctuel propose une valeur ; un intervalle de confiance propose un ensemble de valeurs, construit à partir des observations, qui contient la quantité inconnue avec une probabilité garantie. La notion est l'objet du chapitre suivant ; la définition ci-dessous suffit à la question 7 de l'Exercice 1 (PC2).

Définition I-2.2.2: Intervalle de confiance

Soient $(Z, \mathcal{Z}, \{\mathbb{P}_\theta, \theta \in \Theta\})$ un modèle statistique paramétrique, $g : \Theta \rightarrow \mathbb{R}$ une fonction et $\alpha \in]0, 1[$. Un *intervalle de confiance de niveau $1 - \alpha$* pour $g(\theta)$ est un intervalle aléatoire $[A(Z), B(Z)]$, dont les bornes A et B sont des statistiques à valeurs dans $\mathbb{R} \cup \{-\infty, +\infty\}$, tel que

$$\mathbb{P}_\theta(A(Z) \leq g(\theta) \leq B(Z)) \geq 1 - \alpha \quad \text{pour tout } \theta \in \Theta.$$

La probabilité porte sur l'intervalle, qui est aléatoire, et non sur $g(\theta)$, qui est un nombre inconnu mais fixé : l'énoncé garantit que, quelle que soit la valeur du paramètre, la procédure qui produit l'intervalle capture $g(\theta)$ avec une probabilité au moins égale à $1 - \alpha$. Le procédé le plus courant consiste à trouver une variable aléatoire, fonction des observations et de $g(\theta)$, dont la loi sous $\mathbb{P}_\theta$ ne dépend pas de θ ; les quantiles d'ordre $\alpha/2$ et $1 - \alpha/2$ de cette loi l'encadrent avec probabilité $1 - \alpha$, et l'on obtient l'intervalle en résolvant cet encadrement par rapport à $g(\theta)$. Dans l'Exercice 1 (PC2), la variable $n\hat{\sigma}^2/\sigma^2$, de loi $\chi^2(n-2)$ sous $\mathbb{P}_\theta$ quelle que soit la valeur du paramètre, joue ce rôle pour $g(\theta) = \sigma^2$.