Chapitre 10

Convergence d'estimateurs

Au chapitre précèdent, nous avons introduit la notion d'estimateur $\hat{\theta}_n = \theta_n(X_1, \ldots, X_n)$ de la valeur θ du paramètre caractérisant la loi des $X_i \rightsquigarrow \mathcal{L}(\theta)$, et nous avons donné une stratégie (le *maximum de vraissemblance*) pour former des fonctions θ_n et, partant, suggérer des estimateurs $\hat{\theta}_n$. Il convient à présent de déterminer dans quelle mesure un estimateur $\hat{\theta}_n$ converge effectivement vers la (bonne) valeur θ. Comme $\hat{\theta}_n$ est une v.a., l'écart entre $\hat{\theta}_n$ et θ dépend à la fois de $\omega \in \Omega$ (le "hasard") et de n (l'"asymptotique"). Conformément à l'usage en analyse classique, nous exprimerons que $\hat{\theta}_n$ est proche de θ pour n grand en définissant ce que nous entendons par $\theta = \lim_{n\to+\infty} \hat{\theta}_n$. Voici deux manière, non equivalentes, d'exprimer qu'une v.a. $\overline{Y}$ est la limite d'une suite de v.a. Y_n.

10.1 Convergence d'une suite de v.a.

10.1.1 Convergence en probabilité

Définition : On dit qu'une suite de v.a. $(Y_n)_{n\geq 0}$ *converge en probabilité vers la v.a.* $\overline{Y}$, et on note $Y_n \xrightarrow{P} \overline{Y}$ si et seulement si $\forall \delta > 0$, $\lim_{n\to+\infty} \mathbb{P}(\{|Y_n - \overline{Y}| > \delta\}) = 0$.
On dit que l'estimateur $\hat{\theta}_n$ de θ est *consistent* si et seulement si $\hat{\theta}_n \xrightarrow{P} \theta$.

10.1.2 Convergence dans L^q

Rappelons que l'inégalité de Markov (6.1) assure que $\mathbb{P}(\{|X| \geq \delta\}) \leq \frac{1}{\varphi(\delta)}\mathbb{E}(\varphi(|X|))$ pour n'importe quelle fonction croissante $\varphi : \mathbb{R}_*^+ \longrightarrow \mathbb{R}_*^+$. Une stratégie pour montrer la convergence en probabilité des Y_n vers $\overline{Y}$ est alors de trouver une fonction φ telle qu'on puisse s'assurer que $\lim_{n\to+\infty} \mathbb{E}(\varphi(|Y_n - \overline{Y}|))=0$. Lorsqu'on peut choisir $\varphi(x) = x^q$ avec $q \geq 1$, on dit que les Y_n convergent *dans* L^q vers $\overline{Y}$:

Définition : Soit $q \geq 1$; on dit qu'une suite de v.a. $(Y_n)_{n\geq 0}$ *converge dans* L^q *vers la v.a.* $\overline{Y}$, et on note $Y_n \xrightarrow{L^q} \overline{Y}$ si et seulement si $\lim_{n\to+\infty} \mathbb{E}(|Y_n - \overline{Y}|^q) = 0$.

Nous venons de remarquer que l'inégalité de Markov montre qu'on a

Proposition 10.1 *Si $Y_n \xrightarrow{L^q} \overline{Y}$, alors $Y_n \xrightarrow{P} \overline{Y}$.*

Nous voyons que la convergence en probabilité ne dit rien des valeurs de $Y_n - \overline{Y}$ lorsque $|Y_n - \overline{Y}| > \delta$; ces valeurs peuvent donc être très grandes, quand bien même elles seraient de faible probabilité. Il est donc douteux que cette proposition admette une réciproque. Voici un exemple montrant que la réciproque de la proposition (10.1) est effectivement fausse en général.

Contre-exemple à la réciproque

Considérons une situation où $\overline{Y} = 0$; soit $q \geq 1$ quelconque fixé. Choisissons $Y_n := nZ_n$, avec $Z_n \rightsquigarrow \mathcal{B}(1, p_n)$. Observons que, comme $Z_n \rightsquigarrow \mathcal{B}(1, p_n)$, Z_n ne prend que les valeurs 0 et 1, et donc $|Z_n|^q = Z_n$. On a donc

$$\mathbb{P}(\{|Y_n - \overline{Y}|^q > \delta\}) = \mathbb{P}(\{n^q Z_n > \delta\}) = \mathbb{P}\left(\left\{Z_n > \frac{\delta}{n^q}\right\}\right) \stackrel{\sharp}{=} \mathbb{P}(\{Z_n = 1\}) = p_n,$$

où ($\sharp$) est vrai dès que $\frac{\delta}{n^q} \leq 1$.

Nous avons donc convergence en probabilité des Y_n vers $\overline{Y} = 0$ dès lors qu'on suppose que $\lim p_n = 0$.

En revanche $\mathbb{E}(|Y_n - \overline{Y}|^q) = \mathbb{E}(n^q Z_n) = n^q p_n$ qui sera constamment égale à 1 (et ne convergera donc pas vers 0) si l'on pose $p_n = \frac{1}{n^q}$ qui tend bien vers 0, ce qui, comme nous l'avons vu, assure la convergence en probabilité. Nous avons donc bien formé ainsi un exemple de suite $(Y_n)_{n\geq 1}$ qui converge en probabilité vers $\overline{Y} = 0$ et ne converge pas dans L^q.

10.1.3 Cas L^2 : convergence en moyenne quadratique

Approfondissons le cas important où $q = 2$ déjà abordé au paragraphe 6.3.4 ; c'est le cas qui est le plus commode du point de vue des calculs, et c'est celui où la variabilité est mesurée par la variance σ^2 qui nous est à présent familière. Un peu de terminologie :

Définition : On dit que l'estimateur $\hat{\theta}_n$ *converge en moyenne quadratique vers* θ si et seulement si $\hat{\theta}_n \xrightarrow{L^2} \theta$. On dit aussi que *l'erreur quadratique* $(\mathbb{E}(|\hat{\theta}_n - \theta|^2))^{\frac{1}{2}}$ tends vers 0.

Remarquons tout d'abord que l'identité du Huygens assure que

$$\mathbb{E}\left(|\hat{\theta}_n - \theta|^2\right) = \mathrm{Var}\,(\hat{\theta}_n) + \left(\theta - \mathbb{E}(\hat{\theta}_n)\right)^2. \tag{10.1}$$

Par cette identité, nous voyons que pour assurer la convergence en moyenne quadratique de $\hat{\theta}_n$ vers θ il faut et il suffit qu'à la fois $\lim_{n\to+\infty} \mathrm{Var}\,(\hat{\theta}_n) = 0$ et $\lim_{n\to+\infty} \mathbb{E}(\hat{\theta}_n) = \theta$.

Exemple : Nous avons déjà souligné que la Loi des Grands Nombres assure, pour toute suite $(X_n)_{n\geq 0}$ de v.a. i.i.d. d'espérance $\mu := \mathbb{E}(X_i)$ et de variance $\sigma^2 := \mathrm{Var}\,(X_i)$, la convergence en probabilité de $\hat{\theta} := M_n := \frac{1}{n}(X_1 + \ldots + X_n)$ vers l'espérance commune $\mu := \mathbb{E}(X_i)$. Montrons que nous avons aussi convergence dans L^2. En vertu de (10.1) il suffit que nous vérifions à la fois que $\lim_{n\to+\infty} \mathbb{E}(M_n) = \mu$ et $\lim_{n\to+\infty} \mathrm{Var}\,(M_n) = 0$. On a clairement $\mathbb{E}(M_n) = \mu$ pour tout n ; de plus

$$\mathrm{Var}\,(M_n) = \mathrm{Var}\,\left(\frac{1}{n}(X_1 + \ldots + X_n)\right) = \frac{1}{n^2}(\mathrm{Var}\,(X_1) + \ldots + \mathrm{Var}\,(X_n)) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}\ ;$$

donc $\hat{\theta} := M_n$ converge bien, dans L^2, vers μ.

10.2 Biais d'un estimateur

Nous venons de voir que pour l'estimateur $\hat{\theta}_n := M_n$ de μ, on a $\mathbb{E}(\hat{\theta}_n) = \mu$; on dit que la moyenne M_n est un estimateur *sans biais* de l'espérance μ ; plus généralement :

Définition : On appelle *biais* d'un estimateur $\hat{\theta}_n$ de θ le nombre $\theta - \mathbb{E}(\hat{\theta}_n)$. On dit que l'estimateur $\hat{\theta}_n$ de θ est *sans biais* si et seulement si son biais $\theta - \mathbb{E}(\hat{\theta}_n)$ est nul pour tout n.

Remarque : Au vu de (10.1) on comprend l'intérêt, pour réduire l'erreur en moyenne quadratique $\mathbb{E}(|\hat{\theta}_n - \theta|^2)$ de l'estimateur $\hat{\theta}_n$ de θ, de préférer des estimateurs sans biais. Toutefois on peut préférer un estimateur avec biais dans le cas où le choix d'un tel estimateur avec biais permet d'avoir une convergence plus rapide de $\mathrm{Var}\,(\hat{\theta}_n)$ vers 0.

Au chapitre 9 nous avons vu que, pour la loi $\mathcal{L}(a,b) := \mathcal{U}[a,b]$, $\hat{a} := \mathrm{Min}\,\{X_1, \ldots, X_n\}$ est un estimateur consistent de a et $\hat{b} := \mathrm{Max}\,\{X_1, \ldots, X_n\}$ est un estimateur consistent de b. Calculons le biais $b - \mathbb{E}(\mathrm{Max}\,\{X_1, \ldots, X_n\})$ de $\hat{b}$.

Nous avons vu que $F(x) := \mathbb{P}(\{\mathrm{Max}\,\{X_1, \ldots, X_n\} \leq x\}) = \left(\frac{x-a}{b-a}\right)^n$ pour $x \in [a,b]$. Donc, pour $x \in]a,b[$, la densité f est donnée par $f(x) = F'(x) = \frac{n}{(b-a)^n}(x-a)^{n-1}$, d'où, en intégrant par parties

$$\begin{aligned}
\mathbb{E}(\mathrm{Max}\,\{X_1, \ldots, X_n\}) &= \int_a^b x f(x)dx = \frac{1}{(b-a)^n}\int_a x n(x-a)^{n-1}dx \\
&= \frac{1}{(b-a)^n}\left([x(x-a)^n]_{x=a}^b - \int_a (x-a)^n dx\right) \\
&= \frac{1}{(b-a)^n}\left(b(b-a)^n - \frac{1}{n+1}[(x-a)^{n+1}]_{x=a}^b\right) \\
&= \frac{1}{(n+1)(b-a)^n}[(n+1)b(b-a)^n - (b-a)^{n+1}] = \frac{1}{n+1}(nb+a).
\end{aligned}$$

Finalement $\boxed{\overline{b} := \mathbb{E}(\mathrm{Max}\{X_1, \ldots, X_n\}) = b - \frac{b-a}{n+1}}$.

On montrerait de même que $\boxed{\overline{a} := \mathbb{E}(\mathrm{Min}\{X_1, \ldots, X_n\}) = a + \frac{b-a}{n+1}}$. Nous voyons donc que le biais $(a, b) - \hat{\theta}_n$ de cet estimateur est égal $(\frac{b-a}{n+1}, -\frac{b-a}{n+1})$.

10.2.1 Elimination du biais d'un estimateur

Supposons que $a = 0$ soit connu et que seul b soit à estimer. On vérifie aisément que le biais est inchangé : $b - \mathbb{E}(\mathrm{Max}\{X_1, \ldots, X_n\}) = \frac{b}{n+1}$, et donc $\mathbb{E}(\mathrm{Max}\{X_1, \ldots, X_n\}) = \frac{n}{n+1}b$, d'où encore $b = \frac{n+1}{n}\mathbb{E}(\mathrm{Max}\{X_1, \ldots, X_n\}) = \mathbb{E}(\frac{n+1}{n}\mathrm{Max}\{X_1, \ldots, X_n\})$, d'où l'idée de considérer le nouvel estimateur $\boxed{\hat{b}' := \frac{n+1}{n}\mathrm{Max}\{X_1, \ldots, X_n\}.}$

Compte tenu de ce qui précède, $\hat{b}'$ est, lui, sans biais.

Exercice : On ne suppose plus connue la valeur de a.

1. Trouver $\alpha(n)$ et $\beta(n)$ tel que $\hat{b}'' := \alpha(n)\hat{a} + \beta(n)\hat{b}$ soit un estimateur sans biais de b.
2. Calculer $\mathrm{Var}(\hat{b})$ et en déduire $\mathrm{Var}(\hat{b}')$.
3. En déduire que les estimateurs $\hat{b}$ et $\hat{b}'$ sont des estimateurs consistents de b et qu'ils convergent en moyenne quadratique.

.