Théorème d'approximation de Weierstrass par les polynômes de Bernstein

Référence : Zuily-Queffélec : Analyse pour l'agrégation p. 518

Théorème (Théorème de Bernstein - 1912)

Soit $f : [0,1] \to \mathbb{C}$ une fonction continue. On pose

$$\omega(h) := \sup\{|f(u) - f(v)|, |u - v| \leqslant h\}$$

son module de continuité [1] uniforme.
On considère

$$B_n(f,x) = B_n(x) := \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} f\left(\frac{k}{n}\right)$$

le $n^{\text{ième}}$ polynôme de Bernstein de f. Alors

(1) (B_n) converge uniformément vers f sur $[0,1]$.

(2) $\|f - B_n\|_\infty \leqslant C\omega\left(\frac{1}{\sqrt{n}}\right)$ où C est une constante.

(3) L'estimation (2) est optimale : il existe f lipschitzienne telle que $\|f - B_n\|_\infty \geqslant \frac{\delta}{\sqrt{n}}$ pour une constante $\delta > 0$.

Preuve :

(1) Soit $x \in [0,1]$ et $(X_i)_{i\geqslant 1}$ une suite de variables de Bernoulli i.i.d de paramètre x. On note $S_n := X_1 + \cdots + X_n$. Ainsi, S_n suit une loi binomiale [2] et on a alors :

$$\mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right] \overset{3}{=} \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} f\left(\frac{k}{n}\right) = B_n(x) \text{ et } \mathbb{E}[f(x)] = f(x) \text{ (constante, pas d'aléatoire)}$$

Heuristiquement, comme $\frac{S_n}{n} \xrightarrow{\mathbb{P}} x$ [4], on aimerait que $B_n(x)$ soit "proche" de $f(x)$.
Mais on a, d'après précédemment :

$$|f(x) - B_n(x)| = \left|\mathbb{E}\left[f(x) - f\left(\frac{S_n}{n}\right)\right]\right| \leqslant \mathbb{E}\left[\left|f(x) - f\left(\frac{S_n}{n}\right)\right|\right]. \tag{0.1}$$

f est continue sur $[0,1]$ compact donc uniformément continue par le théorème de Heine, donc $\omega(\delta)$ est défini pour tout $\delta > 0$ et $\lim_{\delta \to 0} \omega(\delta) = 0$. On a donc que si $\left|x - \frac{S_n}{n}\right| \leqslant \delta$, $\left|f(x) - f\left(\frac{S_n}{n}\right)\right| \leqslant \omega(\delta)$.

1. Un module de continuité est une application croissante non nulle g de $\mathbb{R}^+$ dans $\mathbb{R}^+$ tq $g(0) = 0$ et pour tout t, t' : $g(t + t') \leqslant g(t) + g(t')$.
2. Note pour la suite : comme elle est à support fini, elle admet des moments de tout ordre.
3. Si X est une variable aléatoire discrète à valeurs dans un ensemble dénombrable $S \subset F$, alors : $\mathbb{E}\big[\varphi(X)\big] = \sum_{x \in S} \varphi(x) \mathbb{P}_X(\{x\})$.

C'est le théorème de transfert. Logiquement les hypothèses du théorème de transfert imposent que la fonction soit à valeurs réelles. Mais ça ne gène pas qu'elle soit complexe, au pire on fait Re+iIm. On n'a pas non plus de problème d'intégrabilité car on a un variable aléatoire discrète sur un espace fini.

4. C'est la LGN faible. Une Bernoulli a un moment d'ordre 2.

Par ailleurs, $\left|f(x) - f\left(\frac{S_n}{n}\right)\right| \leqslant 2\|f\|_\infty$ donc, pour $\delta > 0$,

$$\mathbb{E}\left[\left|f(x) - f\left(\frac{S_n}{n}\right)\right|\right] \leqslant \omega(\delta)\underbrace{\mathbb{P}\left(\left|x - \frac{S_n}{n}\right| \leqslant \delta\right)}_{\leqslant 1} + 2\|f\|_\infty \mathbb{P}\left(\left|x - \frac{S_n}{n}\right| > \delta\right).$$

Mais, par l'inégalité de Tchebychev [5], on a

$$\begin{aligned}\mathbb{P}\left(\left|\frac{S_n}{n} - x\right| > \delta\right) &\leqslant \frac{\mathrm{Var}\left(\frac{S_n}{n}\right)}{\delta^2} = \frac{\mathrm{Var}(S_n)}{n^2\delta^2} = \frac{nx(1-x)}{n\delta^2} \\ &\leqslant \frac{1}{4n\delta^2}.\end{aligned}$$

On en déduit que, pour tout $x \in [0,1]$,

$$|f(x) - B_n(x)| \leqslant \omega(\delta) + \frac{\|f\|_\infty}{2n\delta^2},$$

d'où

$$\limsup_{n\to+\infty} \|f - B_n\|_\infty \leqslant \omega(\delta).$$

Comme $\lim_{\delta\to 0} \omega(\delta) = 0$, $\lim_{n\to+\infty} \|f - B_n\|_\infty = 0$.

(2) Montrons déjà que $\omega(\lambda h) \leqslant (\lambda+1)\omega(h)$:
D'après le **Lemme** et par récurrence, on a $\omega(nh) \leqslant n\omega(h)$ pour $n \in \mathbb{N}$. De plus, ω est croissante donc, pour $\lambda \in \mathbb{R}, \omega(\lambda h) \leqslant \omega(\lceil\lambda\rceil h) \leqslant \lceil\lambda\rceil\omega(h) \leqslant (\lambda+1)\omega(h)$.
Ici, $\omega\left(\left|x - \frac{S_n}{n}\right|\right) = \omega\left(\frac{1}{\sqrt{n}}\sqrt{n}\left|x - \frac{S_n}{n}\right|\right) \leqslant \left(\sqrt{n}\left|x - \frac{S_n}{n}\right| + 1\right)\omega\left(\frac{1}{\sqrt{n}}\right)$.
Déjà, par définition de ω,

$$\mathbb{E}\left[\left|f(x) - f\left(\frac{S_n}{n}\right)\right|\right] \leqslant \mathbb{E}\left[\omega\left(\left|x - \frac{S_n}{n}\right|\right)\right].$$

On en déduit, avec (0.1) et [6]

$$\begin{aligned}|f(x) - B_n(x)| &\leqslant \omega\left(\frac{1}{\sqrt{n}}\right)\mathbb{E}\left[\sqrt{n}\left|x - \frac{S_n}{n}\right| + 1\right] = \omega\left(\frac{1}{\sqrt{n}}\right)\left(1 + \sqrt{n}\left\|x - \frac{S_n}{n}\right\|_1\right) \\ &\leqslant \omega\left(\frac{1}{\sqrt{n}}\right)\left(1 + \sqrt{n}\left\|x - \frac{S_n}{n}\right\|_2\right)\end{aligned}$$

par décroissance des $L^p(\mathbb{P})$ (ou inégalité de Hölder). Or

$$\begin{aligned}\left\|x - \frac{S_n}{n}\right\|_2^2 &= \mathbb{E}\left[\left(x - \frac{S_n}{n}\right)^2\right] \\ &= \underbrace{\mathrm{Var}\left(x - \frac{S_n}{n}\right)}_{\mathrm{Var}\left(\frac{S_n}{n}\right)} + \left(\mathbb{E}\left[x - \frac{S_n}{n}\right]\right)^2 \\ &= \frac{1}{n^2}nx(1-x) + \underbrace{\left(x - \frac{1}{n}nx\right)^2}_{0} = \frac{x(1-x)}{n}.\end{aligned}$$

D'où

$$\begin{aligned}|f(x) - B_n(x)| &\leqslant \omega\left(\frac{1}{\sqrt{n}}\right)\left(1 + \sqrt{n}\sqrt{\frac{x(1-x)}{n}}\right) = \omega\left(\frac{1}{\sqrt{n}}\right)(1 + \underbrace{\sqrt{x(1-x)}}_{\leqslant\sqrt{\frac{1}{4}} = \frac{1}{2}}) \\ &\leqslant \frac{3}{2}\omega\left(\frac{1}{\sqrt{n}}\right).\end{aligned}$$

D'où $\|f - B_n\|_\infty \leqslant \frac{3}{2}\omega\left(\frac{1}{\sqrt{n}}\right)$.

5. Soit X une v.a. d'espérance μ et de variance finie σ^2. On a, pour tout réel strictement positif t, $\mathbb{P}(|X - \mu| > t) \leqslant \frac{\sigma^2}{t^2}$

6. Si on se met dans $L^1(\mathbb{P})$ alors l'espérance c'est $\|\cdot\|_1$ etc. On peut appliquer ce qu'on connaît sur les L^p classiques, d'autant que ici la "mesure" est finie ($= 1$).

(3) On pose $f : x \mapsto \left|x - \dfrac{1}{2}\right|$, on a $\omega(h) \leqslant h$ [7]. Par ailleurs,

$$\|f - B_n\|_\infty \geqslant \left|f\left(\frac{1}{2}\right) - B_n\left(\frac{1}{2}\right)\right| = \left|B_n\left(\frac{1}{2}\right)\right| = \mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right] = \mathbb{E}\left[\left|\frac{S_n}{n} - \frac{1}{2}\right|\right] = \frac{1}{2n}\mathbb{E}[|2S_n - n|]$$

Tout cela avec la probabilité $x = \dfrac{1}{2}$.

On réécrit $\|f - B_n\|_\infty \geqslant \dfrac{1}{2n}\mathbb{E}\left[|\varepsilon_1 + \cdots + \varepsilon_n|\right]$ avec $\varepsilon_j := 2X_j - 1$ variables de Rademacher i.i.d. D'où [8]

$$\begin{aligned}\|f - B_n\|_\infty &\geqslant \frac{1}{2n}\|\varepsilon_1 + \cdots + \varepsilon_n\|_1 \\ &\geqslant \frac{1}{2n\sqrt{\mathrm{e}}}\|\varepsilon_1 + \cdots + \varepsilon_n\|_2\end{aligned}$$

Or $\|\varepsilon_1 + \cdots + \varepsilon_n\|_2^2 = \underbrace{\mathrm{Var}(\underbrace{\varepsilon_1 + \cdots + \varepsilon_n}_{2S_n - n})}_{4\mathrm{Var}(S_n) = 4n\frac{1}{2}(1-\frac{1}{2}) = n} + (\underbrace{\mathbb{E}[\varepsilon_1 + \cdots + \varepsilon_n]}_{0})^2 = n$. D'où

$$\|f - B_n\|_\infty \geqslant \frac{1}{2\sqrt{n\mathrm{e}}} \geqslant \frac{1}{2\sqrt{\mathrm{e}}}\omega\left(\frac{1}{\sqrt{n}}\right).$$

■

Proposition (Inégalité de Khintchine - 1923 ?)

Soit $\varepsilon_1, \ldots, \varepsilon_n$ des variables de Rademacher (*i.e.* valant ± 1 avec probabilité $\dfrac{1}{2}$) i.i.d. Soit f fonction combinaison linéaire des ε_i. Alors $\|f\|_2 \leqslant \sqrt{e}\mathbb{E}[|f|]$.

Preuve :

On a $f = \sum_j a_j\varepsilon_j$. On sait que $\|f\|_2 = \sum_j a_j^2$ (se calcule facilement avec la définition car $\varepsilon_k^2 = 1$).

De plus, on peut supposer $\|f\|_2^2 = 1$ (quitte à tout rediviser par $\|f\|_2$).

Posons $g := \prod_{j=1}^{n}(1 + \mathrm{i}a_j\varepsilon_j)$. Alors pour presque tout x, et sachant que $1 + u \leqslant \exp(u)$,

$$\begin{aligned}|g(x)| &= \prod_{j=1}^{n}\sqrt{1 + a_j^2\varepsilon_j^2(x)} = \prod_{j=1}^{n}\sqrt{1 + a_j^2} \\ &\leqslant \prod_{j=1}^{n}\sqrt{\exp(a_j^2)} = \sqrt{\exp\left(\sum a_j^2\right)} = \sqrt{e}.\end{aligned}$$

D'où $\|g\|_\infty \leqslant \sqrt{e}$. De plus, si $j \in \{1, \ldots, n\}$,

$$\begin{aligned}\mathbb{E}[\varepsilon_j g] &= \mathbb{E}\left[\varepsilon_j(1 + \mathrm{i}a_j\varepsilon_j)\prod_{k\neq j}(1 + \mathrm{i}a_k\varepsilon_k)\right] \\ &= \mathbb{E}[\varepsilon_j(1 + \mathrm{i}a_j\varepsilon_j)]\mathbb{E}\left[\prod_{k\neq j}(1 + \mathrm{i}a_k\varepsilon_k)\right] \quad \text{par indépendance des } \varepsilon_j \\ &= \mathbb{E}[\varepsilon_j(1 + \mathrm{i}a_j\varepsilon_j)]\prod_{k\neq j}\mathbb{E}[1 + \mathrm{i}a_k\varepsilon_k] \\ &= \mathrm{i}a_j \text{ car } \mathbb{E}(\varepsilon_j) = 0.\end{aligned}$$

7. se montre facilement avec $||a| - |b|| \leqslant |a - b|$.

8. On ne peut pas faire en fait avec normes sont équivalentes sur l'espace de dimension finie n $\mathrm{Vect}(\varepsilon_1, \ldots, \varepsilon_n)$ car la constante que l'on obtient ainsi dépend, a priori, de n alors que l'on cherche une minoration indépendante de n.

Or $\mathbb{E}[fg] = \sum_{j=1}^{n} a_j \mathbb{E}[\varepsilon_j g]$, d'où $|\mathbb{E}[fg]| = \left| \mathrm{i} \sum_{j=1}^{n} a_j^2 \right| = 1$.
Or
$$\|f\|_1 \geqslant \frac{|\mathbb{E}[fg]|}{\|g\|_\infty} \geqslant \frac{1}{\sqrt{e}}.$$
■

Lemme

Pour $\delta_1, \delta_2 \in \mathbb{R}^+$, $\omega(\delta_1+\delta_2) \leqslant \omega(\delta_1) + \omega(\delta_2)$

Preuve :
Soit (t, x) tel que $|t-x| \leqslant \delta_1 + \delta_2$.
Si $|x-t| \leqslant \delta - 1$,
$$|f(x) - f(t)| \leqslant \omega(\delta_1) \leqslant \omega(\delta_1) + \omega(\delta_2)$$
Sinon, supposons $x > t$. Alors
$$|f(t) - f(x)| \leqslant |f(t) - f(t+\delta_1)| + |f(t+\delta_1) - f(x)| \leqslant \omega(\delta_1) + \omega(\delta_2)$$
car $x - (t+\delta_1) = x - t - \delta_1 = |x-t| - \delta_1 \leqslant \delta_1 + \delta_2 - \delta_1 = \delta_1$. Donc $|x - (t+\delta_1)| \leqslant \delta_2$. ■

Réponses à de possibles questions

1. (vrai oral) Préciser lemme de transfert (+embêté avec les probas..)
   $\hookrightarrow$
2. (vrai oral) D'où vient l'idée de la preuve ?
   $\hookrightarrow$

Notes :

✓ **A l'oral**, (1) 6'34 (2) 9'30 (3) 14' à allure normale
✓ Loi Bernoulli : $\mathcal{B}(p)$. Espérance : p. Variance : $p(1-p)$. Fonction caractéristique : $1 - p + p\mathrm{e}^{\mathrm{i}t}$.
✓ Loi Binomiale : $\mathcal{B}(n, p)$. Espérance : np. Variance : $np(1-p)$. Fonction caractéristique : $(1 - p + p\mathrm{e}^{\mathrm{i}t})^n$.
✓ Attention ! ZQ part du principe que le module de continuité uniforme vérifie certaine propriété (comme module de continuité). On les revérifie simplement, sauf celle montrée dans le **Lemme**.
✓ Le théorème a été établi en 1885 par Weierstrass. Stone a considérablement généralisé le théorème en 1937 et en simplifia le preuve en 1948.
♣ Karl Weierstrass (1815 - 1897) -pneumonie- est un mathématicien allemand, lauréat de la médaille Copley en 1895. Il créa avec Enneper une classe complète de paramétrisations. Il est souvent cité comme le "père de l'analyse moderne". Il consolida des travaux de Cauchy sur les nombres irrationnels et leur amena une nouvelle compréhension. Ses travaux les plus connus portent sur les fonctions elliptiques. C'est lui qui le premier rendit public un exemple de fonction continue nulle part dérivable.
♣ Marshall Stone (1903 - 1989) est un mathématicien américain célèbre pour ses contributions en analyse réelle, en analyse fonctionnelle et en théorie des algèbres de Boole.
♣ Attention il y a trois mathématiciens Bernstein :
Joseph (1945), israélien,
Felix (1878 - 1956), allemand, théorème de Cantor-Bernstein,
Sergeï (1880 - 1968) soviétique, auteur de cette démonstration. Sa thèse de doctorat soumise en 1904 à la Sorbonne résout le 19e problème de Hilbert. Ses travaux portent sur l'approximation des fonctions et la théorie des probabilités.
♣ Alexandre Khintchine (1878 - 1959) est un mathématicien russe puis soviétique. Il est principalement connu pour son travail sur la théorie des probabilités. Ses travaux portent notamment en mathématiques sur l'analyse réelle, la théorie des probabilités, la théorie ergodique et les fractions continues, et en physique sur la physique statistique.

✓ Rappel : lim sup, lim inf.
Si $(u_n)_{n\geqslant 0}$ est une suite bornée de réels. Alors $v_n = \sup\{u_k \mid k \geqslant n\}$ est décroissante et $w_n = \inf\{u_k \mid k \geqslant n\}$ est croissante. De plus, comme $w_n \leqslant u_n \leqslant v_n$, ce sont donc des suites convergentes.
On pose : $\limsup\limits_{n\to+\infty} u_n = \lim\limits_{n\to+\infty} v_n = \inf v_n$ et $\liminf\limits_{n\to+\infty} u_n = \lim\limits_{n\to+\infty} w_n \sup w_n$.
On a alors, par ex, $\limsup\limits_{n\to+\infty} u_n = +\infty$ si la suite n'est pas majorée, et $\liminf\limits_{n\to+\infty} u_n = -\infty$ si la suite n'est pas minorée.
Exemples : $\limsup\limits_{n\to+\infty}(-1)^n = 1$, $\liminf\limits_{n\to+\infty}(-1)^n = -1$, $\limsup\limits_{n\to+\infty} \sin n = 1$, $\liminf\limits_{n\to+\infty} \sin n = -1$