Théorème de Weierstrass via les polynômes de Bernstein

Julie Parreaux

2018-2019

Référence du développement : Barde–Ledoux [1, p.89] et Zully–Queffelec [2, p.518].

Leçons où on présente le développement : 260 (Espérance et variance) ; 264 (Variables aléatoires discrètes) .
Leçons où il peut être évoqué : 228 (Continuité et dérivabilité).

1 Introduction

Un célèbre théorème de Weierstrass affirme que toute fonction continue sur le segment $[0, 1]$ est limite uniforme sur ce segment d'une suite de polynômes. Bernstein, à l'aide des polynômes qui porte son nom apporte une nouvelle démonstration à ce théorème obtenu initialement à l'aide de la convolution.

De plus, les polynômes de Bernstein donne une démonstration constructive pour laquelle la convergence de la suite est optimale (on ne le montrera pas durant ce développement). C'est un phénomène bien connu : les polynômes de Bernstein épousent bien les propriétés qualitatives de f (continuité, croissance, convexité, ...), mais converge lentement vers ceux-ci car ils n'oscillent pas assez autour de f comme les polynômes de Tchebychev par exemple.

L'objectif du développement est donc de montrer le théorème suivant :

Théorème (Weierstrass)**.** *Pour tout $f : [0, 1] \to \mathbb{R}$ continue, il existe une suite de polynôme convergeant uniformément vers f sur $[0, 1]$.*

2 Le théorème de Weierstrass

Théorème. *Soit $f : [0, 1] \to \mathbb{R}$ une fonction continue, ω son module de continuité uniforme, i.e. $\omega(h) = \sup\{|f(u) - f(v)|; |u - v| \leq h\}$. Pour $n \leq 1$, on considère le polynôme*

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

, le n^{ime} polynômes de Bernstein de f. Alors :

1. *B_n converge uniformément vers f sur $[0, 1]$;*
2. *Plus précisément, on a $||f - B_n||_\infty \leq C\omega\left(\frac{1}{\sqrt{n}}\right)$.*

Schéma du développement

1. Énoncer quelques propriétés sur le module de continuité.
2. Montrer que B_n est un polynôme en x.
3. Montrer que pour tout x, $|f - B_n|_\infty$ est majoré par $\omega(\delta) + \frac{||f||_\infty}{2n\delta^2}$.
4. Montrer la convergence uniforme de B_n vers f.
5. Montrer que $||f - B_n||_\infty \leq C\omega\left(\frac{1}{\sqrt{n}}\right)$.

Démonstration. Remarquons d'abord que ω est bien définie puisque, selon le théorème de Heine, f est uniformément continue sur $[0,1]$, ce qui assure de plus que $\lim_{h\to 0}\omega(h) = 0$.

Étape 1 : énoncer quelques propriétés sur le module de continuité Commençons par rappeler quelques propriétés (que nous ne montrons pas) sur le module de continuité avant de rentré dans le vif du sujet.

Lemme. *Le module de continuité uniforme de f, ω vérifie les propriétés suivantes :*

1. *ω est croissante.*
2. $\forall h_1, h_2 \in [0,1]$, $\omega(h_1 + h_2) \leq \omega(h_1) + \omega(h_2)$.
3. $\forall h \in [0,1]$, $\lambda \in \mathbb{R}_+$ *tels que* $h\lambda \in [0,1]$, $\omega(\lambda h) \leq (\lambda + 1)\omega(h)$

Montrons maintenant le théorème en commençant par montrer que les polynômes de Bernstein B_n convergent uniformément vers f sur $[0,1]$.

Étape 2 : montrons que B_n est bien un polynôme en x. Soient $x \in [0,1]$, $X_1, \dots X_n$ des variables aléatoires discrètes suivant une loi de Bernoulli $\mathcal{B}(x)$ indépendantes. On pose $S_n = \sum_{i=1}^n X_i$ qui suit une loi binomiale $\mathcal{B}(n,x)$ (car la somme de n variables aléatoires indépendantes suivant une loi de Bernoulli de paramètre p donne une loi binomiale de paramètre (n,p)). Par la formule de transfert, on a alors :

$$\mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right] = \sum_{k=0}^{n} \binom{n}{r} (1-x)^{n-k} f\left(\frac{k}{n}\right) = B_n$$

qui est donc bien un polynôme en x.

Étape 3 : montrons que pour tout x, $|f - B_n|_\infty$ est majoré par $\omega(\delta) + \frac{||f||_\infty}{2n\delta^2}$. Soit $||.||_\infty$ la norme sup sur $[0,1]$ et $\delta \in [0,1]$. On a $f(x) - B_n(x) = \mathbb{E}\left[f(x) - f\left(\frac{S_n}{n}\right)\right]$ (par linéarité de l'espérance : comme $\mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right] = B_n$ et que $f(x)$ est une constante pour l'espérance). Ainsi, on a :

$$\begin{aligned}
|f(x) - B_n(x)| &\leq \mathbb{E}\left[\left|f(x) - f\left(\tfrac{S_n}{n}\right)\right|\right] && \text{(propriétés de l'espérance)}\\
&\leq \mathbb{E}\left[\left|f(x) - f\left(\tfrac{S_n}{n}\right)\right|\mathbb{1}_{|x-\frac{S_n}{n}|\leq\delta} + \left|f(x) - f\left(\tfrac{S_n}{n}\right)\right|\mathbb{1}_{|x-\frac{S_n}{n}|>\delta}\right] && \text{(en coupant en deux)}\\
&\leq \underbrace{\mathbb{E}\left[\left|f(x) - f\left(\frac{S_n}{n}\right)\right|\mathbb{1}_{|x-\frac{S_n}{n}|\leq\delta}\right]}_{\text{A}} + \underbrace{\mathbb{E}\left[\left|f(x) - f\left(\frac{S_n}{n}\right)\right|\mathbb{1}_{|x-\frac{S_n}{n}|>\delta}\right]}_{\text{B}} && \text{(inégalité triangulaire)}
\end{aligned}$$

On cherche donc à majorer A et B. On a, par définition du module de continuité uniforme :

$$\text{A} \leq \omega(\delta)$$

Par une majoration brute de f, on obtient par la linéarité de l'espérance :

$$\text{B} \leq 2||f||_\infty \mathbb{E}\left[\mathbb{1}_{|x-\frac{S_n}{n}|>\delta}\right]$$

Soit

$$\begin{aligned}
|f(x) - B_n(x)| &\leq \omega(\delta) + 2||f||_\infty \mathbb{E}\left[\mathbb{1}_{|x-\frac{S_n}{n}|>\delta}\right] && \text{(par les calculs précédents)}\\
&\leq \omega(\delta) + 2||f||_\infty \mathbb{P}\left(|x - \tfrac{S_n}{n}| > \delta\right) && \text{(par l'espérance d'une indicatrice)}
\end{aligned}$$

On souhaite alors majorer $\mathbb{P}\left(|x - \frac{S_n}{n}| > \delta\right)$. Par l'inégalité de Tchebychev [1, p.60], on obtient :

$$\begin{aligned}
\mathbb{P}\left(|x - \tfrac{S_n}{n}| > \delta\right) &= \mathbb{P}(|S_n - nx| > n\delta) && \text{(par multiplication par } n)\\
&\leq \tfrac{\mathrm{Var}(S_n)}{(n\delta)^2} && \text{(car } \mathbb{E}(S_n) = nx)\\
&= \tfrac{nx(1-x)}{(n\delta)^2} && \text{(car } \mathrm{Var}(S_n) = nx(1-x))\\
&\leq \tfrac{1}{4n\delta^2} && \text{(car } x(1-x) \leq \tfrac{1}{4})
\end{aligned}$$

On obtient alors :

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

Etape 4 : montrons la convergence uniforme de B_n vers f. On a [1, p.60],

$$\sup_{0\leq x\leq 1}|f(x)-B_n(x)|\leq \inf_{\delta>0}\left(\omega(\delta)+\frac{||f||_\infty}{2n\delta^2}\right)$$

On en déduit que [2, p.518] :

$$\lim\sup||f-B_n||_\infty\leq\omega(\delta)$$

Or f est continue sur $[0,1]$ compact, donc f est uniformément continue donc $\lim\limits_{\delta\to 0}\omega(\delta)=0$, on en déduit que $\lim\limits_{n\to+\infty}||f-B_n||_\infty=0$.

On montre qu'on peut contrôler cette convergence. De plus, même si on ne le montre pas ici faute de temps elle est optimale (voir la preuve dans la section 3).

Étape 5 : montrons que $||f-B_n||_\infty\leq C\omega\left(\frac{1}{\sqrt{n}}\right)$. On va maintenant vérifier la convergence de la suite de polynôme.

$$\begin{array}{rcll}
|f(x)-B_n(x)| & \leq & \mathbb{E}\left[|f(x)-f\left(\frac{S_n}{n}\right)|\right] & \text{(par ce qui précède)}\\
& \leq & \mathbb{E}\left[\omega\left(\underbrace{\left|x-\frac{S_n}{n}\right|}_{1\leq}\right)\right] & \text{(def de } \omega \text{ et croissance de l'intégrale)}\\
& \leq & \mathbb{E}\left[\left(\underbrace{\sqrt{n}\left|x-\frac{S_n}{n}\right|+1}_{\in\mathbb{R}_+}\right)\omega\left(\underbrace{\frac{1}{\sqrt{n}}}_{\in[0,1]}\right)\right] & \text{(lemme et croissance de l'intégrale)}\\
& \leq & \omega\left(\frac{1}{\sqrt{n}}\right)\mathbb{E}\left[\left(\sqrt{n}|x-\frac{S_n}{n}|+1\right)\right] & \text{(linéarité de l'intégrale)}\\
& \leq & \omega\left(\frac{1}{\sqrt{n}}\right)\left(\sqrt{n}\mathbb{E}\left[|x-\frac{S_n}{n}|\right]+1\right) & \text{(linéarité de l'intégrale)}
\end{array}$$

Il nous faut donc maintenant majorer $\mathbb{E}\left[|x-\frac{S_n}{n}|\right]$.

$$\begin{array}{rcll}
\mathbb{E}\left[|x-\frac{S_n}{n}|\right] & \leq & \mathbb{E}\left[|x-\frac{S_n}{n}|^2\right]^{1/2}\mathbb{E}\left[1^2\right]^{1/2} & \text{(par l'inégalité de Hölder)}\\
& \leq & \mathrm{Var}\left(\frac{S_n}{n}\right)^{1/2} & \text{(définition de la variance)}\\
& \leq & \left(\frac{1}{4n}\right)^{1/2} & \text{(par le calcul précédent)}\\
& \leq & \frac{1}{2\sqrt{n}} &
\end{array}$$

On a ainsi

$$|f(x)-B_n(x)|\leq\omega\left(\frac{1}{\sqrt{n}}\right)\left(\sqrt{n}\frac{1}{2\sqrt{n}}+1\right)=\frac{3}{2}\omega\left(\frac{1}{\sqrt{n}}\right)$$

D'où le résultat. □

Remarque. On peut étendre le théorème de Weierstrass à tout intervalle fermé borné de $\mathbb{R}$ en appliquant ce résultat à $g(x)=f(a+(b-a)x)$.

Remarque. Une fonction qui est limite uniforme de fonctions polynomiales est une fonction polynomiale.

3 Preuve de l'optimalité

Nous avons étudier la vitesse de convergence de la suite des polynômes vers f. Maintenant, nous allons voir qu'elle est optimale, c'est-à-dire, il existe une fonction lipschitzienne f pour laquelle $||f-B_n||_\infty\geq\frac{\delta}{\sqrt{n}}$ où δ est une constante numérique.

Soit $f=\left|x-\frac{1}{2}\right|$. Par inégalité triangulaire renversée, on a $\omega(h)\leq h$ pour tout h.

Soient $X_1, \dots X_n$ une suite de variables aléatoires de Bernoulli de paramètre $\frac{1}{2}$ indépendante et $\epsilon_j = 2X_j - 1$ est alors une suite de variable de Rademacher. Alors, on a

$$\begin{aligned}
||f(x) - B_n(x)||_\infty &\geq \left| f\left(\tfrac{1}{2}\right) - B_n\left(\tfrac{1}{2}\right) \right| \\
&= \left| B_n\left(\tfrac{1}{2}\right) \right| \\
&= \mathbb{E}\left[\tfrac{S_n}{n} - \tfrac{1}{2}\right] \\
&= \tfrac{1}{2n}\mathbb{E}[2S_n - n] \\
&= \tfrac{1}{2n}\mathbb{E}[\epsilon_1 + \dots \epsilon_n]
\end{aligned}$$

D'où via l'inégalité de Khintchine :

$$\begin{aligned}
||f(x) - B_n(x)||_\infty &\geq \tfrac{1}{2n}||\epsilon_1 + \dots \epsilon_n||_1 \\
&\geq \tfrac{1}{2n}||\epsilon_1 + \dots \epsilon_n||_2 \\
&= \tfrac{1}{2\sqrt{2}}\tfrac{1}{\sqrt{n}} \\
&\geq \tfrac{1}{2\sqrt{2}}\omega\left(\tfrac{1}{\sqrt{n}}\right)
\end{aligned}$$

Ce qui prouve l'optimalité de la convergence.

4 Compléments autours du module de continuité et des probabilités

Module de continuité

Nous définissons le module de continuité [2, p.114] qui nous permet de quantifier à quel point la fonction est continue. Plus la notion de continuité est forte (comme la continuité uniforme vs la continuité simple), plus le module est faible.

Définition. Un module de continuité est une application non nulle $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ telle que $\varphi(0) = 0$; φ est continue croissante et φ est sous-additive, i.e. $\varphi(t_1 + t_2) \leq \varphi(t_1) + \varphi(t_2)$, $\forall t_1, t_2$.

Remarque. Par récurrence montre que : $\varphi(rt) \leq r\varphi(t)$ $\forall r \in \mathbb{N}$ et $\forall t \in \mathbb{R}_+$. On obtient alors $\varphi(\lambda t) \leq (\lambda + 1)\varphi(t)$ $\forall \lambda \geq 0$ et $\forall t \in \mathbb{R}_+$ (s'obtient de la croissance de φ).

Lemme. *Le module de continuité uniforme de f, ω vérifie les propriétés suivantes :*

1. *ω est croissante.*
2. $\forall h_1, h_2 \in [0,1]$, $\omega(h_1 + h_2) \leq \omega(h_1) + \omega(h_2)$.
3. $\forall h \in [0,1]$, $\lambda \in \mathbb{R}_+$ *tels que* $h\lambda \in [0,1]$, $\omega(\lambda h) \leq (\lambda + 1)\omega(h)$

Démonstration. On rappelle que nous avons définie le module de continuité uniforme de f tel que $\omega(h) = \sup\{|f(u) - f(v)|; |u - v| \leq h\}$, $\forall h \in [0,1]$.

Montrons 1 : Soient $h_1, h_2 \in [0,1]$ tels que $h_1 \leq h_2$ alors, on a :

$$\{|f(u) - f(v)|; |u - v| \leq h_1\} \subseteq \{|f(u) - f(v)|; |u - v| \leq h_2\}$$

(car si $|u - v| \leq h_1$ alors $|u - v| \leq h_2$ et donc si $|f(u) - f(v)| \in \{|f(u) - f(v)|; |u - v| \leq h_1\}$ alors $|f(u) - f(v)| \in \{|f(u) - f(v)|; |u - v| \leq h_2\}$). D'où le résultat de croissance (par croissance de la borne supérieure).

Montrons 2 : Soient $h_1, h_2, u, v \in [0,1]$ tels que $|u - v| \leq h_1 + h_2$. Alors, il existe $w \in [\min(u,v), \max(u,v)]$ tel que $|u - w| \leq h_1$ et $|w - v| \leq h_2$ (utilisation de l'hypothèse sur $|u - v|$) (écrire la contraposée). On alors

$$\begin{aligned}
|f(u) - f(v)| &\leq |f(u) - f(w)| + |f(w) - f(v)| && \text{(par inégalité triangulaire)} \\
&\leq \omega(h_1) + \omega(h_2) && \text{(par def de la borne supérieure)}
\end{aligned}$$

On conclut en passant au sup : $\omega(h_1 + h_2) \leq \omega(h_1) + \omega(h_2)$ (utilisation de l'hypothèse sur $|u - v|$).

Montrons 3 : D'après le point précédent (par récurrence), on a, pour tout $n \in \mathbb{N}$ et $h \in [0,1]$, $\omega(nh) \leq n\omega(h)$. Soit $\lambda \in \mathbb{R}_+$ tel que $h\lambda \in [0,1]$. On a alors $\lfloor\lambda\rfloor \leq \lambda \leq \lfloor\lambda\rfloor + 1$. On a donc,

$$\begin{array}{rcll} \omega(\lambda h) & \leq & \omega((\lfloor\lambda\rfloor+1)h) & \text{(par croissance de } \omega \text{ et } \lambda \leq \lfloor\lambda\rfloor+1) \\ & \leq & (\lfloor\lambda\rfloor+1)\omega(h) & \text{(par la remarque précédente comme } (\lfloor\lambda\rfloor+1) \in \mathbb{N}) \\ & \leq & (\lambda+1)\omega(h) & \text{(par } \lfloor\lambda\rfloor \leq \lambda) \end{array}$$

D'où le lemme. □

Probabilité

On va alors donné quelques outils permettant de réaliser ce développement.

Espérance et variance On donne quelques résultats sur l'espérance d'une variable aléatoire [1, p.53].

Définition. Soit X une variable aléatoire réelle, définie sur $(\Omega, \mathcal{A}, \Pr)$. Si X est intégrable, on appelle espérance de X (sous la probabilité P) le nombre réel :

$$\mathbb{E}[X] = \int_\Omega X d\Pr$$

On dit que X est centrée, si elle est intégrable et $\mathbb{E}(X) = 0$.

Théorème (Formule de transfert)**.** *Soit X un vecteur aléatoire sur $(\Omega, \mathcal{A}, \Pr)$ à valeurs dans $(\mathbb{R}d, \mathcal{B}(\mathbb{R}^d))$ et soit φ une fonction borélienne de $\mathbb{R}^d$ dans $\mathbb{R}$. Si $\varphi \in L^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \Pr^X)$, alors*

$$\mathbb{E}[\varphi(X)] = \int_\Omega \varphi \circ X(\omega) d\Pr(\omega) = \int_{\mathbb{R}^d} \varphi(\omega) d\overset{X}{\Pr}(\omega)$$

Théorème.

1. *(Inégalité de Jensen) Si φ est connexe sur $\mathbb{R}$ et si X est une variable aléatoire réelle telle que X et $\varphi(X)$ sont intégrables, alors*
$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)]$$
2. *(Inégalité de Hölder) Si $X \in L^p$, $Y \in L^q$, $p, q \geq 1$ deux exposants conjugués, alors $XY \in L^1$ et*
$$\mathbb{E}[XY] \leq \left(\mathbb{E}\left[|X|^p\right]\right)^{1/p} \left(\mathbb{E}\left[|X|^q\right]\right)^{1/q}$$
3. *L'application $p \mapsto \left(\mathbb{E}\left[|X|^p\right]\right)^{1/p}$ est croissante.*
4. *$||.||_p = \left(\mathbb{E}\left[|.|^p\right]\right)^{1/p}$ est une norme sur $L^p(\Omega, \mathcal{A}, \Pr)$, $p \geq 1$.*
5. *On définit $||X||_\infty = \lim\limits_{p\to\infty} ||X||_p$. C'est une norme, appelée norme supremum sur $L^\infty(\Omega, \mathcal{A}, \Pr)$.*

Donnons maintenant quelques éléments sur la variance d'une variable aléatoire réelle.

Définition. Soit X une variable aléatoire réelle dont le carré est intégrable. On appelle variance de X, et on note $\mathrm{Var}(X)$, la quantité :

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

La racine de la variance est appelée l'écart type, noté $\sigma(X)$. Une variable est dite réduite si elle est d'écart type (et de variance) 1.

La variance peut aussi s'exprimer sous la forme suivante :

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

Par exemple, la variance d'une loi de Bernoulli de paramètre (n, p) est de $np(1-p)$.

Théorème (Inégalité de Markov)**.** *Si X est intégrable et $t > 0$, alors*

$$\Pr\{X \geq t\} \leq \frac{\mathbb{E}[X^+]}{t} \leq \frac{\mathbb{E}[|X|]}{t}$$

Corollaire (Inégalité de Tchebychev)**.** *Si $X \in L^2$, pour tout $t > 0$,*

$$\Pr\{|X - \mathbb{E}[X]| \geq t\} \leq \frac{\mathrm{Var}(X)}{t^2}$$

Variables de Rademacher et inégalité de Khintchine Les variables de Rademacher [2, p.512] sont des variables aléatoires discrète pouvant prendre la valeur -1 ou 1 de manière uniforme : si ϵ est une variable de Rademacher, alors $\Pr\{\epsilon = 1\} = \Pr\{\epsilon = -1\} = \frac{1}{2}$. On appelle alors suite de Rademacher une suite de ces variables aléatoires indépendantes et identiquement distribuée.

L'inégalité de Khintchine exprime le fait que les normes L^p sont toutes équivalentes sur des espaces vectoriels de dimension infinie engendré par une suite de Rademacher que l'on note $(\epsilon_n)_n$.

Proposition (Inégalité de Khintchine)**.** *Soit $a_1, \dots, a_N \in \mathbb{C}$, et $X = \sum_{n=1}^{N} a_n \epsilon_n$. Alors, on a*

$$||X||_p \leq C_0 \sqrt{p} ||X||_2 \text{pour} 1 \leq p < \infty, \text{avec} C_0 = \sqrt{2e}$$

$$||X||_1 \geq e^{-1/2} ||X||_2$$

Références

[1] P. Barde and M. Ledoux. *Probabilité*. EDP Science, 2007.

[2] H. Queffélec C. Zuily. *Éléments d'analyse pour l'agrégation*. Masson, 1995.