

SÉMINAIRE DE PROBABILITÉS (STRASBOURG)

JEAN BRETAGNOLLE

CATHERINE HUBER

Lois empiriques et distance de Prokhorov

Séminaire de probabilités (Strasbourg), tome 12 (1978), p. 332-341

http://www.numdam.org/item?id=SPS_1978__12__332_0

© Springer-Verlag, Berlin Heidelberg New York, 1978, tous droits réservés.

L'accès aux archives du séminaire de probabilités (Strasbourg) (<http://portail.mathdoc.fr/SemProba/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

Lois empiriques et distance de Prokhorov

par J. Bretagnolle et C. Huber

1. Introduction

Etant donnée une probabilité sur $(0,1)$, de fonction de répartition F , et \hat{F}_n la fonction de répartition empirique associée à un n -échantillon (X_1, \dots, X_n) d'une variable aléatoire de loi F , soit $\hat{F}_n(t) = n^{-1} \sum_{i=1}^n 1_{(0,t)}(X_i)$, il s'agit d'évaluer la vitesse de convergence de \hat{F}_n vers F , pour la distance de Prokhorov.

Plus précisément, si Π désigne cette distance, dont la définition est rappelée ci-dessous en (2.1), on cherche α tel que

$$(1.1) \quad a \leq P(n^\alpha \Pi(\hat{F}_n, F) > u) \leq b \quad n \geq n_0$$

pour un entier positif n_0 et trois constantes positives $0 < a < b < 1$, $u > 0$. On trouve que α varie entre $1/2$ et $1/3$ suivant le type de la loi F , les vitesses lentes, qui correspondent aux petites valeurs de α , étant atteintes pour des lois singulières.

Au paragraphe 2, on compare les distances de Prokhorov et de Lévy Π et L . On trouve que Π reste toujours comprise entre L et \sqrt{L} . Comme on sait que la vitesse de convergence uniforme de \hat{F}_n vers F est de $\alpha = \frac{1}{2}$, on en déduit facilement qu'il en va de même pour L et, par suite, que pour Π , α reste compris entre $1/4$ et $1/2$. Au paragraphe 4 on démontre qu' α ne peut jamais descendre au dessous de $1/3$ et au paragraphe 5, que tout α de $) \frac{1}{3} \frac{1}{2}$) est atteint.

Les démonstrations reposent sur un encadrement de la probabilité des grands écarts pour une loi multinômiale qui figure au paragraphe 3 (proposition 2). Les résultats ne sont

pas de type asymptotique, les valeurs de n_0 qui figurent en (1.1) étant explicitées et petites.

2. Comparaison des distances de Prokhorov et de Lévy

Si P et Q sont deux probabilités sur $(0,1)$ muni de la tribu \mathcal{B} de ses boréliens, et F et G leurs fonctions de répartition respectives, on désignera par K , L et Π respectivement leurs distances de Kolmogorov, Lévy et Prokhorov, c'est-à-dire

$$K(F,G) = \sup_{t \in (0,1)} |F(t) - G(t)|, \quad L(F,G) = \inf \{ \varepsilon | F(t) \leq G(t+\varepsilon) + \varepsilon \quad \forall t \in (0,1) \} e^{-}$$

$$(2.1) \quad \Pi(F,G) = \inf \{ \varepsilon | P(B) \leq Q(B^\varepsilon) + \varepsilon \quad \forall B \in \mathcal{B} \}$$

où B^ε est le ε -voisinage de B dans $(0,1)$ soit $B^\varepsilon = \{x \in (0,1) | \exists t \in B: |t-x| < \varepsilon\}$.

De ces définitions, il découle clairement que L est inférieure à K et à Π . Or on sait que la vitesse de convergence pour la distance de Kolmogorov est de $\alpha = \frac{1}{2}$ ([1] p.104):

$$(2.2) \quad P(n^{\frac{1}{2}}K(\hat{F}_n, F) > u) \xrightarrow{n \rightarrow \infty} 2 \sum_{k=1}^{\infty} (-1)^{k+1} e^{-2k^2 u^2} \quad u > 0$$

Par suite, la distance de Lévy va au moins à la vitesse $\alpha = 1/2$, et, en général, va exactement à cette vitesse; en particulier dans les cas suivants:

a) S'il existe un intervalle non vide I de $(0,1)$ (chargé par F et tel que, pour un $\varepsilon > 0$, F admette, sur I^ε , une densité f majorée, par rapport à la mesure de Lebesgue, on a un résultat analogue à (2.2) en y remplaçant $K(\hat{F}_n, F)$ par $\sup_{t \in I} |\hat{F}_n(t) - F(t)|$. Soit M le majorant de f sur I^ε .

Dès que n est assez grand pour que $n^{-\frac{1}{2}}u$ soit inférieur à ε , l'évènement $\left\{ \sup_{t \in I} |F(t) - \hat{F}_n(t)| > un^{-\frac{1}{2}} \right\}$ entraîne l'existence d'un t tel que l'on ait $\hat{F}_n(t) > F(t + \frac{un^{-\frac{1}{2}}}{M+1}) + \frac{un^{-\frac{1}{2}}}{M+1}$ ou bien $\hat{F}_n(t) < F(t - \frac{un^{-\frac{1}{2}}}{M+1}) - \frac{un^{-\frac{1}{2}}}{M+1}$, ce qui prouve que $L(\hat{F}_n, F)$ est de l'ordre de $K(\hat{F}_n, F)$.

b) De même, si F a pour support un ensemble fini de points: L'écart minimal entre deux points chargés est, pour n assez grand, supérieur à $un^{-\frac{1}{2}}$ et alors $K(\hat{F}_n, F) > un^{-\frac{1}{2}}$ entraîne la même inégalité pour L .

Proposition 1

Soient deux probabilités sur $(0,1)$. leurs distances de Prokhorov et de Lévy, π et L , satisfont toujours

$$(2.3) \quad \pi^2 \leq L \leq \pi$$

De plus, si l'une des deux probabilités admet une densité minorée, par rapport à la mesure de Lebesgue

$$(2.4) \quad \pi = O(L)$$

démonstration

Soient P et Q deux probabilités sur $(0,1)$, de fonctions de répartition respectives F et G . Supposons que $\pi(F,G)$ soit supérieur à ε . Alors il existe un borélien B tel que $Q(B)$ soit supérieur à $P(B^\varepsilon) + \varepsilon$. Comme B^ε est un ouvert, c'est une réunion d'intervalles ouverts disjoints $]a_j, b_j[$, de longueur supérieure ou égale à 2ε , en nombre $J \leq 1/2\varepsilon$. Comme $\pi \geq L$, $G(a_j + \varepsilon) \geq G(a_j + L) \geq F(a_j) - L$ et $G(b_j - L) \leq F(b_j) + L$ de sorte que

$$(2.5) \quad \varepsilon \leq Q(B) - P(B^\varepsilon) = \sum_{j=1}^J (G(b_j - \varepsilon) - G(a_j + \varepsilon) - F(b_j) + F(a_j)) \\ \leq 2LJ \leq \frac{L}{\varepsilon}$$

ce qui donne la première partie de la proposition.

Supposons maintenant que P admette une densité f minorée par m strictement positif. Alors, reprenant la formulation précédente, $G(a_j + \varepsilon) \geq F(a_j + \varepsilon - L) - L \geq F(a_j) + m(\varepsilon - L) - L$, de sorte que la contribution d'un intervalle $]a_j, b_j[$ dans la somme qui apparaît en (2.5) ne peut être positive que si $L(m+1) - m\varepsilon > 0$, ce qui exige $L(m+1)/m \geq \pi$ et démontre la deuxième partie de la proposition.

Remarque 1

Si le support des lois est un compact de longueur d , l'inégalité devient, d'après la démonstration

$$(2.6) \quad \pi^2 \leq dL \leq d\pi$$

Remarque 2

Il y a effectivement des cas où la distance de Lévy est de l'ordre du carré de la distance de Prokhorov :

Soit m un entier, P la probabilité $P = \frac{1}{m} \sum_{i=1}^m \delta_{\frac{i}{m}}$ et Q la probabilité $Q = \frac{1}{2m^2} \delta_{\frac{1}{2m}} + \sum_{i=1}^m \left(\frac{1}{m} (1 - \frac{1}{m}) \delta_{\frac{i}{m}} + \frac{1}{m^2} \delta_{\frac{2i+1}{2m}} \right)$.
Alors $L(P, Q) = \frac{1}{2m^2}$ et $\pi(P, Q) = \frac{1}{2m}$, comme on le voit en prenant pour B l'ensemble des m points $(2i+1)/2m, i=0, 1, \dots, m-1$.

Remarque 3

De cette proposition, on peut déduire que si F admet une densité qui est à la fois majorée et minorée, la vitesse de convergence de \hat{F}_n vers F en distance de Prokhorov est de $\alpha = \frac{1}{2}$.

Pour obtenir des résultats dans le cas général, on procède comme il est indiqué au début du paragraphe suivant.

3. Grands écarts pour une loi multinômiale

Soit P une probabilité sur $((0, 1), \mathcal{B})$ de fonction de répartition F , $(I_j)_{j=1, \dots, m}$ une partition de $(0, 1)$ en intervalles de longueurs respectives $u_j, u_j \leq u$, $P(I_j) = p_j, j=1, 2, \dots, m$, et A l'évènement

$$(3.1) \quad A = \left\{ \pi(\hat{F}_n, F) > u \right\}$$

A entraîne l'existence d'un sous ensemble I de $\{1, 2, \dots, m\}$ tel que $\hat{F}_n(\bigcup_{i \in I} X_i) > F(\bigcup_{i \in I} X_i - u, X_i + u) + u$. Par suite, si J est l'ensemble des indices des I_j qui contiennent au moins l'un des $X_i, i \in I$, et I' l'ensemble des indices des X_i qui appartiennent à un $I_j, j \in J$, on a

$$\begin{aligned} \hat{F}_n\left(\bigcup_{i \in I'} X_i\right) &\geq \hat{F}_n\left(\bigcup_{i \in I} X_i\right) \geq F\left(\bigcup_{i \in I} X_i - u, X_i + u\right) + u \\ &\geq \sum_j F(I_j) + u \end{aligned}$$

Notant Y_j le nombre des observations qui tombent dans I_j , on voit que A est inclus dans l'évènement $B = \left\{ \sum (Y_j - np_j)^+ \geq nu \right\}$,

et, comme $2 \sum_j (Y_j - np_j)^+ = \sum_j |Y_j - np_j|$, on a la majoration de $P(A)$

$$(3.2) \quad \text{Si } u > u_j, j=1, \dots, m \quad P(A) \leq P\left(\sum_{j=1}^m |Y_j - np_j| \geq 2nu\right)$$

D'autre part, si F charge m intervalles $I_j, j=1, \dots, m$, séparés par des intervalles non chargés de longueurs respectives v_j , $v_j > u, j=1, \dots, m-1$, l'évènement A contient l'évènement B défini ci-dessus. Donc

$$(3.3) \quad \text{Si } u < v_j, j=1, \dots, m-1 \quad P(A) \geq P\left(\sum_{j=1}^m |Y_j - np_j| \geq 2nu\right)$$

On est donc amené à évaluer $F\left(\sum_{j=1}^m |Y_j - np_j| \geq 2nu\right)$, pour une variable $Y = (Y_1, \dots, Y_m)$, multinômiale.

Proposition 2

Soit $Y = (Y_1, \dots, Y_m)$, $m > 1$, une variable multinômiale de paramètres n et $p_j, p_j > 0, j=1, 2, \dots, m, \sum_j p_j = 1$. Alors

$$(3.4) \quad P\left(\sum_{j=1}^m |Y_j - np_j| \geq 2nu\right) \leq 2^m e^{-2nu^2} \quad u \in (0, 1)$$

$$(3.5) \quad \text{Si } p_j = m^{-1}, j=1, 2, \dots, m \text{ et si } nm^{-1} > 1$$

$$P\left(\sum_{j=1}^m |Y_j - np_j| \geq \frac{u}{4} \sqrt{nm}\right) \geq 2^{-3}(1-u)^2 \quad u \in (0, 1)$$

démonstration

On démontrera successivement la majoration puis la minoration en adoptant la notation

$$(3.6) \quad Z = \sum_{j=1}^m (Y_j - np_j)^+ ; \quad S = \sum_{j=1}^m |Y_j - np_j| = 2Z$$

a) majoration

On obtient (3.4) en majorant la transformée de Laplace de Z :

$$(3.7) \quad E(e^{sZ}) \leq 2^m e^{\frac{ns^2}{8}}$$

Pour démontrer (3.7), remarquons que $\exp(a^+) \leq 1 + \exp(a)$ pour

tout a réel, et que, par suite, $E(\exp(sZ)) \leq E\left(\prod_{j=1}^m (1 + \exp(s(Y_j - np_j)))\right)$

qui s'écrit, en développant et en sommant sur les 2^m choix de

$b_j = 0$ ou 1 , soit $b = (b_1, \dots, b_m) \in \{0, 1\}^m$,

$$\sum_b E(\exp(s \sum_j b_j (Y_j - np_j))) \stackrel{\text{déf.}}{=} \sum_b B(b)$$

où $B(b)$ est, pour b fixé, la transformée de Laplace au point s d'une variable binômiale centrée de paramètres n et $p = \sum_{j=1}^m p_j b_j$, soit $B(b) = (pe^{sq} + qe^{-sp})^n = (f(p, s))^n$, par définition de la fonction f . Il en résulte que

$$E(\exp(sZ)) \leq 2^m (\sup_{p \in (0, 1)} f(p, s))^n$$

On obtient, en dérivant par rapport à p , le maximum de f en p à s fixé, qui vaut

$$g(s) = \frac{1}{s} (e^s - 1) \cdot \exp\left(\frac{s}{e^s - 1} - 1\right)$$

et est atteint au point $p(s) = \frac{1}{s} - \frac{1}{e^s - 1}$, qui est toujours compris entre 0 et 1 ($p(s) \geq 0$ car $e^s - 1 \geq s$, et $p(s) \leq 1$, c'est évident lorsque $s \geq 1$, et c'est dû à ce que $1 - s \leq e^{-s}$, lorsque $s \in [0, 1]$).

Comme la fonction f est invariante par la transformation $(p, s) \mapsto (1-p, -s)$, $g(s) = g(-s)$ et par suite, si $h = \text{Log } g$, c'est à dire $h(s) = s(e^s - 1)^{-1} - 1 + \text{Log}(s^{-1}(e^s - 1))$, on a $h(s) = \frac{1}{2}(h(s) + h(-s)) = \frac{1}{2}s \cdot \text{sh}(s) \cdot (\text{ch}(s) - 1)^{-1} - 1 + \frac{1}{2} \text{Log}(2s^{-2}(\text{ch}(s) - 1))$.

Par dérivation de h et développement en série entière, on vérifie que $4h'(s) \leq s$; donc, comme $g(0) = 1$, on obtient la majoration $g(s) \leq \exp(s^2/8)$, ce qui prouve (3.7).

Optimisant en s l'inégalité, valable pour tout s réel, $P(Z \geq nu) \cdot \exp(snu) \leq E(\exp(sZ)) \leq 2^m \exp(ns^2/8)$, on obtient, pour $s = 4u$,

$$P(Z > nu) \leq 2^m e^{-2nu^2}$$

La majoration (3.4) en résulte puisque $S = 2Z$.

b) minoration

Pour tout $0 < a < b$, $ES \leq a P(S \leq a) + bP(a < S \leq b) + \int_{s>b} s dP(s)$.

Donc

$$(3.8) \quad P(S > a) \geq \left(ES - a - \frac{1}{b}E(S^2) \right) \frac{1}{b}$$

Comme (Nous allons le montrer) $E(S^2)$ est de l'ordre de $(ES)^2$ uniformément en m , on obtiendra un minorant en choisissant a et b de l'ordre de ES .

Montrons que $E(S^2)$ est de l'ordre de $(ES)^2$:

On sait déjà que $ES \leq (E(S^2))^{1/2}$ (croissance des moments). Reste à montrer que pour un $c > 0$, $ES \geq c(E(S^2))^{1/2}$. Or $ES = m E|Y-np|$, où Y , sans indice, est une variable binômiale de paramètres n et p , et $E(S^2) \leq m^2 E((Y-np)^2)$. Si $U = |Y-np|$, il suffit donc de prouver que $EU \geq c(E(U^2))^{1/2}$ et, pour cela, que $E(U^2) \geq c(E(U^4))^{1/2}$, car non seulement les moments vont en croissant, mais ils croissent de plus en plus vite; plus précisément, la fonction $\alpha \mapsto \text{Log}(E(X^{1/\alpha})^\alpha)$ est convexe, pour toute variable aléatoire X .

Posons $A = 1/\alpha$. On a besoin de $A = 1, 2, 4$, soit $\alpha = 1, \frac{1}{2}, \frac{1}{4}$. Or $\frac{1}{2} = \frac{1}{3} \cdot 1 + \frac{2}{3} \cdot \frac{1}{4}$. Donc

$$\text{Log}((E(U^2))^{1/2}) \leq \frac{1}{3} \text{Log } EU + \frac{2}{3} \text{Log} ((E(U^4))^{1/4})$$

inégalité qui s'écrit aussi

$$(3.9) \quad \frac{E(U^2)}{(EU)^2} \leq \frac{E(U^4)}{(E(U^2))^2}$$

Si B est une variable de Bernoulli de paramètre p , $E((Y-np)^4) = n E((B-np)^4) + 3n(n-1) (E(B-np)^2)^2 = npq(p^3+q^3) + 3n(n-1)p^2q^2$ quantité qui est inférieure ou égale à $(npq)^2(3 + \frac{1}{npq})$. Comme $E((Y-np)^2) = npq$, on en déduit que

$$E((Y-np)^4) \leq \left(3 + \frac{1}{npq}\right) (E((Y-np)^2))^2$$

Utilisant (3.9) et $U = |Y-np|$, on a $E(U) \gg (3 + \frac{1}{npq})^{-\frac{1}{2}} (E(U^2))^{\frac{1}{2}}$
 et par conséquent la même inégalité où S remplace U . Reportant
 dans (3.8) où on fait $a = u.E(S)$, u dans $]0,1[$, et $b = v.E(S)$
 et optimisant en v , on obtient $v = 2(3 + \frac{1}{npq})(1-u)^{-1}$, soit

$$(3.10) \quad P(S \gg \frac{u(nm)^{\frac{1}{2}}}{2(3+2mn^{-1})}) \gg \frac{(1-u)^2}{4(3+2mn^{-1})}$$

ce qui donne bien la minoration (3.5) lorsque nm^{-1} est supérieur à 1.

4. Vitesse minimale de convergence: $\alpha=1/3$

Des inégalités (3.2) et (3.4), il résulte que pour toute fonction de répartition F , on a

$$(4.1) \quad P(\pi(\hat{F}_n, F) \gg \frac{c}{m}) \leq 2^m e^{-2nc^2 m^{-2}} \quad c \gg 1, m \in \mathbb{N}$$

Pour éliminer la contrainte "m entier", remarquons que si $m \leq v < m+1$, $P(\pi(\hat{F}_n, F) \gg cv^{-1}) \leq P(\pi(\hat{F}_n, F) \gg c(m+1)^{-1}) \leq$
 $\leq 2^{m+1} \exp(-2nc^2(m+1)^{-2}) \leq 2^{v+1} \exp(-2nc^2(v+1)^{-2})$, ce qui permet de remplacer (4.1) par

$$(4.2) \quad P(\pi(\hat{F}_n, F) \gg \frac{c}{v}) \leq 2^{v+1} e^{-2nc^2(v+1)^{-2}} \quad c \gg 1, v > 0$$

Appelons $\exp(f(c,v))$ le second membre de (4.2).

Soit B un réel positif quelconque.

Choisissons $v+1 = n^{1/3} (2c^2 (\log 2)^{-1})^{1/3} - B$, soit $v = n^{1/3} A c^{2/3} - B'$
 en notant

$$(4.3) \quad A = \left(\frac{2}{\log 2}\right)^{1/3}; \quad B' = B+1$$

Alors, $f(c,v)$ est inférieur à $-2B \log 2$.

Pour optimiser l'inégalité (4.2), minimisons $\frac{c}{v}$ en c sous les contraintes $c \geq 1$ et $v > 0$, qui sont vérifiées pour tout $n \geq 1$ dès que

$$(4.4) \quad c \geq \max(1, (B'A^{-1})^{3/2})$$

Or c/v vaut $g(c) = \frac{c}{An^{1/3}c^{2/3}} \cdot \frac{1}{1-B'(An^{1/3}c^{2/3})^{-1}}$ qui reste

inférieur à 2^{premier} facteur dès que c est supérieur ou égal à $(2B'A^{-1})^{3/2}$. Donc, à condition de choisir

$$(4.5) \quad c = c_0 = \max(1, (2B'A^{-1})^{3/2})$$

on obtient

$$(4.6) \quad P(n^{1/3} \pi(\hat{F}_n, F) \geq 2A^{-1}c_0^{1/3}) \leq 2^{-2B} \quad n \geq 1, B > 0$$

On en déduit le théorème

Théorème 1

Soit \mathfrak{F} l'ensemble de toutes les fonctions de répartition sur $(0,1)$, F un élément de \mathfrak{F} et \hat{F}_n la fonction de répartition empirique d'ordre n correspondante. Alors il existe une constante positive C telle que

$$(4.7) \quad \bigvee_{F \in \mathfrak{F}} P(n^{1/3} \pi(\hat{F}_n, F) \geq u) \leq C e^{-2^{-3}u^3} \quad n \geq 1, u \geq 4$$

5. Lois singulières

On a vu que les vitesses de convergence se situent entre $1/3$ et $1/2$ et que la vitesse $1/2$ est atteinte sur une grande classe de fonctions de répartition. Nous allons, dans ce paragraphe, construire des fonctions de répartition pour lesquelles α est dans $]1/3, 1/2[$.

Soit, pour un t de $]0, \frac{1}{2}[$, la probabilité singulière F , ou F_t , sur $(0, 1)$, qui, pour tout entier k , charge également $m = 2^k$ intervalles de longueur commune t^k , séparés par des intervalles égaux de longueur $(1-2t)t^{k-1}$.

D'après (3.3) et (3.5), dès que $2^{-3}u(mn^{-1})^{\frac{1}{2}}$ est inférieur à $(1-2t)t^{k-1}$ et nm^{-1} est supérieur à 1, $P(\pi(\widehat{F}_n, F) \geq 2^{-3}u(mn^{-1})^{\frac{1}{2}}) \geq 2^{-3}(1-u)^2$, pour tout u de $]0, 1[$.

Optimisant en l'entier k sous les contraintes $n2^{-k} > 1$ et $2^{-3}un^{-\frac{1}{2}}2^{k/2} < (1-2t)t^{k-1}$, qui se résument en

$$(5.1) \quad k < \min \left(\frac{\text{Log } n}{\text{Log } 2}, \frac{\text{Log } n + 2\text{Log } 8(1-2t) - 2\text{Log}(ut)}{\text{Log } 2 - 2 \text{Log } t} \right)$$

Choisissant u égal à $((1-2t)t^{-1} \wedge 1) = u(t)$, la condition (5.1) est réalisée pour

$$(5.2) \quad k = \frac{\text{Log } n}{\text{Log } 2 - 2\text{Log } t}$$

Pour cette valeur de k , on obtient

$$(5.3) \quad P(n^{\frac{1}{2}}(1 - \frac{\text{Log } 2}{\text{Log } 2 - 2\text{Log } t}) \pi(\widehat{F}_n, F_t) \geq \frac{u(t)}{4}) \geq 2^{-3}(1-u(t))^2$$

Pour chaque t de $]0, \frac{1}{2}[$, la vitesse de convergence est donc supérieure à $\alpha(t) = \frac{1}{2}(1 - \text{Log } 2(\text{Log } 2 - 2\text{Log } t)^{-1})$. Or $\alpha(t)$ décroît de $\frac{1}{2}$ à $1/3$ quand t croît de 0 à $\frac{1}{2}$ (exclus).

On démontre, de la même manière qu'au paragraphe 4, que la probabilité qui figure au premier membre de (5.3) est majorée, ce qui prouve que la vitesse de convergence de \widehat{F}_n vers F_t est effectivement $\alpha(t)$.