

STATISTIQUE ET ANALYSE DES DONNÉES

BERNARD-BERTRAND GAREL

Calcul pratique de la distance de Prokhorov

Statistique et analyse des données, tome 6, n° 3 (1981), p. 35-46.

http://www.numdam.org/item?id=SAD_1981__6_3_35_0

© Association pour la statistique et ses utilisations, 1981, tous droits réservés.

L'accès aux archives de la revue « Statistique et analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/legal.php>). 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/>

Statistique et Analyse des données
1981 - 3 - pp. 34-45

CALCUL PRATIQUE DE LA DISTANCE DE PROKHOROV

Bernard - Bertrand GAREL

Faculté des Sciences et des Techniques

B.P. 1104 - 73011 CHAMBERY Cédex

Laboratoire IMAG associé au C.N.R.S. (L.A.7)

B.P. 53 - 38041 GRENOBLE Cédex

Résumé : Pour calculer la distance de Prokhorov $\Pi(P, Q)$ entre deux probabilités P et Q , on construit deux suites de probabilités (P_n) et (Q_n) à support fini qui convergent étroitement vers P et Q . Puis on approche $\Pi(P, Q)$ par $\Pi(P_n, Q_n)$.

Abstract : P and Q being any two probabilities, it seems to be very difficult to calculate their exact Prokhorov distance $\Pi(P, Q)$. So we approximate this distance $\Pi(P, Q)$ by the Prokhorov distance $\Pi(P_n, Q_n)$ between P_n and Q_n where (P_n) and (Q_n) are two sequences of probabilities on finite sets which converge weakly towards P and Q respectively.

Mots clés : distance de Prokhorov, convergence étroite, programmation linéaire, algorithme de Ford et Fulkerson.

0 - INTRODUCTION..

Il est admis que le calcul exact de la distance de Prokhorov entre deux probabilités n'est pas chose facile ; ce qui restreint son utilisation pour des applications.

C'est en 1956 que Prokhorov [6] a introduit cette distance sur l'ensemble $M(S)$ des mesures sur S et qu'il a montré l'équivalence entre la convergence étroite et la convergence au sens de cette distance. Elle a été alors utilisée pour étudier des vitesses de convergence. En 1965, Strassen [9] a démontré une équivalence importante de la définition ce qui a permis à Hampel [5] de l'utiliser pour une notion qualitative de robustesse. Plus récemment des liens entre distances entre probabilités et distances entre variables aléatoires ont été établis (cf [11], [10]).

Nous rappelons ici la définition de cette distance dans un cadre général, ainsi que les variantes connues.

Soit (S, d) un espace métrique séparable et complet ; $\mathcal{B}(S)$ sa tribu des boréliens. Soit A un sous ensemble non vide de S . On définit respectivement $A^\epsilon = \{s \in S / \exists a \in A \text{ avec } d(s,A) < \epsilon\}$ et $A^{\leq \epsilon} = \{s \in S / \exists a \in A \text{ avec } d(s,A) \leq \epsilon\}$. L'ensemble A^ϵ est ouvert, alors que $A^{\leq \epsilon}$ est fermé. Soient P et Q deux probabilités définies sur $(S, \mathcal{B}(S))$. La distance de Prokhorov (cf [6]) est définie par

$$(1) \quad \Pi(P,Q) = \inf\{\epsilon / \forall A \text{ fermé, } P(A) \leq Q(A^\epsilon) + \epsilon\} .$$

Si l'on pose $\epsilon(\alpha) = \inf \{\alpha / P(A) \leq Q(A^\alpha) + \alpha\}$, on peut écrire (cf [7], et [11])

$$(2) \quad \Pi(P,Q) = \sup \{\epsilon(A) / A \text{ fermé}\}$$

Si l'on pose $\beta(\alpha) = \inf \{\beta / \forall A \text{ fermé } P(A) \leq Q(A^\alpha) + \beta\}$

on peut écrire (cf.[2] et [8])

$$(3) \quad \Pi(P,Q) = \inf\{\alpha / \beta(\alpha) \leq \alpha\} .$$

Dans les expressions ci-dessus il est possible de remplacer A^α par $A^{\leq \alpha}$ et $\forall A^{\text{fermé}}$ par $\forall A \in \mathcal{B}(S)$ sans changer la valeur de $\Pi(P,Q)$ (cf [2] ou [4]).

La distance de Prokhorov n'est pas invariante par transformation linéaire. Dans un certain nombre de cas, il est possible de calculer sa valeur exacte (cf [4], [7]). Mais il n'existait pas à notre connaissance un procédé général. C'est pourquoi, nous proposons la méthode suivante. Soient P et Q deux probabilité quelconques. Nous allons construire deux suites (P_n) et (Q_n) de probabilités à support fini qui convergent étroitement vers P et Q . Puis nous allons calculer $\Pi(P_n, Q_n)$ pour n suffisamment grand. Ce qui nous donnera une valeur approchée de $\Pi(P,Q)$. Ce qui suppose d'une part de pouvoir construire (P_n) respectivement (Q_n) en ayant à chaque instant une information sur $\Pi(P_n, P)$ (resp. $\Pi(Q_n, Q)$). Puis de pouvoir calculer la distance de Prokhorov pour deux lois à supports finis.

Dans une première partie nous allons exposer la résolution théorique de cette méthode et la construction de l'algorithme. La seconde partie sera consacrée à la mise en oeuvre de la méthode et aux exemples.

I - RESOLUTION THEORIQUE DU CALCUL APPROCHE

Nous allons supposer dans ce qui suit que $S = \mathbb{R}$. Mais la méthode peut s'appliquer à tout espace métrique séparable et complet. Soient P et Q deux probabilités sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Les constructions de la suite (P_n) et l'information sur $\Pi(P_n, P)$ nous sont données par le lemme suivant :

4 Lemme (cf [6]). Soit P une probabilité définie sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Soit ϕ une fonction mesurable de \mathbb{R} dans \mathbb{R} . Soient $\delta > 0$ et $H = \{x / |\phi(x) - x| \leq \delta\}$. Si $P(H) \geq 1 - \epsilon$ alors $\Pi(P, P_\phi) \leq \max(\epsilon, \delta)$ où P_ϕ est la probabilité image de P par ϕ .

Soit n un entier positif donné. On peut trouver un intervalle fermé borné $K_n \subset \mathbb{R}$ tel que $\max(P(\mathbb{R} - K_n), Q(\mathbb{R} - K_n)) < 1/2n$.

Soient k_{\min} et k_{\max} les extrémités de cet intervalle et $\ell = k_{\max} - k_{\min}$ sa longueur. On divise K_n en $[n \cdot \ell] + 1 = m+1$ intervalles disjoints, où $[.]$ représente la partie entière. Posons $I_j = [k_{\min} + (j-1)/n, k_{\min} + j/n]$ $j = 1, \dots, m$
 $I_{m+1} = [k_{\min} + m/n, k_{\max}]$. Soit alors x_j le milieu de l'intervalle I_j pour $j = 1, \dots, m$ et $x_{m+1} = x_m + \frac{1}{n}$.

On définit la fonction ϕ_n par :

$$\forall x \in I_j, \phi_n(x) = x_j \quad j = 1, \dots, m+1$$

$$\forall x \in \mathbb{R} - K_n, \phi_n(x) = x_{m+1}.$$

Il est clair que ϕ_n est mesurable et qu'en appliquant le lemme 4 on obtient :

5 Corollaire : Soit P_n (resp. Q_n) la mesure image de P (resp. Q) par ϕ_n . Alors $\Pi(P, P_n) \leq 1/2n$ (resp. $\Pi(Q, Q_n) \leq 1/2n$).

Le problème se résume maintenant au calcul de $\Pi(P_n, Q_n)$. Pour y arriver nous utilisons la variante (3) de la définition (1). Soit $S_n = \{x_1, \dots, x_{m+1}\}$. Pour $B \subset S_n$ posons $B_n^{\alpha} = B^{\alpha} \cap S_n$. Pour $0 \leq \alpha < 1$ donné, il faut tout d'abord calculer $\beta(\alpha) = \inf \{ \beta / \forall B \in \mathcal{B}(\mathbb{R}), P_n(B) \leq Q_n(B^{\alpha}) + \beta \}$.

Posons $P_n(x_i) = p_i$ $Q_n(x_i) = q_i$

$$d_{ij} = \begin{cases} 1 & \text{si } |x_i - x_j| \leq \alpha \\ 0 & \text{sinon} \end{cases} \quad \text{pour } i, j = 1, \dots, m+1$$

On notera (u, v) un élément de $\mathbb{R}^{2(m+1)}$ de composantes $(u_i)_{i=1, \dots, m+1}, (v_j)_{j=1, \dots, m+1}$
 Soit $W(u, v)$ la forme linéaire définie sur $\mathbb{R}^{2(m+1)}$ par $W(u, v) = \sum_i (p_i u_i + q_i v_i)$

6 Lemme (cf [8]) : On a les égalités suivantes

$$\beta(\alpha) = \inf \{ \beta / \forall B \subset S_n, P_n(B) \leq Q_n(B^{\alpha}) + \beta \}$$

$$= 1 - \text{Min} \sum_i (p_i u_i + q_i v_i)$$

sous les contraintes : $u_i \geq 0 \quad v_j \geq 0$

$$\forall i, j \quad u_i + v_j \geq d_{ij} \quad \left. \vphantom{\sum_i} \right\} \quad (7)$$

Indications : la première égalité nous dit simplement qu'on peut se restreindre au support de P_n et Q_n . La deuxième égalité est nettement plus importante. On montre tout d'abord que pour obtenir le minimum de la forme linéaire W sous les contraintes (7) on peut se restreindre aux valeurs de W sur les sommets du cube unité situés dans la zone des contraintes. On construit alors une partition de cet ensemble. Puis on prend le minimum en deux temps : sur chaque élément de la partition. Puis sur l'ensemble des minimums obtenus (cf [4]). On écrit alors la

valeur de $W(u,v)$ sous forme d'une différence de probabilités de deux sous-ensembles de S_n).

En raison de l'espacement des x_i de $\frac{1}{n}$ l'application $\alpha \rightarrow Q_n(\beta_n^{\alpha})$ est constante par morceaux et a des sauts pour des α de la forme $\frac{k}{n}$, $k = 1, \dots, m$. On donne à α de telles valeurs. Cependant pour des gains de temps calcul, il est nécessaire de restreindre davantage le champ des valeurs possibles de α . On le fait grâce aux inégalités suivantes : si P et Q sont deux probabilités quelconques et A un Borélien donné :

$$(8) \quad \epsilon(A) \leq \Pi(P,Q) \leq \sup_{A \in \mathcal{B}} |P(A) - Q(A)|$$

On donnera donc à α des valeurs k/n telles que

$$0 \leq \epsilon(B) \leq k/n \leq \sup_{B \in S_n} |P_n(B) - Q_n(B)| \leq 1$$

Le minimum de $\sum_i (p_i u_i + q_i v_i)$ s'obtient sur ordinateur grâce à l'algorithme de Ford et Fulkerson (cf [3]) qui recherche le flot maximum noté $F_{\max}(\alpha)$ qui peut entrer dans un réseau de conduits. Les caractéristiques de cet algorithme sont décrites en II ci dessous. D'après le lemme 6 on a $\beta(\alpha) = 1 - F_{\max}(\alpha)$. On cherche alors $\inf \{ \alpha / \beta(\alpha) \leq \alpha \}$. Grâce au programme on trouve :

$$\alpha_{\min} = \min \{ \alpha / \alpha \text{ de la forme } \frac{k}{n} \text{ tel que } \beta(\alpha) \leq \alpha \}$$

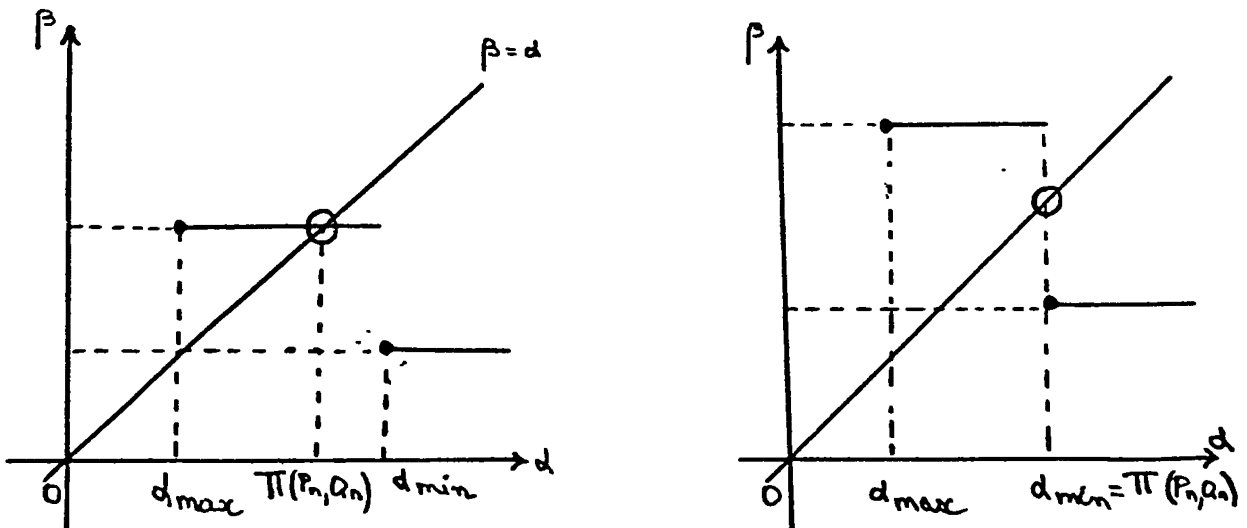
$$\alpha_{\max} = \max \{ \alpha / \alpha \text{ de la forme } \frac{k}{n} \text{ tel que } \beta(\alpha) > \alpha \}$$

Alors en remarquant que $\beta(\alpha)$ est une fonction décroissante de α , on peut conclure que

$$(9) \quad \inf \{ \alpha / \beta(\alpha) \leq \alpha \text{ (sans restriction)} \} = \min [\alpha_{\min}; \beta(\alpha_{\max})]$$

sur la forme des α

Les graphes (10) montrent les deux cas possibles :



Graphes 10

On peut enfin majorer l'erreur d'estimation grâce aux inégalités :

$$\Pi(P_n, Q_n) - \Pi(P_n, P) - \Pi(Q_n, Q) \leq \Pi(P, Q) \leq \Pi(P, P_n) + \Pi(P_n, Q_n) + \Pi(Q_n, Q)$$

$$(11) \quad \text{soit} \quad |\Pi(P, Q) - \Pi(P_n, Q_n)| \leq 1/n$$

II - EXEMPLES ET MISE EN OEUVRE

1°) Soient P et Q les probabilités sur \mathbb{R}^+ définies par

$$f(x) = e^{-x} \cdot \mathbb{1}_{\mathbb{R}^+}(x) \quad \text{et} \quad g(x) = \frac{1}{2} e^{-x/2} \cdot \mathbb{1}_{\mathbb{R}^+}(x)$$

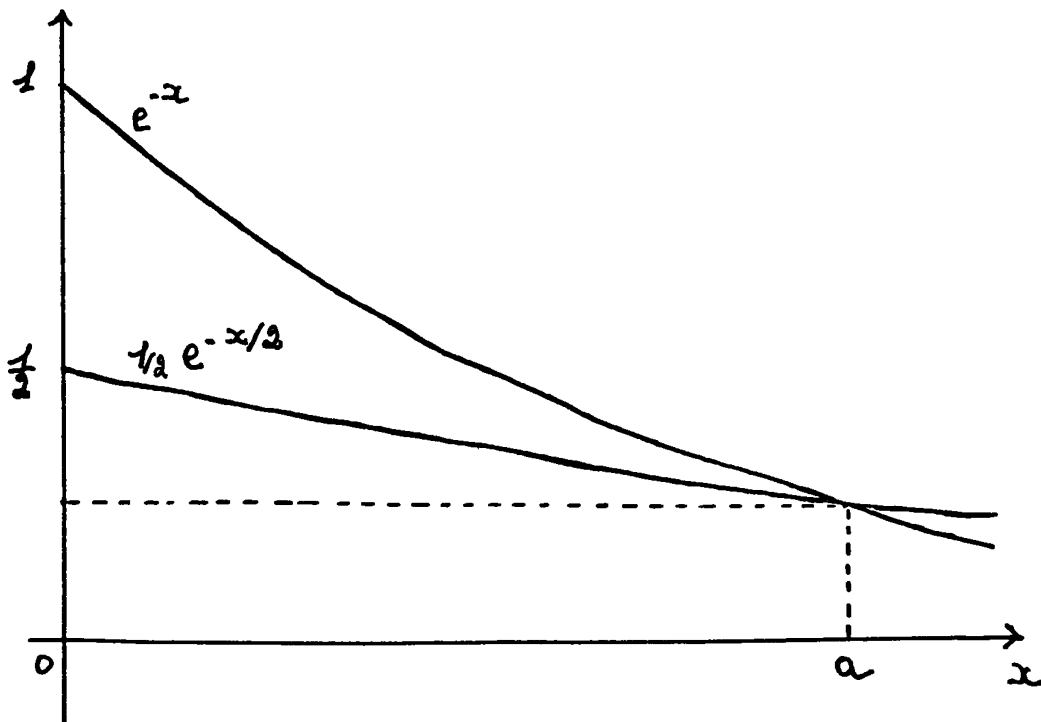
Appellons a le point d'intersection des deux courbes. Alors il est clair que

$\text{Sup}_A |P(A) - Q(A)| = P[0, a] - Q[0, a] = 0,25$. D'autre part on peut calculer facilement $\varepsilon [0, a] \neq 0,202$. On a donc, d'après (8) : $0,202 \leq \Pi(P, Q) \leq 0,25$.

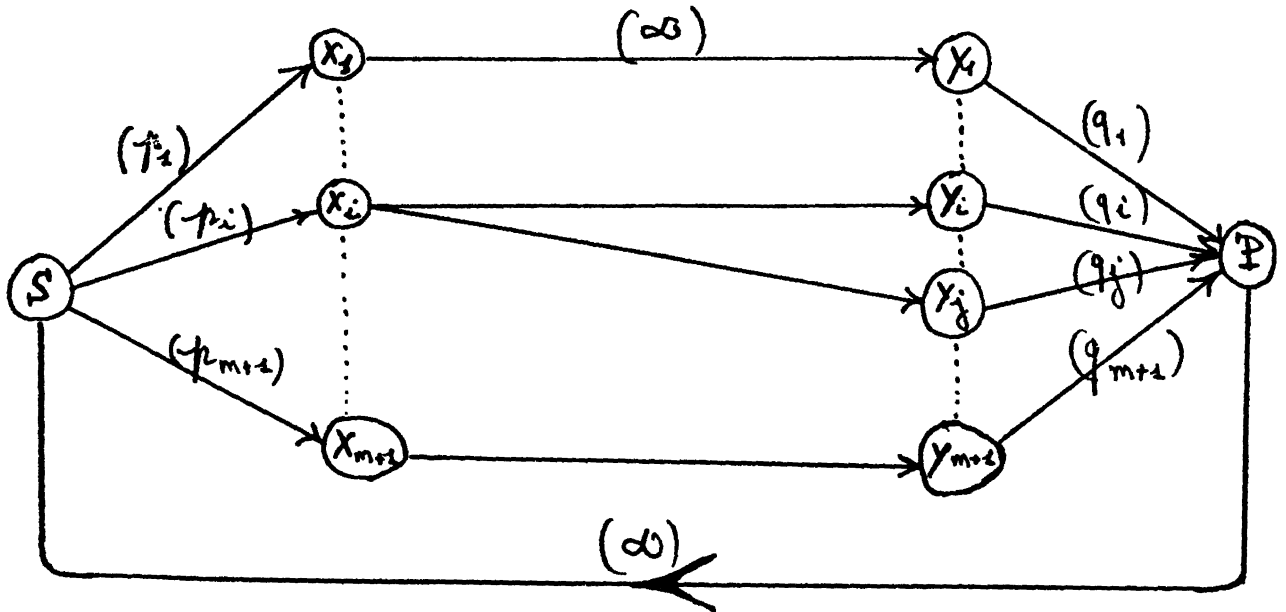
Nous avons choisi $n = 100$. Alors $K_n = [0; 10,597]$ ce qui nous donne $m+1 = 1060$, $I_{m+1} = [10,59; 10,597]$ et $x_{m+1} = 10,595$. Les probabilités P_n et Q_n sont définies sur un ensemble de 1060 points ; chaque point étant équidistant des deux points qui l'entourent à une distance de $1/100$. Le calcul de

$$\int_{(i-1)/n}^{i/n} e^x dx \quad \text{pour } i = 1, \dots, m$$

est effectué en double précision, puis multiplié par 2^{34} , produit dont on ne garde que la partie entière et que l'on pose = p_i de façon à n'avoir pour p_i que des valeurs entières. 2^{34} représente le plus grand nombre entier qu'il est possible d'entrer dans l'ordinateur. Nous procéderons de la même manière pour le calcul des q_i . La somme des erreurs d'arrondis est dans les deux cas inférieure à 4×10^{-8} et n'a aucune influence sur la précision du résultat. Elle a été reportée sur P_{m+1} ainsi que $\int_{10,597}^{+\infty} e^{-x} dx$. Le calcul de $\beta(\alpha)$ est effectué en quelques secondes par l'ordinateur.



Pour expliquer les modalités de ce calcul, construisons le graphe suivant :



Sur les arcs de ce graphe nous faisons circuler des flots bornés sur chaque arcs par le nombre indiqué entre parenthèse. Montrons que le flot maximum entre S et P prend pour valeur $\min \sum_i (p_i u_i + q_i v_i)$ sous les contraintes $u_i \geq 0, v_i \geq 0$ $u_i + v_j \geq d_{ij}$, où $d_{ij} \geq 0$. En effet le dual de ce problème est

$$(D) \begin{cases} \text{maximiser } \sum_{i,j} d_{ij} x_{ij} \\ \text{sous les contraintes : } \sum_i d_{ij} x_{ij} \leq q_j \quad ; \quad \sum_j d_{ij} x_{ij} \leq p_i \\ x_{ij} \geq 0 \end{cases}$$

Nous avons $d_{ij} = 1$ si $|i-j| \leq n\alpha$
 $d_{ij} = 0$ si $|i-j| > n\alpha$

Si $d_{ij} = 0$, on peut supprimer la variable x_{ij} correspondante.

D'où (D) s'écrit :

$$\begin{cases} \text{maximiser } \sum_{\substack{i,j \\ |i-j| \leq n\alpha}} x_{ij} \quad \text{où } x_{ij} \geq 0 \\ \text{et } \sum_i x_{ij} \leq q_j \quad \sum_j x_{ij} \leq p_i \\ |i-j| \leq n\alpha \quad |i-j| \leq n\alpha \end{cases}$$

problème que décrit le graphe ci-dessus si x_{ij} est la valeur du flot allant de X_i à Y_j .

La capacité maximum de l'axe S, X_i exprime que $\sum_{\substack{i,j \\ |i-j| \leq n\alpha}} x_{ij} \leq p_i$. La capacité maximum

de l'axe Y_j, P exprime que $\sum_i x_{ij} \leq q_j$
 $|i-j| \leq n\alpha$

Pour trouver le flot maximum, on utilise l'algorithme de Ford et Fulkerson (cf [3])

qui est suffisamment connu pour que nous n'en fassions qu'un descriptif sommaire dans le cas qui nous intéresse. On appelle sommet insaturé : un sommet X_i tel que la somme des flots sortants est strictement inférieure à p_i , ou un sommet Y_j tel que la somme des flots entrants est strictement inférieure à q_j . A chaque itération on cherche une suite d'arcs partant d'un sommet insaturé X pour aller à un sommet insaturé Y. Cette suite d'arcs est alternée ; les arcs parcourus dans le sens contraire à l'orientation (du côté Y vers le côté X) doivent porter un flot non nul.

Si une telle suite existe on peut augmenter le flot total. La valeur de l'augmentation est entière puisqu'obtenue par somme ou soustraction d'entiers. Ceci permet de borner aisément le nombre d'itérations, car le flot maximum est borné par $\sum_i p_i$ soit 2^{34} . Le nombre d'itérations est en réalité bien inférieur.

Au lieu de partir d'un flot nul, il est avantageux de partir d'un flot aussi grand que possible et ce flot nous est fourni par $\sum_{i=1}^{m+1} \min(p_i, q_i)$. On relie cette valeur à l'inégalité (8) de la façon suivante :

d'après le lemme 6 et l'inégalité (8)

$$\begin{aligned}
 F_{\ell_{\max}}(\alpha) &= 1 - \beta(\alpha) \geq 1 - \sup_B |P_n(B) - Q_n(B)| \\
 &= 1 - \sum_{J_1} (p_i - q_i) \quad \text{où } J_1 = \{i/p_i > q_i\} \\
 &\quad \quad \quad J_2 = \{i/p_i \leq q_i\} \\
 &= (1 - \sum_{J_1} p_i) + \sum_{J_1} q_i = \sum_{J_2} p_i + \sum_{J_1} q_i \\
 &= \sum_i \min(p_i, q_i)
 \end{aligned}$$

Le tableau ci-dessous résume les caractéristiques de notre exemple et les résultats du programme.

$$\begin{aligned}
 \sup_A |P(A) - Q(A)| &= 0,25. \quad \text{Pour } x_0 = 1,590182 \quad \varepsilon([0, x_0]) = 0,203888 \\
 n &= 100 \quad m+1 = 1060 \quad K_n = [0; 10,597]
 \end{aligned}$$

α	$F_{\ell_{\max}}(\alpha)$	$\beta(\alpha)$	$\Pi(P_n, Q_n)$
.19	0,79326	0,20614	0,20468
$\alpha_{\max} \rightarrow$.20	0,79532	0,20468	
$\alpha_{\min} \rightarrow$.21	0,79735	0,20265	
.22	0,79937	0,20063	

2°) Prenons un deuxième exemple. Dans ce qui suit P et Q sont des lois $N(0, \sigma)$ avec $\sigma \geq 1$. Notons tout d'abord qu'il serait vain d'espérer utiliser la symétrie des densités pour réduire l'étude à \mathbb{R}^+ . Appelons en effet P^+ et Q^+ la restriction de telles lois à \mathbb{R}^+ et P^- et Q^- leurs restrictions à \mathbb{R}^- . Soient $a > 0$ et ε_1 tel que

$$P^+([0, a]) \leq Q^+([0, a + \varepsilon_1]) + \varepsilon_1$$

$$\text{donc } P^-([-a, 0]) \leq Q^-([-a - \varepsilon_1, 0]) + \varepsilon_1$$

alors

$$\begin{aligned} P([-a, +a]) &= P^+([0, a]) + P^-([-a, 0]) \leq Q^+[0, a + \varepsilon_1] + Q^-[-a - \varepsilon_1, 0] + 2\varepsilon_1 \\ &= Q([-a, +a]) + 2\varepsilon_1! \end{aligned}$$

Soient P et Q deux lois normales centrées de variances respectives σ_1^2 , σ_2^2 avec $\sigma_1 > \sigma_2$. Appelons t_0 le point d'intersection d'abscisse > 0 des graphes des densités. Il est facile de voir que

$$\sup_{A \in \mathcal{B}} |P(A) - Q(A)| = P([-t_0, t_0]) - Q([-t_0, t_0])$$

et que t_0 nous est fourni par :

$$\frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{-\frac{1}{2} \frac{t^2}{\sigma_1^2}\right\} = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{-\frac{1}{2} \frac{t^2}{\sigma_2^2}\right\} \Leftrightarrow \frac{\sigma_2}{\sigma_1} = \exp\left\{\frac{t^2}{2} \left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}\right)\right\}$$

$$\text{soit } t_0 = \sqrt{2 \frac{\text{Log } \frac{\sigma_2}{\sigma_1}}{1/\sigma_1^2 - 1/\sigma_2^2}}$$

Si l'on appelle Φ la fonction $\int_0^t \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$ qui tabule la loi normale on a : $P([-t_0, t_0]) - Q([-t_0, t_0]) = 2[\Phi(\frac{t_0}{\sigma_1}) - \Phi(\frac{t_0}{\sigma_2})]$. Si l'on cherche $\varepsilon \in ([-t_0, t_0])$ on a à résoudre $2\Phi(\frac{t_0}{\sigma_1}) - 2\Phi(\frac{t_0 + \varepsilon}{\sigma_2}) = \varepsilon$ soit une équation de la forme $\Psi(\varepsilon) = \varepsilon$ avec $\Psi'(\varepsilon) = \frac{1}{\sigma_2} e^{-\varepsilon^2}$ qui justifie l'obtention d'une solution par une méthode itérative.

Le tableau 13 nous donne les valeurs numériques des inégalités (8).

Nous avons alors alternativement choisi $n = 10$ et $n = 50$. L'intervalle k_n est symétrique par rapport à l'origine et nous donnons au tableau la valeur k_{\max} . On a alors divisé k_n en $2([n \cdot k_{\max}] + 1)$ intervalles. On a reporté $\int_{-\infty}^{k_{\min}} f(x) dx$ sur p_0 et $\int_{k_{\max}}^{+\infty} f(x) dx$ sur p_{2m+2} . Ces valeurs ainsi que $\int_{i-1/n}^{i/n} f(x) dx$ sont calculées par l'ordinateur au moyen de la formule d'approximation polynomiale (cf [1]). Le programme a tourné avec des valeurs entières comme dans le cas II 1°. Les résultats sont consignés au tableau 14 et nous avons pu tracer la courbe 15.

CALCULS PREALABLES AUX CALCULS SUR ORDINATEUR

$P : N(0,1)$ $Q : N(0,\sigma)$ $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ $g(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$ t_0 point d'intersection > 0 de f et g
 Nous posons $A_0 = [-t_0, +t_0]$, $\epsilon(A_0) = \epsilon_0$ $t_0 = [2 \text{ Log} \sigma / (1 - \frac{1}{\sigma^2})]^{1/2}$

Tableau 13

σ	t_0	Sup P(A) - Q(A) = P(A ₀) - Q(A ₀)	P(A ₀)	$\epsilon_0 + Q(A_0, \epsilon_0)$	ϵ_0
1,2	1,0924	0,0880	0,7253	0,7260	0,062
2	1,3596	0,3226	0,8260	0,8265	0,248
3	1,5722	0,48432	0,8841	0,8854	0,396
4	1,7197	0,58177	0,9145	0,9152	0,495
5	1,8311	0,64712	0,9329	0,9332	0,565
6	1,9199	0,69411	0,9451	0,9457	0,618
7	1,9932	0,72960	0,9538	0,9542	0,659
8	2,0555	0,75739	0,9602	0,9607	0,691
9	2,1094	0,77978	0,9651	0,9657	0,719
10	2,1598	0,79413	0,9690	0,9690	0,741

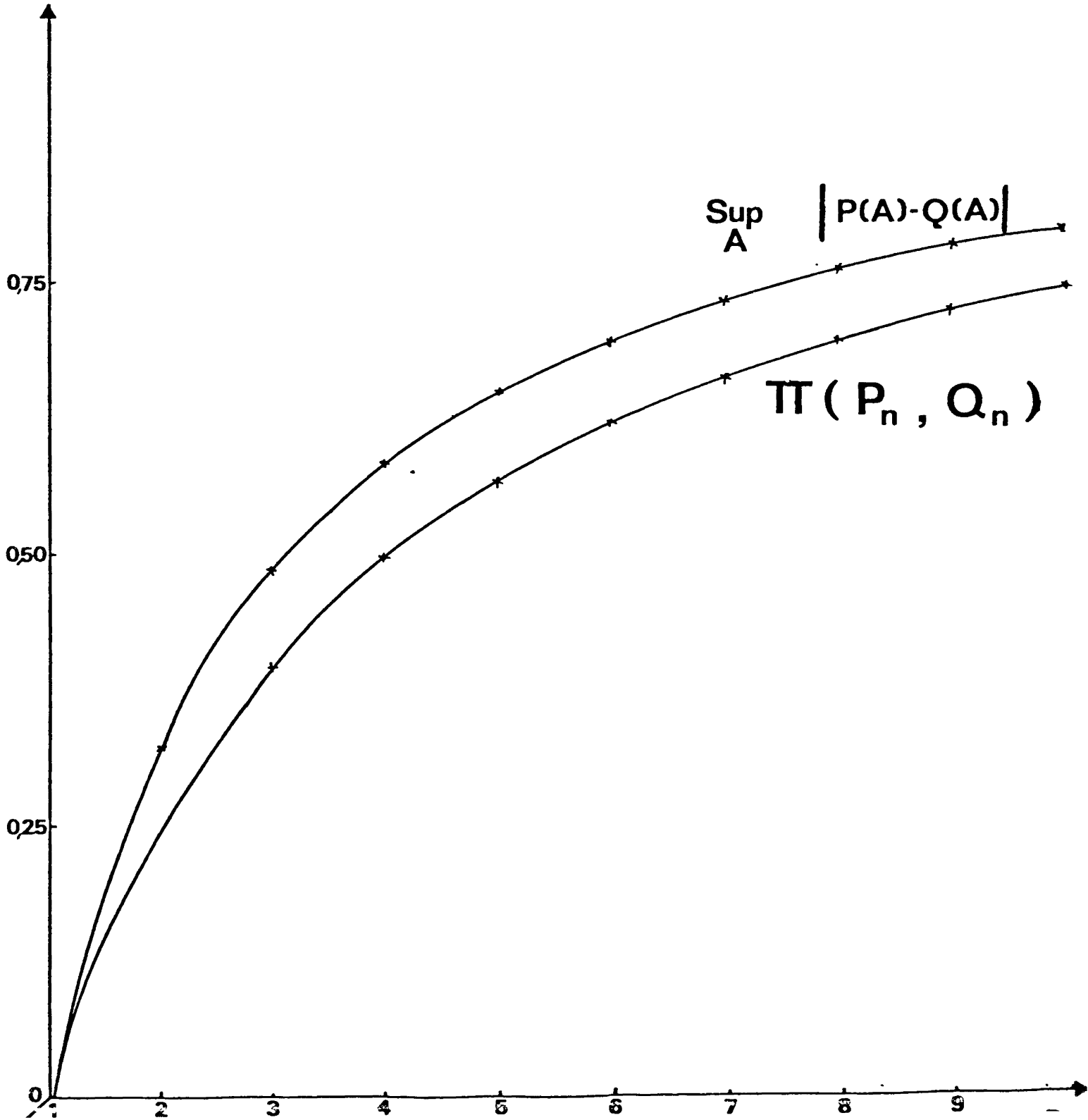
RESULTATS DES CALCULS SUR ORDINATEUR

TABLEAU 14

σ	n.	abscisse max.	α	$F_{\ell_{\max}}$	$\beta = 1 - F_{\ell_{\max}}$	$\Pi(P_n, Q_n)$
3	10	5,88	.39	0,58314	0,41686	
"	"	"	.40	0,60468	0,39532	.4
"	50	7,73	.39	0,60024	0,39976	
"	"	"	.40	0,60448	0,39552	.39976
4	10	7,84	.49	0,48921	0,51079	
"	"	"	.50	0,50636	0,49364	.5
"	50	10,30	.49	0,50293	0,49707	
"	"	"	.50	0,50633	0,49367	.49707
5	10	9,80	.56	0,42621	0,57379	
"	50	12,88	.56	0,43450	0,56550	
"	"	"	.58	0,43794	0,56266	.56550
6	50	15,46	.60	0,38033	0,61967	.61967
7	10	13,72	.6	0,33518	0,66482	.66482
"	50	18,03	.64	0,33913	0,66057	
"	"	"	.66	0,34155	0,65845	.66
8	50	20,61	.69	0,30742	0,69258	.69258
9	10	17,64	.7	0,28007	0,71993	.71993
10	10	19,60	.7	0,25599	0,74400	.74400

COURBE 15

Approximation de la distance de Prokhorov entre $P \sim N(0,1)$ et $Q \sim N(0,\sigma)$ en fonction de σ .



Remarque : pour $n = 10$, $\sigma = 3$ le flot maximum est obtenu en 26 itérations et le calcul a nécessité 1,94 secondes.

Pour $n = 50$, $\sigma = 3$ le flot maximum est obtenu en 125 itérations et le calcul a nécessité 29,94 secondes.

15 Remarque : La mise en oeuvre de l'algorithme de Ford et Fulkerson n'impose nullement que les points x_1, \dots, x_{m+1} chargés par P_n soient les mêmes que ceux chargés par Q_n , ni que les nombres de ces points soient identiques. Le descriptif précédent qui convient particulièrement lorsque P et Q ont des densités de même support, pourra être modifié lorsque les supports sont disjoints ou que l'une au moins des deux lois est discrète. En particulier, lorsque P et Q sont des lois sur des ensembles finis quelconques, l'algorithme de Ford et Fulkerson permet de calculer la valeur exacte de leur distance et Prokhorov.

Merci à C. LACOTE et B. VAN CUTSEM pour leur aide durant de ce travail.

BIBLIOGRAPHIE

- [1] ABRAMOWITZ et STEGUM: Handbook of mathematical functions - National bureau of Standards (1970)
- [2] DUDLEY R. : A.M.S. vol 39, n°5, 1968, p. 1563-72
- [3] FORD et FULKERSON (Algorithme de) : dans SIMONNARD, programmation linéaire Dunod (1962)
- [4] GAREL B.B. : la distance de Prokhorov. Rapport de recherche laboratoire I.M.A.G. (1982) Sera expédié à toute personne qui en fera la demande.
- [5] HAMPEL F.R.: A.M.S. vol. 42, n° 6, 1971, p. 1887-96
- [6] PROKHOROV Y. : Theory of Prob. and its Appl. I n° 2, 1956, p. 157-214
- [7] REY W.J. ; Draft of robust and quasi robust methods ; communication personnelle
- [8] SCHAY G. : Annals of probability, vol 2, n° I, 1974 p.163-6.
- [9] STRASSEN V. : A.M.S., n° 36, (1965) p. 423-39
- [10] SZULGA A. : On minimal metrics in the space of random variables 13th E.M.S. WROCLAW.
- [11] ZOLOTAREV V. : Theory of prob. and its Appl. XX, n° 4, 1975, p. 819-32.