



HAL
open science

Comparaison de deux estimations indépendantes de la densité par la méthode de l'histogramme aléatoire. Lois limites.

Michel Béra

► To cite this version:

Michel Béra. Comparaison de deux estimations indépendantes de la densité par la méthode de l'histogramme aléatoire. Lois limites.. Annales de l'ISUP, 1982, XXVII (1), pp.49-70. hal-03696642

HAL Id: hal-03696642

<https://hal.science/hal-03696642>

Submitted on 16 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

COMPARAISON DE DEUX ESTIMATIONS INDEPENDANTES
DE LA DENSITE PAR LA METHODE DE L'HISTOGRAMME
ALEATOIRE. LOIS LIMITES.

Michel Bera, Assistant-agrégé à l'Université de Paris VI

Key Words: Non parametric methods, density estimation

AMS 1977 Subject classification: Primary 62G10, Secondary 60F05, 60F10, 62E15, 62E20, 62H10.

Summary: This paper deals with the problem of the comparison of two independent estimations of a density f . Given a random variable X valued on $[0,1]^S$, with absolutely continuous d.f. F , we consider two independent samples $(X_{1i}/i=1,n)$ and $(X_{2i}/i=1,n)$ of X , and the related estimators f_{n1} and f_{n2} of $f = dF/dx$, defined by the histogram method. A limit theorem is given for the extreme values of $(f_{n1} - f_{n2})/f_{n1}^{1/2}$ over $[0,1]^S$, valid under very large conditions for f . This theorem can be used for testing the null hypothesis $P_{f_1} = P_{f_2}$. A similar limit theorem is obtained for partial extreme values.

I.S.U.P.

(*) 10 rue J.B.Dumas
75017 Paris

Tour 45-55, 3ème et.
Université Paris VI

Introduction

L'objet du présent article est de comparer deux estimations $f_n^{(1)}$ et $f_n^{(2)}$ d'une même densité f , obtenues en appliquant la méthode de l'histogramme à deux échantillons indépendants de même taille n .

Sous des conditions très larges sur la densité inconnue f , nous obtenons une loi limite pour les valeurs extrêmes de

$$(f_n^{(1)}(x) - f_n^{(2)}(x)) / (f_n^{(1)}(x))^{\frac{1}{2}}$$

sur le pavé $(0,1)^S$.

Notations et définitions

- Dans tout ce qui suit, $L_1 x$ définit la 1^{ère} itérée de la fonction logarithme. Ainsi ($Lx = \text{Log} x$ et $L_2 x = \text{Log}(\text{Log} x)$).

- d_n est une fonction croissante de n à valeurs entières vérifiant $\lim_n d_n = +\infty$, et k est défini par $k = d_n^S$.

- y_1 et y_2 sont deux réels fixés et θ_1 et θ_2 deux fonctions de n définies par

$$\theta_a = (2Lk - L_2 k + y_a)^{\frac{1}{2}} \quad \text{avec } a=1,2.$$

- si I_n est un ensemble d'indices i dépendant de n et T_{in}^1 et T_{in}^2 deux familles de variables aléatoires indicées par i , nous dirons que T_{in}^1 est majoré uniformément en probabi-

lité par T_{in}^2 si $P(\bigcap_{i \in I_n} (T_{in}^1 < T_{in}^2)) \rightarrow 1$ quand $n \rightarrow +\infty$.

- X est une variable aléatoire à valeurs dans $(0,1)^S$, de loi P_f . Nous supposons que la fonction de répartition associée F admet une dérivée f , que nous supposons encadrée sur $(0,1)^S$ par deux réels strictement positifs: $m < f < M$.

Découpons le pavé $(0,1)^S$ en k blocs $B_{nr_1 \dots r_s}$ définis par

$$B_{nr_1 \dots r_s} = \left[\frac{r_1-1}{d_n}, \frac{r_1}{d_n} \right] \times \dots \times \left[\frac{r_s-1}{d_n}, \frac{r_s}{d_n} \right]$$

où chaque r_j ($j=1, \dots, s$) varie de 1 à d_n .

Considérons deux échantillons indépendants de même taille n , (X_{11}, \dots, X_{1n}) et (X_{21}, \dots, X_{2n}) de la loi P_f .

Soit $\gamma_{nr_1 \dots r_s}^1$ (respectivement $\gamma_{nr_1 \dots r_s}^2$) le nombre des X_{1i} (respectivement X_{2i}) contenus dans le bloc $B_{nr_1 \dots r_s}$.

Les deux histogrammes associés aux échantillons ci-dessus sont définis comme suit (a est égal à 1 ou 2):

- si $x \in B_{nr_1 \dots r_s}$, $f_n^{(a)}(x) = \frac{k}{n} \gamma_{nr_1 \dots r_s}^a$.

Etant donné un ordre sur l'ensemble des pavés qui jouxtent la frontière de $[0,1]^S$, on prolonge dans cet ordre l'estimation $f_n^{(a)}(x)$ au centre de ces pavés sur la portion de

frontière qu'ils recouvrent, le prolongement ne s'effectuant que sur ce qui n'aura pas été prolongé précédemment.

$$\text{Notons } Z_1 = (n/k)^{\frac{1}{2}} \max_{x \in [0,1]^S} (f_n^{(1)}(x) - f_n^{(2)}(x)) / (f_n^{(1)}(x))^{\frac{1}{2}}$$

$$\text{et } Z_2 = -(n/k)^{\frac{1}{2}} \min_{x \in [0,1]^S} (f_n^{(1)}(x) - f_n^{(2)}(x)) / (f_n^{(1)}(x))^{\frac{1}{2}}.$$

Il a été montré (Bera(1)) : $P(\inf_{x \in [0,1]^S} (f_n^{(1)}(x)) = 0) \rightarrow 0$

lorsque $n \rightarrow \infty$, ce qui permet d'introduire une division par $(f_n^{(1)}(x))^{\frac{1}{2}}$ dans la définition de Z_1 et Z_2 .

Théorème 1

Si $\frac{k(Lk)^3}{n} \rightarrow 0$ lorsque n tend vers l'infini, alors

$$P(Z_1 < \theta_1 \sqrt{2}, Z_2 < \theta_2 \sqrt{2}) \rightarrow \exp\left(\frac{-1}{2\sqrt{\pi}} \left(e^{-y_1/2} + e^{-y_2/2} \right)\right).$$

(On dira que Z_1 et Z_2 suivent asymptotiquement la loi L.)

Corollaire: Z_1 et Z_2 sont asymptotiquement indépendants au sens de Paul Levy.

Nous allons établir ce théorème en utilisant deux variables Z'_1 et Z'_2 suffisamment proches respectivement de Z_1 et Z_2 .
Introduisons quelques nouvelles notations.

Soit i un indice qui varie de 1 à k déterminant le bloc

$B_{nr_1 \dots r_s}$ et notons successivement

- C_i l'événement $(X \in B_{nr_1 \dots r_s})$,

$$- p_i = P(X \in C_i) = \int_{B_{nr_1 \dots r_s}} f(t) dt,$$

$$- \bar{p} = \max_i p_i \quad \text{et} \quad \underline{p} = \min_i p_i,$$

$$- \gamma_n^i = \gamma_{nr_1 \dots r_s}^a \quad \text{pour} \quad a = 1, 2.$$

Nous pouvons écrire

$$Z_1 = \max_i ((\gamma_n^i - 2\delta_n^i) / (\gamma_n^i)^{\frac{1}{2}})$$

et

$$Z_2 = -\min_i ((\gamma_n^i - 2\delta_n^i) / (\gamma_n^i)^{\frac{1}{2}}).$$

Par ailleurs, lorsque n tend vers l'infini,

$$\bar{p} \ll k^{-1} \sup_{x \in [0,1]^s} f(x) \quad \text{et donc} \quad \bar{p} Lk \text{ tend vers } 0,$$

$$\underline{p} \gg k^{-1} \inf_{x \in [0,1]^s} f(x), \quad \text{soit} \quad (Lk)^3 / \underline{np} < C k (Lk)^3 / n, \quad \text{et donc}$$

$$(Lk)^3 / \underline{np} \text{ tend vers } 0.$$

Posons

$$Z_1' = \max_{i \in [1, k]} ((\gamma_n^i - 2\delta_n^i) / (np_i)^{\frac{1}{2}}),$$

et

$$Z_2' = -\min_{i \in [1, k]} ((\gamma_n^i - 2\delta_n^i) / (np_i)^{\frac{1}{2}}).$$

Nous établirons le théorème 1 à partir des lemmes suivants:

Lemme 1

Si, lorsque n tend vers l'infini, $\bar{p} Lk$ tend vers 0, ce que

nous appellerons condition (I), et $(Lk)^3/np$ tend vers 0, ce que nous appellerons condition (II), alors le couple (Z'_1, Z'_2) suit asymptotiquement la loi L.

Lemme 2

Sous les conditions du Lemme 1, A étant un réel fixé supérieur à $\sqrt{2}$, $|\gamma_n^i - z_n^i| \leq 2A(np_i Lk)^{\frac{1}{2}}$, la majoration en probabilité étant uniforme en i.

Lemme 3

Si le couple (Z'_1, Z'_2) suit asymptotiquement la loi L, et si les quantités $(Lk)^{\frac{1}{2}} (Z_a - Z'_a)$ tendent en probabilité vers 0 lorsque n tend vers l'infini, pour $a = 1, 2$, alors le couple (Z_1, Z_2) suit également asymptotiquement la loi L.

Lemme 4

Sous les conditions du lemme 1, A étant un réel fixé supérieur à $\sqrt{2}$, $\gamma_n^i > np_i - A(np_i Lk)^{\frac{1}{2}}$, et

$$\gamma_n^i \leq np_i + A(np_i Lk)^{\frac{1}{2}},$$

les majorations en probabilité étant uniformes en i.

Les résultats des lemmes 3 et 4 ont été établis par Bera dans (1). Admettons pour l'instant le lemme 2 et montrons que sous les conditions de ce lemme les conditions du lemme 4 sont alors vérifiées.

Considérons $T_1 = (Lk)^{\frac{1}{2}} |Z_1 - Z'_1|$.

Nous avons alors

$$T_1 \leq \max_i (|y_n^i - z_n^i| (y_n^i)^{\frac{1}{2}} - (np_i)^{\frac{1}{2}} |Lk/np_i y_n^i)^{\frac{1}{2}}).$$

Les majorations qui suivent sont uniformes en i :

$$|y_n^i - z_n^i| < 2A(np_i Lk)^{\frac{1}{2}},$$

et

$$|(y_n^i)^{\frac{1}{2}} - (np_i)^{\frac{1}{2}}| < A(np_i Lk)^{\frac{1}{2}} / ((y_n^i)^{\frac{1}{2}} + (np_i)^{\frac{1}{2}}).$$

Comme $y_n^i > D_1 np_i$, où D_1 est un réel fixé inférieur strictement à 1 d'après le lemme 3, il vient

$$((y_n^i)^{\frac{1}{2}} + (np_i)^{\frac{1}{2}})^{-1} < D_2 (np_i)^{-\frac{1}{2}},$$

ce qui montre que T_1 est uniformément majoré en probabilité par $\max_i (2A(np_i Lk)^{\frac{1}{2}} A(np_i Lk)^{\frac{1}{2}} D_2 (np_i)^{-\frac{1}{2}} (Lk)^{\frac{1}{2}} (np_i)^{-\frac{1}{2}} D_1^{-1} (np_i)^{-\frac{1}{2}}$, soit

$$T_1 \leq \max_i (D_3 ((Lk)^3 / np_i)^{\frac{1}{2}}) = D_3 ((Lk)^3 / \underline{np})^{\frac{1}{2}},$$

et donc $T_1 \xrightarrow{p} 0$ lorsque n tend vers l'infini. On montre de même que $T_2 \xrightarrow{p} 0$.

Il reste à montrer les lemmes 1 et 2. La démonstration se fait en cinq étapes.

1ère étape

On peut écrire $P(Z_1' < \theta_1 \sqrt{2}, Z_2' < \theta_2 \sqrt{2}) = P(\bigcap_i G_i)$, où G_i est l'événement

$$(- (2np_i)^{\frac{1}{2}} \theta_1 < y_n^i - z_n^i < (2np_i)^{\frac{1}{2}} \theta_2).$$

Comme $P(\bigcap_i G_i) = 1 - P(\bigcup_i \bar{G}_i)$, où

$$P(\bigcup_i \bar{G}_i) = S_1 - S_2 + \dots + (-1)^{k-1} S_k,$$

et

$$S_r = \sum_{1 \leq i_1 < i_2 < \dots < i_r \leq k} P(\bar{G}_{i_1} \dots \bar{G}_{i_r}).$$

Soit H un entier fixé. Les inégalités de Bonferroni permettent d'encadrer la somme définissant $P(\bigcup_i \bar{G}_i)$ par les deux sommes partielles à l'ordre $H-1$ et à l'ordre H .

Considérons un entier r compris entre 1 et H . Dans ce qui suit, le r -uplet (i_1, \dots, i_r) est supposé vérifier

$$1 \leq i_1 < \dots < i_r \leq k.$$

Nous dirons qu'une fonction $\varepsilon(n)$, qui dépend du r -uplet (i_1, \dots, i_r) , tend vers zéro uniformément, et nous noterons $\varepsilon(n) \rightarrow 0$, s'il existe une fonction $\varepsilon'(n)$ qui ne dépend que de n , tend vers zéro lorsque n tend vers l'infini, et vérifie $|\varepsilon(n)| < \varepsilon'(n)$.

Nous allons montrer que $P(\bar{G}_{i_1} \dots \bar{G}_{i_r})$ est de la forme

$$(1 + \varepsilon(n)) (e^{-y_1/2} + e^{-y_2/2}) / (2\sqrt{\pi}k)^r, \text{ avec } \varepsilon(n) \rightarrow 0.$$

Nous aurons alors

$$S_r = (1 + \varepsilon_r(n)) ((e^{-y_1/2} + e^{-y_2/2}) / (2\sqrt{\pi}))^r / r!, \text{ avec } \varepsilon_r(n) \rightarrow 0.$$

Notons

$$S'_r = ((e^{-y_1/2} + e^{-y_2/2}) / (2\sqrt{\pi}))^r / r!. \text{ Soit } \varepsilon' \text{ fixé positif.}$$

y_1 et y_2 étant fixés, il existe un réel H_0 qui vérifie

$$|\exp(-(2\sqrt{\pi})^{-1} (e^{-y_1/2} + e^{-y_2/2})) - 1 + S'_1 - S'_2 + \dots + (-1)^{K-1} S'_K| < \varepsilon'$$

pour tout K supérieur ou égal à H_0 . Imposons à H d'être

supérieur strictement à H_0 et posons $\varepsilon_H(n) = \sup_r \varepsilon_r(n)$,

où r varie de 1 à H .

Alors $|(1-S_1 + \dots + (-1)^H S_H) - (1-S'_1 + \dots + (-1)^H S'_H)|$ est majoré

par $\varepsilon_H(n)(1+S'_1 + \dots + S'_H)$ et donc par $\varepsilon_H(n)\psi(y_1, y_2)$, où

$$\psi(y_1, y_2) = \exp((2\sqrt{\pi})^{-1} (e^{-y_1/2} + e^{-y_2/2})).$$

La même majoration est appliquée à la différence des deux sommes alternées effectuée à l'ordre $H-1$. Par conséquent

$P(\bigcap_i G_i)$ se trouve encadré par deux quantités dont la limite

lorsque n tend vers l'infini se trouve à une distance

ε' arbitraire de $\exp(-(2\sqrt{\pi})^{-1} (e^{-y_1/2} + e^{-y_2/2}))$ et ceci

achèvera la démonstration du Lemme 1.

2ème étape

Soit A un réel strictement positif. Nous notons pour simplifier ($a = 1$ ou 2)

$$n_q^{(a)} = a \delta_n^i \quad \text{et} \quad \pi_q = p_{i_q},$$

$$a_{F_1}^{(q)} = (n_q^{(a)} < n\pi_q (1 + A(Lk/n\pi_q)^{\frac{1}{2}})),$$

$$a_{F_2}^{(q)} = (n_q^{(a)} > n\pi_q (1 - A(Lk/n\pi_q)^{\frac{1}{2}})),$$

$$E^{(q)} = {}_1F_1^{(q)} \cap {}_2F_2^{(q)} \cap {}_1F_2^{(q)} \cap {}_2F_1^{(q)},$$

$$E = \bigcap_{q=1}^r E^{(q)},$$

$$\text{et } \bar{G}'_{i_q} = \bar{G}_{i_q} \cap E^{(q)}.$$

La difficulté vient de l'impossibilité d'une évaluation directe de $P(\bar{G}_{i_1} \dots \bar{G}_{i_r})$. Nous allons donc remplacer par une quantité équivalente $P(\bar{G}'_{i_1} \dots \bar{G}'_{i_r})$.

Nous avons notamment

$$P(\bar{G}'_{i_1} \dots \bar{G}'_{i_r}) \leq P(\bar{G}_{i_1} \dots \bar{G}_{i_r}),$$

et

$$P(\bar{G}_{i_1} \dots \bar{G}_{i_r}) = P(\bar{G}_{i_1} \dots \bar{G}_{i_r} | E) + P(\bar{G}_{i_1} \dots \bar{G}_{i_r} | \bar{E}),$$

ce qui entraîne

$$P(\bar{G}_{i_1} \dots \bar{G}_{i_r}) < P(\bar{G}'_{i_1} \dots \bar{G}'_{i_r}) + P(\bar{E}).$$

Par conséquent,

$P(\bar{G}_{i_1} \dots \bar{G}_{i_r})$ et $P(\bar{G}'_{i_1} \dots \bar{G}'_{i_r})$ seront équivalents si $P(\bar{E})$ est de la forme

$$\varepsilon(n)P(\bar{G}'_{i_1} \dots \bar{G}'_{i_r}), \text{ avec } \varepsilon(n) \xrightarrow{U} 0.$$

Sous les conditions (I) et (II) du Lemme 1, les majorations suivantes ont été établies (cf Bera dans (I)):

$$P\left(\binom{q}{a \ b}\right) < (1+\varepsilon(n))(\exp(-(A^2/2)Lk(1+\varepsilon(n))))/A(2\pi Lk)^{\frac{1}{2}},$$

où a et $b = 1, 2$ et q varie de 1 à r , avec $\varepsilon(n) \xrightarrow{U} 0$.

Ceci entraîne

$$P(\bar{E}) \leq 4H(1+\varepsilon(n))(\exp(-(A^2/2)Lk(1+\varepsilon(n))))/A(2\pi Lk)^{\frac{1}{2}}, \text{ avec } \varepsilon(n) \rightarrow 0 \text{ quand } n \text{ tend vers l'infini.}$$

3ème étape

Les deux r -uplets $(n_1^{(1)}, \dots, n_r^{(1)})$ et $(n_1^{(2)}, \dots, n_r^{(2)})$ ont une loi de probabilité conjointe définie par le produit de deux lois multinomiales. Par conséquent,

$$P(\bar{G}_{i_1}! \dots \bar{G}_{i_r}!) = \sum_{(n_1^{(1)}, n_1^{(2)}) \in \bar{G}_{i_1}!} \dots \sum_{(n_r^{(1)}, n_r^{(2)}) \in \bar{G}_{i_r}!} \Omega,$$

où $\Omega = \Omega_1 \cdot \Omega_2$, et

$$\Omega_a = (n! / n_1^{(a)}! \dots n_r^{(a)}! (n - N_r^{(a)})!) \pi_1^{n_1^{(a)}} \dots \pi_r^{n_r^{(a)}} (1 - W_r)^{n - N_r^{(a)}},$$

où nous notons

$$N_r^{(a)} = n_1^{(a)} + \dots + n_r^{(a)},$$

et $W_r = \pi_1 + \dots + \pi_r$, a étant un entier égal à 1 ou 2.

Sous les conditions (I) et (II) du lemme 1, chacune des deux lois multinomiales est équivalente à un produit de lois de Poisson lorsque les conditions $F_{ab}^{(q)}$ sont vérifiées pour $a=1,2$ et $b=1,2$ et $q=1$ à r (cf Bera (1)). Cette équivalence est uniforme dans le choix des r -uplets $(i_1 \dots i_r)$.

On peut alors écrire

$$\Omega = (1 + \varepsilon(n)) \prod_{q=1}^r \exp(-2n\pi_q) (n\pi_q)^{n_q^{(1)} + n_q^{(2)}} / n_q^{(1)}! n_q^{(2)}!,$$

avec $\varepsilon(n) \rightarrow 0$.

Intervertissons somme et produit dans l'expression qui définit $P(\bar{G}_{i_1}! \dots \bar{G}_{i_r}!)$. Il vient alors:

$$P(\bar{G}_{i_1}! \dots \bar{G}_{i_r}!) = (1 + \varepsilon(n)) \prod_{q=1}^r \left(\sum_{(n_q^{(1)}, n_q^{(2)}) \in \bar{G}_{i_q}!} Y_q \right),$$

où $\varepsilon(n) \rightarrow 0$, et

$$Y_q = \exp(-2n\pi_q) (n\pi_q)^{n_q^{(1)} + n_q^{(2)}} / n_q^{(1)}! n_q^{(2)}!.$$

4ème étape

La difficulté réside en fait principalement dans l'évaluation asymptotique de la quantité $R =$

$$\sum_{(n_q^{(1)}, n_q^{(2)}) \in \bar{G}_{i_q}!} Y_q.$$

Simplifions à nouveau les notations et posons

$$m_1 = n \binom{(1)}{q}, \quad m_2 = n \binom{(2)}{q}, \quad p = \pi_q,$$

et

$$X(m_1, m_2) = \exp(-2np) \cdot (np)^{m_1+m_2} / m_1! m_2! .$$

La somme R se décompose en deux sommes doubles R_1 et R_2 ,

avec $R_1 = \sum_{D_1} X(m_1, m_2)$, où D_1 est le domaine défini par les

trois inégalités

$$m_1 - m_2 > (2np)^{\frac{1}{2}} \theta_2 = M'_1 ,$$

$$m_1 < np(1+A(Lk/np)^{\frac{1}{2}}) = M'_2 ,$$

$$m_2 > np(1-A(Lk/np)^{\frac{1}{2}}) = M'_3 ,$$

et $R_2 = \sum_{D_2} X(m_1, m_2)$, où D_2 est le domaine défini par les trois inégalités

$$m_1 - m_2 \leq (2np)^{\frac{1}{2}} \theta_1 = M'_4 ,$$

$$m_1 > M'_3 \text{ et } m_2 < M'_2 .$$

Les bornes imposées dans les trois inégalités qui définissent chaque domaine D_1 et D_2 ne sont pas des valeurs entières en général. Il existe des réels h_1, h_2 , et h_3, h_4 majorés en valeur absolue par 1 tels que les quatre expressions $M_i = M'_i + h_i$ (i varie de 1 à 4) soient les bornes entières atteintes dans D_1 et D_2 par m_1, m_2 et $m_1 - m_2$.

D_1 et D_2 sont donc définis par

$$D_1 = (m_1 - m_2 \gg M_1, m_1 \leq M_2, m_2 \gg M_3),$$

$$D_2 = (m_1 - m_2 \leq M_4, m_1 \gg M_3, m_2 \leq M_2) .$$

Pour faciliter la compréhension de ce qui suit, nous avons représenté sur un graphique présenté en Annexe les points

de coordonnées (m_1, m_2) . Les domaines D_1 et D_2 sont constitués par les points à coordonnées entières contenus dans deux triangles. Pour la suite de l'exposé, nous décrirons les différents domaines étudiés à l'aide de ce graphique. Ainsi à D_1 correspond le triangle $(A_1 A_2 E_1)$.

Pour évaluer la somme R_1 , nous allons montrer l'équivalence de cette quantité et de la somme des $X(m_1, m_2)$ sur le domaine défini par la bande B comprise entre la droite $(A_1 A_2)$ (inclusive) d'équation $m_1 - m_2 = M_1$ et la droite $(C_1 C_2)$ (exclusive) d'équation $m_1 - m_2 = M_2 - M_3$.

Remarquons que $M_2 = np(1+A(Lk/np)^{\frac{1}{2}}) - h_2$, avec $0 \leq h_2 < 1$, et $M_3 = np(1-A(Lk/np)^{\frac{1}{2}}) + h_3$, avec $0 \leq h_3 < 1$. Par conséquent $M_2 - M_3 = 2A(npLk)^{\frac{1}{2}} - h$ avec $0 \leq h < 2$.

Nous allons utiliser une méthode de grandes déviations permettant d'évaluer cette nouvelle somme comme la différence de la somme des $X(m_1, m_2)$ pour les points (m_1, m_2) situés dans le demi-plan situé sous la droite $(A_1 A_2)$, d'équation $m_1 - m_2 \geq M_1$ et de la somme des $X(m_1, m_2)$ pour les points (m_1, m_2) situés dans le demi-plan situé sous la droite $(C_1 C_2)$ d'équation $m_1 - m_2 \geq M_2 - M_3$.

Nous utiliserons la fonction g_α définie sur les réels par l'équation $g_\alpha(x) = x^\alpha e^{-x}$. Rappelons que g_α est croissante sur l'intervalle $[0, \alpha]$ et décroissante sur l'intervalle $[\alpha, +\infty)$.

Posons $T^{(1)} = P(X > M_2)$ où X suit une loi de Poisson de paramètre np . Si α est un entier fixé minoré par M_2 , g_α est

croissante dans l'intervalle $[\lfloor np \rfloor, \lfloor np \rfloor + 1]$, où $\lfloor np \rfloor$ est la partie entière de np . Si X' et X'' suivent respectivement des lois de Poisson de paramètre $\lfloor np \rfloor$ et $\lfloor np \rfloor + 1$, il vient alors l'inégalité suivante:

$$U^{(1)} = P(X' > M_2) < T^{(1)} < V^{(1)} = P(X'' > M_2).$$

Nous allons évaluer $U^{(1)}$ et $V^{(1)}$ par la méthode de grandes déviations (voir par exemple Cramer(2) ou Feller(3)).

Théorème de grandes déviations

Considérons une variable aléatoire Y dont la fonction caractéristique est holomorphe dans un voisinage ouvert de l'origine. Soit (Y_1, \dots, Y_N) un échantillon de taille N de Y et soit $F_N(x)$ la fonction de répartition de la somme réduite $(Y_1 + \dots + Y_N - NE(Y)) / \sigma(Y)N^{1/2}$, où $E(Y)$ et $\sigma(Y)$ sont respectivement l'espérance et l'écart-type de Y . Alors, si x est une fonction de N vérifiant $x = o(N^{1/6})$ et si \mathcal{N} est la fonction de répartition de la loi normale centrée réduite, $1 - F_N(x) = (1 + \varepsilon(N))(1 - \mathcal{N}(x))$, la fonction $\varepsilon(N)$ ne dépendant que de la distribution de Y et de N .

Une variable aléatoire qui suit une loi de Poisson de paramètre N peut s'interpréter comme la somme de N variables indépendantes suivant une loi de Poisson de paramètre 1. Appliquons donc le théorème des grandes déviations aux variables X' et X'' avec $N = \lfloor np \rfloor + c$, où $c=0$ ou 1, et Y qui suit une loi de Poisson de paramètre 1 ($E(Y)=1$, $\sigma(Y)=1$) et enfin $x(N) = (M_2 - N) / N^{1/2}$.

D'après ce qui précède, si $c' = np - \lfloor np \rfloor$,

$x(N) = (A(npLk)^{\frac{1}{2}} - h_2 - c + c') / (np + c - c')^{\frac{1}{2}}$, et donc

$x(N) = A(Lk)^{\frac{1}{2}} (1 + O(1/(np)^{\frac{1}{2}}))$. Ceci entraîne que

$x(N)N^{-1/6}$ est majoré par $A((Lk)^3/np)^{1/6} (1 + O(1/(np)^{\frac{1}{2}}))$,

et donc $x(N) = N^{1/6} \varepsilon(N)$, où $\varepsilon(N) \rightarrow 0$.

Par ailleurs, lorsque t tend vers l'infini, on a

$$1 - \eta(t) = \int_t^{+\infty} e^{-x^2/2} dx = t^{-1} e^{-t^2/2} (1 + O(1/t)).$$

Ainsi, $U^{(1)}$ et $V^{(1)}$ sont tous deux de la forme

$$(1 + \varepsilon(n)) \exp(-(A^2/2)Lk) / A(2\pi Lk)^{\frac{1}{2}},$$

ainsi que $T^{(1)}$ qu'ils encadrent.

Nous avons ainsi obtenu une majoration pour la somme $T^{(1)}$

des $X(m_1, m_2)$ pour les points (m_1, m_2) situés dans le demi-plan situé à la droite de la droite $(A_1 E_1)$ d'équation

$m_1 > M_2$. Cette somme $T^{(1)}$ majore la somme $S_1^{(1)}$ des $X(m_1, m_2)$

pour les points (m_1, m_2) situés dans l'intersection du domaine précédent et de la bande B .

On obtient de même une majoration pour la somme $S_1^{(2)}$ des

$X(m_1, m_2)$ situés dans B et vérifiant $m_2 < M_3$. Ainsi

$$S_1^{(1)} \leq (1 + \varepsilon(n)) \exp(-(A^2/2)Lk) / A(2\pi Lk)^{\frac{1}{2}}, \quad \varepsilon(n) \rightarrow 0,$$

et

$$S_1^{(2)} \leq (1 + \varepsilon(n)) \exp(-(A^2/2)Lk) / A(2\pi Lk)^{\frac{1}{2}}, \quad \varepsilon(n) \rightarrow 0.$$

Etablissons maintenant une majoration pour la somme $S_1^{(3)}$

des $X(m_1, m_2)$ pour les points (m_1, m_2) situés dans le domaine

défini comme l'intersection de la bande B et du demi-plan

d'équation $m_1 + m_2 > 2M_2$ situé au dessus de la droite $(B_1 C_1)$.

La somme $S_1^{(3)}$ se majore par la somme complète des $X(m_1, m_2)$ sur le demi-plan d'équation $m_1 + m_2 > 2M_2$, qui peut s'écrire $P(X > 2M_2)$, où X suit une loi de Poisson de paramètre $2np$.

Comme $m_1 + m_2 > 2M_2 = 2np + 2A(npLk)^{\frac{1}{2}} - 2h_2$, avec h_2 compris entre 0 et 1, la fonction $g(x) = x^{m_1+m_2} e^{-2x}$ est croissante sur l'intervalle $[\lfloor 2np \rfloor, \lfloor 2np \rfloor + 1]$. Utilisons comme pour $T^{(1)}$ une technique d'encadrement et une méthode de grandes déviations avec $N = \lfloor 2np \rfloor + h$ ($h = 0$ ou 1) et $x(N) = A(2Lk)^{\frac{1}{2}} (1 + O(1/(np)^{\frac{1}{2}}))$.

On obtient la majoration

$$S_1^{(3)} \leq (1 + \varepsilon(n)) \exp(-A^2 Lk) / 2A(\pi Lk)^{\frac{1}{2}}, \quad \varepsilon(n) \rightarrow 0.$$

Une majoration analogue est obtenue pour la somme $S_1^{(4)}$ des $X(m_1, m_2)$ pour les points de la bande B vérifiant $m_1 + m_2 < 2M_3$. On obtient la même majoration

$$S_1^{(4)} \leq (1 + \varepsilon(n)) \exp(-A^2 Lk) / 2A(\pi Lk)^{\frac{1}{2}}, \quad \varepsilon(n) \rightarrow 0.$$

Bien plus délicat est le problème posé par la somme $S_1^{(5)}$ des $X(m_1, m_2)$ pour les points (m_1, m_2) compris dans le rectangle restant $(B_1 B_2 C_1 C_2)$. En effet, dans ce domaine, la fonction g n'est plus monotone au voisinage de $2np$. C'est pourquoi nous allons majorer directement l'écart entre chaque terme de cette somme et le terme correspondant obtenu en remplaçant dans $X(m_1, m_2)$ np par $\beta = \lfloor np \rfloor$. Notons $t_1 = np - \beta$. On peut écrire en notant cette fois $g(x)$ la fonction $x^{m_1+m_2} e^{-2x}$ ($\neq g_{m_1+m_2}(x)$),

$$g(\beta+t_1) - g(\beta) = t_1 g'(c) = t_1 (m_1+m_2-2c)(c)^{m_1+m_2-1} e^{-2c},$$

$$= t_1 ((m_1+m_2-2c)/c) g(c), \text{ où } c \text{ est un réel compris entre } \beta \text{ et } \beta+t_1.$$

Dans le domaine qui définit $S_1^{(5)}$, il existe un réel W fixé indépendant de n et de (m_1, m_2) tel que $(m_1+m_2-2c)/c$ soit majoré par $u_n = W(Lk/np)^{\frac{1}{2}}$. Choisissons n_0 tel que pour tout n supérieur à n_0 , u_n soit inférieur à $1/2$, et supposons $n > n_0$. Nous pouvons itérer la technique des valeurs intermédiaires, obtenant pour tout ℓ entier la majoration

$$g(\beta+t_1) - g(\beta) \leq u_n^{\ell+1} g(c_\ell) + (u_n/(1-u_n)) g(\beta), \text{ où } c_\ell \text{ est un réel de l'intervalle } [\beta, \beta+t_1].$$

Comme g est continue et donc bornée sur cet intervalle, on peut passer à la limite inférieure sur ℓ , et par conséquent obtenir

$$g(\beta+t_1) - g(\beta) \leq 2u_n g(\beta). \text{ On minore de même } g(\beta+t_1) - g(\beta) \text{ par } -2u_n g(\beta).$$

L'encadrement ainsi obtenu étant uniforme dans le choix du point (m_1, m_2) sur le domaine qui définit $S_1^{(5)}$, cette somme est équivalente à la somme $U_1^{(5)}$ obtenue en remplaçant np par $\beta = \lfloor np \rfloor$ car $u_n \rightarrow 0$.

Définissons de même les deux sommes $U_1^{(3)}$ et $U_1^{(4)}$ obtenues en remplaçant $X(m_1, m_2)$ par $X'(m_1, m_2)$ dans les sommes respectives $S_1^{(3)}$ et $S_1^{(4)}$, où

$$X'(m_1, m_2) = \exp(-2\lfloor np \rfloor) (\lfloor np \rfloor)^{m_1+m_2-1} / m_1! m_2!$$

On remplace cette fois np par sa partie entière.

La somme $U_1^{(3)} + U_1^{(4)} + U_1^{(5)}$ peut s'écrire sous la forme

$$P(X_1 - X_2 \in [M_3'', M_4'']), \text{ où } M_3'' \text{ est à une distance inférieure}$$



à 2 de $(2np)^{\frac{1}{2}}\theta_2$, M_4'' à une distance inférieure à 2 de $2A(npLk)^{\frac{1}{2}}$, X_1 et X_2 sont deux variables indépendantes qui suivent une loi de Poisson de paramètre $N = [np]$. Décomposons X_1 et X_2 en somme de N variables de Poisson indépendantes, de paramètre 1, les $2N$ variables étant indépendantes.

De $X_1 = X_{11} + \dots + X_{1N}$ et $X_2 = X_{21} + \dots + X_{2N}$ on tire $X_1 - X_2 = (X_{11} - X_{21}) + \dots + (X_{1N} - X_{2N})$, ce qui revient à considérer $X_1 - X_2$ comme la somme de N variables indépendantes (Y_1, \dots, Y_N) d'espérance nulle et de variance 2. Les conditions du théorème de grandes déviations sont vérifiées, et ainsi

$$P(X_1 - X_2 \in [M_3'', M_4'']) = (1 - F_N(x_1(N))) - (1 - F_N(x_2(N))), \text{ où } x_1(N) = \theta_2(1 + O(1/(np)^{\frac{1}{2}})) \text{ et } x_2(N) = A(2Lk)^{\frac{1}{2}}(1 + O(1/(np)^{\frac{1}{2}})).$$

Nous obtenons enfin

$$1 - F_N(x_1(N)) = (1 + \varepsilon(N)) \exp(-(y_2/2)) / 2k\sqrt{\pi}, \quad \varepsilon(n) \rightarrow 0, \\ 1 - F_N(x_2(N)) = (1 + \varepsilon(N)) \exp(-A^2Lk) / 2A(\pi Lk)^{\frac{1}{2}}, \quad \varepsilon(n) \rightarrow 0.$$

Supposons A supérieur à $\sqrt{2}$. Alors $\exp(-A^2Lk)$ et $\exp(-(A^2/2)Lk)$ sont des infiniments petits devant k^{-1} , et

$$U_1^{(3)} + U_1^{(4)} + U_1^{(5)} = (1 + \varepsilon(n)) \exp(-(y_2/2)) / 2k\sqrt{\pi}, \quad \varepsilon(n) \rightarrow 0.$$

$U_1^{(3)}$ et $U_1^{(4)}$ sont majorés par les mêmes quantités que $S_1^{(3)}$ et $S_1^{(4)}$, à savoir $(\exp(-A^2Lk) / 2A(\pi Lk)^{\frac{1}{2}}) (1 + \varepsilon(n))$. (cette fois

on change np en $[np]$, ce qui est presque la même chose que de changer $2np$ en $[2np]$).

Nous avons montré à ce stade:

$$U_1^{(3)} = \varepsilon(n)\psi(n), \quad U_1^{(4)} = \varepsilon(n)\psi(n), \quad \text{où } \psi(n) = \exp(-(y_2/2)) / 2k\sqrt{\pi}$$

et $\varepsilon(n) \rightarrow 0$. ceci entraîne $U_1^{(5)} = (1 + \varepsilon(n))\psi(n)$, $\varepsilon(n) \rightarrow 0$.

Mais $S_1^{(5)} = U_1^{(5)}(1 + \varepsilon(n))$, $\varepsilon(n) \rightarrow 0$, et $S_1^{(1)}$, $S_1^{(2)}$, $S_1^{(3)}$ et $S_1^{(4)}$ sont de la forme $\varepsilon(n)\psi(n)$, avec $\varepsilon(n) \rightarrow 0$.

Comme $R_1 = S_1^{(3)} + S_1^{(4)} + S_1^{(5)} - S_1^{(1)} - S_1^{(2)}$,

on en déduit enfin

$$R_1 = (1 + \varepsilon(n))\psi(n) = (1 + \varepsilon(n))\exp(-y_2/2) / 2k\sqrt{\pi}, \quad \varepsilon(n) \rightarrow 0.$$

Un raisonnement symétrique permet d'obtenir

$$R_2 = (1 + \varepsilon(n))\exp(-y_1/2) / 2k\sqrt{\pi}, \quad \varepsilon(n) \rightarrow 0.$$

5ème étape

Les résultats obtenus permettent d'obtenir

$$R = (1 + \varepsilon(n))(2k\sqrt{\pi})^{-1} (e^{-y_1/2} + e^{-y_2/2}), \quad \varepsilon(n) \rightarrow 0.$$

Ceci donne l'équivalent uniforme dans le choix de $(i_1 \dots i_r)$

$$P(\bar{G}_{i_1}^1 \dots \bar{G}_{i_r}^1) = (1 + \varepsilon(n))(2k\sqrt{\pi})^{-r} (e^{-y_1/2} + e^{-y_2/2})^r, \quad \varepsilon(n) \rightarrow 0.$$

L'équivalence entre $P(\bar{G}_{i_1}^1 \dots \bar{G}_{i_r}^1)$ et $P(\bar{G}_{i_1} \dots \bar{G}_{i_r})$ sera assurée dès que $P(\bar{E})$ sera infiniment petit devant $P(\bar{G}_{i_1}^1 \dots \bar{G}_{i_r}^1)$, ce qui est le cas en raison des résultats de l'étape 2 dès que $\exp(-A^2/2)Lk$ est un $o(k^{-H})$, soit $A > (2H)^{\frac{1}{2}}$.

Nous montrons maintenant le lemme 2.

Il faut établir $P(\bigcap_{i=1}^k (|\gamma_n^i - \delta_n^i| < 2A(np_i Lk)^{\frac{1}{2}})) \rightarrow 1$.

Ceci revient à majorer, en raisonnant encore sur l'événement contraire, les quantités $P(\bar{G}_i'')$ définies par

$$P(\bar{G}_i'') = P(|\gamma_n^i - \delta_n^i| > 2A(np_i Lk)^{\frac{1}{2}}), \quad \text{où } \gamma_n^i \text{ et } \delta_n^i \text{ sont deux variables binomiales indépendantes.}$$

Notons $\bar{G}_i^{(3)} = \bar{G}_i'' \cap E^{q(i)}$, où $q(i)$ est le q correspondant

à i comme défini dans les notations de l'étape 1 (i_q est l'un des r indices du r -uplet (i_1, \dots, i_r)).

Comme dans le lemme 1, $P(\bar{G}_i^{(3)}) \leq P(\bar{G}_i'')$ et

$P(\bar{G}_i'') \leq P(\bar{G}_i^{(3)}) + P(\bar{E}^q(i))$, où

$P(\bar{E}^q(i)) \leq 4(1+\varepsilon(n))(\exp(-(A^2/2)Lk(1+\varepsilon(n))))/A(2\pi Lk)^{\frac{1}{2}}$,

avec $\varepsilon(n) \rightarrow 0$. Nous évaluons toujours $P(\bar{G}_i^{(3)})$ par l'équi-

valence entre la loi multinomiale et la loi de Poisson sous le conditionnement défini par $E^q(i)$. Il vient

$P(\bar{G}_i^{(3)}) = (1+\varepsilon(n))P(|m_1 - m_2| > 2A(np_i Lk)^{\frac{1}{2}})$, $\varepsilon(n) \rightarrow 0$,

soit $P(\bar{G}_i^{(3)}) = 2(1+\varepsilon(n))P(m_1 - m_2 > 2A(np_i Lk)^{\frac{1}{2}})$, $\varepsilon(n) \rightarrow 0$,

où nous reprenons maintenant les notations de l'étape 4

de la démonstration du lemme 1, et m_1 et m_2 suivent une loi de Poisson de paramètre p_i . D'après les résultats de cette

4ème étape, $P(\bar{G}_i^{(3)}) = 2(1+\varepsilon(n))\exp(-A^2 Lk)/2A(\pi Lk)^{\frac{1}{2}}$, $\varepsilon(n) \rightarrow 0$,

et finalement, A' étant un réel fixé,

$P(\bar{G}_i'') \leq A' \exp(-(A^2/2)Lk(1+\varepsilon(n)))/(Lk)^{\frac{1}{2}}$, $\varepsilon(n) \rightarrow 0$, ce qui

achève la démonstration du lemme 2 dès que $A > \sqrt{2}$, car

alors $\sum_{i=1}^k P(\bar{G}_i'') \rightarrow 0$.

II - Nous donnons maintenant un théorème concernant la loi

limite conjointe des valeurs extrêmes partielles de

$(f_n^{(1)}(x) - f_n^{(2)}(x))/(f_n^{(1)}(x))^{\frac{1}{2}}$.

Considérons une partition de $[1, k]$ en plusieurs familles

R_1, \dots, R_ℓ où ℓ est un entier fixé. Supposons que le nombre

d'éléments de chaque R_m , que nous noterons $\#R_m$, vérifie

$\#R_m = \lambda_m(n)k$, avec $\lambda_m(n) \rightarrow \lambda_m^0$ réel fixé lorsque $n \rightarrow \infty$.

Soit $y_{11}, \dots, y_{1\ell}, y_{21}, \dots, y_{2\ell}$ un ensemble de 2ℓ réels fixés

et notons θ_{am} l'expression $(2Lk - L_2 k + y_{am})^{\frac{1}{2}}$ pour $a=1,2$ et $m \in [1, \ell]$.

Soit U_m la partie de $[0,1]^S$ définie par l'union des

$B_{nr_1 \dots r_s}$ lorsque l'indice i décrit l'ensemble R_m .

Définissons enfin les valeurs extrêmes partielles

$$Z_1^{(m)} = (n/k)^{\frac{1}{2}} \max_{x \in U_m} ((f_n^{(1)}(x) - f_n^{(2)}(x)) / (f_n^{(1)}(x))^{\frac{1}{2}}),$$

$$\text{soit } Z_1^{(m)} = \max_{i \in R_m} ((\gamma_n^i - 2\gamma_n^i) / (\gamma_n^i)^{\frac{1}{2}}),$$

et

$$Z_2^{(m)} = -(n/k)^{\frac{1}{2}} \min_{x \in U_m} ((f_n^{(1)}(x) - f_n^{(2)}(x)) / (f_n^{(1)}(x))^{\frac{1}{2}}),$$

$$\text{soit } Z_2^{(m)} = -\min_{i \in R_m} ((\gamma_n^i - 2\gamma_n^i) / (\gamma_n^i)^{\frac{1}{2}}).$$

Théorème II

Si $k(Lk)^3/n \rightarrow 0$ lorsque $n \rightarrow \infty$,

$$P(\bigcap_m (Z_1^{(m)} < \theta_{1m} \sqrt{2}), \bigcap_m (Z_2^{(m)} < \theta_{2m} \sqrt{2}))$$

$$\rightarrow \exp(-1/2\sqrt{\pi}) \sum_{m=1}^{\ell} \int_m^0 (e^{-y_{1m}^2/2} + e^{-y_{2m}^2/2}).$$

Ce théorème généralise le théorème 1 et montre l'indépendance asymptotique des statistiques $Z_a^{(m)}$ ($a=1,2$ et $m \in [1, \ell]$).

La démonstration suit le même schéma que celle du théorème 1. Nous généralisons la définition de l'événement G_i intro-

duit dans l'étape 1 de la démonstration du lemme 1 en

$$\text{posant } G_i = (-(2np_i)^{\frac{1}{2}} \theta_{1m} \leq \gamma_n^i - 2\gamma_n^i \leq (2np_i)^{\frac{1}{2}} \theta_{2m}),$$

où m est l'indice qui vérifie $i \in R_m$.

Pour évaluer la quantité $P(\bar{G}_{i_1} \dots \bar{G}_{i_r})$, nous supposons

que dans le r -uplet (i_1, \dots, i_r) ,

q_1 des i_α proviennent de R_1 ,

...

q_ℓ des i_α proviennent de R_ℓ ,

avec bien entendu $q_1 + \dots + q_\ell = r$.

Reprenons maintenant les notations simplifiées du lemme 1,

4ème étape. Si $i_\alpha \in R_m$, l'expression (2)

$$\sum_{(n_\alpha^{(1)}, n_\alpha^{(2)}) \in \bar{G}_{i_\alpha}} \exp(-2n\pi_\alpha) (n\pi_\alpha)^{n_\alpha^{(1)} + n_\alpha^{(2)}} / n_\alpha^{(1)}! n_\alpha^{(2)}!$$

est de la forme

$$(1 + \varepsilon(n)) (1/2\sqrt{\pi}) k^{-1} (e^{-y_{1m}/2} + e^{-y_{2m}/2}), \text{ avec } \varepsilon(n) \rightarrow 0.$$

Les majorations utilisées dans le lemme 1 permettent alors

d'obtenir sous les conditions du théorème II

$$P(\bar{G}_{i_1} \dots \bar{G}_{i_r}) = (1 + \varepsilon(n)) k^{-r} (2\sqrt{\pi})^{-r} \prod_{j=1}^{\ell} (e^{-y_{1j}/2} + e^{-y_{2j}/2})^{q_j}.$$

Il reste à évaluer la somme

$$S_r = \sum_{1 \leq i_1 < \dots < i_r \leq k} P(\bar{G}_{i_1} \dots \bar{G}_{i_r}).$$

Le nombre des r -uplets (i_1, \dots, i_r) analogues à celui étudié

plus haut est donné par l'expression

$$C_{\#R_1}^{q_1} \dots C_{\#R_\ell}^{q_\ell} = C_{\mathcal{M}_1(n)k}^{q_1} \dots C_{\mathcal{M}_\ell(n)k}^{q_\ell}.$$

Ainsi,

$$S_r = (1 + \varepsilon(n)) (2k\sqrt{\pi})^{-r} \sum_{q_1} \dots \sum_{q_\ell} C_{\mathcal{M}_1(n)k}^{q_1} \dots C_{\mathcal{M}_\ell(n)k}^{q_\ell} \cdot W, \text{ où}$$

$$W = \prod_{j=1}^{\ell} (e^{-y_{1j}/2} + e^{-y_{2j}/2}), \text{ avec } q_1 + \dots + q_\ell = r \text{ et } \varepsilon(n) \rightarrow 0.$$

Remarquons maintenant que lorsque n tend vers l'infini, $C_{\Lambda_j(n)k}^{q_j} k^{-q_j} \rightarrow (\Lambda_j^0)^{q_j} / q_j!$, et que le nombre des termes de la somme S_r est borné par une constante qui ne dépend pas

de n . Par conséquent,

$$S_r \rightarrow (2\sqrt{\pi})^{-r} (1/r!) \left(\sum_{j=1}^r \Lambda_j^0 (e^{-y_1 j/2} + e^{-y_2 j/2}) \right)^r, \text{ ce qui}$$

achève la démonstration du théorème II.

Bibliographie

- (1) Beré, Thèse 3ème cycle, Université de Paris 6, 1977.
- (2) Cramer, Sur un nouveau théorème limite de la théorie des probabilités, Actual. Sci. et Ind., n°736, Paris, 1938.
- (3) Feller W., Generalization of a probability theorem of Cramer, Trans. Amer. Math. Soc., 54, 1943, pp. 361-372.

Annexe

Nous donnons en Annexe plusieurs diagrammes permettant de suivre les développements de la partie 4 de la démonstration du lemme 1 du théorème I. Les coordonnées des points de la figure 1 sont les suivantes:

$$A_1 = (M_2, M_2 - M_1) \sim (np + A(npLk)^{\frac{1}{2}}, np + A(npLk)^{\frac{1}{2}} - (2np)^{\frac{1}{2}}\theta_2)$$

$$A_2 = (M_3 - M_1, M_3) \sim (np - A(npLk)^{\frac{1}{2}} - (2np)^{\frac{1}{2}}\theta_2, np - A(npLk)^{\frac{1}{2}})$$

$$B_1 = (M_2 + M_1/2, M_2 - M_1/2) \sim (np - A(npLk)^{\frac{1}{2}} + \frac{1}{2}(2np)^{\frac{1}{2}}\theta_2, np + A(npLk)^{\frac{1}{2}} - \frac{1}{2}(2np)^{\frac{1}{2}}\theta_2)$$

$$B_2 = (M_3 + M_1/2, M_3 - M_1/2) \sim (np - A(npLk)^{\frac{1}{2}} + \frac{1}{2}(2np)^{\frac{1}{2}}\theta_2, np - A(npLk)^{\frac{1}{2}} - \frac{1}{2}(2np)^{\frac{1}{2}}\theta_2)$$

$$E_1 = (M_2, M_3) \simeq (np + A(npLk)^{\frac{1}{2}}, np - A(npLk)^{\frac{1}{2}})$$

$$H_1 = (M_2, np) \simeq (np + A(npLk)^{\frac{1}{2}}, np)$$

$$H_2 = (M_3, np) \simeq (np, np - A(npLk)^{\frac{1}{2}})$$

$$C_0 = (np, np)$$

Le symbole \simeq signifie un écart inférieur à 2 entre les coordonnées, provenant des problèmes de borne entière rencontrés dans le lemme 1.

Les trois droites qui semblent concourir en C_1 et C_2 ne sont pas concourantes. Ces points sont entourés d'un petit cercle.

Reçu en Octobre 1981

Figure 1:

Etude locale de la
somme R_1 -

(Les petits graphiques
suivants décrivent
chaque somme
intermédiaire)

