



HAL
open science

Lois limites des écarts extrêmes associés aux histogrammes et à diverses statistiques d'ordre dans l'estimation d'une densité de probabilité

Michel Béra

► **To cite this version:**

Michel Béra. Lois limites des écarts extrêmes associés aux histogrammes et à diverses statistiques d'ordre dans l'estimation d'une densité de probabilité. Statistiques [math.ST]. Université Pierre et Marie Curie - Paris VI, 1977. Français. NNT: . tel-00751542

HAL Id: tel-00751542

<https://theses.hal.science/tel-00751542>

Submitted on 13 Nov 2012

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.

THESE présentée

pour l'obtention

du

DIPLOME de DOCTEUR de 3ÈME CYCLE

à

L'UNIVERSITE PIERRE ET MARIE CURIE

- Paris 6 -

Spécialité : Mathématiques

Mention : Statistiques

par M. Michel BERA

Sujet de la thèse: Lois limites des écarts extrêmes associés aux histogrammes et à diverses statistiques d'ordre dans l'estimation d'une densité de probabilité

Soutenue le : 13 juin 1977 devant la commission composée de

M DUGUE	Président
M GEFFROY	Rapporteur
M BENZECRI	Examineur
M KREWERAS	Invité
M DEHEUVELS	Invité
M CHEVALIER	Invité

THESE présentée

pour l'obtention

du

DIPLOME de DOCTEUR de 3 ÈME CYCLE

à

L'UNIVERSITE PIERRE ET MARIE CURIE

- Paris 6 -

Spécialité : Mathématiques

Mention : Statistique

par M. Michel BERA .

Sujet de la thèse: Lois limites des écarts extrêmes associés aux histogrammes et à diverses statistiques d'ordre dans l'estimation d'une densité de probabilité

Soutenue le : 13 juin 1977 devant la commission composée de:

M DUGUE

Président

M GEFFROY

Rapporteur

M BENZECRI

Examineur

M KREWERAS

Invité

M DEHEUVELS

Invité

M CHEVALIER

Invité

Je tiens à remercier Mr le professeur D. Dugué qui m'a fait l'honneur de présider ce jury, messieurs les professeurs Benzecri, Kreweras, Geffroy ainsi que messieurs Chevalier et Dehevels qui ont accepté d'en être membres.

Je tiens en particulier à exprimer ma profonde reconnaissance au professeur Geffroy pour l'aide importante qu'il a apportée à la réalisation de ce travail, et à Jacques Chevalier dont les conseils m'ont beaucoup aidé à mes débuts.

INTRODUCTION

L'étude que nous présentons ici est essentiellement consacrée aux problèmes de lois limites associées à un échantillon (X_1, X_2, \dots, X_n) ; Nous étudierons également l'application de ces lois limites à l'estimation de la densité dans \mathbb{R}^S .

Notre travail est divisé en trois chapitres :

Dans le premier chapitre, nous étudions certains aspects de la loi multinomiale, et nous dégagons une loi limite sur les valeurs extrêmes des problèmes d'occupation : C_1, C_2, \dots, C_k étant k événements s'excluant mutuellement, de probabilités respectives p_1, p_2, \dots, p_k , on considère n épreuves et on étudie les valeurs extrêmes de la variable aléatoire :

$$c_i^{(1)} = \frac{v_n^i - np_i}{\sqrt{np_i}} \quad \text{où } i \in [1, k] \quad (0-1)$$

La loi limite, déjà obtenue par Révész (12) est obtenue ici sous des conditions plus faibles. Nous passons ainsi en particulier de :

$$k < n^{1-\epsilon} \quad \text{à} \quad \frac{kL^3k}{n} \rightarrow 0 \quad (\text{th } (1-1))$$

Par ailleurs, notre loi limite fait apparaître des propriétés d'indépendance asymptotique entre $\min(c_i^{(1)})$ et $\max(c_i^{(1)})$. Nous étudions ensuite la loi limite conjointe des extremums partiels de $(0-1)$, lorsque i décrit un ensemble d'indices R_j . (th (1-2)).

Nous dégageons enfin des applications des th (1 - 1) et (1 - 2) à l'estimation de la densité dans \mathbb{R}^S par la méthode de l'histogramme dans deux cas: (f est une densité de probabilité uniformément continue)

- sur un compact de \mathbb{R}^S , découpé en pavés de taille égale, nous étudions la loi limite des valeurs extrêmes de:

$$H'_n = \frac{f - f_n}{\sqrt{f_n}} \quad (\text{th (1 - 3 - 1)})$$

- sur un domaine croissant vers le support de f , les pavés étant cette fois de probabilité égale et déterminés par f , nous étudions la loi limite des valeurs extrêmes de:

$$H_n = \frac{f - f_n}{f_n} \quad (\text{th (1 - 3 - 2)})$$

Ces théorèmes généralisent ceux de Revesz (12)(13). Si le th (1 - 3 - 1) peut servir à déterminer un intervalle de confiance pour f , le th (1 - 3 - 2) permet de construire des tests.

Dans le second chapitre, nous étudions la loi conjointe des écarts $\sigma_i = Y_{i+1}^* - Y_i^*$, où les (Y_i^*) sont extraits d'un échantillon ordonné de taille n de la loi uniforme sur $|0,1|$, et déterminent un découpage de $|X_1^*, X_n^*|$ en k blocs équivalents (voir Lecoutre (9)).

Nous donnons la loi limite des valeurs extrêmes de

$$c_i^{(2)} = \sqrt{\frac{n}{k}} (k\sigma_i - 1) \quad (\text{th (2 - 1)})$$

qui fait apparaître des propriétés d'indépendance entre $\max(c_i^{(2)})$ et $\min(c_i^{(2)})$. Nous généralisons cette loi limite au cas des extremums partiels (i décrit un ensemble R_j) (th (2 - 2)). Enfin nous généralisons la notion de bloc équilibré, les lois limites demeurant valides. Ces théorèmes généralisent ceux de Tusnady (15)

Le troisième chapitre est consacré à l'estimation de la densité par la méthode de l'histogramme aléatoire (voir Tusnady (15) et plus récemment Abou-Jaoudé (1)). Nous appliquons les résultats du chapitre 2

et nous étudions les valeurs extrêmes de:

$$H_n = \frac{f - f_n}{f_n}$$

où x décrit un intervalle I_n , déterminé par l'échantillon, qui croît vers le support de f à l'infini. Nous étudions les cas particuliers:

- f minorée sur un compact: th (3 - 1)
- f décroissante sur $|a, +\infty|$ th (3 - 2)
- f nulle en a , croissante sur $|a, b|$ th (3 - 3)

Nous utilisons alors le th (2 - 2) pour conclure sur tout le support de f , considéré comme réunion d'intervalles disjoints: th (4 - 3).

Ces théorèmes sont entièrement originaux.

Les lois limites que nous établissons à partir de la loi multinomiale et de la loi de Dirichlet semblent provenir du caractère poissonnien des processus d'échantillonnage: plus précisément, il est possible d'obtenir des conditions plus fines (suffisantes) en travaillant sur la loi de Poisson qu'en utilisant le théorème central limite et la loi de Gauss qui en découle. Cette propriété, prévue par le Pr Geffroy, et utilisée par Zéboulon (18) dans une thèse préparée avec le Pr Geffroy, nous semble être très générale, et devrait nous permettre dans des travaux ultérieurs d'obtenir des résultats intéressants dans plusieurs domaines.

CHAPITRE PREMIER

Dans ce chapitre, nous nous proposons d'étudier la loi multinomiale et son approximation par une loi de Poisson. Par une méthode de conditionnement, nous dégagerons une loi limite sur les valeurs extrêmes des problèmes d'occupation (th (1 - 1)).

Cette loi limite, déjà obtenue par Revesz (12), est ici obtenue sous des conditions plus faibles que celles de Revesz. De plus, nous donnons la loi conjointe du min et du max, faisant apparaître une indépendance asymptotique entre ces deux statistiques.

Nous montrons ensuite un prolongement de notre loi limite aux valeurs extrêmes partielles (th (1 - 2)). (Nous comptons utiliser ce résultat dans des travaux ultérieurs)

Enfin nous rappelons brièvement les applications de Revesz de ces théorèmes à l'estimation de la densité sur \mathbb{R} par la méthode de l'histogramme. Nous utiliserons nos propres résultats pour énoncer deux théorèmes limites plus généraux relatifs à l'estimation de la densité sur \mathbb{R}^s par la méthode de l'histogramme: (th (1 - 3 - 1) et th (1 - 3 - 2))

§ I - 1 Notations et exposé du problème

Soient (C_1, C_2, \dots, C_k) k événements disjoints de probabilités respectives (p_1, p_2, \dots, p_k) .

On considère n épreuves et on pose:

$$\nu_n^i = \text{nombre de réalisations de } C_i \quad (1-1-1)$$

Nous noterons dans tout ce travail θ_j l'expression:

$$\theta_j = (2Lk - L_2^k + y_j)^{1/2} \quad (1-1-2)$$

où $y_j \in \mathbb{R}$ et $j = 1$ ou 2 .

Définissons les événements:

$$L_i = \{ \nu_n^i < np_i - \sqrt{np_i} \theta_1 \} \quad (1-1-3)$$

$$H_i = \{ \nu_n^i > np_i + \sqrt{np_i} \theta_2 \} \quad (1-1-4)$$

Nous supposerons que k est une fonction de n qui tend vers l'infini avec n . Par ailleurs, si nous posons:

$$\bar{p} = \max_{i \in [1, k]} p_i \quad (1-1-5)$$

$$\underline{p} = \min_{i \in [1, k]} p_i \quad (1-1-6)$$

Nous supposerons que \underline{p} et \bar{p} vérifient:

$$\bar{p} Lk \rightarrow 0 \quad (1-1-7)$$

$$\frac{Lk}{np} \rightarrow 0 \quad (1-1-8)$$

On remarquera que (1-1-8) entraîne:

$$\frac{kLk}{n} \rightarrow 0 \quad \text{ou} \quad k = o\left(\frac{n}{Ln}\right) \quad (1-1-9)$$

Nous nous proposons dans ce chapitre d'étudier les statistiques réduites:

$$Z_1 = - \min_{i \in [1, k]} \left(\frac{v_n^i - np_i}{\sqrt{np_i}} \right) \quad (1 - 1 - 10)$$

$$Z_2 = \max_{i \in [1, k]} \left(\frac{v_n^i - np_i}{\sqrt{np_i}} \right) \quad (1 - 1 - 11)$$

Définissons les événements:

$$\bar{G}_i = L_i + H_i \quad (1 - 1 - 12)$$

Nous allons établir le th (1 - 1):

THEOREME 1 - 1

$$\text{Si } \bar{p} L_k \rightarrow 0 \quad \text{et} \quad \frac{(Lk)^3}{np} \rightarrow 0.$$

$$P (Z_1 < \theta_1, Z_2 < \theta_2) \underset{n \rightarrow \infty}{\rightarrow} \exp \left[- \frac{1}{2\sqrt{\pi}} \left(e^{-\left(\frac{y_1}{2}\right)} + e^{-\left(\frac{y_1}{2}\right)} \right) \right]$$

où θ_1 et θ_2 sont définis en (1 - 1 - 2)

On remarquera l'indépendance asymptotique au sens de Lévy de Z_1 et Z_2 .

Nous allons utiliser la méthode de Revesz [12] et Woodrofe [17].

Comme

$$P (Z_1 < \theta_1, Z_2 < \theta_2) = P (\bigcap_i G_i) \quad (1 - 1 - 13)$$

et

$$P (\bigcap_i G_i) = 1 - P (\bigcup_i \bar{G}_i) \quad (1 - 1 - 14)$$

La formule de Poincaré permet d'écrire:

$$P \left(\bigcup_i \bar{G}_i \right) = S_1 - S_2 + S_3 - \dots + (-1)^{k-1} S_k \quad (1-1-15)$$

où

$$S_r = \sum_{1 \leq i_1 < i_2 < \dots < i_r \leq k} P \left(\bar{G}_{i_1} \dots \bar{G}_{i_r} \right) \quad (1-1-16)$$

Les inégalités de Bonferroni permettent d'encadrer $P \left(\bigcup_i \bar{G}_i \right)$ par les sommes partielles de (1-1-15) à l'ordre r et $r+1$.

Soit un entier H fixé; étudions pour tous les r inférieurs à H les quantités:

$$P \left(\bar{G}_{i_1} \dots \bar{G}_{i_r} \right)$$

Nous allons montrer que si (1-1-7) est vérifié et si

$$\frac{(Lk)^3}{n^p} \rightarrow 0 \quad (1-1-16)$$

$$k^r P \left(\bar{G}_{i_1} \dots \bar{G}_{i_r} \right) = \left[\frac{1}{2\sqrt{\pi}} \left(e^{-\left(\frac{y_1}{2}\right)} + e^{-\left(\frac{y_2}{2}\right)} \right) \right]^r (1 + \varepsilon(n))$$

$$(1-1-17)$$

Rappelons que nous notons $\varepsilon(n)$ une fonction qui tend vers 0, lorsque n tend vers l'infini, d'une manière uniforme et qui ne dépend que de n .

Alors énonçons la proposition 1-1:

PROPOSITION 1 - 1

$$S_r = \left(C_k^r k^{-r} \right) \left(\frac{1}{2\sqrt{\pi}} \left(e^{-\left(\frac{y_1}{2}\right)} + e^{-\left(\frac{y_2}{2}\right)} \right) \right)^r (1 + \varepsilon(n)) \quad (1-1-18)$$

Comme

$$C_k^r k^{-r} \rightarrow \frac{1}{r!} \quad \text{lorsque } n \rightarrow \infty$$

$$S_r \rightarrow \frac{1}{r!} \left[\frac{1}{2\sqrt{\pi}} \left(e^{-\left(\frac{y_1}{2}\right)^2} + e^{-\left(\frac{y_2}{2}\right)^2} \right) \right]^r \quad (1-1-19)$$

Les sommes partielles de (1-1-15) convergeront vers les sommes partielles du développement en série de:

$$1 - \exp \left[-\frac{1}{2\sqrt{\pi}} \left(e^{-\left(\frac{y_1}{2}\right)^2} + e^{-\left(\frac{y_2}{2}\right)^2} \right) \right] \quad (1-1-20)$$

d'où l'énoncé du théorème (1-1).

Nous allons donc établir la proposition (1-1).

§ I - 2 Etude de $P(\bar{G}_{i_1} \dots \bar{G}_{i_r})$. Autres notations.

Posons pour simplifier les notations:

$$n_1 = \nu_n^{i_1}, \quad n_2 = \nu_n^{i_2}, \quad \dots, \quad n_r = \nu_n^{i_r} \quad (1-2-1)$$

$$N_r = n_1 + n_2 + \dots + n_r \quad (1-2-2)$$

$$\pi_1 = p_{i_1}, \quad \dots, \quad \pi_r = p_{i_r} \quad (1-2-3)$$

$$W_r = \pi_1 + \dots + \pi_r \quad (1-2-4)$$

Définissons les événements (q décrit ici $[1, r]$)

$$F_1^{(q)} = \left\{ n_q < n\pi_q \left(1 + A \left(\frac{Lk}{n\pi_q} \right)^{1/2} \right) \right\} \quad (1-2-5-1)$$

$$F_2^{(q)} = \left\{ n_q > n\pi_q \left(1 - A \left(\frac{Lk}{n\pi_q} \right)^{1/2} \right) \right\} \quad (1-2-5-2)$$

Il est clair que les n_q sont des entiers. les deux événements ci-dessus peuvent s'exprimer au moyen de la fonction partie entière avec des symboles \leq ; nous y reviendrons plus loin. Convenons de noter:

$\lfloor x \rfloor$: la partie entière de x .

Définissons de même:

$$E_1^* = \bigcap_q F_1^{(q)} \quad \text{et} \quad E_2^* = \bigcap_q F_2^{(q)} \quad (1-2-6)$$

Nous poserons encore:

$$M_q = n\pi_q + \sqrt{n\pi_q} \theta_2 \quad (1-2-7)$$

$$m_q = n\pi_q - \sqrt{n\pi_q} \theta_1 \quad (1-2-8)$$

$$L_{i_q} = \{ 0 \leq n_q < m_q \} \quad (1-2-9)$$

$$H_{i_q} = \{ n_q > M_q \} \quad (1-2-10)$$

Supposons A (défini en (1 - 2 - 5 - 1) et (1 - 2 - 5 - 2))
réel positif fixé supérieur à $\sqrt{2}$ strictement.

Définissons alors pour tout q de $[1, r]$;

$$L_{i_q}^* = L_{i_q} \cap F_2^{(q)} \quad (1-2-11)$$

soit:

$$L_{i_q}^* = \{ n\pi_q (1 - (\frac{Lk}{n\pi_q})^{1/2A}) < n_q < n\pi_q - \sqrt{n\pi_q} \theta_1 \} \quad (1-2-12)$$

De même on définira:

$$H_{i_q}^* = H_{i_q} \cap F_1^{(q)} \quad (1-2-13)$$

soit:

$$H_{i_q}^* = \{ n\pi_q + \sqrt{n\pi_q} \theta_2 < n_q < n\pi_q (1 + (\frac{Lk}{n\pi_q})^{1/2A}) \} \quad (1-2-14)$$

Enfin on définit $G_{i_q}^*$ par:

$$G_{i_q}^* = L_{i_q}^* + H_{i_q}^* \quad (1-2-15)$$

Ces notations seront utilisées plus loin dans notre méthode de conditionnement.

Nous avons donc:

$$\begin{aligned}
 P (\bar{G}_{i_1} \dots \bar{G}_{i_r}) &= \\
 &= \sum_{\bar{G}_{i_1}} \dots \sum_{\bar{G}_{i_r}} \frac{n!}{n_1! n_2! \dots n_r! (n - N_r)!} \pi_1^{n_1} \pi_2^{n_2} \dots \pi_r^{n_r} (1 - W_r)^{(n - N_r)} \\
 & \hspace{15em} (1 - 2 - 16)
 \end{aligned}$$

Nous allons rapprocher cette loi multinomiale d'une loi de Poisson.

§ I - 3 Loi multinomiale et loi de Poisson

Supposons que dans (1 - 2 - 16) chaque n_i puisse s'écrire

$$n_i = n\pi_i \left(1 + \rho_i \left(\frac{Lk}{n\pi_i} \right)^{1/2} \right) \quad (1 - 3 - 1)$$

où ρ_i est un réel vérifiant $|\rho_i| \leq C$ où C est fixé. Alors énonçons la proposition (1-2):

PROPOSITION 1 - 2

Sous les conditions (1 - 1 - 7): $Lk \bar{p} \rightarrow 0$, (1 - 1 - 8): $\frac{Lk}{np} \rightarrow 0$ et

(1 - 3 - 1), on a:

$$\begin{aligned}
 &\frac{n!}{n_1! n_2! \dots n_r! (n - N_r)!} \pi_1^{n_1} \pi_2^{n_2} \dots \pi_r^{n_r} (1 - W_r)^{(n - N_r)} = \\
 &= (1 + \epsilon(n)) \frac{\pi_1^{n_1} \pi_2^{n_2} \dots \pi_r^{n_r}}{n_1! n_2! \dots n_r!} n^{N_r} \exp(-nW_r) \quad (1 - 3 - 2)
 \end{aligned}$$

$$= (1 + \epsilon(n)) \prod_{q \in [1, r]} \left[\frac{(n\pi_q)^{n_q}}{n_q!} \exp(-n\pi_q) \right] \quad (1 - 3 - 3)$$

Démontrons la proposition (1 - 2). Il suffit de montrer que:

$$T_1 = \frac{n!}{(n - N_r)!} \frac{(1 - W_r)^{(n - N_r)}}{n^{N_r}} \exp(-nW_r) \quad (1 - 3 - 4)$$

est de la forme $1 + \varepsilon(n)$

Comme

$$\rho_i \left(\frac{Lk}{n\pi_j} \right)^{1/2} < C \left(\frac{Lk}{n\underline{p}} \right)^{1/2} \quad \text{qui tend vers 0 d'après (1-1-8),}$$

ε étant fixé, au delà d'un certain n

$$n_j < n\pi_j (1 + \varepsilon) \quad \text{et} \quad N_r < nH\bar{p} (1 + \varepsilon) \quad (1-3-5)$$

car $r \leq H$

Mais $\bar{p} \rightarrow 0$ quand $n \rightarrow \infty$ d'après (1-1-7) et donc:

$$n - N_r > \frac{n}{2} \quad \text{au delà d'un certain rang, et ainsi:}$$

$$\left(\frac{n}{n - N_r} \right)^{1/2} = 1 + \varepsilon(n) \quad (1-3-6)$$

On applique alors la formule de Stirling à (1-3-4) pour obtenir:

$$T_1 = \exp [nLn - n - (n - N_r)L(n - N_r) + n - N_r + (n - N_r)L(1 - W_r) - N_rLn + nW_r] (1 + \varepsilon(n)) \quad (1-3-7)$$

que nous écrivons:

$$T_1 = [\exp(U_1)] (1 + \varepsilon(n))$$

mais

$$U_1 = -n \left[\left(1 - \frac{N_r}{n}\right)L\left(1 - \frac{N_r}{n}\right) - \left(1 - \frac{N_r}{n}\right)L(1 - W_r) - W_r + \frac{N_r}{n} \right] \quad (1-3-8)$$

soit

$$U_1 = -n (1 - W_r) \left[\left(1 - \frac{N_r}{n}\right)(1 - W_r)^{-1} L \left[\left(1 - \frac{N_r}{n}\right)(1 - W_r)^{-1} \right] + 1 - \left(1 - \frac{N_r}{n}\right)(1 - W_r)^{-1} \right] \quad (1-3-9)$$

Définissons ψ_n par:

$$\psi_n = - \left(1 - \frac{N_r}{n}\right)(1 - W_r)^{-1} + 1 \quad (1-3-10)$$

U_1 peut s'écrire ainsi:

$$U_1 = -n (1 - W_r) [(1 - \psi_n)L(1 - \psi_n) + \psi_n]$$

mais d'après (1 - 3 - 1)

$$N_r = nW_r + \sqrt{nLk} \sum_i \rho_i \sqrt{\pi_i} \quad (1 - 3 - 11)$$

et

$$1 - \frac{N_r}{n} = 1 - W_r - \left(\frac{Lk}{n}\right)^{1/2} \sum_i \rho_i \sqrt{\pi_i} \quad (1 - 3 - 12)$$

et enfin

$$\psi_n = \left(\frac{Lk}{n}\right)^{1/2} (1 - W_r)^{-1} \sum_i \rho_i \sqrt{\pi_i} \quad (1 - 3 - 13)$$

Donc

$$\psi_n < \left(\frac{Lk}{n}\right)^{1/2} (1 - H\bar{p})^{-1} HC (\bar{p})^{1/2} \quad (1 - 3 - 14)$$

expression qui tend uniformément vers 0 ;

On peut alors développer U_1 :

$$U_1 = -n (1 - W_r) \frac{\psi_n^2}{2} (1 + \varepsilon(n)) \quad (1 - 3 - 15)$$

$|U_1|$ est majoré par

$$(1 - H\bar{p})^{-1} H^2 C^2 (Lk \cdot \bar{p}) (1 + \varepsilon(n)) \quad (1 - 3 - 16)$$

et d'après (1 - 1 - 7), U_1 tend vers 0 uniformément.

Ainsi $T_1 = 1 + \varepsilon(n)$, ce qui termine la démonstration ■

§ I - 4 Réduction de l'intervalle de variation des n_i

L'essentiel de la somme qui définit $P(\bar{G}_{i_1} \dots \bar{G}_{i_r})$ se trouve rassemblé dans le pavé

$$\prod_i \{ m_i < n_i < M_i \}$$

Nous allons montrer cette propriété, et conditionner en conséquence les événements $\bar{G}_{i_1} \dots \bar{G}_{i_r}$ par les événements $E_1^* E_2^*$.

Remarquons que:

$$\bar{G}_{i_q}^* = \bar{G}_{i_q} \cap F_1^{(q)} \cap F_2^{(q)} \quad (1-4-1)$$

Et donc que:

$$(\bar{G}_{i_1} \dots \bar{G}_{i_r}) \cap (E_1^* E_2^*) = \bar{G}_{i_1}^* \dots \bar{G}_{i_r}^* \quad (1-4-2)$$

On peut alors écrire:

$$\begin{aligned} P(\bar{G}_{i_1} \dots \bar{G}_{i_r}) &= P(\bar{G}_{i_1}^* \dots \bar{G}_{i_r}^*) \\ &+ P(\bar{G}_{i_1}^* \dots \bar{G}_{i_r}^* | \overline{E_1^* E_2^*}) P(\overline{E_1^* E_2^*}) \quad (1-4-3) \end{aligned}$$

Donc que:

$$\begin{aligned} P(\bar{G}_{i_1} \dots \bar{G}_{i_r}) &\leq P(\bar{G}_{i_1}^* \dots \bar{G}_{i_r}^*) + P(\overline{E_1^*}) + P(\overline{E_2^*}) \quad (1-4-4) \\ P(\bar{G}_{i_1} \dots \bar{G}_{i_r}) &\leq P(\bar{G}_{i_1}^* \dots \bar{G}_{i_r}^*) + \sum_{q \in [1, r]} [P(\overline{F_1^{(q)}}) + P(\overline{F_2^{(q)}})] \\ &\quad (1-4-5) \end{aligned}$$

Par ailleurs

$$P(\bar{G}_{i_1} \dots \bar{G}_{i_r}) \geq P(\bar{G}_{i_1}^* \dots \bar{G}_{i_r}^*) \quad (1-4-6)$$

Nous allons majorer $P(\overline{F_1^{(q)}})$ et $P(\overline{F_2^{(q)}})$; il nous restera à évaluer $P(\bar{G}_{i_1}^* \dots \bar{G}_{i_r}^*)$, ce qui est possible d'après la proposition (1-2):

il suffit de prendre $C = A$ où C est défini en (1-3-1) et A en (1-2-5). On effectuera alors l'évaluation directe sur la loi de Poisson. Enfin, on choisira A de telle sorte que le majorant de

$$\sum_{q \in [1, r]} [P(\overline{F_1^{(q)}}) + P(\overline{F_2^{(q)}})]$$

soit infiniment petit devant $P(\bar{G}_{i_1}^* \dots \bar{G}_{i_r}^*)$, ce qui achèvera la

démonstration de la proposition (1-1).

PROPOSITION 1 - 3

Sous les hypothèses (1 - 1 - 7) : $Lk \cdot \bar{p} \rightarrow 0$ et (1 - 1 - 8) :

$\frac{Lk}{n\bar{p}} \rightarrow 0$, $P (\bar{F}_1^{(q)})$ et $P (\bar{F}_2^{(q)})$ sont majorés par la quantité:

$$T = \frac{1}{A} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{LK}} \left(\exp\left(-A^2 \frac{Lk}{2} (1 + \varepsilon(n)) \right) \right) (1 + \varepsilon(n)) \quad (1 - 4 - 7)$$

Remarquons que cette majoration est uniforme.

Démonstration de la proposition 1-3:

La variable n_q suit une loi binômiale $B(n, \pi_q)$. Appelons X une variable suivant cette loi.

Posons:

$$\bar{D} = n\pi_q \left(1 + A \left(\frac{Lk}{n\pi_q} \right)^{1/2} \right) + \iota_1, \quad \underline{D} = n\pi_q \left(1 - A \left(\frac{Lk}{n\pi_q} \right)^{1/2} \right) + \iota_2 \quad (1 - 4 - 8)$$

où la présence de ι_1 et de ι_2 est due au fait que n_q est nécessairement un entier: plus précisément, ι_1 et ι_2 vérifient:

$$P (\bar{F}_1^{(q)}) = P (X = \bar{D}) + P (X = \bar{D} + 1) + \dots \quad (1 - 4 - 9)$$

$$P (\underline{F}_2^{(q)}) = P (X = \underline{D}) + P (X = \underline{D} - 1) + \dots \quad (1 - 4 - 10)$$

Par définition de $F_1^{(q)}$ et $F_2^{(q)}$, les réels ι_1 et ι_2 sont majorés en valeur absolue par 2.

Nous allons utiliser une technique due à Feller [5].

Définissons:

$$u_j = \frac{P(X = j)}{P(X = j - 1)} \quad (1 - 4 - 11)$$

Comme:

$$u_j = \frac{n - j + 1}{j} \frac{\pi_q}{1 - \pi_q} \quad (1 - 4 - 12)$$

u_j est décroissant en j .

$u_j = 1$ est obtenu pour $j \sim n\pi_q + \pi_q$

donc si $j = \bar{D} = n\pi_q + A\sqrt{n\pi_q Lk} + \epsilon_1$,

$u_{\bar{D}} < 1$ car $n\pi_q \rightarrow \infty$

Par ailleurs:

$$\frac{P(X = \bar{D} + 1)}{P(X = \bar{D})} = u_{\bar{D} + 1} < u_{\bar{D}} \quad (1-4-13)$$

$$P(X = \bar{D} + 1) \leq u_{\bar{D}} P(X = \bar{D}) \quad (1-4-14)$$

$$P(X = \bar{D} + 2) \leq u_{\bar{D}} P(X = \bar{D} + 1) \quad (1-4-15)$$

et enfin:

$$P(\bar{F}_1^{(q)}) \leq \frac{1 - (u_{\bar{D}})^{\Omega}}{1 - u_{\bar{D}}} \cdot P(X = \bar{D}) \leq \frac{1}{1 - u_{\bar{D}}} \cdot P(X = \bar{D}) \quad (1-4-16)$$

puisque $u_{\bar{D}} < 1$.

Il nous faut maintenant travailler

$$P(X = \bar{D}) \text{ et } \frac{1}{1 - u_{\bar{D}}};$$

remarquons que $n - \bar{D} \rightarrow \infty$ et que $\bar{D} \rightarrow \infty$; par conséquent:

$$P(X = \bar{D}) = C_n^{\bar{D}} (\pi_q)^{\bar{D}} (1 - \pi_q)^{n - \bar{D}} \quad (1-4-17)$$

Appliquons la formule de Stirling: il vient:

$$P(X = \bar{D}) = \frac{1}{\sqrt{2\pi}} \left(\frac{n}{n - \bar{D}}\right)^{1/2} (\bar{D})^{-1/2} \exp(U_0) (1 + \epsilon(n)) \quad (1-4-18)$$

et

$$U_0 = n \ln n - (n - \bar{D}) \ln(n - \bar{D}) - \bar{D} \ln \bar{D} + \bar{D} \ln \pi_q + (n - \bar{D}) \ln(1 - \pi_q) \quad (1-4-19)$$

Nous allons transformer l'expression U_0 .

$$U_0 = n \left[\frac{\bar{D}}{n} L\left(\frac{n\pi_q}{\bar{D}}\right) - \left(1 - \frac{\bar{D}}{n}\right) L\left(1 - \frac{\bar{D}}{n}\right) + \left(1 - \frac{\bar{D}}{n}\right) L(1 - \pi_q) \right] \quad (1-4-20)$$

posons

$$\bar{D} = n\pi_q (1 + t_1) \quad (1-4-21)$$

Alors d'après (1-4-8), nous pouvons écrire:

$$t_1 = A \left(\frac{Lk}{n\pi_q}\right)^{1/2} (1 + \varepsilon(n)) \quad (1-4-21-1)$$

en effet: t_1 étant borné par 2 en valeur absolue,

$$\frac{t_1}{n\pi_q} \text{ est un } o\left(\left(\frac{Lk}{n\pi_q}\right)^{1/2}\right) \text{ car } n\pi_q Lk \rightarrow \infty \quad (1-4-21-2)$$

Par conséquent:

$$U_0 = n \left[-\pi_q(1 + t_1)L(1 + t_1) - (1 - \pi_q(1 + t_1))L(1 - \pi_q(1 + t_1)) \right. \\ \left. + (1 - \pi_q(1 + t_1))L(1 - \pi_q) \right]$$

$$U_0 = n \left[-\pi_q(1 + t_1)L(1 + t_1) \right. \\ \left. - (1 - \pi_q(1 + t_1))L((1 - \pi_q(1 + t_1))(1 - \pi_q)^{-1}) \right]$$

$$U_0 = -n \left[\pi_q(1 + t_1)L(1 + t_1) \right. \\ \left. + (1 - \pi_q)[1 - t_1\pi_q(1 - \pi_q)^{-1}]L[1 - t_1\pi_q(1 - \pi_q)^{-1}] \right] \quad (1-4-22)$$

D'après (1-1-8) et (1-4-21-1), $t_1 \rightarrow 0$.

Nous pouvons donc effectuer un développement limité de U_0 :

$$U_0 = -n \left[\pi_q \frac{t_1^2}{2} (1 + \varepsilon(n)) + \frac{t_1^2 \pi_q}{2} (1 - \pi_q)^{-1} (1 + \varepsilon(n)) \right] \quad (1-4-23)$$

Mais $\pi_q \rightarrow 0$ et par conséquent:

$$U_0 = -n\pi_q \frac{t_1^2}{2} (1 + \varepsilon(n))$$

$$U_0 = -A^2 \frac{Lk}{2} (1 + \varepsilon(n)) \quad (1-4-24)$$

Revenons à $u_{\bar{D}}$:

$$u_{\bar{D}} = \frac{n - \bar{D} + 1}{\bar{D}} \frac{\pi_q}{1 - \pi_q} \quad (1 - 4 - 25)$$

avec la notation en t introduite en (1 - 4 - 21) il vient

$$u_{\bar{D}} = \frac{1}{1 + t_1} [1 - t_1 \pi_q (1 - \pi_q)^{-1} + \frac{1}{n} (1 - \pi_q)^{-1}] \quad (1 - 4 - 26)$$

$$1 - u_{\bar{D}} = \frac{t_1}{1 + t_1} [1 + \pi_q (1 - \pi_q)^{-1} - \frac{1}{nt_1} (1 - \pi_q)^{-1}] \quad (1 - 4 - 27)$$

mais: $nt_1 = nA \left(\frac{Lk}{n\pi_q}\right)^{1/2} (1 + \epsilon(n)) = \left(\frac{nLk}{\pi_q}\right)^{1/2} (1 + \epsilon(n))A \rightarrow \infty$

donc

$$1 - u_{\bar{D}} = t_1 (1 + \epsilon(n)) \quad (1 - 4 - 28)$$

et

$$\frac{1}{1 - u_{\bar{D}}} = \frac{1}{t_1} (1 + \epsilon(n))$$

soit

$$\frac{1}{1 - u_{\bar{D}}} = \frac{1}{A} \left(\frac{n\pi_q}{Lk}\right)^{1/2} (1 + \epsilon(n)) \quad (1 - 4 - 29)$$

D'après (1 - 4 - 24) :

$$P(X = \bar{D}) = \frac{1}{\sqrt{2\pi}} \left(\frac{n}{n - \bar{D}}\right)^{1/2} (\bar{D})^{-1/2} \exp \left[-A^2 \frac{Lk}{2} (1 + \epsilon(n)) \right] \quad (1 - 4 - 30)$$

Comme \bar{D} est infiniment petit devant n ,

$$\frac{n}{n - \bar{D}} = 1 + \epsilon(n) \quad (\bar{p} \rightarrow 0) \quad (1 - 4 - 31)$$

enfin $t_1 \rightarrow 0$ d'où

$$(\bar{D})^{-1/2} = \frac{1}{\sqrt{n\pi_q}} (1 + \epsilon(n))$$

$$P(X = \bar{D}) = \frac{1}{\sqrt{2\pi}} \left(\frac{n}{n - \bar{D}} \right)^{1/2} (\bar{D})^{-1/2} \exp \left[-A^2 \frac{Lk}{2} (1 + \varepsilon(n)) \right] \quad (1 - 4 - 32)$$

et en rapprochant (1 - 4 - 32), (1 - 4 - 29) et (1 - 4 - 16) on obtient:

$$P(\bar{F}_1^{(q)}) \leq \frac{1}{A} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{Lk}} \left(\exp \left[-A^2 \frac{Lk}{2} (1 + \varepsilon(n)) \right] \right) (1 + \varepsilon(n)) \quad (1 - 4 - 33)$$

Traisons de même $P(\bar{F}_2^{(q)})$; cette fois $u_D < 1$

$$\frac{P(X = \underline{D})}{P(X = \underline{D} - 1)} = u_D \quad ; \quad \frac{P(X = \underline{D} - 1)}{P(X = \underline{D} - 2)} = u_D - 1 > \frac{u_D}{u_D - 1} \quad (1 - 4 - 34)$$

Soit:

$$P(\bar{F}_2^{(q)}) \leq P(X = \underline{D}) \frac{1 - (1/u_D)^{\Omega'}}{1 - (1/u_D)} \leq P(X = \underline{D}) \frac{u_D}{u_D - 1} \quad (1 - 4 - 36)$$

Car

$$P(X = \underline{D} - 2) \leq \left(\frac{1}{u_D} \right)^2 P(X = \underline{D}) \quad (1 - 4 - 35)$$

Pour passer de $P(X = \bar{D})$ à $P(X = \underline{D})$ il suffit de changer A en $-A$, en définissant t_2 par:

$$\underline{D} = n\pi_q (1 - t_2) \quad (1 - 4 - 39)$$

d'après (1 - 4 - 8), t_2 vérifie:

$$t_2 = A \frac{Lk}{n\pi_q}^{1/2} (1 + \varepsilon(n)) \quad (1 - 4 - 39 - 1)$$

il vient donc:

$$P(X = \underline{D}) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{n\pi_q}} \left(\exp \left[-A^2 \frac{Lk}{2} (1 + \varepsilon(n)) \right] \right) (1 + \varepsilon(n)) \quad (1 - 4 - 37)$$

on trouve ensuite:

$$u_D = \frac{1}{1 - t_2} \left[1 + t_2 \pi_q (1 - \pi_q)^{-1} + \frac{1}{n} (1 - \pi_q)^{-1} \right] \quad (1 - 4 - 38)$$

En continuant, il vient:

$$u_{\underline{D}} - 1 = \frac{t_2}{1 - t_2} \left[1 + \pi_q (1 - \pi_q)^{-1} + \frac{1}{nt_2} (1 - \pi_q)^{-1} \right] \quad (1 - 4 - 40)$$

et

$$\frac{u_{\underline{D}}}{u_{\underline{D}} - 1} = A^{-1} \frac{\left(\frac{n\pi_q}{Lk}\right)^{1/2}}{Lk} (1 + \epsilon(n)) \quad (1 - 4 - 41)$$

soit enfin:

$$P(\overline{F}_2^{(q)}) \leq \frac{1}{A} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{Lk}} \left(\exp \left[-A^2 \frac{Lk}{2} (1 + \epsilon(n)) \right] \right) (1 + \epsilon(n)) \quad (1 - 4 - 42)$$

La proposition 1 - 3 est ainsi démontrée.

§ I - 5 Calcul d'un équivalent de $P(\overline{G}_{i_1}^* \dots \overline{G}_{i_r}^*)$

Comme nous l'avons indiqué plus haut, nous pouvons appliquer pour $P(\overline{G}_{i_1}^* \dots \overline{G}_{i_r}^*)$ la proposition 1 - 2 avec $C = A$.

Ainsi:

$$P(\overline{G}_{i_1}^* \dots \overline{G}_{i_r}^*) = (1 + \epsilon(n)) U_2 \quad \text{où } U_2 \text{ est égal à :}$$

$$U_2 = \sum_{\overline{G}_{i_1}^*} \dots \sum_{\overline{G}_{i_r}^*} \left[\prod_{q \in [1, r]} \frac{(n\pi_q)^{n_q}}{n_q!} \exp(-n\pi_q) \right] \quad (1 - 5 - 1)$$

soit:

$$U_2 = \prod_{q \in [1, r]} \left[\sum_{\overline{G}_{i_q}^*} \frac{(n\pi_q)^{n_q}}{n_q!} \exp(-n\pi_q) \right] \quad (1 - 5 - 2)$$

mais

$$\sum_{\overline{G}_{i_q}^*} = \sum_{L_{i_q}^*} + \sum_{H_{i_q}^*} \quad (1 - 5 - 3)$$

Rappelons la formule de la fonction gamma incomplète (cf Renyi [11])

Si λ_0 est un réel positif fixé:

$$e^{-\lambda_0} \left[1 + \frac{\lambda_0}{1!} + \dots + \frac{\lambda_0^q}{q!} \right] = 1 - \frac{1}{q!} \int_0^{\lambda_0} u^q e^{-u} du$$

$$= \frac{1}{q!} \int_{\lambda_0}^{+\infty} u^q e^{-u} du \quad (1-5-4)$$

par ailleurs

$$e^{-\lambda_0} \left[\frac{\lambda_0^{q+1}}{(q+1)!} + \frac{\lambda_0^{q+2}}{(q+2)!} + \dots \right] = 1 - e^{-\lambda_0} \left[1 + \frac{\lambda_0}{1!} + \dots + \frac{\lambda_0^q}{q!} \right]$$

(1-5-5)

Alors notons

$$m_q^{\star} = n\pi_q \left(1 - A\left(\frac{Lk}{n\pi_q}\right)^{1/2} \right) + \iota_1^{\prime} \quad (1-5-6)$$

$$M_q^{\star} = n\pi_q \left(1 - A\left(\frac{Lk}{n\pi_q}\right)^{1/2} \right) + \iota_2^{\prime} \quad (1-5-7)$$

$$m_q^{\star} = n\pi_q - \sqrt{n\pi_q} \theta_1 + \iota_3^{\prime} \quad (1-5-7-1)$$

$$M_q^{\star} = n\pi_q + \sqrt{n\pi_q} \theta_2 + \iota_4^{\prime} \quad (1-5-7-2)$$

où les nouvelles variables m_q^{\star} , M_q^{\star} , m_q^{\star} et M_q^{\star} sont définies pour tenir compte du fait que la variable n_q est nécessairement un entier.

Ainsi les différents ι_j^{\prime} , j variant de 1 à 4, sont bornés en valeur absolue par 2.

Alors:

$$\Sigma_{L_i^{\star}} = I_1 - I_2 \quad ; \quad \Sigma_{H_i^{\star}} = I_3 - I_4 \quad (1-5-8)$$

où I_1 , I_2 , I_3 et I_4 sont définis par :

$$I_1 = \frac{1}{m_q^{\star!}} \int_{n\pi_q}^{+\infty} u^{m_q^{\star}} e^{-u} du \quad (1-5-9)$$

$$I_2 = \frac{1}{m_q^{\prime\star!}} \int_{n\pi_q}^{+\infty} u^{m_q^{\prime\star}} e^{-u} du \quad (1-5-10)$$

$$I_3 = \frac{1}{M_q^{\star!}} \int_0^{n\pi_q} u^{M_q^{\star}} e^{-u} du \quad (1-5-11)$$

$$I_4 = \frac{1}{M_q^{\prime\star!}} \int_0^{n\pi_q} u^{M_q^{\prime\star}} e^{-u} du \quad (1-5-12)$$

Nous allons utiliser la méthode de Laplace pour construire un équivalent asymptotique des intégrales du type des I_j , $j = 1, 2, 3, 4$;

PROPOSITION 1 - 4

Soient deux fonctions t et $r(t)$, $r(t) > t$;

On définit:

Cas 1 $\gamma(t)$ par $r(t) = t(1 + \gamma(t))$
 $\lambda(t)$ par $\lambda(t) = t[\gamma(t) - L(1 + \gamma(t))]$

et

$$J_1 = \frac{1}{\Gamma(t+1)} \int_r^{+\infty} u^t e^{-u} du \quad (1-5-14)$$

Cas 2 $\gamma(t)$ par $r(t) = t(1 - \gamma(t))$
 $\lambda(t)$ par $\lambda(t) = t[-\gamma(t) - L(1 - \gamma(t))]$

et

$$J_2 = \frac{1}{\Gamma(t+1)} \int_0^r u^t e^{-u} du \quad (1-5-19)$$

On suppose dans chaque cas que lorsque $t \rightarrow \infty$:

$$r(t) \rightarrow \infty \quad (1-5-15)$$

$$\gamma(t) \rightarrow 0^+ \quad (1-5-16)$$

$$\lambda(t) \rightarrow \infty \quad (1-5-17)$$

Alors dans chaque cas J_1 et J_2 vérifient:

$$J_1 \sim J_2 \sim \frac{1}{2\sqrt{\pi}} e^{-\lambda} \frac{1}{\sqrt{\lambda}} \quad \text{lorsque } t \rightarrow \infty \quad (1-5-20)$$

Démonstration de la proposition 1 - 4 :

Posons dans J_1 défini par (1 - 5 - 18) $u = t(1 + v)$, où v est une nouvelle variable.

Alors:

$$\begin{aligned} J_1 &= \frac{1}{\Gamma(t+1)} t^{t+1} \int_{\gamma}^{+\infty} e^{t[L(1+v) - 1 - v]} dv \\ &= \frac{t^{t+1}}{\Gamma(t+1)} J_1' \end{aligned} \quad (1 - 5 - 21)$$

où

$$J_1' = \int_{\gamma}^{+\infty} e^{t[L(1+v) - 1 - v]} dv \quad (1 - 5 - 22)$$

définissons: $h(x) = L(1+x) - 1 - x$ (1 - 5 - 23)
et $y(x)$, fonction positive de x par:

$$y(x) = \sqrt{-1 - h(x)} \quad (1 - 5 - 24)$$

$-1 - h(x)$ est une fonction croissante de x pour $x > 0$ qui s'annule pour $x = 0$; $y(x)$ est bien définie. c'est une fonction monotone croissante qui admet une fonction réciproque dérivable

$$x = \psi(y)$$

Alors dans J_1' prenons $y(v)$ comme nouvelle variable:

$$dv = \psi'(y) dy \quad ; \quad w = y(v)$$

et enfin comme:

$$\begin{aligned} J_1' &= \int_{\gamma}^{+\infty} e^{t[-1 - y^2(v)]} dv \\ J_1' &= \int_{y(\gamma)}^{+\infty} e^{t[-1 - w^2]} \psi'(w) dw \quad (1 - 5 - 25) \\ J_1' &= e^{-t} J_1'' \end{aligned}$$

$$\text{où } J_1'' = \int_{y(\gamma)}^{+\infty} e^{-tw^2} \psi'(w) dw \quad (1-5-26)$$

mais

$$\psi'(y) = \frac{1}{y'(x)} = -2 \frac{\sqrt{-1-h(x)}}{h'(x)} = 2y \frac{1+x}{x} \quad (1-5-27)$$

Au voisinage de 0, $\psi'(y)$ est continue.

Au voisinage de l'infini, $\psi'(y) \sim 2y$

Il existe donc B_1 et B_2 fixés dans \mathbb{R}^+ vérifiant:

$$\forall y > B_1, \psi'(y) < B_2 y \quad (1-5-28)$$

Changeons encore de variable dans J_1'' et posons:

$$z = tw^2 \quad (1-5-29)$$

$$\frac{dz}{dw} = 2\sqrt{z}t \quad (1-5-30)$$

$$J_1'' = \frac{1}{2\sqrt{t}} J_1^{(3)} \quad (1-5-31)$$

où

$$J_1^{(3)} = \int_{ty^2(\gamma)}^{+\infty} \psi'\left(\frac{z}{t}\right)^{1/2} e^{-z} \frac{dz}{\sqrt{z}}$$

$$J_1^{(3)} = \int_{\lambda(t)}^{+\infty} \psi'\left(\frac{z}{t}\right)^{1/2} e^{-z} \frac{dz}{\sqrt{z}} \quad (1-5-32)$$

Nous allons couper en trois $J_1^{(3)}$:

$$\begin{aligned} \int_{\lambda(t)}^{+\infty} &= \int_{\lambda(t)}^{2\lambda(t)} + \int_{2\lambda(t)}^{tB_1^2} + \int_{tB_1^2}^{+\infty} \\ &= \Omega_1 + \Omega_2 + \Omega_3 \end{aligned} \quad (1-5-32-1)$$

Comme $\frac{\lambda(t)}{t} \rightarrow 0$, Ω_1 s'estime aisément: $t \in [\lambda(t), 2\lambda(t)]$

et donc:

$$\psi'\left(\frac{z}{t}\right)^{1/2} \sim \sqrt{z} \quad (1-5-35)$$

sur ce même intervalle.

donc

$$\int_{\lambda(t)}^{2\lambda(t)} \psi' \left(\left(\frac{z}{t} \right)^{1/2} \right) e^{-z} \frac{dz}{\sqrt{z}} \sim \sqrt{2} \int_{\lambda(t)}^{2\lambda(t)} e^{-z} \frac{dz}{\sqrt{z}} \quad (1-5-36)$$

Comme

$$\int_{\lambda(t)}^{+\infty} e^{-z} \frac{dz}{\sqrt{z}} = e^{-\lambda} \frac{1}{\sqrt{\lambda}} (1 + \varepsilon(t)) \quad (1-5-37)$$

$$\int_{\lambda(t)}^{2\lambda(t)} e^{-z} \frac{dz}{\sqrt{z}} = e^{-\lambda} \frac{1}{\sqrt{\lambda}} (1 + \varepsilon(t)) - e^{-2\lambda} \frac{1}{\sqrt{2\lambda}} (1 + \varepsilon(t)) \quad (1-5-38)$$

$$\int_{\lambda(t)}^{2\lambda(t)} e^{-z} \frac{dz}{\sqrt{z}} = e^{-\lambda} \frac{1}{\sqrt{\lambda}} (1 + \varepsilon(t)) \quad (1-5-39)$$

donc

$$\Omega_1 = \int_{\lambda(t)}^{2\lambda(t)} \psi' \left(\left(\frac{z}{t} \right)^{1/2} \right) e^{-z} \frac{dz}{\sqrt{z}} \sim \sqrt{2} e^{-\lambda} \frac{1}{\sqrt{\lambda}} \quad (1-5-40)$$

Dans Ω_2 , comme $z \in [2\lambda(t), tB_1^2]$, $\left(\frac{z}{t} \right)^{1/2} \in \left[\left(\frac{2\lambda(t)}{t} \right)^{1/2}, B_1 \right]$, qui est inclus dans $[0, B_1]$.

Soit M le maximum de ψ' sur $[0, B_1]$:

$$\Omega_2 \leq \int_{2\lambda(t)}^{tB_1^2} M e^{-z} \frac{dz}{\sqrt{z}} \leq M \int_{2\lambda(t)}^{+\infty} e^{-z} \frac{dz}{\sqrt{z}} \sim M e^{-2\lambda} \frac{1}{\sqrt{2\lambda}} \quad (1-5-40-1)$$

$$\Omega_2 \text{ est majoré par } \Omega_1 \rho_1 \text{ où } \rho_1 \sim \frac{M}{2} e^{-\lambda} \rightarrow 0$$

donc $\Omega_2 = o(\Omega_1)$ (1-5-40-2)

Dans Ω_3 , comme $z \in [tB_1^2, +\infty[$, $\left(\frac{z}{t} \right)^{1/2} \in [B_1, +\infty[$, et

$$\psi' \left(\left(\frac{z}{t} \right)^{1/2} \right) \leq B_2 \left(\frac{z}{t} \right)^{1/2} \text{ d'après } (1-5-28)$$

$$\Omega_3 = \int_{tB_1^2}^{+\infty} \psi' \left(\left(\frac{z}{t} \right)^{1/2} \right) \frac{dz}{\sqrt{z}} \leq B_2 \frac{1}{\sqrt{t}} \int_{tB_1^2}^{+\infty} e^{-z} dz = B_2 \frac{1}{\sqrt{t}} e^{-tB_1^2}$$

(1-5-40-3)

Ω_3 est majoré par $\Omega_1 \rho_2$

$$\text{où } \rho_2 = B_2 \frac{1}{\sqrt{t}} e^{-tB_1^2} \frac{1}{\sqrt{2}} e^{\lambda} \sqrt{\lambda} = B_2 \left(\frac{\lambda}{2t}\right)^{1/2} e^{\lambda(t) - tB_1^2}$$

$$\rho_2 = B_2 \left(\frac{\lambda}{2t}\right)^{1/2} \exp \left[-tB_1^2 \left(1 - \frac{1}{B_1^2} \frac{\lambda(t)}{t}\right) \right] \quad (1-5-40-4)$$

comme $\frac{\lambda(t)}{t} \rightarrow 0$, $\rho_2 \rightarrow 0$ et $\Omega_3 = o(\Omega_1)$;

Appliquant la formule de Stirling de la fonction gamma, il vient:

$$\Gamma(t+1) \sim \frac{t}{e} t^t \sqrt{2\pi t} \quad (1-5-42)$$

on obtient:

$$J_1^{(3)} \sim \sqrt{2} e^{-\lambda} \frac{1}{\sqrt{\lambda}} \Rightarrow J_1'' \sim \frac{1}{\sqrt{2t}} e^{-\lambda} \frac{1}{\sqrt{\lambda}}$$

$$\Rightarrow J_1' \sim e^{-t} \frac{1}{\sqrt{2t}} e^{-\lambda} \frac{1}{\sqrt{\lambda}} \Rightarrow J_1 \sim \frac{1}{2\sqrt{\pi}} e^{-\lambda} \frac{1}{\sqrt{\lambda}} \quad (1-5-43)$$

Ce qui montre la première partie de notre proposition.

Passons à J_2 :

$$J_2 = \frac{1}{\Gamma(t+1)} \int_0^r u^t e^{-u} du$$

posons: $u = t(1-v)$; $du = -t dv$ (1-5-44)

$$J_2 = \frac{t^{t+1}}{\Gamma(t+1)} J_2' \quad \text{où} \quad J_2' = \int_{\gamma}^1 e^t [L(1-v) - 1 + v] dv \quad (1-5-45)$$

posons cette fois:

$$h(x) = L(1-x) - 1 + x = -1 - y^2(x) \quad (1-5-46)$$

comme en (1-5-24) $y(x)$ est bien défini par:

$$y(x) = \sqrt{-1 - h(x)}$$

y est une fonction croissante qui s'annule pour $x = 0$, elle admet une fonction réciproque

$$x = \psi(y)$$

Prenons $y(v)$ comme nouvelle variable: $w = y(v)$ dans J_2' ; il vient:
 $dv = \psi'(y) dy$

$$J_2' = \int_{y(\gamma)}^{+\infty} e^{t[-1-w^2]} \psi'(w) dw \quad (1-5-47)$$

la convergence en $+\infty$ de J_2' est assurée par les propriétés de régularité de ψ' (cf (1-5-28)). Nous avons ainsi:

$$J_2' = e^{-t} J_2'' \quad (1-5-48)$$

$$J_2'' = \int_{y(\gamma)}^{+\infty} e^{-tw^2} \psi'(w) dw \quad (1-5-49)$$

Enfin

$$\psi'(y) = \frac{1}{y'(x)} = 2y \frac{1-x}{x}$$

Au voisinage de $x = 0$, $y \sim \frac{x}{\sqrt{2}}$ donc $\psi'(y)$ existe en $y = 0$;

Au voisinage de $x = 1$ ($y \rightarrow \infty$)

$$\frac{\psi'(y)}{y} = 2 \frac{1-x}{x} \text{ est majoré par } B_2 \text{ pour } y > B_1$$

ce qui confirme la convergence de J_2'

$$\int_{B_1}^{+\infty} e^{-tw^2} \psi'(w) dw < B_2 \int_{B_1}^{+\infty} e^{-tw^2} w dw \text{ qui converge.}$$

On pose enfin $z = tw^2$: $dz = 2 dw \sqrt{zt}$ (1-5-50)
 et on obtient:

$$J_2'' = \frac{1}{2\sqrt{t}} J_2^{(3)} \quad \text{où} \quad J_2^{(3)} = \int_{ty^2(\gamma)}^{+\infty} \psi'\left(\left(\frac{z}{t}\right)^{1/2}\right) e^{-z} \frac{dz}{\sqrt{z}} \quad (1-5-51)$$

soit:

$$J_2^{(3)} = \int_{\lambda(t)}^{+\infty} \psi'\left(\left(\frac{z}{t}\right)^{1/2}\right) e^{-z} \frac{dz}{\sqrt{z}} \quad (1-5-52)$$

On applique les résultats du paragraphe précédent ((1-5-32) à (1-5-43)) pour obtenir

$$J_2^{(3)} \sim \sqrt{2} e^{-\lambda} \frac{1}{\sqrt{\lambda}} \quad \text{et enfin} \quad J_2 \sim \frac{1}{2\sqrt{\pi}} e^{-\lambda} \frac{1}{\sqrt{\lambda}}$$

Ce qui achève notre démonstration de la proposition 1 - 4 ■

Appliquons la proposition 1 - 4 à I_1, I_2, I_3, I_4 :

d'après les définitions des bornes m_q^* , m'_q , M_q^* , et M'_q , et compte-tenu du fait que les ι_j sont bornés, il existe des fonctions $\epsilon_j(n)$ tendant vers 0 avec n , vérifiant:

$$m_q^* = n\pi_q - \sqrt{n\pi_q} \theta_1^{(q)} \quad (1 - 5 - 52 - 1)$$

$$\text{où} \quad \theta_1^{(q)} = (2Lk - L_2k + y_1 + \epsilon_1^{(q)})^{1/2} \quad (1 - 5 - 52 - 2)$$

$$\text{avec heureusement} \quad |\epsilon_1^{(q)}| < \epsilon_1(n) \quad (1 - 5 - 52 - 3)$$

Pour le voir il suffit de remarquer que les encadrements de π_q s'effectuent de manière uniforme: en identifiant les deux expressions (1 - 5 - 7 - 1) et (1 - 5 - 52 - 1) on obtient (1 - 5 - 52 - 2).

de même, on aura:

$$m'_q = n_q \left(1 - A \left(\frac{Lk + \epsilon_2^{(q)}}{n\pi_q}\right)^{1/2}\right) \quad (1 - 5 - 52 - 4)$$

$$\text{où} \quad |\epsilon_2^{(q)}| < \epsilon_2(n) \quad (1 - 5 - 52 - 5)$$

Nous aurons bien entendu des expressions analogues pour M_q^* et M'_q . Le point important est que la manière dont les $\epsilon_j(n)$ tendent vers 0 ne dépend pas de q.

Avec ces nouvelles notations, nous pouvons obtenir pour I_1 :

$$t = n\pi_q - \sqrt{n\pi_q} \theta_1^{(q)} \quad (1 - 5 - 53 - 1)$$

$$r = n\pi_q \quad (1 - 5 - 53 - 2)$$

$$\gamma = \left(\theta_1^{(q)} / \sqrt{n\pi_q} \right) \left(1 - \theta_1^{(q)} / \sqrt{n\pi_q} \right)^{-1} \quad (1 - 5 - 53 - 3)$$

et enfin, par un développement limité rapide

$$\lambda = \frac{(\theta_1^{(q)})^2}{2} \left(1 + \frac{1}{3} \frac{\theta_1^{(q)}}{\sqrt{n\pi_q}} (1 + \varepsilon(n)) \right) \quad (1 - 5 - 53 - 4)$$

D'après la condition (1 - 1 - 8) : $\frac{Lk}{n_p} \rightarrow 0$

soit $t \rightarrow \infty$ et $r \rightarrow \infty$

enfin si $\frac{(Lk)^3}{n_p} \rightarrow 0$ (1 - 1 - 16) alors $\frac{(\theta_1^{(q)})^3}{\sqrt{n\pi_q}} \rightarrow 0$

et

$$\lambda = \frac{(\theta_1^{(q)})^2}{2} + \varepsilon(n) \quad (1 - 5 - 54)$$

La formule (1 - 5 - 52 - 2) permet alors d'écrire:

$$\lambda = Lk - \frac{1}{2} L_2 k + \frac{y_1}{2} + \varepsilon(n) \quad (1 - 5 - 56)$$

et

$$\sqrt{\lambda} = \sqrt{Lk} (1 + \varepsilon(n)) \quad (1 - 5 - 57)$$

Donc

$$e^{-\lambda} = \frac{1}{k} \sqrt{Lk} \left[e^{-(y_1/2)} \right] (1 + \varepsilon(n)) \quad (1 - 5 - 58)$$

et

$$I_1 \sim \frac{1}{2\sqrt{\pi}} \frac{1}{k} e^{-(y_1/2)} \quad (1 - 5 - 59)$$

Passons à I_2 : c'est une intégrale du cas 1 de la proposition 1 - 4; nous remarquons qu'il suffit de remplacer dans (1 - 5 - 53)

$$\theta_1^{(q)} \text{ par } A (Lk + \varepsilon_2^{(q)})^{1/2}$$

rappelons que $A > \sqrt{2}$. Alors le calcul de t et de r s'effectue de manière analogue, et si

$$\frac{(Lk)^3}{n_p} \rightarrow 0$$

$$\lambda = \frac{A^2}{2} Lk (1 + \varepsilon(n)) \quad (1 - 5 - 60)$$

par conséquent

il existe un réel ε_0 fixé tel que:

$$I_2 < \frac{1}{2\sqrt{\pi}} \frac{\sqrt{2}}{A} \frac{1}{\sqrt{Lk}} \exp. \left[-\frac{A^2}{2} Lk (1 - \varepsilon_0) \right] \quad (1-5-61)$$

et

$$I_2 / I_1 \rightarrow 0 \quad (1-5-63)$$

Pour I_3 on obtient: (Cas 2)

$$t = n\pi_q + \sqrt{n\pi_q} \theta_2^{(q)} \quad (1-5-64-1)$$

$$r = n\pi_q \quad (1-5-64-2)$$

$$\gamma = \left(\theta_2^{(q)} / \sqrt{n\pi_q} \right) \left(1 + \theta_2^{(q)} / \sqrt{n\pi_q} \right)^{-1} \quad (1-5-64-3)$$

Si (1-1-16) est vérifiée, $\frac{(Lk)^3}{np} \rightarrow 0$, alors:

$$\lambda = \frac{(\theta_2^{(q)})^2}{2} + \varepsilon(n)$$

$$I_3 \sim \frac{1}{2\sqrt{\pi}} \frac{1}{k} e^{-(y_1/2)} \quad (1-5-65)$$

On montrerait de même que $I_4 / I_3 \rightarrow 0$ (1-5-66)

Par conséquent

$$\Sigma_{L_{i_q}^*} \sim \frac{1}{2\sqrt{\pi}} \frac{1}{k} e^{-(y_1/2)} ; \quad \Sigma_{H_{i_q}^*} \sim \frac{1}{2\sqrt{\pi}} \frac{1}{k} e^{-(y_2/2)} \quad (1-5-67)$$

Et donc:

$$\Sigma_{\bar{G}_{i_q}^*} \sim \frac{1}{2\sqrt{\pi}} \frac{1}{k} \left[e^{-(y_1/2)} + e^{-(y_2/2)} \right] \quad (1-5-68)$$

L'équivalent étant uniforme, on en déduit:

$$P(\bar{G}_{i_1}^* \dots \bar{G}_{i_r}^*) \sim \frac{1}{k^r} \left[\frac{1}{2\sqrt{\pi}} \left(e^{-(y_1/2)} + e^{-(y_2/2)} \right) \right]^r$$

(1-5-69)

Etudions le majorant de l'expression Ω_1 :

$$\Omega_1 = \sum_{q \in [1, r]} [P(\bar{F}_1^{(q)}) + P(\bar{F}_2^{(q)})]$$

Ω_1 est majoré d'après la proposition 1-3 par:

$$\Omega_2 = \frac{2H-1}{A} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{Lk}} \exp \left[-A^2 \frac{Lk}{2} (1 + \varepsilon(n)) \right] (1 + \varepsilon(n)) \quad (1-5-70)$$

Il suffit donc de prendre $\frac{A^2}{2} > H$ pour avoir

$$\Omega_1 = o \left(P(\bar{G}_{i_1}^* \dots \bar{G}_{i_r}^*) \right) \quad (1-5-71)$$

On en déduit alors que

$$k^r P(\bar{G}_{i_1} \dots \bar{G}_{i_r}) \sim \left[\frac{1}{2\sqrt{\pi}} \left(e^{-(y_1/2)} + e^{-(y_2/2)} \right) \right]^r$$

(1-1-17)

et le théorème (1-1) est démontré ■

Les conditions obtenues pour ce théorème améliorent celles de Revesz [12].
A la page 37 de [12], le théorème 2 de Revesz s'énonce ainsi:

$$\text{si } n > k^{1+\varepsilon_2} \quad \text{où } \varepsilon_2 > 0 \quad \text{et } \frac{1}{n^\alpha} \leq \underline{p}, \quad \bar{p} \leq \frac{1}{n^\beta}$$

où $0 < \beta \leq 1 \leq \alpha < 1 + \varepsilon_2$, Alors

$$P(Z_2 < \theta_2) \rightarrow \exp \left[-\frac{1}{2\sqrt{\pi}} e^{-(y_2/2)} \right] \quad (1-5-72)$$

et en posant $Z'_2 = \max(-Z_1, Z_2)$

$$P(Z'_2 < \theta_1) \rightarrow \exp \left[-\frac{1}{\sqrt{\pi}} e^{-(y_1/2)} \right] \quad (1-5-73)$$

(1-5-73) résulte facilement de notre théorème (1-1).

Notre amélioration principale réside dans le passage de la condition

$$n > k^{1+\varepsilon_2} \quad \text{à la condition } \frac{(Lk)^3}{n\underline{p}} \rightarrow 0$$

§ I - 6 Généralisation du théorème (1 - 1)

Nous allons étendre le théorème (1 - 1) en étudiant les valeurs extrêmes partielles :

Considérons un découpage de l'ensemble $[1, k]$ en plusieurs familles; plus précisément posons

$$R = [1, k] = R_1 \cup R_2 \cup \dots \cup R_\ell \quad (1 - 6 - 1)$$

où ℓ est un entier fixé.

On suppose que chaque R_m vérifie:

$$\#R_m = \Lambda_m(n) k \quad (1 - 6 - 2)$$

où nous notons $\#R_m$ le cardinal de R_m ; chaque $\Lambda_m(n)$ vérifie

$$\Lambda_m(n) \rightarrow \Lambda_m^0 \text{ lorsque } n \rightarrow \infty \quad (1 - 6 - 3)$$

Définissons

$$Z_1^{(m)} = - \min_{i \in R_m} \left[\frac{\nu_n^i - np_i}{\sqrt{np_i}} \right] \quad (1 - 6 - 4)$$

$$Z_2^{(m)} = \max_{i \in R_m} \left[\frac{\nu_n^i - np_i}{\sqrt{np_i}} \right] \quad (1 - 6 - 5)$$

$$\theta_{im} = (2Lk - L_2k + y_{im})^{1/2} \quad i = 1, 2 \quad (1 - 6 - 6)$$

Alors nous pouvons énoncer le théorème (1 - 2):

Théorème (1 - 2)

$$\text{Si } \bar{p} Lk \rightarrow 0 \text{ et } \frac{(Lk)^3}{np} \rightarrow 0$$

$$P \left[\bigcap_m \{ Z_1^{(m)} < \theta_{1m} \}, \bigcap_m \{ Z_2^{(m)} < \theta_{2m} \} \right] \rightarrow \exp \left[- \frac{1}{2\sqrt{\pi}} \sum_{m=1}^{\ell} \Lambda_m^0 \left(e^{-(y_{1m}/2)} + e^{-(y_{2m}/2)} \right) \right] \quad (1 - 6 - 7)$$

Il y a donc indépendance asymptotique de bloc à bloc et de maximum à minimum. Ce résultat généralise celui de Revesz [12] et Woodroffe [17]. Cette indépendance est due au comportement poissonnien de la loi multinomiale.

Démonstration du théorème (1 - 2) :

Nous étendons les notations de L_i , H_i et \bar{G}_i en posant cette fois :

$$\begin{aligned} L_i &= \{ \nu_n^i < np_i - \sqrt{np_i} \theta_{1m} \} \\ H_i &= \{ \nu_n^i > np_i + \sqrt{np_i} \theta_{2m} \} \\ \bar{G}_i &= L_i + H_i \end{aligned} \quad \text{si } i \in R_m \quad (1 - 6 - 8)$$

Nous sommes cette fois encore amenés à considérer :

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

Supposons que dans $P(\bar{G}_{i_1} \dots \bar{G}_{i_r})$,

$$\begin{aligned} q_1 \text{ des événements } (\bar{G}_{i_\alpha}) \text{ proviennent de } R_1, \\ \dots \\ q_\ell \text{ des événements } (\bar{G}_{i_\alpha}) \text{ proviennent de } R_\ell, \end{aligned}$$

avec bien entendu : $q_1 + q_2 + \dots + q_\ell = r$ (1 - 6 - 9)

Alors nous savons que si $Lk \bar{p} \rightarrow 0$ et si $\frac{(Lk)^3}{np} \rightarrow 0$

$$\sum_{\bar{G}_{i_\alpha}^*} \sim \frac{1}{2\sqrt{\pi}} \frac{1}{k} (e^{-(y_{1m}/2)} + e^{-(y_{2m}/2)}) \text{ si } i_\alpha \in R_m$$

(cf (1 - 5 - 68))

Le conditionnement du § I - 4 est toujours applicable, ce qui permet de déduire que :

$$P(\bar{G}_{i_1} \dots \bar{G}_{i_r}) \sim \frac{1}{k^r} \left(\frac{1}{2\sqrt{\pi}} \right)^r \prod_{j=1}^r [e^{-(y_{1j}/2)} + e^{-(y_{2j}/2)}]^{q_j}$$

(1 - 6 - 10)

Nous pouvons construire $C_{\Lambda_1(n)k}^{q_1} C_{\Lambda_2(n)k}^{q_2} \dots C_{\Lambda_\ell(n)k}^{q_\ell}$ (1 - 6 - 11)

tels $\bar{G}_{i_1} \dots \bar{G}_{i_r}$, mais lorsque $n \rightarrow \infty$, q_j étant majoré par r , donc par H

$$C_{\Lambda_j(n)k}^{q_j} \frac{1}{k^{q_j}} \rightarrow \frac{(\Lambda_j^0)^{q_j}}{q_j!} \quad (1 - 6 - 12)$$

Comme $\prod_{j=1}^{\ell} \left(\frac{1}{k} \frac{1}{2\sqrt{\pi}} \right)^{q_j} = \frac{1}{k^r} \left(\frac{1}{2\sqrt{\pi}} \right)^r$ (1 - 6 - 13)

Nous en déduisons l'équivalent de la somme S_r :

$$S_r \sim \left(\frac{1}{2\sqrt{\pi}} \right)^r \frac{1}{r!} \sum_{q_1} \sum_{q_2} \dots \sum_{q_\ell} \left[\frac{r!}{q_1! q_2! \dots q_\ell!} \prod_j [(\Lambda_j^0)^{q_j} (e^{-(y_{1j}/2)} + e^{-(y_{2j}/2)})^{q_j}] \right] \quad (1 - 6 - 14)$$

où bien entendu $\sum q_j = r$.

$$S_r \sim \left(\frac{1}{2\sqrt{\pi}} \right)^r \frac{1}{r!} \left(\sum_j \Lambda_j^0 (e^{-(y_{1j}/2)} + e^{-(y_{2j}/2)}) \right)^r \quad (1 - 6 - 15)$$

Alors comme:

$$P \left[\prod_j \{ Z_1^{(m)} < \theta_{1m} \} , \prod_j \{ Z_2^{(m)} < \theta_{2m} \} \right] \\ = 1 - S_1 + S_2 - S_3 + \dots + (-1)^k S_k$$

On applique la technique des paragraphes précédents et le théorème (1 - 2) est démontré. ■

§ I - 7 Application à l'estimation de la densité

Nous nous proposons ici d'appliquer les résultats des théorèmes (1 - 1) et (1 - 2) à l'estimation de la densité par la méthode de l'histogramme.

Soit un pavage de \mathbb{R}^S formé par les pavés $\{ B_j , j \in J \}$; soit une variable X à valeurs dans \mathbb{R}^S dont la densité de probabilité est une fonction f que nous supposons uniformément continue sur \mathbb{R}^S .

On définit à partir d'un échantillon (X_1, X_2, \dots, X_n) les réels $\{ \nu_j, j \in J \}$:

$$\nu_j = \{ \text{nombre des } X_i \text{ contenus dans } B_j \} \quad (1-7-1)$$

L'estimateur f_n de f est défini par :

$$f_n(x) = \frac{1}{n} \frac{1}{\mu(B_j)} \nu_j \quad \text{si } x \in B_j$$

où μ représente la mesure de Lebesgue sur \mathbb{R}^S .

Geffroy [7] et Bertrand-Retali [2] ont donné des C.N.S. de convergence presque complète, en probabilité de f_n vers f uniformément sur \mathbb{R}^S .

Revesz s'est intéressé au problème des lois limites [12][13]. Nous allons généraliser grâce au théorème (1-1) deux théorèmes de Revesz.

Énonçons les deux théorèmes de Revesz :

(f est une densité de probabilité sur \mathbb{R} , strictement positive sur son support, [ce qui ne signifie pas forcément minorée par un réel strictement positif si le support n'est pas compact] et enfin Lipchitzienne sur ce support)

Théorème A (Revesz [13])

On définit un pavage de $\text{Supp}(f) = [A_0, B_0]$ par les segments

$$B_j = [x_j, x_{j+1}] \quad j = 0, 1, \dots, k-1$$

où x_j est défini par :

$$\int_{-\infty}^{x_j} f(t) dt = \frac{j}{k} \quad (x_0 = A_0, x_k = B_0) \quad (1-7-2)$$

k est une fonction qui vérifie :

$$n^{1/3} \ln n < k < n^{1-\varepsilon_1} \quad (1-7-3)$$

où ε_1 est un réel positif donné.

Alors il existe un intervalle $]x_\alpha, x_\beta[$ croissant vers $]A_0, B_0[$ lorsque n tend vers l'infini; on suppose de plus que f est minorée sur $]x_\alpha, x_\beta[$:

θ_i étant toujours défini par: $\theta_i = (2Lk - L_2k + y_i)^{1/2}$,
sur l'intervalle $]x_\alpha, x_\beta[$,

$$\lim_{n \rightarrow \infty} P\left(\left(\frac{n}{k}\right)^{1/2} Q_1 < \theta_1\right) = \exp\left(-\frac{1}{2\sqrt{\pi}} e^{-(y_1/2)}\right) \quad (1-7-4)$$

$$\lim_{n \rightarrow \infty} P\left(\left(\frac{n}{k}\right)^{1/2} Q'_1 < \theta_2\right) = \exp\left(-\frac{1}{\sqrt{\pi}} e^{-(y_2/2)}\right) \quad (1-7-5)$$

où

$$Q_1 = \sup_{]x_\alpha, x_\beta[} \frac{f_n - f}{f} \quad (1-7-6)$$

$$Q'_1 = \sup_{]x_\alpha, x_\beta[} \left| \frac{f_n - f}{f} \right| \quad (1-7-7)$$

Théorème B (Revesz [12])

On découpe cette fois \mathbb{R} en intervalles de longueur $1/k$:

$$\dots, x_{-2}, x_{-1}, x_0, x_1, x_2, \dots$$

ξ_1 et ξ_2 sont deux constantes qui vérifient:

$$\frac{1}{2} < \xi_1 < \xi_2 < \frac{2}{3} \quad (1-7-8)$$

$$n^{1-\xi_2} \geq k \geq n^{1-\xi_1} \quad (1-7-9)$$

$[a, b]$ étant le support de f , éventuellement égal à $]-\infty, +\infty[$, nous supposons que x_α et x_β sont des termes de la suite (x_j) définie ci-dessus, qui vérifient

$$]x_\alpha, x_\beta[= [a, b] \cap \left[-n^{-(\xi_1/2)}, n^{(\xi_1/2)}\right] \quad (1-7-10)$$

vérifiant également $f(x) \geq \frac{1}{Ln}$ sur $]x_\alpha, x_\beta[$. (1-7-11)

α et β sont ainsi deux fonctions de n .

On demande enfin que x_α et x_β vérifient

$$\int_{x_\alpha}^{x_\beta} f(t) dt \rightarrow 1 \quad \text{lorsque } n \rightarrow \infty \quad (1-7-12)$$

Posons $Z_0 = \max_{[x_\alpha, x_\beta]} \left(\frac{f_n - f}{\sqrt{f}} \right)$ et $Z'_0 = \max_{[x_\alpha, x_\beta]} \left| \frac{f_n - f}{\sqrt{f}} \right|$

Alors:

$$\lim_{n \rightarrow \infty} P[Z_0 < (2L(\beta - \alpha) - L_2(\beta - \alpha) + y)^{1/2}] = \exp \left[- \frac{1}{2\sqrt{\pi}} e^{-(y/2)} \right]$$

$$\lim_{n \rightarrow \infty} P[Z'_0 < (2L(\beta - \alpha) - L_2(\beta - \alpha) + y)^{1/2}] = \exp \left[- \frac{1}{\sqrt{\pi}} e^{-(y/2)} \right]$$

(1 - 7 - 13) et (1 - 7 - 14)

Nous allons proposer deux théorèmes sur l'estimation de la densité par histogramme dans \mathbb{R}^S , associés à deux pavages différents de \mathbb{R}^S . L'un des pavages dépendant de f , le théorème correspondant sera utilisé pour construire des tests.

f sera supposée désormais strictement positive sur son support et lipchitzienne.

Nous munissons \mathbb{R}^S de la norme $\| \cdot \|_\infty$

$$\|x\|_\infty = \sup_{i \in [1, s]} |x_i|$$

Dans notre premier théorème, nous estimerons f sur un compact que nous pouvons toujours supposer ramené à $[0, 1]^S$ formé des pavés $B_{nr_1 \dots r_s}$

$$B_{nr_1 \dots r_s} = \left[\frac{r_1 - 1}{d_n}, \frac{r_1}{d_n} \right] \times \dots \times \left[\frac{r_s - 1}{d_n}, \frac{r_s}{d_n} \right] \quad (1 - 7 - 15)$$

$$v_{nr_1 \dots r_s} = \{ \text{le nombre des } X_i \text{ contenus dans } B_{nr_1 \dots r_s} \} \quad (1 - 7 - 16)$$

On pose $k = d_n^S$ (1 - 7 - 17)

Nous supposons que $[0, 1]^S$ est inclus dans le support de f .

Remarquons au passage que ceci n'entraîne pas: $P_f([0, 1]^S) = 1$.

Alors

Théorème 1 - 3 - 1

$$\text{Si } \frac{k(Lk)^3}{n} \rightarrow 0 \text{ et } \frac{nLk}{k^{1 + (2/s)}} \rightarrow 0 \quad (1 - 7 - 18)$$

Notant

$$Z_1 = \left(\frac{n}{k} \right)^{1/2} \min_{[0, 1]^S} \left(\frac{f_n - f}{\sqrt{f_n}} \right) \quad \text{et} \quad Z_2 = \left(\frac{n}{k} \right)^{1/2} \max_{[0, 1]^S} \left(\frac{f_n - f}{\sqrt{f_n}} \right)$$

On a dès lors:

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

On trouvera en Annexe A la démonstration de la proposition suivante:

Proposition 1- 5

Soient (Z_1, Z_2) , (Z_1^i, Z_2^i) quatre statistiques dépendant de n .
On suppose que:

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

Alors si:

$$\sqrt{Lk} (Z_i - Z_i^i) \xrightarrow{p} 0 \quad \text{lorsque } n \rightarrow \infty \quad (i = 1 , 2)$$

$$\lim_{n \rightarrow \infty} P(Z_1^i < \theta_1 , Z_2^i < \theta_2) = \exp \left[- \frac{1}{2\sqrt{\pi}} \left(e^{-(y_1/2)} + e^{-(y_2/2)} \right) \right]$$

Démonstration du théorème 1-3-1 :

Nous pouvons écrire:

$$\frac{f_n - f}{\sqrt{f_n}} = \frac{f_n - \bar{f}_n}{(\bar{f}_n)^{1/2}} + \frac{f_n - f}{\sqrt{f_n}} + \left(\frac{f_n - \bar{f}_n}{\sqrt{f_n}} - \frac{f_n - \bar{f}_n}{(\bar{f}_n)^{1/2}} \right) \quad (1 - 7 - 18 - 1)$$

Notons $p_{nr_1 \dots r_s} = P_f (B_{nr_1 \dots r_s}) = \int_{B_{nr_1 \dots r_s}} f(t) dt \quad (1 - 7 - 19)$

et $\bar{f}_n = \frac{1}{\mu(B_{nr_1 \dots r_s})} \int_{B_{nr_1 \dots r_s}} f(t) dt \quad (1 - 7 - 20)$

Alors $\bar{f}_n = \frac{1}{\mu(B_{nr_1 \dots r_s})} p_{nr_1 \dots r_s} = k p_{nr_1 \dots r_s} \quad (1 - 7 - 21)$

f est minorée sur $[0,1]^S$ par δ et majorée par Δ .

Ainsi notant:

$$\bar{p} = \sup(p_{nr_1 \dots r_s}) \quad \text{et} \quad \underline{p} = \inf(p_{nr_1 \dots r_s})$$

$$\bar{p} Lk < \frac{\Delta Lk}{k} \quad \text{qui tend vers } 0 \quad \text{et} \quad \frac{k(Lk)^3}{n} \rightarrow 0 \quad \text{et donc} \quad \frac{(Lk)^3}{np} \rightarrow 0$$

On peut appliquer le théorème 1 - 1 aux quantités

$$\frac{v_{nr_1 \dots r_s} - np_{nr_1 \dots r_s}}{\sqrt{np_{nr_1 \dots r_s}}} = \binom{n}{k}^{1/2} \frac{f_n - \bar{f}_n}{(\bar{f}_n)^{1/2}} \quad (1-7-22)$$

Proposition 1 - 6

Si A est un réel fixé supérieur à $\sqrt{2}$,

$$v_{nr_1 \dots r_s} < p \cdot n p_{nr_1 \dots r_s} \left(1 + A \left(\frac{Lk}{np_{nr_1 \dots r_s}}\right)^{1/2}\right)$$

$$v_{nr_1 \dots r_s} > p \cdot n p_{nr_1 \dots r_s} \left(1 - A \left(\frac{Lk}{np_{nr_1 \dots r_s}}\right)^{1/2}\right)$$

uniformément en $r_1 \dots r_s$

Démonstration de la proposition 1 - 6 :

Il suffit de reprendre les calculs du théorème 1-1 en remplaçant θ_1 et θ_2 par $A\sqrt{Lk}$

Si $\frac{k(Lk)^3}{n} \rightarrow 0$, on obtient dans le calcul de I_1, I_2, I_3, I_4

$$\lambda = \frac{A^2}{2} Lk + o(1) \quad (1-7-23)$$

D'où

$$P(\bar{G}_{i_1} \dots \bar{G}_{i_r}) \leq \left(\frac{C'}{\alpha_1^k}\right) \text{ où } C' \text{ est fixé et } \alpha_1 > 1 \quad (1-7-24)$$

$$P(\bar{G}_i) \leq \frac{C'}{\alpha_1^k} \text{ et } P\left(\bigcap_i G_i\right) \geq 1 - \frac{C'}{\alpha_1^k - 1} \quad (1-7-25)$$

Donc

$$P\left(\bigcap_i G_i\right) \rightarrow 1$$

Ce qui équivaut à la proposition 1 - 6 ■

Nous déduisons de la proposition 1 - 6 que:

$$f_n < p \bar{f}_n \left(1 + A \left(\frac{Lk}{np_{nr_1 \dots r_s}}\right)^{1/2}\right) \quad (1-7-26)$$

$$f_n <_p \bar{f}_n \left(1 + A \left(\frac{kLk}{n}\right)^{1/2} (\bar{f}_n)^{-1/2}\right) \quad (1-7-27)$$

et de même

$$f_n >_p \bar{f}_n \left(1 - A \left(\frac{kLk}{n}\right)^{1/2} (\bar{f}_n)^{-1/2}\right) \quad (1-7-28)$$

D'après (1-7-22) et la proposition 1-5, il suffit pour montrer le théorème 1-3-1 de montrer:

$$\sup_{[0,1]^s} \left(\frac{nLk}{k}\right)^{1/2} \left| \frac{\bar{f}_n - f}{\sqrt{\bar{f}_n}} \right| \xrightarrow{p} 0 \quad (1-7-29)$$

$$\sup_{[0,1]^s} \left(\frac{nLk}{k}\right)^{1/2} \left| \frac{f_n - \bar{f}_n}{\sqrt{\bar{f}_n}} - \frac{f_n - \bar{f}_n}{(\bar{f}_n)^{1/2}} \right| \xrightarrow{p} 0 \quad (1-7-30)$$

Montrons (1-7-29) :

$$f(x) - \bar{f}_n(x) = k \int_{B_{nr_1 \dots r_s}} (f(x) - f_n(t)) dt \quad (1-7-31)$$

donc

$$|f(x) - \bar{f}_n(x)| < k \int_{B_{nr_1 \dots r_s}} T_2 \|t - x\|_\infty dt < \frac{T_2}{d_n} \quad (1-7-32)$$

où T_2 est fixé car f est supposée Lipchitzienne. Comme

$$f_n >_p \bar{f}_n \left(1 - A \left(\frac{kLk}{n}\right)^{1/2} (\bar{f}_n)^{-1/2}\right) \quad (1-7-33)$$

$$f_n >_p \delta \left(1 - A \left(\frac{kLk}{n}\right)^{1/2} \frac{1}{\sqrt{\delta}}\right) = \delta (1 - \epsilon(n)) \quad (1-7-34)$$

On majore ainsi uniformément

$$\left(\frac{nLk}{k}\right)^{1/2} \left| \frac{\bar{f}_n - f}{\sqrt{\bar{f}_n}} \right| < \left(\frac{nLk}{k}\right)^{1/2} \frac{T_2}{d_n} \frac{1}{\sqrt{\delta}} \quad (1-7-35)$$

Comme $d_n = k^{1/s}$ on a bien (1-7-29) grâce à (1-7-18)

Passons à (1-7-30) :

$$\left| \frac{f_n - \bar{f}_n}{\sqrt{\bar{f}_n}} - \frac{f_n - \bar{f}_n}{(\bar{f}_n)^{1/2}} \right| = \frac{(f_n - \bar{f}_n)^2}{(f_n \bar{f}_n)^{1/2}} \frac{1}{\sqrt{\bar{f}_n} + (\bar{f}_n)^{1/2}} \quad (1-7-36)$$

$$\frac{1}{\sqrt{f_n}} < \frac{1}{(\bar{f}_n)^{1/2}} (1 + \varepsilon(n)) \quad \text{Car} \quad \frac{kLk}{n} \frac{1}{\delta} \rightarrow 0$$

$$(f_n - \bar{f}_n)^2 < \frac{A^2 kLk}{n} \bar{f}_n \quad (1 - 7 - 37)$$

d'où

$$\left(\frac{nLk}{k}\right)^{1/2} \left| \frac{f_n - \bar{f}_n}{\sqrt{f_n}} - \frac{f_n - \bar{f}_n}{(\bar{f}_n)^{1/2}} \right| < \left(\frac{k(Lk)^3}{n}\right)^{1/2} A^2 \frac{1}{2\sqrt{\delta}} \quad (1 - 7 - 38)$$

majorant qui tend bien uniformément vers 0 si on a (1 - 7 - 18) ■

Notre deuxième théorème utilise un pavage de \mathbb{R}^s associé canoniquement à f . Nous supposons cette fois l'existence des densités conditionnelles de f . Posons toujours:

$$d_n = k^{1/s}$$

Soient (e_1, e_2, \dots, e_s) les s directions orthogonales de \mathbb{R}^s . Découpons \mathbb{R}^s par des hyperplans perpendiculaires à e_1 , en d_n tranches de probabilités égales (au sens de f):

$$(Q^{(1)}, Q^{(2)}, \dots, Q^{(d_n)}) \quad (1 - 7 - 39)$$

Recommençons l'opération dans chaque tranche $(Q^{(i)})$ en découplant perpendiculairement à e_2 :

$$Q^{(i)} = \bigcup_j Q^{(i,j)} \quad j \in [1, d_n]$$

$$\text{avec } P_f(Q^{(i,j)}) = \frac{1}{d_n^2} \quad (1 - 7 - 39)$$

De proche en proche on aboutit ainsi à un pavage de \mathbb{R}^s en blocs

$$\{ Q^{(i_1 \dots i_s)} \mid i_1 \in [1, d_n], \dots, i_s \in [1, d_n] \} \quad (1 - 7 - 40)$$

avec par construction:

$$P_f(Q^{(i_1 \dots i_s)}) = \frac{1}{d_n^s} = \frac{1}{k} \quad (1 - 7 - 41)$$

Nous allons être amenés à faire une hypothèse supplémentaire de régularité sur f au voisinage de l'infini.

Notons: $f_j(x_j | x_1 \dots x_{j-1}) \quad j \in [2, s]$

la densité conditionnelle de x_j par rapport à $x_1 \dots x_{j-1}$.

Nous supposons qu'il existe une fonction $\delta(n)$, qui tend vers 0 lorsque n tend vers l'infini, et un pavé $\vartheta(n)$

$$\vartheta(n) = \vartheta_1 \times \vartheta_2 \times \dots \times \vartheta_s \quad \text{où} \quad \vartheta_s = [-v_j, v_j]$$

tel que $\forall (x_1 \dots x_s) \in \vartheta(n) \quad (\text{i.e. } x_j \in \vartheta_j \text{ pour tout } j \text{ de } 1 \text{ à } s)$

$$1) \quad f_j(x_j | x_1 \dots x_{j-1}) > \delta(n) \quad (1 - 7 - 42)$$

$$2) \quad P_f(\vartheta(n)) \rightarrow 1 \text{ lorsque } n \rightarrow \infty \quad (1 - 7 - 43)$$

Remarquons que ces conditions sont toujours vérifiées pour $s = 1$.

Notons $\{v_j^{(i_1 \dots i_s)} \mid j \in [1, s]\}$ les longueurs des cotés du pavé $Q^{(i_1 \dots i_s)}$

Alors si f vérifie (1 - 7 - 42), si $Q^{(i_1 \dots i_s)}$ est inclus dans $\vartheta(n)$,

$$v_j^{(i_1 \dots i_s)} < \frac{1}{d_n} \frac{1}{\delta(n)} \quad (1 - 7 - 44)$$

En effet si nous notons la densité projetée de f sur \mathbb{R}^p : $f^{(p)}(x_1 \dots x_p)$

$$\int_{Q^{(i_1 \dots i_p)}} f^{(p)}(x_1 \dots x_p) dx_1 \dots dx_p = \left(\frac{1}{d_n}\right)^p \quad (1 - 7 - 45)$$

Mais

$$\int_{Q^{(i_1 \dots i_p)}} = \int_{Q^{(i_1 \dots i_{p-1})}} \Upsilon(x_1 \dots x_{p-1}) f^{(p-1)}(x_1 \dots x_{p-1}) dx_1 \dots dx_{p-1}$$

où

$$\Upsilon(x_1 \dots x_{p-1}) = \int_{x_p}^{x_p^{(i_1 \dots i_p)} + v_p^{(i_1 \dots i_p)}} f_p(x_p | x_1 \dots x_{p-1}) dx_p$$

et par conséquent:

$$\int_{Q^{(i_1 \dots i_p)}} > v_p^{(i_1 \dots i_p)} \delta(n) \left(\frac{1}{d_n}\right)^{p-1} \text{ d'où le résultat.}$$

Ainsi si $Q^{(i_1 \dots i_p)}$ est inclus dans $\vartheta(n)$, et si f vérifie (1-7-42)

$$\forall (x,y) \in Q^{(i_1 \dots i_s)}, \|x - y\|_\infty < \frac{1}{d_n} \frac{1}{\delta(n)} \quad (1-7-46)$$

Soit Θ_n l'ensemble des $(i_1 \dots i_s)$ tels que:

$$Q^{(i_1 \dots i_s)} \text{ soit inclus dans } \vartheta(n).$$

De même, soit $\bar{\Theta}_n$ l'ensemble des $(i_1 \dots i_s)$ tels que

$$Q^{(i_1 \dots i_s)} \cap \vartheta(n) \neq \emptyset$$

$$\text{Par construction: } (\#\bar{\Theta}_n) = (\#\Theta_n) (1 + \varepsilon(n)) \quad (1-7-47)$$

Mais:

$$\bigcup_{\Theta_n} Q^{(i_1 \dots i_s)} \subset \vartheta(n) \subset \bigcup_{\bar{\Theta}_n} Q^{(i_1 \dots i_s)} \quad (1-7-48)$$

Comme $P_f(\vartheta(n)) \rightarrow 1$ si (1-7-43) est vérifié, on a

$$\#\bar{\Theta}_n = k (1 - \varepsilon(n)) \quad \text{et donc} \quad \#\Theta_n = k (1 - \varepsilon(n)) \quad (1-7-49)$$

$$\text{soit enfin: } P_f\left(\bigcup_{\Theta_n} Q^{(i_1 \dots i_s)}\right) \rightarrow 1 \quad (1-7-50)$$

Nous noterons alors:

$$Q_n^* = \bigcup_{\Theta_n} Q^{(i_1 \dots i_s)} \quad (1-7-51)$$

Notre théorème 1-3-2 s'énonce alors:

Théorème 1 - 3 - 2

Soit f une densité strictement positive sur son support (mais non forcément minorée sur celui-ci) et Lipchitzienne. On suppose de plus que f vérifie (1-7-42) et (1-7-43) pour une certaine fonction $\delta(n)$. Supposons encore que k et n vérifient:

$$\frac{k(Lk)^3}{n} \rightarrow 0 \quad \text{et} \quad \frac{nLk}{k^{1+(2/s)} \delta(n)^{2+2s}} \rightarrow 0 \quad (1-7-53)$$

Notons cette fois:

$$Z_1 = -\left(\frac{n}{k}\right)^{1/2} \min_{Q_n^*} \left[\frac{f_n - f}{f_n} \right] ; \quad Z_2 = \left(\frac{n}{k}\right)^{1/2} \max_{Q_n^*} \left[\frac{f_n - f}{f_n} \right] \quad (1-7-52)$$

Alors:

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

Notons comme dans (1-7-20)

$$\bar{f}_n(x) = \frac{1}{\mu(Q^{(i_1 \dots i_s)})} \int_{Q^{(i_1 \dots i_s)}} f(t) dt = \frac{1}{k \mu(Q^{(i_1 \dots i_s)})} \quad (1-7-54)$$

si $x \in Q^{(i_1 \dots i_s)}$.

Nous avons en notant $\nu_n^{(i_1 \dots i_s)}$ le nombre de points de l'échantillon

(X_1, X_2, \dots, X_n) contenus dans $Q^{(i_1 \dots i_s)}$:

$$\frac{f_n - \bar{f}_n}{\bar{f}_n} = \frac{1}{n} \frac{\nu_n^{(i_1 \dots i_s)} - n P_f(Q^{(i_1 \dots i_s)})}{P_f(Q^{(i_1 \dots i_s)})} \quad (1-7-55)$$

Comme par construction:

$$P_f(Q^{(i_1 \dots i_s)}) = \frac{1}{k}$$

$$\left(\frac{n}{k}\right)^{1/2} \frac{f_n - \bar{f}_n}{\bar{f}_n} = \frac{\nu_n^{(i_1 \dots i_s)} - n P_f(Q^{(i_1 \dots i_s)})}{(n P_f(Q^{(i_1 \dots i_s)}))^{1/2}} \quad (1-7-56)$$

Nous pouvons donc appliquer à (1-7-56) le théorème 1-1 si $\bar{p} Lk \rightarrow 0$

et $\frac{(Lk)^3}{np} \rightarrow 0$, ce qui est impliqué par la condition $\frac{k(Lk)^3}{n} \rightarrow 0$ car $\bar{p} = \frac{1}{k}$;

Nous allons de nouveau écrire:

$$\frac{f_n - f}{f_n} = \frac{f_n - \bar{f}_n}{\bar{f}_n} + \frac{\bar{f}_n - f}{f_n} + \left(\frac{f_n - \bar{f}_n}{f_n} - \frac{f_n - \bar{f}_n}{\bar{f}_n} \right) \quad (1-7-57)$$

Nous chercherons encore des conditions suffisantes pour avoir

$$\left(\frac{nLk}{k}\right)^{1/2} \left| \frac{f_n - \bar{f}}{f_n} \right| \xrightarrow{p} 0 \quad \left(\text{Uniformément sur les } Q^{(i_1 \dots i_s)} \right)$$

$$\left(\frac{nLk}{k}\right)^{1/2} \frac{(f_n - \bar{f}_n)^2}{f_n \bar{f}_n} \xrightarrow{p} 0 \quad (1-7-58) \text{ et } (1-7-59)$$

La proposition 1-5 fournit alors le résultat. D'après la proposition 1-6 nous avons (uniformément sur Θ_n)

$$\left| \nu_n^{(i_1 \dots i_s)} - n P_f(Q^{(i_1 \dots i_s)}) \right| < A \left(n P_f(Q^{(i_1 \dots i_s)}) Lk \right)^{1/2} \quad (1-7-60)$$

soit, uniformément sur Θ_n :

$$\left| f_n - \bar{f}_n \right| < A \left(\frac{Lk}{kn}\right)^{1/2} \frac{1}{\mu(Q^{(i_1 \dots i_s)})} = A \left(\frac{kLk}{n}\right)^{1/2} \bar{f}_n \quad (1-7-61)$$

Ainsi:

$$f_n > \bar{f}_n \left(1 - A \left(\frac{kLk}{n}\right)^{1/2}\right)$$

et (1-7-59) devient:

$$\left(\frac{nLk}{k}\right)^{1/2} \frac{(f_n - \bar{f}_n)^2}{f_n \bar{f}_n} < A^2 \left(\frac{k(Lk)^3}{n}\right)^{1/2} (1 + \epsilon(n)) \quad (1-7-62)$$

et l'expression de droite tend vers 0 d'après (1-7-53).

Il nous reste à étudier (1-7-58)

$$f \text{ étant Lipchitzienne, si } x \text{ et } y \text{ se trouvent dans } Q^{(i_1 \dots i_s)}$$

$$\left| f(x) - f(y) \right| < T_2 \|x - y\|_\infty \quad (1-7-63)$$

mais (1-7-46) entraîne alors:

$$\left| f(x) - \bar{f}_n(x) \right| < T_2 \frac{1}{d_n} \frac{1}{\delta(n)} \quad (1-7-64)$$

donc:

$$\left| f(x) - \bar{f}_n(x) \right| < \frac{1}{\mu(Q^{(i_1 \dots i_s)})} \int_{Q^{(i_1 \dots i_s)}} T_2 \frac{1}{d_n} \frac{1}{\delta(n)} dt = T_2 \frac{1}{d_n} \frac{1}{\delta(n)} \quad (1-7-65)$$

Par construction, si $x \in Q_n^*$, $f(x) > \delta^s(n)$ car

$$f(x_1 \dots x_s) = f_s(x_s | x_1 \dots x_{s-1}) \dots f_2(x_2 | x_1) f_1(x_1)$$

et par hypothèse: $f_j(x_j | x_1 \dots x_{j-1}) > \delta(n)$
donc:

$$\bar{f}_n(x) > \delta^s(n) \quad \text{et} \quad f_n > \delta^s(n) \left(1 - A \left(\frac{kLk}{n}\right)^{1/2}\right) \quad (1-7-66)$$

Ainsi majore-t-on (1-7-58) :

$$\left(\frac{nLk}{k}\right)^{1/2} \left| \frac{\bar{f}_n - f}{f_n} \right| < T_2 \left(\frac{nLk}{k}\right)^{1/2} \frac{1}{k^{1/s}} \frac{1}{(\delta(n))^{1+s}} (1 + \varepsilon(n)) \quad (1-7-67)$$

Le terme de droite tend bien vers 0 si

$$\frac{nLk}{k^{1+(2/s)}} \frac{1}{\delta(n)^{2+2s}} \rightarrow 0 \quad (1-7-68)$$

Remarque 1 :

Supposons que nous limitons l'estimation de f à un compact K_0 vérifiant $P_f(K_0) = \Lambda_0$ (1-7-69)

Nous pouvons toujours considérer l'ensemble Θ'_n des $(i_1 \dots i_s)$ tels que $Q^{(i_1 \dots i_s)}$ soit inclus dans K_0 et définir Q'_n inclus dans K_0 par:

$$Q'_n = \bigcup_{\Theta'_n} Q^{(i_1 \dots i_s)} \quad (1-7-70)$$

puisque nous travaillons sur un compact, la condition (1-7-42) est vérifiée pour δ fixé:

$$f_j(x_j | x_1 \dots x_{j-1}) > \delta$$

par ailleurs, $P_f(Q'_n) \rightarrow \Lambda_0$, et donc $\#\Theta'_n = \Lambda_0 (1 + \varepsilon(n)) k$ (1-7-71)

On a donc en appliquant le théorème 1-2 le résultat suivant :

Corollaire 1 - 3 - 2

Soit f une densité strictement positive (et Lipchitzienne, rappelons-le sur un compact K_0 de \mathbb{R}^s . Si nous posons

$$Z_1 = -\left(\frac{n}{k}\right)^{1/2} \min_{Q'_n} \left(\frac{f_n - f}{f_n} \right) , \quad Z_2 = \left(\frac{n}{k}\right)^{1/2} \max_{Q'_n} \left(\frac{f_n - f}{f_n} \right)$$

et si k et n vérifient

$$\frac{k(Lk)^3}{n} \rightarrow 0 \quad \text{et} \quad \frac{nLk}{n^{1+(2/s)}} \rightarrow 0$$

Alors

$$P(Z_1 < \theta_1 , Z_2 < \theta_2) \rightarrow \exp \left[- \frac{P_f(K_0)}{2\sqrt{\pi}} \left(e^{-(y_1/2)} + e^{-(y_2/2)} \right) \right]$$

On étendrait sans peine le résultat ci-dessus à une famille de compacts disjoints de \mathbb{R}^s : (K_0, \dots, K_ℓ) où ℓ est fixé, en utilisant le théorème 1-2 (cf chapitre 3).

Remarque 2

Dans le cas $s = 1$, notre théorème 1-3-2 est semblable au théorème A de Revesz, mais sous des conditions suffisantes plus faibles: on passe de

$$k < n^{1 - \epsilon_1} \quad \tilde{a} \quad \frac{k(Lk)^3}{n} \rightarrow 0$$

Nous donnons la loi produit des statistiques Z_1 et Z_2 et nous remplaçons

$$\frac{f - f_n}{f} \quad \text{par} \quad \frac{f - f_n}{f_n} , \quad \text{quantité bien plus utilisable par le statisticien.}$$

Enfin nous proposons un théorème dans le cas de \mathbb{R}^s .

Cependant, f étant inconnue a-priori, son découpage en blocs $Q^{(i_1 \dots i_s)}$ n'est pas connu non plus. Nous pensons cependant utiliser ce théorème dans des travaux ultérieurs pour établir d'autres lois limites plus maniables concrètement.

Pour ce qui est du théorème 1 - 3 - 1 , nous avons préféré nous limiter à un compact et fournir un théorème utilisable (en effet cette fois le pas est connu, ainsi que l'ensemble d'estimation) par le statisticien.

On remarquera que la quantité $P_f([0,1]^S)$ n'intervient pas dans la formulation de la loi limite. là encore, nous passons de:

$$k \leq n^{1 - \xi_2}, \quad \xi_2 > 1/2 \quad \text{à} \quad \frac{k(Lk)^3}{n} \rightarrow 0$$

en remplaçant:

$$\frac{f_n - f}{\sqrt{f}} \quad \text{par} \quad \frac{f_n - f}{\sqrt{f_n}}, \quad \text{utilisable concrètement.}$$

CHAPITRE SECOND

Dans ce chapitre, nous nous proposons d'étudier une loi limite sur les valeurs extrêmes de statistiques d'écart associées à l'échantillon ordonné sur $[0,1]$: th 2-1 ; nous emploierons une technique semblable à celle du chapitre 1 en montrant l'approximation de la loi de Dirichlet par une loi de Poisson. Le théorème 2-1 généralise un lemme de Tusnady [15] en donnant la loi conjointe des min et des max.

Nous prolongerons ensuite le théorème 2-1 aux valeurs extrêmes partielles: ce théorème fait apparaître une indépendance asymptotique intéressante entre les différentes statistiques : th 2-2 ; Nous assouplirons enfin les conditions sur le nombre de recouvrements élémentaires dans chaque bloc.

§ II - 1 Notations et exposé du problème

On considère un échantillon de taille n de la loi uniforme sur $[0,1]$, et l'échantillon ordonné associé

$$(X_1^*, X_2^*, \dots, X_n^*) \quad X_1^* \leq X_2^* \leq \dots \leq X_n^*$$

On appelle recouvrement élémentaire l'intervalle $[X_j^*, X_{j+1}^*]$, où j varie de 1 à n .

On partitionne alors l'intervalle $[X_1^*, X_n^*]$ en k blocs $[Y_i^*, Y_{i+1}^*]$, où i varie de 1 à k .

Chaque Y_i^* est un $X_{\beta(i)}^*$, $i \in [1, k]$, avec $\beta(1) = 1$ et $\beta(k+1) = n$.

Nous notons toujours $\lfloor x \rfloor$ la fonction partie entière de x :

$$x - 1 < \lfloor x \rfloor \leq x$$

Il est toujours possible de s'arranger pour que les k blocs $[Y_i^*, Y_{i+1}^*]$ soient formés chacun de

$$m_i = \lfloor \frac{n}{k} \rfloor + \eta_i \quad \eta_i = -1, 0 \text{ ou } 1 \quad (2-1-1)$$

recouvrements élémentaires. En effet comme

$$\frac{n}{k} - 1 < \lfloor \frac{n}{k} \rfloor \leq \frac{n}{k} \quad , \quad \frac{n}{k} - 2 < \lfloor \frac{n}{k} \rfloor - 1 \leq \frac{n}{k} - 1$$

$$n - k < k \lfloor \frac{n}{k} \rfloor \leq n \quad \text{et} \quad k(\lfloor \frac{n}{k} \rfloor - 1) \leq n - k$$

Soit

$$n < k(\lfloor \frac{n}{k} \rfloor + 1)$$

Nous pouvons trouver des entiers m_1, m_2, \dots, m_k vérifiant:

$$\sum_{i=1}^k m_i = n - 1 \quad (2-1-2)$$

Nous allons nous intéresser aux variables

$$\sigma_i = Y_{i+1}^* - Y_i^* \quad \text{où } i \in [1, k] \quad (2-1-3)$$

et aux statistiques

$$Z_1 = - \min_i \left(\frac{\sigma_i - 1/k}{1/\sqrt{nk}} \right) \quad \text{et} \quad Z_2 = \max_i \left(\frac{\sigma_i - 1/k}{1/\sqrt{nk}} \right)$$

$$(2-1-4) \quad \text{et} \quad (2-1-5)$$

Nous posons toujours

$$\theta_j = (2L_1k - L_2k + y_j)^{1/2} \quad j = 1, 2 \quad (2-1-6)$$

et nous supposons que

$$k \rightarrow \infty, \quad n \rightarrow \infty, \quad \frac{kLk}{n} \rightarrow 0 \quad (2-1-7)$$

Enonçons alors le théorème 2-1:

Théorème 2-1

$$\text{Si } \frac{k(Lk)^3}{n} \rightarrow 0 \quad (2-1-8)$$

$$P(Z_1 < \theta_1, Z_2 < \theta_2) \rightarrow \exp \left[-\frac{1}{2\sqrt{\pi}} \left(e^{-(y_1/2)} + e^{-(y_2/2)} \right) \right] \quad (2-1-9)$$

Pour démontrer ce théorème, nous allons utiliser une méthode analogue à celle du chapitre 1 pour le théorème 1-1.

D'après Wilks [16], p 238, la distribution des variables $(\sigma_1, \dots, \sigma_k)$ est la loi de Dirichlet:

$$D(m_1, m_2, \dots, m_k, n - m_1 - \dots - m_k + 1)$$

Soit la loi $D(m_1, \dots, m_k, 2)$ dont la densité est d'après (2-1-2)

$$f_0(t_1, \dots, t_k) = \frac{\Gamma(m_1 + \dots + m_k + 2)}{\Gamma(m_1) \dots \Gamma(m_k) \Gamma(2)} t_1^{m_1-1} \dots t_k^{m_k-1} (1 - t_1 - \dots - t_k)$$

Soit d'après (2-1-2)

$$f_0(t_1, \dots, t_k) = \frac{\Gamma(n+1)}{\Gamma(m_1) \dots \Gamma(m_k)} t_1^{m_1-1} \dots t_k^{m_k-1} (1 - t_1 - \dots - t_k)$$

L'espérance de σ_i est $\frac{m_i}{n+1}$, équivalente à $\frac{1}{k}$ (2-1-10)

La variance de σ_i est $\frac{m_i (n+1 - m_i)}{(n+1)^2 (n+2)}$, équivalente à $\frac{1}{nk}$ (2-1-11)

Nous considérons donc les variables réduites:

$$\frac{\sigma_i - 1/k}{1/\sqrt{nk}} \quad (2-1-13)$$

Définissons les événements $(i \in [1, k])$

$$L_i = \left\{ \sigma_i < \frac{1}{k} - \frac{1}{\sqrt{nk}} \theta_1 \right\} \quad (2-1-14)$$

$$H_i = \left\{ \sigma_i > \frac{1}{k} + \frac{1}{\sqrt{nk}} \theta_2 \right\} \quad (2-1-15)$$

$$\bar{G}_i = L_i + H_i \quad (2-1-16)$$

L'événement $\{ Z_1 < \theta_1, Z_2 < \theta_2 \}$ s'écrit $\left\{ \bigcap_{i=1}^k \bar{G}_i \right\}$ $(2-1-17)$

Nous appliquons la formule de Poincaré:

$$P\left(\bigcup_i \bar{G}_i \right) = S_1 - S_2 + \dots + (-1)^{k-1} S_k \quad (2-1-18)$$

où

$$S_r = \sum_{1 \leq i_1 < i_2 < \dots < i_r \leq k} P(\bar{G}_{i_1} \dots \bar{G}_{i_r}) \quad (2-1-19)$$

Nous allons montrer que si $\frac{k(Lk)^3}{n} \rightarrow 0$ et $r < H$ fixé:

Proposition 2 - 1

$$k^r P(\bar{G}_{i_1} \dots \bar{G}_{i_r}) = \left[\frac{1}{2\sqrt{\pi}} \left(e^{-(y_1/2)} + e^{-(y_2/2)} \right) \right]^r (1 + \epsilon(n))$$

uniformément en $(i_1 \dots i_r)$.

On achève alors la démonstration du théorème 2-1 comme celle du théorème 1-1: $(1-1-18)$ à $(1-1-20)$.

§ II - 2 Etude de $P(\bar{G}_{i_1} \dots \bar{G}_{i_r})$. Autres notations

Nous poserons: $x_1 = \sigma_{i_1}, x_2 = \sigma_{i_2}, \dots, x_r = \sigma_{i_r}$ $(2-2-1)$

$$M_r = m_{i_1} + m_{i_2} + \dots + m_{i_r} \quad (2-2-2)$$

$$T_r = x_1 + x_2 + \dots + x_r \quad (2-2-3)$$

$$a_1 = m_{i_1}, \dots, a_r = m_{i_r} \quad (2-2-4)$$

Ainsi $M_r = a_1 + \dots + a_r \quad (2-2-5)$

Ces premières notations étant définies, nous pouvons écrire:

$$P(\bar{G}_{i_1} \dots \bar{G}_{i_r}) = \int_{\bar{G}_{i_1}} \int_{\bar{G}_{i_2}} \dots \int_{\bar{G}_{i_r}} f_1(x_1, \dots, x_r) dx_1 dx_2 \dots dx_r \quad (2-2-6)$$

où toujours d'après Wilks [16] p 238 th 8-7-7 :

$$f_1(x_1, \dots, x_r) = \frac{\Gamma(n+1)}{\Gamma(a_1) \dots \Gamma(a_r) \Gamma(n+1-a_r)} x_1^{a_1-1} \dots x_r^{a_r-1} (1-T_r)^{n-M_r} \quad (2-2-7)$$

Définissons encore les événements:

$$F_1^{(q)} = \left\{ x_q < \frac{1}{k} \left(1 + \left(\frac{kLk}{n} \right)^{1/2} A \right) \right\} \quad (2-2-8)$$

$$F_2^{(q)} = \left\{ x_q > \frac{1}{k} \left(1 - \left(\frac{kLk}{n} \right)^{1/2} A \right) \right\} \quad (2-2-9)$$

$$E_1^* = \bigcap_i F_1^{(i)} \quad (2-2-10)$$

où $i \in [1, r]$

$$E_2^* = \bigcap_i F_2^{(i)} \quad (2-2-11)$$

A est toujours une constante supérieure à $\sqrt{2}$.

Nous définissons également:

$$L_{i_q}^* = \left\{ \frac{1}{k} \left(1 - A \left(\frac{kLk}{n} \right)^{1/2} \right) < x_q < \frac{1}{k} - \frac{1}{\sqrt{nk}} \theta_1 \right\} \quad (2-2-12)$$

Soit

$$L_{i_q}^* = L_{i_q} \cap F_2^{(q)}$$

de même soit:

$$H_{iq}^* = \left\{ \frac{1}{k} + \frac{1}{\sqrt{nk}} \theta_2 < x_q < \frac{1}{k} \left(1 + A \left(\frac{kLk}{n} \right)^{1/2} \right) \right\} \quad (2-2-13)$$

Soit

$$H_{iq}^* = H_{iq} \cap F_1^{(q)}$$

Enfin on notera:

$$\bar{G}_{iq}^* = L_{iq}^* + H_{iq}^* \quad (2-2-14)$$

Nous allons rapprocher la loi de Dirichlet de la loi de Poisson:

§ II - 3 Loi de Dirichlet et loi de Poisson

Supposons que dans (2-2-7) chaque x_i puisse s'écrire:

$$x_i = \frac{1}{k} \left(1 + \rho_i \left(\frac{kLk}{n} \right)^{1/2} \right) \quad (2-3-1)$$

où $|\rho_i| \leq C$ où C est un réel fixé.

Nous avons alors la proposition 2 - 2 :

Proposition 2 - 2

Si $\frac{kLk}{n} \rightarrow 0$ et si (2-3-1) est réalisé, la loi de densité f_1

définie en (2-2-7) vérifie:

$$f_1(x_1, \dots, x_r) = \left[\frac{x_1^{a_1-1}}{\Gamma(a_1)} \frac{x_2^{a_2-1}}{\Gamma(a_2)} \dots \frac{x_r^{a_r-1}}{\Gamma(a_r)} n^{M_r} \exp(-nT_r) \right] (1 + \varepsilon(n))$$

(2-3-2)

$\varepsilon(n)$ étant toujours uniforme.

On reconnaîtra mieux une loi de Poisson en mettant le terme de droite de

(2-3-2) sous la forme

$$\prod_{i=1}^r \frac{(nx_i)^{a_i-1}}{(a_i-1)!} n e^{-nx_i} \quad (2-3-3)$$

en effet:

$$\left[\sum_i (a_i - 1) \right] + r = \sum_i a_i = M_r$$

Démontrons maintenant la proposition 2-2; il suffit de montrer que:

$$\Omega_0 = \frac{n!(1 - T_r)^{n-M_r}}{(n - M_r)! n^{M_r}} \exp(nT_r)$$

est de la forme $1 + \varepsilon(n)$ (2 - 3 - 4)

Remarquons, que contrairement au théorème 1-1, les T_r varient, M_r étant fixé. Par analogie avec la proposition 1-2, posons encore:

$$\psi_n = 1 - \left(1 - \frac{M_r}{n}\right)(1 - T_r)^{-1} \quad (2 - 3 - 5)$$

D'après (2 - 1 - 1)

$$m_i = \left\lfloor \frac{n}{k} \right\rfloor + \eta_i \Rightarrow a_j = \frac{n}{k} + \eta'_j$$

où $|\eta_i|$ et $|\eta'_j|$ sont majorés par 2 .

Donc

$$M_r = a_1 + \dots + a_r = \frac{rn}{k} + \Lambda_r \quad \text{où} \quad |\Lambda_r| < 2r < 2H \quad (2 - 3 - 6)$$

Mais, d'après (2 - 3 - 1)

$$x_i = \frac{1}{k} + \rho_i \left(\frac{Lk}{kn}\right)^{1/2}$$

d'où

$$T_r = \frac{r}{k} + \left(\frac{Lk}{kn}\right)^{1/2} R' \quad \text{avec} \quad |R'| < CH \quad (2 - 3 - 7)$$

Donc $n - M_r$ tend vers l'infini d'après (2 - 3 - 6) et on peut appliquer la formule de Stirling à (2 - 3 - 4), obtenant dès lors:

$$\Omega_0 = \exp(U_3) (1 + \varepsilon(n))$$

où

$$U_3 = -n(1 - T_r) \left[\left(1 - \frac{M_r}{n}\right)(1 - T_r)^{-1} L \left(\left(1 - \frac{M_r}{n}\right)(1 - T_r)^{-1} \right) + 1 - \left(1 - \frac{M_r}{n}\right)(1 - T_r)^{-1} \right] \quad (2 - 3 - 8)$$

Soit d'après (2 - 3 - 5)

$$U_3 = -n(1 - T_r) \left[(1 - \psi_n)L(1 - \psi_n) + \psi_n \right]$$

Explicitons ψ_n :

$$\psi_n = 1 - \left(1 - \frac{r}{k} - \frac{\Lambda_r}{n}\right) \left(1 - \frac{r}{k} - \left(\frac{Lk}{kn}\right)^{1/2} R'\right)^{-1} \quad (2-3-9)$$

Ce qui peut s'écrire:

$$\psi_n = \left(-\left(\frac{Lk}{kn}\right)^{1/2} R' + \frac{\Lambda_r}{n}\right) \left(1 - \frac{r}{k} - \left(\frac{Lk}{kn}\right)^{1/2} R'\right)^{-1} \quad (2-3-10)$$

Comme $\frac{kLk}{n} \rightarrow 0$, $\psi_n \rightarrow 0$;

Effectuons un développement limité de U_3 :

$$U_3 = -n(1 - T_r) \left(\frac{\psi_n^2}{2} (1 + \varepsilon(n))\right) \quad (2-3-11)$$

Comme

$$\psi_n = -\left(\frac{Lk}{kn}\right)^{1/2} R' (1 + \varepsilon(n)) \quad (2-3-12)$$

et que R' est borné par CH , U_3 est équivalent à $-\frac{R'^2 Lk}{2k}$ et tend vers 0, ce qu'il fallait montrer ■

§ II - 4 Réduction de l'intervalle de variation des x_i

Comme pour le théorème 1-1, nous allons conditionner les événements $\bar{G}_{i_1} \dots \bar{G}_{i_r}$ par $E_1^* E_2^*$ et montrer que

$$P(\bar{G}_{i_1} \dots \bar{G}_{i_r}) = (1 - \varepsilon(n)) P(\bar{G}_{i_1}^* \dots \bar{G}_{i_r}^*) \quad (2-4-1)$$

Il nous faut majorer, comme au § I-4:

$$\sum_{q \in [1, r]} [P(\bar{F}_1^{(q)}) + P(\bar{F}_2^{(q)})]$$

Etudions les événements $\bar{F}_1^{(i)}$ et $\bar{F}_2^{(i)}$ et montrons la proposition 2-3:

Proposition 2-3

Si $\frac{kLk}{n} \rightarrow 0$, les inégalités suivantes sont vérifiées:

$P(\bar{F}_1^{(i)})$ et $P(\bar{F}_2^{(i)})$ sont majorés uniformément par:

$$\frac{1}{A} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{Lk}} \exp \left[-\frac{A^2}{2} Lk (1 + \varepsilon(n)) \right] (1 + \varepsilon(n)) \quad (2-4-2)$$

On remarquera que la quantité qui majore $P(\bar{F}_1^{(i)})$ et $P(\bar{F}_2^{(i)})$ est la même que celle qui intervient dans la proposition 1-3, ce qui nous donne à penser qu'une loi plus générale est vraisemblable.

Démontrons maintenant la proposition 2-3:

Traitons d'abord l'événement $\bar{F}_1^{(i)}$:

$$F_1^{(i)} = \left\{ x_i < \frac{1}{k} \left(1 + A \left(\frac{kLk}{n} \right)^{1/2} \right) \right\}$$

D'après Wilks [16] p 180 th 7-7-4, x_i suit la loi Bêta $B(a_i, n+1-a_i)$, dont la densité est:

$$f_2(x) = \frac{n!}{(a_i - 1)!(n - a_i)!} x^{a_i-1} (1-x)^{n-a_i} \quad (2-4-3)$$

Ainsi

$$P(F_1^{(i)}) = \int_0^{1 - \omega_1} f_2(x) dx \quad (2-4-4)$$

où

$$\omega_1 = 1 - \frac{1}{k} \left(1 + \left(\frac{kLk}{n} \right)^{1/2} A \right) \quad (2-4-5)$$

de même

$$P(\bar{F}_1^{(i)}) = \int_{1 - \omega_1}^1 f_2(x) dx = 1 - \int_0^{1 - \omega_1} f_2(x) dx \quad (2-4-6)$$

Utilisons la formule sommatoire de la loi Bêta: Kendall [8], Vol 1 p 124

$$\begin{aligned} & \frac{n!}{(a_i - 1)!(n - a_i)!} \int_0^{1 - \omega_2} x^{n-a_i} (1-x)^{a_i-1} dx \\ &= \sum_{j=0}^{a_i-1} C_n^j \omega_2^j (1 - \omega_2)^{n-j} \quad (2-4-7) \end{aligned}$$

Comme

$$\int_{1-\omega_1}^1 f_2(x) dx = \int_0^{\omega_1} f_2(1-x) dx \quad (2-4-8)$$

$$P(\bar{F}_1^{(i)}) = \frac{n!}{(a_i - 1)!(n - a_i)!} \int_0^{\omega_1} x^{n-a_i} (1-x)^{a_i-1} dx \quad (2-4-9)$$

Posons

$$\omega_2 = 1 - \omega_1 = \frac{1}{k} \left(1 + \left(\frac{kLk}{n}\right)^{1/2} A\right) \quad (2-4-10)$$

Alors

$$P(\bar{F}_1^{(i)}) = \sum_{j=0}^{a_i-1} C_n^j \omega_2^j (1-\omega_2)^{n-j} \quad (2-4-11)$$

On reconnaît la loi binomiale $B(n, \omega_2)$. Appelons X une v.a.r. suivant cette loi; nous pouvons écrire:

$$P(\bar{F}_1^{(i)}) = P(X=0) + \dots + P(X=a_i-1) \quad (2-4-12)$$

$$u_j = \frac{P(X=j)}{P(X=j-1)} \quad (2-4-13)$$

$$u_j = \frac{n-j+1}{j} \frac{\omega_2}{1-\omega_2} \quad \text{décroit quand } j \text{ croît.}$$

Or

$$u_j = 1 \text{ pour } j_0 \text{ voisin de } (n+1)\omega_2$$

Donc

$$j_0 \sim \frac{n+1}{k} + \left(\frac{(n+1)^2 Lk}{nk}\right)^{1/2}$$

Or

$$a_i - 1 = \frac{n}{k} + \eta_i' - 1 \quad \text{avec } |\eta_i'| < 2$$

Donc $j_0 - a_i$ tend vers $+\infty$ et nous sommes dans la partie des u_j supérieurs à 1.

Comme au chapitre 1 nous obtenons:

$$P(\bar{F}_1^{(i)}) \leq P(X = a_i - 1) \frac{1}{1 - (1/u_{a_i-1})} \quad (2-4-13)$$

D'où

$$P(\overline{F}_1^{(i)}) \leq P(X = a_i - 1) \frac{u_{a_i-1}}{u_{a_i-1} - 1} \quad (2-4-14)$$

$$P(\overline{F}_1^{(i)}) \leq P(X = a_i) \frac{u_{a_i}}{u_{a_i} - 1} \quad (2-4-15)$$

Car

$$u_{a_i} > 1 \Rightarrow P(X = a_i) > P(X = a_i - 1) \quad (2-4-16)$$

Effectuons un développement limité de u_{a_i} :

$$u_{a_i} = \frac{n+1-a_i}{a_i} \frac{\omega_2}{1-\omega_2}$$

$$u_{a_i} = \left(1 - \frac{1}{k} + \frac{1-\eta_i^!}{n}\right) \left(1 - \frac{1}{k} \left(1 + \left(\frac{kLk}{n}\right)^{1/2} A\right)\right)^{-1}$$

$$\left(1 + \left(\frac{kLk}{n}\right)^{1/2} A\right) \left(1 + \eta_i^! \frac{k}{n}\right)^{-1} \quad (2-4-17)$$

Posons

$$\eta_i'' = 1 - \eta_i^! \quad , \quad \text{donc} \quad |\eta_i''| \leq 3$$

$$u_{a_i} = \left(1 - \frac{1}{k} - \frac{1}{k} A \left(\frac{kLk}{n}\right)^{1/2} + \frac{1}{k} A \left(\frac{kLk}{n}\right)^{1/2} + \eta_i'' \frac{1}{n}\right)$$

$$\left(1 - \frac{1}{k} - \frac{1}{k} A \left(\frac{kLk}{n}\right)^{1/2}\right)^{-1} \left(1 + \left(\frac{kLk}{n}\right)^{1/2} A\right) \left(1 + \eta_i^! \frac{k}{n}\right)^{-1}$$

$$(2-4-18)$$

$$u_{a_i} = \left(1 + \left(\frac{1}{k} \left(\frac{kLk}{n}\right)^{1/2} A + \frac{\eta_i''}{n}\right) \left(1 - \frac{1}{k} - \frac{1}{k} \left(\frac{kLk}{n}\right)^{1/2} A\right)^{-1}\right)$$

$$\left(1 + \left(\frac{kLk}{n}\right)^{1/2} A\right) \left(1 + \eta_i^! \frac{k}{n}\right)^{-1} \quad (2-4-19)$$

Donc

$$u_{a_i} = \left(1 + \left(\frac{Lk}{kn}\right)^{1/2} (1 + \varepsilon(n))\right) \left(1 + \left(\frac{kLk}{n}\right)^{1/2} A\right) \left(1 + \eta_i^! \frac{k}{n}\right)^{-1}$$

$$(2-4-20)$$

Car $\frac{\eta_i''}{n} = \eta_i'' \left(\frac{k}{nLk}\right)^{1/2} \left(\frac{Lk}{kn}\right)^{1/2}$ et $\left(\frac{k}{nLk}\right)^{1/2} \rightarrow 0$

Nous sommes amenés à comparer les infiniments petits: $\left(\frac{Lk}{kn}\right)^{1/2}$, $\left(\frac{kLk}{n}\right)^{1/2}$ et $\frac{k}{n}$;

Comme $\left(\frac{kLk}{n}\right)^{1/2} = \left(\frac{Lk}{kn}\right)^{1/2} k$, $\left(\frac{Lk}{kn}\right) = o\left(\left(\frac{kLk}{n}\right)^{1/2}\right)$ (2-4-21)

et $u_{a_i} = \left(1 + A \left(\frac{kLk}{n}\right)^{1/2} (1 + \varepsilon(n))\right) \left(1 + \eta_i' \frac{k}{n}\right)^{-1}$ (2-4-22)

Mais $\left(\frac{kLk}{n}\right)^{1/2} = \frac{k}{n} \left(\frac{nLk}{k}\right)^{1/2}$ et donc $\frac{k}{n} = o\left(\left(\frac{kLk}{n}\right)^{1/2}\right)$ (2-4-23)

soit finalement:

$$u_{a_i} = 1 + \left(\frac{kLk}{n}\right)^{1/2} (1 + \varepsilon(n)) \quad (2-4-24)$$

par conséquent

$$\frac{u_{a_i}}{u_{a_i} - 1} = \frac{1}{A} \left(\frac{n}{kLk}\right)^{1/2} (1 + \varepsilon(n)) \quad (2-4-25)$$

Il nous faut maintenant étudier

$$P(X = a_i) = C_n^{a_i} \omega_2^{a_i} (1 - \omega_2)^{n-a_i}$$

Appliquons la formule de Stirling: en effet $n - a_i \rightarrow \infty$ et $a_i \rightarrow \infty$

Nous obtenons:

$$P(X = a_i) = \frac{1}{\sqrt{2\pi}} \left(\frac{n}{n - a_i}\right)^{1/2} \frac{1}{\sqrt{a_i}} \exp(U_4) (1 + \varepsilon(n)) \quad (2-4-26)$$

où

$$U_4 = nLn - (n - a_i)L(n - a_i) - a_i La_i + a_i L\omega_2 + (n - a_i)L(1 - \omega_2) \quad (2-4-27)$$

Effectuons quelques transformations sur U_4 :

$$U_4 = n \left[\frac{a_i}{n} L\left(\frac{n\omega_2}{a_i}\right) - \left(1 - \frac{a_i}{n}\right)L\left(1 - \frac{a_i}{n}\right) + \left(1 - \frac{a_i}{n}\right)L(1 - \omega_2) \right] \quad (2-4-28)$$

Comme

$$a_i = \frac{n}{k} + \eta_i' = \frac{n}{k} \left(1 + \eta_i' \frac{k}{n}\right) \quad \text{et} \quad \omega_2 = \frac{1}{k} \left(1 + \left(\frac{kLk}{n}\right)^{1/2} A\right)$$

On peut poser

$$a_i = n\omega_2 \left(1 + \left(\frac{kLk}{n}\right)^{1/2} A\right)^{-1} \left(1 + \eta_i' \frac{k}{n}\right) \quad (2-4-29)$$

mais d'après (2-4-23)

$$a_i = n\omega_2 \left(1 - \left(\frac{kLk}{n}\right)^{1/2} A (1 + \varepsilon(n))\right) \quad (2-4-30)$$

posons enfin

$$a_i = n\omega_2 (1 + t) \quad (2-4-31)$$

soit:

$$t = - \left(\frac{kLk}{n}\right)^{1/2} A (1 + \varepsilon(n)) \quad , \text{ et donc } t \rightarrow 0 \quad (2-4-32)$$

La similitude des calculs avec ceux effectués au chapitre 1 permet de poser:

$$U_4 = n \left[\omega_2 (1 + t)L(1 + t) + (1 - \omega_2)(1 - t\omega_2(1 - \omega_2)^{-1})L(1 - t\omega_2(1 - \omega_2)^{-1}) \right] \quad (2-4-33)$$

puis en développant:

$$U_4 = -n \left[\omega_2 \frac{t^2}{2} (1 + \varepsilon(n)) + \frac{t^2 \omega_2^2}{2} (1 - \omega_2)^{-1} (1 + \varepsilon(n)) \right] \quad (2-4-34)$$

Comme $\omega_2 \rightarrow 0$

$$U_4 = -n \frac{t^2 \omega_2^2}{2} (1 + \varepsilon(n)) \quad (2-4-35)$$

et enfin:

$$U_4 = - \frac{A^2}{2} Lk (1 + \varepsilon(n)) \quad (2-4-36)$$

soit:

$$P(X = a_i) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{a_i}} \exp \left[- \frac{A^2}{2} Lk (1 + \varepsilon(n)) \right] (1 + \varepsilon(n)) \quad (2-4-37)$$

et comme:

$$a_i = \frac{n}{k} \left(1 + \eta_i' \frac{k}{n}\right)$$

$$P(X = a_i) = \frac{1}{\sqrt{2\pi}} \left(\frac{k}{n}\right)^{1/2} \exp \left[-\frac{A^2}{2} Lk (1 + \epsilon(n)) \right] (1 + \epsilon(n)) \quad (2-4-38)$$

et en rapprochant (2-4-38) de (2-4-25) on a bien:

$$P(\bar{F}_1^{(i)}) \leq \frac{1}{A} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{Lk}} \exp \left[-\frac{A^2}{2} Lk (1 + \epsilon(n)) \right] (1 + \epsilon(n)) \quad (2-4-39)$$

la majoration étant uniforme.

Traisons maintenant $P(\bar{F}_2^{(i)})$: Nous définissons

$$\omega_1' = 1 - \frac{1}{k} \left(1 - \left(\frac{kLk}{n}\right)^{1/2} A\right) \quad (2-4-40)$$

$$\omega_2' = 1 - \omega_1' = \frac{1}{k} \left(1 - \left(\frac{kLk}{n}\right)^{1/2} A\right) \quad (2-4-41)$$

L'événement $\bar{F}_2^{(i)}$ s'écrit:

$$\{x_i > \frac{1}{k} \left(1 - \left(\frac{kLk}{n}\right)^{1/2} A\right)\} \text{ soit } \{x_i > \omega_2'\} \quad (2-4-42)$$

$$P(\bar{F}_2^{(i)}) = \int_0^{\omega_2'} f_2(x) dx \quad \text{où } f_2(x) \text{ est défini en } (2-4-3)$$

$$P(\bar{F}_2^{(i)}) = \int_0^{1 - \omega_1'} f_2(x) dx$$

$$P(\bar{F}_2^{(i)}) = \frac{n!}{(a_i - 1)!(n - a_i)!} \int_0^{1 - \omega_1'} x^{a_i - 1} (1 - x)^{n - a_i} dx \quad (2-4-43)$$

mais

$$\int_0^{1 - \omega_1'} x^{a_i - 1} (1 - x)^{n - a_i} dx = \int_{\omega_1'}^1 (1 - x)^{a_i - 1} x^{n - a_i} dx \quad (2-4-44)$$

et

$$\int_{\omega_1'}^1 = \int_{1 - \omega_2'}^1 = \int_0^1 - \int_0^{1 - \omega_2'}$$

soit:

$$P(\bar{F}_2^{(i)}) = 1 - \frac{n!}{(a_i - 1)!(n - a_i)!} \int_0^{1 - \omega_2'} (1 - x)^{a_i - 1} x^{n - a_i} dx$$

(2-4-45)

et d'après (2-4-7)

$$P(\bar{F}_2^{(i)}) = 1 - \sum_{j=0}^{a_i-1} C_n^j \omega_2^j (1 - \omega_2)^{n-j} \quad (2-4-46)$$

soit:

$$P(\bar{F}_2^{(i)}) = \sum_{j=a_i}^n C_n^j \omega_2^j (1 - \omega_2)^{n-j} \quad (2-4-47)$$

On reconnaît la queue de distribution d'une loi binômiale $B(n, \omega_2)$.

Associons à nouveau une v.a.r. X à cette loi:

$$P(\bar{F}_2^{(i)}) = P(X = a_i) + P(X = a_i + 1) + \dots + P(X = n) \quad (2-4-48)$$

posons encore:

$$u_j = \frac{P(X = j)}{P(X = j - 1)} ; \quad u_j = \frac{n - j + 1}{j} \frac{\omega_2}{1 - \omega_2}$$

On peut poser aussi:

$$a_i = n \omega_2' \left(1 + \left(\frac{kLk}{n}\right)^{1/2} A (1 + \varepsilon(n))\right) \quad (2-4-49)$$

par des calculs analogues à (2-4-30)

soit $j_0 = (n+1)\omega_2'$, correspondant à $u_{j_0} \sim 1$:

cette fois $a_i > j_0$; on est dans la zone des u_j inférieurs à 1.

On obtient alors toujours par la même technique:

$$P(\bar{F}_2^{(i)}) \leq \frac{1}{1 - u_{a_i}} P(X = a_i) \quad (2-4-50)$$

Sans recommencer tous les calculs, on peut remarquer que l'on passe de ω_2 à ω_2' en changeant A en $-A$; Nous avons donc:

$$u_{a_i} = 1 - \left(\frac{kLk}{n}\right)^{1/2} A (1 + \varepsilon(n)) \quad (2-4-51)$$

soit:

$$\frac{1}{1 - u_{a_i}} = \frac{1}{A} \left(\frac{n}{kLk}\right)^{1/2} (1 + \varepsilon(n)) \quad (2-4-52)$$

de même le développement de $P(X = a_i)$ se fait comme de (2-4-26) à (2-4-38) avec:

$$a_i = n \omega_2' (1 + t') \quad \text{où } t' = \left(\frac{kLk}{n}\right)^{1/2} A (1 + \varepsilon(n)) \quad (2-4-53)$$

on obtient donc:

$$P(X = a_i) = \frac{1}{\sqrt{2\pi}} \left(\frac{k}{n}\right)^{1/2} \exp \left[-\frac{A^2}{2} Lk (1 + \varepsilon(n)) \right] (1 + \varepsilon(n)) \quad (2-4-54)$$

soit

$$P(\bar{F}_2^{(i)}) \leq \frac{1}{A} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{Lk}} \exp \left[-\frac{A^2}{2} Lk (1 + \varepsilon(n)) \right] (1 + \varepsilon(n)) \quad (2-4-55)$$

Ce qui achève la démonstration de la proposition 2-3 ■

§ II - 5 Calcul d'un équivalent de $P(\bar{G}_{i_1}^* \dots \bar{G}_{i_r}^*)$

Comme au chapitre 1, nous allons appliquer pour le calcul d'un équivalent de $P(\bar{G}_{i_1}^* \dots \bar{G}_{i_r}^*)$ la proposition 2-2 avec $C = A$;

$$P(\bar{G}_{i_1}^* \dots \bar{G}_{i_r}^*) = (1 + \varepsilon(n)) U_5 \quad \text{où } U_5 \text{ est égal à:}$$

$$U_5 = \int_{\bar{G}_{i_1}^*} \dots \int_{\bar{G}_{i_r}^*} \frac{x_1^{a_1-1}}{\Gamma(a_1)} \dots \frac{x_r^{a_r-1}}{\Gamma(a_r)} n^{M_r} \exp(-nT_r) dx_1 \dots dx_r \quad (2-5-1)$$

soit:

$$U_5 = \prod_{q=1}^r \left[\int_{\bar{G}_{i_q}^*} \frac{(nx_q)^{a_q-1}}{\Gamma(a_q)} e^{-nx_q} n dx_q \right] \quad (2-5-2)$$

mais

$$\int_{\bar{G}_{i_q}^*} = \int_{L_{i_q}^*} + \int_{H_{i_q}^*} \quad (2-5-3)$$

et en posant $u = nx_q$ dans ces intégrales

$$\int_{L_{i_q}^*} = \int_0^{m_q} - \int_0^{m'_q} ; \quad \int_{H_{i_q}^*} = \int_{M_q}^{+\infty} - \int_{M'_q}^{+\infty} \quad (2-5-4) \text{ et } (2-5-5)$$

où m_q, m'_q, M_q, M'_q sont définis par:

$$m_q = \frac{n}{k} \left(1 - \left(\frac{n}{k}\right)^{1/2} \theta_1\right) ; \quad m'_q = \frac{n}{k} \left(1 - \left(\frac{kLk}{n}\right)^{1/2} A\right) \quad (2-5-5-1)$$

et

$$M_q = \frac{n}{k} \left(1 + \left(\frac{n}{k}\right)^{1/2} \theta_2\right) ; M'_q = \frac{n}{k} \left(1 + \left(\frac{kLk}{n}\right)^{1/2} A\right) \quad (2-5-5-2)$$

Soit:

$$\int_{L_{i_q}^*} = I_1 - I_2 \quad ; \quad \int_{H_{i_q}^*} = I_3 - I_4 \quad (2-5-6)$$

où I_1, I_2, I_3, I_4 sont définis ainsi par:

$$I_1 = \int_0^{m_q} \frac{u^{a_q-1}}{\Gamma(a_q)} e^{-u} du \quad (2-5-7)$$

$$I_2 = \int_0^{m'_q} \frac{u^{a_q-1}}{\Gamma(a_q)} e^{-u} du \quad (2-5-8)$$

$$I_3 = \int_{M_q}^{+\infty} \frac{u^{a_q-1}}{\Gamma(a_q)} e^{-u} du \quad (2-5-9)$$

$$I_4 = \int_{M'_q}^{+\infty} \frac{u^{a_q-1}}{\Gamma(a_q)} e^{-u} du \quad (2-5-10)$$

Nous allons utiliser les résultats de la proposition 1-4 du chapitre 1 pour évaluer les I_j , $j = 1, 2, 3, 4$. Reprenons les notations de la proposition 1-4:

Pour I_1 : cas 2

$$t = a_q - 1 = \frac{n}{k} \left(1 + (\eta'_q - 1) \frac{k}{n}\right) \quad (2-5-11)$$

$$r(t) = \frac{n}{k} \left(1 - \left(\frac{k}{n}\right)^{1/2} \theta_1\right) \quad (2-5-12)$$

$$r(t) = t(1 - \gamma(t)) \quad (2-5-13)$$

$$1 - \gamma(t) = \left(1 - \left(\frac{k}{n}\right)^{1/2} \theta_1\right) \left(1 - \eta''_q \frac{k}{n}\right)^{-1} \quad (2-5-14)$$

$$1 - \gamma(t) = \left(1 + \eta''_q \frac{k}{n} - \left(\frac{k}{n}\right)^{1/2} \theta_1 - \eta''_q \frac{k}{n}\right) \left(1 - \eta''_q \frac{k}{n}\right)^{-1}$$

$$1 - \gamma(t) = 1 - \left(\frac{k}{n}\right)^{1/2} \theta_1 - \eta''_q \frac{k}{n} \left(1 - \eta''_q \frac{k}{n}\right)^{-1}$$

et

$$\gamma(t) = \left(\frac{k}{n}\right)^{1/2} \theta_1 - \eta_q'' \frac{k}{n} \left(1 - \eta_q'' \frac{k}{n}\right)^{-1} \quad (2-5-15)$$

Calculons maintenant:

$$\lambda(t) = t [-\gamma(t) - L(1 - \gamma(t))]$$

$$\lambda(t) = -t\gamma(t) - tL\left(1 - \left(\frac{k}{n}\right)^{1/2} \theta_1\right) + tL\left(1 - \eta_q'' \frac{k}{n}\right) \quad (2-5-16)$$

$$\begin{aligned} \lambda(t) = & -\frac{n}{k} \left(\frac{k}{n}\right)^{1/2} \theta_1 - \eta_q'' \frac{k}{n} + \frac{n}{k} \left(1 - \eta_q'' \frac{k}{n}\right) \left(-\eta_q'' \frac{k}{n} - \frac{1}{2} \eta_q''^2 \frac{k^2}{n^2} (1 + \varepsilon(n))\right) \\ & + \frac{n}{k} \left(1 - \eta_q'' \frac{k}{n}\right) \left(\frac{k}{n}\right)^{1/2} \theta_1 + \frac{1}{2} \theta_1^2 \frac{k}{n} + \frac{1}{3} \left(\frac{k}{n}\right)^{3/2} \theta_1^3 (1 + \varepsilon(n)) \end{aligned}$$

$$\begin{aligned} \lambda(t) = & \frac{1}{2} \theta_1^2 + \frac{1}{3} \left(\frac{k}{n}\right)^{1/2} \theta_1^3 (1 + \varepsilon(n)) - \eta_q'' \left(\frac{k}{n}\right)^{1/2} \theta_1 - \frac{1}{2} \frac{k}{n} \eta_q''^2 \theta_1^2 \\ & - \frac{1}{3} \eta_q''^3 \left(\frac{k}{n}\right)^{3/2} \theta_1^3 (1 + \varepsilon(n)) - \frac{1}{2} \eta_q''^2 \frac{k}{n} (1 + \varepsilon(n)) + \eta_q''^2 \frac{k}{n} \\ & + \frac{1}{2} \eta_q''^3 \frac{k^2}{n^2} (1 + \varepsilon(n)) \quad (2-5-17) \end{aligned}$$

Soit en regroupant les termes:

$$\lambda(t) = \frac{1}{2} \theta_1^2 + \frac{1}{3} \left(\frac{k}{n}\right)^{1/2} \theta_1^3 (1 + \varepsilon(n)) + \varepsilon(n) \quad (2-5-18)$$

Comme $\theta_1 \sim \sqrt{2Lk}$, l'hypothèse $\frac{k(Lk)^3}{n} \rightarrow 0$ assure donc:

$$\lambda(t) = Lk - \frac{L_2 k}{2} + \frac{y_1}{2} + \varepsilon(n) \quad (2-5-19)$$

on a bien:

$$r(t) \rightarrow \infty, \quad \gamma(t) \rightarrow 0^+, \quad \lambda(t) \rightarrow \infty$$

donc:

$$I_1 \sim \frac{1}{2\sqrt{\pi}} e^{-\lambda} \frac{1}{\sqrt{\lambda}} \quad (2-5-20)$$

soit:

$$I_1 \sim \frac{1}{2\sqrt{\pi}} \frac{1}{k} e^{-(y_1/2)} \quad (2-5-21)$$

Le calcul de I_2 s'effectue simplement en changeant dans tous les calculs précédents θ_1 en $A\sqrt{Lk}$.

On obtient alors, si $\frac{k(Lk)^3}{n} \rightarrow 0$

$$r(t) = \frac{n}{k} \left(1 - \left(\frac{k}{n}\right)^{1/2} A \sqrt{Lk}\right) \rightarrow \infty, \quad \gamma(t) \sim \left(\frac{kLk}{n}\right)^{1/2} A \rightarrow 0$$

$$\lambda(t) = \frac{A^2}{2} Lk + \varepsilon(n) \rightarrow \infty \quad (2-5-22)$$

et d'après la proposition 1-4:

$$I_2 \sim \frac{1}{2\sqrt{\pi}} \frac{1}{k} A^{2/2} \left(\frac{k}{A^2}\right)^{1/2} \frac{1}{\sqrt{Lk}} \quad (2-5-23)$$

et donc $I_2 = o(I_1)$ si $A \geq \sqrt{2}$ (2-5-24)
dans ce dernier cas, on a dès lors:

$$\left. \begin{array}{l} \\ \\ \\ \end{array} \right\} L_{i_q}^* = \frac{1}{2\sqrt{\pi}} \frac{1}{k} e^{-(y_1/2)} (1 + \varepsilon(n)) \quad (2-5-25)$$

de manière uniforme.

Passons à I_3 et I_4 : Cas 1 de la proposition 1-4;

Pour I_3 nous obtenons:

$$t = a_q - 1 = \frac{n}{k} \left(1 + \eta_q' \frac{k}{n}\right) \quad |\eta_q'| < 3 \quad (2-5-26)$$

$$r(t) = \frac{n}{k} \left(1 + \left(\frac{k}{n}\right)^{1/2} \theta_2\right) \quad \text{et donc } r(t) \rightarrow \infty \quad (2-5-27)$$

$$r(t) = t(1 + \gamma(t)) \quad (2-5-28)$$

$$\lambda(t) = t [\gamma(t) - L(1 + \gamma(t))] \quad (2-5-29)$$

On trouve comme pour I_1 :

$$\gamma(t) = \left(\frac{k}{n}\right)^{1/2} \theta_2 (1 + \varepsilon(n)) \quad \text{donc } \gamma(t) \rightarrow 0 \quad (2-5-30)$$

$$\lambda(t) = \frac{1}{2} \theta_2^2 - \frac{1}{3} \left(\frac{k}{n}\right)^{1/2} \theta_2^3 (1 + \varepsilon(n)) + \varepsilon(n) \quad (2-5-31)$$

Par conséquent si $\frac{k(Lk)^3}{n} \rightarrow 0$

$$\lambda(t) = Lk - \frac{L_2 k}{2} + \frac{y_2}{2} + \varepsilon(n) \rightarrow +\infty \quad (2-5-32)$$

et en appliquant la proposition 1-4:

$$I_3 \sim \frac{1}{2\sqrt{\pi}} \frac{1}{k} e^{-(y_2/2)} \quad (2-5-33)$$

le calcul de I_4 s'effectue comme celui de I_3 en remplaçant θ_2 par $A\sqrt{Lk}$; il vient ainsi:

$$\lambda(t) = \frac{A^2}{2} Lk + \varepsilon(n) \quad \text{qui tend vers } +\infty \quad (2-5-34)$$

donc

$$I_4 \sim \frac{1}{2\sqrt{\pi}} \frac{1}{k} A^{2/2} \left(\frac{2}{A^2}\right) \frac{1}{\sqrt{Lk}} \quad (2-5-35)$$

ainsi I_4 est un $o(I_3)$ si $A \geq \sqrt{2}$ et dans ce cas

$$\left. \vphantom{\int} \right\} H_{i_q}^* = \frac{1}{2\sqrt{\pi}} \frac{1}{k} e^{-(y_2/2)} (1 + \varepsilon(n)) \quad (2-5-36)$$

de manière uniforme. Ceci donc nous donne:

$$\left. \vphantom{\int} \right\} \bar{G}_{i_q}^* = \left(\frac{1}{2\sqrt{\pi}} \frac{1}{k} (e^{-(y_1/2)} + e^{-(y_2/2)}) \right) (1 + \varepsilon(n)) \quad (2-5-37)$$

En remontant, nous aurons comme r est majoré par H fixé:

$$P(\bar{G}_{i_1}^* \dots \bar{G}_{i_r}^*) = (1 + \varepsilon(n)) \left(\frac{1}{2\sqrt{\pi}} \frac{1}{k} (e^{-(y_1/2)} + e^{-(y_2/2)}) \right)^r$$

Ce qui était bien le résultat demandé ■

§ II - 6 Fin de la démonstration de la proposition 2 - 1

Il nous faut rassembler les morceaux: puisque $P(\bar{G}_{i_1}^* \dots \bar{G}_{i_r}^*)$ est de l'ordre de $\frac{1}{k^r}$, nous pouvons utiliser la proposition 2-3 pour majorer $P(\overline{E_1^* E_2^*})$:

$$P(\overline{E_1^* E_2^*}) \leq \sum_q [P(\bar{F}_1^{(q)}) + P(\bar{F}_2^{(q)})] \quad (2-6-1)$$

$$P(\overline{E_1^* E_2^*}) \leq \frac{2H}{A} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{Lk}} \exp \left[-\frac{A^2}{2} Lk (1 + \epsilon(n)) \right] (1 + \epsilon(n)) \quad (2-6-2)$$

et $\exp\left(-\frac{A^2}{2} Lk\right)$ est un $o\left(\frac{1}{k^r}\right)$ si $\frac{A^2}{2} > H$. Supposons donc $A > \sqrt{2H}$

pour $H \geq 1$, on a bien $A \geq \sqrt{2}$ (2-6-3)

Alors:

$$P(\bar{G}_{i_1} \dots \bar{G}_{i_r}) \leq P(\bar{G}_{i_1}^* \dots \bar{G}_{i_r}^*) (1 + \epsilon(n)) \quad (2-6-4)$$

$$P(\bar{G}_{i_1} \dots \bar{G}_{i_r}) > P(\bar{G}_{i_1}^* \dots \bar{G}_{i_r}^*) \quad (2-6-5)$$

Donc

$$P(\bar{G}_{i_1} \dots \bar{G}_{i_r}) = (1 + \epsilon(n)) \left[\frac{1}{2\sqrt{\pi}} \frac{1}{k} \left(e^{-(y_1/2)} + e^{-(y_2/2)} \right) \right]^r \quad (2-6-6)$$

Ce qui démontre la proposition 2-1 ■

On achève alors la démonstration du th 2-1 comme celle du th 1-1.

§ II - 7 Généralisation du théorème 2-1 : th 2-2 et th 2-3

Les résultats obtenus sur les valeurs extrêmes partielles au chapitre 1 s'étendent au théorème 2-1 :

Considérons un découpage de $(1, 2, \dots, k)$ en plusieurs familles;
 soit $R = [1, k] = R_1 \cup R_2 \cup \dots \cup R_\lambda$ (2-7-1)
 où λ est un entier fixé.

Nous supposons que chaque R_m vérifie:

$$|R_m| = \Lambda_m(n) k \quad (2-7-2)$$

où $\Lambda_m(n) \rightarrow \Lambda_m^0$ lorsque $n \rightarrow +\infty$

Définissons:

$$Z_1^{(m)} = \min_{j \in R_m} \left(\frac{\sigma_j - 1/k}{1/\sqrt{nk}} \right) \quad (2-7-3)$$

$$Z_2^{(m)} = \max_{j \in R_m} \left(\frac{\sigma_j - 1/k}{1/\sqrt{nk}} \right) \quad (2-7-4)$$

définissons à nouveau:

$$\theta_{im} = (2Lk - L_2k + y_{im})^{1/2} \quad i = 1, 2 \quad (2-7-5)$$

où les réels y_{im} sont fixés. Alors le théorème 2-3 s'énonce:

Théorème 2 - 3

Si k vérifie $\frac{k(Lk)^3}{n} \rightarrow 0$,

$$P \left[\bigcap_m \{ Z_1^{(m)} < \theta_{1m} \}, \bigcap_m \{ Z_2^{(m)} < \theta_{2m} \} \right]$$

$$\rightarrow \exp \left[- \frac{1}{2\sqrt{\pi}} \sum_{m=1}^{\infty} \Lambda_m^0 \left(e^{-(y_{1m}/2)} + e^{-(y_{2m}/2)} \right) \right] \quad (2-7-6)$$

Il y a encore indépendance asymptotique entre les blocs définis par les R_m , et de plus entre les extrémums respectifs définis sur ces blocs. Ce résultat généralise bien celui de Tusnady [15], qui se limite à un seul bloc.

Nous ne donnerons pas le détail de la démonstration du th 2-2, qui est calquée sur celle du th 1-2. Il suffit de poser cette fois:

$$L_i = \left\{ \sigma_i < \frac{1}{k} - \frac{1}{\sqrt{nk}} \theta_{im} \right\} \quad \text{si } i \in R_m \quad (2-7-7)$$

$$H_i = \left\{ \sigma_i > \frac{1}{k} + \frac{1}{\sqrt{nk}} \theta_{2m} \right\} \quad \text{si } i \in R_m \quad (2-7-8) \quad 67$$

$$\bar{G}_i = L_i + H_i \quad (2-7-9)$$

On dénombre comme au th 1-2 les événements $P(\bar{G}_{i_1} \dots \bar{G}_{i_r})$ en fonction de leur provenance, faisant apparaître la limite uniforme vers laquelle tend S_r comme un développement multinomial:

$$S_r \sim \left(\frac{1}{2\sqrt{\pi}} \right)^r \frac{1}{r!} \left(\sum_{m=1}^{\ell} \Lambda_m^0 (e^{-(y_{1m}/2)} + e^{-(y_{2m}/2)}) \right)^r \quad (2-7-10)$$

où $r < H$ fixé. Comme:

$$\begin{aligned} P \left[\bigcap_m \{ Z_1^{(m)} < \theta_{1m} \}, \bigcap_m \{ Z_2^{(m)} < \theta_{2m} \} \right] \\ = 1 - S_1 + S_2 - \dots + (-1)^k S_k \end{aligned}$$

En appliquant les encadrements successifs de Bonferroni on arrive au résultat. ■

Assouplissement des conditions sur la taille des blocs

Nous nous sommes dans un premier temps limités à un découpage en k blocs de même taille à une unité près:

$$m_i = \frac{n}{k} + \eta_i^! \quad \text{avec} \quad |\eta_i^!| < 2$$

Il est aisé de passer de

$$|\eta_i^!| < 2 \quad (\forall i, \forall n) \quad \text{à} \quad |\eta_i^!| < H' \text{ fixé} \quad (\forall i, \forall n) \quad (2-7-12)$$

$$\text{Peut-on imposer à } \bar{\eta}(n) = \bar{\eta} = \sup_i (|\eta_i^!|) \quad (2-7-13)$$

des conditions permettant de mener à bien les démonstrations du théorème 2-1 (et donc du théorème 2-2) ?

$\bar{\eta}$ intervient ainsi dans la proposition 2-2 : on majore

$$\Lambda_r < H\bar{\eta}$$

(2 - 7 - 14)

Enonçons alors notre théorème 2-3:

Théorème 2-3

$$\text{Si } \bar{\eta} = \sup_i (|\eta_i''|) \text{ vérifie: } \bar{\eta} = o\left(\frac{n}{kLk}\right)^{1/2} \quad (2 - 7 - 15)$$

(blocs "bien équilibrés"), Alors les énoncés des th 2-1 et 2-2 demeurent valides.

Ceci permet dans la pratique d'assouplir les découpages définis par l'échantillon (cf chapitre 3).

Démonstration du th 2-3 : $\frac{n}{kLk}$ est un infiniment petit de $\frac{nLk}{k}$;

Il nous faut reprendre pas à pas les étapes où $\bar{\eta}$ intervient:

- en (2 - 3 - 10), nous passons de

$$\psi_n = \left(-\left(\frac{Lk}{kn}\right)^{1/2} R' + \frac{\Lambda_r}{n} \right) \left(1 - \frac{r}{k} - \left(\frac{Lk}{kn}\right)^{1/2} R' \right)^{-1}$$

à

$$\psi_n = -\left(\frac{Lk}{kn}\right)^{1/2} R' (1 + \epsilon(n)) \quad (2 - 3 - 12)$$

Cela est possible si $\frac{\Lambda_r}{n} = o\left(\left(\frac{Lk}{kn}\right)^{1/2}\right)$ (2 - 7 - 16)

mais si (2 - 7 - 15) est vérifié

$$\frac{\bar{\eta}}{n} = o\left(\left(\frac{Lk}{kn}\right)^{1/2}\right) \text{ et d'après (2 - 7 - 14) tout va bien.}$$

- dans la majoration de $P(\bar{F}_1^{(i)})$, il faut comparer j_0 et a_i .

Comme $\left(\frac{nLk}{k}\right)^{1/2}$ est par hypothèse un infiniment grand de $\bar{\eta}$, on a bien

$j_0 > a_i$ pour n suffisamment grand. Par ailleurs il faut en (2 - 4 - 18) que

$$\frac{\eta_i''}{n} \text{ soit infiniment petit devant } \left(\frac{Lk}{kn}\right)^{1/2} \quad (2 - 7 - 17)$$

cette condition est bien vérifiée sous les hypothèses du th 2-3 ;

Il nous faut également en (2 - 4 - 23) :

$$\bar{\eta} \frac{k}{n} = o\left(\left(\frac{kLk}{n}\right)^{1/2}\right) \quad (2 - 7 - 18)$$

ici encore tout va bien. de même, en (2 - 4 - 29) , il n'y a pas de problèmes.

donc $\bar{\eta} \frac{k}{n} \rightarrow 0$ et $P(\bar{F}_1^{(i)})$, $P(\bar{F}_2^{(i)})$ vérifient la proposition 2-3 .

Il nous reste à vérifier le développement asymptotique des I_j ($j = 1,2,3,4$)
(cf § II - 5) :

Dans le développement (2 - 5 - 17) de $\lambda(t)$, nous isolons les termes

contenant $\eta_q'' = 1 - \eta_q'$; $\eta_q'' \frac{k}{n} \rightarrow 0$ et on peut développer les termes :

$$\eta_q'' \left(\frac{k}{n}\right)^{1/2} \theta_1 , \quad \frac{1}{2} \frac{k}{n} \eta_q'' \theta_1^2 \quad \text{et} \quad \frac{1}{3} \left(\frac{k}{n}\right)^{3/2} \theta_1^3 \eta_q'' \quad (2 - 7 - 19)$$

tendent vers 0 car $\bar{\eta} \left(\frac{k}{n}\right)^{1/2} \theta_1 < \bar{\eta} A \left(\frac{kLk}{n}\right)^{1/2}$ qui tend vers 0 par hypothèse.

enfin :

$$\eta_q''^2 \frac{k}{n} \text{ est un infiniment petit devant } \frac{1}{Lk} \quad (2 - 7 - 20)$$

$$\eta_q''^3 \frac{k^2}{n^2} \text{ est un infiniment petit devant } \left(\frac{n}{n(Lk)^3}\right)^{1/2} \quad (2 - 7 - 21)$$

et on a donc bien sous les hypothèses du th 2-3 :

$$\lambda(t) = \frac{1}{2} \theta_1^2 + \frac{1}{3} \left(\frac{k}{n}\right)^{1/2} \theta_1^3 (1 + \varepsilon(n)) + \varepsilon(n) \quad (2 - 7 - 22)$$

soit comme par hypothèse :

$$\frac{k(Lk)^3}{n} \rightarrow 0 \quad : \quad \lambda(t) = Lk - \frac{L_2 k}{2} + \frac{y_1}{2} + \varepsilon(n) \quad (2 - 7 - 23)$$

ce qui termine la démonstration ■

Remarque: comme $\frac{k(Lk)^3}{n} \rightarrow 0$, $\frac{kLk}{n}$ est un infiniment petit devant $\frac{1}{(Lk)^2}$

et donc L_k est un infiniment petit devant $\left(\frac{n}{kL_k}\right)^{1/2}$;

on pourra ainsi prendre

$$\bar{\eta} = L_k$$

CHAPITRE TROISIEME

Dans ce chapitre, nous nous proposons d'appliquer les résultats du chapitre 2 à l'estimation de la densité par la méthode de l'histogramme aléatoire. Nous nous intéresserons à la recherche de lois limites pour l'expression: $(f - f_n)/f_n$, où f_n est l'estimateur de la densité de f .

Nous commencerons par étudier le comportement de cette expression dans trois cas:

- f minorée sur un compact th 3-1
- f décroissante sur un intervalle $[a, +\infty[$ th 3-2
- f s'annulant en a et croissante sur $[a, b]$ th 3-3

Dans chaque cas nous chercherons à construire un intervalle défini cette fois par l'échantillon, qui tende lorsque n tend vers l'infini vers le support de f , et tel que la loi limite soit vérifiée.

Enfin nous utiliserons les propriétés d'indépendance asymptotique remarquées au chapitre 2 pour obtenir une loi limite sur l'ensemble du support de f , f étant cette fois beaucoup plus générale: th 3-3, et nous conclurons en posant le problème d'optimisation des vitesses de convergence, en fonction de h , k et n .

§ III - 1 Notations et définitions

Dans tout ce chapitre f est la densité de probabilité d'une variable aléatoire à valeurs dans \mathbb{R} . Nous serons amenés à faire certaines hypothèses de régularité sur f dans la suite de ce chapitre. Pour le moment, et dans tout ce qui suit, nous imposons f Lipchitzienne.

Associons à f un échantillon de la loi P_f :

$$(X_1, X_2, \dots, X_n)$$

et notons cette fois encore l'échantillon ordonné associé :

$$(X_1^*, X_2^*, \dots, X_n^*) \quad \text{où} \quad X_1^* \leq X_2^* \leq \dots \leq X_n^*$$

Nous partitionnons l'intervalle $[X_1^*, X_n^*]$ en k blocs

$$[Y_i^*, Y_{i+1}^*] \quad i \in [1, k]$$

Chaque bloc $[Y_i^*, Y_{i+1}^*]$ contenant m_i cellules élémentaires $[X_j^*, X_{j+1}^*]$, les m_i vérifiant :

$$m_i = \frac{n}{k} + \eta_i \quad (3-1-1)$$

nous noterons encore : $\bar{\eta} = \sup_i (|\eta_i|)$ (3-1-2)

bien entendu les m_i vérifient :

$$\sum_i m_i = n - 1 \quad (3-1-3)$$

Soit $\Delta_i = Y_{i+1}^* - Y_i^*$ (3-1-4)

la méthode de l'estimation de la densité par histogramme aléatoire consiste à définir $f_n(x)$ par :

$$- f_n(x) = 0 \quad \text{sur} \quad]-\infty, X_1^*] \cup [X_n^*, +\infty[$$

$$- f_n(x) = \frac{1}{k\Delta_i} \quad \text{si} \quad x \in [Y_i^*, Y_{i+1}^*[\quad (3-1-5)$$

La convergence uniforme de l'estimateur f_n sur \mathbb{R} a été étudiée par Lecoutre [9] et Abou-Jaoudé [1]. Lecoutre étudie f_n dans le cas où f est continue et positive sur un compact (donc minorée). Il montre alors que la condition:

$$k = o\left(\frac{n}{\ln n}\right) \quad (\text{ou ce qui est équivalent: } \frac{k \ln k}{n} \rightarrow 0) \quad (3-1-6)$$

est nécessaire et suffisante pour que

$$\sup_x |f_n(x) - \hat{f}(x)| \xrightarrow{\chi} 0 \quad \text{lorsque } n \rightarrow \infty$$

où χ signifie "en probabilité" dans le cas "nécessaire" et "presque complètement sûrement" dans le cas "suffisant".

Les lois limites ont été étudiées par Tusnady [15]. Tusnady reprend une expression de Revesz:

$$Q_1 = \max_{I_n} \left[\frac{f - f_n}{f} \right] \quad (3-1-7)$$

et

$$Q'_1 = \max_{I_n} \left| \frac{f - f_n}{f} \right|$$

où I_n est un intervalle qui croît vers le support de f lorsque $n \rightarrow \infty$. De plus, I_n est déterminé par l'échantillon (X_1, \dots, X_n) .

Tusnady définit alors une famille (T_1) de fonctions f vérifiant:

- i) $f > 0$ sur l'intérieur de $\text{supp}(f)$ (support de f) (3 - 1 - 8)
- ii) f dérivable à dérivée bornée (donc f Lipchitzienne) (3 - 1 - 9)
- iii) $\forall \alpha^* \in]0, \frac{1}{2}[$, $\forall x$ vérifiant $\alpha^* \leq F(x) \leq 1 - \alpha^*$:

$$|f'(x)| \leq \frac{C^*}{\alpha^*} f^2(x) \quad \text{où } C^* \text{ est une constante fixée (3 - 1 - 10)}$$

l'intervalle I_n est défini par une fonction $h(n)$ à valeurs dans \mathbb{N} et croissante, vérifiant

$$h(n) = o(n) \quad (3-1-11)$$

Tusnady définit des blocs $[Y_i^*, Y_{i+1}^*]$ de manière analogue à (3-1-1), mais en prenant $\bar{\eta}$ borné, et en travaillant non pas sur $[X_1^*, X_n^*]$, mais sur

$$I_n = [X_h^*, X_{n-h}^*]$$

le théorème de Tusnady s'énonce alors:

Théorème C:

Si h, k et n vérifient:

$$\frac{k(Lk)^3}{n} \rightarrow 0 \quad (3-1-12)$$

$$\frac{n}{k} \left(\frac{-}{-}\right)^{3/2} \frac{1}{h} L\left(\frac{-}{-}\right) \rightarrow 0 \quad (3-1-13)$$

si $f \in (T_1)$, alors quand $n \rightarrow \infty$

$$P\left[\left(\frac{2nLk}{k}\right)^{1/2} Q_1 < 2Lk - \frac{L_2k}{2} + t\right] \rightarrow \exp\left(-\frac{1}{2\sqrt{\pi}} e^{-t}\right) \quad (3-1-14)$$

et

$$P\left[\left(\frac{2nLk}{k}\right)^{1/2} Q_1' < 2Lk - \frac{L_2k}{2} + t\right] \rightarrow \exp\left(-\frac{1}{\sqrt{\pi}} e^{-t}\right) \quad (3-1-15)$$

Remarquons que cette loi limite est mise sous une autre forme que celle de Revesz (cf chapitre 1): il suffit de remarquer que:

$$\sqrt{2Lk} (2Lk - L_2k + 2t)^{1/2} = 2Lk - \frac{L_2k}{2} + t + \varepsilon(n) \quad (3-1-16)$$

Il est simple de montrer que notre théorème 2-1 entraîne celui de Tusnady, dans le cas où f est la loi uniforme sur $[0,1]$: on retrouve le Lemme 2 de Tusnady [15]; En effet dans le cas où la loi est uniforme, posons:

$$Z_1 = -\left(\frac{n}{k}\right)^{1/2} \min_i (k\Delta_i - 1) \quad ; \quad Z_2 = \left(\frac{n}{k}\right)^{1/2} \max_i (k\Delta_i - 1) \quad (3-1-17)$$

et (3-1-18)

On peut si (3-1-12) est vérifié, écrire: (th 2-1)

$$P\left[\sqrt{2Lk} Z_1 < 2Lk - \frac{L_2k}{2} + t_1, \sqrt{2Lk} Z_2 < 2Lk - \frac{L_2k}{2} + t_2\right] \rightarrow \exp\left[-\frac{1}{2\sqrt{\pi}} (e^{-t_1} + e^{-t_2})\right] \quad (3-1-19)$$

Il nous faut montrer que la loi limite (3 - 1 - 19) est encore valable si l'on remplace Z_1 et Z_2 par:

$$W_1 = - \left(\frac{n}{k}\right)^{1/2} \min_x \left(\frac{1 - f_n}{1}\right) = - \left(\frac{n}{k}\right)^{1/2} \min_i \left(1 - \frac{1}{k\Delta_i}\right) \quad (3 - 1 - 20)$$

$$W_2 = \left(\frac{n}{k}\right)^{1/2} \max_i \left(\frac{1 - f_n}{1}\right) = \left(\frac{n}{k}\right)^{1/2} \max_i \left(1 - \frac{1}{k\Delta_i}\right) \quad (3 - 1 - 21)$$

Nous allons montrer que:

$$\sqrt{Lk} (W_1 - Z_1) \xrightarrow{p} 0 \quad \text{et} \quad \sqrt{Lk} (W_2 - Z_2) \xrightarrow{p} 0 \quad (3 - 1 - 22)$$

et utiliser la proposition 1-5.

Proposition 3-1

$$\forall i \in [1, k] \quad \text{si} \quad \frac{k(Lk)^3}{n} \rightarrow 0, \quad$$

$$P\left(\bigcap_i \left\{ k\Delta_i < 1 + \left(\frac{kLk}{n}\right)^{1/2} A \right\}\right) \rightarrow 1 \quad (3 - 1 - 23)$$

$$P\left(\bigcap_i \left\{ k\Delta_i > 1 - \left(\frac{kLk}{n}\right)^{1/2} A \right\}\right) \rightarrow 1 \quad (3 - 1 - 24)$$

$$\text{où } A > \sqrt{2} \quad (3 - 1 - 25)$$

Il s'agit là d'une transposition de la proposition 1-6 ; Il suffit en effet pour vérifier la proposition 3-1 de reprendre les calculs du th 2-1 en remplaçant θ_1 et θ_2 par $A\sqrt{Lk}$.

Comme en (1 - 7 - 25), on obtient:

$$P\left(\bigcap_i G_i\right) \geq 1 - \frac{C'}{k\alpha_1^{1-I}} \quad \text{où } \alpha_1 > 1 \quad (3 - 1 - 26)$$

et donc:

$$P\left(\bigcap_i G_i\right) \rightarrow 1 \quad (3 - 1 - 27)$$

(bien entendu ces résultats s'étendent au cas où les m_i sont tels que

$$\bar{\eta} = o\left(\left(\frac{n}{kLk}\right)^{1/2}\right)$$

D'après la remarque faite en (1 - 7 - 30) il suffit de majorer uniformément en i :

$$\left(\frac{nLk}{k}\right)^{1/2} \left[(k\Delta_i - 1) - \left(1 - \frac{1}{k\Delta_i}\right) \right] = \left(\frac{nLk}{k}\right)^{1/2} (k\Delta_i - 1)^2 \frac{1}{k\Delta_i}$$

(3 - 1 - 28)

Comme la proposition 3-1 entraîne:

$$\min_i k\Delta_i \xrightarrow{p} 1 \quad \text{car} \quad \frac{kLk}{n} \rightarrow 0$$

(3 - 1 - 29)

$$\max_i (k\Delta_i - 1)^2 < A^2 \frac{kLk}{n}$$

(3 - 1 - 30)

Ainsi $\forall i$, et surtout uniformément en i :

$$\left(\frac{nLk}{k}\right)^{1/2} (k\Delta_i - 1)^2 < \frac{k(Lk)^{3/2}}{n} \left(\frac{n}{k}\right)^{1/2} = \left(\frac{k(Lk)^3}{n}\right)^{1/2}$$

(3 - 1 - 31)

qui tend bien vers 0 sous les hypothèses de Tusnady ■

Nous préférons, comme au chapitre 1, substituer à l'expression

$$\frac{f - f_n}{f}$$

l'expression :

$$\frac{f - f_n}{f_n} = H_n$$

Comme Tusnady, nous chercherons à définir certaines familles de densités de probabilité, pour lesquelles nous saurons définir un intervalle I_n sur lequel nos théorèmes de lois limites s'appliqueront; cet intervalle devra être déterminé par l'échantillon. Nous verrons que le fait d'avoir remplacé au dénominateur f par f_n perturbe profondément les conditions entre h , k et n dans le cas où $I_n \rightarrow [a, +\infty[$.

§ III - 2 Cas où f est minorée sur un compact K_0 (avec éventuellement $P_f(K_0) < 1$)

Supposons f strictement positive sur un compact: f étant par hypothèse continue admet un minimum strictement positif. supposons pour simplifier que le compact soit un intervalle $[a,b]$; (a et b étant connus)

Posons dès lors:

$$\alpha = F(a) \quad \text{et} \quad \beta = F(b) \quad , \text{ où } F \text{ est la fonction de répartition}$$

associée à f . Considérons le pavage associé à l'échantillon de taille n (X_1, \dots, X_n). Nous savons que les variables

$$(F(X_1^*), F(X_2^*), \dots, F(X_n^*))$$

constituent un échantillon de taille n de la loi uniforme sur $[0,1]$.

Si nous ne considérons que les blocs $[Y_i^*, Y_{i+1}^*]$ tels que:

$$a \leq Y_i^* \leq Y_{i+1}^* \leq b \quad (3-2-2)$$

comme d'après les travaux de Smirnov [14], les statistiques ordonnées $X_{\alpha n}^*$ et $X_{\beta n}^*$ convergent presque complètement sûrement lorsque $n \rightarrow \infty$ vers a et b , le nombre des blocs de la forme (3-2-2) vérifie:

$$\# R = (\beta - \alpha) k (1 + \varepsilon(n)) \quad (3-2-3)$$

(où R est le nombre de ces blocs).

Ainsi, si

$$\frac{k(Lk)^3}{n} \rightarrow 0 \quad \text{et} \quad \bar{\eta} = o\left(\left(\frac{n}{kLk}\right)^{1/2}\right), \text{ nous pouvons appliquer le th 2-2}$$

avec $\lambda = 1$ aux quantités:

$$\sigma_i = F(Y_{i+1}^*) - F(Y_i^*) \quad (3-2-4)$$

$$\text{avec } \Lambda_0 = \beta - \alpha \quad (3-2-5)$$

$$\text{Par ailleurs } f(x) > \delta \text{ sur } [a,b] \quad (3-2-6)$$

D'après le théorème des accroissements finis:

$$F(Y_{j+1}^*) - F(Y_j^*) = f(z_j) (Y_{j+1}^* - Y_j^*) \quad (3-2-7)$$

comme $Y_{j+1}^* - Y_j^* = \Delta_j$, (3-2-7) s'écrit:

$$\sigma_j > \delta \Delta_j \quad (3-2-8)$$

D'après la proposition 3-1:

$$\max_j \sigma_j < \frac{1}{p} + A \left(\frac{Lk}{nk} \right)^{1/2} \quad \text{dès que } A > \sqrt{2}$$

Ainsi:

$$\max_j \Delta_j < \frac{1}{p} \left(\frac{1}{\delta} + A \left(\frac{Lk}{nk} \right)^{1/2} \right) < \frac{2}{k\delta} \quad \text{pour } n \text{ grand} \quad (3-2-9)$$

et donc:

$$\Delta_j \rightarrow 0 \quad \text{uniformément en } j \quad \text{pour tous les blocs vérifiant } (3-2-2).$$

Par ailleurs nous pouvons écrire:

$$H_n = \frac{f - f_n}{f_n} = \frac{f - \bar{f}_n}{\bar{f}_n} + \frac{\bar{f}_n - f_n}{f_n} \quad (3-2-10)$$

où

$$\bar{f}_n(x) = \frac{1}{\Delta_j} \int_{Y_j^*}^{Y_{j+1}^*} f(t) dt \quad \text{pour } x \in [Y_j^*, Y_{j+1}^*] \quad (3-2-11)$$

Il nous faut étudier:

$$\left(\frac{nLk}{k} \right)^{1/2} H_n$$

Occupons nous de la partie $f - \bar{f}_n$ (non probabiliste):

$$f_n = \frac{1}{k\Delta_j} \quad \text{et} \quad f(x) = \frac{1}{\Delta_j} \int_{Y_j^*}^{Y_{j+1}^*} f(x) dt \quad (3-2-12)$$

Ainsi:

$$\frac{f - \bar{f}_n}{f_n} = k \int_{Y_j^*}^{Y_{j+1}^*} (f(x) - f(t)) dt \quad (3-2-13)$$

f étant Lipchitzienne par hypothèse:

$$|f(x) - f(t)| < T_2 \cdot |x - t| < T_2 \Delta_j \quad (3-2-14)$$

Ainsi:

$$\frac{f - \bar{f}_n}{f_n} \leq k T_2 \Delta_j^2 \quad (3-2-15)$$

et d'après (3-2-9)

$$\max_x \left(\frac{f - \bar{f}_n}{f_n} \right) \leq \frac{4T_2}{p} \frac{1}{\delta^2} \frac{1}{k} = \frac{T_2'}{k} \quad (3-2-16)$$

D'après la proposition 1-5, il nous suffit d'assurer:

$$\left[\max_x \left(\frac{f - \bar{f}_n}{f_n} \right) \right] \left(\frac{nLk}{k} \right)^{1/2} \rightarrow 0 \text{ lorsque } n \rightarrow \infty \quad (3-2-17)$$

Ceci est vérifié d'après (3-2-16) si $\frac{nLk}{k^3} \rightarrow 0$ (Condition C_0)

(3-2-18)

La seconde quantité:

$\frac{\bar{f}_n - f_n}{f_n}$ est constante¹ sur chaque bloc $[Y_j^*, Y_{j+1}^*]$, où elle vaut:

$$k \left[\int_{Y_j^*}^{Y_{j+1}^*} f(t) dt \right] - 1 = k\sigma_j - 1 \quad (3-2-19)$$

Donc

$$\left(\frac{n}{k} \right)^{1/2} \left(\frac{\bar{f}_n - f_n}{f_n} \right) = \frac{\sigma_j - 1/k}{1/\sqrt{nk}} \quad (3-2-20)$$

Il nous reste à déterminer un intervalle I_n qui tende vers $[a, b]$ lorsque $n \rightarrow \infty$: ce sera l'intervalle $[Y_\lambda^*, Y_\mu^*]$ où λ et μ sont déterminés par (3-2-2). mais $\alpha = F(a)$ et $\beta = F(b)$ ne sont pas connus, dépendant de f ;

Nous travaillerons donc avec les estimateurs de Glivenko-Cantelli:

$$\alpha_n = F_n(a) \text{ et } \beta_n = F_n(b) \quad (3-2-21)$$

où

$$F_n(t) = \frac{1}{n} \{ \text{nombre des } X_i \leq t \} \quad (3-2-22)$$

Alors comme:

$$\lim_{n \rightarrow \infty} P \left(\max_x |F_n(x) - F(x)| < \frac{\xi}{\sqrt{n}} \right) = \sum_{-\infty}^{+\infty} (-1)^k e^{-2k^2 \xi^2} \quad (3-2-23)$$

$$\beta_n - \alpha_n = (\beta - \alpha)(1 + \epsilon(n)) \quad (3-2-24)$$

Nous partagerons donc l'intervalle $[Y_\lambda^*, Y_\mu^*]$ en blocs du type $[Y_j^*, Y_{j+1}^*]$.

¹ Remarque: f_n n'est en fait défini que sur l'intervalle $[Y_j^*, Y_{j+1}^*]$: il y a en effet saut à chaque extrémité Y_j^* : nous prendrons par pure convention dans les calculs la valeur égale au centre de l'intervalle, éventuellement prolongée par continuité sur les extrémités suivant les besoins des calculs.

Nous pouvons ainsi énoncer le théorème 3 - 1:

Théorème 3 - 1

Soient n et k vérifiant:

$$\frac{nLk}{k^3} \rightarrow 0 \quad (C_0) \quad \frac{k(Lk)^3}{n} \rightarrow 0 \quad \text{et} \quad \bar{\eta} = o\left(\left(\frac{n}{kLk}\right)^{1/2}\right)$$

On considère alors l'intervalle $I_n = [Y_\lambda^*, Y_\mu^*]$ déterminé par (3 - 2 - 2).
Alors:

$$P\left(-\left(\frac{n}{k}\right)^{1/2} \min_{I_n} H_n < \theta_1, \left(\frac{n}{k}\right)^{1/2} \max_{I_n} H_n < \theta_2\right) \\ \rightarrow \exp\left[-\frac{\beta - \alpha}{2\sqrt{\pi}} \left(e^{-(y_1/2)} + e^{-(y_2/2)}\right)\right]$$

β et α étant estimés par $\beta_n = F_n(b)$ et $\alpha_n = F_n(a)$.

Remarquons que la condition (C_0) intervient dans la plupart des théorèmes sur l'estimation de la densité; elle est à rapprocher de la condition (1 - 7 - 3) ou (1 - 7 - 9) de Révész, ou de la condition (1 - 7 - 18) de notre théorème 1 - 3 - 1 (Cas où $s = 1$).

On pourrait aussi se poser le problème dans le cas où a et b ne seraient pas connus a-priori: si f est minorée "un peu autour" du compact K_0 , on peut alors se permettre de laisser un des blocs $[Y_j^*, Y_{j+1}^*]$ dépasser les bornes du compact inconnu: par exemple en estimant sur l'intervalle formé par les blocs $[Y_j^*, Y_{j+1}^*]$ contenus dans l'intervalle $[X_{n\alpha}^*, X_{n\beta}^*]$: cette fois ce sont a et b qui sont inconnus, et α et β qui sont fixés; il est facile alors d'énoncer un théorème analogue au théorème 3 - 1 dans le cas où f est minorée par δ' sur $]a - \epsilon, b + \epsilon[$; en effet pour n suffisamment grand on a

$$a - \epsilon < X_{n\alpha}^* < X_{n\beta}^* < b - \epsilon$$

et nous appliquons les calculs précédents.

Remarque:

Un encadrement de β et α précis peut être effectué à partir des résultats de Massey [10] qui précisent la loi limite de

$$\max_x |F_n(x) - F(x)|$$

§ III - 3 Etude de l'estimation de f sur un segment $[a, +\infty[$ (1ère partie)

Dans ce paragraphe, nous allons faire quelques hypothèses supplémentaires sur f :

f sera supposée strictement positive sur $[a, +\infty[$ (3 - 3 - 0)

f sera supposée de classe C^2 , vérifiant également

$f'(x) < 0$ sur $[a, +\infty[$ (3 - 3 - 1)

On définit la fonction:

$$\varphi(x) = - \frac{f'(x)}{f(x)} \quad (3 - 3 - 2)$$

que nous supposerons monotone sur $[a, +\infty[$

On définit la fonction:

$$\psi(x) = x \varphi(x) \quad (3 - 3 - 3)$$

que nous supposerons également monotone sur $[a, +\infty[$

On appellera (F) la famille des densités de probabilité vérifiant ces conditions.

Nous noterons $\xi(n)$ la quantité:

$$\xi(n) = \left(\frac{k^3}{nLk} \right)^{1/2} \quad (3 - 3 - 4)$$

la condition (C_0) s'écrit ainsi (cf th 3-1):

$$(C_0) \quad \xi(n) \rightarrow \infty \quad (3 - 3 - 5)$$

Remarque:

$\psi(x)$ ne peut être un $o(1)$

En effet dans ce cas, nous aurions:

$$|Lf(x)| = o(|Lx|)$$

(3 - 3 - 6)

et $f(x)$ serait un infiniment grand devant $\frac{1}{x^\varepsilon}$ pour tout $\varepsilon > 0$;

mais alors l'intégrale $\int_a^{+\infty} f(t) dt$ ne serait pas définie, ce qui est impossible.

Nous allons mettre l'expression H_n sous une forme plus maniable:

$$H_n = \frac{f(x) - f_n(x)}{f_n(x)} \quad \text{soit} \quad H_n = H_n(x)$$

(cf p79) Comme au paragraphe précédent, nous définissons une partition en blocs $[Y_j^*, Y_{j+1}^*]$ contenus dans $[a, +\infty[$, soit un intervalle $[Y_\lambda^*, Y_\mu^*]$ inclus dans $[a, +\infty[$. sur chaque bloc $[Y_j^*, Y_{j+1}^*]$, la quantité $f_n(x)$ est constante. or f est continue sur $[a, +\infty[$.

Il existe donc pour chaque j deux points (x_j, x'_j) tels que:

$$H_n(x_j) = \max_{[Y_j^*, Y_{j+1}^*]} \left[\frac{f(x) - f_n(x)}{f_n(x)} \right] \quad (3 - 3 - 7)$$

$$H_n(x'_j) = \min_{[Y_j^*, Y_{j+1}^*]} \left[\frac{f(x) - f_n(x)}{f_n(x)} \right] \quad (3 - 3 - 8)$$

Le nombre des blocs $[Y_j^*, Y_{j+1}^*]$ est de la forme:

$$\# R = k (1 - \alpha)(1 + \varepsilon(n)) \quad (3 - 3 - 9)$$

(R étant la famille des indices caractérisant les blocs qui nous intéressent)
 α est toujours défini par $F(a) = \alpha$. Nous pouvons donc appliquer les résultats du chapitre 2.

Le théorème des accroissements finis nous donne:

$$F(Y_{j+1}^*) - F(Y_j^*) = f(z_j) (Y_{j+1}^* - Y_j^*) \quad (3 - 3 - 10)$$

ou encore avec les notations introduites plus haut:

$$\sigma_j = f(z_j) \Delta_j \quad (3 - 3 - 11)$$

ainsi sur $[Y_j^*, Y_{j+1}^*]$:

$$f_n(x) = \frac{1}{k\Delta_j} = \frac{f(z_j)}{k\sigma_j} \quad (3-3-12)$$

et ainsi

$$\frac{f - f_n}{f_n} = -1 + k\sigma_j \frac{f(x)}{f(z_j)} \quad (3-3-13)$$

Donc nous pouvons écrire:

$$\begin{aligned} \max_{x \in [Y_\lambda^*, Y_\mu^*]} \left(\frac{f(x) - f_n(x)}{f_n(x)} \right) &= \max_{j \in R} \max_{x \in [Y_j^*, Y_{j+1}^*]} \left(\frac{f(x) - f_n(x)}{f_n(x)} \right) \\ &= \max_{j \in R} \left(\frac{f(x_j) - f_n(x)}{f_n(x)} \right) \\ &= \max_{j \in R} \left(-1 + \frac{f(x_j)}{f(z_j)} k\sigma_j \right) \end{aligned} \quad (3-3-14)$$

et de même:

$$\min_{x \in [Y_\lambda^*, Y_\mu^*]} \left(\frac{f(x) - f_n(x)}{f_n(x)} \right) = \min_{j \in R} \left(-1 + \frac{f(x'_j)}{f(z_j)} k\sigma_j \right) \quad (3-3-15)$$

Le théorème 3-2 nous donne la loi limite des quantités:

$$\max_{j \in J} (-1 + k\sigma_j) \quad , \quad \min_{j \in J} (-1 + k\sigma_j) \quad (3-3-16)$$

D'après la proposition 1-5, pour affirmer que les quantités (3-3-14) et (3-3-15) d'une part, (3-3-16) d'autre part ont même loi limite, il suffit de prouver: en posant

$$\left(\frac{nLk}{k} \right)^{1/2} k\sigma_j \left(1 - \frac{f(x_j)}{f(z_j)} \right) = A_j \quad (3-3-17)$$

il nous faut assurer:

$$\max_j A_j \xrightarrow{p} 0 \quad \text{si} \quad n \rightarrow \infty$$

Comme $k\sigma_j \xrightarrow{p} 1$ uniformément en j , il suffit d'étudier:

$$A_j^! = \left(\frac{nLk}{k} \right)^{1/2} \left(\frac{f(z_j) - f(x_j)}{f(z_j)} \right) \quad (3-3-18)$$

D'après le théorème des accroissements finis:

il existe $\xi_j \in]Y_j^*, Y_{j+1}^*[$ vérifiant :

$$f(z_j) - f(x_j) = f'(\xi_j) (z_j - x_j) \quad (3-3-19)$$

Nous allons définir un réel r_n appartenant à $[a, +\infty[$ possédant la propriété suivante:

- sur l'intervalle $[a, r_n]$, la quantité A_j tend en probabilité vers 0, le théorème 3-2 est applicable;
- lorsque $n \rightarrow \infty$, $r_n \rightarrow \infty$

Nous cherchons à définir un intervalle I_n à partir des statistiques ordonnées (X_1^*, \dots, X_n^*) ; d'après les travaux de Geffroy [6] et de Chevalier [3], nous pouvons minorer en probabilité la quantité $F(r_n)$ par la statistique ordonnée $F(X_{\gamma(n)}^*)$

où $\gamma(n) = n - h(n)$, avec $h(n) = o(n)$ et $h(n) \rightarrow \infty$ quand $n \rightarrow \infty$ h vérifiant enfin:

$$F(r_n) = 1 - \frac{h}{n} + \frac{\sqrt{h}}{n\epsilon(n)} \quad \text{où } \epsilon(n) \rightarrow 0 \quad (3-3-20)$$

Ainsi toute condition sur r_n se transforme-t-elle en une condition sur h . Nous prendrons un minorant "large" en imposant:

$$\frac{h}{n} = (1 - F(r_n)) C \quad \text{où } C \text{ est } \underline{\text{fixé}} \text{ supérieur à } 1 \quad (3-3-21)$$

Nous supposons que f est monotone décroissante sur $[a, +\infty[$; par conséquent:

$$\sigma_j > f(r_n) \Delta_j \quad (3-3-22)$$

Comme $k\sigma_j \xrightarrow{p} 1$ uniformément en j :

$$\Delta_j < \frac{1}{p} \frac{1 + \epsilon}{f(r_n) k} \quad (\epsilon \text{ ext fixé}) \quad (3-3-23)$$

D'autre part, sur l'intervalle $[a, r_n]$ la fonction

$$\varphi(x) = - \frac{f'(x)}{f(x)}$$

est majorée par un nombre dépendant de r_n : $M(r_n)$

$$\forall x \in [a, r_n], \varphi(x) < M(r_n) \quad (3-3-24)$$

Nous allons reprendre une transformation de Tusnady [15].

Comme

$$\begin{aligned} |f(z_j) - f(x_j)| &\leq |f'(\xi_j)| |z_j - x_j| \\ &\leq |f'(\xi_j)| \Delta_j \\ &\leq |f(\xi_j)| \Delta_j M(r_n) \end{aligned} \quad (3-3-25)$$

soit:

$$\left| \frac{f(z_j) - f(x_j)}{f(z_j)} \right| \leq \left| \frac{f(\xi_j)}{f(z_j)} \right| \Delta_j M(r_n) \quad (3-3-26)$$

mais :

$$\left| \frac{f(\xi_j)}{f(z_j)} \right| \leq 1 + \left| \frac{f(\xi_j) - f(z_j)}{f(z_j)} \right| \quad (3-3-27)$$

on peut ainsi itérer le processus en définissant $\xi_j^{(1)}, \dots, \xi_j^{(l)}, \dots$
et l'on obtient:

$$\begin{aligned} \left| \frac{f(z_j) - f(x_j)}{f(z_j)} \right| &\leq \Delta_j M(r_n) + [\Delta_j M(r_n)]^2 + \dots + [\Delta_j M(r_n)]^m \\ &\quad + [\Delta_j M(r_n)]^m \left| \frac{f(\xi_j^{(m-1)})}{f(z_j)} \right| \end{aligned} \quad (3-3-28)$$

et comme

$$\left| \frac{f(\xi_j^{(m-1)})}{f(z_j)} \right| < M'(r_n) \quad \text{sur } [a, r_n] \quad (3-3-29)$$

Si uniformément en j:

$$\Delta_j M(r_n) < \frac{1 - \epsilon}{p} \quad \text{on peut passer à la limite en } m \text{ et obtenir:}$$

$$\left| \frac{f(z_j) - f(x_j)}{f(z_j)} \right| \leq \frac{\Delta_j M(r_n)}{1 - \Delta_j M(r_n)} \quad (\epsilon \text{ est fixé}) \quad (3-3-30)$$

Ce sera ainsi le cas d'après (3-3-23) si

$$\frac{M(r_n)}{f(r_n)} \frac{1}{k} < \frac{1 - \epsilon}{p} \quad \text{où } \epsilon \text{ est fixé.} \quad (3-3-31)$$

Dans ce cas, on peut majorer $A_j^!$ uniformément par:

$$A_j^! < \frac{nLk}{p} \frac{1}{k} \frac{M(r_n)}{f(r_n)} \frac{1}{k} (1 + \epsilon(n)) \quad (3-3-32)$$

Pour assurer $A_j^! \rightarrow 0$ il suffit que le terme de droite de (3-3-32)

tende vers 0 à l'infini. Mais dans ce cas, (3-3-31) est bien vérifié.

Il nous faut assurer ainsi:

$$\frac{1}{\xi(n)} \frac{M(r_n)}{f(r_n)} \rightarrow 0 \quad (3-3-33)$$

§ III- 4 Etude de l'estimation de f sur un segment $[a, +\infty[$ (2ème partie)

Dans ce paragraphe, nous allons séparer les différents cas de (F):
La fonction φ étant monotone positive, deux cas sont possibles:

- A: φ est croissante (3 - 4 - 1)
- B: φ est décroissante (3 - 4 - 2)

Dans l'hypothèse A deux cas sont à nouveau possibles:

- a: $\varphi(x) \rightarrow \infty$ (3 - 4 - 3)
- b: $\varphi(x)$ est majorée, et donc $\varphi(x) \rightarrow L > 0$ (3 - 4 - 4)

Traisons le cas A : famille des fonctions "à décroissance rapide"

Considérons la fonction décroissante:

$$x(x) = \frac{1}{\varphi(x)} \quad (3-4-5)$$

x étant positive tend vers une limite L' lorsque x tend vers l'infini.
Si : $x'(x) \rightarrow 0$ et $x''(x) \rightarrow 0$

Alors d'après Dieudonné [4],

$$\int_x^{+\infty} f(t) dt \sim - \frac{f^2(x)}{f'(x)} \quad (3-4-6)$$

et donc:

$$1 - F(r_n) \sim - \frac{f^2(r_n)}{f'(r_n)} \quad (3-4-7)$$

Or d'après l'hypothèse A, $M(r_n) = - \frac{f'(r_n)}{f(r_n)}$ (3 - 4 - 8)

la condition (3 - 3 - 33) devient:

$$\frac{1}{\xi(n)} \frac{f'(r_n)}{f^2(r_n)} \rightarrow 0 \quad (3-4-9)$$

Soit, d'après (3 - 3 - 21) et (3 - 4 - 7)

$$\frac{n}{h} = o(\xi(n)) \quad \text{ou} \quad Lk \frac{n^3}{k^3 h^2} \rightarrow 0 \quad (\text{condition } (C_1)) \quad (3 - 4 - 10)$$

Traitons maintenant le cas B :

φ est décroissante positive: deux cas sont possibles:

- a: φ est minorée par une constante positive et donc $\varphi(x) \rightarrow L > 0$ lorsque x tend vers l'infini.

Alors $M(r_n) < C_1^*$ fixé et il suffit que:

$$\frac{1}{\xi(n)} \frac{1}{f(r_n)} \rightarrow 0 \quad (3 - 4 - 11)$$

Nous avons d'autre part:

$$\varphi(x) > L \quad \text{soit} \quad - \frac{f'(x)}{f(x)} > L \quad (3 - 4 - 12)$$

d'où

$$- \frac{f'(x)}{f(x)} > \frac{L}{f(x)} \quad (3 - 4 - 13)$$

l'équivalent (3 - 4 - 7) est encore valable et donc

$$1 - F(r_n) \sim - \frac{f'(r_n)}{f^2(r_n)} \quad (3 - 4 - 14)$$

Ceci entraîne

$$C_2^* \frac{n}{h} > \frac{1}{f(r_n)}$$

et nous retrouvons la condition (C_1) .

- b: φ tend vers 0 :

Alors ψ étant monotone, deux cas sont possibles:

- 1: $\psi(x)$ est croissante
- 2: $\psi(x)$ est décroissante

Traitons le cas 2 : ψ étant décroissante et positive admet une limite L' lorsque x tend vers l'infini. si $L' < 1$, il existe ε fixé avec $L' + \varepsilon < 1$, $\varepsilon > 0$ et on a:

$$-x \frac{f'}{f} < L' + \varepsilon$$

pour x suffisamment grand. ceci entraînerait alors:

$$f > \frac{C_3^*}{x^{L'+\varepsilon}} \quad \text{et } f \text{ ne serait pas intégrable sur } [a, +\infty[, \text{ ce qui est}$$

impossible par hypothèse. Ainsi $L' \geq 1$.

Remarquons en passant que le cas $L' = 1$ est totalement indéterminé:

$$\text{si } -\frac{f'}{f} = \frac{1}{x} + \frac{B}{xLx} \quad \text{où } B \text{ est fixé positif} \quad (3-4-15)$$

alors

$$f(x) = \frac{C_3^*}{x} \left(\frac{1}{Lx}\right)^B \text{ sera intégrable ou non suivant que } B \text{ sera supérieur}$$

ou inférieur à 1. Or dans les deux cas, $\psi(x) \rightarrow 1$.

Nous limiterons donc notre étude à la famille (F_1) des fonctions "à décroissance assez rapide" qui vérifient

$$\lim_{x \rightarrow \infty} \psi(x) > 1 \quad (3-4-16)$$

Si $f \in (F_1)$, nous pouvons supposer que

$$\psi(x) > L' > 1 \quad (3-4-17)$$

où L' est un réel fixé.

Remarque la famille des lois de Pearson d'équation

$$\frac{f'}{f} = \frac{ax + b}{a'x^2 + b'x + c'} \quad (3-4-18)$$

est dans (F_1) : il suffit de remarquer que les trois cas litigieux:

$$(1) \quad \frac{f'}{f} = \frac{-1}{x - \alpha} \quad (3-4-19)$$

$$(2) \quad \frac{f'}{f} = \frac{A}{x - r_1} + \frac{B}{x - r_2}, \quad A + B = -1 \quad (3-4-20)$$

$$(3) \quad \frac{f'}{f} = \frac{-x + c}{(x + a)^2 + b^2} \quad (3-4-21)$$

conduisent à des densités non-intégrables.

Une autre famille de fonctions de (F_1) est formée des lois Log-normales de densités:

$$f(x) = A x^\alpha \exp(-\beta (Lx)^2) \quad \text{avec } \beta > 0 \quad (3-4-22)$$

en effet

$$-x \frac{f'}{f} = -\alpha + 2\beta Lx : \text{ on est dans le cas } \underline{B} \underline{b} \underline{1} .$$

Ainsi le cas B se divise dans le cas b en:

$$\underline{1\text{-bis}}: \psi(x) \rightarrow L' > 1$$

$$\underline{2\text{-bis}}: \psi(x) \rightarrow +\infty$$

enfin, comme $\varphi(x) \rightarrow 0$, $M(r_n)$ est majoré par une constante C_0^* fixée, et il suffit que

$$\frac{1}{\xi(n)} \frac{1}{f(r_n)} \rightarrow 0 : \text{ c'est } f(r_n) \text{ qu'il faut travailler.}$$

Plaçons-nous dans le cas B b 1-bis : Nous pouvons écrire car $L' > 1$

$$\psi(x) = L' + \varepsilon(x) \quad (3-4-23)$$

avec $\varepsilon(x) \rightarrow 0$ lorsque $x \rightarrow \infty$. Par conséquent:

$$\varphi(x) = \frac{L'}{x} + \frac{\varepsilon(x)}{x} \quad (3-4-24)$$

Il existe ainsi deux constantes C_1^* et C_2^* fixés vérifiant pour x assez grand:

$$\frac{C_1^*}{x} < -\frac{f'}{f} < \frac{C_2^*}{x}, \quad C_1^* \text{ et } C_2^* \text{ étant supérieurs à } 1 \quad (3-4-25)$$

Intégrant cette inégalité sur $[a, +\infty[$, il vient:

$$D_2 \frac{1}{x^{C_2^*}} < f(x) < D_1 \frac{1}{x^{C_1^*}} \quad (3-4-26)$$

et en intégrant à nouveau:

$$\frac{D_2}{C_2^* - 1} \frac{1}{r_n^{C_2^* - 1}} < 1 - F(r_n) < \frac{D_1}{C_1^* - 1} \frac{1}{r_n^{C_1^* - 1}} \quad (3-4-27)$$

où D_1 et D_2 sont deux nouvelles constantes d'intégration.

Comme d'après (3 - 3 - 21) :

$$\frac{h}{n} = (1 - F(r_n)) C \quad \text{où} \quad C > 1 \quad (3 - 4 - 28)$$

nous avons

$$r_n < C_3^* \left(\frac{-}{h}\right)^{\frac{n C_1^* - 1}{h}} \quad (3 - 4 - 29)$$

soit:

$$\frac{1}{f(r_n)} < D_3 \left(\frac{-}{h}\right)^{\frac{n C_1^* - 1}{h}} \quad (3 - 4 - 30)$$

nous voulons:

$$\frac{1}{\xi(n)} \frac{1}{f(r_n)} \rightarrow 0 \quad (3 - 4 - 31)$$

Il suffit pour cela que pour toutes les valeurs de C_1^* , C_2^* :

$$\frac{1}{\xi(n)} \left(\frac{-}{h}\right)^{n D_4} \rightarrow 0 \quad \text{où} \quad D_4 = \frac{C_2^*}{C_1^* - 1} \quad (3 - 4 - 32)$$

et (3 - 4 - 33)

ceci entraîne: pour tout réel positif D_4 :

$$D_4 L\left(\frac{-}{h}\right) - L(\xi(n)) \rightarrow -\infty \quad (3 - 4 - 34)$$

$$L\left(\frac{-}{h}\right) \left[D_4 - \frac{L(\xi(n))}{L\left(\frac{-}{h}\right)} \right] \rightarrow -\infty \quad (3 - 4 - 35)$$

soit:

$$L\left(\frac{-}{h}\right) = o(L(\xi(n))) \quad \text{Condition } (C_2) \quad (3 - 4 - 36)$$

Plaçons-nous maintenant dans le cas B b 2-bis : Alors nous pouvons écrire

$$\varphi(x) = \frac{1}{x \varepsilon(x)} \quad \text{où} \quad \varepsilon(x) \rightarrow 0 \quad \text{et} \quad \varepsilon_1(x) = \frac{1}{x \varepsilon(x)} = \varphi(x) \rightarrow 0 \quad (3 - 4 - 37)$$

$\varepsilon_1(x)$ et $\varepsilon(x)$ sont monotones décroissants. (en effet $\varepsilon(x) = \frac{1}{\psi(x)}$)

Quel que soit $C_1^* > 2$ fixé, il existe un $x_0(C_1^*)$ tel que pour tout $x > x_0(C_1^*)$:

$$\frac{C_1^*}{x} < \frac{1}{x \varepsilon(x)} = - \frac{f'}{f} \quad (3 - 4 - 38)$$

Intégrons cette inégalité entre $x_0(C_1^*)$ et $x > x_0(C_1^*)$:

$$C_1^* L\left(\frac{x}{x_0(C_1^*)}\right) < -Lf(x) + Lf(x_0(C_1^*)) \quad (3-4-39)$$

soit:

$$f(x) < D_1 \frac{1}{C_1^* x} \quad \text{pour } x \text{ suffisamment grand} \quad (3-4-40)$$

Mais

$$-\frac{f'}{f} < \frac{1}{x} \frac{1}{\varepsilon(r_n)} \quad \text{si } x \in [a, r_n] \quad (3-4-41)$$

et

$$f(r_n) > D_2 \frac{1}{r_n (1/\varepsilon(r_n))} = D_2 \exp\left[-\frac{Lr_n}{\varepsilon(r_n)}\right] \quad (3-4-42)$$

intégrons (3-4-40) entre r_n et $+\infty$: il vient

$$1 - F(r_n) < D_2' \frac{1}{C_1^{*-1} r_n} \quad (3-4-43)$$

soit:

$$r_n < D_3 \left(\frac{n}{h}\right)^{\frac{1}{C_1^{*-1}}} \quad (3-4-45)$$

comme pour n suffisamment grand:

$$\frac{1}{f(r_n)} < \frac{1}{D_2} \exp\left[\frac{Lr_n}{\varepsilon(r_n)}\right] \quad (3-4-46)$$

$$\frac{1}{f(r_n)} < \frac{1}{D_2} \exp\left[r_n Lr_n\right] \quad \text{car } \frac{1}{r_n \varepsilon(r_n)} = \varepsilon_1(r_n) < 1 \quad (3-4-47)$$

et

$$\frac{1}{f(r_n)} < \frac{1}{D_2} \exp\left[D_3 \left(\frac{n}{h}\right)^{\frac{1}{C_1^{*-1}}} \left(LD_3 + \frac{1}{C_1^{*-1}} L\left(\frac{n}{h}\right)\right)\right]$$

d'où

$$\frac{1}{f(r_n)} < \frac{1}{D_2} \exp\left[D_4 \left(\frac{n}{h}\right)^{\frac{1}{C_1^{*-1}}} L\left(\frac{n}{h}\right)\right] \quad (3-4-48)$$

nous obtenons ainsi la condition:

$$\frac{1}{\xi(n)} \exp \left[D_4 \left(\frac{n}{h} \right)^{\frac{1}{C_1^* - 1}} L \left(\frac{n}{h} \right) \right] \rightarrow 0 \quad (3 - 4 - 49)$$

ou encore:

$$D_4 \left(\frac{n}{h} \right)^{\frac{1}{C_1^* - 1}} L \left(\frac{n}{h} \right) - L(\xi(n)) \rightarrow -\infty \quad (3 - 4 - 50)$$

(ici D_4 dépend de C_1^* , mais est fixé si C_1^* est fixé)

Il suffit donc qu'il existe un D_5 fixé positif vérifiant:

$$\frac{n}{h} < [L(\xi(n))]^{D_5} \quad (3 - 4 - 51)$$

Ceci s'écrit:

$$L \left(\frac{n}{h} \right) = O(L_2 \xi(n)) \quad \text{Condition } (C_3) \quad (3 - 4 - 52)$$

Remarque:

Le fait de connaître plus précisément f permet bien souvent d'améliorer les conditions entre n , k et h ; mais on objectera avec raison qu'une estimation paramétrique devient alors plus efficace

Ainsi dans le cas d'une loi Log-normale:

$$f(x) = A x^\alpha \exp(-\beta (Lx)^2) \quad \beta > 0 \quad (3 - 4 - 53)$$

$$1 - F(r_n) = \int_{r_n}^{+\infty} t^\alpha \exp(-\beta (Lt)^2) dt \quad (3 - 4 - 54)$$

posons $Lt = u$ (3 - 4 - 55)

$$1 - F(r_n) = \int_{Lr_n}^{+\infty} e^{(\alpha + 1)u - \beta u^2} du \quad (3 - 4 - 56)$$

soit:

$$h(u) = \exp((\alpha + 1)u - \beta u^2) \quad (3 - 4 - 57)$$

alors:

$$-\frac{h'(u)}{h(u)} = -(\alpha + 1) + 2\beta u \quad (3 - 4 - 58)$$

Alors nous avons:

$$\int_{Lr_n}^{+\infty} h(u) du \sim -\frac{h^2(Lr_n)}{h'(Lr_n)} \quad (3 - 4 - 59)$$

$$1 - F(r_n) = C^{-1} \frac{h}{n} \sim \frac{1}{2\beta L r_n} r_n^{\alpha+1} \exp(-\beta(Lr_n)^2) \quad (3-4-60)$$

comme :

$$f(r_n) = A r_n^\alpha \exp(-\beta(Lr_n)^2) \quad (3-4-61)$$

$$C^{-1} \frac{h}{n} \sim \frac{1}{2\beta L r_n} r_n f(r_n)$$

Il suffira donc que :

$$\frac{1}{\xi(n)} \frac{r_n}{L r_n} \frac{n}{h} \rightarrow 0 \quad (3-4-63)$$

en prenant le logarithme de (3-4-60) il vient :

$$-\frac{n}{h} - LC \sim -L(2\beta) - L_2 r_n + (\alpha + 1)Lr_n - \beta(Lr_n)^2 \quad (3-4-64)$$

d'où

$$\frac{n}{h} \sim \beta(Lr_n)^2 \quad (3-4-65)$$

soit :

$$Lr_n \sim \left[\frac{1}{\beta} \frac{n}{h} \right]^{1/2} \quad \text{et} \quad Lr_n \sim \frac{1}{2} L_2 \frac{n}{h} \quad (3-4-66)$$

prenant le logarithme de (3-4-63) il vient :

$$Lr_n + \frac{n}{h} - L_2 r_n - L\xi(n) \rightarrow -\infty$$

Comme

$$Lr_n = o\left(\frac{n}{h}\right) \quad (3-4-67)$$

il suffit que

$$\frac{n}{h} < D_1 L\xi(n) \quad \text{où} \quad D_1 < 1 \quad (3-4-68)$$

condition plus forte que (C₁) mais plus faible que (C₂) .

Dans la pratique on estimera comme en III-2 l'intervalle [a, +∞[par [X_λ^{*}, X_{n-h(n)}^{*}] où h(n) vérifie (C₁), (C₂) ou (C₃)

Comme $h(n)$ est un $o(n)$, le nombre des blocs $[Y_j^*, Y_{j+1}^*]$ contenus dans cet intervalle sera de la forme

$$\# R = k (1 - \alpha)(1 + \epsilon(n)) \quad (3 - 4 - 70)$$

ce qui permet d'appliquer les résultats du chapitre 2. Nous pouvons ainsi énoncer le théorème 2:

Théorème 2

Soient k, n et h vérifiant avec $\bar{\eta}$:

$$\frac{k(Lk)^3}{n} \rightarrow 0, \quad \bar{\eta} = o\left(\left(\frac{n}{kLk}\right)^{1/2}\right) \quad \text{et} \quad \frac{k^3}{nLk} = \xi^2(n) \rightarrow \infty \quad (C_0)$$

Soit f de classe C^2 vérifiant les propriétés de monotonie à l'infini définies en III-3 : famille (F)

- Alors si $f \in (F_2)$, famille des fonctions à décroissance rapide ($\varphi(x) \neq 0$), il suffit que:

$$\frac{n}{h} = o(\xi(n)) \quad (C_1)$$

- Alors si $f \in (F_1)$, famille des fonctions à décroissance assez rapide, ($\lim_{x \rightarrow \infty} \psi(x) > 1$), deux cas sont possibles:

- si $\psi(x) \rightarrow L'$, il suffit que:

$$\frac{n}{Lh} = o(L\xi(n)) \quad (C_2)$$

- si $\psi(x) \rightarrow \infty$, il suffit que:

$$\frac{n}{Lh} = O(L_2\xi(n)) \quad (C_3)$$

Dans chacun de ces cas: en posant $I_n = [X_{\lambda}^*, X_{n-h(n)}^*]$, estimant α par $\alpha_n = F_n(a)$

$$P\left(-\left(\frac{n}{k}\right)^{1/2} \min_{I_n} H_n < \theta_1, \left(\frac{n}{k}\right)^{1/2} \max_{I_n} H_n < \theta_2\right)$$

$$\rightarrow \exp\left[-\frac{1-\alpha}{2\sqrt{\pi}} \left(e^{-(y_1/2)} + e^{-(y_2/2)}\right)\right]$$

Remarques :

- On a bien entendu $(F_2) \quad (F_1) \quad (F)$
 et $(C_3) \Rightarrow (C_2) \Rightarrow (C_1)$

- la condition (C_2) suffit pour les lois de Pearson.

- la condition (C_1) est équivalente à celle de Tusnady: le fait de diviser par f_n au lieu de f fait perdre beaucoup d'information dans le cas des décroissances lentes à l'infini.

- la fonction $h(n)$ peut toujours être définie car $\zeta(n) \rightarrow \infty$;

- comme $h(n) = o(n)$, $I_n \xrightarrow{p} [a, +\infty[$ quand $n \rightarrow \infty$

§ III- 5 Etude au voisinage d'un zéro de f

Nous supposons ici que nous estimons f sur un intervalle $[a, b]$ où f est positive strictement sur $]a, b[$ et vérifie $f(a) = 0$.

Nous allons chercher un intervalle I_n défini à partir des statistiques ordonnées (X_1^*, \dots, X_n^*) qui tende vers $[a, b]$ lorsque $n \rightarrow \infty$, et sur lequel la loi limite de H_n soit vérifiée.

Nous pouvons toujours estimer: $\alpha = F(a)$ et $\beta = F(b)$ par $\alpha_n = F_n(a)$ et $\beta_n = F_n(b)$.

De même, nous nous ramenons aisément à $a = 0$: nous chercherons d'abord à déterminer un intervalle $[r_n, b]$ sur lequel la loi limite reste valable: Alors si h est une fonction qui vérifie:

$$h(n) \rightarrow \infty, \quad h(n) = o(n) \quad (3-5-2)$$

$$\frac{h}{n} = C [F(r_n) - F(a)] \quad \text{où} \quad C > 1 \quad (3-5-2)$$

la statistique $X_{\alpha_n + h(n)}^*$ majore en probabilité r_n et:

$$X_{\alpha_n + h(n)}^* \rightarrow a = 0 \quad \text{lorsque} \quad n \rightarrow \infty$$

par ailleurs le nombre de blocs $[Y_j^*, Y_{j+1}^*]$ contenus dans $[X_{\alpha}^* n + h(n), b]$ est égal à

$$\# R = k (\beta - \alpha)(1 + \epsilon(n)) \quad (3-5-4)$$

Et nous pouvons appliquer les résultats du chapitre 2. Soit : $[Y_{\lambda}^*, Y_{\mu}^*] = U_R [Y_j^*, Y_{j+1}^*]$

Pour obtenir un équivalent de

$$F(r_n) - F(0) = \int_0^{r_n} f(t) dt \quad (3-5-5)$$

posons

$$t = e^{-u} \quad (3-5-6)$$

Alors:

$$C^{-1} \frac{h}{n} = \int_{-Lr_n}^{+\infty} f(e^{-u}) e^{-u} du \quad (3-5-7)$$

soit

$$g(u) = f(e^{-u}) e^{-u} \quad (3-5-8)$$

nous étant ainsi ramenés au voisinage de l'infini, nous pouvons reprendre les techniques du paragraphe précédent:

le $\varphi(u)$ associé à g est égal à:

$$\varphi(u) = 1 + e^{-u} \frac{f'(e^{-u})}{f(e^{-u})} \quad (3-5-9)$$

soit en notation x :

$$\hat{\varphi}(x) = 1 + x \frac{f'(x)}{f(x)} \quad (3-5-10)$$

nous supposons f croissante sur $[a, b]$, et $\hat{\varphi}$ monotone sur ce segment.

alors $\varphi(u)$ est monotone et minorée et par conséquent:

$$\int_{Lr_n}^{+\infty} f(e^{-u}) e^{-u} du \sim - \frac{g^2(u)}{g'(u)} \quad (3-5-11)$$

d'où

$$C \frac{h}{n} \sim \frac{\varphi(u)}{g(u)} \sim \frac{1}{r_n f(r_n)} + \frac{f'(r_n)}{f^2(r_n)} \quad (3-5-12)$$

nous cherchons à majorer l'expression

$$\frac{1}{\xi(n)} \frac{M(r_n)}{f(r_n)}$$

où $M(r_n)$ est le sup sur $[r_n, b]$ de $\frac{f'(x)}{f(x)}$

- si $\varphi(u)$ est croissante, ce qui revient à supposer $\hat{\varphi}(x)$ décroissante:

$$M(r_n) = \frac{f'(r_n)}{f(r_n)} \quad (3-5-13)$$

comme:

$$C \frac{n}{h} > \frac{f'(r_n)}{f(r_n)} \frac{1}{f(r_n)} \quad (3-5-14)$$

la loi limite de H_n sera assurée si

$$\frac{n}{h} = o(\xi(n)) \quad (C_1) \quad (3-5-15)$$

- si $\varphi(u)$ est décroissante, $M(r_n) < D_1$ fixé $(3-5-16)$

mais:

$$\frac{1}{r_n f(r_n)} + \frac{f'(r_n)}{(f(r_n))^2} > \frac{1}{r_n f(r_n)} \quad (3-5-17)$$

et comme $r_n \rightarrow 0$

$$\frac{1}{r_n f(r_n)} > \frac{1}{f(r_n)} \quad (3-5-18)$$

d'où

$$\frac{1}{f(r_n)} < C \frac{n}{h}$$

et donc il suffit pour assurer la loi limite de H_n de prendre:

$$\frac{n}{h} = o(\xi(n)) \quad (C_1)$$

on s'aperçoit ainsi que pour assurer la loi limite de H_n au voisinage d'un zéro de f , la condition (C_1) est suffisante. Cette différence avec l'étude au voisinage de l'infini est probablement due au fait qu'une fonction continue qui s'annule à distance finie est intégrable, tandis qu'une fonction qui s'annule "à l'infini" ne l'est pas forcément.

Nous pouvons donc énoncer le théorème 3-3:

Théorème 3 - 3

Soit f de classe C^2 , vérifiant: $x \frac{f'}{f}$ monotone sur l'intervalle $[a, b]$, et vérifiant $f(a) = 0$.

Si h, k et n vérifient avec $\bar{\eta}$:

$$\frac{k(Lk)^3}{n} \rightarrow 0, \quad \bar{\eta} = o\left(\left(\frac{n}{kLk}\right)^{1/2}\right), \quad h \rightarrow \infty, \quad h = o(n)$$

et: $\frac{k^3}{nLk} = \xi^2(n) \rightarrow \infty$; Alors si: $\frac{n}{h} = o(\xi(n)) \quad (C_1)$

On a:

$$P\left(-\left(\frac{n}{k}\right)^{1/2} \min_{I_n} H_n < \theta_1, \left(\frac{n}{k}\right)^{1/2} \max_{I_n} H_n < \theta_2\right) \\ \rightarrow \exp\left[-\frac{\beta - \alpha}{2\sqrt{\pi}} \left(e^{-\theta_1^2/2} + e^{-\theta_2^2/2}\right)\right]$$

où $I_n = [Y_{\lambda}^*, Y_{\mu}^*]$, $\alpha = F(a)$ et $\beta = F(b)$ étant estimés par

$$\alpha_n = F_n(a) \quad \text{et} \quad \beta_n = F_n(b)$$

Il est clair que l'on peut énoncer un théorème analogue pour f décroissante sur $[a, b]$ et s'annulant en $x = b$

§ III - 6 Cas général: estimation sur le support global de f

Nous avons traité dans les paragraphes précédents des phénomènes locaux: estimation au voisinage d'un zéro, sur un voisinage de l'infini, sur un compact où f est minorée.

Nous allons maintenant utiliser l'indépendance asymptotique des blocs constatée au théorème 2-2 pour définir une loi limite de H_n sur l'ensemble du support de f .

Considérons la famille (H) des densités de probabilité définies sur \mathbb{R} vérifiant:

Il existe une partition de $]-\infty, +\infty[$ définie par

$$-\infty = a_0 < a_1 < a_2 \dots < a_l = +\infty \quad (3-6-1)$$

telle que sur chaque intervalle $[a_i, a_{i+1}]$, la fonction f définissant la densité soit dans l'un des trois cas suivants:

- 1) f minorée sur $[a_i, a_{i+1}]$ (a_i et a_{i+1} étant à distance finie)
- 2) f s'annule en a_i (a_i étant à distance finie), est croissante sur $[a_i, a_{i+1}]$, et vérifie les propriétés de monotonie du § III-5 (respectivement: f s'annule en a_{i+1} (à distance finie, lui aussi) et f est décroissante sur $[a_i, a_{i+1}]$)
- 3) f vérifie sur $[a_i, +\infty[$ les propriétés de monotonie des § III-2 et III-3 (respectivement sur $]-\infty, a_i]$)

Alors si f appartient à (H) , on peut construire à partir des statistiques ordonnées (X_i^*) un découpage formé par les (Y_λ^*) associés à chaque $[a_i, a_{i+1}]$. il faudra ainsi considérer des fonctions h_j dans les cas 2) et 3).

Par exemple, au voisinage d'un zéro a_i de f , on écartera $h_j(n)$ (où $h_j(n) = o(\xi(n))$) statistiques ordonnées de part et d'autre de a_i .

On associe ainsi à chaque intervalle $[a_i, a_{i+1}]$ un intervalle empirique $[X_{\mu(i)}^*, X_{\mu(i+1)}^*] = \Xi_i$ (3-6-2)

intervalle qui converge lorsque n tend vers l'infini vers $[a_i, a_{i+1}]$. Sur cet intervalle empirique, on construit alors le découpage en blocs équilibrés, chaque bloc comprenant

$$m_j = \frac{n}{k} + \eta_j \quad (3-6-3)$$

mailles élémentaires. le nombre de blocs contenus dans $[X_{\mu(i)}^*, X_{\mu(i+1)}^*] = \Xi_i$ est égal à:

$$\# R_j = k (\alpha_{i+1} - \alpha_i)(1 + \varepsilon(n)) \quad (3-6-4)$$

où $\alpha_i = F_n(a_i)$.

Si alors dans chaque intervalle Ξ_i les fonctions h, k et n vérifient les conditions (C_i) suffisantes pour assurer localement la loi limite de H_n sur $[a_i, a_{i+1}]$, la proposition 1-5 montre alors que :

les expressions:

$$\left(-\binom{n}{k}^{1/2} \min_{\mathbb{R}_i} \left[\frac{f(x) - f_n(x)}{f_n(x)} \right], \binom{n}{k}^{1/2} \max_{\mathbb{R}_i} \left[\frac{f(x) - f_n(x)}{f_n(x)} \right] \right)$$

où $i \in [0, \ell - 1]$

(3 - 6 - 5)

ont même loi que les statistiques:

$$\left(-\min_{\mathbb{R}_i} \left(\frac{\sigma_j - 1/k}{1/\sqrt{nk}} \right), \max_{\mathbb{R}_i} \left(\frac{\sigma_j - 1/k}{1/\sqrt{nk}} \right) \right)$$

où $i \in [0, \ell - 1]$

(3 - 6 - 6)

Nous pouvons appliquer le théorème 2-2 à l'expression (3 - 6 - 6) d'après (3 - 6 - 4); définissons toujours:

$$\theta_{ji} = (2Lk - L_2k + y_{ji})^{1/2} \quad j = 1, 2 \quad (3 - 6 - 7)$$

On a alors, si l'ensemble des conditions entre les h_j , k et n sont respectées:

$$\begin{aligned} & P \left[\bigcap_i \left\{ -\min_{\mathbb{R}_i} \left(\binom{n}{k}^{1/2} H_n \right) < \theta_{1i} \right\}, \bigcap_i \left\{ \max_{\mathbb{R}_i} \left(\binom{n}{k}^{1/2} H_n \right) < \theta_{2i} \right\} \right] \\ & \rightarrow \exp \left[-\frac{1}{2\sqrt{\pi}} \sum_{i=0}^{\ell-1} (\alpha_{i+1} - \alpha_i) \left(e^{-(y_{1i}/2)} + e^{-(y_{2i}/2)} \right) \right] \end{aligned}$$

(3 - 6 - 8)

où $\alpha_0 = 0$, $\alpha_i = F(a_i)$ et $\alpha_\ell = 1$

Nous voulons maintenant prendre le sup et le inf de $\binom{n}{k}^{1/2} H_n$ lorsque x décrit

$$I_n = \bigcup_i \mathbb{R}_i \quad (3 - 6 - 9)$$

Posons pour simplifier:

$$\bar{W}_i = \max_{\mathbb{R}_i} \left[\binom{n}{k}^{1/2} H_n \right], \quad \underline{W}_i = -\min_{\mathbb{R}_i} \left[\binom{n}{k}^{1/2} H_n \right] \quad (3 - 6 - 10)$$

$$\bar{W} = \max_{I_n} \left[\binom{n}{k}^{1/2} H_n \right], \quad \underline{W} = -\min_{I_n} \left[\binom{n}{k}^{1/2} H_n \right] \quad (3 - 6 - 11)$$

(3 - 6 - 12)

par conséquent:

$$\bar{W} = \max_i \bar{W}_i \quad \text{et} \quad \underline{W} = \max_i \underline{W}_i$$

Nous cherchons la loi limite de (\bar{W}, \underline{W}) : elle prendra la forme maintenant habituelle:

$$P(\bar{W} < \theta_1, \underline{W} < \theta_2) \rightarrow \tau(y_1, y_2) \quad (3-6-13)$$

où τ reste à préciser. L'événement

$$\{ \bar{W} < \theta_1, \underline{W} < \theta_2 \} \quad (3-6-14)$$

peut s'écrire:

$$\{ \bigcap_i \{ \bar{W}_i < \theta_1 \}, \bigcap_i \{ \underline{W}_i < \theta_2 \} \} \quad (3-6-15)$$

Appliquons alors la loi (3-6-8): sous les conditions suffisantes locales entre les h_i , n , k et $\bar{\eta}$ on a:

$$\sum_{i=0}^{k-1} (\alpha_{i+1} - \alpha_i) (e^{-(y_1/2)} + e^{-(y_2/2)}) = e^{-(y_1/2)} + e^{-(y_2/2)} \quad (3-6-16)$$

et par conséquent nous avons le théorème 3-4:

Théorème 3 - 4

Soit f une fonction de (H) ; on suppose que dans chaque intervalle $[a_i, a_{i+1}]$ associé à f les fonctions h_i , n , k et $\bar{\eta}$ vérifient les conditions suffisantes (C_i) ; de plus, on suppose que:

$$\frac{k(Lk)^3}{n} \rightarrow 0, \quad \frac{k^3}{nLk} \rightarrow 0 \quad \text{et} \quad \bar{\eta} = o\left(\frac{n}{k}\right)^{1/2}$$

Alors il existe un compact I_n , réunion d'intervalles Ξ_i définis sur chaque $[a_i, a_{i+1}]$ à partir de l'échantillon (X_1, \dots, X_n) , $I_n \rightarrow \text{support}(f)$ quand $n \rightarrow \infty$ et:

$$P\left(\left(\frac{n}{k}\right)^{1/2} \min_{I_n} \left(\frac{f(x) - f_n(x)}{f_n(x)} \right) < \theta_1, \left(\frac{n}{k}\right)^{1/2} \max_{I_n} \left(\frac{f(x) - f_n(x)}{f_n(x)} \right) < \theta_2 \right) \\ \rightarrow \exp \left[-\frac{1}{2\sqrt{\pi}} \left(e^{-(y_1/2)} + e^{-(y_2/2)} \right) \right] \quad \text{lorsque} \quad n \rightarrow \infty$$

Remarque: cette méthode convient parfaitement aux fonctions:

- nulles en a et positives sur $]a, +\infty[$.
- nulles en a et nulles en b , positives sur $]a, b[$
- nulles en a et positives sur $] -\infty, a[$
- positives sur $] -\infty, +\infty[$

et bien entendu:

- positives sur $]a, b[$ et minorées sur ce segment.

dans les cas où f a un zéro "au milieu" du support, on estimera dans un premier temps la position du zéro de f (voir par exemple les travaux de Chevalier [3] sur l'estimation du support d'une densité de probabilité), puis on laissera de côté h_j statistiques ordonnées de part et d'autre de $X_{\lambda_n}^*$, où λ_n est défini par:

$$\lambda_n = F_n(a_n) \quad , \quad a_n \text{ estimant le zéro de } f \quad .$$

Une autre critique que l'on peut formuler à l'égard des théorèmes de ce chapitre est liée au problème des vitesses de convergence;

Si nous mettons en évidence une loi limite sur un intervalle fixé par l'échantillon, nous ignorons a-priori à partir de quel moment cette loi limite est effective (i.e. peut être utilisée concrètement par le statisticien dans un cas pratique) la vitesse de convergence des diverses expressions qui conduisent aux th 3-1, 3-2, 3-3 et 3-4 dépend en particulier de f (par exemple, par la constante de Lipchitz qui intervient dans les calculs); de même, lorsque nous passons de

$$\{ f_1 = o(f_2^\alpha) \text{ pour tout } \alpha \} \quad \tilde{a} \quad \{ Lf_1 = o(Lf_2) \}$$

nous ignorons la vitesse avec laquelle f_1/f_2^α tend vers 0: cela dépend de f .

Une étude des vitesses de convergence vers les lois limites devra donc être pratique, en particulier il sera possible d'établir des étalonnages de la méthode à partir de lois connues.

Une autre approche consiste à chercher à harmoniser h , k et n (en supposant un seul h pour tous les cas), par exemple en posant

$$k = n^{\alpha_1} \quad \text{et} \quad h = n^{\alpha_2} \quad \alpha_1 \text{ et } \alpha_2 \text{ étant fixés} \quad (3 - 6 - 17)$$

alors il faut que les trois conditions suivantes soient réalisées:

$$0 < \alpha_1 < 1 \quad ; \quad 0 < \alpha_2 < 1 \quad ; \quad 3\alpha_1 + 2\alpha_2 > 3$$

ce qui détermine un triangle dans le plan (α_1, α_2) ; A chaque loi P_f correspond un point de ce triangle "optimum" : ce pourrait être l'objet d'une étude concrète.

Remarquons enfin que le nombre des points "laissés de côté" au voisinage de l'infini ou au voisinage d'un zéro de f est toujours infiniment grand, dans le cas examiné ci-dessus, devant le nombre de points d'un bloc; cette règle n'est pas générale cependant pour toutes les fonctions h et k , comme le montre l'exemple suivant:

$$k = \frac{n}{(\ln)^3 L_2 n} \quad , \quad h = (\ln)^5 (L_2 n)^{3/2 + \varepsilon} \quad \text{où } \varepsilon > 0$$

(cet exemple ne doit pas être bon pour la vitesse de convergence !)

Annexe

A) Démonstration de la proposition : 1 - 5

Soient (Z_1, Z_2) , (Z_1', Z_2') quatre statistiques dépendant de n

On suppose que :

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

Alors si :

$$\sqrt{Lk} (Z_i - Z_i') \xrightarrow{p} 0 \quad \text{quand } n \rightarrow \infty \quad (i = 1, 2)$$

$$\lim_{n \rightarrow \infty} P(Z_1' < \theta_1, Z_2' < \theta_2) = \exp \left[-\frac{1}{2\sqrt{\pi}} (e^{-(y_1/2)} + e^{-(y_2/2)}) \right]$$

D : Définissons les évènements

$$E_j : \left\{ \sqrt{Lk} |Z_j - W_j| < \frac{\varepsilon}{\sqrt{2}} \right\} \quad j = 1, 2 \quad (A - 1)$$

$$E_3 = E_1 \cap E_2 \quad \text{donc} \quad \bar{E}_3 = \bar{E}_1 \cup \bar{E}_2$$

$$\text{et } P(\bar{E}_3) \rightarrow 0 \quad \text{quand } n \rightarrow \infty$$

$$\theta_j \text{ désigne toujours } (2Lk - L_2k + y_j)^{1/2} \quad j = 1, 2$$

Soit

$$\theta_j' = 2Lk - \frac{L_2k}{2} + \frac{y_j}{2} \quad j = 1, 2 \quad (A - 2)$$

Alors.

$$\begin{aligned}
 & P(\sqrt{2Lk} Z_1 < \theta_1', \sqrt{2Lk} Z_2 < \theta_2') \\
 & \leq P(\sqrt{2Lk} Z_1 < \theta_1', \sqrt{2Lk} Z_2 < \theta_2' | E_3) P(E_3) + P(\overline{E_3}) \quad (A-3)
 \end{aligned}$$

mais si E_3 est réalisé

$$\{\sqrt{2Lk} Z_i < \theta_i'\} \Rightarrow \{\sqrt{2Lk} Z_i' < \theta_i' + \varepsilon\} \quad (A-4)$$

Donc

$$\begin{aligned}
 & P(\sqrt{2Lk} Z_1 < \theta_1', \sqrt{2Lk} Z_2 < \theta_2') \\
 & \leq P(\sqrt{2Lk} Z_1' < \theta_1' + \varepsilon, \sqrt{2Lk} Z_2' < \theta_2' + \varepsilon) + P(\overline{E_3}) \quad (A-5)
 \end{aligned}$$

Pour n suffisamment grand

$$P(\sqrt{2Lk} Z_1 < \theta_1', \sqrt{2Lk} Z_2 < \theta_2') > \exp\left[-\frac{1}{2\sqrt{\pi}} (e^{-(y_1/2)} + e^{-(y_2/2)})\right] - \varepsilon \quad (A-6)$$

$$P(\overline{E_3}) < \varepsilon \quad (A-7)$$

et donc

$$\begin{aligned}
 & P(\sqrt{2Lk} Z_1' < \theta_1' + \varepsilon, \sqrt{2Lk} Z_2' < \theta_2' + \varepsilon) \quad (A-8) \\
 & \geq \exp\left[-\frac{1}{2\sqrt{\pi}} (e^{-(y_1/2)} + e^{-(y_2/2)})\right] - 2\varepsilon
 \end{aligned}$$

et en changeant

y_j en $y_j - 2\varepsilon$ on a

$$P(\sqrt{2Lk} Z_1' < \theta_1', \sqrt{2Lk} Z_2' < \theta_2') \geq \exp\left[-\frac{1}{2\sqrt{\pi}} (e^{-(y_1/2)+\varepsilon} + e^{-(y_2/2)+\varepsilon})\right] - 2\varepsilon \quad (A-9)$$

Mais de même :

$$\begin{aligned}
 & P(\sqrt{2Lk} Z_1' < \theta_1', \sqrt{2Lk} Z_2' < \theta_2') \\
 & \leq P(\sqrt{2Lk} Z_1' < \theta_1', \sqrt{2Lk} Z_2' < \theta_2' | E_3) + P(\bar{E}_3) \quad (A - 10)
 \end{aligned}$$

et si E_3 est réalisé

$$\{\sqrt{2Lk} Z_i' < \theta_i'\} \implies \{\sqrt{2Lk} Z_i' < \theta_i' + \varepsilon\} \quad (A - 11)$$

donc

$$\begin{aligned}
 & P(\sqrt{2Lk} Z_1' < \theta_1', \sqrt{2Lk} Z_2' < \theta_2' | E_3) \\
 & \leq P(\sqrt{2Lk} Z_1' < \theta_1' + \varepsilon, \sqrt{2Lk} Z_2' < \theta_2' + \varepsilon) \quad (A - 12)
 \end{aligned}$$

Comme pour n suffisamment grand

$$\begin{aligned}
 & P(\sqrt{2Lk} Z_1' < \theta_1' + \varepsilon, \sqrt{2Lk} Z_2' < \theta_2' + \varepsilon) \\
 & < \exp\left[-\frac{1}{2\sqrt{\pi}} \left(e^{-\frac{(y_1/2) - (\varepsilon/2)}{\sqrt{2Lk}}} + e^{-\frac{(y_2/2) - (\varepsilon/2)}{\sqrt{2Lk}}} \right)\right] + \varepsilon \quad (A - 13)
 \end{aligned}$$

$$\begin{aligned}
 & P(\sqrt{2Lk} Z_1' < \theta_1', \sqrt{2Lk} Z_2' < \theta_2') \\
 & < \exp\left[-\frac{1}{2\sqrt{\pi}} \left(e^{-\frac{(y_1/2) - (\varepsilon/2)}{\sqrt{2Lk}}} + e^{-\frac{(y_2/2) - (\varepsilon/2)}{\sqrt{2Lk}}} \right)\right] + 2\varepsilon \quad (A - 14)
 \end{aligned}$$

ε étant arbitraire, le résultat découle de (A - 9) et (A - 14)

BIBLIOGRAPHIE

- [1] ABOU-JAUDE
Thèse d'état, Paris, (1977)
- [2] BERTRAND-RETALI
Convergence uniforme stochastique d'un estimateur d'une densité de probabilité dans \mathbb{R}^S . C.R.Acad. Sci. Paris, t. 27°, série A (4 février 1974) pp. 451-453 .
- [3] CHEVALIER J.
Contribution à l'étude des statistiques d'ordre, de l'estimation géométrique et du maximum de vraisemblance . Thèse d'état, Paris, (1975)
- [4] DIEUDONNE J.
Calcul infinitésimal, Hermann, (1968)
- [5] FELLER W.
An introduction to probability theory and its applications. Wiley, 3ème édition, (1968)
- [6] GEFFRROY J.
Contribution à la théorie des valeurs extrêmes . Thèse d'état, Paris, (1958)
- [7] GEFFRROY J.
Sur l'estimation d'une densité dans un espace métrique . C.R. Acad. Sci. Paris, t. 278, série A (27 mai 1974) pp. 1449-1452
- [8] KENDALL
The Advanced Theory of Statistics, Griffin, (1968), 3ème édition.
- [9] LECOUTRE
Convergence et optimisation de certains estimateurs des densités de probabilité . Thèse 3ème cycle, Paris, (1975)
- [10] MASSEY
A note on the estimation of a distribution function by confidence limits A.M.S. Vol 21 , pp. 116-119
- [11] RENYI
Calcul des Probabilités, Dunod, (1966)
- [12] REVESZ P.
Testing of density functions. Periodica Math. Hung. 1 (1971) pp 35-44 .
- [13] REVESZ P.
Empirical density function . Periodica Math. Hung. 2 (1972) pp 85-110 .
- [14] SMIRNOV
Limit distributions for the terms of a variational series. Amer. Math. Soc. Transl. Ser. 1, n° 11, (1962) pp. 82-143
- [15] TUSNADY G.
Testing density functions. Periodica Math. Hung. 5 (1974) pp 161-169
- [16] WILKS
Mathematical Statistics, Wiley, (1962).
- [17] WOODROOFE
On the maximum deviation of the sample density. Ann. Math. Stat. 38 (1966) pp. 475-481 .
- [18] ZEBULON H.
Contribution à l'étude des convergences stochastiques des mesures aléatoires. Thèse 3ème cycle, Paris, (1975)