STATISTIQUE ET ANALYSE DES DONNÉES

ORESTE NASI

Lois asymptotiques de statistiques de rang sous des alternatives continues ou discrètes

Statistique et analyse des données, tome 1, nº 2 (1976), p. 18-38

<http://www.numdam.org/item?id=SAD_1976__1_2_18_0>

LOIS ASYMPTOTIQUES DE STATISTIQUES DE RANG SOUS DES ALTERNATIVES CONTINUES OU DISCRETES.

Par Oreste NASI
UNIVERSITE DE METZ

Nous traitons de la convergence en loi de suites de statistiques de rang ou de rang randomisées, sous des hypothèses alternatives plus larges que celles de HAJEK et SIDAK pour les échantillons de loi continue et que celles de VORLICKOVA pour les échantillons discrets. Il en résulte des propriétés sur la puissance asymptotique de nouveaux tests de rang, et la construction de tests asymptotiquement les plus puissants pour différents tests d'hypothèses.

Pour simplifier les notations et alleger les démonstrations nous considérons des suites d'échantillons $\{X^N = (X_{N1}, X_{N2}, \dots, X_{NN})\}$ indicées par l'ensemble des entiers $\{N \in \mathbb{N}^*\}$, mais il est aisé de voir que les résultats donnés sont valables pour des suites d'échantillons $\{X^{N_n} = (X_{n1}, \dots, X_{nN_n})\}$ et des suites de statistiques $S_{N_n}(X^{N_n})$ de ces échantillons, où $\{N_n\}$ $n \in \mathbb{N}$, est une suite croissante d'entiers tendant vers $+\infty$.

1. INTRODUCTION

Soit $X^N = (X_{N1}, \dots, X_{NN})$ un échantillon à valeurs dans $\mathbb{R}^N$. On appelle <u>vecteur rang</u> de X^N la statistique $R^N(X^N) = (R_{N1}, \dots, R_{NN})$ où :

$$R_{Ni}(X^N) = \sum_{j=1}^{N} u(X_{Ni} - X_{Nj}) \quad , \quad 1 \leq i \leq N$$

avec $u(x) = 1$ si $x \geq 0$ et $u(x) = 0$ si $x < 0$

Soit $U^N = (U_{N1}, \dots, U_{NN})$ un échantillon indépendant de X^N, constitué de variables mutuellement indépendantes et de loi uniforme sur $]0,1[$. La statistique $R^{*N}(X^N, U^N) = (R^*_{N1}, \dots, R^*_{NN})$ suivante est appelée <u>vecteur de rang randomisé</u> de X^N :

$$R^*_{Ni}(X^N, U^N) = \sum_{j=1}^{N} u^*(X_{Ni} - X_{Nj} \, , \, U_{Ni} - U_{Nj}) \quad , \quad 1 \leq i \leq N$$

où $u^*(x,y) = 1$ si $x > 0$ ou si $X=0$ et $y \geq 0$ et $u^*(x,y)=0$ si $x<0$ ou $x=0$ et $y < 0$.

Considérons les suites de statistiques :

$$(1.1) \qquad S_N(X^N) = \sum_{i=1}^{N} (c_{Ni} - \bar{c}_N) \quad a_N(R_{Ni}) \qquad \forall N \geq 1$$

$$(1.2) \qquad S_N^*(X^N, U^N) = \sum_{i=1}^{N} (c_{Ni} - \bar{c}_N) \quad a_N(R_{Ni}^*)$$

où $c^N = (c_{N1}, \ldots, c_{NN}) \in \mathbb{R}^N$, $\bar{c}_N = \frac{1}{N} \sum_{i=1}^{N} c_{Ni}$, a_N est une application de $\{1,2,\ldots,N\}$ dans $\mathbb{R}$.

Ces statistiques sont appelées respectivement statistiques de rang linéaires simples et statistiques de rang linéaires simples randomisées, suivant la terminologie de HAJEK ([1]) ; les C^N sont appelés coefficients et les a_N "scores" de ces statistiques.

Les c^N vérifient les <u>conditions de NOETHER</u> si et seulement si :

<u>DEFINITION 1</u>

$$\max_{1 \leq i \leq N} (c_{Ni} - \bar{c}_N)^2 > 0 \quad \forall N \geq 1$$

$$\lim_{N \to +\infty} \sum_{i=1}^{N} (c_{Ni} - \bar{c}_N)^2 \Big/ \max_{1 \leq i \leq N} (c_{Ni} - \bar{c}_N)^2 = +\infty$$

<u>DEFINITION 2</u> : Les scores a_N vérifient <u>la condition de convergence par rapport à une fonction</u> φ de $]0,1[$ dans $\mathbb{R}$ si et seulement si :

$$(1.3) \qquad \lim_{N \to +\infty} \int_0^1 \left[a_N(1 + [uN]) - \varphi(u)\right]^2 du = 0 \quad ,$$

où $[x]$ désigne la partie entière de x.

Pour toute application $\varphi :]0,1[\longrightarrow \mathbb{R}$, notons :

$$(1.4) \qquad T_N^{\varphi} = \sum_{i=1}^{N} (c_{Ni} - \bar{c}_N) \quad \varphi(F_{Ni}(X_{Ni})) \qquad \forall N \geq 1$$

où F_{Ni} désigne la fonction de répartition de X_{Ni}. Notons également $L^2(]0,1[)$ l'ensemble $\{\varphi :]0,1[\longrightarrow \mathbb{R} \setminus \int_0^1 \varphi^2(u)\, du < +\infty\}$.

THEOREME 1 : (HAJEK et SIDAK).

Si :

i) les coefficients c^N vérifient les conditions de NOETHER,

ii) les scores a_N vérifient la condition de convergence par rapport à une fonction φ non constante de $L^2(]0,1[)$,

iii) pour tout N, $X_{N1},\ldots, X_{NN}$ sont mutuellement indépendantes et de même fonction de répartition F_N continue,

alors les statistiques (1.1) et (1.4) satisfont à :

$$\frac{S_N - T_N^{\varphi}}{\sigma_N} \xrightarrow{\text{proba}} 0.$$

$$\mathcal{L}\left(\frac{S_N}{\sigma_N}\right) \longrightarrow \mathcal{N}(0,1), \quad \text{avec :}$$

$$\sigma_N^2 = \sum_{i=1}^{N} (c_{Ni} - \bar{c}_N)^2 \int_0^1 (\varphi(u) - \bar{\varphi})^2 \, du \quad , \quad \bar{\varphi} = \int_0^1 \varphi(u)\, du. \tag{1.5}$$

Notations : Soient $\{X_N\}$, $\{Y_N\}$, X et Y des suites de v.a et des v.a.; $\mathcal{L}(X_N) \longrightarrow \mathcal{L}(X)$ signifie que la suite $\{X_N\}$ converge en loi vers X. De même $\mathcal{L}(X_N,Y_N) \longrightarrow \mathcal{L}(X,Y)$ désignera la convergence en loi dans $\mathbb{R}^2$. (m,σ^2) est la loi normale de moyenne m et de variance σ^2 et $\mathcal{N}(m_1,m_2,\sigma_1^2,\sigma_2^2,\sigma_{12})$ la loi normale à deux dimensions de paramètres m_1, m_2, σ_1^2, σ_2^2 et σ_{12}.

La démonstration du théorème se trouve dans [2] (théorème V.1.6.a).

THEOREME 2 : Sous les conditions i) et ii) du théorème 1, et si pour tout N, $X_{N1},\ldots,X_{NN}$ sont mutuellement indépendantes et de même fonction de répartition F_N, alors les statistiques (1.2) vérifient : $\mathcal{L}\left(\frac{S_N}{\sigma_N}\right) \longrightarrow \mathcal{N}(0,1)$.

démonstration du théorème 29.A de [1] il résulte que R^{*N} est de même loi que le vecteur de rang R^N quand celui-ci est le rang d'un échantillon dont les X_{Ni} sont de loi continue. Il suffit donc d'appliquer le théorème 1.

2. QUANTITE D'INFORMATION ET FONCTION D'UNE FAMILLE DE DENSITES.

Soit une famille de densités $\{f(x,\theta)\}$ ainsi définie :

(2.1) $\theta \in I$, intervalle ouvert de $\mathbb{R}$ contenant 0.

(2.2) $\forall\, \theta \in I$, $f(x,\theta)$ [notée parfois $f_\theta(x)$] est une densité par rapport à une même mesure μ sur $(\mathbb{R}, \mathcal{B})$.

Par $f'(x,\theta)$ ou $f'_\theta(x)$ nous désignerons la dérivée partielle en θ de $f(x,\theta)$:

$$f'(x,\theta) = \lim_{u \to 0} \frac{f(x,\theta+u) - f(x,\theta)}{u}$$

DEFINITION 1 : pour tout θ de I, on appelle quantité d'information de la densité $f(x,\theta)$ l'expression (quand elle est définie) :

$$I_f(\theta) = \int_{\mathbb{R}} \left[\frac{f'(x,\theta)}{f(x,\theta)}\right]^2 f(x,\theta)\, d\mu\,(x).$$

Notons par $F(x,\theta)$, (ou $F_\theta(x)$) la fonction de répartition de $f(x,\theta)$:

$$F(x,\theta) = \int_{]-\infty,x]} f(t,\theta)\, d\mu(t)$$ et par $F^{-1}(u,\theta)$ (ou $F_\theta^{-1}(u)$) sa fonction "inverse" :

$$F^{-1}(u,\theta) = \inf\{x \setminus F(x,\theta) \geq u\} \quad , \forall\, u \in\,]0,1[.$$

DEFINITION 2 : on appelle fonction Ψ_f de HAJEK-SIDAK de la densité $f(x,\theta)$ l'application (quand elle est définie) de $]0,1[$ dans $\mathbb{R}$ donnée par :

$$\Psi_f(u,\theta) = \frac{f'_\theta(F_\theta^{-1}(u))}{f_\theta\,(F_\theta^{-1}(u))} \qquad \forall\, u \in\,]0,1[.$$

exemple : soit $\mathcal{C}$ un ensemble discret de $\mathbb{R}$ dont les éléments sont rangés par ordre croissant :

(2.4) $\mathcal{C} = \{x_h \,,\, h \in H \subset \mathbb{Z}\}$

(2.5) $h > h' \Rightarrow x_h > x_{h'} \quad \forall (h,h') \in H^2.$

Soit μ la mesure de dénombrement sur : $\forall B \in \mathcal{B}$, $\mu(B) = \mathrm{Card}(B \cap \mathcal{C})$.

La famille de densités définie par (2.1) et (2.2) vérifie alors :

$$(2.6) \qquad I_f(\theta) = \sum_{h \in H_\theta} \frac{[f'(x_h,\theta)]^2}{f(x_h,\theta)} \qquad \forall \theta \in I ,$$

où $H_\theta = \{h \in H \;/\; f(x_h,\theta) > 0\}$. Pour tout h de H_θ, notons $I_h^\theta =]F(x_{h-1},\theta), F(x_h,\theta)]$, avec éventuellement $I_{h_0}^\theta =]0, F(xh_0,\theta)]$ et $I_{h_1}^\theta =]F(x_{h_1-1},\theta), 1[$ si H_θ possède un plus petit élément h_0 ou un plus grand élément h_1 $\{I_h^\theta\}$, $h \in H_\theta$, forme une partition de $]0,1[$ et :

$$F^{-1}(u,\theta) = x_h \quad \forall u \in I_h^\theta , \quad \forall h \in H_\theta, \text{ d'où}$$

$$(2.7) \qquad \Psi_f(u,\theta) = \frac{f'(x_h,\theta)}{f(x_h,\theta)} \qquad \forall\, u \in I_h^\theta ,\ \forall\, h \in H_\theta$$

<u>notations</u> dans la suite <u>λ désignera la mesure de Lebesgue</u> sur $\mathbb{R}$ et <u>$\mu_{\mathcal{C}}$ la mesure de dénombrement</u> par rapport à un ensemble $\mathcal{C}$ discret de $\mathbb{R}$.

<u>PROPOSITION 1</u> : Supposons la famille $\{f(x,\theta)\}$ définie par rapport à la mesure λ ou à la mesure $\mu_{\mathcal{C}}$, où $\mathcal{C}$ satisfait à (2.4) et (2.5). Soit $\theta \in I$ tel que :

i) $f'(x,\theta)$ existe μ-presque sûrement pour tout x,

ii) $I_f(\theta) < +\infty$,

alors :
$$\int_0^1 \Psi_f^2(u,\theta)\, du = I_f(\theta)$$

<u>Démonstration</u> : Si $\mu = \lambda$, considérons une v.a. $X_\theta : (\Omega,\mathcal{C},P) \to (\mathbb{R},\mathcal{B})$ de densité $f(x,\theta)$ par rapport à μ. Sa fonction de répartition $F_\theta(x)$ est continue et $F_\theta(X_\theta)$ est de loi uniforme sur $]0,1[$. En utilisant le théorème du transfert on a :

$$\int_0^1 \Psi^2{}_f(u,\theta)\, du = \int_\Omega \left[\frac{f'_\theta(F_\theta^{-1}[F_\theta(X_\theta)])}{f_\theta(F_\theta^{-1}[F_\theta(X_\theta)])}\right]^2 dp$$

$$= \int_\Omega \left[\frac{f'_\theta(X_\theta)}{f_\theta(X_\theta)}\right]^2 dp = \int_{\mathbb{R}} \left[\frac{f'_\theta(x)}{f_\theta(x)}\right]^2 f_\theta(x)\, dx = I_f(\theta)$$

Si $\mu = \mu_{\mathcal{C}}$, la conclusion de la proposition découle aisément de (2.7) et (2.6).

PROPOSITION 2 :

Soit $\{f(x,\theta)\}$ définie par rapport à λ ou à $\mu_{\mathcal{C}}$ comme pour la propositic 1, et θ tel que :

i) il existe un voisinage V de θ tel que $f'(x,t)$ existe pour tout t de B et u-presque tout x.

ii) il existe un voisinage V de θ où $G(t) = \int_{\mathbb{R}} |f'(x,t)|\, d\mu(x)$ est définie et continue en t alors

$$\int_0^1 \psi_f(u,\theta)\, du = 0.$$

démonstration : si $\mu = \lambda$, on a, en utilisant la v.a X_θ définie plus haut :

$$\int_0^1 \psi_f(u,\theta)\, du = \int_\Omega \frac{f'_\theta(F_\theta^{-1}[F_\theta(X_\theta)])}{f_\theta(F_\theta^{-1}[F_\theta(X_\theta)])}\, dP = \int \frac{f'_\theta(X_\theta)}{f_\theta(X_\theta)}\, dP,$$

d'où :

$$\int_0^1 \psi_f(u,\theta)\, du = \int_{\mathbb{R}} f'_\theta(x)\, d\mu(x). \tag{2.8}$$

Cette dernière égalité se vérifie aisément également dans le cas où $\mu=\mu_{\mathcal{C}}$ d'après (2.7). Posons, pour une suite $\{t_n\}$, $n \geq 1$, de réels strictement positifs tendant vers 0 quand $n \longrightarrow +\infty$,

$$h_n(x) = \frac{f(x,\theta+t_n) - f(x,\theta)}{t_n} \quad \forall\, n \geq 1 \text{ , et}$$

$$h(x) = f'(x,\theta).$$

D'après l'hypothèse i), $\lim_{n \to +\infty} h_n = h$ pour μ-presque tout x ; d'autre part :

$$\int_{\mathbb{R}} |h_n(x)|\, d\mu(x) = \frac{1}{t_n} \int_{\mathbb{R}} |f(x,\theta + t_n) - f(x,\theta)|\, d\mu(x)$$

$$\leq \frac{1}{t_n} \int_{\theta}^{\theta+t_n} \int_{\mathbb{R}} |f'(x,t)|\, d\mu(x)\,dt = \int_{\mathbb{R}} |f'(x,\theta_n)|\, d\mu(x) \quad , \quad \theta < \theta_n < \theta + t_n,$$

en vertu de l'hypothèse ii) et du théorème de la moyenne. Par conséquent :

$$\lim_{n \to +\infty} \int_{\mathbb{R}} |h_n(x)|\, d\mu(x) \leq \int_{\mathbb{R}} |h(x)|\, d\mu(x).$$

Le théorème de convergence II.4.2 de [2] donne donc :

$$\lim_{n \to +\infty} \int_{\mathbb{R}} h_n(x)\, d\mu\,(x) = \int_{\mathbb{R}} h(x)\, d\mu(x) = \int_{\mathbb{R}} f'_{\theta}(x)\, d\mu(x)$$

La partie gauche de cette égalité est nulle car $f(x,\theta + t_n)$ et $f(x,\theta)$ sont des densités. Ce qui prouve, avec (2.8), la proposition.

<u>Remarque</u> : les propositions 1 et 2 généralisent les lemmes I.2.4.d , I.2.4.a et b de [2] qui ne concernent que les familles $f(x,\theta) = g(x-\theta)$ et $f(x,\theta) = e^{-\theta}g(xe^{-\theta})$ où g est une densité par rapport à la mesure de Lebesgue.

3. CONVERGENCE DES STATISTIQUES DE RANG SOUS DES ALTERNATIVES CONTINUES

Nous allons utiliser dans la suite les conditions suivantes sur la famille $\{f(x,\theta)\}$:

(C_1) Si E_θ désigne l'ensemble $\{x \in \mathbb{R} \setminus f(x,\theta) > 0\}$, $\forall \theta \in I$,
$\exists E \subset \mathbb{R} \setminus \forall \theta \in I$, $E \subset E_\theta$ et $\mu(E_\theta - E) = 0$.

Il existe un voisinage V de 0 tel que :

(C_2) $f'(x,\theta)$ existe et est continue en θ sur V pour tout x de E

(C_3) $I_f(\theta)$ est une fonction de θ strictement positive et continue sur V

(C_4) $\int_E |f'(x,\theta)|\, d\mu(x)$ est une fonction de θ continue sur V;

(C_5) Pour toutes suites $\{v_N\}$ et $\{t_N\}$ telles que :

(3.1) $$\lim_{N \to +\infty} v_N = +\infty \quad , \quad \lim_{N \to +\infty} t_N = 0 \quad , \text{ on a :}$$

$$\lim_{N \to +\infty} \int_{\left\{x \in E \setminus \left|\frac{f'(x,t_N)}{f(x,t_N)}\right| > v_N\right\}} \left(\frac{f'(x,t_N)}{f(x,t_N)}\right)^2 d\mu(x) = 0$$

Soit la suite de densités :

(3.2) $$q_N(x^N) = \prod_{i=1}^{N} f(x_{Ni}, t_{Ni}) \quad \forall x^N = (x_{N1}, \ldots, x_{NN}) \in \mathbb{R}^N, \ \forall N \geq$$

où $t^N = (t_{N1}, \ldots, t_{NN}) \in \mathbb{R}^N$ pour tout N et :

(3.3) $$\lim_{N \to +\infty} \max_{1 \leq i \leq N} t_{Ni}^2 = 0$$

(3.4) $$\lim_{N \to +\infty} I_f(0) \sum_{i=1}^{N} (t_{Ni} - \bar{t}_N)^2 = b^2 \quad \text{avec} \quad 0 < b < +\infty \ , \bar{t}_N = \frac{1}{N} \sum_{i=1}^{N} t_N$$

Considérons les statistiques de rang S_N données par (1.1).

<u>THEOREME</u> : Si

i) la famille $\{f(x,\theta)\}$ définie par rapport à la mesure λ de Lebesgue vérifie (C_1) à (C_5),

ii) les coefficients c^N satisfont aux conditions de Noether,

iii) les scores a_N satisfont à la condition de convergence par rapport à une fonction φ non constante de $L^2(]0,1[)$,

iv) $\lim_{\theta \to 0} \int_0^1 \varphi(u) \ \psi_f(u,\theta)\, du = \int_0^1 \varphi(u) \ \psi_f(u,0)\, du$

alors, sous les densités q_N données en (3.2) où les t^N vérifient (3.3) et (3.4), on a :

$$\mathcal{L}\left(\frac{S_N - \mu_N}{\sigma_N}\right) \to \mathcal{N}(0,1), \quad \text{où}$$

(3.5) $$\mu_N = \sum_{i=1}^{N} (C_{Ni} - \bar{C}_N)(t_{Ni} - \bar{t}_N) \int_0^1 \varphi(u) \ \psi_f(u,0)\, du \ , \text{ et}$$

σ_N^2 est donné par (1.5).

Démonstration : Nous la donnons plus bas au paragraphe 5. Par "sous q_N" nous entendons que les échantillons X^N sont de densités q_N.

Remarques : Ce théorème étend le théorème analogue de [2] qui concerne $f(x,\theta) = g(x,\theta)$ et $f(x,\theta) = e^{-\theta} g(x\, e^{-\theta})$ où g est une densité par rapport à la mesure de Lebesgue ([2], théorème VI.2.4).

Dans [3] nous donnons une condition (C'_5) assez simple à vérifier pour de nombreuses familles de densités et qui implique (C_5).

4. CONVERGENCE DES STATISTIQUES DE RANG RANDOMISEES SOUS DES ALTERNATIVES DISCRETES

Supposons maintenant que la famille $\{f(x,\theta)\}$ est définie par rapport à la mesure $\mu_{\mathcal{C}}$ sur un ensemble discret $\mathcal{C}$ qui satisfait à (2.4),(2.5) et à :

(4.1) $$\inf_{\substack{h,h' \in H \\ h \neq h'}} \{|x_h - x_{h'}|\} = \ell > 0.$$

Pour simplifier, supposons également que $\{f(x,\theta)\}$ satisfait à :

(C'_1) $$\forall\, \theta \in I,\ \forall\, h \in H,\ f(x_h,\theta) > 0 \quad \text{et} \quad \sum_{h \in H} f(x_h,\theta) = 1.$$

On vérifie aisément que (C'_1) implique (C_1) avec $E = \mathcal{C}$

Considérons les statistiques randomisées S_N données en (1.2).

THEOREME. Si

a) la famille $\{f(x,\theta)\}$ définie par rapport à la mesure $\mu_{\mathcal{C}}$ satisfait à (C'_1) et (C_2) à (C_5),

b) les hypothèses ii) à iv) du théorème du paragraphe 4 sont vérifiées, alors les densités q_N données en (3.2) où les t^N vérifient (3.3) et (3.4), on a :

$$\mathcal{L}\left(\frac{S^*_N - u_N}{\sigma_N}\right) \longrightarrow \mathcal{N}(0,1)$$

Démonstration : elle est donnée au paragraphe 5.

Remarques : Ce théorème étend celui de Vorlickova ([4], théorème 4.2) où au lieu de (3.4) il est supposé que $\lim\limits_{N \to +\infty} I_f(0) \sum\limits_{i=1}^{N} t^2_{Ni} = b^2 (0 < b < +\infty)$. Cette condition est plus forte que (3.4) et de plus ne permet d'obtenir des tests asymptotiquement les plus puissants (voir paragraphe 6) que pour des alternatives q_N telles que $\bar{t}_N=0$ pour tout N, au lieu de $\lim_{N\to+\infty} \bar{t}_N = 0$ comme le permet notre résultat.

Quand X^N admet la densité q_N par rapport à $\mu\mathscr{C}$, on a :

$$p\{X_{Ni} = x_h\} = f(x_h, t_{Ni}) \quad \forall\, i=1,2,\dots,N \quad , \quad \forall\, h \in H.$$

Par ailleurs on peut écrire également μ_N sous la forme :

$$\mu_N = \sum_{i=1}^{N} (c_{Ni} - \bar{c}_N)(t_{Ni} - \bar{t}_N) \sum_{h\in H} \frac{f'(x_h,0)}{f(x_h,0)} \int_{I_h} \varphi(u)\, du \ ,$$

où $I_h = \left]\sum_{k\leq h-1} f(x_k,0) , \sum_{k\leq h} f(x_k,0)\right]$, $\forall\, h \in H$.

Dans [4] nous montrons que si $\mathscr{C}$ est un ensemble fini, (C_1),(C_2) et (C_3) impliquent (C_5).

5. DEMONSTRATIONS DES THEOREMES

Soient les densités et les statistiques :

(5.1) $$\bar{p}_N(x^N) = \prod_{i=1}^{N} f(x_{Ni}, \bar{t}_N) \quad , \quad \forall\, N \geq 1$$

(5.2) $$T_N(X^N) = \sum_{i=1}^{N} (t_{Ni} - \bar{t}_N) \frac{f'(X_{Ni},\bar{t}_N)}{f(X_{Ni},\bar{t}_N)} \quad , \quad \forall\, N \geq 1$$

Lemme 1 : Si la famille $\{f(x,\theta)\}$ est définie par rapport à λ ou à $\mu\mathscr{C}$, ($\mathscr{C}$ satisfaisant à 4.1) et vérifie les conditions (C_1) à (C_5), et si les t^N satisfont à (3.3) et (3.4), alors

$$\mathscr{L}(T_N) \longrightarrow \mathcal{N}(0,b^2) \quad \text{sous } \bar{p}_N.$$

démonstration : pour alleger les notations, nous négligerons d'écrire un certain nombre d'indices N. Les variables :

$$Y_{Ni} = (t_{Ni}-\bar{t}_N) \frac{f'(X_{Ni},\bar{t}_N)}{f(X_{Ni},\bar{t}_N)} \qquad N \geq 1 \quad , \quad 1 \leq i \leq N$$

sont centrées dès que N est suffisemment grand. Cela résulte de la proposition 2 du paragraphe 2 et de (3.3) qui implique $\lim\limits_{N\to+\infty} \bar{t}_N = 0$:

$$E(Y_i)=(t_i-\bar{t})\int_{\mathbb{R}} f'(x,t_N)\,d\mu(x) = (t_i-\bar{t})\int_{\mathbb{R}} \psi f(u,\bar{t}_N)\,du = 0.$$

$$\Longrightarrow \mathrm{Var}(Y_i) = (t_i-\bar{t})^2\int_E \frac{[f'(x,t_N)]^2}{f(x,\bar{t}_N)}\,d\mu(x) = (t_i-\bar{t})\ I_f(\bar{t}_N) < +\infty$$

Posons :
$$s_{Ni}^2 = \mathrm{Var}\,(Y_{Ni}) = (t_{Ni}-\bar{t}_N)^2\ I_f(\bar{t}_N)$$
$$s_N^2 = \sum_{i=1}^N s_{Ni}^2 = \sum_{i=1}^N (t_{Ni}-\bar{t}_N)^2\ I_f(\bar{t}_N)$$

$$A_N(\varepsilon) = \frac{1}{s_N^2}\sum_{i=1}^N \int_{\{|x|>\varepsilon s_N\}} x^2\,dF_{Y_{Ni}}(x) = \frac{1}{s_N^2}\sum_{i=1}^N (t_i-\bar{t})^2\int_{\{|x|>\varepsilon \frac{s_N}{|t_i-\bar{t}|}\}} x^2\,dF_{G_N}(x) \quad , \forall\ \varepsilon > 0$$

où F_{G_N} est la fonction de répartition commune aux $G_N(X_{Ni}) = \dfrac{f'(X_{Ni},\bar{t}_N)}{f(X_{Ni},\bar{t}_N)}$. On a

$$\frac{s_N^2}{(t_i-\bar{t})^2} \geqslant \frac{I_f(\bar{t}_N)\sum_{i=1}^N (t_i-\bar{t})^2}{\max\limits_{1\leqslant i\leqslant N}(t_i-\bar{t})^2} = m_N^2$$

avec $\lim\limits_{N\to+\infty} m_N^2 = +\infty$, d'après (3.3), (3.4) et (C_3). D'où, sous $\bar{p}_N$:

$$A_N(\varepsilon) \leqslant \frac{1}{s_N^2}\sum_{i=1}^N (t_i-\bar{t})^2\int_{\{|x|>\varepsilon|m_N|\}} x^2\,dF_{G_N}(x)$$

$$= \frac{1}{I_f(\bar{t}_N)}\int_{\{x\in E/|\frac{f'(x,\bar{t}_N)}{f(x,\bar{t}_N)}|>\varepsilon|m_N|\}} \frac{[f'(x,\bar{t}_N)]^2}{f(x,\bar{t}_N)}\,d\mu(x).$$

Les conditions (C_3) et (C_5) impliquent $\lim_{N \to +\infty} A_N(\varepsilon) = 0$. Il résulte du théorème de la limite centrale de Lindeberg que :

$$\mathcal{L}\left(\frac{\sum_{i=1}^{N} Y_{Ni}}{s_N}\right) \longrightarrow \eta(0,1).$$

Comme $\lim_{N \to +\infty} s_N = b$ d'après (3.4) et $\sum_{i=1}^{N} Y_{Ni} = T_N$, le lemme est démontré.

Soit la suite de statistiques :

$$W_N(x^N) = 2 \sum_{i=1}^{N} \left\{ \left[\frac{f(X_{Ni}, t_{Ni})}{f(X_{Ni}, \bar{t}_N)}\right]^{1/2} - 1 \right\} \quad \forall N \geq 1.$$

Lemme 2 : si les hypothèses du lemme 1 sont vérifiées, on an sous $\bar{p}_N$:

1. $\lim_{N \to +\infty} E(W_N) = -\frac{b^2}{4}$

2. $\lim_{N \to +\infty} \mathrm{Var}(W_N - T_N) = 0.$

Démonstration : 1. Sous $\bar{p}_N$, et toujours en négligeant un certain nombre d'indices N

$$E(W_N) = 2 \sum_{i=1}^{N} \int_{\mathbb{R}} \frac{f^{1/2}(x,t_i) - f^{1/2}(x,\bar{t})}{f^{1/2}(x,\bar{t})} f(x,\bar{t})\, d\mu(x)$$

$$= \sum_{i=1}^{N} \int_E \left[f(x,t_i) - f(x,\bar{t}) - \left\{ f^{1/2}(x,t_i) - f^{1/2}(x,\bar{t}) \right\}^2 \right] d\mu(x)$$

$$= -\sum_{i=1}^{N} (t_i - \bar{t})^2 \int_E \left[\frac{f^{1/2}(x,t_i) - f^{1/2}(x,\bar{t})}{t_i - \bar{t}} \right]^2 d\mu(x) \tag{5.3}$$

Soit pour une suite $\{(\theta_N, \theta'_N)\}$ de $\mathbb{R}^2$ telle que $\theta_N \neq \theta'_N$ pour tout N et $\lim_{N \to +\infty} (\theta_N, \theta'_N) = (0,0)$,

$$h_N(x) = \frac{f^{1/2}(x,\theta_N) - f^{1/2}(x,\theta'_N)}{\theta_N - \theta'_N} \quad \forall x \in E, \forall N \geq 1$$

$$h(x) = \frac{1}{2} \frac{f'(x,0)}{f^{1/2}(x,0)}$$

On a :

$$(5.4)\qquad \lim_{N\to+\infty} h_N(x) = h(x) \quad \text{pour tout } x \text{ de } E \text{ et :}$$

$$\int_E h_N^2(x)\, d\mu(x) \leqslant \frac{1}{4(\theta_N-\theta'_N)}\int_{\theta'_N}^{\theta_N}\int_E \frac{[f'(x,t)]^2}{f(x,t)}\, d\mu(x)\, dt$$

$$= \frac{1}{4(\theta_N-\theta'_N)}\int_{\theta'_N}^{\theta_N} I_f(t)\, dt.$$

Comme $I_f(\theta)$ est continue au voisinage de 0, il en résulte :

$$(5.5)\qquad \lim_{N\to+\infty}\int_E h_N^2(x)\, d\mu(x) \leqslant \frac{1}{4} I_f(0) = \int_E h^2(x)\, d\mu(x),$$

d'où d'après le théorème II.4.2 de [2]:

$$\lim_{(\theta,\theta')\to(0,0)}\int_E\left[\frac{f^{1/2}(x,\theta)-f^{1/2}(x,\theta')}{\theta-\theta'}\right]^2 d\mu(x) = \frac{1}{4} I_f(0)$$

$$\Longrightarrow \lim_{N\to+\infty}\max_{1\leqslant i\leqslant N}\int_E \frac{[f^{1/2}(x,t_i)-f^{1/2}(x,\bar t)]^2}{t_i-\bar t}\, d\mu(x) = \frac{1}{4} I_f(0),$$

d'après (3.3). Ce qui reporté dans (5.3) donne la limite de $E(W_N)$ si on tient compte de (3.4).

2. Quelques calculs donnent :

$$\mathrm{Var}(W_N-T_N) \leqslant 4\sum_{i=1}^{N}\int_E\left[\frac{f^{1/2}(x,t_i)}{f^{1/2}(x,\bar t)} - 1 - \frac{t_i-\bar t}{2}\,\frac{f'(x,\bar t)}{f(x,\bar t)}\right]^2 d\mu(x)$$

$$\leqslant 8\sum_{i=1}^{N}(t_i-\bar t)^2\left[A_{Ni} + \frac{1}{4}B_N\right]$$

où $$A_{Ni} = \int_E\left[\frac{f^{1/2}(x,t_{Ni})-f^{1/2}(x,\bar t_N)}{t_{Ni}-\bar t_N} - \frac{1}{2}\,\frac{f'(x,0)}{f^{1/2}(x,0)}\right]^2 d\mu(x)$$

$$B_N = \int_E\left[\frac{f'(x,0)}{f^{1/2}(x,0)} - \frac{f'(x,\bar t_N)}{f^{1/2}(x,\bar t_N)}\right]^2 d\mu(x).$$

Les relations (5.4) et (5.5) impliquent, d'après le théorème V.1.3 de [2] :

$$\lim_{\substack{(\theta,\theta')\to(0,0)\\ \theta\neq\theta'}} \int_E \left[\frac{f^{1/2}(x,\theta) - f^{1/2}(x,\theta')}{\theta-\theta'} - \frac{1}{2}\,\frac{f'(x,0)}{f^{1/2}(x,0)}\right]^2 d\mu(x) = 0$$

$$\Longrightarrow \lim_{N\to+\infty} \max_{1\leq i\leq N} A_{Ni} = 0 \text{ , d'après (3.3)}$$

D'autre part, en posant :

$$h_N(x) = \frac{f'(x,\bar{t}_N)}{f^{1/2}(x,\bar{t}_N)} \;,\quad \forall N\geq 1 \quad \text{et} \quad h(x) = \frac{f'(x,0)}{f^{1/2}(x,0)} \;;$$

on montre aisément que $\lim_{N\to+\infty} h_N(x) = h(x)$ pour tout x de E et

$$\lim_{N\to+\infty} \int_E h_N^2(x)\,d\mu(x) = \int_E h^2(x)\,d\mu(x)$$

En utilisant à nouveau le théorème V.1.3 de [2], on en déduit que $\lim_{N\to+\infty} B_N=0$ q.f.d.

Soit la suite de statistiques :

$$L_N(X^N) = \begin{cases} \dfrac{q_N(X^N)}{\bar{p}_N(X^N)} & \text{si } \bar{p}_N > 0 \\[2ex] 1 & \text{si } \bar{p}_N = q_N = 0 \\ +\infty & \text{si } \bar{p}_N = 0 < q_N \end{cases}$$

<u>Lemme 3</u> : sous les hypothèses du lemme 1 et sous $\bar{p}_N$, on a :

$$\mathrm{Log}\, L_N(X^N) - T_N(X^N) + \frac{b^2}{2} \xrightarrow{\text{proba}} 0$$

$$\mathcal{L}(\log L_N) \longrightarrow \eta\left(-\frac{b^2}{2}, b^2\right)$$

<u>démonstration</u> : les T_N sont centrées car $X_{N-1},\dots,X_{NN}$ sont équidistribuées et $\sum_{i=1}^{N} (t_{ni}-\bar{t}_N) = 0$. D'où d'après le lemme 2,

$$W_N - T_N + \frac{b^2}{4} \xrightarrow{\text{proba}} 0 \;, \tag{5.6}$$

et d'après le lemme 1, $\mathcal{L}(W_N) \longrightarrow \eta\left(-\frac{b^2}{4}, b^2\right)$.

Il résulte du 2ème lemme de Le Cam ([2] , lemme VI.1.3) que, sous $\bar{p}_N$,

$$\mathrm{Log}\, L_N - W_N + \frac{b^2}{4} \xrightarrow{\text{proba}} 0 \tag{5.7}$$

et $\mathcal{L}(\text{Log } L_N) \to \eta(-\frac{b^2}{2}, b^2)$. Les relations (5.6) et (5.7) donnent $\log L_N - T_N + \frac{b^2}{2} \xrightarrow{\text{proba}} 0$.

Démonstration du théorème du paragraphe 3.

Il suffit de montrer le théorème pour des C_{Ni} et t_{Ni} vérifiant $\sum_{i=1}^{N}(c_{Ni}-\bar{c}_N)=1$ et

$$\lim_{N\to+\infty} \sum_{i=1}^{N}(c_{Ni}-\bar{c}_N)(t_{Ni}-\bar{t}_N) = b_{12} < +\infty \tag{5.8}$$

donc de montrer que, sous q_N, $\mathcal{L}(\frac{S_N-m}{\sigma}) \longrightarrow \eta(0,1)$ avec

$$m = b_{12}\int_0^1 \varphi(u)\, \psi_f(u,0)\,du \quad \text{et } \sigma^2 = \int_0^1 (\varphi(u)-\bar{\varphi})^2\, d\mu. \tag{5.9}$$

Les variables T_N^{φ} et T_N données par (1.4) et (5.2) sont centrées sous $\bar{p}_N$. De plus $\text{Var}(T_N^{\varphi}) = \sigma^2$ et

$$\text{Var}(T_N) = \sum_{i=1}^{N} (t_i-\bar{t})^2 \quad E\left\{\left[\frac{f'(X_i,\bar{t})}{f(X_i,t)}\right]^2\right\}$$

car $E\left\{\frac{f'(X_i,\bar{t}_N)}{f(X_i,\bar{t}_N)}\right\} = \int_E f'(x,\bar{t}_N)\,d\mu(x) = 0$ dès que N est assez grand, d'après la proposition 2 du paragraphe 2. Des relations (3.3), (3.4) et de la condition (C_3) il résulte donc :

$$\lim_{N\to+\infty} \text{Var}(T_n) = \lim_{N\to+\infty} \sum_{i=1}^{N} (t_{Ni} - \bar{t}_N)^2\, I_f(\bar{t}_N) = b^2.$$

d'autre part :

$$E(T_N^{\varphi}T_N) = \sum_{i=1}^{N} (c_i-\bar{c})(t_i-\bar{t}) \; E\left\{\varphi(F_{Ni}(X_{Ni}))\, \frac{f'(X_{Ni},\bar{t}_N)}{f(X_{Ni},\bar{t}_N)}\right\}$$

$$= \sum_{i=1}^{N} (c_i-\bar{c})(t_i-\bar{t}) \int_0^1 \varphi(u)\, \frac{f'(F_N^{-1}(u),\bar{t}_N)}{f(F_N^{-1}(u),\bar{t}_N)}\,du$$

$$= \sum_{i=1}^{N} (c_i-\bar{c})(t_i-\bar{t}) \int_0^1 \varphi(u)\, \psi_f(u,\bar{t}_N)\,du$$

$$\lim_{N\to+\infty} E(T_N^{\varphi}T_N) = b_{12}\int_0^1 \varphi(u)\, \psi_f(u,0)\,du = m\,,$$

d'après l'hypothèse iv) du théorème et les relations (5.8) et (5.9). Donc

$$\lim_{N \to +\infty} \mathrm{Var}(u\, T_N^{\varphi} + vT_N) = u\,\sigma^2 + v^2 b^2 + uvm \quad \forall (u,v) \in \mathbb{R}^2.$$

Il en résulte (cf; démonstration du théorème VI.2.4 de [2]) que $\mathcal{L}(T_N^{\varphi}, T_N) \longrightarrow \eta(0,0,\sigma^2,b^2,m)$ sous $\bar{p}_N$. Mais le théorème 1 du paragraphe 1 et le lemme 3 donnent $\frac{S_N - T_N}{\sigma} \xrightarrow{\text{proba}} 0$ et $\mathrm{Log}\, L_N - T_N + \frac{b^2}{2} \xrightarrow{\text{proba}} 0$. D'où :

$$\mathcal{L}(S_N, \mathrm{Log}\, L_N) \longrightarrow \eta(0, -\frac{b^2}{2}, \sigma^2, b^2, m) \quad \text{sous } \bar{p}_N.$$

En utilisant le 3ème lemme de Le Cam ([2],Lemme VI.1.4) on en déduit que <u>sous les densités q_N, $\mathcal{L}(S_N) \longrightarrow \eta(m,\sigma^2)$</u>, ce qui prouve le théorème.

<u>Démonstration du théorème du paragraphe 4.</u>

Il est aisé de voir que , sous l'hypothèse (4.1) on a $R^{*N}(X^N,U^N) = R'^N(X^N,U^N)$ presque sûrement, où $R'^N = (R'_{N1},\ldots,R'_{NN})$ est le vecteur rang de $X'^N = (X'_{N1},\ldots,X'_{NN})$ ainsi défini :

$$X'_{Ni} = X_{Ni} + lU_{Ni} \qquad \forall N \geq 1 \;,\; \forall i = 1,2,\ldots,N$$

Il suffit donc de démontrer le théorème pour les variables $S'_N = \sum_{i=1}^{N} (c_{Ni} - \bar{c}_N)\, a_N(R'_{Ni})$. La démonstration en est pas à pas identique à celle du théorème du cas continu, en remplaçant T_N^{φ} par :

$$T'^{\varphi}_N = \sum_{i=1}^{N} (c_{Ni} - \bar{c}_N)\, \varphi(G_{Ni}(X'_{Ni})) \;,$$

où G_{Ni} est la fonction de répartition (continue) de X'_{Ni}. Sous $\bar{p}_N$ on a $G_{Ni} = G_{N2} = \ldots G_{NN} = G_N$, donc :

$$E(T'^{\varphi}_N T_N) = \sum_{i=1}^{N} (c_i - \bar{c})(t_i - \bar{t})\; E\left\{ \varphi(G_N(X_i + l\, U_i))\, \frac{f'(X_i,\bar{t})}{f(X_i,\bar{t})} \right\}$$

$$= \sum_{i=1}^{N} (c_i - \bar{c})(t_i - \bar{t}) \int_0^{l} \int_{\mathcal{C}} \varphi(G_N(u+v))\, f'(u,\bar{t})\, \frac{1}{l}\, d\mu(u)\; dv$$

$$= \sum_{i=1}^{N} (c_i - \bar{c})(t_i - \bar{t}) \sum_{h \in H} f'(x_h,\bar{t}) \int_0^{l} \frac{\varphi(G_N(x_h+v))}{l}\, dv \;,$$

compte tenu de l'indépendance de X_{Ni} et $l\, U_{Ni}$ qui donne également :

$$G_N(u) = \sum_{j / x_j \leq u-l} f(x_j,\bar{t}) + \int_{u-l}^{u} \frac{u-y}{l}\, f(y,\bar{t})\, d\mu(y)$$

$$\Rightarrow \forall v,\ G_N(x_h + v-\ell,\ \bar{t}) + \int_{x_h+v-1}^{x_h+v} \frac{x_h+v-y}{\ell}\, f(y,\bar{t})\, d\mu(y)$$

$$\Rightarrow \forall\, v \in \left]0,\ell\right[,\quad G_N(x_h+v) = F(x_{h-1},\ \bar{t}) + \frac{v}{1}\, f(x_h,\bar{t})$$

En reportant dans (5.10) on obtient :

$$E(T_N^{\varphi} T_N) = \sum_{i=1}^{N} (c_i-\bar{c})(t_i-\bar{t}) \sum_{h\in H} \frac{f'(x_h,\bar{t})}{f(x_h,\bar{t})} \int_{F(x_{h-1},\bar{t})}^{F(x_h,\bar{t})} \varphi(u)\,du$$

$$= \sum_{i=1}^{N} (c_i-\bar{c})(t_i-\bar{t}) \int_0^1 \psi_f(u,\bar{t})\, \varphi(u)\,du$$

d'après (2.7). Par conséquent :

$$\lim_{N\to+\infty} E(T_N^{\varphi} T_N) = \lim_{N\to+\infty} \sum_{i=1}^{N} (c_i-\bar{c})(t_i-\bar{t}) \int_0^1 \psi_f(u,\bar{t})\varphi(u)\,du = m$$

Ensuite, il suffit de reprendre la démonstration du cas continu, en remplaçant S_N par S'_N.

6. TESTS ASYMPTOTIQUEMENT UNIFORMEMENT LES PLUS PUISSANTS.

Soit X^N un échantillon à valeurs dans $\mathbb{R}^N$ pour tout $N \geq 1$, et $\mathcal{P}_N = \{P_N\}$ et Q_N respectivement un ensemble de lois de probabilité et une probabilité sur $(\mathbb{R}^N, \mathcal{B}^N)$ tels que $Q_N \notin \mathcal{P}_N$.
Considérons la suite de couples d'hypothèses :

$$H_0^N : \text{la loi de } X^N \in \mathcal{P}_N$$

$$H_1^N : \text{la loi de } X^N \text{ est } Q_N$$

et supposons que, pour tout N, la décision du test de H_0^N contre H_1^N s'effectue d'après les valeurs d'une fonction de test $\phi_N = \mathbb{R}^N \to [0,1]$.

DEFINITION 1 : La suite de fonctions de test ϕ_N détermine <u>un test asymptotiquement</u> de seuil α de H_0^N contre H_1^N si et seulement si :

$$(6.1) \qquad \limsup_{N\to+\infty} \sup_{P_N\in\mathcal{P}_N} \int \phi_N\, dP_N = \alpha$$

DEFINITION 2 : La puissance asymptotique du test de H_0^N contre H_1^N déterminé par les fonctions de test ϕ_N est :

$$\beta = \liminf_{N \to +\infty} \int \phi_N \, dQ_N .$$

Pour tout $N \geqslant 1$, désignons par $\beta(\alpha, H_0^N, Q_N)$ la puissance d'un test uniformément le plus puissant de H_0^N contre H_1^N, de seuil α ;

DEFINITION 3 : La suite de fonctions de tests ϕ_N détermine un test asymptotiquement uniformément le plus puissant (a.u.p.p.) au seuil de H_0^N contre H_1^N si et seulement si on a (6.1) et

$$\lim_{N \to +\infty} \left[\beta(\alpha, H_0^N, Q_N) - Q_N) - \int \phi_N \, d\, Q_N\right] = 0$$

Soient maintenant les suites d'hypothèses :

$$\begin{cases} H_0^N : X_{N1}, \dots, X_{NN} \text{ sont mutuellement indépendantes et de même loi continue.} \\ H_1^N : X^N = (X_{N1}, \dots, X_{NN}) \text{ est de densité } q_N(x^N) = \prod_{i=1}^{N} f(x_i, t_{Ni}). \end{cases}$$

$$\begin{cases} H_0^{*N} : X_{N1}, \dots, X_{NN} \text{ sont mutuellement indépendantes et de même loi} \\ H_1^{*N} : X^N = (X_{N1}, \dots, X_{NN}) \text{ est de densité } q_N(x^N) = \prod_{i=1}^{N} f(x_i, t_{Ni}), \end{cases}$$

et les fonctions de test :

$$(6.2) \qquad \begin{cases} \phi_N = 1 & \text{si} \quad \dfrac{S_N^t}{\sigma_N^t} \geqslant k[1-\alpha] \\ = 0 & \text{si} \quad \dfrac{S_N^t}{\sigma_N^t} < k[1-\alpha] \end{cases}$$

$$(6.3) \qquad \begin{cases} \phi_N = 1 & \text{si} \quad \dfrac{S_N^{*t}}{\sigma_N^t} \geqslant k[1-\alpha] \\ = 0 & \text{si} \quad \dfrac{S_N^{*t}}{\sigma_N^t} < k[1-\alpha] \end{cases}$$

où $S_N^t = \sum_{i=1}^{N} (t_{Ni} - \bar{t}_N) \, a_N(R_{Ni})$, $S_N^{*t} \sum_{i=1}^{N} (t_{Ni} - \bar{t}_N) a_N(R_{Ni}^*)$, $(\sigma_N^t)^2 = I_f(0) \sum_{i=1}^{N} (t_{Ni} - \bar{t}_N)^2$

$k[y]$ le quantile d'ordre y de la fonction de répartition Φ de $\mathcal{N}(0,1)$: $\Phi(k[y]) = y$, $\forall y \in]0,1[$.

THEOREME 1. Si

i) la famille $\{f(x,\theta)\}$ est définie par rapport à la mesure de Lebesgue et satisfait aux conditions (C_1) à (C_5),

ii) les scores a_N vérifient la condition de convergence par rapport à $\psi_f(u,0)$

iii) les t^N satisfont à (3.3) et (3.4)

iv) $\lim_{\theta \to 0} \int_0^1 \psi_f(u,\theta)\, \psi_f(u,0)\,du = \int_0^1 \psi_f^2(u,0)\,du$,

la suite de fonctions de tests ϕ_N donnée par (6.2) détermine un test a.u.p.p. de seuil α de H_0^N contre H_1^N, de puissance asymptotique $\beta = 1 - \Phi(k[1-\alpha] - b)$.

THEOREME 2 : Si

a) la famille $\{f(x,\theta)\}$ est définie par rapport à la mesure μ_φ sur un ensemble discret $\mathcal{C}$ donné par (2.4), (2.5) et (4.1), et satisfait aux conditions (C'_1) et (C_2) à (C_5),

b) les hypothèses ii) à iv) du théorème 1 sont satisfaites,

La suite de fonctions de tests ϕ_N^* donnée par (6.3) détermine un test a.u.p.p de seuil de H_0^{*N} contre H_1^{*N} de puissance asymptotique $\beta = 1 - \Phi(k[1-\alpha] - b)$.

Démonstration : Ces deux théorèmes découlent des résultats des paragraphes 1, 3 et 4 suivant une démonstration analogue à celle de [2] (théorème VII 1.3) concernant le cas mesure de Lebesgue avec $f(x,\theta) = g(x-\theta)$ et $f(x,\theta) = e^{-\theta} g(xe^{-\theta})$.

Les deux exemples d'application suivants concernent les doubles échantillons, l'un dans le cas continu, l'autre dans le cas discret.

Considérons des couples d'entiers $\{(m_n, m'_n)\}$, $n \geq 1$; posons $N_n = m_n + m'_n$ et supposons que $\lim_{n \to +\infty} N_n = +\infty$. Soient :

$$t_{Ni} = \Delta_{N_n} \neq 0 \qquad \forall\ 1 \leq i \leq m_n$$

$$= 0 \qquad \forall\ m_{n+1} \leq i \leq m_n + m'_n.$$

Pour que les hypothèses (3.3) et (3.4) soient vérifiées, nous supposerons de plus que $\lim_{n \to +\infty} \Delta_{N_n} = 0$ et $\lim_{n \to +\infty} \frac{m_n m'_n}{m_n + m'_n} \Delta^2_{N_n} = c$, c fini et non nul.

Dans la suite nous négligeons un certain nombre d'indices n.

Exemple 1 : Soient $f(x,\theta) = \frac{1}{a+\theta} e^{-\frac{x}{a+\theta}}$ si $x > 0$, 0 sinon, avec $a > 0$ et $\theta \in]-a,+\infty[$,

H_0^N : $X_1,\dots,X_m, X_{m+1},\dots,X_{m+m'}$ sont mutuellement indépendantes et de même loi continue.

H_1^N : $(X_1,\dots,X_m)$ et $(X_{m+1},\dots,X_{m+m'})$ sont deux vecteurs aléatoires indépendants constitués de v.a. mutuellement indépendantes, de même densité $f(x,\Delta_N)$ pour le premier, $f(x,0)$ pour le second.

Le test de H_0^N contre H_1^N de région de rejet :

$$\varepsilon(\Delta_N) \frac{-\sum_{i=1}^{m} \mathrm{Log}\left(1-\frac{R_{Ni}}{m+m'+1}\right) + C_n}{|D_n|} \geqslant k \, [1-\alpha]$$

où $\varepsilon(\Delta_N) = +1$ ou -1 suivant que Δ_N est > 0 ou < 0, R^N est le vecteur rang de $X^N = (X_1,\dots,X_{m+m'})$, $C_n = \frac{m}{m+m'} \sum_{i=1}^{N} \mathrm{Log}\left(1-\frac{i}{m+m'+1}\right)$, $D_n^2 = \frac{mm'}{m+m'}$,

est a.u.p.p. de seuil α.

Pour montrer ce résultat, on vérifie assez aisément les hypothèses du théorème 1. Notons que les a_N satisfont à l'hypothèse de convergence par rapport à $\Psi_f(u,0) = -\frac{1}{a}\left[1+\log(1-u)\right]$ en vertu du lemme V.1.6.b de [2].

Exemple 2 : Soient $f(x,\theta) = \binom{k}{x}(p+\theta)^x(1-p-\theta)^{k-x}$, $\forall x \in \{0,1,\dots,k\}$, $\forall \theta \in]-p,1-p[$, $p \in]0,1[$;

H_0^{*N} : $X_1,\dots,X_m,\dots,X_{m+m'}$ sont mutuellement indépendantes et de même loi.

H_1^{*N} : $(X_1,\dots,X_m)$ et $(X_{m+1},\dots,X_{m+m'})$ sont deux vecteurs aléatoires indépendants, constitués de v.a. mutuellement indépendantes, de même loi binomiale

$\mathcal{B}(k,p+\Delta_N)$ pour le premier, $\mathcal{B}(k,p)$ pour le second.

Soit le test de H_0^{*N} contre H_1^{*N} de région de rejet :

$$\mathcal{E}(\Delta_N)\ \frac{\sum_{i=1}^{m} H(R^*_{Ni}) - C'_n}{|D'_n|} \geqslant k\left[1-\alpha\right]$$

où R^{*N} est un vecteur rang randomisé de $X^N = (X_1,\ldots,X_{m+m'})$, la fonction H est définie par :

$$H(i) = \frac{h-kp}{p(1-p)} \quad \text{si } \frac{i}{m+m'+1} \in I_h = \left]F(h-1)\ ,\ F(h)\right] \quad 0 \leqslant h \leqslant k\ ,$$

F désignant la fonction de répartition de $\mathcal{B}(k,p)$; $C'_n = \frac{m}{p(1-p)}\left[\frac{m+m'+1}{2} - kp\right]$,

$$D_n^2 = \frac{mm'}{(m+m')p^2(1-p)^2}\left[\sum_{h=0}^{k} (h-kp)^2\ p^h\ (1-p)^{k-h}\right].$$

Ce test est a.u.p.p. de seuil α.

REFERENCES

[1] HAJEK J. Non parametric statistics, Holden-Day, San Francisco (1969)

[2] HAJEK J. et SIDAK Z. Theory of rank tests, Academia Publishing House of the Czechoslovak Academy of Sciences, Prague (196

[3] NASI O. Thèse de 3ème cycle, Université de Nancy I (1975)

[4] VORLICKOVA D. Asymtotic properties of rank tests under discrete distributions, Zeitschrift für Warscheinlichkeits theorie und verw gebiete, vol.14, 275-289, (1970).