Atti Accademia Nazionale dei Lincei

Classe Scienze Fisiche Matematiche Naturali

Rendiconti

Paul Deheuvels

Majoration et minoration presque sûre optimale des éléments de la statistique ordonnée d'un échantillon croissant de variables aléatoires indépendantes

Atti della Accademia Nazionale dei Lincei. Classe di Scienze Fisiche, Matematiche e Naturali. Rendiconti, Serie 8, Vol. **56** (1974), n.5, p. 707–719.

Accademia Nazionale dei Lincei

<http://www.bdim.eu/item?id=RLINA_1974_8_56_5_707_0>

Atti della Accademia Nazionale dei Lincei. Classe di Scienze Fisiche, Matematiche e Naturali. Rendiconti, Accademia Nazionale dei Lincei, 1974.

Calcolo delle probabilità. — *Majoration et minoration presque sûre optimale des éléments de la statistique ordonnée d'un échantillon croissant de variables aléatoires indépendantes.* Nota di PAUL DEHEUVELS, presentata [*] dal Socio straniero A. LICHNEROWICZ.

RIASSUNTO. — Se $\{X_n ; n = 1, 2, \cdots\}$ è una successione di variabili aleatorie indipendenti aventi una stessa legge e si designa con $X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n,n}$ la statistica ordinata rispetto all'ordine crescente delle $X_1, X_2, \cdots, X_n$, qui vengono studiate le successioni numeriche $M_{n,m(n)}$, $m_{n,m(n)}$ tali che per n sufficientemente grande si abbia

$$m_{n,m(n)} \leq X_{m(n),n} \leq M_{n,m(n)} .$$

Principalmente si studiano siffatti inquadramenti per $m(n) = m$ costante, da un lato per scale di funzioni a crescenza di tipo logaritmico e, dall'altro, dando condizioni necessarie e sufficienti relative alle successioni m ed M affinché esse realizzino un tale inquadramento. Questo problema è già stato risolto in (1), (2) per $X_{1,n}$ ed $X_{n,n}$, e qui si danno dimostrazioni esplicite per i risultati ottenuti.

I metodi qui utilizzati possono venire generalizzati al caso di $m(n)$ variabile, come si mostrerà un successivo lavoro.

CHAPITRE I: ENCADREMENTS PRESQUE SURS DES EXTREMA D'UNE SUITE DE VARIABLES ALÉATOIRES INDÉPENDANTES DE MÊME LOI

1. INTRODUCTION

Soit $\{X_n , n \geq 1\}$ une suite de v.a. indépendantes de même fonction de répartition F.

Soit $\{U_n , n \geq 1\}$ une suite de v.a. indépendantes uniformément distribuées sur $(0,1)$. Si on pose:

$$G(t) = \operatorname{Inf}\{x \in \mathscr{R} \,\|\, F(x) \geq t\}, \tag{1,1,1}$$

on sait que $\{G(U_n), n \geq 1\}$ est identique en loi à $\{X_n , n \geq 1\}$. Par conséquent, encadrer $\operatorname{Inf}(X_1, \cdots, X_n)$ peut se ramener à encadrer $\operatorname{Inf}(U_1, \cdots, U_n)$, et de même pour $\operatorname{Sup}(X_1, \cdots, X_n)$. Dans ce qui suit, on étudiera donc:

$$Z_n = \operatorname{Inf}(U_1, \cdots, U_n).$$

La minoration p.s. de Z_n est triviale par Borel-Cantelli (2). La majoration p.s. de Z_n a été etudiée dans (1) et (2). Nous y avions obtenu [1]:

p.s. pour $n \to \infty$,

$$Z_n < \frac{1}{n}(\operatorname{Log}_2 n + 2 \operatorname{Log}_3 n + \operatorname{Log}_4 n + \cdots + (1 + A) \operatorname{Log}_p n), \quad A > 0, \tag{1,1,2}$$

(*) Nella seduta del 28 maggio 1974.

(1) Nous rectifions ici deux énoncés inexacts de (1) et (2) concernant la majoration de Z_n. Les démonstrations de (1) et (2) conduisent seulement à (1,1,2) et (1,1,3).

p.s., il existe une infinité d'indices n tels que:

$$(1,1,3) \qquad Z_n > \frac{1}{n}\left(\mathrm{Log}_2\, n + \mathrm{Log}_3\, n + \mathrm{Log}_4\, n + \cdots + \mathrm{Log}_p\, n\right).$$

Nous montrons ici que la majoration (1,1,2) est optimale.

Il est utile de se ramener à la majoration de suites de temps de passage définis par:

$$(1,1,4) \qquad \tau_\varepsilon = \mathrm{Inf}\{n \geq 1 \mid Z_n \leq \varepsilon\}.$$

Les propriétés des temps de passage ont été étudiées dans (1) (2) (3). Nous renvoyons à ces références. Il est équivalent de montrer (1,1,2) et: p.s. pour $n \to \infty$,

$$(1,1,5) \qquad \tau_{\varepsilon_n} < n\left(\mathrm{Log}_2\, n + 2\,\mathrm{Log}_3\, n + \mathrm{Log}_4\, n + \cdots + (1+A)\,\mathrm{Log}_p\, n\right), \qquad \varepsilon_n = 1/n.$$

On montrera ici cette majoration, et le fait que, p.s., il existe une infinité de n tels qu'elle ne soit pas vérifiée pour $A = 0$ [1].

2. Minoration p.s. de Z_n

Proposition 1. *Une* C.N.S. *pour que, p.s., pour* $n \to \infty$, $Z_n > u_n$, *est que*:

$$(1,2,1) \qquad \sum_{n=1}^{\infty} u_n < \infty. \quad \textit{En particulier, si } A > 0,$$

$$(1,2,2) \qquad \frac{1}{n\,(\mathrm{Log}\, n)\,(\mathrm{Log}_2\, n)\cdots(\mathrm{Log}_p\, n)^{1+A}} < Z_n$$

et p.s., pour $A = 0$, *il existe une infinité d'indices* n *tels que* (1,2,2) *ne soit pas vérifié.*

Preuve: Nous renvoyons à (2).

3. Majoration p.s. de Z_n

Posons tout d'abord $N(n, A) = n\left(\mathrm{Log}_2\, n + 2\,\mathrm{Log}_3\, n + \mathrm{Log}_4\, n + \cdots + (1 + A)\,\mathrm{Log}_p\, n\right)$, et $n_k = \exp\left(kR\,\frac{\mathrm{Log}_{p-1} k}{\mathrm{Log}\, k}\right)$. Nous allons donner deux démonstrations pour (1,1,2) dans le cas $A > 0$, et une seule dans le cas $A = 0$.

a) *Cas où* $A > 0$.

La démonstration la plus simple suit le plan de (1) et (2): On considère:

$$C_n = \{\tau_{\varepsilon_n} > n\left(\mathrm{Log}_2\, n + \mathrm{Log}_3\, n + \cdots + (1 + A)\,\mathrm{Log}_p\, n\right),\ \tau_{\varepsilon_{n+1}} - \tau_{\varepsilon_n} > 0\}$$

$$D_n = \{\tau_{\varepsilon_n} \leq n\left(\mathrm{Log}_2\, n + \mathrm{Log}_3\, n + \cdots + (1 + A\,(\mathrm{Log}_p\, n)\right),$$
$$\tau_{\varepsilon_{n+1}} > (n+1)\left(\mathrm{Log}_2 (n+1) + 2\,\mathrm{Log}_3 (n+1) + \cdots + (1+A)\,\mathrm{Log}_p (n+1)\right)\}.$$

On constante aisément, que si $A > 0$,

$$\sum_{n=1}^{\infty} P(G_n) \quad , \quad \text{et} \ \sum_{n=1}^{\infty} P(D_n)$$

sont finis. Il suffit alors d'appliquer Borel–Cantelli.

Cette démonstration de (1,1,2) est intéressante, car elle se généralise facilement.

En effet, posons, si $\{u_n \, , n \geq 1\}$ est une suite décroissant vers 0:

$$E_n = \{\tau_{u_n} > n \, , \tau_{u_{n+1}} - \tau_{u_n} > 0\} \, , \qquad F_n = \{\tau_{u_n} \leq n \, , \tau_{u_{n+1}} > n+1\} \, .$$

On obtient:

$$P(E_n) = \frac{u_n - u_{n+1}}{u_n} (1 - u_n)^n ,$$

$$P(F_n) = \sum_{i=1}^{n} (1 - u_n)^{i-1} (u_n - u_{n+1}) (1 - u_{n+1})^{n-i} \, .$$

Il est aisé de constater que les séries $\sum_{n=1}^{\infty} P(E_n)$ et $\sum_{n=1}^{\infty} P(F_n)$ sont de même nature que les séries:

$$(1,3,1) \qquad \sum_{n=1}^{\infty} \frac{u_n - u_{n+1}}{u_n} \exp(-n u_n) \quad \text{et} \quad \sum_{n=1}^{\infty} (\exp(-n u_{n+1}) - \exp(-n u_n)).$$

La convergence de ces deux séries étant donc une condition suffisante pour que p.s., pour $n \to \infty$, $Z_n < u_n$.

La deuxième démonstration fait intervenir les suites n_k, et les méthodes utilisées seront commodes pour $A = 0$. On utilise les faits suivants:

1) $n_{k+1} - n_k \# n_k \left(\exp\left(R \dfrac{\mathrm{Log}_{p-1} k}{\mathrm{Log}\, k}\right) - 1\right) = n_k R \dfrac{\mathrm{Log}_p n_k}{\mathrm{Log}_2 n_k} + n_k o \left[\dfrac{\mathrm{Log}_p n_k}{\mathrm{Log}_2 n_k}\right]^2$

2) $N(n_{k+1}, A) \# N(n_k, A+R)$.

On en déduit que, pour montrer (1,1,2), il est nécessaire et suffisant de montrer: que, p.s. pour k,

$$(1,3,2) \qquad \tau_{\varepsilon_{n_k}} < N(n_k, A - R) ,$$

et ceci, pour R arbitrairement petit.

Or:

$$P(\tau_{\varepsilon_{n_k}} > N(n_k, A - R)) \sim \exp\left(-\frac{N(n_k, A - R)}{n_k}\right) \sim$$

$$\sim \frac{1}{k R \dfrac{\mathrm{Log}_{p-1} k}{\mathrm{Log}\, k} (\mathrm{Log}\, k)^2 (\mathrm{Log}_3 k) \cdots (\mathrm{Log}_{p-2} k)^{1+A-R}} \, .$$

La série $\sum_{k=1}^{\infty} P(\tau_{\varepsilon_{n_k}} > N(n_k, A - R))$ converge donc pour $A > 0$, et $R < A$, d'où le résultat par Borel–Cantelli.

On résume ces résultats dans:

THÉORÈME 1. $\forall A > 0$, $\forall p \geq 4$, *p.s. pour* $n \to \infty$, (1,1,2) *est vérifié. De plus, pour que, p.s. pour* $n \to \infty$, $Z_n < u_n$, *il suffit que*:

$$\sum_{n=1}^{\infty} \frac{u_n - u_{n+1}}{u_n} \exp(-nu_n) < \infty \tag{1,3,3}$$

$$\textit{et} \quad \sum_{n=1}^{\infty} (\exp(-nu_{n+1}) - \exp(-nu_n)) < \infty .$$

b) *Cas où* $A = 0$.

Avec les hypothèses précédentes, compte tenu du fait que, p.s. pour $n \to \infty$, C_n n'a pas lieu, on en déduit que, pour montrer que, p.s. il existe une infinité d'indices n tels que $Z_n > N(n, 0)$, il suffit de montrer que, $\forall R > 0$, $\forall p \geq 4$, il existe p.s. une infinité d'indices k tels que:

$$\{N(n_k, -2R) \leq \tau_{\varepsilon_{n_k}} \leq N(n_k, -R)\} = A_k \quad \text{a lieu.} \tag{1,3,4}$$

Pour montrer ce résultat en se ramenant à Borel-Cantelli, bien que les A_k soient liés, on utilise le lemme:

LEMME 1. *Si* B_k *est une famille d'événements tels que*:

$$\operatorname{LimInf}_{n\infty} \frac{\sum_{i,j=1}^{n} P(B_i B_j)}{\sum_{i,j=1}^{n} P(B_i) P(B_j)} = 1 , \quad \text{et} \quad \sum_{n=1}^{\infty} P(B_n) = \infty , \tag{1,3,5}$$

alors, p.s., une infinité de B_n *a lieu.*

Preuve. Voir par exemple (3), p. 368.

Considérons alors $P(A_k A_{k+r})$, $r \geq 1$:

$$P(A_k A_{k+r}) =$$

$$= \sum_{N(n_k, -2R) \leq s \leq N(n_k, -R)} \varepsilon_{n_k} (1 - \varepsilon_{n_k})^{s-1} (\varepsilon_{n_k} - \varepsilon_{n_{k+r}}) \varepsilon_{n_k}^{-1} [(1 - \varepsilon_{n_k})^{t-s-1}]_{t=N(n_{k+r}, +2R)}^{t=N(n_{k+r}, +R)}$$

Soit encore:

$$P(A_k A_{k+r}) \# P(A_{k+r}) \exp - \left(1 - \exp\left(- Rr \frac{\operatorname{Log}_{p-1} k}{\operatorname{Log} k}\right)\right) \times \tag{1,3,6}$$

$$\times (\operatorname{Log} k + 2 \operatorname{Log}_2 k + \cdots + (1 - 2R) \operatorname{Log}_{p-1} k) ,$$

ceci, pourvu que $r = o(k)$. Constatons alors, que, d'une part, pour k fixé, $P(A_k A_{k+r})$ est une fonction décroissante de r, d'autre part, Pour $\quad r \leq \mathrm{D}\,\dfrac{(\operatorname{Log} k)^{1/4}}{\operatorname{Log}_{p-1} k}$,

$$\mathrm{P}(\mathrm{A}_k \mathrm{A}_{k+r}) \sim \mathrm{P}(\mathrm{A}_k)\,\frac{1}{(\operatorname{Log}_{p-2} k)^{\mathrm{R}r}},$$

Pour $\quad r > \mathrm{D}\,\dfrac{(\operatorname{Log} k)^{1/4}}{\operatorname{Log}_{p-1} k}$,

$$\mathrm{P}(\mathrm{A}_k \mathrm{A}_{k+r}) < \mathrm{P}(\mathrm{A}_k) \exp\left(-\mathrm{D}\,(\operatorname{Log} k)^{1/4}\right) < \frac{\mathrm{P}(\mathrm{A}_k)}{(\operatorname{Log} k)^2},$$

pour k assez grand.

Considérons alors:

1) $$\sum_{r=0}^{(\operatorname{Log} k)^{1/4}/\operatorname{Log}_{p-1} k} \mathrm{P}(\mathrm{A}_k \mathrm{A}_{k+r}) \sim \mathrm{P}(\mathrm{A}_k)$$

2) $$\sum_{r=(\operatorname{Log} k)^{1/4}/\operatorname{Log}_{p-1} k}^{\frac{(\operatorname{Log}_2 k)(\operatorname{Log} k)\mathrm{H}}{\operatorname{Log}_{p-1} k}} \mathrm{P}(\mathrm{A}_k \mathrm{A}_{k+r}) < \mathrm{P}(\mathrm{A}_k)\,\frac{(\operatorname{Log}_2 k)(\operatorname{Log} k)\,\mathrm{H}}{\operatorname{Log}_{p-1} k}\,\frac{1}{(\operatorname{Log} k)^2}$$

3) $$\sum_{r=\frac{(\operatorname{Log}_2 k)(\operatorname{Log} k)\mathrm{H}}{\operatorname{Log}_{p-1} k}}^{\mathrm{N}} \mathrm{P}(\mathrm{A}_k \mathrm{A}_{k+r}) \qquad \sum_{r=\frac{(\operatorname{Log}_2 k)(\operatorname{Log} k)\mathrm{H}}{\operatorname{Log}_{p-1} k}}^{\mathrm{N}} \mathrm{P}(\mathrm{A}_k)\,\mathrm{P}(\mathrm{A}_{k+r})$$

si $\mathrm{HA} > 1$.

Il suffit alors de constater que

$$\sum_{k=1}^{n} \mathrm{P}(\mathrm{A}_k) \sim \frac{(\operatorname{Log}_{p-2} n)^{2\mathrm{R}}}{2\,\mathrm{R}^2\,(\operatorname{Log}_{p-1} n)} \to \infty,$$

et de combiner 1), 2), 3), pour obtenir (1,3,5) et le résultat cherché.

Revenons maintenant sur (1,3,3). Supposons tout d'abord que $nu_n \to \infty$. Dans ce cas, par la convergence de la première série, comparée avec celle de la seconde que la deuxième série est suffisante. Elle se met alors sous la forme équivalente:

$$\sum_{n=1}^{\infty} n\,(u_n - u_{n+1}) \exp(-nu_n) < \infty. \tag{1,3,7}$$

Faisons alors l'hypothèse:

$$\exists \mathrm{A}, \mathrm{B}, 0 < \mathrm{A}\,\frac{n_{k+1}-n_k}{n_k} \leq \frac{u_{n_k} - u_{n_{k+1}}}{n_{n_k}} \leq \mathrm{B}\,\frac{n_{k+1}-n_k}{n_k} < \infty \tag{1,3,8}$$

où ici, n_k est définie par récurrence en posant $n_{k+1} = n_k + \dfrac{1}{u_{n_k}}$. (1,3,7) est alors équivalent à:

$$\sum_{k=1}^{\infty} \exp(-n_k u_{n_k}) < \infty. \tag{1,3,9}$$

On montre ainsi l'équivalence de (1,3,1) et de (1,3,2). Supposons maintenant que:

$$(1,3,10) \qquad \sum_{k=1}^{\infty} \exp\left(-n_k u_{n_k}\right) = \infty .$$

Compte tenu des hypothèses précèdentes, il existe une sous-suite m_k extraite de la suite n_k, telle que:

$$(1,3,11) \qquad \sum_{k=1}^{\infty} \exp\left(-m_k u_{m_k}\right) = \infty , \qquad \text{et} \qquad m_{k+1} = m_k + \frac{\varphi(k)}{u_{m_k}} ,$$

$$\text{avec} \quad \varphi(k) \to \infty .$$

Considérons alors les événements:

$$(1,3,12) \qquad B_k = \{ m_k < \tau_{\varepsilon_{m_k}} < m_{k+1} \} .$$

(1,3,10) implique que $\sum_{k=1}^{\infty} P(B_k) = \infty$. Le même calcul que pour (1,3,6) montre alors que:

$$(1,3,12) \qquad P(B_k B_{k+r}) \sim P(B_{k+r}) \exp\left(-m_m u_{m_k} \frac{u_{m_k} - u_{m_{k+r}}}{u_{m_k}}\right).$$

La suite de la démonstration est analogue à ce qui précède. Il suffit de constater que $P(B_k B_{k+r})$ est décroissante, et de décomposer en trois parties pour se ramener à (1,3,5). On montre ainsi que (1,3,7) est nécessaire et suffisante. Nous résumons ces résultats:

THÉORÈME 2. $\forall p \geq 4$, *il existe p.s. une infinité d'indices n tels que* (1,1,2) *ne soit pas vérifié avec* $A = 0$.

De plus, si $\{u_n , n \geq 1\}$ *est une suite décroissante telle que* $nu_n \to \infty$, *et* (1,3,8) *soit vrai, une* C.N.S. *pour qu'existe une infinité d'indices n tels que* $Z_n > u_n$, *est que* (1,3,7) *soit vérifié.*

Remarque. Les majorations obtenues par (1,1,3) en faisant varier p, sont obtenues dans une échelle de développements qui ne peut être améliorée en utilisant des fonctions simples. En effet, l'échelle des séries de Bertrand utilisée correspond, en termes de fonctions croissantes, aux fonctions $\exp(\exp(\cdots \exp(x)) \cdots))$. Toute amélioration de la précision des développements, possible par l'utilisation de (1,3,7), necessite l'introduction de fonctions croissant plus rapidement que $\exp(\cdots \exp(x)) \cdots))$, à tous les ordres.

Remarquons également que la condition (1,3,7), si elle est vérifiée implique toujours que, p.s. il existe une infinité d'indices n tels que:

$$(1,3,13) \qquad \tau_{\varepsilon_{u_{n+1}}} - \tau_{\varepsilon_{u_n}} > n , \qquad \text{et par conséquent} \quad Z_n > u_n .$$

Ces conditions mènent à la majoration (1,1,3) qui n'est pas optimale. La présence du coefficient 2 dans (1,1,2) est surprenante, elel s'explique cependant en considérant la loi limite de $n\tau_{\varepsilon_n}$.

4. EXTREMA DE SUITES DE V.A. INDÉPENDANTES DE MÊME LOI

On déduit de la Proposition 1, des Théorèmes 1, 2 et de (1,1,1), que:

COROLLAIRE 1. $\forall p \geq 4, \forall A > 0$, *p.s. pour n assez grand,*

$$(1,4,1)\qquad G\left[\frac{1}{n(\operatorname{Log} n)(\operatorname{Log}_2 n)\cdots(\operatorname{Log}_p n)^{1+A}}\right] \leq \operatorname{Inf}(X_1,\cdots,X_n) \leq$$

$$\leq G\left[\frac{1}{n}(\operatorname{Log}_2 n + 2\operatorname{Log}_3 n + \operatorname{Log}_4 n + \cdots + (1+A)\operatorname{Log}_p n\right],$$

De plus, pour $A = 0$, *il existe p.s. une infinité de fois où chacun des deux membres de* (1,4,1) *ne sont pas vérifiés.*

Nous énonçons dans le paragraphe suivant les résultats obtenus pour des lois usuelles.

a) *Loi uniforme sur* (0, 1). C'est l'énoncé même de (1,2,2) et (1,1,2).

b) *Loi exponentielle*:

$$F(x) = 1 - e^{-x} \quad , \quad G(x) = -\operatorname{Log}(1-x).$$

1) *Minimum*: Mêmes encadrements que pour la loi uniforme.

2) *Maximum*:

$$\overline{Y}_n = \operatorname{Sup}(X_1,\cdots,X_n) < \operatorname{Log} n + \operatorname{Log}_2 n + \operatorname{Log}_3 n + \cdots + (1+A)\operatorname{Log}_p n.$$

$$\overline{Y}_n > \operatorname{Log} n - \operatorname{Log}_3 n - \frac{2\operatorname{Log}_3 n}{\operatorname{Log}_2 n} - \frac{\operatorname{Log}_4 n}{\operatorname{Log}_2 n} - \frac{\operatorname{Log}_5 n}{\operatorname{Log}_2 n} - \cdots - \frac{\operatorname{Log}_p n}{\operatorname{Log}_2 n}.$$

c) *Loi gamma*:

$$F(x) = \int_0^x t^{a-1} e^{-t}\,\Gamma(a)^{-1}\,dt.$$

1) *Minimum*:

$$\underline{Y}_n = \operatorname{Inf}(X_1,\cdots,X_n) > \left[\frac{\Gamma(a)}{n(\operatorname{Log} n)\operatorname{Log}_2 n\cdots(\operatorname{Log}_p n)^{1+A}}\right]^{1/a}.$$

$$\underline{Y}_n\ \frac{(\operatorname{Log}_2 n)^{1/a}}{n^{1/a}}\left\{1 + \frac{2\operatorname{Log} n_3}{a\operatorname{Log} n_2} + \frac{\operatorname{Log}_4 n}{a\operatorname{Log}_2 n} + \frac{\operatorname{Log}_5 n}{a\operatorname{Log}_2 n} + \cdots + (1+A)\frac{\operatorname{Log}_p n}{a\operatorname{Log}_2 n}\right\}(\Gamma(a))^{1/a}.$$

2) *Maximum*:

$$\overline{Y}_n < \operatorname{Log} n + a \operatorname{Log}_2 n + \operatorname{Log}_3 n + \cdots + (1 + A) \operatorname{Log}_p n$$

$$\overline{Y}_n > \operatorname{Log} n + (a - 1) \operatorname{Log}_2 n - \operatorname{Log}_3 n - \frac{2 \operatorname{Log}_3 n}{\operatorname{Log}_2 n} - \frac{\operatorname{Log}_4 n}{\operatorname{Log}_2 n} -$$

$$- \cdots - (1 + A) \frac{\operatorname{Log}_p n}{\operatorname{Log}_2 n} - \operatorname{Log} (\Gamma (a)) .$$

d) *Loi du* χ^2.

Loi à n degrés de liberté, c'est 2 fois une loi Gamma définie en c), avec $a = n/2$. Il suffit donc de multiplier par 2 les formules précédentes.

e) *Loi de Cauchy*:

$$F(x) = \frac{1}{\pi} \left(\frac{\pi}{2} + \operatorname{Arctg} \pi x \right)$$

$$-\underline{Y}_n , \overline{Y}_n < n (\operatorname{Log} n) (\operatorname{Log}_2 n) \cdots (\operatorname{Log}_p n)^{1+A} ;$$

$$-\underline{Y}_n , \overline{Y}_n > \frac{n}{\operatorname{Log}_2 n} \left[1 - \frac{2 \operatorname{Log}_3 n}{\operatorname{Log}_2 n} - \cdots - (1 + A) \frac{\operatorname{Log}_p n}{\operatorname{Log}_2 n} \right]$$

f) *Loi normale*

$$N(0 , 1)^{(2)} : F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(1/2)t^2} dt$$

$$\overline{Y}_n < \sqrt{2 \operatorname{Log} n} + \frac{\operatorname{Log}_2 n}{2 \sqrt{2 \operatorname{Log} n}} + \frac{\operatorname{Log}_3 n}{\sqrt{2 \operatorname{Log} n}} + \cdots + (1 + A) \frac{\operatorname{Log}_p n}{\sqrt{2 \operatorname{Log} n}} .$$

$$\overline{Y}_n > \sqrt{2 \operatorname{Log} n} - \frac{\operatorname{Log}_2 n}{2 \sqrt{2 \operatorname{Log} n}} - \frac{\operatorname{Log}_3 n}{\sqrt{2 \operatorname{Log} n}} - \frac{\operatorname{Log} 2\pi}{\sqrt{2 \operatorname{Log} n}} -$$

$$- \frac{2 \operatorname{Log}_3 n}{(\operatorname{Log}_2 n) \sqrt{2 \operatorname{Log} n}} - \frac{\operatorname{Log}_4 n}{(\operatorname{Log}_2 n) \sqrt{2 \operatorname{Log} n}} - \cdots - (1 + A) \frac{\operatorname{Log}_p n}{(\operatorname{Log}_2 n) \sqrt{2 \operatorname{Log} n}} .$$

f) *Loi de Poisson.*

$$F(n) = \sum_{i=1}^{n} \frac{a^n}{n!} e^{-a}.$$

Dans ce cas, l'échelle utilisée est insuffisamment précise. En développant G, on obtient des suites numériques encadrant $\overline{Y}_n$, p.s. pour $n \to \infty$, avec une précision de l'ordre de: $(\operatorname{Log} n)/(\operatorname{Log}_2 n)^K$, K arbitrairement grand. On ne peut distinguer ainsi la majoration et la minoration de $\overline{Y}_n$.

(2) Ces résultats rectifient un énoncé inexact de (4) pour les v.a. normales indépendantes.

CHAPITRE II: ENCADREMENTS PRESQUE SURS DES ÉLÉMENTS DE LA STATISTIQUE ORDONNÉE D'UN ÉCHANTILLON CROISSANT DE VARIABLES ALÉATOIRES INDÉPENDANTES DE MÊME LOI

1. INTRODUCTION

On conserve les définitions et notations du Chapitre I. On pose:

$$X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n,n} \tag{2,1,1}$$

la statistique ordonnée de $X_1, \cdots, X_n$. Nous étudions ici des encadrements presque sûrs de $X_{m(n),n}$, où $\{m(n), n \geq 1\}$ est une suite d'indices entiers ≥ 1, croissante au sens large. Dans le Chapitre I, nous avons résolu ce problème pour $m(n) = 1$.

Comme dans le Chapitre I, on se ramène à des varaibles uniformément distribuées sur $(0, 1)$, et on pose de même:

$$U_{1,n} < U_{2,n} < \cdots < U_{n,n} \quad \text{la statistique ordonnée de } U_1, \cdots, U_n. \tag{2,1,2}$$

Soit $\{\varepsilon_n, n \geq 1\}$ une suite positive décroissant strictement vers 0, et $\{m(n), n \geq 1\}$ une suite d'indices entiers plus grands que 1, et croissante au sens large. On pose:

$$\forall \varepsilon > 0, \qquad \tau_\varepsilon^{(m)} = \operatorname{Inf}\{p \mid \operatorname{Card}\{i \leq p \mid X_i \leq \varepsilon\} \geq m\}. \tag{2,1,3}$$

Cette définition généralise (1,1,4) avec $\tau_\varepsilon = \tau_\varepsilon^{(1)}$.

Comme précédemment, majorer $U_{m(n),n}$ peut se ramener à majorer $\tau_{\varepsilon_n}^{(m(n))}$. Nous étudions donc tout d'abord les propriétés des $\tau_{\varepsilon_n}^{(m(n))}$.

2. PROPRIÉTÉS DES TEMPS DE PASSAGE

PROPOSITION 1. *La suite* $\{\tau_{\varepsilon_n}^{(m(n))}, n \geq 1\}$ *est définie et croît indéfiniment p.s.*

Preuve. Pour $\varepsilon > 0$ fixé, il existe une infinité de i, p.s., tels que $X_i \leq \varepsilon$. D'autre part pour n donné, $\operatorname{Inf}(X_1, \cdots, X_n) > 0$, p.s..

THÉORÈME 3. *La suite* $\{\tau_{\varepsilon_{n+1}}^{(m(n+1))} - \tau_{\varepsilon_n}^{(m(n))}, n \geq 1\}$ *est une suite de v.a. indépendantes positives, défines p.s.. De plus*:

$$P\left(\tau_{\varepsilon_{n+1}}^{(m(n+1))} - \tau_{\varepsilon_n}^{(m(n))} = r\right) = \tag{2,2,1}$$

$$= \frac{(\varepsilon_{n+1})^{m(n+1)}}{(\varepsilon_n)^{m(n)}} \left\{ \sum_{p=\operatorname{Sup}(0,m(n+1)-r)}^{\operatorname{Inf}(m(n),m(n+1)-1)} C_{m(n)}^{P} C_{r-1}^{m(n+1)-p-1} (\varepsilon_n - \varepsilon_{n+1})^{m(n)-p} (1-\varepsilon_{n+1})^{r-p-m(n+1)} \right\}$$

$$\text{si} \quad r \geq m(n+1) - m(n).$$

Cette expression reste valable si $m(n) = m(n+1)$ *et* $r = 0$, *en remplaçant la sommation par* 1.

$$P(\tau_{\varepsilon_n}^{(m)} = r) = C_{r-1}^{m-1}(1-\varepsilon_n)^{r-m}(\varepsilon_n)^m, \qquad r \geq m. \tag{2,2,2}$$

Preuve. La démonstration est une généralisation de la démonstration de la même propriété pour $m(n) = m(n+1) = 1$, exposée dans (1).

3. Cas où $m(n) = m$ fixé

C'est le cas où les résultats seont les plus simples à obtenir par généralisation des méthodes utilisées dans le Chapitre I.

a) *Majoration de* $U_{m,n}$:

Comme précédemment, montrer que $U_{m,n} \leq u_n$ équivaut à montrer que $\tau_{u_n}^{(m)} \leq n$.

On obtient:

THÉORÈME 4. $\forall A > 0$, $\forall p \geq 4$, p.s. pour $n \to \infty$, on a:

$$U_{m,n} \leq \frac{1}{n}\{\operatorname{Log}_2 n + (m+1)\operatorname{Log}_3 n + \operatorname{Log}_4 n + \cdots + (1+A)\operatorname{Log}_p n\}. \tag{2,3,1}$$

Preuve.

1) *Première démonstration*. On introduit comme dans 1,3) *a*) les suites $n_k = \exp\left(kR\frac{\operatorname{Log}_{p-1} k}{\operatorname{Log} k}\right)$.

On montre que, pour montrer (2,3,1), il suffit de montrer la convergence de la série:

$$\sum_{k=1}^{\infty} P\left(\tau^{(m)}_{\frac{1}{n_k}(\operatorname{Log}_2 n_k + \cdots + (1+A)\operatorname{Log}_p n_k)} > n_k\right) \tag{2,3,2}$$

pour $A > 0$ et $R > 0$ arbitrairement petit.

Mais, cette série, par (2,2,2) est de même nature que

$$\sum_{k=1}^{\infty} n_k^{m-1} u_{n_k}^{m-1} \exp(-n_k u_{n_k}),$$

u_{n_k} ayant la valeur correspondant à (2,3,2). Le résultat suit alors par la même démonstration que pour le Théorème 1.

2) *Deuxième démonstration*. Cette démonstration se ramène également au cas $m = 1$. On pose:

$$E_n = \{\tau_{u_n}^{(m)} > n,\ \tau_{u_{n+1}}^{(m)} - \tau_{u_n}^{(m)} > 0\}, \qquad F_n = \{\tau_{u_n}^{(m)} \leq n,\ \tau_{u_{n+1}}^{(m)} > n+1\}.$$

Supposons maintenant que $\{u_n, n \geq 1\}$ soit une suite décroissant vers 0, telle que:

$$\lim_{m\infty} n(u_n - u_{n+1}) = 0 \quad , \quad \lim_{m\infty} u_n/u_{n+1} = 1 . \tag{2,3,3}$$

On obtient que $\sum_n P(E_n)$ est de même nature que

$$\sum_n n^{m-1} u_n^{m-2} (u_n - u_{n+1}) \exp(-nu_n) ,$$

et de même, $\sum_n P(F_n)$ est de même nature que

$$\sum_n n^m u_n^{m-1} (u_n - u_{n+1}) \exp(-nu_n) .$$

On obtient alors aisément (2,3,1), en remplacant u_n par la valeur correspondante, les deux séries étant convergentes. On obtient mieux.

THÉORÈME 4 BIS: *Si* $\{u_n, n \geq 1\}$ *est une suite décroissant vers* 0, *vérifiant* (2,3,3), *une condition suffisante pour que, p.s., pour* $n \to \infty$, $U_{m,n} \leq u_n$, *est que*:

$$\sum_n n^{m-1} u_n^{m-2} (u_n - u_{n+1}) \exp(-nu_n) < \infty \quad \text{et} \tag{2,3,4}$$

$$\sum_n n^m u_n^{m-1} (u_n - u_{n+1}) \exp(-nu_n) < \infty .$$

De même que pour le cas $m = 1$, qui a été précédemment été exposé en détail, on obtient par la même démonstration que:

THÉORÈME 5. *Presque sûrement, il existe une infinité d'indices n tels que* (2,3,1) *ne soit pas vérifié, pour* $A = 0$.

On obtient aisi les majorations optimales. Un raisonnement analogue permet d'obtenir avec des conditions très générales sur la suite u_n, que les conditions (2,3,4) sont nécessaires et suffisantes pour que $U_{n,n} < u_n$.

b) *Minoration de* $U_{m,n}$:

La minoration est ici beaucoup plus délicate que dans le cas où $m = 1$. Cependant, nous avons développé dans (1) une méthode qui se généralise très facilement à ce cas.

Considérons en effet le cas $m = 1$, on constate aisément que (1,2,2) est équivalent à

$$\liminf_{u \to \infty} u\tau_u (\mathrm{Log}(1/u)) (\mathrm{Log}_2(1/u)) \cdots (\mathrm{Log}_p(1/u))^{1+A} = \infty \quad \text{p.s.} \tag{2,3,5}$$

Pour montrer ce résultat, pour tout $A > 0$, il suffit de le montrer pour τ_{1/a^n}, a étant choisi arbitrairement > 1. On se ramène alors à des séries de Bertrand.

Constatons alors que pour le cas où m est fixé, on a la même simplification: Si $n\varepsilon_n \to 0$, on constate aisément que, par (2,2,2):

$$(2,3,6) \qquad P(\tau_{1/a}^{(m)} n < a^n (n (\mathrm{Log}\, n) \cdots (\mathrm{Log}_{p-1} n)^{1+A})^{-1/m}) = \\ = \frac{o(1)}{n (\mathrm{Log}\, n) (\mathrm{Log}_2\, n) \cdots (\mathrm{Log}_{p-1} n)^{1+A}} .$$

On en déduit par le même raisonnement:

THÉORÈME 6. $\forall A > 0$, $\forall p \geq 3$, *p.s. pour n assez grand*:

$$(2,3,7) \qquad U_{m,p} > \frac{1}{n ((\mathrm{Log}\, n) (\mathrm{Log}_2\, n) \cdots (\mathrm{Log}_p\, n)^{1+A})^{1/m}} .$$

La réciproque est analogue à la démonstration du Théorème 2. Montrons la dans le cas $m = 1$ pour simplifier. Supposons que $\{u_n , n \geq 1\}$ soit une suite décroissant vers 0, et telle que:

$$(2,3,8) \qquad \operatorname*{Lim}_{n\infty} n u_n = 0 , \qquad \text{et} \qquad \sum_{k=1}^{\infty} P(n_{k-1} < \tau_{u_{n_k}} \leq n_k) = \{n_k , k \geq 1\}$$

suite d'indices croissante.

Posons $A_k = \{n_{k-1} < \tau_{u_{n_k}} \leq n_k\}$. Estimons $P(A_k A_{k+r})$: Par (2,2,2) $(m = 1)$:

$$P(A_k A_{k+r}) \# \sum_{i=n_{k-1}}^{n_k - 1} u_{n_k} (1 - u_{n_k})^{i-1} (u_{n_k} - u_{n_{k+r}}) \frac{u_{n_{k+r}}}{u_{n_k}} \times$$

$$\times ((1 - u_{n_{k+r}})^{n_{k+r-1}-i} - (1 - u_{n_{k+r}})^{n_{k+r}-i}) \frac{1}{u_{n_{k+r}}}$$

$$P(A_k A_{k+r}) \# (n_k - n_{k-1}) (u_{n_k} - u_{n_{k+r}}) (n_{k+r} - n_{k+r-1}) u_{n_{k+r}} .$$

Prenons maintenant $n_k = \dfrac{a^k}{(k (\mathrm{Log}\, k) \cdots (\mathrm{Log}_{p-1} k)^{1+A})^{1/m}}$ et $u_{n_k} = 1/a^k$. On constate $(m = 1)$ que (2,3,8) est vérifié, et que, pour $r = o(k)$, $r \geq 1$,

$$(2,3,9) \qquad P(A_k A_{k+r}) = P(A_k) P(A_{k+r}) (1 + o(a^{-r})) .$$

On en déduit que (1,3,5) et le Lemme 1 s'appliquent. La démonstration est identique en tous points pour $m > 1$. On en déduit:

THÉORÈME 7. *P.s., il existe une infinité d'indices n tels que* (2,3,7) *ne soit pas vérifié pour* $A = 0$.

Ce résultat montre que (2,3,7) réalise les meilleures majorations possibles dans l'échelle considérée.

Donnons maintenant une autre démonstration du Théorème 6, qui permet d'obtenir des C.N.S. portant sur u_n, pour que $U_{m,n} > u_n$ p.s. pour n assez grand. On obtient facilement que:

$$(2,3,10) \qquad P(U_n \leq u_n , \mathrm{Card}\, \{i < n \mid X_i \leq u_n\} \geq m - 1) = o(n^{m-1} u_n^m) .$$

On peut montrer par des méthodes analogues à celles du Théorème 4, que la condition:

$$\sum_{n=1} n^{m-1} u_n^m < \infty \,, \tag{2,3,11}$$

est nécessaire et suffisante pour que $U_{m,n} > u_n$, p.s. pour n assez grand.

c) *Cas de v.a. indépendantes de même loi*:

On obtient comme corollaire de ce qui précède, que:

COROLLAIRE 2. *P.s. pour n assez grand*, $A > 0 \, , p \geq 4$,

$$G \left[\frac{1}{n ((\operatorname{Log} n) (\operatorname{Log}_2 n) \cdots (\operatorname{Log}_p n)^{1+A})^{1/m}} \right] \leq X_{m,n} \leq \tag{2,3,12}$$

$$\leq G \left[\frac{1}{n} (\operatorname{Log}_2 n + (m+1) \operatorname{Log}_3 n + \cdots + (1+A) \operatorname{Log}_p n) \right]$$

De plus, p.s. pour $A = 0$, *les deux membres de* (2,3,12) *ne sont pas vérifiés une infinité de fois.*

4. CAS OÙ $m(n)$ VARIE

Les mçmes méthodes, en particulier celles des Théorèmes 4 et suivants s'appliquent sans difficulté. Cependant, il y a lieu de distinguer suivant la vitesse de croissance de $m(n)$. Ceci peut être aisément fait, si $m(n)$ peut être développé suivant l'échelle logarithmique:

$$m(n) = An^a + B (\operatorname{Log} n)^b + C (\operatorname{Log}_2 n) + \cdots . \tag{2,3,12}$$

Les développements ainsi obtenus feront l'objet d'une publication ultérieure. On peut cependant remarquer le rôle frontière joué par les suites croissant comme $\operatorname{Log}_2 n$.

BIBLIOGRAPHIE

[1] P. DEHEUVELS (1974) – «Ann. Inst. H. Poincaré», *10* (1), 89–114.

[2] P. DEHEUVELS – «Comptes Rendus, Acad. Sci. Paris», *278*, Ser. A, 513–516.

[3] A. RENYI – *Calcul des Probabilités*, Dunod.

[4] P. DEHEUVELS – «Comptes Rendus, Acad. Sci. Paris», *278*, Ser. A, 989–992 (pour une bibliographie plus exhaustive, consulter (1), (2), (4)).