ESAIM: Probability and Statistics December 1999, Vol. 3, p. 151–162
URL: http://www.emath.fr/ps/

A NEW LARGE DEVIATION INEQUALITY FOR U-STATISTICS OF ORDER 2

Jean Bretagnolle[1]

Abstract. We prove a new large deviation inequality with applications when projecting a density on a wavelet basis.

Résumé. Nous prouvons une inégalité de grandes déviations applicable à la projection d'une densité sur une base d'ondelettes.

AMS Subject Classification. 60F10, 62-xx.

Received July 2, 1998. Revised September 14, 1999.

1. Introduction and main result

Let F be some law on $\mathbb{R}$. When g is a measurable function from $\mathbb{R}^d$ to $\mathbb{R}$, $E(g)$ and $Var(g)$ denote expectation and variance with respect to the Probability $F^{\otimes d}$. Let f be a bounded and symmetric function from $\mathbb{R}^2$ to $\mathbb{R}$. Following *Arcones and Giné* [1], we construct its canonical projections: ξ and η being independent with law F

$$\begin{aligned} \pi_1 f(x) &= Ef(x,\eta) - Ef(\xi,\eta) \cdot \\ \pi_2 f(x,y) &= f(x,y) - Ef(x,\eta) - Ef(\xi,y) + Ef(\xi,\eta) \cdot \end{aligned} \tag{1}$$

Let ξ_i, $i = 1, 2, \cdots, n$ be a n-sample of F $(n \geq 2)$. We consider the U-statistics (without any normalisation)

$$\begin{aligned} U_n^{(2)}(f) &= \Sigma_{1 \leq i \neq j \leq n} f(\xi_i, \xi_j) \cdot \\ U_n^{(1)}(\pi_1 f) &= \Sigma_{1 \leq i \leq n} \pi_1 f(\xi_i) \cdot \\ U_n^{(2)}(\pi_2 f) &= \Sigma_{1 \leq i \neq j \leq n} \pi_2 f(\xi_i, \xi_j), \quad \text{thus} \\ U_n^{(2)}(f - Ef)) &= 2(n-1) U_n^{(1)}(\pi_1 f) + U_n^{(2)}(\pi_2 f) \cdot \end{aligned} \tag{2}$$

We are interested in a large deviation inequality for the latter U-statistic when f is centered and bounded.

First, if $|f| \leq c$ and $Ef^2 = \sigma^2$, the usual Bernstein type inequality is

$$P(U_n^{(2)}(f - Ef) \geq n(n-1)t) \leq \exp\left(-[n/2]t^2/\{2\sigma^2 + 2ct/3\}\right). \tag{AG1}$$

But we can consider $U_n^{(1)}(\pi_1 f)$ (which is a sum of i.i.d. $\mathbb{R}$-valued random variables) as the main part and it can be interesting to bound separately the second part $U_n^{(2)}(\pi_2 f)$. Now, as $\pi_2 f$ is canonical of order 2, if $|\pi_2 f| \leq c$

Keywords and phrases: Large deviations, U-statistics.

[1] URAD 0743, Université d'Orsay, bâtiment 425, 91405 Orsay Cedex, France; e-mail: Jean.Bretagnolle@math.u-psud.fr

and $\sigma^2 = E(\pi_2 f)^2$, there exist two constants c_1, c_2 such that

$$P(|U_n^{(2)}(\pi_2 f)| \geq (n-1)t) \leq c_1 \exp(-c_2 t/\{\sigma + c^{2/3} t^{1/3} n^{-1/3}\}) \tag{AG2}$$

((AG1) and (AG2) can be found in Arcones and Giné [1]).

The normalized version of inequality (AG2) can be restated

$$P(|U_n^{(2)}(\pi_2 f)|) \geq a_3 \text{ Max } (n\sigma x, c\sqrt{n} x^{3/2})) \leq a_4 \exp(-x). \tag{AG2'}$$

The aim of the paper is to give a new large deviation inequality (Th. 1 and Cor. 1 later). To every partition D we associate two functionnals $\|f\|_D$ and $v_D(F)$ (see Def. 4 later) such that

$$P(|U_n^{(2)}(\pi_2 f)| \geq a'_3 x \|f\|_D \text{ Max } (n\sqrt{v_D(F)}, 1)) \leq a'_4 \exp(-2\sqrt{x}) \tag{3'}$$

where all a_3,a'_3 are universal constants. For a comparison between the two inequalities, see the discussion after Theorem 2 *infra*.

We need now some definitions: let $(I_\lambda \mid \lambda \in D)$ be a Borelian partition of $\mathbb{R}$ finite or enumerable, where I_λ denotes the subset and its indicator. Let τ be a permutation of D. Its *graph* g_τ is $\{(\lambda, \tau(\lambda)) | \lambda \in D\}$, a subset of $D \times D$. A *collection* G is an enumerable set $(\tau_s \mid s \in G)$ of permutations such that $D \times D \subset \cup_{s \in G} g_{\tau_s}$. $\mathbb{H}$ is the family of collections. For any matrix $M = (a_{\lambda,\mu} \mid \lambda \in D;\ \mu \in D)$ we set

$$\|M\|_G = \Sigma_{i \in G} \text{ Sup }_{\lambda \in D} |a_{\lambda, \tau_i(\lambda)}| \text{ and } \|M\|_D = \text{ Inf }_{G \in \mathbb{H}} \|M\|_G.$$

Let f be a bounded real valued function defined on $\mathbb{R} \times \mathbb{R}$. We set

$$\begin{aligned} M(D,f) &= (a_{\lambda,\mu}) \text{ where } a_{\lambda,\mu} = \text{ Sup } (f^2(x,y) I_\lambda(x) I_\mu(y)) \\ \|f\|_D^2 &= \|M(D,f)\|_D \\ v_D(F) &= \Sigma_\lambda P(\xi \in I_\lambda)^2. \end{aligned}$$

Our main result is

Theorem 1. *There exists some constant C ($C = 80\pi$ holds) such that for every integer $k > 0$, for every partition D, for every symmetric f*

$$E(U_n^{(2)}(\pi_2 f)^{2k}) \leq C^{2k} \times (E\mathcal{N}^{2k})^4 \times \|f\|_D^{2k} \text{ Max } (n^{2k}(v_D(F))^k, n^2 v_D(F)).$$

where $\mathcal{N}$ denotes the standard normal distribution.

Corollary 1. *There exists some constant C such that for every partition D, for every symmetric f, for every $x > 0$*

$$P(|U_n^{(2)}(\pi_2 f)| \geq Cx \|f\|_D \text{Max } (n\sqrt{v_D(F)}, 1)) \leq \exp(6 - 2\sqrt{x}).$$

As we will see in the following *Discussion*,

Remark 1: The classical inequality (AG) is strictly better whenever $x \leq n\sigma^2/c^2$.
Remark 2: Nevertheless, our inequality can work when the classical one does not.
Remark 3: Finally, up to some logarithm, $\sqrt{x}$ is the best possible rate.

Main application: Let $(\Psi_1$,$\Psi)$be a wavelet and ε_λ, $\varepsilon_{\ell,\lambda}$ be the associated basis:

$$\varepsilon_\lambda(x) = \Psi_1(x+\lambda)/\|\Psi_1\|_2, \varepsilon_{\ell,\lambda}(x) = 2^{\ell/2}\Psi(2^\ell x + \lambda)/\|\Psi\|_2 (\lambda \in \mathbb{Z}, \ell \in \mathbb{N}).$$

Let p be some density of Probability on $\mathbb{R}$ equipped with Lebesgue measure dx, assumed to be square integrable in applications.

Let β_λ be $\int p(x)\varepsilon_\lambda(x)dx$ and $\gamma_{\ell,\lambda}$ be $\int \varepsilon_{\ell,\lambda}(x)p(x)dx$.

We set $p_L = \Sigma_\lambda \beta_\lambda \varepsilon_\lambda + \Sigma_{\ell\leq L}\Sigma_\lambda \gamma_{\ell,\lambda}\varepsilon_{\ell,\lambda}$ (the projection of p up to level of resolution L) and want to estimate its square norm

$$\theta_L \doteq \|p_L\|_2^2 = \Sigma_\lambda \beta_\lambda^2 + \Sigma_{\ell\leq L}(\Sigma_\lambda \gamma_{\ell,\lambda}^2)) \cdot \tag{4}$$

When $(\xi_i \mid 1 \leq i \leq n)$ is a n-sample with density p $(n \geq 2)$, the empirical estimators of coefficients are $\beta_{\lambda,n} = \Sigma_i \varepsilon_\lambda(\xi_i)/n$, $\gamma_{\ell,\lambda,n} = \Sigma_i \varepsilon_{\ell,\lambda}(\xi_i)/n$. Let $p_{L,n}$ be

$$p_{L,n} = \Sigma_\lambda \beta_{\lambda,n}\varepsilon_\lambda + \Sigma_{\ell\leq L}\Sigma_\lambda \gamma_{\ell,\lambda,n}\varepsilon_{\ell,\lambda}.$$

The "natural estimator" for θ_L is

$$\Sigma_\lambda \beta_{\lambda,n}^2 + \Sigma_{\ell\leq L}(\Sigma_\lambda \gamma_{\ell,\lambda,n}^2).$$

But the latter has positive bias, and the unbiased estimator is

$$\widehat{\theta}_{L,n} = (n(n-1))^{-1}\Sigma_{1\leq i\neq j\leq n}\{\Sigma_\lambda \varepsilon_\lambda(\xi_i)\varepsilon_\lambda(\xi_j) + \Sigma_{\ell\leq L}(\Sigma_\lambda \varepsilon_{\ell,\lambda}(\xi_i)\varepsilon_{\ell,\lambda}(\xi_j))\} \cdot \tag{5}$$

Let $\Delta_{n,L} = \widehat{\theta}_{L,n} - \theta_L$. According to (5), $\Delta_{n,L}$ can be decomposed into canonical U-statistics in the following way: let Φ_1 be $\Psi_1/\|\Psi_1\|_2$ and let Φ be $\Psi/\|\Psi\|_2$,

$$f(x,y) = \Sigma_\lambda \Phi_1(x+\lambda)\Phi_1(y+\lambda),\ \ f_\ell(x,y) = \Sigma_\lambda \Phi(2^\ell x+\lambda)\Phi(2^\ell y+\lambda) \tag{3}$$

$$\delta_n = (2/n)U_n^{(1)}(\pi_1 f) + (1/n(n-1))U_n^{(2)}(\pi_2 f) \tag{4}$$

$$\delta_{\ell,n} = (2\cdot 2^\ell/n)U_n^{(1)}(\pi_1 f_\ell) + (2^\ell/n(n-1))U_n^{(2)}(\pi_2 f_\ell). \tag{6}$$

Then $\Delta_{n,L} = \delta_n + \Sigma_{\ell\leq L}\delta_{\ell,n}$. In the decomposition above, the sum of U-statistics of order 1 is equal to the part up to level of resolution L of $\int p(p_n - p)$. It can be bounded in Probability by classical Bernstein's inequality. The control of each U-statistic of order 2 will be performed by our inequality. This is very useful, either to estimate θ_L [3], or in model selection: in this problem, the authors consider a wide family of finite dimensional projections. Our study is quite general, but we observe that if Ψ_1 and Ψ are with compact support, for every L the family $(\beta_\lambda, \alpha_{\ell,\lambda} \mid \ell \leq L)$ is in fact with finite dimension, and hypothesis (H) *infra* holds. Thus our result can be used in adaptive estimation of quadratic functionals in a density model (see for example [4] where white noise is treated), where the theory needs good bounds up to $2^\ell = O(n^2)$. These bounds cannot be obtained by classical Hoeffding's bounds (see the chapter "Discussion").

We assume in the whole paper

Hypothesis (H).

$$|\Phi(x)| \leq \Sigma_{u\in\mathbb{Z}}\omega_u 1_{u\leq x<u+1} \text{ with } \Sigma_{u\in\mathbb{Z}}\omega_u < \infty$$
$$\text{and we set } M^2(\Phi) = \Sigma_{(u,v,w)\in\mathbb{Z}^3}\ \omega_u\omega_v\omega_{u+w}\omega_{v+w}. \tag{H}$$

In the particular case when the law has density p, for the normalised U-statistic, we have

Theorem 2. *We assume that the function Φ with $\|\Phi\|_2 = 1$ satisfies hypothesis (H), and that the law F has a density p with $\|p\|_2 = (\int p^2(x)dx)^{1/2} < \infty$. Set*

$$\|p\|_{2,\ell} = 2^{\ell/2}\left[\Sigma_{k\in\mathbb{Z}}\left(\int_{k2^{-\ell}}^{(k+1)2^{-\ell}} p(x)dx\right)^2\right]^{1/2}$$

$$\delta^{(2)}_{\ell,n} = (2^\ell/n(n-1))U^{(2)}_n(\pi_2 f_\ell) \text{ where } f_\ell(x,y) = \Sigma_\lambda \Phi(2^\ell x+\lambda)\Phi(2^\ell y+\lambda)$$
$$Z_{\ell,n} = \sqrt{n(n-1)}2^{-\ell/2}\delta^{(2)}_{\ell,n}.$$

Then, if $n\|p\|_{2,\ell} \geq 2^{\ell/2}$, *we have, for* C *as in Theorem 1*

$$P(|Z_{\ell,n}| \geq 2C \times M(\Phi) \times x) \leq \exp(6-2\sqrt{x})\cdot$$

In this formula, $Var(Z_{\ell,n})$ *does not depend on* n. *Moreover, there exists positive and finite constants* $a(\Phi)$, $b(\Phi)$ *depending only on* Φ *such that*

$$\lim_{\ell\to\infty} \|p\|_{2,\ell} = \|p\|_2\cdot$$
$$\lim_{\ell\to\infty} \|\pi_2 f_\ell\|_\infty = b(\Phi) \text{ and } a(\Phi)\|p\|_2^2 \leq \liminf_{\ell\to\infty} Var(Z_{\ell,n})\cdot$$
$$\|\pi_2 f_\ell\|_\infty \leq 4M(\Phi) \text{ and } Var(Z_{\ell,n}) \leq \|p\|_2^2 M^2(\Phi)\cdot$$

Remark 4: Such a result is interesting only if it works, for given n, ℓ, *uniformly* for large classes of densities. Obviously we need some uniform control of $\|p\|_2$, but this is not sufficient in view of condition $n\|p\|_{2,\ell} \geq 2^{\ell/2}$.

Thus we need some extra condition. If for example we assumme that the support of p is contained in some interval $[x, x+M]$ we get $\|p\|_{2,\ell} \geq 1/(M+1)$ and the bound works if $n^2 \geq (M+1)2^\ell$.

I would like to thanks the two anonymous referees whose remarks and suggestions have much improved the presentation.

2. Discussion

a) About Corollary 1.

We consider inequality stated in Corollary 1 and assertion (AG2"). We assumme $\|f\|_\infty = 1$. Without further knowledge about the law F, we can only bound $\sigma^2 = Var(\pi_2(f))$ by Ef^2 and $c = \|\pi_2 f\|_\infty$ by 4. Up to some change of a_3, (AG2') is restated

$$P\left(|U^{(2)}_n(\pi_2 f)| \geq a_3 \text{ Max } \left(nx\sqrt{Ef^2}, \sqrt{n}x^{3/2}\right)\right) \leq a_4 \exp(-x).$$

On the other hand, we have the obvious inequality

$$Ef^2 \leq \|f\|_D^2 v_D(F).$$

Proof of Remark 1: Let us assumme $x \leq n\sigma^2$. Thus $\text{Max}(n\sigma x, c\sqrt{n}x^{3/2}) = n\sigma x \leq n\|f\|_2 x \leq x\|f\|_D$ Max $(n\sqrt{v_D(F)}, 1)$ and, up to constants (AG2') is *always* better than the bound of Corollary 1 whenever $x \leq n\sigma^2$.

Nevertheless, our inequality provides a possible alternative if there exists some partition D such that $n^2 v_D \geq 1$ and $\|f\|_D^2 v_D(F) \approx Ef^2$ (up to constant). We exhibit two extremal cases in the case when the law is the uniform one:

1) $f(x,y) = \mathbb{1}_{0\leq x<p; 0\leq y<p}$, where $0 < p < 1$. We set $q = 1-p$. We have $\|f\|_\infty = 1$, $Ef^2 = p^2$. Choosing $I_1 = (0,p)$ $I_\lambda =)p+(\lambda-2)q/K, p+(\lambda-1)q/K)$ for $1 < \lambda \leq K+1$, we get $v_D = p^2 + q^2/K$ and, for a convenient choice of G, $\|f\|_D^2 = 1$. Moreover $Var(\pi_2 f) = p^2q^2$ and $\|\pi_2 f\|_\infty = max(p^2, q^2)$.

When p is small, up to constants, in the classical inequality we can use either precise true parameters or rough estimates ($\|f\|_\infty$ and Ef^2 for $\|\pi_2 f\|_\infty$ and $Var(\pi_2 f)$) and we get

$$P(|U^{(2)}_n(\pi_2 f)| \geq a \text{ Max } (nxp, \sqrt{n}x^{3/2})) \leq a_4 \exp(-x) \qquad \text{(AG2')}$$

Our result is

$$P(|U^{(2)}_n(\pi_2 f)| \geq Cx \text{ Max } (np, 1)) \leq \exp(6-2\sqrt{x}). \qquad \text{(Cor 1)}$$

In this setting, we can assume $x > 1$, n large.

If $1 < x < np^2$, the classical inequality is better, but (up to constants) only with respect to the exponent of x.

A contrario the classical result does not work in the case when $np^2 = o(1)$ but not our one provided that np is large, and this justify the Remark 2.

We will see that it is a quite general result in the *main application*.

Remark 3: Assuming $np = 1$ and denoting f_n the corresponding function, when $n \to \infty$, it is easy to prove that $U_n^{(2)}(\pi_2 f_n)$ converges in law to $Y^2 - 3Y + 1$ where the law of Y is the Poisson law with parameter 1. Thus

$$\liminf_{x\to\infty} \lim_n \{-\operatorname{Log} P(U_n^{(2)}(\pi_2 f_n) \geq x)/\sqrt{x} \operatorname{Log}(\sqrt{x})\} \geq 1$$

proving that the power 1/2 is the best possible.

2) Let be $g(x,y) = 1\!\!1_{0\leq x\leq 1;\ 0\leq y\leq 1} - 1\!\!1_{p\leq x\leq 1-p;\ p\leq y\leq 1-p}$ (p small) and $\varepsilon(x) = 1\!\!1_{x\leq 1/2} - 1\!\!1_{1/2<x}$ and finally $f(x,y) = \varepsilon(x)\varepsilon(y)g(x,y)$. Then, for the uniform law, $f = \pi_2 f$ and $Var(\pi_2 f) = Var(f) = 4p(1-p)$. Obviously, for every D, $\|f\|_D^2 = |D|$ and $v_D(F) \geq 1/D$, thus $\|f\|_D \max(n\sqrt{v_D(F)}, 1) \geq n$ (and n is obtained by the partition with *one* element).

The classical inequality gives for some universal a

$$P(|U_n^{(2)}(\pi_2 f)| \geq a \operatorname{Max}(n\sqrt{p}x, \sqrt{n}x^{3/2})) \leq a_4 \exp(-x)$$

and, *whatever be p*, our inequality provides only

$$P(|U_n^{(2)}(\pi_2 f)| \geq Cxn) \leq \exp(6 - 2\sqrt{x})$$

a very poor result!

b) *About the main application*:

b1) We consider firstly the case of the Haar basis ($\Phi(x) = 1\!\!1_{0\leq x<1}$ or $\Phi(x) = 1\!\!1_{0\leq x<1/2} - 1\!\!1_{1/2\leq x<1}$), with uniform law on the interval $]0,1]$. In the first case, at the level ℓ, setting $D = ([\lambda 2^{-\ell}, (\lambda+1)2^{-\ell}[\mid \lambda \in \mathbb{Z})$, $f_\ell = \Sigma_{\lambda\in\mathbb{Z}} \Phi(2^\ell x + \lambda)\Phi(2^\ell y + \lambda)$, we have $M(\Phi) = \|f_\ell\|_D = 1$, $\|\pi_2(f_\ell)\|_\infty = (1 - 2^{-\ell})$, $v_D(F) = 2^{-\ell}$ and $Var(\pi_2(f_\ell)) = (1-2^{-\ell})^2 2^{-\ell}$.

Thus for every $\ell \geq 1$ we have

$$2^\ell v_D(F) = \|f\|_D = 1.$$

$$1/2 \leq Var(Z_{\ell,n}) = 2^\ell Var(\pi_2 f_\ell) \leq 1.$$

$$1/2 \leq \|\pi_2 f_\ell\|_\infty \leq 1$$

and, whenever $2^\ell \leq n^2$, Theorem 2 provides

$$P\left(|Z_\ell|/\sqrt{Var(Z_{\ell,n})} \geq 2Cx\right) \leq \exp(6 - 2\sqrt{x}).$$

The classical one provides

$$P\left(|Z_\ell|/\sqrt{Var(Z_{\ell,n})} \geq a_3(x + (2^\ell/n)^{1/2}x^{3/2})\right) \leq a_4 \exp(-x)$$

and does not work if $n = o(2^\ell)$.

Remark 5: Massart (private communication) thinks that using Talagrand's inequality the best possible bandwith is $2^\ell = O(n^{3/2})$ and this is the principal motivation of this work.

b2) Finally, let $\kappa^2_{n,\ell}$ the chi-square (with $2^\ell - 1$ degrees of freedom) associated to the partition: with $N_\lambda = \Sigma_{1\le i\le n}\mathbb{1}_{\lambda 2^{-\ell}<\xi_i\le(\lambda+1)2^{-\ell}}$

$$\kappa^2_{n,\ell} = \Sigma_\lambda(N_\lambda - EN_\lambda)^2/EN_\lambda.$$

The centered and normalised $\kappa^2_{n,\ell}$ is equal to $Z_\ell/\sqrt{Var(Z_{\ell,n})}$. Thus our result provides a large deviation inequality for $\kappa^2_{n,\ell}$ even in the case when $n \approx 2^{\ell/2}$. Remark that the mean number of visits EN_λ can be $O(1/n)$!

The second case is the best possible: for every ℓ we have $\|f_\ell\|_\infty = 1$, f_ℓ is canonical and $\|f\|^2_D v_D(F) = Var(Z_\ell)$.

b3) We consider now the general case in the main application:

Using the final assertions of Theorem 2, we see that we have asymptotically the same conclusion as in the case of b1: whenever ℓ is large, our Z, up to constants depending only on Φ and the law p, is the normalised U-statistic of order 2 corresponding to some canonical function the Sup norm of which is equivalent to 1.

Thus we get a large deviation inequality which cannot be obtained using the classical result for $2^{\ell/2} \ll n \ll 2^\ell$.

3. Proofs

The proof is based on *De la Peña's* inequalities [2]. As all bounds are continuous with respect to $\|f\|_D$, $v_D(F)$, it suffices to prove that, if G is a collection such that $\Sigma_{i\in G}$ Sup $_{\lambda\in D}$ Sup $(f^2(x,y)I_\lambda(x)I_{\tau_i(\lambda)}(y)) = 1$, then

$$E(V_n^{2k}) \le C^{2k} \times (\mathbb{E}\mathcal{N}^{2k})^4 \times \text{ Max } (n^2 v, (n^2 v)^k) \tag{8}$$

where, to simplify notations, we set

$$V_n = \Sigma_{1\le i\ne j\le n}\pi_2 f(\xi_i,\xi_j) \text{ and } v = v_D(F)\cdot \tag{9}$$

1: Symmetrization

Let $\varepsilon_i, \varepsilon_i', \mathcal{N}_i, \mathcal{N}_i', \xi_i, \eta_i$ be six independent n-samples: the common law of ε's is the law of the centered sign, the common law of the $\mathcal{N}$'s is the normal $\mathcal{N}(0,1)$, the common law of ξ's and η's is the law F.

Using the first Theorem of *De la Peña*, as $\pi_2 f$ is canonical, we get:

For every Γ even, increasing on $\mathbb{R}^+$ and convex

$$E\Gamma(V_n) \le E\Gamma(4\Sigma_{1\le i,j\le n; i\ne j}\pi_2 f(\xi_i,\eta_j))\cdot$$

Using the classical symmetrization inequalities (see [2] again), we have

$$E\Gamma(V_n) \le E\Gamma(16\Sigma_{1\le i,j\le n; i\ne j}\varepsilon_i\varepsilon_j'\pi_2 f(\xi_i,\eta_j)).$$

As the ε_i, ε_i' can be viewed as conditional expectations of $\sqrt{\pi/2}\mathcal{N}_i$, $\sqrt{\pi/2}\mathcal{N}_i'$, using convexity again we get

$$E\Gamma(V_n) \le E\Gamma(8\pi\Sigma_{1\le i,j\le n; i\ne j}\mathcal{N}_i\mathcal{N}_j'\pi_2 f(\xi_i,\eta_j)).$$

We set now

$$W_n = \Sigma_{1\le i,j\le n; i\ne j}\pi_2 f(\xi_i,\eta_j)^2. \tag{10}$$

In law, $(\Sigma_{1\le i,j\le n; i\ne j}\mathcal{N}_i\mathcal{N}_j'\pi_2 f(\xi_i,\eta_j))^2 = \mathcal{N}^2\Sigma_i(\Sigma_j\pi_2 f(\xi_i,\eta_j)\mathcal{N}_j)^2)$ or $\mathcal{N}^2\Sigma_k\lambda_k\mathcal{N}_k^2$, where $\Sigma_k\lambda_k = W_n$ with $\lambda_k \ge 0$, thus, by convexity:

Lemma 1. *For every $k \in \mathbb{N}$, we have*

$$EV_n^{2k} \le (8\pi)^{2k}(E\mathcal{N}^{2k})^2 EW_n^k.$$

2: Bounds for functions

k is a natural integer. The current indexes i,j of the sample belongs to $[1,n]$. The current s belongs to G, other current indexes as λ, μ,.. belong to D. We have $|f| \leq \sqrt{h}$ where

$$h(x,y) = \Sigma_{\lambda,\mu} a_{\lambda,\mu} I_\lambda(x) I_\mu(y). \tag{11}$$

Thus $|\pi_2 f(x,y)| \leq \int\int(\sqrt{h}(x,y) + \sqrt{h}(x,t) + \sqrt{h}(z,y) + \sqrt{h}(z,t))F(dz)F(dt)$ and finally

$$\pi_2 f(x,y)^2 \leq 4\int(h(x,y) + h(x,t) + h(z,y) + h(z,t))F(dz)F(dt)\cdot$$

Thus by convexity

Lemma 2. *For h defined in (11) and natural integer k we have*

$$EW_n^k \leq (16)^k E((\Sigma_{1\leq i\neq j\leq n} h(\xi_i,\eta_j))^k). \tag{12}$$

3: Bounds for moments

We define the numbers of visits of I_λ by each of the two samples as

$$X_\lambda = \Sigma_{1\leq i\leq n} I_\lambda(\xi_i) \ \text{ and } \ Y_\lambda = \Sigma_{1\leq i\leq n} I_\lambda(\eta_i)\cdot \tag{13}$$

Let τ be the current permutation of G and π_τ be $\text{Sup}_{\ \lambda} a_{\lambda,\tau(\lambda)}$. We have obviously

$$\Sigma_{1\leq i\neq j\leq n} h(\xi_i,\eta_j) \leq \Sigma_\tau \pi_\tau(\Sigma_\lambda X_\lambda Y_{\tau(\lambda)})\cdot$$

As $\Sigma_\tau\pi_\tau = 1$, by convexity again and (13) we obtain

$$EW_n^k \leq (16)^k \ \text{Sup}_{\ \tau} E((\Sigma_\lambda X_\lambda Y_{\tau(\lambda)})^k)\cdot$$

Appendix 1 contains the proof of the main technical result, namely:

Lemma 3. *With previous notations, for every τ, for every integer $k \geq 1$, we have*

$$E(\Sigma_\lambda X_\lambda Y_{\tau(\lambda)})^k \leq 6^k \text{Max } (n^2 v, (n^2 v)^k)(E\mathcal{N}^{2k})^2.$$

Collecting the previous bounds, proof of Theorem 1 is achieved. □

4: Proof of Corollary 1

A) We assume that $n^2 v \geq 1$. Let X be $U_n^{(2)}(f)/C\|f\|_D n\sqrt{v}$.
Appendix 2 contains the proof of the quite obvious

Lemma 4. *If for every natural integer k we have $\mathbb{E}X^{2k} \leq (\mathbb{E}\mathcal{N}^{2k})^4$, then*

$$\mathbb{P}(|X| \geq x) \leq \exp\ \left(6 - 2\sqrt{x}\right).$$

B) Now, if $n^2 v \leq 1$, let Y be $U_n^{(2)}(f)/C\|f\|_D$. For the same reason we have

$$\mathbb{P}(|Y| \geq x) \leq \ \exp\ \left(6 - 2\sqrt{x}\right).$$

This achieves the proof. □

5: Proof of Theorem 2

Let Δ be some positive integer and D be the partition $(I_{\lambda,D} = (\lambda/\Delta, (\lambda+1)/\Delta(|\ \lambda \in \mathbb{Z})$. In what follows indexes λ, u, v, s belong to $\mathbb{Z}$.

Let $p_{\lambda,D}$ be $P(\xi \in I_{\lambda,D})$ and p_D be the density $p_D(x) = \Sigma \Delta p_{\lambda,D} I_{\lambda,D}$. Expectation with respect to p_D is denoted E_D. We have

$$\lim_{\Delta\to\infty} \|p_D - p\|_2 = 0,\ \Delta v_D = \|p_D\|_2^2 \text{ and thus } \lim_{\Delta\to\infty} \Delta v_D = \|p\|_2^2.$$

We set $\Phi(x) = \Sigma_{u\in\mathbb{Z}} \gamma_u(x-u) 1\!\!1_{u\le x<u+1}$, where the support of γ_u is included in $[0,1[$ (thus defining the γ_u's); we have $\|\gamma_u\|_\infty \le \omega_u$. As $\|\Phi\|_2 = 1$, there exists some u_o with $\|\gamma_{u_o}\|^2 > 0$.

We set $f_D(x,y) = \Sigma_\lambda \Phi(\Delta x + \lambda)\Phi(\Delta y + \lambda)$.

a) *Bounds for* $Ef_D^2(\xi,\eta)$ *and* $\|f_D\|_D$:

We begin by bounding from below the quantity $Ef_D^2(\xi,\eta)$.

We have $Ef_D^2(\xi,\eta) = \Sigma_{\lambda,\mu}(E\Phi(\Delta\xi+\lambda)\Phi(\Delta\eta+\mu))^2 \ge \Sigma_\lambda (E\Phi^2(\Delta x+\lambda))^2$, thus $Ef_D^2(\xi,\eta) \ge \Sigma_\lambda (E\gamma_{u_o}^2(\Delta\xi+\lambda-u_o)I_{u_o-\lambda,D})^2$. A classical computation gives $(EgI_{\mu,D})^2 - (E_D g I_{\mu,D})^2 \ge -2\|gI_{\mu,D}\|_\infty \|gI_{\mu,D}\|_2 p_{\mu,D} (\int_{I_{\mu,D}} (p-p_D)^2 dx)^{1/2}$ then $Ef_D^2(\xi,\eta) \ge \Sigma_\lambda (E_D \gamma_{u_o}^2(\Delta\xi+\lambda-u_o)I_{u_o-\lambda,D})^2 - 2\omega_{u_o}(\|\gamma_{u_o}^2\|_2/\sqrt{\Delta})\sqrt{v_D}\|p-p_D\|_2$. As $(E_D\gamma_{u_o}^2 (\Delta\xi+\lambda-u_o)I_{u_o-\lambda,D})^2 = p_{u_o-\lambda,D}^2 \|\gamma_{u_o}\|_2^4$, we get

$$\Delta E f_D^2(\xi,\eta) \ge \Delta v_D \|\gamma_{u_o}\|_2^4 - 2\omega_{u_o}^3 \|p-p_D\|_2 \sqrt{\Delta v_D} \text{ and}$$

$$\liminf_{\Delta\to\infty} \Delta E f_D^2(\xi,\eta) \ge \|p_2\|^2 a(\Phi) := \|p\|_2^2 \|\gamma_{u_o}\|_2^4 > 0\,\cdot \tag{15}$$

On the other hand, $|f_D(x,y)| \le g(x,y) := \Sigma_\lambda |\Phi(\Delta x+\lambda)\Phi(\Delta y+\lambda)|$. Using (H), we have $g(x,y) \le \Sigma_{\lambda,u,v}\omega_{u+\lambda}\omega_{v+\lambda} I_u(x)I_v(y)$.

Setting $\sqrt{a_s} = \Sigma_u \omega_u \omega_{u+s}$, $g(x,y) \le \Sigma_{\lambda,s}\sqrt{a_s} I_\lambda(x) I_{\lambda+s}(y)$, then

$$g^2(x,y) \le \Sigma_s a_s (\Sigma_\lambda I_\lambda(x) I_{\lambda+s}(y))\,\cdot \tag{16}$$

But $\Sigma_s a_s = \Sigma_{u,v,s}\omega_u\omega_v\omega_{u+s}\omega_{v+s} = M(\Phi)^2 \ge \Sigma_{u,v}\omega_u\omega_v\omega_u\omega_v = (\Sigma_u \omega_u^2)^2 \ge 1$ because $\|\Phi\|_2 = 1$. Taking for G the collection of $\lambda{:}\to \lambda+s$, we get

$$\|f_D\|_D^2 \le M^2(\Phi) \text{ and } M^2(\Phi) \ge 1\,\cdot \tag{17}$$

We recall the obvious upper bound

$$Ef_D^2 \le v_D M^2(\Phi) \le \|p\|_2^2 M^2(\Phi)/\Delta\,\cdot \tag{18}$$

b) *Bounds for* $\|f_D\|_\infty$:

Obviously, $\|f_D\|_\infty$ does not depend on D, and is less than $M(\Phi)$:

$$\text{There exists some } b(\Phi) \text{ with } 0 < b(\Phi) = \|f_D\|_\infty \le M(\Phi). \tag{19}$$

d) *Bounds for* $\|\pi_2 f_D\|_2$ *and* $\|\pi_2 f_D\|_\infty$:

As for Δ large, $\text{Sup}_\lambda p_{\lambda,D} = o(1/\sqrt{\Delta})$ we have $|E\Phi(\Delta\eta+\lambda)| = o(1/\sqrt{\Delta})$. Thus $\|Ef(x,\eta)\|_\infty := \text{Sup}_x |Ef(x,\eta)| \to 0$. We have $\pi_2 f_D(x,y) = f_D(x,y) - Ef(x,\eta) - Ef(\xi,y) + Ef(\xi,\eta)$, and asymptotically we have

$$\lim_{\Delta\to\infty} \|\pi_2 f_D\|_\infty = b(\Phi) \text{ (and obviously by (19) } \|\pi_2 f_D\|_\infty \le 4M(\Phi). \tag{20}$$

We have $E(f_D)^2 \ge Var(\pi_2 f_D) \ge \Sigma_\lambda (Var\Phi^2(\Delta\xi+\lambda))^2 = \Sigma_\lambda (E(\Phi^2(\Delta\xi+\lambda) - E(\Phi(\Delta\xi+\lambda)^2)^+)^2$. Using the fact that $\Delta E(\Phi(\Delta\xi+\lambda)^2$ goes uniformly to 0 and (15), we get

$$a(\Phi)\|p\|_2^2 \le \liminf_{\Delta\to\infty} \Delta Var(\pi_2 f_D) \tag{21}$$

$$\Delta Var(\pi_2 f_D) \leq \|p\|_2^2 M^2(\Phi).$$

e) *Proof of Theorem 2*

We set now $\Delta = 2^\ell$. Using $nC/\sqrt{n/(n-1)} \leq 2C$, $n^2 v_D \geq 1$, $\|f_D\|_D \leq M(\Phi)$ and $\Delta v_D \leq \|p\|_2^2$, Corollary 1 gives the exponential upper bound.

Appendix 1: Proof of Lemma 3

In this appendix where partition D is fixed, we use notation

$$v_D = v,\ p_\lambda = P(\xi \in I_\lambda)\ ,\ \eta_\lambda = \mathbb{E}(X_\lambda) = np_\lambda \ \text{ thus } \ \Sigma \eta_\lambda^2 = n^2 v\cdot$$

We recall that τ is some permutation of D, and that λ is the current point of D.

We consider two laws on $\mathbb{N}^D \times \mathbb{N}^D$, the current point of which is $(\mathbf{X}, \mathbf{Y})$. In all cases, $\mathbf{X}$ and $\mathbf{Y}$ are independent with the same law.

In the first case, the law of $\mathbf{X}$ is $\mathcal{M}(n, \mathbf{p})$ $(n \geq 2)$, the Multinomial where $\mathbf{p} = (p_\lambda)$, with associated expectation E.

In the second one, the X_λ are independent, with Poisson law, and mean value $E_\eta(\mathbf{X}) = \eta$, where $\eta = (\eta_\lambda) = n\mathbf{p}$. The associated expectation is $\mathbb{E}_\eta$.

We consider the mapping U from $\mathbb{N}^D \times \mathbb{N}^D$ to $\mathbb{R}$ defined by $U(\mathbf{X}, \mathbf{Y}) = \Sigma_\lambda X_\lambda Y_{\sigma(\lambda)}$. We will first prove that for every positive integer k $E(U^k)$ is less than $E_\eta(U^k)$ and then furnish an upper bound for this moment.

1: Reduction to the Poisson case

In what follows, E_μ denotes the expectation associated to the Poisson law with parameter μ. $X^{[k]}$is the Polynomial $X(X-1)\cdot\cdot(X-k+1)$, for which $E_\mu X^{[k]} = \mu^k$.

Definitions: A mapping ψ from $\mathbb{N}$ to $\mathbb{N}$ is strongly positive if

$$\psi(X) = \Sigma_k a_k X^{[k]},\ \text{with } a_k \geq 0 \text{ for every } k. \tag{d1}$$

A mapping Ψ from $\mathbb{N}^D$ to $\mathbb{N}$ is strongly positive if there exist some enumerable I, a family $(\psi_{\lambda,i} \mid \lambda; i \in I)$, a family $(a_i \mid i \in I)$, where each $\psi_{\lambda,i}$ is strongly positive and each a_i is positive, such that

$$\Psi(\mathbf{X}) = \Sigma_i a_i \Pi_\lambda \psi_{\lambda,i}(X_\lambda)\ \cdot \tag{d2}$$

Lemma 5. *If Ψ is strongly positive, then*

$$E(\Psi(\mathbf{X})) \leq E_\eta(\Psi(\mathbf{X})). \tag{a1}$$

For every k, X^k and $\{X(X-1)\}^k$ are strongly positive. Moreover for $k > 0$, we have

$$E_\mu\{X(X-1)\}^k \leq \text{Max } (\mu^{2k}, \mu^2) \times (E\mathcal{N}^{2k})^2. \tag{a2}$$

Remark: The upper bound in (a2) is increasing of k and μ $(k \geq 1$ and $\mu > 0)$.

Proof of (a1): By d2, it suffices to prove the formula when $\Psi(\mathbf{X}) = \Pi X_\lambda^{[k_\lambda]}$. For such a Ψ, $E\Psi(\mathbf{X}) = 0$ if $\Sigma_\lambda k_\lambda > n$, and $(n!/(n - \Sigma_\lambda k_\lambda)!)\Pi_\lambda p_\lambda^{k_\lambda}$ else, obviously less than $\Pi_\lambda (np_\lambda)^{k_\lambda} = E_\eta \psi(\mathbf{X})$.

Proof of (a2): The fact that X^k is strongly positive (in our sense) is well-known.

Let T_k be $\{X(X-1)\}^k$. T_1 is $X^{[2]}$ and $E_\mu T_1 = \mu^2$. Assume $k > 1$; with $x = X - 2$, $T_k = X^{[2]}\{(x+2)(x+1)\}^{k-1}$. But $\{(x+2)(x+1)\}^{k-1}$ is polynomial with respect to x, with positive coefficients, thus strongly positive with respect to x: $\{(x+2)(x+1)\}^{k-1} = \Sigma_{0\le j\le 2k-2}\gamma_{j,k}x^{[j]}$; finally $X^{[2]}x^{[j]} = X^{[j+2]}$ and T^k is strongly positive. Moreover, $E_\mu T_k = E_\mu \Sigma_{0\le j\le 2k-2}\gamma_{j,k}X^{[2+j]} = \mu^2\Sigma_{0\le j\le 2k-2}\gamma_{j,k}EX^{[j]} = \mu^2 E_\mu\{(X+2)(X+1)\}^{k-1}$.

Let g_k be $g_k(X) = T_k(X)\mathbb{1}_{X>1}$. g_k is convex, and if X is a Poisson r.v, almost surely $g_k(X) = T_k$. For $k=1$, $E_\mu g_1(X) = \mu^2$; let k be >1. We have obtained $E_\mu\{X(X-1)\}^k = \mu^2 E_\mu g_{k-1}(X+2)$. Let Y be independent of X, Poisson with parameter 2; by Jensen, conditionally on $X=x$, $g_{k-1}(x+2) \le E_2 g_{k-1}(x+Y)$, thus, as the law of $X+Y$ is Poisson with parameter $2+\mu$, we get $E_\mu\{X(X-1)\}^k \le \mu^2 E_{\mu+2}g_{k-1}(X)$, thus, recursively

$$\text{if } k>0, \text{ then } E_\mu\{X(X-1)\}^k \le [\mu(\mu+2)\cdots(\mu+2k-2)]^2.$$

The product $\mu(\mu+2)\cdots(\mu+2k-2)$ is bounded by $\text{Max}\,(\mu^k,\mu)\times(1\cdot3\cdot5\cdots(2k-1)) = \text{Max}\,(\mu^k,\mu)\times E\mathcal{N}^{2k}$ and the proof is achieved for a2.

Now we return to the proof. U^k being a sum with positive coefficients of products of powers of the almost surely positive X's and the Y's is obviously strongly positive with respect to the X's and Y's; by independence and (a1), we obtain for every $k\ge0$:

$$EU^k \le E_\eta U^k\,\cdot \tag{a3}$$

2: The Poisson case

For every pair x,y of natural integers, we have easily

$$xy \le x(x-1)+y(y-1)+\mathbb{1}_{x=1}\times\mathbb{1}_{y=1}\,\cdot \tag{a4}$$

Let us define now

$$\begin{aligned} &Z = \Sigma_\lambda X_\lambda(X_\lambda-1),\ Z' = \Sigma_\lambda Y_\lambda(Y_\lambda-1)\\ &T = \Sigma_\lambda \mathbb{1}_{X_\lambda=1}\times\mathbb{1}_{Y_{\tau(\lambda)}=1}\end{aligned} \tag{d3}$$

Using the fact that τ is a permutation, by a4 we have $U \le Z+Z'+T$, then, as the laws of Z and Z' are the same

$$E_\eta U^k \le 3^k\ \text{Max}\,(E_\eta Z^k, E_\eta T^k)\,\cdot \tag{a5}$$

3: Bound for the first term

As $E_\eta Z = n^2 v$, we assume that $k>1$. We set

$$E(k,\mu) = \text{ Max}\,(\mu^{2k},\mu^2)(E\mathcal{N}^{2k})^2\ if\ k>0\ and\ 1\ else\cdot$$

By a2, we have

$$E_\eta Z^k \le \Sigma_{k_\lambda\ge0;\Sigma_\lambda k_\lambda=k}\{k!/\Pi_\lambda k_\lambda!\}\Pi_\lambda E(k_\lambda,\eta_\lambda)\,\cdot \tag{a6}$$

First case: If for each λ, $\eta_\lambda\ge1$, then, as $\Pi_\lambda E\mathcal{N}^{2k_\lambda} \le E\mathcal{N}^{2\Sigma_\lambda k_\lambda}$, we have

$$E_\eta Z^k \le (E\mathcal{N}^{2k})^2\Sigma_{k_\lambda\ge0;\Sigma_\lambda k_\lambda=k}\{k!/\Pi_\lambda k_\lambda!\}\eta_\lambda^{2k_\lambda} \tag{5}$$

thus, in the first case, for every $k\ge0$

$$E_\eta Z^k \le (n^2 v)^k(E\mathcal{N}^{2k})^2\,\cdot \tag{A1}$$

Second case: For each λ, $\eta_\lambda \leq 1$:

Let A be a non-void subset of $[0, D[\cap\mathbb{Z}$, and $M(A,k)$ be the subset of $\mathbb{N}^A$ given by $(k_i \mid i \in A;\ k_i > 0\ for\ each\ i \in A;\ \Sigma_{i\in A} k_i = k)$. We set

$$S(A,k) = \Sigma_{M(A,k)}\{k!/\Pi_i k_i!\}\Pi_i E(k_i, \eta_i)\cdot$$

The general term of $S(A,k)$ is $\{k!/\Pi_i k_i!\}\Pi_i \eta_i^2 \Pi_i E(\mathcal{N}^{2k_i})^2$.

Let ν_j be $(E\mathcal{N}^{2j})^2/j!$ $(j \in \mathbb{N})$. Elementary computation gives

$$If\ 1 \leq j \leq k,\ \text{then}\ \nu_j\nu_k \leq \nu_{j-1}\nu_{k+1}. \tag{a7}$$

Thus the general term of $S(A,k)$ is bounded by $\Pi_i\eta_i^2 E(\mathcal{N}^{2k})^2$ (obtained outside of $M(A,k)$, when all k_i are 0 except one).

On the other hand, it is well-known that $|M(A,k)| = \binom{k-1}{a-1}$, where $a = |A|$. Finally

$$E_\eta(Z^k) \leq \Sigma_{\text{non void}} S(A,k) \leq (E(\mathcal{N}^{2k})^2 \times \{\Sigma_{a>0}\tbinom{k-1}{a-1}(\Sigma_{|A|=a}\Pi_{i\in A}\eta_i^2)\}\cdot$$

As $\Sigma_{|A|=a}\Pi_{i\in A}\eta_i^2 \leq (n^2v)^a$ and $\Sigma_{a>0}\binom{k-1}{a-1} = 2^{k-1}$, we get in the second case, for any $k \geq 1$

$$E_\eta Z^k \leq 2^{k-1}\ \text{Max}\ (n^2v, (n^2v)^k)(E\mathcal{N}^{2k})^2\ \cdot \tag{A2}$$

General case: We divide $[0, D[$ into two (non void) subsets:
$\Delta_1 = (\lambda \mid \eta_\lambda < 1)$ and $\Delta_2 = (\lambda \mid \eta_\lambda \geq 1)$, and set $v_i = \Sigma_{\lambda\in\Delta_i}\eta_\lambda^2$. Using (A1, A2), we obtain

$$E(Z^k) \leq (E\mathcal{N}^{2k})^2(n^2v_2)^k + \Sigma_{j>0}\tbinom{k}{j}2^{j-1}(E\mathcal{N}^{2j})^2(E\mathcal{N}^{2k-2j})^2\ \text{Max}\ (n^2v_1, (n^2v_1)^j)(n^2v_2)^{k-j}.$$

The latter bound is increasing of v_i, each bounded by v. Thus finally, in any case, for $k \geq 1$

$$E_\eta Z^k \leq 2^k\ \text{Max}\ (n^2v, (n^2v)^k)(E\mathcal{N}^{2k})^2\ \cdot \tag{A3}$$

4: Bound for the second term

We can bound $E_\eta \mathbb{1}_{X_\lambda=1} \times \mathbb{1}_{Y_{\tau(\lambda)}=1}$ by $b_\lambda^2 := \eta_\lambda\eta_{\tau(\lambda)}$. We take notations of *Second case* of previous paragraph. Setting now

$$S'(A,k) = \Sigma_{M(A,k)}\{k!/\Pi_i k_i!\}\Pi_i b_i^2,\ \text{we have}\ E_\eta T^k = \Sigma_{A\ \text{non void}} S'(A,k).$$

The current term of $S'(A,k)$ is bounded by $k!\Pi_i b_i^2$. Thus we obtain here

$$E_\eta T^k \leq k! 2^{k-1}\ \text{Max}\ (w, w^k)\ \text{where}\ w = \Sigma_\lambda b_\lambda^2.$$

As by Cauchy-Schwartz $w \leq n^2v$ and $k! \leq (E\mathcal{N}^{2k})^2$, we get again for $k \geq 1$

$$E_\eta T^k \leq 2^k\ \text{Max}\ (n^2v, (n^2v)^k)(E\mathcal{N}^{2k})^2\ \cdot \tag{A4}$$

Using (a3, a5, A3) and (A4), the proof is finished. □

Appendix2: Proof of Lemma 4

For $k \in \mathbb{N}$, we set $u_k = e^k E\mathcal{N}^{2k}(2k+1)^{-k}$, $r_k = u_{k+1}/u_k = e\{(2k+1)/(2k+3)\}^{k+1}$ and finally $\varphi(x) = 1 + (x+1)\ \text{Log}\ ((2x+1)/(2x+3))$ for $x \geq 0$, then $\varphi(k) = \text{Log}\ (r_k)$. We have

$$\varphi'(x) = \text{Log}\ ((2x+1)/(2x+3)) + 1/(2x+1) + 1/(2x+3).$$

$$\varphi''(x)/2 = 1/(2x+1) - 1/(2x+3) - 1/(2x+1)^2 - 1/(2x+3)^2 \leq 0.$$

As φ' goes to 0 when x goes to ∞, we have $\varphi' \geq 0$. As φ goes to 0 when x goes to ∞, $\varphi \leq 0$. Thus, for $k > 0$, $u_{k+1} \leq u_k \leq u_0 = 1$: we have proved that

$$\text{for every } k \in \mathbb{N}, E\mathcal{N}^{2k} \leq e^{-k}(2k+1)^k. \qquad *$$

Let us assumme that for every $k \in \mathbb{N}$, $EX^{2k} \leq (E\mathcal{N}^{2k})^4$. If $2k+3 \geq \sqrt{x} \geq 2k+1$, *via* Markov's inequality and assertion *, $P(|X| \geq x) \leq e^{-4k} \leq e^{-2\sqrt{x}+6}$. Then the result is proved for $x \geq 1$ and obvious for $0 \leq x \leq 1$.

References

[1] M.A. Arcones and E. Giné, Limit Theorems for U-processes. *Ann. Probab.* **21** (1993) 1494-1542.

[2] V. De la Peña, Decoupling and Khintchine's inequalities for U-statistics. *Ann. Probab.* **20** (1992) 1887-1892.

[3] B. Laurent, Efficient estimation of integral functionals of a density. *Ann. Statist.* **24** (1996) 659-681.

[4] B. Laurent and P. Massart, Adaptative estimation of a quadratic functional by model selection (1998) preprint.