Exponential Inequalities, with Constants, for U-statistics of Order Two

C. Houdré and P. Reynaud-Bouret

Abstract. A martingale proof of a sharp exponential inequality (with constants) is given for U-statistics of order two as well as for double integrals of Poisson processes.

1. Introduction

We wish in these notes to further advance our knowledge of exponential inequalities for U–statistics of order two. These types of inequalities are already present in Hoeffding seminal papers [6], [7] and have seen further development since then. For example, exponential bounds were obtained (in the (sub)Gaussian case) by Hanson and Wright [5], by Bretagnolle [1], and most recently by Giné, Latała, and Zinn [4] (and the many references therein). As indicated in [4], the exponential bound there is optimal since it involves a mixture of exponents corresponding to a Gaussian chaos of order two behavior, and (up to logarithmic factors) to the product of a normal and of a Poisson random variable and to the product of two independent Poisson random variables. These various behaviors can be obtained as limits in law of triangular arrays of canonical U-statistics of degree two (with possibly varying kernels).

The methods of proof of [4] rely on precise moment inequalities of Rosenthal type which are of independent interest (and which are valid for U–statistics of arbitrary order). In case of order two, these moment inequalities together with Talagrand inequality for empirical processes provided exponential bounds. Here, we present a different proof of their result which also provide information about the constants which is often needed in statistical applications [9]. Our approach still rely on Talagrand inequality but replaces the moment estimates by martingales types inequalities. As also indicated [4] the moment estimates and the exponential inequality are equivalent to one another and so our approach also provides sharp moment estimates. The methods presented here are robust enough that they can

---

1991 *Mathematics Subject Classification.* AMS numbers 60E15, 60G42, 62E17.
Key words and phrases. U–statistics, exponential inequalities, Poisson integrals.
Research supported in part by a NSF Grant.

be adapted to provide exponential inequalities for double integrals with respect to Poisson processes.

2. Background

Let us recall some known facts about U-statistics of order two. Throughout these notes, let $T_1, \dots, T_n$, be independent real random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

A canonical U-statistics of order two is generally defined for all positive integer n as

$$\sum_{i=1}^{n} \sum_{j=1}^{n} f_{i,j}(T_i, T_j), \tag{2.1}$$

where the $f_{i,j} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ are Borel measurable functions.

We will not be concerned in this work with the diagonal part

$$\sum_{i=1}^{n} f_{i,i}(T_i, T_i),$$

nor with the part of (2.1) made of sums of independent random variables. Indeed for these parts, exponential tail inequalities are well known and a "$u/2$ argument", combined with our results, provides exponential bounds for canonical U–statistics (of order two). Hence we will deal with degenerate U-statistics of order two, defined for all integer $n \geq 2$, by

$$\begin{aligned} \mathcal{U}_n = \sum_{i=1}^{n} \sum_{j \neq i} \Big[& f_{i,j}(T_i, T_j) - \mathbb{E}(f_{i,j}(T_i, T_j)|T_j) - \\ & - \mathbb{E}(f_{i,j}(T_i, T_j)|T_i) + \mathbb{E}(f_{i,j}(T_i, T_j)) \Big]. \end{aligned} \tag{2.2}$$

This is equivalent to considering for all integer $n \geq 2$,

$$U_n = \sum_{i=2}^{n} \sum_{j=1}^{i-1} g_{i,j}(T_i, T_j), \tag{2.3}$$

where the $g_{i,j} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ are Borel measurable functions verifying

$$\mathbb{E}(g_{i,j}(T_i, T_j)|T_i) = 0 \text{ and } \mathbb{E}(g_{i,j}(T_i, T_j)|T_j) = 0, \tag{2.4}$$

and where $\mathbb{E}$ is the expectation with respect to $\mathbb{P}$. Indeed it is sufficient to take $g_{i,j}(T_i, T_j) = f_{i,j}(T_i, T_j) + f_{j,i}(T_j, T_i) - \mathbb{E}(f_{i,j}(T_i, T_j) + f_{j,i}(T_j, T_i)|T_i) - \mathbb{E}(f_{i,j}(T_i, T_j) + f_{j,i}(T_j, T_i)|T_j) + \mathbb{E}(f_{i,j}(T_i, T_j) + f_{j,i}(T_j, T_i))$.

Throughout these notes, U_n is now given by (2.3) *and satisfies* (2.4).

For any $n \geq 1$, let $\mathcal{F}_n$ be the σ-field generated by $\{T_1, \ldots, T_n\}$, $\mathcal{F}_0 = \{\Omega, \emptyset\}$ and for any $n \geq 2$, let

$$X_n = \sum_{j=1}^{n-1} g_{n,j}(T_n, T_j).$$

As in (2.3), U_n is only defined for $n \geq 2$, we set $U_1 = 0$ and also $X_1 = 0$. The following is an easy, known, but important lemma:

Lemma 2.1. $(U_n, n \in \mathbb{N})$ *is a discrete time martingale with respect to the filtration* $(\mathcal{F}_n, n \in \mathbb{N})$ *and for all* n, $\mathbb{E}(X_n|\mathcal{F}_{n-1}) = 0$.

Proof. Let $n \geq 2$. Then clearly, X_n is $\mathcal{F}_n$-measurable. Moreover

$$\mathbb{E}(X_n|\mathcal{F}_{n-1}) = \sum_{j=1}^{n-1} \mathbb{E}(g_{n,j}(T_n, T_j)|\mathcal{F}_{n-1}) = \sum_{j=1}^{n-1} \mathbb{E}(g_{n,j}(T_n, T_j)|T_j) = 0,$$

since the T_i's are independent random variables and by (2.4). Finally, since $U_n = \sum_{i=1}^n X_i$, $\mathbb{E}(U_n|\mathcal{F}_{n-1}) = U_{n-1} + \mathbb{E}(X_n|\mathcal{F}_{n-1}) = U_{n-1}$. □

Throughout the sequel, and for all i *and* j, *we use the notation*

$$\mathbb{E}_{(i)}(g_{i,j}(T_i, T_j)) = \mathbb{E}(g_{i,j}(T_i, T_j)|T_j)$$

and

$$\mathbb{E}^{(j)}(g_{i,j}(T_i, T_j)) = \mathbb{E}(g_{i,j}(T_i, T_j)|T_i).$$

3. Exponential Inequalities

Let V_n^2 be the angle bracket [12, p. 148] of U_n, i.e. let $V_n^2 = \sum_{i=1}^n \mathbb{E}(X_i^2|\mathcal{F}_{i-1})$ and let also $B_n = \sup_{i \leq n} |X_i|$. Let us present a first result which is not quite the one obtained in [4] (because of the extra term F present below) but which already provides some knowledge of the constants.

Theorem 3.1. *Let* $u > 0, \varepsilon > 0$ *and let* $|g_{i,j}| \leq A$, *for all* i, j. *Then*

$$\begin{aligned}\mathbb{P}\Big[U_n \geq (1+\varepsilon)C\sqrt{2u} + \left(2\sqrt{\kappa}D + \frac{1+\varepsilon}{3}F\right)u \\ + \left(\sqrt{2}\kappa(\varepsilon) + \frac{2\sqrt{\kappa}}{3}\right)Bu^{3/2} + \frac{\kappa(\varepsilon)}{3}Au^2\Big] \\ \leq 3e^{-u} \wedge 1. \qquad (3.1)\end{aligned}$$

Above,

$$C^2 = \sum_{i=2}^{n}\sum_{j=1}^{i-1} \mathbb{E}(g_{i,j}(T_i, T_j)^2), \qquad (3.2)$$

$$(3.3)\quad D = \sup\Bigg\{\mathbb{E}\left(\sum_{i=2}^{n}\sum_{j=1}^{i-1} g_{i,j}(T_i,T_j)a_i(T_i)b_j(T_j)\right) : \mathbb{E}\left(\sum_{i=2}^{n} a_i(T_i)^2\right) \le 1, \mathbb{E}\left(\sum_{j=1}^{n-1} b_j(T_j)^2\right) \le 1\Bigg\},$$

$$(3.4)\quad F = \mathbb{E}\left(\sup_{i,t}\left|\sum_{j=1}^{i-1} g_{i,j}(t,T_j)\right|\right),$$

$$(3.5)\quad B^2 = \max\left\{\sup_{t,i}\left(\sum_{j=1}^{i-1}\mathbb{E}^{(j)}(g_{i,j}(t,T_j)^2)\right), \sup_{t,j}\left(\sum_{i=j+1}^{n}\mathbb{E}_{(i)}(g_{i,j}(T_i,t)^2)\right)\right\},$$

while κ and $\kappa(\varepsilon)$ can be chosen respectively equal to 4 and $(2.5+32\varepsilon^{-1})$.

As a preparation for the proof, we first obtain bounds on V_n^2 and B_n.

Lemma 3.2. *Let $u>0$ and let $\varepsilon>0$. With probability larger than $1-2e^{-u}$,*

$$\sqrt{V_n^2} \le (1+\varepsilon)C + D\sqrt{2\kappa u} + \kappa(\varepsilon)Bu$$

and

$$B_n \le (1+\varepsilon)F + B\sqrt{2\kappa u} + \kappa(\varepsilon)Au,$$

where κ and $\kappa(\varepsilon)$ can be chosen respectively equal to 4 and $(2.5+32\varepsilon^{-1})$.

To prove this lemma, we use Talagrand's inequality [16] for empirical processes, and more precisely the version with constants obtained by Massart [11] (see also Ledoux [10]).

(Talagrand's inequality) Let $X_1=(X_1^1,\dots,X_1^N),\dots,X_n=(X_n^1,\dots,X_n^N)$ be independent random variables with values in $[-b,b]^N$, for some positive real b. Let

$$(3.6)\quad Z = \sup_{1\le t\le N}\left|\sum_{i=1}^{n}\left(X_i^t-\mathbb{E}(X_i^t)\right)\right|,$$

and let

$$(3.7)\quad v = \sup_{1\le t\le N}\sum_{i=1}^{n}\mathrm{Var}(X_i^t).$$

Then for all $\varepsilon>0, z>0$

$$(3.8)\quad \mathbb{P}[Z \ge (1+\varepsilon)\mathbb{E}(Z) + \sqrt{2\kappa v z} + \kappa(\varepsilon)bz] \le e^{-z},$$

where κ and $\kappa(\varepsilon)$ can be respectively chosen equal to 4 and $2.5+32/\varepsilon$.

Proof. **[Lemma 3.2]** It is easy to see by the independence property of the variables that

$$V_n^2 = \sum_{i=2}^{n} \mathbb{E}_{(i)} \left(\left[\sum_{j=1}^{i-1} g_{i,j}(T_i, T_j) \right]^2 \right).$$

Therefore, by duality, we have that:

$$\begin{aligned}\sqrt{V_n^2} &= \sup_{\sum_{i=2}^{n} \mathbb{E}(a_i(T_i)^2)=1} \left| \sum_{i=2}^{n} \mathbb{E}_{(i)} \left(a_i(T_i) \sum_{j=1}^{i-1} g_{i,j}(T_i, T_j) \right) \right| \\ &= \sup_{\sum_{i=2}^{n} \mathbb{E}(a_i(T_i)^2)=1} \left| \sum_{j=1}^{n-1} \sum_{i=j+1}^{n} \mathbb{E}_{(i)}(a_i(T_i) g_{i,j}(T_i, T_j)) \right|,\end{aligned}$$

and

$$B_n = \sup_{i \le n} |X_i| \le \sup_{i \le n} \sup_{t} \left| \sum_{j=1}^{i-1} g_{i,j}(t, T_j) \right| := \tilde{B}_n.$$

By density, we can restrict the previous suprema to a countable deterministic dense subset of parameters : for V_n^2, the set of $\mathbb{L}^2$ functions is separable and for $\tilde{B}_n$, the set of t is $\mathbb{R}$ which is also separable. By monotone limit, we can restrict ourselves to take a finite subset of parameters and then pass to the limit. These suprema can then be interpreted as suprema of the form $\sup_{u \in \mathcal{T}} \sum_{j=1}^{n-1} X_j^u$, where $\mathcal{T}$ is finite and the $(X_j^u, u \in \mathcal{T})$'s are centered, independent and bounded. Therefore, applying Talagrand's inequality, and passing to the limit give the following results:

Let $u > 0$ and let $\varepsilon > 0$. With probability larger than $1 - e^{-u}$,

$$\sqrt{V_n^2} \le (1 + \varepsilon)\mathbb{E}(\sqrt{V_n^2}) + \sqrt{2\kappa v_1 u} + \kappa(\varepsilon) b_1 u, \tag{3.9}$$

where

$$v_1 = \sup_{\sum_{i=2}^{n} \mathbb{E}(a_i(T_i)^2)=1} \sum_{j=1}^{n-1} \mathrm{Var}^{(j)} \left(\sum_{i=j+1}^{n} \mathbb{E}_{(i)}(a_i(T_i) g_{i,j}(T_i, T_j)) \right)$$

and

$$b_1 = \sup_{t, j, \sum_{i=2}^{n} \mathbb{E}(a_i(T_i)^2)=1} \left| \sum_{i=j+1}^{n} \mathbb{E}_{(i)}(a_i(T_i) g_{i,j}(T_i, t)) \right|.$$

For $\tilde{B}_n$ we have with probability larger than $1 - e^{-u}$,

$$\tilde{B}_n \le (1 + \varepsilon)\mathbb{E}(\tilde{B}_n) + \sqrt{2\kappa v_2 u} + \kappa(\varepsilon) b_2 u, \tag{3.10}$$

where

$$v_2 = \sup_{i,t} \sum_{j=1}^{i-1} \mathrm{Var}^{(j)} \left(g_{i,j}(t, T_j) \right)$$

and

$$b_2 = \sup_{t,j,x,i} \left| g_{i,j}(x, t) \right|.$$

So (3.9) and (3.10) hold true together on an event of probability larger than $1-2e^{-u}$. Using (2.4), we have $\mathbb{E}(\sqrt{V_n^2}) \leq \sqrt{\mathbb{E}(V_n^2)} = C$, $v_1 = D^2$, $b_1 \leq B$, $\mathbb{E}(\tilde{B}_n) = F$, $v_2 \leq B^2$ and $b_2 = A$. The result follows. □

Proof. **[Theorem 3.1]** First, define b and v by

$$\sqrt{v} = (1+\varepsilon)C + D\sqrt{2\kappa u} + \kappa(\varepsilon)Bu$$

and

$$b = (1+\varepsilon)F + B\sqrt{2\kappa u} + \kappa(\varepsilon)Au.$$

Next, let us now return to U_n. More precisely, let us define the stopping time T by $T+1 = \inf\{k \in \mathbb{N}, V_k > v \text{ or } \tilde{B}_k > b\}$. Then U_n^T, the martingale U_n stopped in T, is also a martingale with respect to the same filtration. As V_k and $\tilde{B}_k$ are nondecreasing, the angle bracket and the jumps of this new martingale are respectively bounded by v and b. Therefore, (see [12, Lemma VII-2-8, p. 154]), for all $\lambda > 0$,

$$\left(e^{\lambda U_n^T - \phi_b(\lambda)v}, n \in \mathbb{N}\right) \tag{3.11}$$

is a super-martingale where $\phi_b(\lambda) = (e^{\lambda b} - \lambda b - 1)/b^2$. Finally, performing some classical computation on the Laplace transform of U_n^T, we get via Chebyshev's inequality

$$\mathbb{P}\left(U_n^T \geq \sqrt{2vu} + \frac{b}{3}u\right) \leq e^{-u}.$$

Hence

$$\begin{aligned}\mathbb{P}\left(U_n \geq \sqrt{2vu} + \frac{b}{3}u\right) &\leq \mathbb{P}\left(U_n^T \geq \sqrt{2vu} + \frac{b}{3}u\right) + \mathbb{P}(T+1 \leq n) \\ &\leq 3e^{-u}\end{aligned}$$

by Lemma 3.2. □

As already indicated, Theorem 3.1 does not quite recover the exponential bound of [4] because of the extra term F. With a little more work, F can be removed. At first, we need the following simple lemma.

Lemma 3.3. *Let $(Y_n, n \in \mathbb{N})$ be a martingale. For all $k \geq 2$, let*

$$A_n^k = \sum_{i=1}^n \mathbb{E}\left((Y_i - Y_{i-1})^k | \mathcal{F}_{i-1}\right).$$

Then for all integer $n \geq 1$ and for all λ such that for all $i \leq n$, $\mathbb{E}[\exp(|\lambda(Y_i - Y_{i-1})|)] < +\infty$,

$$\mathcal{E}_n = \exp\left(\lambda Y_n - \sum_{k\geq 2} \frac{\lambda^k}{k!} A_n^k\right) \tag{3.12}$$

is a super-martingale.

Proof. For all integer $n \geq 1$,

$$\mathbb{E}(\mathcal{E}_n|\mathcal{F}_{n-1}) = \mathcal{E}_{n-1}\mathbb{E}(e^{\lambda(Y_n - Y_{n-1})}|\mathcal{F}_{n-1}) \exp\left(-\sum_{k\geq 2}\frac{\lambda^k}{k!}\mathbb{E}\left((Y_n - Y_{n-1})^k|\mathcal{F}_{n-1}\right)\right),$$

But

$$\mathbb{E}(e^{\lambda(Y_n - Y_{n-1})}|\mathcal{F}_{n-1}) = 1 + \mathbb{E}\left(\sum_{k\geq 2}\frac{\lambda^k}{k!}(Y_n - Y_{n-1})^k|\mathcal{F}_{n-1}\right).$$

The partial sums are dominated by $\exp(|\lambda(Y_n - Y_{n-1})|)$ which is integrable by assumption. Therefore, by dominated convergence for conditional expectations, we can exchange sum and expectation to obtain:

$$\begin{aligned}\mathbb{E}\left(e^{\lambda(Y_n - Y_{n-1})}|\mathcal{F}_{n-1}\right) &= 1 + \sum_{k\geq 2}\frac{\lambda^k}{k!}\mathbb{E}\left((Y_n - Y_{n-1})^k|\mathcal{F}_{n-1}\right) \\ &\leq \exp\left(\sum_{k\geq 2}\frac{\lambda^k}{k!}\mathbb{E}\left((Y_n - Y_{n-1})^k|\mathcal{F}_{n-1}\right)\right),\end{aligned}$$

giving the result. □

A_n^2 is the classical angle bracket. Assume $Y_0 = 0$. If the A_n^k are bounded by $w_n^k \geq 0$, we have for all $\lambda > 0$,

$$\mathbb{E}(e^{\lambda Y_n}) \leq \exp\left(\sum_{k\geq 2}\frac{\lambda^k}{k!}w_n^k\right), \tag{3.13}$$

since $\mathbb{E}(\mathcal{E}_n) \leq \mathbb{E}(\mathcal{E}_0) = 1$. This result is due to Pinelis [13, Theorem 8.5].

We now state our main result which recovers the exponential bound of [4] with estimates on the constants.

Theorem 3.4. *Let A, B, C, D be as in* Theorem 3.1. *For all $\varepsilon, u > 0$,*

$$\mathbb{P}(U_n \geq 2(1+\varepsilon)^{3/2}C\sqrt{u} + \eta(\varepsilon)Du + \beta(\varepsilon)Bu^{3/2} + \gamma(\varepsilon)Au^2) \leq 2.77e^{-u} \tag{3.14}$$

where

- $\eta(\varepsilon) = \sqrt{2\kappa}(2 + \varepsilon + \varepsilon^{-1})$,
- $\beta(\varepsilon) = e(1 + \varepsilon^{-1})^2\kappa(\varepsilon) + \left[(\sqrt{2\kappa}(2 + \varepsilon + \varepsilon^{-1})) \vee \frac{(1+\varepsilon)^2}{\sqrt{2}}\right]$,
- $\gamma(\varepsilon) = (e(1 + \varepsilon^{-1})^2\kappa(\varepsilon)) \vee \frac{(1+\varepsilon)^2}{3}$,
- $\kappa = 4$,
- $\kappa(\varepsilon) = 2.5 + 32\varepsilon^{-1}$.

Proof. The A_n^k corresponding to the martingale U_n are

$$\sum_{i=2}^{n} \mathbb{E}_{(i)} \left[\left(\sum_{j=1}^{i-1} g_{i,j}(T_i,T_j) \right)^k \right] \leq V_n^k = \sum_{i=2}^{n} \mathbb{E}_{(i)} \left[\left| \sum_{j=1}^{i-1} g_{i,j}(T_i,T_j) \right|^k \right].$$

We now wish to estimate the V_n^k and this is the purpose of:

Lemma 3.5. *Let $\varepsilon > 0$ and $u > 0$. One has with probability larger than $1-1.77e^{-u}$, for all $k \geq 2$*

$$(V_n^k)^{1/k} \leq (1+\varepsilon)(\mathbb{E}(V_n^k))^{1/k} + \sigma_k\sqrt{2\kappa k u} + \kappa(\varepsilon) b_k k u,$$

where

$$\sigma_k^2 = \sup_{\sum_{i=2}^{n} \mathbb{E}(|a_i(T_i)|^{k/(k-1)})=1} \left\{ \sum_{j=1}^{n-1} \mathbb{E}\left(\left[\sum_{i=j+1}^{n} \mathbb{E}_{(i)}(a_i(T_i) g_{i,j}(T_i,T_j) \right]^2 \right) \right\},$$

$$b_k = \sup_{\sum_{i=2}^{n} \mathbb{E}(|a_i(T_i)|^{k/(k-1)})=1, j\leq n-1} \sup_t \left| \sum_{i=j+1}^{n} \mathbb{E}_{(i)} \left[g_{i,j}(T_i,t) a_i(T_i) \right] \right|$$

and where κ and $\kappa(\varepsilon)$ can be chosen respectively equal to 4 and $2.5+32/\varepsilon$.

Proof. **[Lemma 3.5]** By Hölder's inequality, we have:

$$(V_n^k)^{1/k} = \sup_{\sum_{i=1}^{n} \mathbb{E}(|a_i(T_i)|^{k/(k-1)})=1} \left\{ \sum_{j=1}^{n-1} \sum_{i=j+1}^{n} \mathbb{E}_{(i)} \left(g_{i,j}(T_i,T_j) a_i(T_i) \right) \right\}.$$

Using the same method as before, we can view the V_n^k's as a limit of suprema of the form

$$\sup_{u\in\mathcal{T}} \sum_{j=1}^{n-1} X_j^u$$

where $\mathcal{T}$ is finite and where the $(X_j^u, u \in \mathcal{T})$'s are independent centered and bounded real random variables. Therefore we can again apply Talagrand's inequality (3.8): for all $k \geq 2$, all $z > 0$ and all $\varepsilon > 0$

$$\mathbb{P}\left((V_n^k)^{1/k} \geq (1+\varepsilon)\mathbb{E}((V_n^k)^{1/k}) + \sigma_k\sqrt{2\kappa z} + \kappa(\varepsilon) b_k z \right) \leq e^{-z}. \tag{3.15}$$

Applying (3.15) to $z = ku$ and summing over k, it follows that:

$$\mathbb{P}\left(\forall k \geq 2,\ (V_n^k)^{1/k} \geq (1+\varepsilon)\mathbb{E}((V_n^k)^{1/k}) + \sigma_k\sqrt{2\kappa k u} + \kappa(\varepsilon) b_k k u \right) \leq \sum_{k\geq 2} e^{-ku}.$$

In fact the above left hand side is more precisely dominated by

$$1 \wedge \sum_{k\geq 2} e^{-ku} \leq 1 \wedge \frac{1}{e^u(e^u-1)} \leq 1 \wedge \frac{1}{ue^u} \leq 1.77e^{-u}.$$

Finally, $\mathbb{E}((V_n^k)^{1/k}) \leq (\mathbb{E}(V_n^k))^{1/k}$ and the result follows. □

We now bound the σ_k's and the b_k's. The easiest to bound are the b_k's: by Hölder's inequality,

$$b_k \leq \sup_{j,t} \left(\sum_{i=j+1}^{n} \mathbb{E}_{(i)}(|g_{i,j}(T_i,t)|^k) \right)^{1/k} \leq (B^2 A^{k-2})^{1/k},$$

where again B is given by (3.5) and since the $g_{i,j}$'s are bounded by A. The variance term is a bit more intricate.

$$\begin{aligned}
\sigma_k &= \sup_{\substack{\sum_{i=2}^n \mathbb{E}(|a_i(T_i)|^{k/(k-1)}) = 1 \\ \sum_{j=1}^{n-1} \mathbb{E}(|b_j(T_j)|^2) = 1}} \sum_{j=1}^{n-1} \mathbb{E}^{(j)} \left[\sum_{i=j+1}^{n} \mathbb{E}_{(i)}(g_{i,j}(T_i,T_j)a_i(T_i)b_j(T_j)) \right] \\
&= \sup_{\substack{\sum_{i=2}^n \mathbb{E}(|a_i(T_i)|^{k/(k-1)}) = 1 \\ \sum_{j=1}^{n-1} \mathbb{E}(|b_j(T_j)|^2) = 1}} \sum_{i=2}^{n} \mathbb{E}_{(i)} \left[\sum_{j=1}^{i-1} \mathbb{E}^{(j)}(g_{i,j}(T_i,T_j)b_j(T_j))a_i(T_i) \right] \\
&= \sup_{\sum_{j=1}^{n-1} \mathbb{E}(|b_j(T_j)|^2)=1} \left[\sum_{i=2}^{n} \mathbb{E}_{(i)} \left[\sum_{j=1}^{i-1} \mathbb{E}^{(j)} \left(g_{i,j}(T_i,T_j)b_j(T_j) \right) \right]^k \right]^{1/k} \\
&\leq (B^{k-2}D^2)^{1/k},
\end{aligned}$$

with D given by (3.3).

Next, since x^k is a convex function of x, applying the convexity property to $\left(\frac{\theta_1+\theta_2}{1+\varepsilon}\right)^k = \left(\frac{\theta_1}{1+\varepsilon} + \frac{\varepsilon\theta_2}{1+\varepsilon}\right)^k$, it easily follows that:

$$\forall k > 1, \theta_1, \theta_2, \varepsilon > 0, (\theta_1+\theta_2)^k \leq (1+\varepsilon)^{k-1}\theta_1^k + (1+\varepsilon^{-1})^{k-1}\theta_2^k, \tag{3.16}$$

Using this previous inequality several times, with probability larger than $1 - 1.77e^{-u}$, for all $k \geq 2$, V_n^k is bounded by w_n^k, where w_n^k is given by

$$\begin{aligned}
w_n^k = (1+\varepsilon)^{2k-1}\mathbb{E}(V_n^k) + (2+\varepsilon+\varepsilon^{-1})^{k-1} D^2 B^{k-2} (\sqrt{2\kappa k u})^k \\
+ (1+\varepsilon^{-1})^{2k-2} B^2 A^{k-2} \kappa(\varepsilon)^k (ku)^k.
\end{aligned}$$

As in the proof of Theorem 3.1, let $T+1 = \inf\{p \in \mathbb{N}, \exists k, V_p^k \geq w_n^k\}$ and note that since the V_n^k are nondecreasing, by Lemma 3.5 $\mathbb{P}(T < n) \leq 1.77e^{-u}$. Then stopping U_n at T, gives by Equation (3.13)

$$\mathbb{E}(e^{\lambda U_n^T}) \leq \exp\left(\sum_{k\geq 2} \frac{\lambda^k}{k!} w_n^k \right).$$

It remains to simplify this last bound and to use Chebyshev's inequality.

$$\begin{aligned} q_n &= \sum_{k\geq 2} \frac{\lambda^k}{k!} w_n^k \\ &\leq \sum_{k\geq 2} \frac{\lambda^k}{k!}(1+\varepsilon)^{2k-1}\mathbb{E}(V_n^k) + \\ &\quad + \sum_{k\geq 2} \frac{\lambda^k}{k!}(2+\varepsilon+\varepsilon^{-1})^{k-1} D^2 B^{k-2}(\sqrt{2\kappa k u})^k + \\ &\quad + \sum_{k\geq 2} \frac{\lambda^k}{k!}(1+\varepsilon^{-1})^{2k-2} B^2 A^{k-2}\kappa(\varepsilon)^k (ku)^k. \end{aligned}$$

Let us respectively denote by α, β and γ, each one of the three previous sums. For the last sum, since for all k, $k! \geq (k/e)^k$ (see Stirling's formula with correction [3, p. 54]), setting $\delta(\varepsilon) = e(1+\varepsilon^{-1})^2\kappa(\varepsilon)$, we get

$$\gamma \leq \sum_{k\geq 2} (\delta(\varepsilon))^k B^2 A^{k-2}(\lambda u)^k = \frac{\lambda^2 (B\delta(\varepsilon)u)^2}{1-(A\delta(\varepsilon)u)\lambda},$$

for $\lambda < (A\delta(\varepsilon)u)^{-1}$.

For the middle sum, since for all $k \geq 2$, $k! \geq k^{k/2}$ (again, see [3, p. 54]) and since moreover $2+\varepsilon+\varepsilon^{-1} \geq 4$, setting $\eta(\varepsilon) = \sqrt{2\kappa}(2+\varepsilon+\varepsilon^{-1})$, we similarly get

$$\beta \leq \frac{\lambda^2 (D\eta(\varepsilon)\sqrt{u}/2)^2}{1-(B\eta(\varepsilon)\sqrt{u})\lambda},$$

for $\lambda < (B\eta(\varepsilon)\sqrt{u})^{-1}$.

The estimation of the first sum is more intricate:

$$\alpha = \frac{1}{1+\varepsilon}\sum_{i=1}^n \mathbb{E}_{(i)}\left(\mathbb{E}(\exp(\mu|C_i|)|T_i) - \mu\mathbb{E}(|C_i||T_i) - 1\right), \tag{3.17}$$

where $C_i = \sum_{j=1}^{i-1} g_{i,j}(T_i,T_j)$ and $\mu = \lambda(1+\varepsilon)^2$. As $e^\theta - \theta - 1 \geq 0$, for all θ, adding $\mathbb{E}(\exp(-\mu|C_i|)|T_i) + \mu\mathbb{E}(|C_i||T_i) - 1$ to (3.17), we get

$$\alpha \leq \frac{1}{1+\varepsilon}\sum_{i=1}^n \mathbb{E}_{(i)}\left(\mathbb{E}(\exp(\mu C_i)|T_i) - 1 + \mathbb{E}(\exp(-\mu C_i)|T_i) - 1\right).$$

Let us recall:

(Bernstein's inequality) Let $X_1,\ldots,X_n$ be n independent centered variables with values in $[-A,A]$. Let $S_n = X_1+\cdots+X_n$ and let $v = \mathrm{Var}(S_n)$. Then for all $\mu > 0$,

$$\mathbb{E}(e^{\mu S_n}) \leq e^{\frac{\mu^2 v}{2-2\mu\frac{A}{3}}}.$$

Given T_i, C_i and $-C_i$ are sums of centered bounded i.i.d. quantities, it follows from Bernstein's inequality that

$$\alpha \leq \frac{2}{1+\varepsilon} \sum_{i=1}^{n} \mathbb{E}_{(i)} \left(e^{\frac{\mu^2 v_i(T_i)}{2-2\mu\frac{A}{3}}} - 1 \right), \tag{3.18}$$

where $v_i(T_i) = \sum_{j=1}^{i-1} \mathbb{E}^{(j)}(g_{i,j}(T_i, T_j)^2)$. But $v_i(T_i) \leq B^2$, thus

$$\sum_{i=1}^{n} \mathbb{E}_{(i)}(v_i(T_i)^k) \leq C^2 B^{2(k-1)},$$

where C is given by (3.2). Using these facts in (3.18) leads to

$$\alpha \leq \frac{(1+\varepsilon)^3 C^2 \lambda^2}{1 - \lambda(1+\varepsilon)^2 A/3 - \lambda^2(1+\varepsilon)^4 B^2/2}.$$

The last expression can be upper bounded by:

$$\alpha \leq \frac{(1+\varepsilon)^3 C^2 \lambda^2}{1 - (1+\varepsilon)^2 \lambda(A/3 + B/\sqrt{2})},$$

for $\lambda \leq [(1+\varepsilon)^2(A/3 + B/\sqrt{2})]^{-1}$. Finally one has,

$$\mathbb{E}(e^{\lambda U_n^T}) \leq \exp\left(\frac{\lambda^2 W^2}{1 - \lambda c}\right), \tag{3.19}$$

where

$$W = (1+\varepsilon)^{3/2} C + \eta(\varepsilon) D\sqrt{u}/2 + \delta(\varepsilon) B u,$$

and

$$c = \max\left((1+\varepsilon)^2(A/3 + B/\sqrt{2}), \eta(\varepsilon) B\sqrt{u}, \delta(\varepsilon) A u\right).$$

Next, Chebyshev's inequality $\mathbb{P}(U_n^T \geq s) \leq e^{-\lambda s}\mathbb{E}(e^{\lambda U_n^T})$, in conjunction with (3.19) and for $\lambda = \frac{\sqrt{u}}{W + c\sqrt{u}}$, give

$$\mathbb{P}(U_n^T \geq 2W\sqrt{u} + cu) \leq e^{-u}.$$

Proceeding as in the end of the proof of Theorem 3.1, one then gets the bound

$$\mathbb{P}(U_n \geq 2W\sqrt{u} + cu) \leq 2.77 e^{-u}.$$

This inequality implies the result for $u > 1$, but if $u \leq 1$, $2.77\exp(-u) > 1$. This finishes the proof of the theorem. □

Both Theorem 3.1 and Theorem 3.4 present some interest. The quadratic term in the first one is, as ε tends to 0, of the form $C\sqrt{2u}$ which is the optimal rate for the Central Limit Theorem since the variance term C^2 represents the true variance of the process.

The quadratic term in the second theorem is larger: it is of the form $2C\sqrt{u}$, the extra factor $\sqrt{2}$ coming from the use of symmetrization in the proof. This theorem gives precise constants which are unspecified in the result of [4]. Moreover

Theorem 3.4 has better order of magnitude than Theorem 3.1, as can be seen in the following example originating in statistics (see [9]).

Let $T_1, \dots, T_n$, be uniformly distributed on $[0,1)$. Let m be a regular partition of $[0,1)$, i.e. $[0,1) = \cup_{i=1}^d [\frac{i-1}{d}, \frac{i}{d})$.

We set

$$\forall (x,y) \in [0,1)^2, g(x,y) = d \sum_{I \in m} (1\!\!1_I(x) - 1/d)(1\!\!1_I(y) - 1/d).$$

Let U_n be the corresponding U-statistics (see the appendix of [9]). One has

$$A \leq 4d,\ B^2 \leq 2nd,\ C^2 \leq \frac{n(n-1)}{2} d,\ D \leq \frac{(n-1)}{2}.$$

F can also be computed (using Laplace transform) and is of the order of $d \ln n + n$.

For all ε and u positive, the following concentration inequalities hold true

- by applying Theorem 3.1: with probability smaller than $3e^{-u}$ one has

$$\frac{1}{n(n-1)} \sum_{i \neq j} g(T_i, T_j) = \frac{2U_n}{n(n-1)} \leq$$

$$2(1+\varepsilon)\sqrt{\frac{d}{n(n-1)} u} + \square \left(\frac{1}{n} + \frac{d \ln n}{n^2} \right) u +$$

$$+ \square \frac{\sqrt{d/n}}{n-1} u^{3/2} + \square \frac{d}{n(n-1)} u^2.$$

- by applying Theorem 3.4: with probability smaller than $2.77e^{-u}$ one has

$$\frac{2U_n}{n(n-1)} \leq 2(1+\varepsilon)^3 \sqrt{\frac{2d}{n(n-1)} u} + \square \frac{1}{n} u + \square \frac{\sqrt{d/n}}{n-1} u^{3/2} + \square \frac{d}{n(n-1)} u^2.$$

(The squares represent known but intricate constants.) The second inequality is sharper in the second term. In particular if d is of order n^2, the second one remains bounded while the first one tends to infinity with n.

4. The Poisson framework

The methodology of the previous sections can be easily adapted to obtain similar results for double integrals of Poisson processes. Let N be a time Poisson process with compensator Λ, and let $(M_t = N_t - \Lambda_t, t \geq 0)$ be the corresponding martingale.

The U-statistic or the double integral for the Poisson process is defined by

$$Z_t = \int_0^t \int_0^{y^-} f(x,y) dM_x dM_y$$

for $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ a Borel function. We do not need degeneracy assumptions, since we integrate with respect to a martingale and this implies that the expectations are zero.

Then we can easily obtain the corresponding version of Theorem 3.1.

Theorem 4.1. *Let $u, \varepsilon > 0$. If f is bounded by A, then*

$$\mathbb{P}\left[Z_t \geq (1+\varepsilon)C\sqrt{2u} + \left(2\sqrt{\kappa}D + \frac{1+\varepsilon}{3}F\right)u + \right.$$
$$\left. + \left(\sqrt{2}\kappa(\varepsilon) + \frac{2\sqrt{\kappa}}{3}\right)Bu^{3/2} + \frac{\kappa(\varepsilon)}{3}Au^2\right] \leq 3e^{-u},$$

where

$$C^2 = \int_0^t \int_0^y f(x,y)^2 d\Lambda_x \; d\Lambda_y,$$
$$D = \sup_{\int_0^t a_x^2 d\Lambda_x = 1, \int_0^t b_y^2 d\Lambda_y = 1} \int_0^t a_x \int_x^t b_y f(x,y) d\Lambda_y d\Lambda_x,$$
$$F = \mathbb{E}\left(\sup_{y \leq t}\left|\int_0^t \mathbb{1}_{x<y} f(x,y) dM_x\right|\right),$$

and

$$B^2 = \max\left\{\sup_{y\leq t}\int_0^y f(x,y)^2 d\Lambda_x, \sup_{x\leq t}\int_x^t f(x,y)^2 d\Lambda_y\right\}.$$

where $\kappa = 6$ and $\kappa(\varepsilon) = 1.25 + 32/\varepsilon$ are given by [15, Corollary 2].

Proof. Perform similar computations in continuous time, replacing Talagrand's inequality by [15, Corollary 2] and (3.11) by the corresponding Lemma derived by van de Geer in [17] or in [8, Theorem 23.17]. □

To conclude, we also state the Poisson version of Theorem 3.4.

Theorem 4.2. *For all $\varepsilon, u > 0$,*

$$\mathbb{P}(Z_t \geq 2(1+\varepsilon)^{3/2}C\sqrt{u} + 2\eta(\varepsilon)Du + \beta(\varepsilon)Bu^{3/2} + \gamma(\varepsilon)Au^2) \leq 2.77e^{-u}$$

where

- $\eta(\varepsilon) = \sqrt{2\kappa}(2 + \varepsilon + \varepsilon^{-1})$,
- $\beta(\varepsilon) = e(1+\varepsilon^{-1})^2\kappa(\varepsilon) + (\sqrt{2\kappa}(2+\varepsilon+\varepsilon^{-1})) \vee \frac{(1+\varepsilon)^2}{\sqrt{2}}$,
- $\gamma(\varepsilon) = (e(1+\varepsilon^{-1})^2\kappa(\varepsilon)) \vee \frac{(1+\varepsilon)^2}{3}$,
- $\kappa = 6$,
- $\kappa(\varepsilon) = 1.25 + 32/\varepsilon$.

Proof. Perform similar computations in continuous time, replacing Talagrand's inequality by [15, Corollary 2] and replacing Lemma 3.3 by its corresponding continuous time version [14, Proposition 4]. □

Potential statistical applications of the two previous theorems would be to construct tests for the Poisson intensity.

5. Concluding Remarks

In [4], the exponential bound is obtained for decoupled U-statistics, i.e. of the form

$$\sum_{i=1}^{n}\sum_{j=1}^{n} f_{i,j}(T_i, T'_j), \tag{5.1}$$

where $T_1, \ldots, T_n, T'_1, \ldots, T'_n$ are independent random variables. The decoupling inequality of de la Peña and Montgomery-Smith [2] states that, whenever $f_{i,i} = 0$ and $f_{i,j} = f_{j,i}$ for all i, j,

$$\mathbb{P}\left(\left|\sum_{i=1}^{n}\sum_{j=1}^{n} f_{i,j}(T_i, T_j)\right| \geq z\right) \leq C_2 \mathbb{P}\left(C_2\left|\sum_{i=1}^{n}\sum_{j=1}^{n} f_{i,j}(T_i, T'_j)\right| \geq z\right), \tag{5.2}$$

for all $z > 0$ and for some unspecified constant $C_2 > 0$.

Our methods provide an exponential upper bound for the left hand side of (5.2) while [4] provides an exponential upper bound for its right hand side. However, Theorem 3.1 and Theorem 3.4 immediately imply their versions for decouples U–statistics. Indeed, it is enough to take $n' = 2n$, $g'_{i,n+i} = g_{i,j}$ and $g'_{i,j} = 0$ if $1 \leq i, j \leq n$ or $n + 1 \leq i, j \leq 2n$.

The martingale part of the approach presented in these notes adapts easily to higher order U-statistics. However, we are lacking the corresponding version of (3.8). Even for suprema of U-statistics of order two, which will then imply results on U-statistics of order three, (3.8) is unknown. This problem deserves a closer attention.

References

[1] J. Bretagnolle. A new large deviation inequality for U-statistics of order 2. *ESAIM: Probability and Statistics*, 1999.

[2] V. De la Peña and S. Montgomery-Smith. Decoupling inequalities for the tail probabilities of multivariate u-statistics. *Ann. Probab.*, 23:806–816, 1995.

[3] W. Feller. *An introduction to probability theory and its applications.*, volume 1. Wiley, 3 edition, 1968.

[4] E. Giné, R. Latała, and J. Zinn. Exponential and moment inequalities for U-statistics. *High Dimensional Prob. II*, 2000.

[5] D.L. Hanson and F.T. Wright. A bound on tail probabilities for quadratic forms in independent random variables. *Ann. Math. Stat.*, 42(3):1079–1083, 1971.

[6] W. Hoeffding. A class of statistics with asymptotically normal distribution. *Ann. Math. Stat.*, 19(3), 1948.

[7] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58, 1963.

[8] O. Kallenberg. *Foundations of modern probability.* Springer, 1997.

[9] B. Laurent. Adaptive estimation of a quadratic functional of a density by model selection. Prépublication, Département de Mathématiques, Université Paris XI, 2003.

[10] M. Ledoux. On Talagrand deviation inequalities for product measures. In *ESAIM: Probability and statistics 1*, 1996.

[11] P. Massart. About the constants in Talagrand's concentration inequalities for empirical processes. *Ann. Proba.*, 2(28):863–884, 2000.

[12] J. Neveu. *Discrete-Parameter Martingales. Translated by T. P. Speed.* American Elsevier Publishing Company, Inc., 1975.

[13] I. Pinelis. Optimum bounds for the distributions of martingales in Banach spaces. *Ann. Prob.*, 22:1679–1706, 1994.

[14] P. Reynaud-Bouret. Exponential inequalities for counting processes. Technical Report 1102-001, Georgia Institute of Technology, School of Mathematics, ftp://ftp.math.gatech.edu/pub/preprints/ available_preprints_som.html, 2002.

[15] P. Reynaud-Bouret. Adaptive estimation of the intensity of inhomogeneous Poisson processes via concentration inequalities. *P.T.R.F.*, 2003. D.O.I: 10.1007/s00440-003-0259-1.

[16] M. Talagrand. New concentration inequalities in product spaces. *Invent. Math.*, 126(3):505–563, 1996.

[17] S. van de Geer. Exponential inequalities for martingales, with application to maximum likelihood estimation for counting processes. *Ann. Stat.*, 23(5):1779–1801, 1995.

Laboratoire d' Analyse et de Mathématiques Appliquées, CNRS UMR 8050, Université Paris XII, 94010 Créteil Cedex, France, and School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, USA.
E-mail address: houdre@math.gatech.edu

School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, USA, and DMA, École Normale Supérieure, 45 rue d'Ulm, 75230 Paris Cedex 05, France.
E-mail address: Patricia.Reynaud-Bouret@ens.fr