Electronic Journal of Probability

Vol. 14 (2009), Paper no. 28, pages 752–779.

Journal URL
http://www.math.washington.edu/~ejpecp/

Exponential inequalities for sums of weakly dependent variables

Bernard Delyon
IRMAR, Université Rennes 1
Campus de Beaulieu, 35042 Rennes Cedex, France
bernard.delyon@univ-rennes1.fr

Abstract

We give new exponential inequalities and Gaussian approximation results for sums of weakly dependent variables. These results lead to generalizations of Bernstein and Hoeffding inequalities, where an extra control term is added; this term contains conditional moments of the variables.

Key words: Mixing, exponential inequality, random fields, weak dependence.

AMS 2000 Subject Classification: Primary 60F99; Secondary: 60G60.

Submitted to EJP on June 10, 2008, final version accepted March 16, 2009.

1 Introduction

In the whole paper $(X_i)_{1\le i\le n}$ is a sequence of *centered* random variables. Our objective is to give new exponential inequalities and Gaussian approximation results for the sum $S = X_1 + \cdots + X_n$ (and other functions of $(X_1,\dots X_n)$) in the case where first and second order mixing conditions are assumed (first order mixing conditions involve conditionnal means and second order ones involve conditionnal covariances).

The essential application of exponential inequalities is to give small event probabilities; typically, we would like here to extend the Hoeffding and Bernstein inequalities to mixing processes, that is

$$P(S \ge A) \le \exp\left(-\frac{2A^2}{\sum_i b_i^2 + \rho}\right), \qquad a_i \le X_i \le a_i + b_i \tag{1}$$

$$P(S \ge A) \le \exp\left(-\frac{A^2}{2E[S^2] + 2Am/3 + \rho}\right), \qquad m = \sup_k \|X_k\|_\infty \tag{2}$$

where $\rho = 0$ if the variables are independent. We shall obtain here a value of ρ which depends on conditional moments of the variables. We also want to provide inequalities which generalize what is already known for martingales. Actually Equation (2) is not satisfied for a martingale ($E[S^2]$ has to be replaced with a bound on the total variation); this will lead us to two different alternatives

$$P(S \ge A) \le \exp\left(-\frac{A^2}{2v + 2Am/3 + \rho}\right), \tag{3}$$

where v is a bound on some kind of quadratic variation and $\rho = 0$ in the case of a martingale (cf. Theorem 4 for a precise statement), or

$$P(S \ge A) \le \exp\left(-\frac{A^2}{2E[S^2] + \sqrt{2A\rho_1/3}}\right), \tag{4}$$

where ρ_1 is a third order quantity (i.e. if S is replaced with tS, ρ_1 becomes $t^3\rho_1$) which involves conditional moments and is in the independent case $\rho_1 = \sum Var(X_k)\|X_k\|_\infty + \|X_k^3\|_\infty/2$ (Theorem 7 and Theorem 9). For instance, if we are close to the independent case and S is a normalized sum, that is $X_k = U_k/\sqrt{k}$ where $(U_k)_k$ is a sequence of weakly dependent bounded random variables, ρ_1 has order $1/\sqrt{n}$ and the second term in the denominator is residual as long as $A \ll \sqrt{n}$.

Bounds like Equation (3) will be obtained through what we call here the first order approach whereas (4) will require the second order approach. We present the main ideas below.

First order approach. Considering the sequence $X_1,\dots X_n$ as a time series, for instance a martingale, it is natural to introduce the σ-fields

$$\mathcal{F}_k = \sigma(X_1,\dots X_k). \tag{5}$$

It will appear that the remainder ρ will involve essentially the L_∞ norm of the conditional expectations $E[X_j|\mathcal{F}_k]$, $k < j$. If the sequence is a random field this filtration will generally not be very useful because of the arbitrariness of the order on the variables, and we shall need to proceed as follows: to each index k we associate a reordering of the sequence which corresponds (hopefully) to increasing dependence with X_k; this brings on a new sequence, depending on k: X_j^k. More precisely:

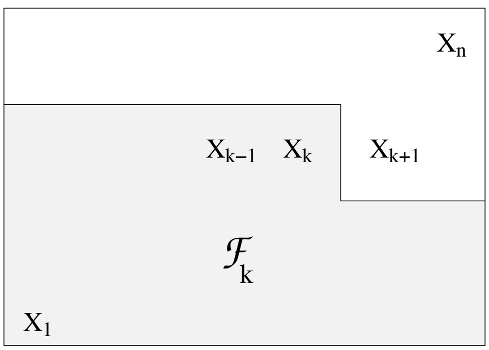
For any $1 \leq k \leq n$ is given a sequence X_j^k, $j = 1, \dots k$, which is a reordering of $(X_j,\ j = 1, \dots k)$ with $X_k^k = X_k$. We attach to each k a family of σ-algebras $(\mathcal{F}_j^k)_{j \leq k}$ such that

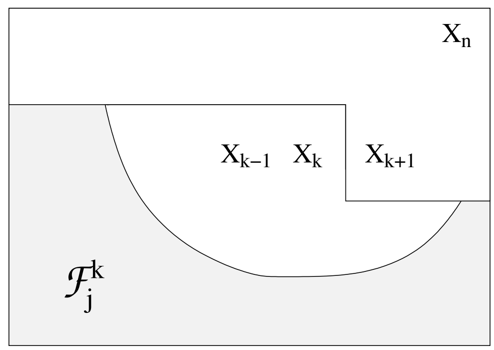
$$\mathcal{F}_j^k \supset \sigma(X_i^k, i \leq j), \quad j \leq k. \tag{6}$$

In particular, we have

$$\mathcal{F}_k^k \supset \mathcal{F}_k$$
$$\mathcal{F}_{k-1}^k \supset \mathcal{F}_{k-1}.$$

If (X_i) is a time series, it is natural to set $X_i^k = X_i$ and $\mathcal{F}_j^k = \mathcal{F}_j = \sigma(X_i, i \leq j)$: the superscripts can be dropped. Later on, the term "times series" will refer to to this situation, whereas the general case will be rather refered to as "random fields".





When dealing with mixing random fields of $\mathbb{R}^d$, each index k corresponds to some point P_k of the space where X_k sits; for each k, the sequence (X_j^k) will be typically obtained by sorting the original sequence $(X_j)_{j \leq k}$ in decreasing order of the distance $d(P_j, P_k)$. A simple example is the case of m-dependent fields indexed by $\mathbb{Z}^d$, that is, a process $X_a, a \in \mathbb{Z}^d$, such that the set of variables $X_A = \{X_a : a \in A\}$ is independent of X_B if the sets A and B have distance at least m; they are typically fields of the form $X_a = h(Y_{a+C})$ where Y_a is an i.i.d. random field, C a finite neighborhood of 0 in $\mathbb{Z}^d$, and h a measurable function of $|C|$ variables.

We would like to point out that this framework covers quite different situations. For instance, in the Erdös-Rényi model of an unoriented random graph with n vertices, edges are represented by $\binom{n}{2}$ i.i.d. Bernoulli variables Y_{ab}, $1 \leq a < b \leq n$, with the convention $Y_{ab} = Y_{ba}$ and $Y_{aa} = 0$ (see [2] and references therein). The number of triangles (for instance) in such a model is

$$\sum_{a,b,c} Y_{ab} Y_{bc} Y_{ac} = \sum_{a,b,c} X_{abc}.$$

The process X is here an m-dependent process on the set of three element subsets of $\{1, \dots n\}$. We shall treat this example in Section 3.5. As pointed out by Lumley [14], a similiar situation appears in the case of U-statistics; we shall however treat them with a martingale method.

Within this framework, we are able to control exponential moments of S with the help of formulas which generalize the Hoeffding and Bernstein inequalities for independent variables (Theorem 2 and Theorem 3). The bound of Theorem 2 involves $v = \sum_k \|E[X_k^2|\mathcal{F}_{k-1}^k]\|_\infty$, and a remainder term involving conditional expectations $\|E[X_k|\mathcal{F}_i^k]\|_\infty$. This is slightly unsatisfactory since it is known that the key quantity in the case of a martingale is the quadratic variation $\langle X\rangle = \sum_{k=1}^n E[X_k^2|\mathcal{F}_{k-1}]$, and in most cases effective bounds will actually involve $\|\langle X\rangle\|_\infty$, which is smaller than v. This is corrected in Theorem 1, where we give a result which generalizes what is known from martingale theory and improves on classical papers concerned with mixing [6]. However, inspection of the bounds shows that this improvement is really effective only if the conditional expectations $|E[X_k|\mathcal{F}_i^k]|$ are significantly smaller than $|X_k|$; if not, the only way to improve accuracy is to use the second order approach of Section 3 briefly discussed below.

Second order results. By this terminology, we mean the following fact: the Hoeffding inequality (Equation 1 with $\rho = 0$) for instance is obtained from the exponential inequality

$$E[e^{tS}] \le e^{t^2\sum_i b_i^2/8}. \tag{7}$$

One obvious drawback of this upper bound is that when t tends to 0, it does not look like $1+t^2E[S^2]$. One would rather expect something like

$$E[e^{tS}] \le e^{t^2E[S^2]/2+Ct^3} \tag{8}$$

which has more interesting scaling properties; this approach would hopefully lead to significant improvements in a moderate deviation domain; this is what has been done in [5], but there S is an arbitrary function of independent variables, or of a Markov chain. In order to get closer, like in Equation (34) and (41) below, we have to pay with higher order extra terms: the remainder terms will not only contain conditional expectations $E[X_k|\mathcal{F}_i^k]$ but also conditional covariances; this will force us to consider for each pair of indices (i,j) another reordering of the sequence which corresponds to increasing dependence with the pair (X_i,X_j), and to introduce σ-fields $\mathcal{H}_k^{ij}$; we postpone details to Section 3.

In this context we shall give exponential inequalities and Gaussian approximation; we give in particular a bound for $|E[h(X)]-E[h(N)]|$ where h is any function of n variables with all third derivatives bounded and N is a Gaussian random variable with same covariance matrix as X.

The paper is organized as follows. The two forthcoming sections deal with first order and second order exponential inequalities. A classical use of the exponential inequalities leads to Theorem 4 which generalizes the Bernstein and Hoeffding inequalities. An application to concentration inequalities and triangle counts is given in Section 2.2.

Section 3 is concerned with the second order approach, with applications to bounded difference inequalities and triangle counts.

In Section 4 we give some estimates under mixing assumptions.

2 First order approach

2.1 General bounds

This section is devoted to bounds for the Laplace transform of S. The corresponding deviation probabilities will be obtained in Section 2.2.1 through classical arguments.

In Theorem 1 we present bounds which generalize known results concerning martingales. Since they only use the linear sequence of σ-fields $\mathcal{F}_k$, they are essentially interesting for time series.

In Theorem 2 we give a Hoeffding bound which is valid in both cases (time series and random fields), and Theorem 3 gives a Bennett bound for random fields which does not exactly generalizes (9) because the quadratic variation $\langle X\rangle$ is changed into the more drastic upper bound v.

The applicability of the following theorem depends on the way one can bound the quadratic variations involved. In the forthcoming examples, we shall consider only Equation (9) through a bound on $\|\langle X\rangle\|_\infty$; however Equations (10) and (11) have the advantage of not involving m.

Theorem 1. *We are in the setting described in the introduction, with the filtration defined by (5). The variables X_k are centered. We define*

$$\begin{aligned}
m &= \sup_{1\le k\le n} \operatorname{ess\,sup} X_k\\
[X] &= \sum_{k=1}^{n} X_k^2\\
\langle X\rangle &= \sum_{k=1}^{n} E[X_k^2|\mathcal{F}_{k-1}]\\
[X_+] &= \sum_{k=1}^{n} (X_k)_+^2\\
\langle X_-\rangle &= \sum_{k=1}^{n} E[(X_k)_-^2|\mathcal{F}_{k-1}]\\
q &= \sum_{k=1}^{n}\sum_{i=1}^{k-1} \|X_i\|_\infty \|E[X_k|\mathcal{F}_i]\|_\infty
\end{aligned}$$

where the notation x_+^2 (resp. x_-^2) stands for $x^2 1_{x>0}$ (resp. $x^2 1_{x<0}$). Then

$$E\left[\exp\left(S-\frac{\langle X\rangle}{m^2}(e^m-m-1)\right)\right] \le e^{4q} \tag{9}$$

$$E\left[\exp\left(S-\frac{1}{2}[X_+]-\frac{1}{2}\langle X_-\rangle\right)\right] \le e^{4q} \tag{10}$$

$$E\left[\exp\left(S-\frac{1}{6}[X]-\frac{1}{3}\langle X\rangle\right)\right] \le e^{4q}. \tag{11}$$

Remark. In the martingale case $q=0$. We recommend [1] for an account on recent work concerning exponential inequalities for martingales.

Proof. Consider a pair of functions $\theta(x)$ and $\psi(x)$ such that

$$\psi \ge 0, \qquad e^{x-\theta(x)} \le 1 + x + \psi(x). \tag{12}$$

These functions are meant to be $O(x^2)$ in the neighborhood of 0. Three examples of such functions are

$$\begin{aligned}
&\theta(x) = 0, \quad \psi(x) = e^x - x - 1\\
&\theta(x) = \zeta(x_+), \quad \psi(x) = \zeta(x_-), \quad \zeta(x) = e^{-x} + x - 1\\
&\theta(x) = \frac{x^2}{6}, \quad \psi(x) = \frac{x^2}{3}.
\end{aligned}$$

Inequality (12) for these functions is proved in Proposition 12 of the Appendix. Set

$$T_k = \sum_{i=1}^{k} X_i - \theta(X_i) - \log(1+\xi_i), \text{ where } \xi_i = E[\psi(X_i)|\mathcal{F}_{i-1}].$$

Then

$$\begin{aligned}
E[e^{T_n}] &= E[e^{X_n - \theta(X_n)}(1+\xi_n)^{-1}e^{T_{n-1}}]\\
&\leq E[(1+X_n+\psi(X_n))(1+\xi_n)^{-1}e^{T_{n-1}}]\\
&= E[X_n(1+\xi_n)^{-1}e^{T_{n-1}}] + E[e^{T_{n-1}}]\\
&= E[X_n((1+\xi_n)^{-1}-1)e^{T_{n-1}}] + E\sum_{i=1}^{n-1} X_n(e^{T_i} - e^{T_{i-1}}) + E[e^{T_{n-1}}]\\
&= r_1 + r_2 + E[e^{T_{n-1}}].
\end{aligned}$$

In the martingale case, the first two terms are zero; here we have

$$r_1 = E[E[X_n|\mathcal{F}_{n-1}]e^{T_{n-1}}\xi_n/(1+\xi_n)] \leq E[|E[X_n|\mathcal{F}_{n-1}]|e^{T_{n-1}}](\|\psi(X_n)\|_\infty \wedge 1).$$

The above defined function ψ is convex with $\psi(0)=0$, $\psi(-1)\leq 1$ and $\psi(1)\leq 1$. Hence $|\psi(x)|\wedge 1 \leq |x|$ and therefore

$$r_1 \leq \|X_n\|_\infty \|E[X_n|\mathcal{F}_{n-1}]\|_\infty E[e^{T_{n-1}}].$$

Let $\Delta_i = T_i - T_{i-1}$; the second remainder is bounded as follows:

$$\begin{aligned}
r_2 &= E\sum_{i=1}^{n-1} X_n \tanh(\Delta_i/2)(e^{T_i} + e^{T_{i-1}})\\
&\leq E\sum_{i=1}^{n-1} |E[X_n|\mathcal{F}_i]\Delta_i|(e^{T_i} + e^{T_{i-1}})/2\\
&\leq \sum_{i=1}^{n-1} \|E[X_n|\mathcal{F}_i]\Delta_i\|_\infty \sup_{j\leq n-1} E[e^{T_j}].
\end{aligned}$$

Equation (53) of the Appendix implies that $|\tanh(\Delta_i)| \leq 3\|X_i\|_\infty$, hence

$$r_2 \leq 3\rho_n \sup_{i\leq n-1} E[e^{T_i}], \quad \rho_n = \sum_{i=1}^{n-1} \|X_i\|_\infty \|E[X_n|\mathcal{F}_i]\|_\infty.$$

Finally

$$E[e^{T_n}] \le (1+4\rho_n) \sup_{i\le n-1} E[e^{T_i}] \le \exp(4\rho_n) \sup_{i\le n-1} E[e^{T_i}]$$

and we get by induction that

$$\sup_{i\le k} E[e^{T_i}] \le \exp(4\sum_{i=1}^{k} \rho_i).$$

In particular

$$E\left[\exp\left\{\sum_{i=1}^{n} X_i - \theta(X_i) - \log(1+E[\psi(X_i)|\mathcal{F}_{i-1}])\right\}\right] \le \exp(4q)$$

hence

$$E[\exp\{\sum_{i=1}^{n} X_i - \theta(X_i) - E[\psi(X_i)|\mathcal{F}_{i-1}]\}] \le \exp(4q).$$

This leads to the three bounds by using the three pairs of functions and by noticing that for $m \ge 0$ and $x \le m$

$$\varphi(x) \le \varphi(m), \quad \varphi(x) = \frac{e^x - x - 1}{x^2} \tag{13}$$

which is a consequence of L'Hospital's rule for monotonicity [**?**], and that for $x \ge 0$

$$\zeta(x) \le \frac{x^2}{2}$$

since the function $x^2/2 - \zeta(x)$ has a non-negative derivative. □

Theorem 2. *Assume that we are in the setting described in the introduction, with a family of σ-fields satisfying (6). The variables X_k are centered. We define now q as*

$$q = \sum_{k=1}^{n} \sum_{i=1}^{k-1} \|X_i^k\|_\infty \|E[X_k|\mathcal{F}_i^k]\|_\infty$$

(this is consistent with the definition in Theorem 1). If the variables are lower and upper bounded with probability one:

$$a_i \le X_i \le a_i + b_i \tag{14}$$

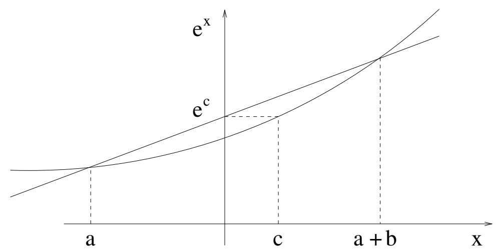
the following inequality holds

$$E\left[\exp\left(S - \frac{1}{8}\sum_i b_i^2\right)\right] \le e^{8q}. \tag{15}$$

In the martingale case ($\mathcal{F}_i^k = \mathcal{F}_i$ and $E[X_i|\mathcal{F}_{i-1}] = 0$ for all i and k), this inequality remains true if we allow a_i and b_i to be an $\mathcal{F}_{i-1}$-measurable random variables.

Proof. We assume first that a_i and b_i are deterministic. We start, as in the proof of the Hoeffding inequality, with the following inequality based on the majoration of the exponential function by the chord over the curve on $[a, a+b]$:

$$e^x \leq \frac{(a+b)e^a - ae^{a+b}}{b} + x\frac{e^{a+b} - e^a}{b}, \quad a \leq x \leq a+b.$$



It is well known that the first term of the right hand side, e^c on the figure, is smaller than $\exp(b^2/8)$ independently of a (this a key step for proving the Hoeffding inequality, see for instance Appendix B of [**?**]). On the other hand, it is clear that $c \leq a+b$ (see the figure or bound e^a with e^{a+b} in the expression of e^c). Hence, if we define c_i and d_i with the equations

$$c_i = \min(\frac{b_i^2}{8}, a_i + b_i), \quad d_i = e^{a_i}\frac{e^{b_i} - 1}{b_i}$$

we have

$$e^x \leq e^{c_i} + d_i x, \quad a_i \leq x \leq a_i + b_i.$$

Now let the random variables T_j and T_j^k be defined as

$$T_k = \sum_{i=1}^{k}(X_i - c_i), \qquad T_j^k = \sum_{i=1}^{j}(X_i^k - c_i^k)$$

where $(c_i^k)_{i\leq k}$ is the corresponding reordering of the sequence $(c_i)_{i\leq k}$. We obtain

$$\begin{aligned} E[e^{T_n}] &= E[e^{X_n}e^{T_{n-1}-c_n}] \\ &\leq E[(e^{c_n} + d_n X_n)e^{T_{n-1}-c_n}] \end{aligned}$$

In the martingale case, the term involving d_n vanishes and this equation gives immediately the result. We assume now that we are not necessarily in this case but a_i and b_i are deterministic. We can assume in addition, without loss of generality, that a_i and b_i are chosen so that (14) is tight. Notice that in this case we also have $a_i \leq 0 \leq a_i + b_i$ since $E[X_i] = 0$. The previous equation implies

$$\begin{aligned} E[e^{T_n}] &\leq d_n e^{-c_n} E[X_n e^{T_{n-1}}] + E[e^{T_{n-1}}] \\ &= d_n e^{-c_n} E\sum_{i=1}^{n-1} X_n(e^{T_i^n} - e^{T_{i-1}^n}) + E[e^{T_{n-1}}] \\ &= d_n e^{-c_n} r_2 + E[e^{T_{n-1}}]. \end{aligned} \tag{16}$$

Let $\Delta_i = T_i^n - T_{i-1}^n = X_i^n - c_i^n$; bounding r_2 as in the proof of Theorem 1 we get

$$\begin{aligned} r_2 &= E\sum_{i=1}^{n-1} X_n \tanh(\Delta_i/2)(e^{T_i^n} + e^{T_{i-1}^n}) \\ &\leq E\sum_{i=1}^{n-1} |E[X_n|\mathcal{F}_i^n]\Delta_i|(e^{T_i^n} + e^{T_{i-1}^n})/2 \\ &\leq \sum_{i=1}^{n-1} \|E[X_n|\mathcal{F}_i^n]\Delta_i\|_\infty \sup_{j\leq n-1} E[e^{T_j^n}] \end{aligned}$$

and since

$$|\Delta_i| \leq \max(a_i^n + b_i^n - c_i^n, c_i^n - a_i^n) \leq b_i^n \leq 2\|X_i^n\|_\infty$$

(b_i^n is the difference between the essential supremum and the essential infimum) we get that

$$r_2 \leq 2\rho_n \sup_{i\leq n-1} E[e^{T_i^n}], \quad \text{with} \quad \rho_n = \sum_{i=1}^{n-1} \|X_i^n\|_\infty \|E[X_n|\mathcal{F}_i^n]\|_\infty.$$

On the other hand since $a_i \leq 0$ the Equation (55) in Propostion 13 of the Appendix leads to $d_n e^{-c_n} \leq 4$ and (16) becomes finally

$$E[e^{T_n}] \leq (1 + 8\rho_n) \sup_{i\leq n-1} E[e^{T_i^n}] \leq e^{8\rho_n} \sup_{i\leq n-1} E[e^{T_i^n}].$$

For any sequence $\alpha \in \{0,1\}^n$ set

$$T_k(\alpha) = \sum_{i=1}^{k} \alpha_i (X_i - c_i).$$

The bound we obtained is still obviously valid for $T_n(\alpha)$ since the replacement of X_i with $\alpha_i X_i$ does not increases ρ_n, hence

$$\sup_\alpha E[e^{T_n(\alpha)}] \leq e^{8\rho_n} \sup_\alpha \sup_{i\leq n-1} E[e^{T_i^n(\alpha)}] = e^{8\rho_n} \sup_\alpha E[e^{T_{n-1}(\alpha)}]$$

and we get

$$E[e^{T_n}] \leq \exp\left(8\sum_{i=1}^{n} \rho_i\right).$$

We obtain (15) by using that $c_i \leq b_i^2/8$ in the expression of T_n. □

Theorem 3. *Assume that we are in the setting described in the introduction, with a family of σ-fields satifying (6). The variables X_k are centered. We define*

$$v = \sum_{k=1}^{n} \|E[X_k^2|X_{k-1}, \ldots X_1]\|_\infty,$$

m as in Theorem 1, and q as in Theorem 2. Then for any $t > 0$

$$E[e^S] \leq \exp\left(\frac{v}{m^2}(e^m - m - 1) + q\right). \tag{17}$$

Proof. We set

$$S_i = X_1^n + X_2^n + \cdots + X_i^n, \quad i \le n$$

and $S_0 = 0$. Equation (13) implies that

$$e^{X_n} \le 1 + X_n + X_n^2 \varphi(m)$$

hence

$$\begin{aligned}
E[e^S] &\le E[(1 + X_n + X_n^2\varphi(m))e^{S_{n-1}}] \\
&= E\sum_{i=1}^{n-1} X_n(e^{S_i} - e^{S_{i-1}}) + E[(1 + X_n^2\varphi(m))e^{S_{n-1}}] \\
&= E\sum_{i=1}^{n-1} X_n X_i^n \frac{\tanh(X_i^n/2)}{X_i^n}(e^{S_i} + e^{S_{i-1}}) + \varphi(m)E[X_n^2 e^{S_{n-1}}] + E[e^{S_{n-1}}] \\
&\le E\sum_{i=1}^{n-1} \|E[X_n|\mathcal{F}_i^n]X_i^n\|_\infty (e^{S_i^n} + e^{S_{i-1}})/2 + \varphi(m)E[E[X_n^2|\mathcal{F}_{n-1}^n]e^{S_{n-1}}] + E[e^{S_{n-1}}] \\
&\le (1 + q^n + \varphi(m)v^n) \sup_{i \le n-1} E[e^{S_i}] \\
&\le e^{q^n + \varphi(m)v^n} \sup_{i \le n-1} E[e^{S_i}]
\end{aligned}$$

where q^n and v^n are the terms corresponding to $k = n$ in the definition of q and v. This proves the result by induction. □

2.2 Applications

2.2.1 Deviation bounds

In this section we give the deviation inequalities that can be deduced from the preceding exponential inequalities. We generalize the Bernstein inequality in Equations (18) and (22), and the Hoeffding inequality in Equation (21); one could get Bennett inequalities through a similar process, we refer to Appendix B of [**?**]. In the martingale case, Equations (19) and (20) do not assume that the variables are bounded, but sums of squares are involved.

Theorem 4. *With the notations of Theorem 1 we have for any $A, y > 0$*

$$P(S \ge A, \langle X \rangle \le y) \le \exp\left(-\frac{A^2}{2(y + 8q) + 2Am/3}\right) \tag{18}$$

$$P(S \ge A, [X_+] + \langle X_- \rangle \le y) \le \exp\left(-\frac{A^2}{2(y + 8q)}\right) \tag{19}$$

$$P(S \ge A, [X] + 2\langle X \rangle \le 3y) \le \exp\left(-\frac{A^2}{2(y + 8q)}\right). \tag{20}$$

With the notations of Theorem 2 and 3 we have for any $A, y > 0$

$$P(S \geq A, \sum_i b_i^2 \leq 4y) \leq \exp\left(-\frac{A^2}{2y+32q}\right) \tag{21}$$

$$P(S \geq A) \leq \exp\left(-\frac{A^2}{2(v+2q)+2Am/3}\right). \tag{22}$$

In the martingale case, (21) remains true if we allow a_i and b_i to be an $\mathcal{F}_{i-1}$-measurable random variable.

REMARK. Equation (21) is analogous to Corollary 3(a) of [6].

Proof. Applying the bound (9) to the variables tX_i for some $t > 0$, we get

$$\begin{aligned}\log P(S \geq A, \langle X\rangle \leq y) &\leq \log E[\exp\{t(S-A) - \frac{t^2\langle X\rangle - t^2 y}{t^2 m^2}(e^{tm} - tm - 1)\}]\\ &\leq 4t^2 q + \frac{y}{m^2}(e^{tm} - tm - 1) - tA\\ &\leq \frac{y+8q}{m^2}(e^{tm} - tm - 1) - tA.\end{aligned}$$

The optimization of this expression w.r.t. $t \geq 0$ is classical in the theory of Bennett and Bernstein inequalities and delivers (18); see for instance the Appendix B of [**?**]. The second inequality is deduced from (10) with the same method: for $V = [X_+] + \langle X_-\rangle$ or $V = ([X] + 2\langle X\rangle)/3$ one has

$$\log P(S \geq A, V \leq y) \leq \log E[e^{tS - tA - t^2(V-y)/2}] \leq 4t^2 q + y\frac{t^2}{2} - tA$$

and we take $t = A/(y+8q)$.

Equations (21) and (22) are obtained similarly. □

2.2.2 Bounded difference inequalities

The above results lead straightforwardly to bounded difference inequalities by using a classical martingale argument of Maurey [**?**]. Equation (26) is the McDiarmid inequality [**?**]. Equation (25) is a Bernstein inequality in the same context.

Theorem 5. *Let $Y = (Y_1, \ldots Y_n)$ be a zero-mean sequence of independent variables with values in some measured space E. Let f be a measurable function on E^n with real values. Set*

$$\begin{aligned}&S = f(Y) - E[f(Y)]\\ &D_k(y,z) = f(Y_1, \ldots Y_{k-1}, y, Y_{k+1}\ldots Y_n) - f(Y_1, \ldots Y_{k-1}, z, Y_{k+1}\ldots Y_n)\\ &\Phi_k = \sup_{y,z} E[D_k(y,z)|Y_1, \ldots Y_{k-1}]\\ &\Delta_k = f(Y) - E[f(Y)|Y_1, \ldots Y_{k-1}, Y_{k+1}\ldots Y_n] = E[D_k(Y_k, Y_k')|Y]\\ &m = \sup_k \ \operatorname{ess\,sup} \Delta_k\end{aligned} \tag{23}$$

where Y_k' is an independent copy of Y_k. We assume the measurability of Φ_k. Then for any $A, y > 0$

$$P(S \geq A, \sum_k \Phi_k^2 \leq 4y) \leq \exp\left(-\frac{A^2}{2y}\right) \tag{24}$$

$$P(S \geq A, \sum_k E[\Delta_k^2 \,|Y_1, \ldots Y_{k-1}] \leq y) \leq \exp\left(-\frac{A^2}{2y + 2Am/3}\right). \tag{25}$$

In particular

$$P(S \geq A) \leq \exp\left(-\frac{2A^2}{\sum_k \delta_k^2}\right) \tag{26}$$

$$\delta_k = \|D_k(Y_k, Y_k')\|_\infty.$$

Remark. Let us mention that if f has the form $f(Y) = \sup_{g\in\Gamma} g(Y)$ for some finite class of functions Γ then, with obvious notations,

$$\begin{aligned} D_k(y,z) &= \sup_{g\in\Gamma} g(Y_1, \ldots Y_{k-1}, y, Y_{k+1} \ldots Y_n) - \sup_{g\in\Gamma} g(Y_1, \ldots Y_{k-1}, z, Y_{k+1} \ldots Y_n) \\ &\leq \sup_{g\in\Gamma}\{g(Y_1, \ldots Y_{k-1}, y, Y_{k+1} \ldots Y_n) - g(Y_1, \ldots Y_{k-1}, z, Y_{k+1} \ldots Y_n)\} \\ &= \sup_{g\in\Gamma} D_k^g(y,z) \end{aligned}$$

in particular $\delta_k \leq \sup_{g\in\Gamma} \delta_k^g$. This is a classical argument in the theory of concentration inequalities.

Proof. We shall utilize (21) and (18) with

$$X_k = E[f(Y)|\mathcal{F}_k] - E[f(Y)|\mathcal{F}_{k-1}]$$
$$\mathcal{F}_k = \sigma(Y_1, \ldots Y_k).$$

We have already pointed out that $q = 0$ since X_k is a martingale difference. Let us define the random variables

$$L_k = \inf_y E[F_k(y)|Y_1, \ldots Y_{k-1}]$$
$$U_k = \sup_y E[F_k(y)|Y_1, \ldots Y_{k-1}].$$

The equation

$$L_k \leq E[f(Y)|\mathcal{F}_k] \leq U_k$$

implies

$$L_k - E[f(Y)|\mathcal{F}_{k-1}] \leq X_k \leq U_k - E[f(Y)|\mathcal{F}_{k-1}]$$

and since $U_k - L_k = \Phi_k$ we can apply (21) with $b_k = \Phi_k$ and get (24).
Clearly X_k rewrites

$$X_k = E[\Delta_k|\mathcal{F}_k]$$

hence $E[X_k^2 \,|\mathcal{F}_{k-1}] \leq E[\Delta_k^2 \,|\mathcal{F}_{k-1}]$, $\langle X\rangle \leq V$, and (25) follows from (18). $\square$

2.2.3 Inequalities for suprema of U-statistics

For some problems of adaptive estimation and testing, it is very important to be able to control the supremum of U-statistics [9]. We give here a bound in this direction.

Consider a sequence of i.i.d. random variables $Y_1, \dots Y_n$ with values on some measurable space E and a finite familly H of measurable symmetric functions on E^d and set for $h \in H$

$$
\begin{aligned}
&Y_I = \{Y_i, i \in I\}, \qquad I \subset \{1, \dots n\} \\
&Z_h(Y) = \frac{1}{\binom{n}{d}} \sum_{I \subset \{1,\dots n\}} h(Y_I) \\
&S = \sup_{h \in H} Z_h(Y) - E[\sup_{h \in H} Z_h(Y)]. \qquad (27)
\end{aligned}
$$

where the sum is restricted to the subsets with cardinality d; since the kernel h is symmetric there is no ambiguity regarding the notation $h(Y_A)$. We assume that h is centered:

$$E[h(Y_1, \dots Y_d)] = 0.$$

It is well known that if Z_h is non degenerate, that is

$$E[h(Y_1, \dots Y_d)|Y_1] \not\equiv 0,$$

then the variance of Z_h has order n^{-1}, cf [13] p.12. We give in the following corollary a deviation bound for S which corresponds to a Gaussian approximation with variance of the same order of magnitude. In the case of degenerate U-statistics we do not get good bounds; this is apparent in the case H has only one element since an abundant litterature exists concerning deviation of degenerate U-statistics [11**?** ; 4].

The function L below may be bounded by $2\|h\|_\infty$, what should be generally good enough unless Y_1 takes a specific value with high probability in which case (29) may become significantly better:

Corollary 6. *The symmetric function h satisfies for some function* $L(x, y)$

$$|h(y_1, \dots y_d) - h(y_1', y_2, \dots y_d)| \leq L(y_1, y_1').$$

Then, for any $A > 0$

$$P(S \geq A) \leq \exp\left(-\frac{2nA^2}{\mu^2}\right), \quad \mu = d \operatorname{ess\,sup} L(Y_1, Y_2) \qquad (28)$$

$$P(S \geq A) \leq \exp\left(-\frac{nA^2}{d^2 E[L(Y_1, Y_1')^2] + 2A\mu/3}\right). \qquad (29)$$

Proof. With the notation of Theorem 5, we have with $Y^k = (Y_1, \dots Y_{k-1}, Y_k', Y_{k+1}, \dots Y_n)$

$$\Delta_k \leq \sup_h Z_h(Y) - \sup_h Z_h(Y^k) \leq \sup_h (Z_h(Y) - Z_h(Y^k))$$

and

$$
\begin{aligned}
|Z_h(Y) - Z_h(Y^k)| &\le \frac{1}{\binom{n}{d}} \Big| \sum_{I \ni k} h(Y_I) - h(Y_I^k) \Big| \\
&\le \frac{1}{\binom{n}{d}} \binom{n-1}{d-1} L(Y_k, Y_k') \\
&= \frac{d}{n} L(Y_k, Y_k').
\end{aligned}
$$

Hence $\Delta_k \le \frac{d}{n} L(Y_k, Y_k')$ and the result is now just a consequence of Theorem 5, noticing that

$$
\begin{aligned}
v &\le \frac{d^2}{2n} E[L(Y_1, Y_1')^2] \\
\|\Delta_k\|_\infty &\le \frac{\mu}{n}.
\end{aligned}
$$

□

3 Second order approach

As mentionned in the introduction, in order to get better results, we shall have to control conditional expectations of products $X_k X_j$; actually, our procedure will rather lead to products like $X_k X_j^k$; hence we are led to introduce for any pair (k, j) a sequence $X_i^{k,j}$ corresponding to increasing dependence with (X_k, X_j^k). More precisely:

Inhomogeneous setting

(i) *For any $1 \le k \le n$ is given a sequence X_j^k, $j = 1 \dots k$, which is a reordering of $(X_j,\ j = 1, \dots k)$ with $X_k^k = X_k$. We attach to each k the σ-algebras*

$$
\mathcal{F}_j^k = \sigma(X_i^k, i \le j), \quad j \le k. \tag{30}
$$

(ii) *For $1 \le j \le k \le n$, sequences $(X_i^{k,j})_{i<j}$ and $(\mathcal{H}_i^{k,j})_{i<j}$ are defined as above from the sequence $(X_i^k)_{i<j}$.*

In other words the σ-field $\mathcal{F}_j^k$ is made by taking off $\mathcal{F}_{j+1}^k$ the "closest" variable to X_k. The σ-field $\mathcal{H}_i^{k,j}$ is made by continuing this process after $\mathcal{F}_j^k$ (hence $i < j$) in a way which may depend on k and j.

This setup is essentially, in a somewhat more general context, what is considered in [6]. For time series, we have $X_i^k = X_i$ and $\mathcal{F}_i^k = \mathcal{H}_i^{k,j} = \mathcal{F}_i = \sigma(X_l, l \le i)$.

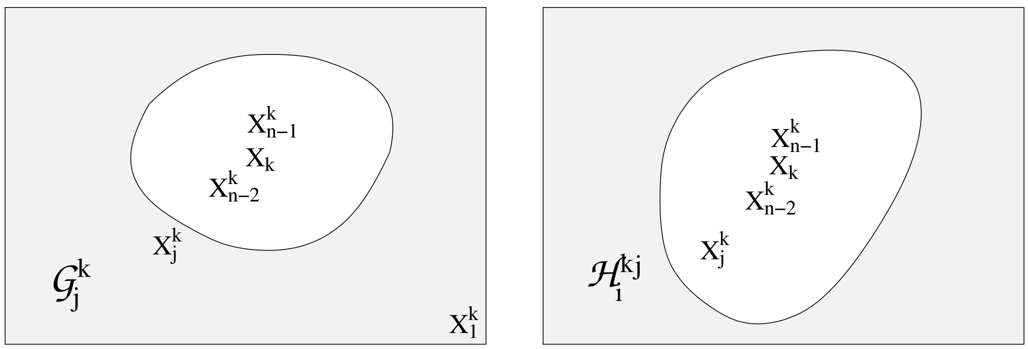
It happens commonly in the theory of random fields that there is no natural order on the variables (this is however typically not the case if there is a time variable). In those cases, instead of building the $\mathcal{F}_i^k$ by reordering the variables which are "before" X_k, it is not more restrictive to assume that for any k there exist a reordering of *all* variables for which the conditional expectations will decrease rapidly enough. In this case the $\mathcal{F}_j^k$ will be replaced by the $\mathcal{G}_j^k$ defined as follows

(i) *For any* $1 \le k \le n$ *is given a sequence* $X_j^k,\ j = 1\dots n$, *which is a reordering of* $(X_j,\ j = 1\dots n)$ *with* $X_n^k = X_k$. *We attach to each* k *the* σ*-algebras*

$$\mathscr{G}_j^k = \sigma(X_i^k, i \le j), \quad j \le n. \tag{31}$$

(ii) *For any* $1 \le j, k \le n$ *is given a sequence* $(X_i^{k,j})_{i<j}$, *which is a reordering of* $(X_i^k)_{i<j}$. *We attach to each such pair* (k, j) *the* σ*-algebras*

$$\mathcal{H}_i^{k,j} = \sigma(X_l^{k,j}, l \le i), \quad i < j. \tag{32}$$



This setting is adequate for dealing with mixing random fields in which case each index k corresponds to some point P_k of the space; for each k, the sequence $(X_j^k)_j$ will be obtained by sorting the original sequence (X_j) in decreasing order of the distance $d(P_j, P_k)$, and $\mathscr{G}_j^k$ will be the σ-field generated by the random variables setting on the j more distant points from P_k, say $P_1^k, \dots P_j^k$; it is natural to define $X_i^{k,j}$ as X_l where P_l is the i-th more distant point from $\{P_k, P_j^k\}$, but the choice $X_i^{k,j} = X_i^k$, $\mathcal{H}_i^{k,j} = \mathscr{G}_i^k$, is typically good enough for random fields over the Euclidean space (see Section 4).

3.1 The homogeneous setting

The following theorem states a Gaussian approximation result (33) and an exponential bound (34). Equation (33) may seem rather overtechnical; the reader can think first of the case $p = 1$, $q = \infty$, which is natural if h has bounded third derivatives; however the proof of (34) utilizes the case $p = \infty$, $q = 1$.

Theorem 7. *Set for* $1 \le p \le +\infty$

$$r_p = \sum_{k,j=1}^{n} \left(\frac{1}{2} \| E[X_k | \mathscr{G}_j^k](X_j^k)^2 \|_p + \sum_{i \le j-1} \| (E[X_k X_j^k | \mathcal{H}_i^{k,j}] - E[X_k X_j^k]) X_i^{k,j} \|_p + \sum_{i \ge j} |E[X_k X_j^k]| \|X_i^k\|_p \right).$$

Let h be a three time differentiable function of n variables, and N be a centered Gaussian vector independent of $X=(X_1,\dots X_n)$ with the same covariance matrix. Then for any $1\le p\le\infty$

$$|E[h(X)]-E[h(N)]|\le\frac{r_p}{3}\sup_{i,j,k}\sup_{0\le t\le 1}\sup_{Y\in\mathcal{X}}\|h'''_{ijk}(tY+\sqrt{1-t^2}N)\|_q \tag{33}$$

where $q=p/(p-1)$, and $\mathcal{X}$ is the set of processes $Y_i=\alpha_iX_i$ where α is allowed to be any sequence of $[0,1]^n$.

The sum $S=\sum_k X_k$ satisfies for any $t>0$

$$E[e^{tS}]\le\exp\left(\sigma_*^2\frac{t^2}{2}+r_\infty\frac{t^3}{3}\right),\qquad \sigma_*^2=\max_{I\subset\{1,2,\dots n\}}E\left[\left(\sum_{i\in I}X_i\right)^2\right]. \tag{34}$$

REMARKS. If the X_k's are independent $r_p=\sum Var(X_k)\|X_k\|_p+\|X_k^3\|_p/2$. Hence under strong enough mixing assumptions, r_p/σ_*^2 is expected to tend to 0 as the number of variables n tends to infinity; in this case, (33) leads in particular to the central limit theorem for S/σ_*.

For the sake of clarity, we shall start the proof with a preparatory lemma: We know that if (Y,Z) is a Gaussian vector and Y is scalar centered, for any differentiable function g one has

$$E[Yg(Z)]=\sum_i E[YZ_i]E[g_i'(Z)],\quad\text{where}\quad g_i'(x)=\frac{d}{dx_i}g(x)$$

under appropriate integrability conditions. We shall see that, when the variables are not Gaussian, bounds of the difference between both sides when Y varies among the coordinates of Z provide an efficient measure, on some sense, of how far Z is from a Gaussian vector of the same covariance (this will appear in Equation (37) below).

Next lemma provides a way to estimate such a bound:

Lemma 8. *Let Y be a centered random variable and Z be a random vector on $\mathbb{R}^n$. For any $1\le j\le n$ is given and sequence $(Z_i^j)_{i<j}$ which is a reordering of $(Z_i)_{i<j}$. Let $\mathcal{A}_j$ and $\mathcal{B}_i^j$, $i<j$, be σ-algebras such that $\mathcal{A}_j\supset\sigma(Z_1,\dots Z_j)$ and $\mathcal{B}_i^j\supset\sigma(Z_1^j,\dots Z_i^j)$ and*

$$r_p=\sum_{j=1}^n\frac{1}{2}\|E[Y|\mathcal{A}_j](Z_j)^2\|_p+\sum_{i=1}^{j-1}\|(E[YZ_j|\mathcal{B}_i^j]-E[YZ_j])Z_i^j\|_p+\sum_{i=j}^n|E[YZ_j]|\;\|Z_i\|_p.$$

Then for any function g two times differentiable on $\mathbb{R}^n$, with first (resp. second) order derivatives denoted by g_j' (resp. g_{ij}''), and $1\le p\le\infty$:

$$|E[Yg(Z)]-\sum_j E[YZ_j]E[g_j'(Z)]|\le r_p\sup_{i,j,\alpha}\|g_{ij}''(\alpha_1Z_1,\dots\alpha_nZ_n)\|_q,\quad\frac{1}{p}+\frac{1}{q}=1 \tag{35}$$

where the supremum over α is taken over all sequences of $[0,1]^n$ having no more than one term different from 0 or 1.

Proof. We set

$$\bar{Z}_j = (Z_1, \dots Z_j, 0, \dots 0)$$

and $\bar{Z}_i^j$ is defined by setting to zero all entries in $\bar{Z}_j$ which are not a Z_l^j for some $l \le i$; hence $\bar{Z}_{j-1}^j = \bar{Z}_{j-1}$ and $\bar{Z}_0^j = 0$.

The key observation is that if all terms of the sums below are integrable we have

$$\begin{aligned} E[Yg(Z)] - \sum_{j=1}^n E[YZ_j]E[g_j'(Z)] =& E\sum_{j=1}^n Y(g(\bar{Z}_j) - g(\bar{Z}_{j-1}) - Z_j g_j'(\bar{Z}_{j-1})) \\ &+ E\sum_{j=1}^n \sum_{i=1}^{j-1} (YZ_j - E[YZ_j])(g_j'(\bar{Z}_i^j) - g_j'(\bar{Z}_{i-1}^j)) \\ &- E\sum_{j=1}^n \sum_{i \ge j} E[YZ_j](g_j'(\bar{Z}_i) - g_j'(\bar{Z}_{i-1})). \end{aligned} \tag{36}$$

Clearly indeed, for any j, the factors of $E[YZ_j]$ are the same on both sides, except a residual $E[YZ_j]g_j'(0)$ on the right hand side which compensates for a remaining term from the otherwise vanishing sum of the $E[YZ_j g_j'(\bar{Z}_i^j)]$ (namely the term $i = 0$); finally the terms $Yg(\bar{Z}_j)$ of the right hand side sum up to $E[Yg(Z)]$ since $E[Y] = 0$.

The general identities

$$f(x+h) - f(x) - hf'(x) = h^2 \int_0^1 (1-t) f''(x+th)dt$$
$$f'(x+h) - f'(x) = h \int_0^1 f''(x+th)dt$$

imply that (e_i is the ith vector of the canonical basis of $\mathbb{R}^n$)

$$g(\bar{Z}_j) - g(\bar{Z}_{j-1}) - Z_j g_j'(\bar{Z}_{j-1}) = Z_j^2 \int_0^1 (1-t) g_{jj}''(\bar{Z}_{j-1} + tZ_j e_j)dt$$
$$g_j'(\bar{Z}_i^j) - g_j'(\bar{Z}_{i-1}^j) = Z_i^j \int_0^1 g_{ij}''(\bar{Z}_{i-1}^j + tZ_i^j e_i)dt$$

hence by the Minkowski inequality

$$\|Z_j^{-2}(g(\bar{Z}^j) - g(\bar{Z}^{j-1}) - Z_j g_j'(\bar{Z}^{j-1}))\|_q \le \sup_\alpha \|g_{ij}''(\alpha_1 Z_1, \dots, \alpha_n Z_n)\|_q \int_0^1 (1-t)dt$$
$$\|(Z_i^j)^{-1}(g_j'(\bar{Z}_i^j) - g_j'(\bar{Z}_{i-1}^j))\|_q \le \sup_\alpha \|g_{ij}''(\alpha_1 Z_1, \dots, \alpha_n Z_n)\|_q$$

and (35) is now only a consequence of (36) and the Hölder inequality. □

Proof of theorem 7. Set for $0 \le t \le 1$

$$\varphi(t) = E[h(tX + \sqrt{1-t^2}N)].$$

Using that for any differentiable function f with bounded derivatives $E[N_k f(N)] = \sum_i E[N_k N_j] E[f_j'(N)]$ we obtain

$$\begin{aligned}
\frac{d}{dt}\varphi(t) =& \sum_k E[X_k h_k'(tX + \sqrt{1-t^2}N)] - \frac{t}{\sqrt{1-t^2}} \sum_k E[N_k h_k'(tX + \sqrt{1-t^2}N)] \\
=& \sum_k E[X_k h_k'(tX + \sqrt{1-t^2}N)] - t \sum_{k,j} E[X_k X_j] E[h_{kj}''(tX + \sqrt{1-t^2}N)]. \qquad (37)
\end{aligned}$$

Applying n times Lemma 8 with $Y = X_k$, $Z_j = X_j^k$, $Z_i^j = X_i^{k,j}$, $\mathcal{A}_j = \mathcal{G}_j^k$, $\mathcal{B}_i^j = \mathcal{H}_i^{k,j}$ and $g(x) = E[h_k'(tx + \sqrt{1-t^2}N)]$, for $k = 1, \dots n$ we get

$$|\varphi'(t)| \le t^2 r_p \sup_{i,j,k,Y \in \mathcal{X}} \|h_{ijk}'''(tY + \sqrt{1-t^2}N)\|_q$$

where $\mathcal{X}$ is the family of the processes of the form $Y_i = \alpha_i X_i$, $\alpha \in [0,1]^n$. This leads to (33).

Concerning (34), notice first that an elementary scaling argument reduces to the case $t = 1$. We consider now the function $h(x) = \exp(\sum x_i)$:

$$\varphi(t) = e^{(1-t^2)\sigma^2/2} E[h(tX)], \quad \sigma^2 = E[S^2]$$

and Equation (37) rewrites:

$$\begin{aligned}
|\varphi'(t)| =& |\sum_k E[X_k h_k'(tX))] e^{\sigma^2(1-t^2)/2} - t \sum_{k,i} E[X_k X_j] E[h_{kj}''(tX)] e^{\sigma^2(1-t^2)/2}| \\
\le& t^2 e^{\sigma^2(1-t^2)/2} r_\infty \sup_{Y \in \mathcal{X}} E[h(tY)] \qquad (38)
\end{aligned}$$

by application of Lemma 8 since $h_{ijk}''' = h$. Hence it we set

$$\psi(t) = e^{\sigma_*^2(1-t^2)/2} \sup_{Y \in \mathcal{X}} E[h(tY)]$$

(the reason for the choice of σ_* rather than σ will appear soon) Equation (38) implies

$$\varphi(t) \le \varphi(0) + r_\infty \int_0^t s^2 \psi(s) ds \le \psi(0) + r_\infty \int_0^t s^2 \psi(s) ds. \qquad (39)$$

The variance of any $Y \in \mathcal{X}$ is smaller than σ_*^2 due to the fact that the convex function on $[0,1]^n$

$$\alpha \mapsto E[\langle \alpha, X \rangle^2]$$

attains necessarily its maximum on extremal points of $[0,1]^n$. Hence (39) is again valid if instead of X we put any $Y \in \mathcal{X}$ in the definition of φ since r_∞ and σ_* would be smaller, thus

$$\psi(t) \le \psi(0) + r_\infty \int_0^t s^2 \psi(s) ds$$

which implies via the Gronwall lemma that

$$\psi(t) \le \psi(0) \exp(r_\infty t^3/3) = \exp(\sigma_*^2/2 + r_\infty t^3/3).$$

If we set $t = 1$ in this equation, we get (34) for $t = 1$. □

3.2 The inhomogeneous setting

In this section we give a companion to Theorem 7 in the inhomogeneous case.

Theorem 9. *Set*

$$\sigma^2 = E[S^2]$$
$$\sigma_m^2 = \sup_{k\le n} E[S_k^2]$$
$$v = \sum_k E[X_k^2] + 2E[X_k S_{k-1}]_+$$
$$w_p = \sum_{k=1}^n \sum_{i=1}^k \left(\frac{1}{2} \|(X_i^k)^2 E[X_k|\mathcal{F}_i^k]\|_p + \sum_{j>i} \|(E[X_k X_j^k|\mathcal{H}_i^{k,j}] - E[X_k X_j^k]) X_i^{k,j}\|_p + \sum_{j\le i} |E[X_k X_j^k]|\ \|X_i^k\|_p \right).$$

Then for any $a \in \mathbb{R}$

$$|E[e^{iaS}] - e^{-a^2\sigma^2/2}| \le |a|^3 w_1 e^{a^2(\sigma_m^2 - \sigma^2)/2}, \tag{40}$$
$$E[e^{aS}] \le \exp\left(v\frac{a^2}{2} + w_\infty \frac{|a|^3}{3} \right). \tag{41}$$

REMARK. In the case of a martingale, $\sigma_m^2 = \sigma^2 = v$ and $w_p = \sum Var(X_k)\|X_k\|_p + \|X_k^3\|_p/2$.

Proof. Let $\lambda \in \mathbb{C}$. Set $S(t) = S_{n-1} + tX_n$, and

$$\varphi(t) = E[e^{\lambda S(t)}].$$

The derivative of φ is :

$$\varphi'(t) = \lambda E[X_n e^{\lambda S(t)}] = \lambda^2 E[X_n S(t)]\varphi(t) + w(t) \tag{42}$$

where, thanks to Lemma 8 with $Y = X_n$, $Z = (X_1^n, \dots X_{n-1}^n, tX_n)$, $g(x) = e^{\lambda \sum x_i}$,

$$|w(t)| \le |\lambda|^3 \sup_{Y\in\mathcal{X}} \|e^{\lambda(Y_1+\cdots+Y_{n-1}+tY_n)}\|_p w_q^n \tag{43}$$

and $\mathcal{X}$ is the family of the processes of the form $Y_i = \alpha_i X_i$ where α is any decreasing sequence of $[0,1]^n$ with no more than one term different from 0 or 1, and w_q^n is the term corresponding to $k = n$ in the expression of w_q. Integrating (42) we get

$$\varphi(t)e^{-\lambda^2 \int_0^t E[X_n S(s)]ds} = \varphi(0) + \int_0^t e^{-\lambda^2 \int_0^s E[X_n S(u)]du} w(s)ds$$

and since $E[X_n S(t)]$ is half the derivative of $\sigma(t)^2 = E[S(t)^2]$, this rewrites

$$\varphi(t)e^{-\lambda^2\sigma(t)^2/2} = \varphi(0)e^{-\lambda^2\sigma(0)^2/2} + \int_0^t e^{-\lambda^2\sigma(s)^2/2} w(s)ds. \tag{44}$$

If $\lambda = ia \in i\mathbb{R}$, taking $p = \infty$ in (43), (44) implies

$$|\varphi(t)e^{a^2\sigma(t)^2/2} - \varphi(0)e^{a^2\sigma(0)^2/2}| \leq |a|^3 w_1^n \exp(a^2 \sup_{0\leq t\leq 1} \sigma(t)^2/2).$$

Now since the function $\sigma(t)$ is convex, its supremum over $[0,1]$ is either $\sigma(0)$ or $\sigma(1)$ hence

$$|\varphi(1)e^{a^2\sigma(1)^2/2} - \varphi(0)e^{a^2\sigma(0)^2/2}| \leq |a|^3 w_1^n e^{a^2\sigma_m^2/2}$$

which implies (40) by induction on n. Now for a fixed real $\lambda \in \mathbb{R}$, let

$$\varphi_*(t) = \sup_{Y\in\mathcal{X}} E[e^{\lambda(Y_1+\cdots+Y_{n-1}+tY_n)}].$$

Equations (43) and (44) with $p = 1$ imply

$$\varphi(t) \leq \varphi_*(0)e^{\lambda^2(\sigma(t)^2-\sigma(0)^2)/2} + |\lambda|^3 w_\infty^n \int_0^t e^{\lambda^2(\sigma(t)^2-\sigma(s)^2)/2}\varphi_*(s)ds.$$

We have for $t \geq s$

$$\sigma(t)^2 - \sigma(s)^2 = 2(t-s)E[X_nS_{n-1}] + (t^2-s^2)E[X_n^2] \leq 2(t-s)E[X_nS_{n-1}]_+ + (t^2-s^2)E[X_n^2].$$

Hence, if we set $u(t) = tE[X_nS_{n-1}]_+ + t^2E[X_n^2]/2$

$$\varphi(t) \leq \varphi_*(0)e^{\lambda^2(u(t)-u(0))} + |\lambda|^3 w_\infty^n \int_0^t e^{\lambda^2(u(t)-u(s))}\varphi_*(s)ds.$$

For any $Y \in \mathcal{X}$, the same bound holds if X is replaced by Y in the definition of φ, since the corresponding values of w_∞^n and $u(t)-u(s)$ will be smaller (either $\alpha_n = 0$ and the corresponding value of $u(t)-u(s)$ is zero, or $Y_n = \alpha_n X_n$ and $Y_i = X_i$, $i < n$). Hence

$$\varphi_*(t) \leq \varphi_*(0)e^{\lambda^2(u(t)-u(0))} + |\lambda|^3 w_\infty^n \int_0^t e^{\lambda^2(u(t)-u(s))}\varphi_*(s)ds$$

or

$$\varphi_*(t)e^{-\lambda^2u(t)} \leq \varphi_*(0)e^{-\lambda^2u(0)} + |\lambda|^3 w_\infty^n \int_0^t e^{-\lambda^2u(s)}\varphi_*(s)ds,$$

and by Gronwall's Lemma:

$$\varphi_*(t)e^{-\lambda^2u(t)} \leq \varphi_*(0)e^{-\lambda^2u(0)^2}e^{t|\lambda|^3w_\infty^n}$$

which gives for $t = 1$

$$\varphi_*(1) \leq \varphi_*(0)e^{\lambda^2(E[X_nS_{n-1}]_+ + E[X_n^2]/2)+|\lambda|^3w_\infty^n}.$$

This proves (41) by induction on n. $\square$

3.3 Applications to deviation bounds

In this section we give the deviation inequalities that can be deduced from the preceding exponential inequalities. The bound (46) is analogous to the bound or Corollary 3 (b) in [6] p.85, but in this paper the variance v is amplified with an extra factor $2e^2$ and w_∞ takes a slightly different value.

Theorem 10. *With the notations of Theorem 7 and Theorem 9, we have for any $A > 0$*

$$P(S \geq A) \leq \exp\left(-\frac{A^2}{2\sigma_*^2 + 2\sqrt{2Ar_\infty/3}}\right), \tag{45}$$

$$P(S \geq A) \leq \exp\left(-\frac{A^2}{2v + 2\sqrt{2Aw_\infty/3}}\right). \tag{46}$$

Proof. Using (34) we get

$$\log P(S \geq A) \leq \log E[e^{t(S-A)}] \leq \sigma_*^2 \frac{t^2}{2} + r_\infty \frac{t^3}{3} - tA.$$

We choose $t = A/(\sigma_*^2 + \sqrt{2Ar_\infty/3})$; in particular $t \leq \sqrt{3A/2r_\infty}$ (value of t if $\sigma_* = 0$) hence

$$\log P(S \geq A) \leq \sigma_*^2 \frac{t^2}{2} + \sqrt{2Ar_\infty/3}\,\frac{t^2}{2} - tA = -\frac{tA}{2} = -\frac{A^2}{2\sigma_*^2 + 2\sqrt{2Ar_\infty/3}}.$$

Equation (46) is obtained similarly on the basis of (41). □

3.4 More on bounded difference inequalities

We give a second order variant of Equation (26).

Theorem 11. *With the notation of Theorem 5, the following inequality holds true*

$$P(S \geq A) \leq \exp\left(-\frac{A^2}{2Var(S) + 2\sqrt{A\sum_k \delta_k^3}}\right).$$

Proof. As in the proof of Theorem 5, we notice that S is the sum of the following martingale increments

$$X_k = E[f(Y)|\mathcal{F}_k] - E[f(Y)|\mathcal{F}_{k-1}] = E[\Delta_k|\mathcal{F}_k]$$
$$\mathcal{F}_k = \sigma(Y_1, \ldots Y_k)$$

but now we use Equation (46). As pointed out right after Theorem 9, one has in the martingale case:

$$w_\infty = \sum Var(X_k)\|X_k\|_\infty + \|X_k^3\|_\infty/2 \leq \frac{3}{2}\sum \|X_k^3\|_\infty$$

and we conclude by noticing that $\|X_k^3\|_\infty \leq \delta_k$. □

3.5 Triangle counts

We shall show that the standard Gaussian approximation is asymptotically valid for triangle counts in the moderate deviation domain.

In the Erdös-Rényi model of an unoriented random graph with n vertices, edges are represented by $\binom{n}{2}$ i.i.d. Bernoulli variables Y_{ab}, $1 \le a < b \le n$, with the convention $Y_{ab} = Y_{ba}$ and $Y_{aa} = 0$. The number of triangles in such a model is

$$Z = \sum_{\{a,b,c\}} Y_{ab} Y_{bc} Y_{ac}.$$

We set $p = E[Y_{12}]$ and

$$\begin{aligned} X_{abc} &= Y_{ab} Y_{bc} Y_{ac} - p^3 \\ S &= \sum_{\{a,b,c\}} Y_{ab} Y_{bc} Y_{ac} - p^3 = \sum_{\{a,b,c\}} X_{abc} = \sum_{\tau \in T} X_\tau \end{aligned} \tag{47}$$

where T is the set of subsets of $\{1,\dots n\}$ with three elements, $|T| = \binom{n}{3}$. Recall that r_∞ rewrites, with these notations

$$r_\infty = \sum_{\tau \in T, j \le |T|} \left(\frac{1}{2} \|E[X_\tau | \mathcal{G}_j^\tau](X_j^\tau)^2\|_\infty + \sum_{i \le j-1} \|(E[X_\tau X_j^\tau | \mathcal{H}_i^{\tau,j}] - E[X_\tau X_j^\tau]) X_i^{\tau,j}\|_\infty \right. \\ \left. + \sum_{i \ge j} |E[X_\tau X_j^\tau]| \|X_i^\tau\|_\infty \right).$$

For any $\tau = \{a,b,c\}$, define A_τ the set elements of T such that at least two points are in common with τ (τ, $\{a,b,d\}$, $\{a,d,c\}$, ...); this makes $n_1 = 1 + 3(n-3)$ elements, and $(X_\sigma)_{\sigma \notin A_\tau}$ is independent of X_τ. We define the ordering X_j^τ by taking first (j large) these X_σ for which $\sigma \in A_\tau$ and define $\mathcal{G}_j^\tau$ according to (31). Since $\mathcal{G}_j^\tau$ is independent of X_τ for $j \le |T| - n_1$, the first term in r_p will be 0 unless $j > |T| - n_1$, hence

$$\sum_{\tau \in T, j \le |T|} \frac{1}{2} \|E[X_\tau | \mathcal{G}_j^\tau](X_j^\tau)^2\|_\infty \le \frac{1}{2} |T| n_1$$

and

$$\sum_{\tau \in T, j \le |T|} \sum_{i \ge j} |E[X_\tau X_j^\tau]| \|X_i^\tau\|_\infty \le |T| n_1^2.$$

Now since $\mathcal{H}_i^{\tau,j} \subset \mathcal{G}_j^\tau$, X_τ is independent of $(X_j^\tau, \mathcal{H}_i^{\tau,j})$ for $j \le |T| - n_1$, and for these j, the second term vanishes, independently of the construction of $\mathcal{H}_i^{\tau,j}$.

We consider now the case $j > |T| - n_1$. In this case, and if $j \ne |T|$, $X_j^\tau = X_{\tau^j}$ for some τ^j which has two points in common with τ, for instance $\{a,b,d\}$. We define the σ-fields $\mathcal{H}_i^{\tau,j}$ by excluding first the σ which has at least two points in common with τ or τ^j, this makes at most $n_2 = 3(n-3) + 1 +$

$2(n-3)=1+5(n-3)$ points to exclude. If $j=|T|$, we take $\mathcal{H}_i^{\tau,|T|}=\mathcal{G}_i^\tau$, and this contributes to at most n_1 non-zero terms in the middle sum; finally:

$$\sum_{\tau\in T, j\le |T|}\sum_{i\le j-1}\|(E[X_\tau X_j^\tau|\mathcal{H}_i^{\tau,j}]-E[X_\tau X_j^\tau])X_i^{\tau,j}\|_\infty\le |T|(n_1n_2+n_1).$$

Hence

$$r_\infty\le |T|\left(\frac{1}{2}n_1+n_1^2+n_1(n_2+1)\right)=|T|n_1(8(n-3)+4)\le 4n^5$$

and Equation (45) rewrites

$$P(S\ge A)\le \exp\left(-\frac{A^2}{2\sigma_*^2+4\sqrt{2n^5A/3}}\right). \tag{48}$$

In this case it is easily verified that $\sigma_*^2=Var(S)$ since covariances are non-negative. Let us recall that (see [2])

$$Var(S)=\binom{n}{3}(p^3-p^6)+\binom{n}{4}\binom{4}{2}(p^5-p^6).$$

This has order n^4 due to the covariance terms. Let us briefly compare with the bound of [2]; this paper delivers a bound for $P(S\ge A)$ which is sligthly larger than

$$\exp\left(-\frac{A^2}{6nE[Z]+16\sqrt{E[Z]}A}\right),\qquad E[Z]=\frac{n(n-1)(n-2)}{6}p^3 \tag{49}$$

(the actual formula is much more complicated; we have used that $\min(a^{-1},b^{-1})\le 2/(a+b)$ to obtain this from Theorem 18 of [2]). One has $6nE[Z]\ge\frac{2}{p^2-p^3}2\sigma_*^2$. For p fixed and n large, the square root term in (48) is residual if $A\ll n^3$; this is the moderate deviation case since the centering term in (47) has order n^3 (notice that $S\le n^3$ w.p.1), and we get the right variance. In (49), a change occurs when $n^{5/2}\ll A$, and if we set $A=Bn^{5/2}$ with B large, (49) leads to $\exp(-cnB)$ while (47) behaves like $\exp(-cnB^2)$.

4 Evaluation of constants under mixing assumptions

We give here informally some arguments to convince the reader that under standard φ-mixing assumptions the constant q has the same order as the variance of the sum, and that w_∞ and r_∞ will be small; under β-mixing assumptions one can control only r_s for $s<\infty$ and w_1.

For more details about mixing we refer to [6], [7] and [3] (particularly Section 8 concerning random fields).

4.1 φ-mixing

The φ-mixing constant between two σ-fields $\mathcal{A}$ and $\mathcal{B}$ is defined as

$$\varphi(\mathcal{A},\mathcal{B}) = \sup_{A\in\mathcal{A},B\in\mathcal{B}} |P(B|A) - P(B)|.$$

It is well known, see reference [?] p.27 or [12] p.278, that this implies that if Z is a zero-mean $\mathcal{B}$-measurable random variable

$$\|E[Z|\mathcal{A}]\|_\infty \le 2\varphi(\mathcal{A},\mathcal{B})\|Z\|_\infty.$$

Assume that X is a field over a part of $\mathbb{Z}^d$: each variable X_k of the field sits on some $P_k \in \mathbb{Z}^d$. $\mathcal{G}_j^k$ is the σ-field generated by the j more distant points from P_k, and we take for simplicity $\mathcal{H}_i^{k,j} = \mathcal{G}_i^k$. Notice that the distance of P_k to P_j^k, the jth closest point, is at least $cj^{1/d}$, for some constant c. This implies that standard φ-mixing assumptions between $\sigma(X_k)$ and $\mathcal{G}_j^k$ rewrite

$$\varphi(\mathcal{G}_j^k,\sigma(X_k)) \le \varphi_{\infty,1}((n-j)^{1/d}) \tag{50}$$

for some decreasing function $\varphi_{\infty,1}$ (for example exponential decay holds for finite range[1] shift-invariant Gibbs random fields [10] pp. 158-159; this contains a lot of examples). The subscripts ∞ and 1 on φ mean that there is no restriction on the number of random variables containned in the first σ-field, $\mathcal{G}_j^k$, and there is only 1 variable in the second, $\sigma(X_k)$; we use this traditionnal notation, in particular for compatibility with [6].

On the other hand for $i < j$, $j-i$ is smaller than the number of points in the annulus $\{x : \|P_j^k - P_k\| \le \|x - P_k\| \le \|P_i^k - P_k\|\}$, in particular, for some c

$$j - i \le c\|P_i^k - P_k\|^{d-1}\|P_j^k - P_i^k\|.$$

Hence $\|P_j^k - P_i^k\|$ is at least $c(j-i)(n-i)^{1/d-1}$ for some c. This implies that standard φ-mixing assumptions between $\sigma(X_k, X_j^k)$ and $\mathcal{G}_i^k$, rewrite

$$\varphi(\mathcal{G}_i^k,\sigma(X_k,X_j^k)) \le \varphi_{\infty,2}((j-i)(n-i)^{1/d-1}), \quad i \le j \tag{51}$$

for some decreasing function $\varphi_{\infty,2}$.

Equations (50) and (51) will imply, for $i \le j \le k$, and any measurable bounded functions f and g

$$\|E[f(X_k)|\mathcal{G}_j^k]\|_\infty \le 2\|f(X_k)\|_\infty\varphi_{\infty,1}((n-j)^{1/d})$$
$$\|E[g(X_k,X_j^k)|\mathcal{G}_i^k] - E[g(X_k,X_j^k)]\|_\infty \le 2\|g(X_k,X_j^k)\|_\infty\varphi_{\infty,2}((j-i)(n-i)^{1/d-1}).$$

The first equation leads to

$$\|E[X_k|\mathcal{G}_j^k]\|_\infty \le 2\|X_k\|_\infty\varphi_{\infty,1}((n-j)^{1/d})$$
$$\|E[X_kX_j^k|\mathcal{G}_i^k] - E[X_kX_j^k]\|_\infty \le 2m\|E[X_k|\mathcal{G}_j^k]\|_\infty \le 4m\|X_k\|_\infty\varphi_{\infty,1}((n-j)^{1/d})$$

[1]This means the existence of a constant c such that if $k,l \in \mathbb{Z}^d$ are such that $d(k,l) > c$ then X_k and X_l are independent conditionally to $(X_j)_{j\notin\{k,l\}}$.

with $m = \sup_i \|X_i\|_\infty$, and the second

$$\|E[X_k X_j^k | \mathcal{G}_i^k - E[X_k X_j^k]\|_\infty \le 2m\|X_k\|_\infty \varphi_{\infty,2}((j-i)(n-i)^{1/d-1}).$$

Hence

$$\|E[X_k X_j^k | \mathcal{G}_i^k] - E[X_k X_j^k]\|_\infty \le 4m\|X_k\|_\infty \min(\varphi_{\infty,2}((j-i)(n-j)^{1/d-1}), \varphi_{\infty,1}((n-j)^{1/d})).$$

We get with $m_1 = \sum \|X_i\|_\infty$, and setting $\varphi(x) = \varphi([x])$

$$q \le \sum_{k \ge i} \|X_i^k\|_\infty \|E[X_k | \mathcal{G}_i^k]\|_\infty \le 2m_1 m \int_0^\infty \varphi_{\infty,1}(x^{1/d})dx = 2dm_1 m \int_0^\infty \varphi_{\infty,1}(x)x^{d-1}dx.$$

The integral is essentially the quantity $B(\phi)$ of [6], and Equations (21) and (22) may be seen as improvements over (b)(i) and (ii) of Corollary 4 of [6]. We get for r_∞

$$\begin{aligned}
r_\infty &\le 3dm^2 m_1 \int_0^\infty \varphi_{\infty,1}(y)y^{d-1}dy + 4m^2 m_1 \int_0^\infty \int_0^\infty \min(\varphi_{\infty,2}((y-z)y^{1/d-1}), \varphi_{\infty,1}(z^{1/d}))1_{z<y}dydz \\
&\le 3dm^2 m_1 \int_0^\infty \varphi_{\infty,1}(y)y^{d-1}dy + 4m^2 m_1 \int_0^\infty \int_0^\infty \varphi_{\infty,2}((y-z)y^{1/d-1}1_{y>2z} + \varphi_{\infty,1}(z^{1/d})1_{z<y<2z})dydz \\
&\le Cm^2 m_1 \int_0^\infty (\varphi_{\infty,1}(y) + \varphi_{\infty,2}(y))y^{2d-1}dy
\end{aligned}$$

for some constant C which depends only on d. The integral in the right hand side is essentially the $D(\phi)$ of [6] page 86. If we refer to the independent case (X_k of order $1/\sqrt{n}$) the factor $m^2 m_1$ is of order $1/\sqrt{n}$, what makes the factor of t^3 in (34) residual as far as t is smaller than $\sqrt{n}$ (moderate deviations).

The same estimates can be obtained for w_∞.

4.2 β-mixing

φ-mixing is not always a realistic assumption: for a Markov chain, φ-mixing implies typically a Doeblin condition; it is satisfied for ergodic finite state Markov chain, but on the other hand, a non-trivial Gaussian autoregressive process is not φ-mixing.

β-mixing is a much more satisfactory measure of independence, cf [3]. The β-mixing constant between two σ-fields $\mathcal{A}$ and $\mathcal{B}$ represent the total variation between the actual measure on $\mathcal{A} \otimes \mathcal{B}$ and the independence (product of marginal measures). A non-singular autoregressive process, as most Markov chains, is β-mixing with exponentially decreasing coefficients [3]. If X is $\mathcal{A}$-measurable and $\beta(\mathcal{A}, \mathcal{B})$ is the β-mixing constant of these σ-fields, one has

$$\|E[X|\mathcal{B}]\|_s \le 2\beta(\mathcal{A}, \mathcal{B})^{1/q}\|X\|_p, \qquad \frac{1}{p} + \frac{1}{q} = \frac{1}{s}$$

(elementary consequence of Theorem 1.4 (a) of [**?**]). This implies that w_1 as well as r_s, $s < \infty$, can be bounded as above using the β-mixing constants.

A Technical inequalities

Proposition 12. *The three pairs of functions*

$$\theta(x) = 0, \quad \psi(x) = e^x - x - 1$$
$$\theta(x) = \zeta(x_+), \quad \psi(x) = \zeta(x_-), \quad \zeta(x) = e^{-x} + x - 1$$
$$\theta(x) = \frac{x^2}{6}, \quad \psi(x) = \frac{x^2}{3}$$

satisfy

$$e^{x-\theta(x)} \leq 1 + x + \psi(x). \tag{52}$$

We have also

$$|\tanh(x - \theta(x) - \log(1 + \psi(y)))| \leq 3m, \quad |x|, |y| \leq m. \tag{53}$$

Proof. Everything will be more or less based on the inequality $e^x \leq 1 + x$.

Equation (52) is obvious for the first pair of functions. In the second case we have only to check for $x > 0$; since in this case $x - \theta(x) = 1 - e^{-x}$ this reduces to proving that:

$$1 - e^{-x} \leq \log(1 + x).$$

The function $\log(1 + x) + e^{-x} - 1$ has a derivative $(1 + x)^{-1} - e^{-x}$ which is ≥ 0 since $e^x \geq 1 + x$; hence the inequality is satisfied. The third case is the non negativity of the function

$$f(x) = 1 + x + \frac{x^2}{3} - e^{x - x^2/6}.$$

This function satisfies

$$f'(x) = 1 + \frac{2x}{3} - (1 - \frac{x}{3})e^{x-x^2/6}$$
$$f''(x) = \frac{2}{3} - ((1 - \frac{x}{3})^2 - \frac{1}{3})e^{x-x^2/6} = \frac{2}{3}\left(1 - (1 - x + x^2/6)e^{x-x^2/6}\right)$$

which is non negative since $1 - (1 - u)e^u \geq 0$; f is convex. Since $f'(0) = 0$ and $f(0) = 0$, we conclude that f is non negative.

For the last inequality, we start with an upper bound on $\psi(y)$. In the first case

$$\psi(y) \leq \psi(|y|) \leq \psi(m) \leq e^m - 1.$$

In the second case $\psi(y) \leq \psi(-m) \leq e^m - 1$ and in the third case $\psi(y) \leq m^2/3 \leq e^m - 1$ (because of the expansion of the exponential); in any case we have

$$\log(1 + \psi(y)) \leq m. \tag{54}$$

On the one hand

$$\tanh(x - \theta(x) - \log(1 + \psi(y))) \leq \tanh(x) \leq m$$

and on the other hand, thanks to (54), using that $\theta(x) \leq x^2/2$

$$\begin{aligned}\tanh(\log(1+\psi(y)) - x + \theta(x) &\leq \tanh(2m + m^2/2)\\ &\leq \min(2m + m^2/2, 1)\\ &\leq 3m.\end{aligned}$$

□

Proposition 13. *For any $a \leq 0 \leq b$ one has*

$$e^a\,\frac{e^b - 1}{b} \leq 4\exp\left(\min\left(\frac{b^2}{8}, a+b\right)\right). \tag{55}$$

Proof. We have indeed

$$\begin{aligned}\exp\left(-\frac{b^2}{8}\right)\frac{e^b-1}{b} &= \exp\left(-\frac{b^2}{8}\right)\frac{(e^{b/2}+1)(e^{b/2}-1)}{b}\\ &= e^{-b^2/8}(e^{b/2}+1)^2\frac{\tanh(b/4)}{b}\\ &\leq (e^{b/2-b^2/16} + e^{-b^2/16})^2\,\frac{1}{4}\\ &\leq \frac{1}{4}(e+1)^2\\ &< 4.\end{aligned}$$

Hence

$$e^a\frac{e^b-1}{b} \leq 4\exp\left(\min\left(\frac{b^2}{8}\right)\right).$$

And clearly

$$e^{-a-b}e^a\frac{e^b-1}{b} = \frac{1-e^{-b}}{b} \leq 1.$$

□

References

[1] B. Bercu, A. Touati, Exponential inequalities for self-normalized martingales with applications, *Ann. Appl. Probab.* 13 (2008), no. 5, 1848–1869. MR2462551

[2] S. Boucheron, G. Lugosi, P. Massart, Concentration inequalities using the entropy method, *Ann. Probab.* 31 (2003), no. 3, 1583–1614. MR1989444

[3] R.C. Bradley, Basic properties of strong mixing conditions. A survey and some open questions. Update of, and a supplement to, the 1986 original. MR0899990 *Probab. Surv.* 2 (2005), 107–144.

[4] J. Bretagnolle, A new large deviation inequality for U-statistics of order 2, *ESAIM: Probability and Statistics* 3 (1999), p. 151-162. MR1742613

[5] O. Catoni, Laplace transform estimates and deviation inequalities. *Ann. Inst. H. Poincaré Probab. Statist.* 39 (2003), no. 1, 1–26. MR1959840

[6] J. Dedecker, Exponential Inequalities and Functional central limit theorems for random fields, *ESAIM* 5 (2001), 77-104. MR1875665

[7] P. Doukhan, *Mixing: Properties and Examples*, *Lecture Notes in Stat.* 85, 1994. MR1312160

[8] P. Doukhan, Paul, P. Massart, E. Rio, Emmanuel, Invariance principles for absolutely regular empirical processes, *Ann. Inst. Henri Poincaré, Probab. Stat.* 31 (1995), no. 2, 393-427. MR1324814

[9] M. Fromont, B. Laurent, Adaptive goodness-of-fit tests in a density model. *Ann. Statist.* 34 (2006), no. 2, 680–720. MR2281881

[10] H-O. Georgii, *Gibbs Measures and Phase Transitions*, de Gruyter Studies in Mathematics 9, Walter de Gruyter & Co., 1988. MR0956646

[11] E. Giné, R. Latala, J. Zinn, Exponential and moment inequalities for U-statistics. *High dimensional probability, II*, 13–38, Progr. Probab., 47 (2000), Birkhauser Boston, Boston, MA, 2000. MR1857312

[12] P. Hall, C.C. Heyde, *Martingale Limit Theory and its Applications*, Academic Press, 1980. MR0624435 MR0624435

[13] A.J. Lee, *U-Statistics : theory and practice*, Statistics: Textbooks and Monographs, 110. M Marcel Dekker, 1990. MR1075417

[14] T. Lumley, An Empirical Process Limit Theorem for Sparsely Correlated Data, *UW Biostatistics Working Paper Series* (2005). Working Paper 255.