Ann. I. H. Poincaré – PR 39, 1 (2003) 1–26

S0246-0203(02)00017-1/FLA

LAPLACE TRANSFORM ESTIMATES AND DEVIATION INEQUALITIES

ESTIMÉES DE LA TRANSFORMÉE DE LAPLACE ET INÉGALITÉS DE DÉVIATION

Olivier CATONI

Laboratoire de probabilités et modèles aléatoires, U.M.R. 7599 du C.N.R.S., case 188, Université Paris 6, bureau 4 E 19, bât Chevaleret, 4, place Jussieu, 75252 Paris cedex 05, France

Received 16 February 2000, revised 11 December 2001

ABSTRACT. – We derive deviation inequalities from non-asymptotic bounds of the log-Laplace transform of a function of N random variables. We assume either that these random variables are independent or that they form a Markov chain. We assume also that the partial finite differences of order one and two of the function are suitably bounded, or more generally that they have some exponential moments. The estimates we get are sharp enough to induce a central limit theorem when N goes to infinity and to prove non-asymptotic "almost Gaussian" deviation bounds.

MSC: 60E15; 60F05; 94A17; 60G42

Keywords: Concentration of product measures; Deviation inequalities; Markov chains; Maximal coupling; Central limit theorem

RÉSUMÉ. – Nous démontrons des inégalités de déviation à partir d'estimées du logarithme de la transformée de Laplace d'une fonction de N variables aléatoires. Nous supposons soit que ces variables sont indépendantes, soit qu'elles forment une chaîne de Markov. Nous supposons aussi que les différences finies partielles d'ordre un et deux de la fonction sont convenablement bornées, ou plus généralement qu'elles ont certains moments exponentiels. Nos estimées sont suffisamment précises pour induire un théorème de la limite centrale quand N tend vers l'infini et pour prouver des inégalités de déviation "presque Gaussiennes".

E-mail address: catoni@ccr.jussieu.fr (O. Catoni).

Introduction

We are going to give "almost Gaussian" finite sample bounds for the log-Laplace transform of some functions of the type $f(X_1, \ldots, X_N)$, where the random variables $(X_1, \ldots, X_N)$ are assumed to be independent, or to form a Markov chain.

We will use throughout the paper a normalisation that parallels the classical case of the sum

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i$$

of real valued random variables. As is usual in these matters, we will deduce from upper log-Laplace estimates finite sample almost sub-Gaussian deviation inequalities for $f(X_1, \ldots, X_N)$. We will also obtain that the central limit theorem holds when N goes to infinity. Although limit laws are not the main subject of this paper, they will be a guide for using a relevant normalisation of constants.

To make sure that this is feasible, it is necessary to make assumptions not only on the first order partial derivatives (or more generally first order partial finite differences) of f, but also on the second order partial derivatives (or more generally on the second order partial finite differences) of f. Indeed, the simple example of

$$f(X_1, \ldots, X_N) = g\left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i\right),$$

should immediately convince the reader that Lipschitz conditions are not enough to enforce a Gaussian limit.

A proper normalisation being chosen, we will be interested in expansions of $Z \stackrel{\text{def}}{=} f(X_1, \ldots, X_N) - \mathbb{E}(f(X_1, \ldots, X_N))$ of the type

$$\log \mathbb{E}\bigl(\exp(\lambda Z)\bigr) = \frac{\lambda^2}{2} \mathbb{V}(Z) + \cdots,$$

where λ is "of order one", $\mathbb{V}(Z) = \mathbb{E}\{[Z - \mathbb{E}(Z)]^2\}$ and the remaining terms are small when N is large.

Our line of proof will be a combination of the martingale difference sequence approach initiated by Hoeffding [6] and Yurinskii [21] and the statistical mechanics philosophy we already used in [4]. The martingale approach to deviation inequalities is also reviewed in [10, p. 30] and [16]. More precisely, we will decompose Z into its martingale difference sequence $Z = \sum_{i=1}^{N} F_i$ and we will take appropriate partial derivatives of the log-Laplace transform

$$(\lambda_1, \ldots, \lambda_N) \mapsto \log \mathbb{E} \exp\left(\sum_{i=1}^{N} \lambda_i F_i\right).$$

We will consider first the case of *independent* random variables $X_1, \ldots, X_N$, ranging in some product of probability spaces $\bigotimes_{i=1}^{N} (\mathcal{X}_i, \mathfrak{B}_i, \mu_i)$. In the first section, we will assume that the partial finite differences of $f(x_1,\dots,x_N)$ of order one and two are bounded. In the second section, we will assume that they have exponential moments instead, and in the third section, we will study the case of Markov chains.

1. Bounded range functionals of independent variables

Let the collection of random variables $X=(X_1,\dots,X_N)$ take its values in some product of measurable spaces $\bigotimes_{i=1}^N(\mathfrak{X}_i,\mathfrak{B}_i)$. We will assume in the following that $(X_1,\dots,X_N)$ is the canonical process. Let $\mathbb{P}=\bigotimes_{i=1}^N\mu_i$ be a product probability measure on $\bigotimes_{i=1}^N(\mathfrak{X}_i,\mathfrak{B}_i)$.

For any bounded measurable function W of $(X_1,\dots,X_N)$, we will define the modified probability distribution $\mathbb{P}_W$ by

$$d\mathbb{P}_W=\frac{\exp(W)}{\mathbb{E}[\exp(W)]}\,d\mathbb{P},$$

and we will use the notation $\mathbb{E}_W$ for the expectation operator with respect to $\mathbb{P}_W$.

On the other hand, if $\mathfrak{F}$ is some sub-sigma algebra of $\bigotimes_{i=1}^N\mathfrak{B}_i$, then $\mathbb{E}^{\mathfrak{F}}$ will be used as a short notation for the conditional expectation with respect to $\mathfrak{F}$.

As $(\mathbb{E}^{\mathfrak{F}})_W=(\mathbb{E}_W)^{\mathfrak{F}}$, we will simply write $\mathbb{E}_W^{\mathfrak{F}}$ for this conditional expectation which we can more explicitely define as

$$\mathbb{E}_W^{\mathfrak{F}}(U)=\frac{\mathbb{E}[U\exp(W)\mid\mathfrak{F}]}{\mathbb{E}[\exp(W)\mid\mathfrak{F}]}.$$

Note that we have $\mathbb{E}_W(\mathbb{E}_W^{\mathfrak{F}})=\mathbb{E}_W$, whereas $\mathbb{E}(\mathbb{E}_W^{\mathfrak{F}})\neq\mathbb{E}_W$ and $\mathbb{E}_W(\mathbb{E}^{\mathfrak{F}})\neq\mathbb{E}_W$ in general. In the same way we will use the notation $\mathbb{V}_W(U)$ for $\mathbb{E}_W\{[U-\mathbb{E}_W(U)]^2\}$ and the notation $\mathbb{M}_W^3(U)$ for $\mathbb{E}_W\{[U-\mathbb{E}_W(U)]^3\}$. Note that for any bounded measurable function W,

$$\begin{aligned}
&\frac{\partial}{\partial\lambda}\log\bigl\{\mathbb{E}\bigl[\exp(\lambda W)\bigr]\bigr\}=\mathbb{E}_W(W),\\
&\frac{\partial^2}{\partial\lambda^2}\log\bigl\{\mathbb{E}\bigl[\exp(\lambda W)\bigr]\bigr\}=\mathbb{V}_W(W),\quad\text{and}\\
&\frac{\partial^3}{\partial\lambda^3}\log\bigl\{\mathbb{E}\bigl[\exp(\lambda W)\bigr]\bigr\}=\mathbb{M}_W^3(W).
\end{aligned}$$

For each $i=1,\dots,N$, let $\mathfrak{F}_i$ be the sigma algebra generated by $(X_1,\dots,X_i)$ and let $\mathfrak{G}_i$ be the sigma algebra generated by $(X_1,\dots,X_{i-1},X_{i+1},\dots,X_N)$.

To stress the role of the independence assumption, we will put the superscript "i" on the equalities and inequalities requiring this assumption.

Let us introduce some notations linked with the martingale differences of a bounded random variable W measurable with respect to $\mathfrak{F}_N$. We will put

$$\begin{aligned}
G_i(W)&=W-\mathbb{E}^{\mathfrak{G}_i}(W),\\
F_i(W)&=\mathbb{E}^{\mathfrak{F}_i}(W)-\mathbb{E}^{\mathfrak{F}_{i-1}}(W)\\
&\overset{\mathrm{i}}{=}\mathbb{E}^{\mathfrak{F}_i}(G_i).
\end{aligned}$$

As we explained in the introduction, we will study the log-Laplace transform of

$$Z = f(X) - \mathbb{E}\big(f(X)\big)$$

of some bounded measurable real valued function $f : \prod_{i=1}^{N} \mathfrak{X}_i \to \mathbb{R}$. We will decompose Z into the sum of its martingale differences

$$Z = \sum_{i=1}^{N} F_i(Z),$$

and use the short notation $Z_i = \mathbb{E}^{\mathfrak{F}_i}(Z)$.

In this context, it is natural to assume that for some positive constants B_j, $j = 1, \dots, N$, for any $(x_1, \dots, x_N) \in \prod_{j=1}^{N} \mathfrak{X}_j$, for any $i = 1, \dots, N$, any $y_i \in \mathfrak{X}_i$,

$$f(x_1, \dots, x_N) - f(x_1, \dots, x_{i-1}, y_i, x_{i+1}, \dots, x_N) \leqslant \frac{B_i}{\sqrt{N}}.$$

The reader should understand that we are interested mainly in the case when the constants B_i are of order one. Although all these scaling factors are not really needed for finite sample bounds, we have found them useful to indicate what should be considered to be *small* and what should not.

To ensure that $f(X)$ is almost Gaussian, we need to make also an assumption on its second partial differences. This corresponds to conditions on the second partial derivatives of f in the case when the random variables $(X_1, \dots, X_N)$ take their values in some finite dimensional vector space, are bounded, and f is a smooth function.

For simplicity, we will use the short notation x_1^N for $(x_1, \dots, x_N)$. Let us put for any $x_1^N \in \prod_{j=1}^{N} \mathfrak{X}_j$, any $y_i \in \mathfrak{X}_i$,

$$\Delta_i f\big(x_1^N, y_i\big) = f\big(x_1^N\big) - f\big(x_1^{i-1}, y_i, x_{i+1}^N\big).$$

For a fixed value of y_i, $\Delta_i f$ may be seen as a function of x_1^N, and when we will write $\Delta_j \Delta_i f(x_1^N, y_i, y_j)$, we will mean that we apply Δ_j to this function and to y_j. (A more accurate but lengthy notation would have been $\Delta_j(\Delta_i f(\cdot, y_i))(x_1^N, y_j)$.)

Let us assume that for some nonnegative exponent ζ, for any $i \neq j$, for some positive constant $C_{i,j}$, and for any $x_1^N \in \prod_{k=1}^{N} \mathfrak{X}_k$, $y_i \in \mathfrak{X}_i$, $y_j \in \mathfrak{X}_j$,

$$\Delta_i \Delta_j f\big(x_1^N, y_j, y_i\big) \leqslant \frac{C_{i,j}}{N^{3/2-\zeta}}.$$

Note that $\Delta_i \Delta_j f(x_1^N, y_j, y_i) = \Delta_j \Delta_i f(x_1^N, y_i, y_j)$ and therefore that we can assume that $C_{i,j} = C_{j,i}$. We will moreover assume by convention that $C_{i,i} = 0$.

The normalisation is made so that $\zeta = 0$ corresponds to the case of

$$f(X_1, \dots, X_N) = \sqrt{N} g\left(\frac{X_1}{N}, \dots, \frac{X_N}{N}\right),$$

where $g:[0,1]^N \to \mathbb{R}$ is a smooth real function of N real bounded arguments with bounded first and second partial derivatives.

Another class of functions satisfying these hypotheses are the functions of the type

$$f(x_1^N) = \frac{1}{N^{3/2}} \sum_{1 \leqslant i < j \leqslant N} \psi_{i,j}(x_i, x_j),$$

where $\psi_{i,j}$ are bounded measurable functions. Here $\zeta = 0$, and we can take

$$B_i = \frac{2}{N}\left(\sum_{j<i} \|\psi_{j,i}\|_\infty + \sum_{j>i} \|\psi_{i,j}\|_\infty \right),$$

and

$$C_{i,j} = C_{j,i} = 4\|\psi_{i,j}\|_\infty, \quad i < j.$$

More generally

$$f(x_1^N) = N^{1/2-r} \sum_{1 \leqslant i_1 < \cdots < i_r \leqslant N} \psi_{(i_1,\dots,i_r)}(x_{i_1}, \dots, x_{i_r})$$

also satisfies our hypotheses, when the functions $\psi_{(i_1,\dots,i_r)}$ are bounded.

THEOREM 1.1. – *Under the previous hypotheses, for any positive* λ,

$$\left| \log \mathbb{E} \exp\bigl(\lambda f(X)\bigr) - \lambda \mathbb{E}\bigl(f(X)\bigr) - \frac{\lambda^2}{2} \mathbb{V}\bigl(f(X)\bigr) \right|$$
$$\leqslant \frac{\lambda^3}{N^{1/2-\zeta}} \frac{B'CB}{4N^2} + \frac{\lambda^3}{\sqrt{N}} \sum_{i=1}^N \frac{B_i^3}{3N}.$$

COROLLARY 1.1. – *Thus $f(X)$ satisfies the following deviation inequalities*:

$$\mathbb{P}\bigl(f(X) \geqslant \mathbb{E}\bigl(f(X)\bigr) + \varepsilon\bigr) \leqslant \exp\left(-\frac{\varepsilon^2}{2(\mathbb{V}(f(X)) + \frac{\eta\varepsilon}{\mathbb{V}(f(X))})} \right),$$
$$\mathbb{P}\bigl(f(X) \leqslant \mathbb{E}\bigl(f(X)\bigr) - \varepsilon\bigr) \leqslant \exp\left(-\frac{\varepsilon^2}{2(\mathbb{V}(f(X)) + \frac{\eta\varepsilon}{\mathbb{V}(f(X))})} \right),$$

with

$$\eta = \frac{1}{2N^{1/2-\zeta}} \frac{B'CB}{N^2} + \frac{2}{3\sqrt{N}} \sum_{i=1}^N \frac{B_i^3}{N}.$$

Remark 1.1. – We obtain for some constant K depending on $\max_{1 \leqslant i \leqslant N} B_i$ and $\max_{1 \leqslant i,j \leqslant N} C_{i,j}$, but not on N, that

$$\left| \log \mathbb{E} \exp\bigl(\lambda f(X)\bigr) - \lambda \mathbb{E}\bigl(f(X)\bigr) - \frac{\lambda^2}{2} \mathbb{V}\bigl(f(X)\bigr) \right| \leqslant \frac{K\lambda^3}{N^{1/2-\zeta}}.$$

Therefore if we consider a sequence of problems indexed by N such that the constants B_i and $C_{i,j}$ stay bounded, and such that $\mathbb{V}(f(X))$ converges, we get a central limit theorem as soon as $\zeta < 1/2$ (with the caveat that the limiting distribution may degenerate to a Dirac mass if the asymptotic variance is 0): $f(X) - \mathbb{E}(f(X))$ converges in distribution to a Gaussian measure.

Remark 1.2. – The critical value $\zeta_c = 1/2$ is sharp, since when

$$f(X) = g\left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i\right),$$

the central limit theorem obviously does not hold in general, and $\zeta = 1/2$.

Proof. – After decomposing Z into the sum of its martingale differences, we can view the log-Laplace transform of Z as a function of N equal "temperatures":

$$\log \mathbb{E} \exp(\lambda Z) = \log \mathbb{E}\left(\exp\left(\sum_{i=1}^{N} \lambda_i F_i(Z)\right)\right), \quad \lambda_i = \lambda, i = 1, \dots, N.$$

The first step is to take three derivatives with respect to λ_i, for i ranging from N backward to 1:

$$\log \mathbb{E} \exp(\lambda Z_i) = \log \mathbb{E} \exp(\lambda Z_{i-1}) + \int_0^{\lambda} \mathbb{E}_{\lambda Z_{i-1} + \alpha F_i(Z)}\left[F_i(Z)\right] d\alpha.$$

Therefore

$$\begin{aligned}
\log \mathbb{E} \exp(\lambda Z) &= \sum_{i=1}^{N} \int_0^{\lambda} \mathbb{E}_{\lambda Z_{i-1} + \alpha F_i(Z)}\left[F_i(Z)\right] d\alpha \\
&= \sum_{i=1}^{N} \int_0^{\lambda} (\lambda - \beta) \mathbb{V}_{\lambda Z_{i-1} + \beta F_i(Z)}\left[F_i(Z)\right] d\beta \\
&= \sum_{i=1}^{N} \frac{\lambda^2}{2} \mathbb{E}_{\lambda Z_{i-1}}\left[F_i(Z)^2\right] + \int_0^{\lambda} \frac{(\lambda - \gamma)^2}{2} \mathbb{M}^3_{\lambda Z_{i-1} + \gamma F_i(Z)}\left[F_i(Z)\right] d\gamma.
\end{aligned}$$

Thus, using the fact that for any real random variable

$$\mathbb{E}\left\{\left[\xi - \mathbb{E}(\xi)\right]^3\right\} \leqslant 2\|\xi\|_\infty \mathbb{E}\left\{\left[\xi - \mathbb{E}(\xi)\right]^2\right\} \leqslant 2\|\xi\|_\infty \mathbb{E}\left(\xi^2\right) \leqslant 2\|\xi\|_\infty^3,$$

we obtain

LEMMA 1.1. –

$$\left|\log \mathbb{E} \exp(\lambda Z) - \frac{\lambda^2}{2} \sum_{i=1}^{N} \mathbb{E}_{\lambda \mathbb{E}^{\mathfrak{F}_{i-1}}(Z)}\left[F_i(Z)^2\right]\right| \leqslant \sum_{i=1}^{N} \frac{\lambda^3 B_i^3}{3 N^{3/2}}.$$

Remark 1.3. – The upper bound can easily be improved in the following way: notice that for any nonnegative γ, $\mathbb{E}_{\lambda Z_{i-1}+\gamma F_i(Z)}[F_i(Z)] \geqslant 0$, because this expression is equal to zero when $\gamma = 0$ and is nondecreasing with respect to γ. As moreover for any real random variable $r \mapsto \mathbb{E}[(\xi - r)^3]$ is nonincreasing, we see that when $\mathbb{E}(\xi) \geqslant 0$, $\mathbb{E}\{[\xi - \mathbb{E}(\xi)]^3\} \leqslant \mathbb{E}(\xi^3) \leqslant \|\xi\|_\infty^3$. This shows that indeed

$$\log \mathbb{E} \exp(\lambda Z) - \frac{\lambda^2}{2} \sum_{i=1}^N \mathbb{E}_{\lambda \mathbb{E}^{\mathfrak{F}_{i-1}}(Z)}\bigl[F_i(Z)^2\bigr] \leqslant \sum_{i=1}^N \frac{\lambda^3 B_i^3}{6 N^{3/2}}.$$

To proceed in the proof of Theorem 1.1, we have now to approximate $\mathbb{E}_{\lambda Z_{i-1}}[F_i(Z)^2]$ by $\mathbb{E}[F_i(Z)^2]$, in order to get the variance of Z, that can be written as

$$\sum_{i=1}^N \mathbb{E}\bigl[F_i(Z)^2\bigr].$$

Let us put for short $V_i = F_i(Z)^2$ and let us introduce its martingale differences:

$$\mathbb{E}_{\lambda Z_{i-1}}\bigl[V_i - \mathbb{E}(V_i)\bigr] = \sum_{j=1}^{i-1} \mathbb{E}_{\lambda Z_{i-1}}\bigl[F_j(V_i)\bigr].$$

To deal with the jth term of this sum, we introduce the conditional expectation with respect to $\mathfrak{G}_j$:

$$\begin{aligned}
\mathbb{E}^{\mathfrak{G}_j}_{\lambda Z_{i-1}}\bigl[F_j(V_i)\bigr] &= \mathbb{E}^{\mathfrak{G}_j}_{\lambda G_j(Z_{i-1})}\bigl[F_j(V_i)\bigr] \\
&= \underbrace{\mathbb{E}^{\mathfrak{G}_j}\bigl[F_j(V_i)\bigr]}_{\overset{i}{=}0} + \int_0^\lambda \mathbb{E}^{\mathfrak{G}_j}_{\alpha G_j(Z_{i-1})}\bigl\{F_j(V_i)\bigl[G_j(Z_{i-1}) \\
&\quad - \mathbb{E}^{\mathfrak{G}_j}_{\alpha G_j(Z_{i-1})} G_j(Z_{i-1})\bigr]\bigr\}\, d\alpha. \qquad (1)
\end{aligned}$$

As a consequence, letting $U = F_j(V_i)$ and $W = (G_j(Z_{i-1}) - \mathbb{E}^{\mathfrak{G}_j}_{\alpha G_j(Z_{i-1})} G_j(Z_{i-1}))$ and applying the Cauchy–Schwartz inequality, we get

$$\bigl|\mathbb{E}^{\mathfrak{G}_j}_{\lambda Z_{i-1}}\bigl[F_j(V_i)\bigr]\bigr| \leqslant \int_0^\lambda \sqrt{\mathbb{E}^{\mathfrak{G}_j}_{\alpha G_j(Z_{i-1})}\bigl(U^2\bigr)\mathbb{E}^{\mathfrak{G}_j}_{\alpha G_j(Z_{i-1})}\bigl(W^2\bigr)}\, d\alpha.$$

Reminding that we are analysing the case when $j < i$, we can now observe that

$$G_j(Z_{i-1}) \overset{i}{=} \mathbb{E}^{\mathfrak{F}_{i-1}} G_j(Z),$$

and therefore that its conditional range is upper bounded by

$$\operatorname{ess\,sup}\bigl(G_j(Z_{i-1}) \mid \mathfrak{G}_j\bigr) - \operatorname{ess\,inf}\bigl(G_j(Z_{i-1}) \mid \mathfrak{G}_j\bigr) \leqslant \frac{B_j}{\sqrt{N}}.$$

This implies that its variance is bounded by $\frac{B_j^2}{4N}$.

(Let us remind that by definition, for any real random variable W on a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$, and any sub sigma algebra $\mathfrak{G} \subset \mathfrak{F}$,

$$\operatorname{ess\,sup}(W \mid \mathfrak{G}) \stackrel{\text{def}}{=} \sup\{\lambda \in \mathbb{R}\colon \mathbb{P}(W \geqslant \lambda \mid \mathfrak{G}) > 0\}.$$

It can equivalently be defined by the relation

$$\{\operatorname{ess\,sup}(W \mid \mathfrak{G}) > \lambda\} = \{\mathbb{P}(W \geqslant \lambda \mid \mathfrak{G}) > 0\}, \quad \lambda \in \mathbb{Q},$$

which shows that it belongs to $\mathbb{L}^0(\Omega, \mathfrak{G}, \mathbb{P})$ – the set of measurable functions factorised by almost sure equality with respect to $\mathbb{P}$. Here and below, inequalities involving conditional expectations and conditional essential suprema and infima are meant without further notice to hold almost surely.)

Let us consider now on some enlarged probability space two independent random variables X'_i and X'_j, such that $(X_1, \ldots, X_N, X'_j, X'_i)$ is distributed according to $\bigotimes_{k=1}^N \mu_k \otimes \mu_j \otimes \mu_i$. We have

$$F_j\left[F_i(Z)^2\right] = F_j\left\{\left[\mathbb{E}^{\mathfrak{F}_i}\left[\Delta_i f(X_1^N, X'_i)\right]\right]^2\right\}.$$

Moreover for any function $h(X_1^N)$,

$$\begin{aligned}\left|F_j(h(X)^2)\right| &= \left|\mathbb{E}^{\mathfrak{F}_j}\Delta_j h^2(X_1^N, X'_j)\right| \\ &= \left|\mathbb{E}^{\mathfrak{F}_j}\left\{\left[h(X_1^N) + h(X_1^{j-1}, X'_j, X_{j+1}^N)\right]\Delta_j h(X_1^N, X'_j)\right\}\right| \\ &\leqslant 2\operatorname{ess\,sup}\left|h(X)\right|\, \mathbb{E}^{\mathfrak{F}_j}\left|\Delta_j h(X_1^N, X'_j)\right|.\end{aligned}$$

Applying this to $h(X) = \mathbb{E}^{\mathfrak{F}_i}(\Delta_i f(X_1^N, X'_i))$, we get

$$\left|F_j\left(F_i(Z)^2\right)\right| \leqslant 2\frac{B_i}{\sqrt{N}}\mathbb{E}^{\mathfrak{F}_j}\left|\mathbb{E}^{\mathfrak{F}_i}\Delta_j\Delta_i f(X_1^N, X'_i, X'_j)\right| \leqslant \frac{2B_i C_{i,j}}{N^{2-\zeta}}.$$

Therefore

$$\left|\mathbb{E}_{\lambda Z_{i-1}} F_j(V_i)\right| \leqslant \lambda\frac{B_i C_{i,j} B_j}{N^{5/2-\zeta}}.$$

This, combined with Lemma 1.1, ends the proof of Theorem 1.1. The derivation of its corollary is standard: it is obtained by taking

$$\lambda = \frac{\varepsilon}{\mathbb{V}(f) + \eta\varepsilon/\mathbb{V}(f)},$$

and by applying successively the theorem to f and $-f$. $\square$

2. Extension to unbounded ranges

The boundedness assumption of the finite differences of f can be relaxed to exponential moment assumptions. To achieve this, we will suitably modify the bounds of the previous section, still assuming that f is bounded, and will afterwards let more general functions be approximated by bounded ones.

THEOREM 2.1. – *Let λ be a positive real constant. Let $\bigotimes_{i=1}^N(\mathfrak{X}_i, \mathfrak{B}_i, \mu_i)$ be some product of probability spaces. Let $f \in L^2\ (\bigotimes_{i=1}^N(\mathfrak{X}_i, \mathfrak{B}_i, \mu_i))$ be such that*

$$\mathbb{E}\{\exp[\lambda f(X)]\} < +\infty,$$

where $X = (X_1, \dots, X_N)$ is the canonical process. Let $(X_1, \dots, X_N, X'_1, \dots, X'_N)$ be the canonical process on $(\bigotimes_{i=1}^N(\mathfrak{X}_i, \mathfrak{B}_i))^{\otimes 2}$ and let $\mathbb{E}$ be the expectation with respect to the probability measure $(\bigotimes_{i=1}^N \mu_i)^{\otimes 2}$ defined on this enlarged space. Let us introduce the short notations

$$\begin{aligned}
&\Delta_i = \Delta_i f(X_1^N, X'_i), \quad 1 \leqslant i \leqslant N,\\
&\Delta_{j,i} = \Delta_j \Delta_i f(X_1^N, X'_i, X'_j), \quad 1 \leqslant j < i \leqslant N,\\
&\Phi(r) = \exp(r) - 1 - r - \frac{r^2}{2} = \int_0^r \frac{(r-u)^2}{2} \exp(u)\, du,
\end{aligned}$$

and let us decide by convention that $0 \times (+\infty) = +\infty$. Then

$$\begin{aligned}
&\left|\log \mathbb{E} \exp[\lambda f(X)] - \lambda \mathbb{E}[f(X)] - \frac{\lambda^2}{2} \mathbb{V}[f(X)]\right|\\
&\leqslant 5 \sum_{i=1}^N \operatorname{ess\,sup} \mathbb{E}^{\mathfrak{G}_i}\left[\Phi\left(\lambda|\Delta_i|\right)\right]\\
&\quad + \sqrt{2}\lambda^2 \sum_{i=1}^N \sum_{j=1}^{i-1} \operatorname{ess\,sup}\left[\mathbb{E}^{\mathfrak{G}_i}(\Delta_i^2)\right]^{1/2} \operatorname{ess\,sup}\left[\mathbb{E}^{\mathfrak{F}_{j-1}}(\Delta_{j,i}^2)\right]^{1/2}\\
&\quad \times \operatorname{ess\,sup}\left\{\mathbb{E}^{\mathfrak{G}_j}\left[\lambda|\Delta_j|\left(\exp(2\lambda|\Delta_j|) - 1\right)\right]\right\}^{1/2}.
\end{aligned}$$

Let us begin with a lemma.

LEMMA 2.1. – *In the case when $f \in \mathbb{L}^\infty$,*

$$\left|\log \mathbb{E} \exp(\lambda Z) - \frac{\lambda^2}{2} \sum_{i=1}^N \mathbb{E}_{\lambda Z_{i-1}}\left[F_i(Z)^2\right]\right| \leqslant 5 \sum_{i=1}^N \operatorname{ess\,sup} \mathbb{E}^{\mathfrak{F}_{i-1}} \Phi\left(\lambda\left|F_i(Z)\right|\right).$$

Proof. – We can notice that for any bounded real random variable ξ

$$\begin{aligned}
\left|\mathbb{E}\{[\xi - \mathbb{E}(\xi)]^3\}\right| &\leqslant \mathbb{E}\left[|\xi|(\xi - \mathbb{E}(\xi))^2\right] + \mathbb{E}(|\xi|)\mathbb{E}\left[(\xi - \mathbb{E}(\xi))^2\right]\\
&\leqslant \mathbb{E}(|\xi|^3) + 2\mathbb{E}(\xi^2)\mathbb{E}(|\xi|) + \mathbb{E}(|\xi|)^3 + \mathbb{E}(|\xi|)\mathbb{E}(\xi^2)\\
&\leqslant 5\mathbb{E}(|\xi|^3).
\end{aligned}$$

(Indeed $\mathbb{E}(\xi^2)\mathbb{E}(|\xi|) \leqslant \mathbb{E}(|\xi|^3)$, from the convexity of $g : \beta \mapsto \log[\mathbb{E}(|\xi|^\beta)]$, which implies that $g(1) - g(0) \leqslant g(3) - g(2)$.)

Coming back to the proof of Lemma 1.1 we can write the following chain of inequalities:

$$\left|\int_o^\lambda \frac{(\lambda - \gamma)^2}{2} \mathbb{M}^3_{\lambda Z_{i-1+\gamma} F_i(z)}\left[F_i(Z)\right] d\gamma\right|$$

$$\leqslant \frac{5}{2}\int_0^\lambda (\lambda-\gamma)^2 \mathbb{E}_{\lambda Z_{i-1}+\gamma F_i(Z)}\big(|F_i(Z)|^3\big)\,d\gamma$$

$$\leqslant \mathbb{E}_{\lambda Z_{i-1}}\left\{\frac{5}{2}\int_0^\lambda (\lambda-\gamma)^2 \mathbb{E}^{\mathfrak{F}_{i-1}}\big[\exp\big(\gamma|F_i(Z)|\big)|F_i(Z)|^3\big]\,d\gamma\right\}$$

$$\leqslant \frac{5}{2}\operatorname{ess\,sup}\mathbb{E}^{\mathfrak{F}_{i-1}}\left(\int_0^\lambda (\lambda-\gamma)^2|F_i(Z)|^3\exp\big(\gamma|F_i(Z)|\big)\,d\gamma\right).$$

(As in the case of Lemma 1.1 – see Remark 1.3 – the upper bound can be improved, noticing that for any real random variable $\mathbb{E}\{[\xi-\mathbb{E}(\xi)]^3\}\leqslant \mathbb{E}(\xi^3)$ as soon as $\mathbb{E}(\xi)\geqslant 0$.) □

Proof of Theorem 2.1. – Let us prove the theorem first when $f\in\mathbb{L}^\infty$. In addition to those introduced in Theorem 2.1, we will need the following short notation:

$$\overset{j}{\Delta}_i=\Delta_i f\big((X_1,\dots,X_{j-1},X'_j,X_{j+1}^N),X'_i\big),\quad 1\leqslant j<i\leqslant N.$$

Let us come back to Eq. (1) and notice that

$$\begin{aligned}F_j(V_i)&=\mathbb{E}^{\mathfrak{F}_j}\big[\mathbb{E}^{\mathfrak{F}_i}\big(\Delta_i+\overset{j}{\Delta}_i\big)\mathbb{E}^{\mathfrak{F}_i}(\Delta_{j,i})\big],\\ G_j(Z_{i-1})&=\mathbb{E}^{\mathfrak{F}_{i-1}}(\Delta_i),\end{aligned}$$

and therefore, from the Cauchy–Schwartz inequality and the convexity of $r\mapsto r^2$, that

$$\big|F_j(V_i)\big|\leqslant\big\{\mathbb{E}^{\mathfrak{F}_j}\big[\big(\Delta_i+\overset{j}{\Delta}_i\big)^2\big]\big\}^{1/2}\big[\mathbb{E}^{\mathfrak{F}_j}\big(\Delta_{j,i}^2\big)\big]^{1/2}.$$

We can thus bound the integrand of $\mathbb{E}^{\mathfrak{G}_j}_{G_j(Z_{i-1})}$ in the last term of Eq. (1) by

$$\big|F_j(V_i)\big[G_j(Z_{i-1})-\mathbb{E}^{\mathfrak{G}_j}_{G_j(Z_{i-1})}G_j(Z_{i-1})\big]\big|\leqslant A\times B,$$

where

$$\begin{aligned}A&=\big[\mathbb{E}^{\mathfrak{F}_j}\big(\Delta_{j,i}^2\big)\big]^{1/2}\exp\left[-\frac{\alpha}{2}G_j(Z_{i-1})\right],\\ B&=\exp\left[\frac{\alpha}{2}G_j(Z_{i-1})\right]\big\{\mathbb{E}^{\mathfrak{F}_j}\big[\big(\Delta_i+\overset{j}{\Delta}_i\big)^2\big]\big\}^{1/2}\big\{\big|G_j(Z_{i-1})\big|+\big|\mathbb{E}^{\mathfrak{G}_j}_{G_j(Z_{i-1})}G_j(Z_{i-1})\big|\big\}.\end{aligned}$$

Applying the Cauchy–Schwartz inequality and noticing that for any real random variable ξ such that $\mathbb{E}(\xi)\geqslant 0$,

$$\mathbb{E}\big[\exp(\alpha\xi)\big]\geqslant\exp\big[\alpha\mathbb{E}(\xi)\big]\geqslant 1,$$

we get

$$\big|\mathbb{E}_{\lambda Z_{i-1}}\big[F_j(V_i)\big]\big|\leqslant\mathbb{E}_{\lambda Z_{i-1}}\left\{\int_0^\lambda\big[\mathbb{E}^{\mathfrak{G}_j}_{\alpha G_j(Z_{i-1})}\big(A^2\big)\big]^{1/2}\big[\mathbb{E}^{\mathfrak{G}_j}_{\alpha G_j(Z_{i-1})}\big(B^2\big)\big]^{1/2}d\alpha\right\}$$

$$
\begin{aligned}
&\leqslant \mathbb{E}_{\lambda Z_{i-1}}\left\{\int_0^\lambda \left\{\mathbb{E}^{\mathfrak{G}_j}\left[\exp\left[\alpha G_j(Z_{i-1})\right] A_0^2\right]\right\}^{1/2}\left[\mathbb{E}^{\mathfrak{G}_j}_{\alpha G_j(Z_{i-1})}\left(B^2\right)\right]^{1/2} d\alpha\right\}\\
&\leqslant 2\mathbb{E}_{\lambda Z_{i-1}}\Bigg\{\left[\mathbb{E}^{\mathfrak{F}_{j-1}}\left(\Delta_{j,i}^2\right)\right]^{1/2}\\
&\quad\times\int_0^\lambda \left\{\mathbb{E}^{\mathfrak{G}_j}_{\alpha G_j(Z_{i-1})}\left[\exp\left(\alpha G_j(Z_{i-1})\right)\left[G_j(Z_{i-1})^2\right.\right.\right.\\
&\quad\left.\left.\left.+\left[\mathbb{E}^{\mathfrak{G}_j}_{\alpha G_j(Z_{i-1})} G_j(Z_{i-1})\right]^2\right]\right]\right\}^{1/2} d\alpha\Bigg\}\\
&\quad\times\left[\operatorname{ess\,sup}\mathbb{E}^{\mathfrak{F}_j}\left(\Delta_i^2\right)+\operatorname{ess\,sup}\mathbb{E}^{\mathfrak{F}_j}\left(\overset{j}{\Delta}{}_i^2\right)\right]^{1/2},
\end{aligned}
$$

where we have also used the fact that $(a+b)^2 \leqslant 2(a^2+b^2)$. To simplify the bound, we can now notice that for any random variable ξ such that $\mathbb{E}(\xi) \geqslant 0$, $[\mathbb{E}(\xi)]^2\, \mathbb{E}[\exp(\alpha\xi)] \leqslant \mathbb{E}[\xi^2 \exp(\alpha\xi)]$ (because $[\mathbb{E}(\xi)]^2 \leqslant [\mathbb{E}_{\alpha\xi}(\xi)]^2 \leqslant \mathbb{E}_{\alpha\xi}(\xi^2)$ which is the inequality we seek). We can also notice that

$$
\operatorname{ess\,sup}\mathbb{E}^{\mathfrak{F}_j}\left(\Delta_i^2\right) = \operatorname{ess\,sup}\mathbb{E}^{\mathfrak{F}_j}\left(\overset{j}{\Delta}{}_i^2\right) \leqslant \operatorname{ess\,sup}\mathbb{E}^{\mathfrak{G}_i}\left(\Delta_i^2\right).
$$

These two remarks lead to

$$
\begin{aligned}
&\left|\mathbb{E}_{\lambda Z_{i-1}}\left[F_j(V_i)\right]\right|\\
&\quad\leqslant 4\mathbb{E}_{\lambda Z_{i-1}}\Bigg\{\left[\mathbb{E}^{\mathfrak{F}_{j-1}}\left(\Delta_{j,i}^2\right)\right]^{1/2}\\
&\qquad\times\int_0^\lambda \left\{\mathbb{E}^{\mathfrak{G}_j}\left[\exp\left(2\alpha G_j(Z_{i-1})\right) G_j(Z_{i-1})^2\right]\right\}^{1/2} d\alpha\Bigg\}\left[\operatorname{ess\,sup}\mathbb{E}^{\mathfrak{G}_i}\left(\Delta_i^2\right)\right]^{1/2}.
\end{aligned}
$$

Remembering that $G_j(Z_{i-1}) = \mathbb{E}^{\mathfrak{F}_{i-1}}(\Delta_j)$, using the convexity of $r \mapsto r^2\exp(2\alpha r)$ on the positive real axis and another round of the Cauchy–Schwartz inequality, we get

$$
\begin{aligned}
&\int_0^\lambda \left\{\mathbb{E}^{\mathfrak{G}_j}\left[\exp\left(2\alpha G_j(Z_{i-1})\right) G_j(Z_{i-1})^2\right]\right\}^{1/2} d\alpha\\
&\quad\leqslant \sqrt{\frac{\lambda}{2}}\left\{\mathbb{E}^{\mathfrak{G}_j}\mathbb{E}^{\mathfrak{F}_{i-1}}\left[\left(\exp(2\lambda|\Delta_j|)-1\right)|\Delta_j|\right]\right\}^{1/2}\\
&\quad\leqslant \sqrt{\frac{\lambda}{2}}\left\{\operatorname{ess\,sup}\mathbb{E}^{\mathfrak{G}_j}\left[\left(\exp(2\lambda|\Delta_j|)-1\right)|\Delta_j|\right]\right\}^{1/2}.
\end{aligned}
$$

Thus

$$
\begin{aligned}
&\left|\mathbb{E}_{\lambda Z_{i-1}}\left[F_j(V_i)\right]\right|\\
&\quad\leqslant 2\sqrt{2}\mathbb{E}_{\lambda Z_{i-1}}\left\{\left[\mathbb{E}^{\mathfrak{F}_{j-1}}\left(\Delta_{j,i}^2\right)\right]^{1/2}\right\}\\
&\qquad\times\left\{\operatorname{ess\,sup}\mathbb{E}^{\mathfrak{G}_j}\left[\lambda|\Delta_j|\left(\exp(2\lambda|\Delta_j|)-1\right)\right]\right\}^{1/2}\left[\operatorname{ess\,sup}\mathbb{E}^{\mathfrak{G}_i}\left(\Delta_i^2\right)\right]^{1/2}.
\end{aligned}
$$

This combined with Lemma 2.1 proves the following slight improvement of Theorem 2.1 in the case when $f \in \mathbb{L}^\infty$:

$$
\begin{aligned}
&\left|\log \mathbb{E} \exp\bigl[\lambda f(X)\bigr] - \lambda \mathbb{E}\bigl[f(X)\bigr] - \frac{\lambda^2}{2}\mathbb{V}\bigl[f(X)\bigr]\right| \\
&\quad \leqslant 5 \sum_{i=1}^N \operatorname{ess\,sup} \mathbb{E}^{\mathfrak{F}_{i-1}}\bigl[\Phi\bigl(\lambda|\Delta_i|\bigr)\bigr] \\
&\qquad + \sqrt{2}\lambda^2 \sum_{i=1}^N \sum_{j=1}^{i-1} \operatorname{ess\,sup} \bigl[\mathbb{E}^{\mathfrak{G}_i}\bigl(\Delta_i^2\bigr)\bigr]^{1/2} \mathbb{E}_{\lambda Z_{i-1}}\bigl\{\bigl[\mathbb{E}^{\mathfrak{F}_{j-1}}\bigl(\Delta_{j,i}^2\bigr)\bigr]^{1/2}\bigr\} \\
&\qquad \times \operatorname{ess\,sup}\bigl\{\mathbb{E}^{\mathfrak{G}_j}\bigl[\lambda|\Delta_j|\bigl(\exp\bigl(2\lambda|\Delta_j|\bigr) - 1\bigr)\bigr]\bigr\}^{1/2}.
\end{aligned} \tag{2}
$$

When f satisfies only the weaker assumptions of Theorem 2.1, we can introduce the bounded functions $f^T(X) = \min\{\max\{f(X), -T\}, T\}$, $T \in \mathbb{N}$.

The terms of the left-hand side of Eq. (2) for f^T converge from the dominated convergence theorem to their counterpart for f. Indeed

$$
\begin{aligned}
&\exp\bigl[\lambda f^T(X)\bigr] \leqslant \max\bigl\{1, \exp\bigl[\lambda f(X)\bigr]\bigr\} \in \mathbb{L}^1, \\
&\bigl|f^T(X)\bigr| \leqslant |f(X)| \in \mathbb{L}^2.
\end{aligned}
$$

In the right-hand side, we can use the bounds

$$
\bigl|\Delta_i f^T(X, X_i')\bigr| \leqslant \bigl|\Delta_i f(X, X_i')\bigr|,
$$

and apply the dominated convergence theorem to prove that, introducing the notations

$$
\begin{aligned}
\Delta_{j,i}(f) &= \Delta_j \Delta_i f(X, X_i', X_j'), \\
Z_i(f) &= \mathbb{E}^{\mathfrak{F}_i}[f(X)] - \mathbb{E}[f(X)],
\end{aligned}
$$

$$
\lim_{T \to +\infty} \mathbb{E}_{\lambda Z_{i-1}(f^T)}\bigl\{\bigl[\mathbb{E}^{\mathfrak{F}_{j-1}}\bigl(\bigl[\Delta_{j,i}(f^T)\bigr]^2\bigr)\bigr]^{1/2}\bigr\} = \mathbb{E}_{\lambda Z_{i-1}(f)}\bigl\{\bigl[\mathbb{E}^{\mathfrak{F}_{j-1}}\bigl([\Delta_{j,i}(f)]^2\bigr)\bigr]^{1/2}\bigr\},
$$

as soon as $\operatorname{ess\,sup} \mathbb{E}^{\mathfrak{G}_i}(\Delta_i^2) < +\infty$ (in the other case, Theorem 2.1 is trivial). Indeed $\mathbb{E}_{\lambda Z_{i-1}(f)} = \mathbb{E}_{\lambda \mathbb{E}^{\mathfrak{F}_{i-1}}[f(X)]}$,

$$
\exp\bigl[\lambda \mathbb{E}^{\mathfrak{F}_{i-1}}\bigl(f^T(X)\bigr)\bigr] \leqslant \max\bigl\{1, \mathbb{E}^{\mathfrak{F}_{i-1}}\bigl[\exp\bigl(\lambda f(X)\bigr)\bigr]\bigr\} \in \mathbb{L}^1
$$

and, putting $\overset{j}{\Delta}_i(f) = \Delta_i f(\overset{j}{X'}, X_i')$, (so that $\Delta_{j,i}(f) = \Delta_i(f) - \overset{j}{\Delta}_i(f)$),

$$
\mathbb{E}^{\mathfrak{F}_{j-1}}\bigl([\Delta_{j,i}(f^T)]^2\bigr) \leqslant 2\mathbb{E}^{\mathfrak{F}_{j-1}}\bigl[\Delta_i(f^T)^2 + \overset{j}{\Delta}_i(f^T)^2\bigr] \leqslant 4 \operatorname{ess\,sup} \mathbb{E}^{\mathfrak{G}_i}\bigl[\Delta_i(f)^2\bigr].
$$

This proves that inequality (2), and consequently Theorem 2.1 which is weaker, holds for f. $\square$

3. Generalisation to Markov chains

We will study here the case when $(X_1, \dots, X_N)$ is a Markov chain. The assumptions on f will be the same as in the first section. Therefore we will assume throughout this section that

$$\Delta_i f\left(x_1^N, y_i\right) \leqslant \frac{B_i}{\sqrt{N}}, \quad i = 1, \dots, N,$$

and that

$$\Delta_j \Delta_i f\left(x_1^N, y_i, y_j\right) \leqslant \frac{C_{i,j}}{N^{3/2-\zeta}}, \quad 1 \leqslant j < i \leqslant N.$$

When the random variables $(X_1, \dots, X_N)$ are dependent, we have to modify the definition of the operators $G_i(Z)$ and of the sigma algebras $\mathfrak{G}_i$. Indeed to generalise the first part of the proof we would like to have the identity

$$\mathbb{E}^{\mathfrak{F}_i}(G_i(Z)) = F_i(Z),$$

where G_i is "as small as possible". This identity does not hold in the general case with the definition we had for G_i, so we will have to change it.

To generalise the second part of the proof, we need to consider a new definition of the sigma algebra $\mathfrak{G}_j$ for which

$$\mathbb{E}^{\mathfrak{G}_j}\left(F_j(W)\right) = 0,$$

and $Z - \mathbb{E}^{\mathfrak{G}_j}(Z)$ is as small as possible.

We propose a solution here where the two objects G_i and $\mathfrak{G}_j$ are built with the help of coupled processes. This use of coupling was inspired by the works of Katalin Marton [11–13].

Let us remind first some general facts about maximal coupling: given some probability space $(\Omega, \mathfrak{S})$, and two probability measures μ and ν on $(\Omega, \mathfrak{S})$, a maximal coupling between μ and ν is a probability measure ρ on $(\Omega \times \Omega, \mathfrak{S} \otimes \mathfrak{S})$ such that $\rho(d\omega_1) = \mu(d\omega_1)$, $\rho(d\omega_2) = \nu(d\omega_2)$ and $\rho(\omega_1 = \omega_2)$ is maximal (where ω_1 and ω_2 are the first and second coordinates on $\Omega \times \Omega$). In words, a maximal coupling is a joint distribution with prescribed marginals and a maximal weight given to the diagonal. The maximal weight of the diagonal is obviously $(\mu \wedge \nu)(\Omega)$, where $\mu \wedge \nu$ is defined as $d(\mu \wedge \nu) = \min\{\frac{d\mu}{d(\mu+\nu)}, \frac{d\nu}{d(\mu+\nu)}\} d(\mu + \nu)$ and is necessarily the trace of any maximal coupling distribution on the diagonal; it is also related to the variation distance $\|\mu - \nu\|_{\text{var}} \stackrel{\text{def}}{=} \frac{1}{2}|\mu - \nu|(\Omega)$ between μ and ν by the formula $(\mu \wedge \nu)(\Omega) = 1 - \|\mu - \nu\|_{\text{var}}$. A maximal coupling between μ and ν is not unique (its off diagonal behaviour being arbitrary, as long as the marginal constraints are satisfied), however one possible explicit construction is the following: consider on the product space $(\Omega \times \Omega \times [0,1], \mathfrak{S}^{\otimes 2} \otimes \mathfrak{B})$, where $\mathfrak{B}$ is the Borel sigma algebra, the product probability measure $\mu \otimes \|\mu - \nu\|_{\text{var}}^{-1}(\nu - \mu)_+ \otimes U$, where $(\nu - \mu)_+ = \nu - (\nu \wedge \mu)$ is the positive part of the signed measure $\nu - \mu$ and where U is the Lebesgue measure on $[0,1]$. Let (ω_1, ω_2, r) be the three coordinates of

$(\Omega \times \Omega \times [0,1])$ and define the random variables $W \stackrel{\text{def}}{=} \omega_1$, $T = F(W, (\omega_2, r))$, where

$$F(\omega_1, (\omega_2, r)) = \begin{cases} \omega_1 & \text{if } r < \frac{d(\mu \wedge \nu)}{d\mu}(\omega_1), \\ \omega_2 & \text{otherwise.} \end{cases} \quad (\omega_1, \omega_2) \in \Omega^2,\ r \in [0,1], \tag{3}$$

Then it is elementary to check that the distribution of (W, T) is a maximal coupling between μ and ν, equal indeed to

$$(\mu \wedge \nu)(dW)\delta_W(dT) + \|\mu - \nu\|_{\text{var}}^{-1}(\mu - \nu)_+(dW) \otimes (\nu - \mu)_+(dT). \tag{4}$$

We will define auxiliary random variables that will be coupled with the process $(X_1, \ldots, X_N)$ in a suitable way. For this, we will enlarge the probability space: instead of working on the canonical space $(\prod_{i=1}^N \mathfrak{X}_i \bigotimes_{i=1}^N \mathfrak{B}_i)$, we will work on some enlarged probability space $(\Omega, \mathfrak{B})$, where we will jointly define the process $(X_1, \ldots, X_N)$, and N other processes $\{\overset{i}{Y}{}_{j=1}^N;\ i = 1, \ldots, N\}$ that will be useful for the construction of the operators $\{G_i;\ i = 1, \ldots, N\}$. In the following the symbol $\mathbb{E}$ will stand for the expectation on Ω.

The basic construction of coupled processes we will need is the following: we consider, on some augmented probability space Ω, $N+1$ stochastic processes $(X_1, \ldots, X_N)$ and $\{(\overset{i}{Y}_1, \ldots, \overset{i}{Y}_N);\ i = 1, \ldots, N\}$ satisfying the following properties:

- The distribution of each $\overset{i}{Y}$ is equal to the distribution of X.
- Almost surely $\overset{i}{Y}{}_1^{i-1} = X_1^{i-1}$.
- Given X, the N processes $\{\overset{i}{Y};\ i = 1, \ldots, N\}$ are independent.
- Given X_1^{i-1}, $\overset{i}{Y}{}_i^N$ is independent of X_i (but not of X_{i+1}^N, the interesting thing will be on the contrary to have a maximal coupling between $\overset{i}{Y}{}_{i+1}^N$ and X_{i+1}^N).

The general method to build such processes is the following:

- *Choice of* Ω: Take for Ω the canonical space of $(X_1^N, (\overset{i}{Y}{}_{j=1}^N)_{i=1}^N)$, that is $(\bigotimes_{i=1}^N (\mathfrak{X}_i, \mathfrak{B}_i))^{\otimes(N+1)}$. For any random variable W defined on Ω, we will use the notation $\mathbb{P}(dW)$ to denote the distribution of W. We will assume without further notice that all the conditional distributions we need exist and have regular versions. This will always be the case when we deal with Polish spaces $(\mathfrak{X}_i, \mathfrak{B}_i)$ (see [17, p. 146]).
- *Construction of the distribution of the pair* $(X, \overset{i}{Y})$: We are going to construct the joint distribution $\mathbb{P}(dX, d\overset{i}{Y})$ of X and $\overset{i}{Y}$ by defining each term of the decomposition into conditional probabilities

$$\mathbb{P}\big(dX, d\overset{i}{Y}\big) = \mathbb{P}\big(dX_1^{i-1}\big)\mathbb{P}\big(d\overset{i}{Y}{}_1^{i-1} \mid X_1^{i-1}\big) \prod_{j=i}^N \mathbb{P}\big(dX_j, d\overset{i}{Y}_j \mid X_1^{j-1}, \overset{i}{Y}{}_1^{j-1}\big).$$

(All the conditional distributions present in this definition are assumed to exist and have regular versions, which is always the case when the base spaces $(\mathfrak{X}_i, \mathfrak{B}_i)$ are

Polish).

The distribution $\mathbb{P}(dX_1^{i-1})$ is given. We define $\mathbb{P}(d\overset{i}{Y}{}_{j=1}^{i-1} \mid X_1^{i-1})$ by letting $\overset{i}{Y}{}_1^{i-1} = X_1^{i-1}$, a.s.

For $j = i$ we build the conditional distribution

$$\mathbb{P}\big(dX_j, d\overset{i}{Y}_j \mid X_1^{j-1}, \overset{i}{Y}{}_1^{j-1}\big)$$

by putting

$$\mathbb{P}\big(dX_i, d\overset{i}{Y}_i \mid X_1^{i-1}, \overset{i}{Y}{}_1^{i-1}\big) = \mathbb{P}\big(dX_i \mid X_1^{i-1}\big) \otimes \mathbb{P}\big(d\overset{i}{Y}_i \mid \overset{i}{Y}{}_1^{i-1}\big),$$

where $\mathbb{P}(d\overset{i}{Y}_i \mid \overset{i}{Y}{}_1^{i-1})$ is defined by the requirement that the marginal $\mathbb{P}(d\overset{i}{Y})$ be the same as $\mathbb{P}(dX)$. For each $j > i$, we choose for

$$\mathbb{P}\big(dX_j, d\overset{i}{Y}_j \mid X_1^{j-1}, \overset{i}{Y}{}_1^{j-1}\big)$$

some maximally coupled distribution with marginals

$$\begin{cases} \mathbb{P}\big(dX_j \mid X_1^{j-1}, \overset{i}{Y}{}_1^{j-1}\big) = \mathbb{P}\big(dX_j \mid X_1^{j-1}\big), \\ \mathbb{P}\big(d\overset{i}{Y}_j \mid X_1^{j-1}, \overset{i}{Y}{}_1^{j-1}\big) = \mathbb{P}\big(d\overset{i}{Y}_j \mid \overset{i}{Y}{}_1^{j-1}\big), \end{cases}$$

where the second marginal is defined by the requirement that $\mathbb{P}(d\overset{i}{Y})$ be the same as $\mathbb{P}(dX)$. (An explicit construction of a maximal coupling distribution is given by Eq. (4)).

- *Last step of the construction*: Once we have built the distribution of each couple of processes $\mathbb{P}(dX, dY)$, separately for each i, we build the joint distribution of $(X, \overset{i}{Y},\ i = 1, \ldots, N)$ on its canonical space. For the time being, we will not really use this joint distribution, but it is simpler to deal with one probability space Ω than with N probability spaces Ω_i, so let us say that, $\mathbb{P}(dX)$ and $\mathbb{P}(d\overset{i}{Y} \mid X)$ being defined as previously explained (on separate probability spaces), we let

$$\mathbb{P}\big(dX, d\overset{1}{Y}, \ldots, d\overset{N}{Y}\big) = \mathbb{P}(dX) \bigotimes_{i=1}^{N} \mathbb{P}\big(d\overset{i}{Y} \mid X\big),$$

on the joint probability space $(\bigotimes_{i=1}^{N} (\mathfrak{X}, \mathfrak{B}_i))^{\otimes (N+1)}$.

It is immediate to see from this construction that

$$\begin{aligned} \mathbb{P}\big(dX_i, d\overset{i}{Y}{}_i^N \mid X_1^{i-1}\big) &= \mathbb{P}\big(dX_i \mid X_1^{i-1}\big)\mathbb{P}\big(d\overset{i}{Y}_i \mid X_1^{i-1}\big)\mathbb{P}\big(d\overset{i}{Y}{}_{i+1}^N \mid X_1^i, \overset{i}{Y}_i\big) \\ &= \mathbb{P}\big(dX_i \mid X_1^{i-1}\big)\mathbb{P}\big(d\overset{i}{Y}_i \mid X_1^{i-1}\big)\mathbb{P}\big(d\overset{i}{Y}{}_{i+1}^N \mid \overset{i}{Y}{}_1^i\big) \\ &= \mathbb{P}\big(dX_i \mid X_1^{i-1}\big) \otimes \mathbb{P}\big(d\overset{i}{Y}{}_i^N \mid \overset{i}{Y}{}_1^{i-1}\big). \end{aligned} \tag{5}$$

This proves that conditionally on $(X_1, \dots, X_{i-1})$, the random variable X_i is independent of the sigma algebra generated by $(\overset{i}{Y}_i, \dots, \overset{i}{Y}_N)$.

Remark 3.1. – We have also exactly in the same way

$$\mathbb{P}\big(dX_i^N, d\overset{i}{Y}_i \mid X_1^{i-1}\big) = \mathbb{P}\big(dX_i^N \mid X_1^{i-1}\big) \otimes \mathbb{P}\big(d\overset{i}{Y}_i \mid \overset{i}{Y}{}_1^{i-1}\big). \tag{6}$$

Remark 3.2. – Instead of building distributions on the canonical space of $(X, \overset{i}{Y})$, we could also have built $\overset{i}{Y}$ as a function of X and auxiliary random variables, in the spirit of formula (3). We could thus have realized $\overset{i}{Y}_{i+1}, \dots, \overset{i}{Y}_N$ as $\overset{i}{Y}_k = F_k(X_k, \omega_k)$, where each random variable ω_k is independent of X_k, conditionally on $(X_1^{k-1}, \overset{i}{Y}{}_1^{k-1})$. However, this construction would not help much in proving the conditional independence statements of Eqs. (5) and (6).

As in the previous sections, $\mathfrak{F}_i$ will be the sigma algebra generated by $(X_1, \dots, X_i)$, and we will put

$$Z(X) = f(X) - \mathbb{E}\big(f(X)\big).$$

For any bounded measurable function $h(X)$ we will define

$$\begin{aligned}
&G_i\big(h(X)\big) = h(X) - \mathbb{E}^{\mathfrak{F}_N}\big(h(\overset{i}{Y})\big),\\
&F_i\big(h(X)\big) = \mathbb{E}^{\mathfrak{F}_i}\big(h(X)\big) - \mathbb{E}^{\mathfrak{F}_{i-1}}\big(h(X)\big) = \mathbb{E}^{\mathfrak{F}_i}\big(G_i\big(h(X)\big)\big).
\end{aligned}$$

The last line holds because

$$\begin{aligned}
\mathbb{E}^{\mathfrak{F}_i}\mathbb{E}^{\mathfrak{F}_N}\big(h(\overset{i}{Y})\big) &= \mathbb{E}\big(h(\overset{i}{Y}) \mid X_1, \dots, X_i\big)\\
&= \mathbb{E}\big(h(\overset{i}{Y}) \mid X_1, \dots, X_{i-1}\big)\\
&= \mathbb{E}^{\mathfrak{F}_{i-1}}\big(h(X)\big).
\end{aligned}$$

Remark 3.3. – In the case when the random variables $X_1, \dots, X_N$ are independent, we can take for $\overset{i}{Y}_i$ an independent copy of X_i and we can put $\overset{i}{Y}{}_{i+1}^N = X_{i+1}^N$ a.s. With this choice, the definition of G_i given here coincides with that given in the first section.

We have

$$\big|G_i(Z)\big| = \big|\mathbb{E}^{\mathfrak{F}_N}\big(f(X) - f(\overset{i}{Y})\big)\big| \leqslant \mathbb{E}^{\mathfrak{F}_N}\left(\sum_{j=1}^N \mathbf{1}\big(X_j \neq \overset{i}{Y}_j\big)\frac{B_j}{\sqrt{N}}\right).$$

Consequently

$$\big|F_i(Z)\big| \leqslant \operatorname{ess\,sup} \mathbb{E}^{\mathfrak{F}_i}\left(\sum_{j=1}^N \mathbf{1}\big(X_j \neq \overset{i}{Y}_j\big)\frac{B_j}{\sqrt{N}}\right).$$

Let us introduce the notation

$$\widetilde{B}_i = \operatorname{ess\,sup} \mathbb{E}\left(\sum_{j=1}^{N} \mathbf{1}\big(X_j \neq \overset{i}{Y}_j\big) B_j \mid \mathfrak{F}_i, \overset{i}{Y}_i \right). \tag{7}$$

We have established that

$$\big|F_i(Z)\big| \leqslant \frac{\widetilde{B}_i}{\sqrt{N}}.$$

We can now proceed exactly in the same way as in the independent case to prove that

LEMMA 3.1. –

$$\left| \log \mathbb{E} \exp(\lambda Z) - \frac{\lambda^2}{2} \sum_{i=1}^{N} \mathbb{E}_{\lambda \mathbb{E}^{\mathfrak{F}_{i-1}}(Z)}\big(F_i(Z)^2\big) \right| \leqslant \sum_{i=1}^{N} \frac{\lambda^3 \widetilde{B}_i^3}{3N^{3/2}}.$$

To go further, we would like to bound

$$\mathbb{E}_{\lambda \mathbb{E}^{\mathfrak{F}_{i-1}}(Z)}\big(F_i(Z)^2\big) - \mathbb{E}\big(F_i(Z)^2\big),$$

which we will decompose as in the independent case into

$$\sum_{j=1}^{i-1} \mathbb{E}_{\lambda Z_{i-1}}\big[F_j\big(F_i(Z)^2\big)\big].$$

Among other things, we will have to bound $\operatorname{ess\,sup} F_j(F_i(Z)^2)$. Let us start with this. For any bounded measurable function $h(X)$, we have

$$\begin{aligned}
\big|F_j\big(h(X)^2\big)\big| &= \big|\mathbb{E}^{\mathfrak{F}_j}\big(h(X)^2 - h(\overset{j}{Y})^2\big)\big| \\
&= \big|\mathbb{E}^{\mathfrak{F}_j}\big((h(X) + h(\overset{j}{Y}))(h(X) - h(\overset{j}{Y}))\big)\big| \\
&\leqslant 2 \operatorname{ess\,sup}\big|h(X)\big| \mathbb{E}^{\mathfrak{F}_j}\big|h(X) - h(\overset{j}{Y})\big|.
\end{aligned}$$

We will apply this to $h(X) = F_i(Z)$, and in this case, we will try to express $h(X) - h(\overset{j}{Y})$ as a difference "of order two" of four coupled processes. Let us build these processes right now, since we cannot proceed without them. We will call them $(X, \overset{j}{Y}, \overset{i}{Y}, \overset{i}{Y})$. The distribution of $(X, \overset{j}{Y})$ and $(X, \overset{i}{Y})$ on their canonical spaces will be as previously defined. Let us repeat this construction here, to make precise the fact that we can build them in such a way that they satisfy the Markov property, when X does:

- We build $\mathbb{P}(dX_1^{i-1}, d\overset{i}{Y}{}_1^{i-1})$ as $\mathbb{P}(dX_1^{i-1})\delta_{X_1^{i-1}}(d\overset{i}{Y}{}_1^{i-1})$, where $\delta_{X_1^{i-1}}$ is the Dirac mass at point X_1^{i-1} in $\prod_{k=1}^{i-1} \mathfrak{X}_k$.
- We then put

$$\mathbb{P}\big(dX_i, d\overset{i}{Y}_i \mid X_1^{i-1}, \overset{i}{Y}{}_1^{i-1}\big) = \mathbb{P}(dX_i \mid X_{i-1}) \otimes \mathbb{P}\big(d\overset{i}{Y}_i \mid \overset{i}{Y}_{i-1}\big),$$

and for $k > i$ we build $\mathbb{P}(dX_k, d\overset{i}{Y}_k \mid (X, \overset{i}{Y})_1^{k-1})$ as some maximal coupling between $\mathbb{P}(dX_k \mid X_{k-1})$ and $\mathbb{P}(d\overset{i}{Y}_k \mid \overset{i}{Y}_{k-1})$, which we choose in a fixed way, independent of $(X, \overset{i}{Y})_1^{k-2}$. Thus built, $(X, \overset{i}{Y})$ is a Markov chain.

- We build $(X, \overset{j}{Y})$ in the same way, with the index i replaced by j. Then we define the distribution of $(\overset{j}{Y}, \overset{i}{Y})$ on its canonical space to be the same as the distribution of $(X, \overset{i}{Y})$.

These preliminaries being set, we are ready to define the distribution of $(X, \overset{j}{Y}, \overset{i}{Y}, \overset{i}{Y})$ on its canonical space. Let us put for convenience $T_k = (X_k, \overset{j}{Y}_k, \overset{i}{Y}_k, \overset{i}{Y}_k)$. We set

$$\mathbb{P}\big(dX_k, d\overset{j}{Y}_k \mid T_1^{k-1}\big) = \mathbb{P}\big(dX_k, d\overset{j}{Y}_k \mid X_{k-1}, \overset{j}{Y}_{k-1}\big),$$

which we have already defined, and we take for

$$\mathbb{P}\big(d\overset{i}{Y}_k, d\overset{i}{U}_k \mid T_1^{k-1}, X_k, \overset{j}{Y}_k\big)$$

some maximally coupled distribution depending only on $(T_{k-1}, X_k, \overset{j}{Y}_k)$ with marginals

$$\mathbb{P}\big(d\overset{i}{Y}_k \mid X_{k-1}, \overset{i}{Y}_{k-1}, X_k\big)$$

and

$$\mathbb{P}\big(d\overset{i}{U}_k \mid \overset{j}{Y}_{k-1}, \overset{i}{U}_{k-1}, \overset{j}{Y}_k\big)$$

which we have already defined.

Remark 3.4. – The processes $\overset{i}{Y}$ and $\overset{j}{Y}$ are independent knowing X, therefore this construction is compatible with the previous one. Indeed

$$\begin{aligned}
\mathbb{P}\big(dX, d\overset{j}{Y}, d\overset{i}{Y}\big) &= \prod_{k=1}^{N} \mathbb{P}\big(dX_k, d\overset{j}{Y}_k, d\overset{i}{Y}_k \mid (X, \overset{j}{Y}, \overset{i}{Y})_1^{k-1}\big) \\
&= \prod_{k=1}^{N} \mathbb{P}(dX_k \mid X_{k-1}) \mathbb{P}\big(d\overset{j}{Y}_k \mid X_k, X_{k-1}, \overset{j}{Y}_{k-1}\big) \mathbb{P}\big(d\overset{i}{Y}_k \mid X_k, X_{k-1}, \overset{i}{Y}_{k-1}\big),
\end{aligned}$$

thus

$$\begin{aligned}
\mathbb{P}\big(d\overset{j}{Y}, d\overset{i}{Y} \mid X\big) &= \prod_{k=1}^{N} \mathbb{P}\big(d\overset{j}{Y}_k \mid X_k, X_{k-1}, \overset{j}{Y}_{k-1}\big) \prod_{k=1}^{N} \mathbb{P}\big(d\overset{i}{Y}_k \mid X_k, X_{k-1}, \overset{i}{Y}_{k-1}\big) \\
&= \mathbb{P}\big(d\overset{j}{Y} \mid X\big) \otimes \mathbb{P}\big(d\overset{i}{Y} \mid X\big).
\end{aligned}$$

The following lemma will be important to carry the computations (let us recall that in this discussion $h(X) \overset{\text{def}}{=} F_i(Z)$ and that $j < i$):

LEMMA 3.2. –

$$h(X) = \mathbb{E}\big(f(X) - f(\overset{i}{Y}) \mid X_1^i\big) = \mathbb{E}\big(f(X) - f(\overset{i}{Y}) \mid X_1^i, \overset{j}{Y}{}_1^i\big)$$

and in the same way

$$h(\overset{j}{Y}) = \mathbb{E}\big(f(\overset{j}{Y}) - f(\overset{i}{Y}) \mid \overset{j}{Y}{}_1^i\big) = \mathbb{E}\big(f(\overset{j}{Y}) - f(\overset{i}{Y}) \mid X_1^i, \overset{j}{Y}{}_1^i\big).$$

Proof. – Let us remark first that

$$\mathbb{E}\big(f(X) \mid X_1^i\big) = \mathbb{E}\big(f(X) \mid X_1^i, \overset{j}{Y}{}_1^i\big),$$

because $(X_{i+1}^N \perp\!\!\!\perp \overset{j}{Y}{}_1^i \mid X_1^i)$, (we use this short notation here and below to mean that "X_{i+1}^N is independent of $\overset{j}{Y}{}_1^i$ conditionally on X_1^i").

Moreover, from the construction of the coupled process T, we see that

$$\mathbb{P}\big(d\overset{i}{Y}{}_1^N, dX_1^i, d\overset{j}{Y}{}_1^i\big) = \prod_{k=1}^{i} \mathbb{P}(dX_k \mid X_{k-1})\mathbb{P}\big(d\overset{j}{Y}_k \mid X_k, X_{k-1}, \overset{j}{Y}_{k-1}\big)$$
$$\times \mathbb{P}\big(d\overset{i}{Y}_k \mid X_k, X_{k-1}, \overset{i}{Y}_{k-1}\big) \prod_{k=i+1}^{N} \mathbb{P}\big(d\overset{i}{Y}_k \mid \overset{i}{Y}_{k-1}\big),$$

and therefore that

$$\mathbb{P}\big(d\overset{i}{Y}{}_1^N \mid X_1^i, \overset{j}{Y}{}_1^i\big) = \prod_{k=1}^{i} \mathbb{P}\big(d\overset{i}{Y}_k \mid X_k, X_{k-1}, \overset{i}{Y}_{k-1}\big) \prod_{k=i+1}^{N} \mathbb{P}\big(d\overset{i}{Y}_k \mid \overset{i}{Y}_{k-1}\big)$$
$$= \mathbb{P}\big(d\overset{i}{Y}{}_1^N \mid X_1^i\big).$$

As the couples of random variables $(X, \overset{i}{Y})$ and $(\overset{j}{Y}, \overset{i}{Y})$ play symmetric roles (they can be chosen to be exchangeable by a proper construction of T, but even without this refinement, the proof applies mutatis mutandis when the roles of $(X, \overset{i}{Y})$ and $(\overset{j}{Y}, \overset{i}{Y})$ are exchanged), we have in the same way

$$\mathbb{E}\big(f(\overset{j}{Y}) \mid X_1^i\big) = \mathbb{E}\big(f(\overset{j}{Y}) \mid X_1^i, \overset{j}{Y}{}_1^i\big),$$
$$\mathbb{E}\big(f(\overset{i}{U}) \mid X_1^i\big) = \mathbb{E}\big(f(\overset{i}{U}) \mid X_1^i, \overset{j}{Y}{}_1^i\big). \qquad \square$$

We deduce from the previous lemma that

$$\mathbb{E}^{\mathfrak{F}_j}\big|h(X) - h(\overset{j}{Y})\big| = \mathbb{E}^{\mathfrak{F}_j}\big|\mathbb{E}\big(f(X) - f(\overset{i}{Y}) - f(\overset{j}{Y}) + f(\overset{i}{U}) \mid (X, \overset{j}{Y})_1^i\big)\big|$$
$$\leqslant \mathbb{E}^{\mathfrak{F}_j}\big(\big|f(X) - f(\overset{i}{Y}) - f(\overset{j}{Y}) + f(\overset{i}{U})\big|\big).$$

To write the right-hand side of this last inequality as far as possible as a function of the second differences $\Delta_\ell \Delta_k f$, we need one more lemma: let us introduce the two stopping times

$$\tau_i = \inf\{k \geqslant i \mid \overset{i}{Y}_k = X_k\},$$
$$\tau_j = \inf\{k \geqslant j \mid \overset{j}{Y}_k = X_k\}.$$

LEMMA 3.3. – *With the previous construction, we have*

$$\mathbb{P}\big(\overset{i}{U}{}_i^N = \overset{i}{Y}{}_i^N \mid \tau_j < i\big) = 1.$$

In other words, on the event $(\tau_j < i)$ it is almost surely true that $\overset{i}{U}{}_i^N = \overset{i}{Y}{}_i^N$.

Proof. – We have obviously $\overset{j}{Y}{}_{\tau_j}^N = X_{\tau_j}^N$ almost surely. Now when $\tau_j < i$, then a.s. $\overset{j}{Y}_{i-1} = X_{i-1} = \overset{i}{Y}_{i-1} = \overset{i}{U}_{i-1}$, and so $\overset{i}{U}_i$ and $\overset{i}{Y}_i$ knowing the past are maximally coupled and have the same marginals, therefore they are almost surely equal. Then we can carry on the same reasoning for $k = i+1, \ldots, N$ and thus prove by induction that for all these values of k, $\overset{i}{Y}_k = \overset{i}{Y}_k$ a.s. □

Resuming the previous chain of inequalities, we can write, as a consequence of this lemma, that

$$\begin{aligned}
&\mathbb{E}^{\mathfrak{F}_j}\big|h(X) - h(\overset{j}{Y})\big| \\
&\quad\leqslant \mathbb{P}^{\mathfrak{F}_j}(\tau_j \geqslant i)\frac{2\widetilde{B}_i}{\sqrt{N}} + \mathbb{E}^{\mathfrak{F}_j}\big(\mathbf{1}(\tau_j < i)\big|f(X) - f\big(X_1^{i-1}, \overset{i}{Y}{}_i^N\big) \\
&\qquad - f\big(X_1^{j-1}, \overset{j}{Y}{}_j^{\tau_j-1}, X_{\tau_j}^N\big) + f\big(X_1^{j-1}, \overset{j}{Y}{}_j^{\tau_j-1}, X_{\tau_j}^{i-1}, \overset{i}{Y}{}_i^N\big)\big|\big) \\
&\quad\leqslant \mathbb{P}^{\mathfrak{F}_j}(\tau_j \geqslant i)\frac{2\widetilde{B}_i}{\sqrt{N}} + \mathbb{E}^{\mathfrak{F}_j}\left(\mathbf{1}(\tau_j < i)\left|\sum_{k=i}^{\tau_i-1} \Delta_k f\big(\big(X_1^k, \overset{i}{Y}{}_{k+1}^N\big), \overset{i}{Y}_k\big)\right.\right. \\
&\qquad \left.\left. - \Delta_k f\big(\big(X_1^{j-1}, \overset{j}{Y}{}_j^{\tau_j-1}, X_{\tau_j}^k, \overset{i}{Y}{}_{k+1}^N\big), \overset{i}{Y}_k\big)\right|\right) \\
&\quad\leqslant \mathbb{P}^{\mathfrak{F}_j}(\tau_j \geqslant i)\frac{2\widetilde{B}_i}{\sqrt{N}} \\
&\qquad + \mathbb{E}^{\mathfrak{F}_j}\left(\mathbf{1}(\tau_j < i)\left|\sum_{k=i}^{\tau_i-1}\sum_{\ell=j}^{\tau_j-1} \Delta_\ell \Delta_k f\big(\big(X_1^\ell, \overset{j}{Y}{}_{\ell+1}^{\tau_j-1}, X_{\tau_j}^k, \overset{i}{Y}{}_{k+1}^N\big), \overset{i}{Y}_k, \overset{j}{Y}_\ell\big)\right|\right) \\
&\quad\leqslant \mathbb{P}^{\mathfrak{F}_j}(\tau_j \geqslant i)\frac{2\widetilde{B}_i}{\sqrt{N}} + \mathbb{E}^{\mathfrak{F}_j}\left(\sum_{k=i}^{\tau_i-1}\sum_{\ell=j}^{\tau_j-1} \frac{C_{\ell,k}}{N^{3/2-\zeta}}\right).
\end{aligned}$$

Let us put

$$\widetilde{C}_{i,j} = \operatorname{ess\,sup} \mathbb{E}^{\mathfrak{F}_j} \left(\sum_{k=i}^{\tau_i - 1} \sum_{\ell=j}^{\tau_j - 1} C_{k,\ell} \right). \tag{8}$$

We get

$$\left| F_j\big(F_i(Z)^2\big) \right| \leqslant \frac{2\widetilde{B}_i \widetilde{C}_{i,j}}{N^{2-\zeta}} + \mathbb{P}^{\mathfrak{F}_j}(\tau_j \geqslant i) \frac{4\widetilde{B}_i^2}{N}.$$

Let us now define $\mathfrak{G}_j$ to be $\mathfrak{S}(\overset{j}{Y})$, the sigma algebra generated by $(\overset{j}{Y}_1, \ldots, \overset{j}{Y}_N)$. We have

$$\mathbb{E}^{\mathfrak{G}_j}\big[F_j\big(F_i(Z)^2\big)\big] = 0,$$

because X_j and $\overset{j}{Y}{}_j^N$ are conditionally independent knowing X_1^{j-1}. Let moreover

$$\begin{aligned}
\widetilde{G}_j &= Z_{i-1}(X) - Z_{i-1}(\overset{j}{Y}) \\
&= \mathbb{E}\big(f(X) \mid X_1^{i-1}\big) - \mathbb{E}\big(f(\overset{j}{Y}) \mid (\overset{j}{Y})_1^{i-1}\big) \\
&= \mathbb{E}\big(f(X) - f(\overset{j}{Y}) \mid X_1^{i-1}, \overset{j}{Y}{}_1^{i-1}\big).
\end{aligned}$$

We have

$$\begin{aligned}
\mathbb{E}_{\lambda Z_{i-1}}\big[F_j\big(F_i(Z)^2\big)\big] &= \mathbb{E}_{\lambda Z_{i-1}}\big\{\mathbb{E}^{\mathfrak{G}_j}_{\lambda Z_{i-1}}\big[F_j\big(F_i(Z)^2\big)\big]\big\} \\
&= \mathbb{E}_{\lambda Z_{i-1}}\big\{\mathbb{E}^{\mathfrak{G}_j}_{\lambda \tilde{G}_j}\big[F_j\big(F_i(Z)^2\big)\big]\big\} \\
&= \mathbb{E}_{\lambda Z_{i-1}} \int_0^\lambda \mathbb{E}^{\mathfrak{G}_j}_{\alpha \tilde{G}_j}\big\{F_j\big(F_i(Z)^2\big)\big[\widetilde{G}_j - \mathbb{E}^{\mathfrak{G}_j}_{\alpha \tilde{G}_j}(\widetilde{G}_j)\big]\big\}\, d\alpha.
\end{aligned}$$

Therefore

$$\begin{aligned}
&\big|\mathbb{E}_{\lambda Z_{i-1}}\big[F_j\big(F_i(Z)^2\big)\big]\big| \\
&\quad\leqslant \operatorname{ess\,sup}\big|F_j\big(F_i(Z)^2\big)\big| \mathbb{E}_{\lambda Z_{i-1}} \int_0^\lambda \mathbb{E}^{\mathfrak{G}_j}_{\alpha \tilde{G}_j}\big|\widetilde{G}_j - \mathbb{E}^{\mathfrak{G}_j}_{\alpha \tilde{G}_j}(\widetilde{G}_j)\big|\, d\alpha \\
&\quad\leqslant 2 \operatorname{ess\,sup}\big|F_j\big(F_i(Z)^2\big)\big| \mathbb{E}_{\lambda Z_{i-1}} \int_0^\lambda \mathbb{E}^{\mathfrak{G}_j}_{\alpha \tilde{G}_j}\big|\widetilde{G}_j\big|\, d\alpha \\
&\quad\leqslant 2 \operatorname{ess\,sup}\big|F_j\big(F_i(Z)^2\big)\big| \operatorname{ess\,sup} \mathbb{E}^{\mathfrak{G}_j}\left(\int_0^\lambda \exp(\alpha \widetilde{G}_j)|\widetilde{G}_j|\, d\alpha\right)\big(\mathbb{E}^{\mathfrak{G}_j}\big(\exp(\lambda|\widetilde{G}_j|)\big)\big)^{-1} \\
&\quad\leqslant 2 \operatorname{ess\,sup}\big|F_j\big(F_i(Z)^2\big)\big| \operatorname{ess\,sup} \mathbb{E}^{\mathfrak{G}_j}\big(\exp(\lambda|\widetilde{G}_j|) - 1\big)\mathbb{E}^{\mathfrak{G}_j}\big[\exp(\lambda|\widetilde{G}_j|)\big].
\end{aligned}$$

Moreover

$$\begin{aligned}\mathbb{E}^{\mathfrak{G}_j}\big[\exp(\lambda|\widetilde{G}_j|)\big] &= \mathbb{E}\big\{\exp\big(\lambda\big|\mathbb{E}\big(f(X)-f(\overset{j}{Y})\mid X_1^{i-1},\overset{j}{Y}{}_1^{i-1}\big)\big|\big)\mid \overset{j}{Y}\big\}\\ &\leqslant \mathbb{E}\big\{\mathbb{E}\big\{\exp\big[\lambda\big|f(X)-f(\overset{j}{Y})\big|\big]\mid X_1^{i-1},\overset{j}{Y}{}_1^{i-1}\big\}\mid \overset{j}{Y}\big\}\\ &= \mathbb{E}\big\{\exp\big[\lambda\big|f(X)-f(\overset{j}{Y})\big|\big]\mid \overset{j}{Y}{}_1^{i-1}\big\}\end{aligned}$$

because $(X_1^{i-1} \perp\!\!\!\perp \overset{j}{Y}{}_i^N \mid \overset{j}{Y}{}_1^{i-1})$

$$\leqslant \operatorname{ess\,sup}\mathbb{E}^{\mathfrak{G}_j}\left[\exp\left(\lambda\sum_{k=j}^{\tau_j-1}\frac{B_k}{\sqrt{N}}\right)\right].$$

Let us put

$$\widetilde{\widetilde{B}}_j(\lambda) = \operatorname{ess\,sup}\frac{\sqrt{N}}{\lambda}\mathbb{E}^{\mathfrak{G}_j}\left[\exp\left(\lambda\sum_{k=j}^{\tau_j-1}\frac{B_k}{\sqrt{N}}\right)-1\right]\mathbb{E}^{\mathfrak{G}_j}\left[\exp\left(\lambda\sum_{k=j}^{\tau_j-1}\frac{B_k}{\sqrt{N}}\right)\right]. \tag{9}$$

We have

$$\mathbb{E}_{\lambda Z_{i-1}}\big(F_j\big(F_i(Z)^2\big)\big) \leqslant \frac{2\lambda}{\sqrt{N}}\widetilde{\widetilde{B}}_j(\lambda)\frac{2\widetilde{B}_i\widetilde{C}_{i,j}}{N^{2-\zeta}} + \operatorname{ess\,sup}\mathbb{P}^{\mathfrak{F}_j}(\tau_j \geqslant i)\frac{4\widetilde{B}_i^2}{N}.$$

Thus

$$\begin{aligned}\sum_{i=1}^{N}\sum_{j=1}^{i-1}\frac{\lambda^2}{2}\mathbb{E}_{\lambda Z_{i-1}}\big[F_j\big(F_i(Z)^2\big)\big] &\leqslant \frac{\lambda^3}{N^{1/2-\zeta}}\sum_{1\leqslant j<i\leqslant N}\frac{2\widetilde{B}_i\widetilde{C}_{i,j}\widetilde{\widetilde{B}}_j(\lambda)}{N^2}\\ &\quad + \frac{\lambda^3}{\sqrt{N}}\sum_{j=1}^{N}\sum_{i=j+1}^{N}\operatorname{ess\,sup}\mathbb{P}^{\mathfrak{F}_j}(\tau_j \geqslant i)\frac{4\widetilde{B}_i^2\widetilde{\widetilde{B}}_j(\lambda)}{N}.\end{aligned}$$

Therefore if we put

$$\check{B}_j = \sqrt{\sum_{i=j+1}^{N}\widetilde{B}_i^2\operatorname{ess\,sup}\mathbb{P}^{\mathfrak{F}_j}(\tau_j \geqslant i)}, \tag{10}$$

we obtain the following theorem:

THEOREM 3.1. – *When* $(X_1,\dots,X_N)$ *satisfies the Markov property and the function* f *satisfies*

$$\sup_{x,y_i}\Delta_i f(x,y_i) \leqslant \frac{B_i}{\sqrt{N}},$$
$$\sup_{x,y_i,y_j}\Delta_j\Delta_i f(x,y_i,y_j) \leqslant \frac{C_{i,j}}{N^{3/2-\zeta}},$$

then

$$\left|\log \mathbb{E}\big(\exp(\lambda Z)\big) - \frac{\lambda^2}{2}\mathbb{E}\big(Z^2\big)\right| \leqslant \frac{\lambda^3}{N^{1/2-\zeta}} \sum_{1\leqslant j<i\leqslant N} \frac{2\widetilde{B}_i \widetilde{C}_{i,j} \widetilde{\widetilde{B}}_j(\lambda)}{N^2} + \frac{\lambda^3}{\sqrt{N}} \sum_{i=1}^N \left(\frac{\widetilde{B}_i^3}{3N} + \frac{4\check{B}_i^2 \widetilde{\widetilde{B}}_i(\lambda)}{N}\right),$$

where the constants $\widetilde{B}_i$ *are defined by* (7)*, the constants* $\widetilde{C}_{i,j}$ *are defined by* (8)*, the constants* $\widetilde{\widetilde{B}}_i(\lambda)$ *are defined by* (9) *and the constants* $\check{B}_i$ *are defined by* (10).

COROLLARY 3.1. – *Let us assume that* $(X_1, \dots, X_N)$ *is a Markov chain such that for some positive constants* A *and* $\rho < 1$

$$\mathbb{P}(\tau_i > i + k \mid \mathfrak{G}_i, X_i) \leqslant A\rho^k, \quad \text{a.s.,} \tag{11}$$

$$\mathbb{P}(\tau_i > i + k \mid \mathfrak{F}_N, \overset{i}{Y}_i) \leqslant A\rho^k, \quad \text{a.s.,} \tag{12}$$

and let us put $B = \max_i B_i$ *and* $C = \max_{i,j} C_{i,j}$. *Then*

$$\begin{aligned}
&\left|\log \mathbb{E}\big(\exp(\lambda Z)\big) - \frac{\lambda^2}{2}\mathbb{E}\big(Z^2\big)\right| \\
&\quad \leqslant \frac{\lambda^3}{N^{1/2-\zeta}} \frac{BCA^3}{(1-\rho)^3} \left(\frac{\rho\log(\rho^{-1})}{2AB} - \frac{\lambda}{\sqrt{N}}\right)_+^{-1} \\
&\qquad + \frac{\lambda^3}{\sqrt{N}} \left(\frac{B^3A^3}{3(1-\rho)^3} + \frac{4B^2A^3}{(1-\rho)^3}\left(\frac{\rho\log(\rho^{-1})}{2AB} - \frac{\lambda}{\sqrt{N}}\right)_+^{-1}\right).
\end{aligned}$$

Consequently

$$\mathbb{P}\big(f(X) \geqslant \mathbb{E}\big(f(X)\big) + \varepsilon\big) \leqslant \exp\left(-\frac{\varepsilon^2}{2(\mathbb{V}(f(X)) + \frac{2\eta\varepsilon}{\mathbb{V}(f(X))})}\right),$$

$$\mathbb{P}\big(f(X) \leqslant \mathbb{E}\big(f(X)\big) - \varepsilon\big) \leqslant \exp\left(-\frac{\varepsilon^2}{2(\mathbb{V}(f(X)) + \frac{2\eta\varepsilon}{\mathbb{V}(f(X))})}\right),$$

where

$$\begin{aligned}
\eta = {} & \frac{1}{N^{1/2-\zeta}} \frac{BCA^3}{(1-\rho)^3}\left(\frac{\rho\log(\rho^{-1})}{2AB} - \frac{\varepsilon}{\mathbb{V}(f(X))\sqrt{N}}\right)_+^{-1} \\
& + \frac{1}{\sqrt{N}}\left(\frac{B^3A^3}{3(1-\rho)^3} + \frac{4B^2A^3}{(1-\rho)^3}\left(\frac{\rho\log(\rho^{-1})}{2AB} - \frac{\varepsilon}{\mathbb{V}(f(X))\sqrt{N}}\right)_+^{-1}\right).
\end{aligned}$$

Remark 3.5. – If we choose the distribution of the pair $(X, \overset{i}{Y})$ to be exchangeable, and this can always be done, then the two conditions (11) and (12) are equivalent and one is of course superfluous.

Remark 3.6. – The hypotheses are for example fulfilled by any irreducible aperiodic homogeneous Markov chain on a finite state space.

Proof of the corollary. – We have

$$\widetilde{B}_i \leqslant B \operatorname{ess\,sup} \mathbb{E}^{\mathfrak{F}_N}(\tau_j - j) = B \sum_{k=0}^{+\infty} \operatorname{ess\,sup} \mathbb{P}^{\mathfrak{F}_N}(\tau_j > j+k) \leqslant AB \sum_{k=0}^{+\infty} \rho^k = \frac{BA}{1-\rho}.$$

In the same way

$$\begin{aligned}\widetilde{C}_{i,j} &\leqslant C \operatorname{ess\,sup} \mathbb{E}^{\widetilde{\mathfrak{F}}_N}\big((\tau_i - i)(\tau_j - j)\big)\\ &= C \operatorname{ess\,sup} \mathbb{E}^{\widetilde{\mathfrak{F}}_N}(\tau_i - i)\mathbb{E}^{\widetilde{\mathfrak{F}}_N}(\tau_j - j) \leqslant \frac{CA^3}{(1-\rho)^2},\end{aligned}$$

where we have used the fact that $(\overset{i}{Y} \perp\!\!\!\perp \overset{j}{Y} \mid X)$. We also have

$$\begin{aligned}\mathbb{E}^{\mathfrak{G}_j}\left(\exp\left(\lambda \sum_{k=j}^{\tau_j - 1} \frac{B_k}{\sqrt{N}}\right) - 1\right) &= \int_0^{+\infty} \mathbb{P}^{\mathfrak{G}_j}\left(\exp\left(\lambda \sum_{k=j}^{\tau_j-1} \frac{B_k}{\sqrt{N}}\right) - 1 \geqslant \xi\right) d\xi\\ &\leqslant \int_0^{+\infty} \mathbb{P}^{\mathfrak{G}_j}\left((\tau_j - j) \geqslant \frac{\sqrt{N}}{\lambda B}\log(1+\xi)\right) d\xi\\ &\leqslant \int_0^{+\infty} \frac{A}{\rho}\exp\left(\frac{\sqrt{N}}{\lambda B}\log(\rho)\log(1+\xi)\right) d\xi\\ &\leqslant \frac{A}{\rho}\left(\frac{\sqrt{N}\log(\rho^{-1})}{\lambda B} - 1\right)_+^{-1}.\end{aligned}$$

Thus

$$\begin{aligned}\frac{\lambda}{\sqrt{N}}\widetilde{\widetilde{B}}_j(\lambda) &\leqslant \frac{A}{\rho}\left(\frac{\sqrt{N}\log(\rho^{-1})}{\lambda B} - 1\right)_+^{-1}\left(1 + \frac{A}{\rho}\left(\frac{\sqrt{N}\log(\rho^{-1})}{\lambda B} - 1\right)_+^{-1}\right)\\ &\leqslant \frac{2A}{\rho}\left(\frac{\sqrt{N}\log(\rho^{-1})}{\lambda B} - \frac{2A}{\rho}\right)_+^{-1} = \left(\frac{\sqrt{N}\rho\log(\rho^{-1})}{2\lambda AB} - 1\right)_+^{-1},\end{aligned}$$

and

$$\widetilde{\widetilde{B}}_j(\lambda) \leqslant \left(\frac{\rho\log(\rho^{-1})}{2AB} - \frac{\lambda}{\sqrt{N}}\right)_+^{-1}.$$

On the other hand

$$\check{B}_j \leqslant \frac{BA}{1-\rho}\sqrt{\sum_{i=j+1}^{N} A\rho^{i-j-1}} \leqslant \frac{BA^{3/2}}{(1-\rho)^{3/2}}.$$

Substituting all these upper bounds in the theorem proves its corollary. □

Conclusion

We have shown that under quite natural boundedness or exponential moment assumptions, it is possible to get non-asymptotic bounds for the distance between the log-Laplace transform of a function of N random variables and the transform of the corresponding Gaussian random variable. In particular, no convexity assumption is required and we can deal not only with independent random variables, but also with a large class of Markov chains. We hope to present some applications of these bounds in forthcoming studies.

Acknowledgements

I am glad to thank Pascal Massart who kindly gave me an early draft of his paper [14], which was a source of inspiration and stimulation to write this one. I am also grateful to the referee who suggested valuable improvements in the presentation of the proofs.

REFERENCES

[1] K. Azuma, Weighted sums of certain dependent random variables, Tôhoku Math. J. (2) 19 (1967) 357–367.
[2] S. Boucheron, G. Lugosi, P. Massart, A sharp concentration inequality with applications, Prépublication d'Orsay numéro 25 (27/9/1999), http://www.math.u-psud.fr/~biblio/html/ppo.html.
[3] O. Catoni, "Universal" aggregation rules with exact bias bounds, preprint PMA-510, http://www.proba.jussieu.fr/mathdoc/preprints/index.html#1999, Ann. Statist., to appear in revised form (under the modified title: Universal aggregation rules with sharp oracle inequalities).
[4] O. Catoni, Gibbs estimators, preprint LMENS-98-21, 1998, http://www.dmi.ens.fr/preprints/.
[5] A. Dembo, O. Zeitouni, Transportation approach to some concentration inequalities in product spaces, Electron. Comm. Probab. 1 (9) (1996) 83–90 (electronic).
[6] W. Hoeffding, Probability inequalities for sums of bounded random variables, J. Amer. Statist. Assoc. 58 (1963) 13–30.
[7] M. Ledoux, Concentration of measure and logarithmic Sobolev inequalities, Notes Berlin, 1997.
[8] M. Ledoux, On Talagrand's deviation inequalities for product measures, ESAIM Probab. Statist. 1 (1995) 63–87.
[9] M. Ledoux, Isoperimetry and Gaussian analysis, in: Lectures on Probability Theory and Statistics (Saint-Flour, 1994), in: Lecture Notes in Math., Vol. 1648, Springer, Berlin, 1996, pp. 165–294.
[10] M. Ledoux, M. Talagrand, Probability in Banach Spaces, Springer, Berlin, 1991.
[11] K. Marton, Measure concentration for a class of random processes, Probab. Theory Related Fields 110 (3) (1998) 427–439.
[12] K. Marton, A measure concentration inequality for contracting Markov chains, Geom. Funct. Anal. 6 (3) (1996) 556–571.
[13] K. Marton, Bounding $\overline{d}$-distance by informational divergence: a method to prove measure concentration, Ann. Probab. 24 (2) (1996) 857–866.
[14] P. Massart, About the constants in Talagrand's concentration inequalities for empirical processes, Ann. Probab., 1999, to appear.
[15] P. Massart, Optimal constants for Hoeffding type inequalities, Prépublication d'Orsay numéro 86 (18/12/1998), http://www.math.u-psud.fr/~biblio/html/ppo.html.
[16] C. McDiarmid, Concentration, in: M. Habib, C. McDiarmid, B. Reed (Eds.), Probabilistic Methods for Algorithmic Discrete Mathematics, Springer, 1998.

[17] K.R. Parthasarathy, Probability Measures on Metric Spaces, Academic Press, New York, 1967.
[18] P.-M. Samson, Concentration of measure inequalities for Markov chains and Φ-mixing processes, Ann. Probab., 1998, to appear.
[19] P.-M. Samson, Inégalités de concentration de la mesure pour des chaînes de Markov et des processus Φ-mélangeants, PhD, Université Paul Sabatier, Toulouse, France, June 1998.
[20] M. Talagrand, Concentration of measure and isoperimetric inequalities in product spaces, Publ. Math. IHES 81 (1995) 73–205.
[21] V.V. Yurinskiĭ, Exponential estimates for large deviations, Teor. Verojatnost. i Primenen. 19 (1974) 152–154 (Russian).