



HAL
open science

Sharp Bounds for the Tails of Functionals of Markov Chains

Patrice Bertail, Stéphane Clémenton

► **To cite this version:**

Patrice Bertail, Stéphane Clémenton. Sharp Bounds for the Tails of Functionals of Markov Chains. 2007. hal-00140591

HAL Id: hal-00140591

<https://hal.science/hal-00140591>

Preprint submitted on 7 Apr 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Sharp Bounds for the Tails of Functionals of Markov Chains

Patrice Bertail

MODALX - Université Paris X Nanterre

&

CREST (INSEE)

Stéphan Cléménçon

MODALX - Université Paris X Nanterre

&

Laboratoire de Probabilités et Modèles Aléatoires

UMR CNRS 7599 - Universités Paris VI et Paris VII

April 6, 2007

Abstract

This paper is devoted to establishing sharp bounds for deviation probabilities of partial sums $\sum_{i=1}^n f(X_i)$, where $X = (X_n)_{n \in \mathbb{N}}$ is a positive recurrent Markov chain and f is a real valued function defined on its state space. Combining the regenerative method to the Esscher transformation, these estimates are shown in particular to generalize probability inequalities proved in the i.i.d. case to the Markovian setting for (not necessarily uniformly) geometrically ergodic chains.

AMS 2000 subject classification: Primary 60E15, 60J27, Secondary 60K05.

Keywords and phrases: Probability inequalities, Markov chain, regenerative method

1 Introduction

Consider a positive recurrent Markov chain $X = (X_n)_{n \in \mathbb{N}}$ with transition probability Π , state space E with countably generated σ -field \mathcal{E} and stationary distribution μ . It is well known that we may then restrict ourselves to the case when the chain X is regenerative, namely when there exists a measurable set A such that $\mu(A) > 0$ and $\Pi(x, \cdot) = \Pi(y, \cdot)$ for all $(x, y) \in A^2$, even if it entails considering a Nummelin extension of the initial chain (see [17]). By virtue of the strong Markov property, the sequence $(\tau_A(n))_{n \in \mathbb{N}}$ of successive passage times to the regeneration set A forms a (possibly delayed) renewal process in this

case and the segments $(X_{1+\tau_A(j)}, \dots, X_{\tau_A(j+1)})$, $j \geq 1$, obtained by dividing the sample path of the chain according to the regeneration times, are i.i.d. r.v.'s, valued in the torus $\mathbb{T} = \cup_{n \geq 1} \mathbb{E}^n$. Furthermore, the return time to the atom A has finite expectation and the stationary distribution may be viewed as a Pitman occupation measure:

$$\forall B \in \mathcal{E}, \mu(B) = \frac{1}{\mathbb{E}_A[\tau_A]} \mathbb{E}_A\left[\sum_{i=1}^{\tau_A} \mathbb{I}_{\{X_i \in B\}}\right],$$

where $\tau_A = \inf\{n \geq 1, X_n \in A\}$, $\mathbb{E}_A[\cdot]$ denotes the expectation conditioned on $X_0 \in A$ and $\mathbb{I}_{\{\mathcal{A}\}}$ the indicator function of any event \mathcal{A} . One may refer to [18] for an account on renewal theory applied to Markov chain analysis. Let $f : (E, \mathcal{E}) \rightarrow \mathbb{R}$ be a measurable function. Asymptotic expansions for the tail probabilities of the partial sums $n^{-1/2} \sum_{i=1}^n f(X_i)$ have been obtained via the *regenerative method* (see [23]), which technique consists roughly speaking in applying appropriate results for i.i.d. r.v.'s to the partial sums over regeneration cycles $\sum_{i=1+\tau_A(j)}^{\tau_A(j+1)} f(X_i)$, $j \geq 1$. Refer in particular to [4], [13] and [3] for such refinements of the CLT theorem. This paper aims at establishing, with the same means, non asymptotic bounds for the probability that the sum $\sum_{i=1}^n f(X_i)$ exceeds a prescribed number x , holding for all x and n , similar to tail bounds proved in the independent framework such as those obtained by W. Hoeffding, G. Bennett or S.V. Nagaev for instance. We point out that the regenerative method is by no means the sole technique for obtaining probability inequalities in the markovian setting. Such non asymptotic results may be established by martingale arguments (see [10]), coupling techniques based on decay rate assumptions for mixing coefficients (see [6] or [21] for instance), from results of the quasi-compact operator theory when the spectral gap property is assumed to hold for the transition kernel (see [12] and [11]) or by methods based on information inequalities as in [14] (see also [22]). However, the regenerative method imposes much less restrictions on the ergodicity properties of the chain than most alternative techniques. In particular it may be used without stipulating Doeblin's condition (*i.e.* uniform ergodicity) to be fulfilled.

The paper is organized as follows. Section 2 first gives an insight into how the regenerative approach may apply for deriving tail bounds in the positive recurrent case, appealing additionally to an exponential change of probability measure as in [24] and then states the main results of the paper. Technical proofs are postponed to section 3.

2 Probability Inequalities for Regenerative Markov Chains

We first introduce further notation. Let ν be some probability distribution on (E, \mathcal{E}) . Here and throughout, we denote by \mathbb{P}_ν (respectively, \mathbb{P}_A) the probability measure on the underlying space such that $X_0 \sim \nu$ (resp., conditioned on $X_0 \in A$) and by $\mathbb{E}_\nu[\cdot]$ (resp., by $\mathbb{E}_A[\cdot]$) the \mathbb{P}_ν -expectation (resp., the \mathbb{P}_A -expectation). Furthermore, it is assumed

throughout the paper that $\mathbb{E}_A[\tau_A^2] < \infty$ or equivalently that $\mathbb{E}_\mu[\tau_A] < \infty$, since by a standard result in renewal theory $\mathbb{P}_\mu(\tau_A = k) = \mu(A)\mathbb{P}_A(\tau_A \geq k)$ for all $k \geq 1$.

In the case when Y_1, \dots, Y_n are i.i.d. random variables such that $\mathbb{E}[Y_1] = 0$, $\mathbb{E}[Y_1^2] = \sigma^2$ and $|Y_1| \leq M$ for some constant M , the following tail estimate has been obtained in [1]

$$\mathbb{P}\left(\sum_{i=1}^n Y_i \geq x\right) \leq \exp\left\{-\frac{n\sigma^2}{M^2} H\left(\frac{Mx}{n\sigma^2}\right)\right\}, \quad (1)$$

with $H(x) = (1+x)\log(1+x) - x$.

When the Y_i 's are not bounded anymore, this inequality may be extended using truncation arguments, as in [9] (see also [16]) at the price of an additional term related to the tail behavior of the Y_i 's on the right-hand side of (1). For arbitrary positive constants x and M , we have the inequality

$$\begin{aligned} \mathbb{P}\left(\sum_{i=1}^n Y_i \geq x\right) &\leq \exp\left\{\frac{x}{M} - \left(\frac{x - nm_M}{M} + \frac{n\sigma_M^2}{M^2}\right) \log\left(1 + \frac{xM}{n\sigma_M^2}\right)\right\} + n\mathbb{P}(Y > M) \\ &\leq \exp\left\{-\frac{n\sigma_M^2}{M^2} H\left(\frac{Mx}{n\sigma_M^2}\right)\right\} + n\mathbb{P}(Y > M), \end{aligned} \quad (2)$$

with $m_M = \mathbb{E}[Y_1 \mathbb{I}_{\{|Y_1| \leq M\}}]$ and $\sigma_M^2 = \mathbb{E}[Y_1^2 \mathbb{I}_{\{|Y_1| \leq M\}}]$.

This paper is devoted to establishing analogous inequalities under the assumption that $Y_i = f(X_i)$, $1 \leq i \leq n$, where $f : (E, \mathcal{E}) \rightarrow \mathbb{R}$ is a μ -integrable function. Without loss of generality, we assume that $\mu(f) = \mathbb{E}_\mu[f(X_1)] = 0$ throughout the article.

2.1 The regenerative approach

Let $l_n = \sum_{i=1}^n \mathbb{I}\{X_i \in A\}$ be the number of renewals (*i.e.* visits to the regeneration set A) up to time n . We denote by $\alpha = \mathbb{E}_A(\tau_A)$ the mean inter-renewal time. For $j \geq 1$, define $S_j(f) = \sum_{i=1+\tau_A(j)}^{\tau_A(j+1)} f(X_i)$ and $s_j = \tau_A(j+1) - \tau_A(j)$. Notice that these are two sequences of i.i.d. random variables with common variance σ_f^2 and σ_1^2 respectively, which are assumed to be finite throughout the article. The regenerative method is based on the preliminary observation that the sum $\sum_{i \leq n} f(X_i)$ may be decomposed as follows on the event $\{l_n \geq 2\}$:

$$\sum_{i=1}^n f(X_i) = S_A(f) + \sum_{j=1}^{l_n-1} S_j(f) + S_n^{(n)}(f), \quad (3)$$

where $S_A(f) = \sum_{i=1}^{\tau_A} f(X_i)$ and $S_n^{(n)}(f) = \sum_{i=1+\tau_A(l_n)}^n f(X_i)$ with the usual convention regarding empty summation. It is noteworthy that the summands in (3) are generally not

independent for fixed n (observe that, when $f \equiv 1$ for instance, they sum up to n). Hence, when applied for establishing tail bounds, the regenerative method comprises three main steps. The first one consists in partitioning the underlying space Ω according all possible fashions for the chain to regenerate up to time n , so that the summands in (3) be mutually independent on each subset of the partition. The matter is next to establish a tail bound on each of these subsets by relying on the resulting independence structure. And the desired tail estimate is finally computed by summing all the bounds previously obtained in a way that the sum obtained may be identified.

Now that an insight into the principle of the method has been given, in the next subsection we turn to developing the argument for proving tail inequalities in a rigorous fashion. Observe first that for all strictly positive real numbers p_1 , p_2 and p_3 such that $1/p_1 + 1/p_2 + 1/p_3 = 1$, we have for all $x > 0$ the immediate bound

$$\begin{aligned} \mathbb{P}_\nu\left(\sum_{i=1}^n f(X_i) \geq x\right) &\leq \mathbb{P}_\nu\left(\sum_{i=1}^n f(X_i) \geq x, l_n \leq 1\right) + \mathbb{P}_\nu(S_A(f) \geq x/p_1) \\ &\quad + \mathbb{P}_\nu\left(\sum_{j=1}^{l_n-1} S_j(f) \geq x/p_2\right) + \mathbb{P}_\nu(S_n^{(n)}(f) \geq x/p_3), \end{aligned} \quad (4)$$

with the convention that $\sum_{j=1}^{l_n-1} S_j(f) = 0$ when $l_n \leq 1$. Whereas bounds for the first two terms on the right hand side of (4), as well as for the last one, may be easily derived from assumptions on the tails of τ_A and $S_A(f)$ under \mathbb{P}_ν and \mathbb{P}_A , special attention must be paid to the tail of $\sum_{j=1}^{l_n-1} S_j(f)$. Beyond the dependence structure among the summands emphasized above, it is noteworthy that the $S_j(f)$'s are generally unbounded (even though one assumes f to be bounded).

2.2 Main result

Although the technique we present here for establishing bounds for the tail behavior of $\sum_{j=1}^{l_n-1} S_j(f)$ is very general, we focus on a simple result for the sake of clarity. Precisely, we derive an inequality extending (2), whose right hand side consists of two components: the first one being of the form of an exponential bound involving truncated moments of the $(S_j(f), s_j)$'s, while the second one is related to their tail behavior. Other probability bounds can be deduced by a slight adaption of the argument described below. And extensions of moment inequalities of Rosenthal type for $\sum_{i \leq n} f(X_i)$ may be established with the same method. We also emphasize that, although the present study is confined to the markovian setting, the method applies in the same way to any (eventually continuous time and not necessarily markovian) stochastic process with a regenerative extension (see [25]).

Theorem 1 Consider a regenerative positive recurrent Markov chain $X = (X_n)_{n \in \mathbb{N}}$ with initial and stationary probability distributions ν and μ . Let A be an accessible atom for X and $f : (E, \mathcal{E}) \rightarrow \mathbb{R}$ some μ -integrable function. Assume further that

- (i) $\alpha = \mathbb{E}_A[\tau_A] < \infty$ and $0 < \sigma_A^2 = \mathbb{E}_A[\tau_A^2] < \infty$,
- (ii) $\mathbb{E}_A[S_A(f)] = 0$ and $0 < \sigma_f^2 = \mathbb{E}_A[S_A(f)^2] < \infty$.

Then, for any vector $M = (M_1, M_2) \in \mathbb{R}_+^{*2}$ of euclidian norm $\|M\| = (M_1^2 + M_2^2)^{1/2}$, there exists a constant C_M depending on σ_A^2 and σ_f^2 such that , for any $n \geq \alpha$,

$$\begin{aligned} \mathbb{P}_\nu\left(\sum_{j=1}^{l_n-1} S_j(f) \geq x\right) &\leq C_M \exp\left\{-\frac{n(1+|\tilde{\rho}|)\tilde{\sigma}^2}{2\|M\|^2} \text{H}\left(\frac{\|M\|\sqrt{2}}{(1+|\tilde{\rho}|)\tilde{\sigma}\tilde{\sigma}_f}x\right)\right\} \\ &+ (n-1)\mathbb{P}_A(|S_A(f)| \geq M_1) + (n-1)\mathbb{P}_A(\tau_A \geq M_2), \end{aligned} \quad (5)$$

with

$$\begin{aligned} \tilde{\sigma}_f^2 &= \text{Var}_A(S_A(f)\mathbb{I}_{\{|S_A(f)| \leq M_1\}}), \quad \tilde{\sigma}_A^2 = \text{Var}_A(\tau_A\mathbb{I}_{\{\tau_A \leq M_2\}}), \\ \tilde{\rho} &= (\tilde{\sigma}_f\tilde{\sigma}_A)^{-1}\text{Cov}_A(S_A(f)\mathbb{I}_{\{|S_A(f)| \leq M_1\}}, \tau_A\mathbb{I}_{\{\tau_A \leq M_2\}}), \\ \tilde{\sigma} &= \sqrt{\tilde{\sigma}_f^2\tilde{\sigma}_A^2/(\tilde{\sigma}_f^2 + \tilde{\sigma}_A^2)}, \end{aligned}$$

Var_A and Cov_A denoting respectively variance and covariance under \mathbb{P}_A and the constant C_M , depending on α , $\tilde{\sigma}_f^2$ and $\tilde{\sigma}^2$ only, being of order $O(\|M\|^2)$ as $\|M\| \rightarrow \infty$. Furthermore, we have, for any $n \geq \alpha$,

$$\begin{aligned} \mathbb{P}_\nu\left(\sum_{j=1}^{l_n-1} S_j(f) \geq x\right) &\leq C_M \exp\left\{-\frac{x^2\tilde{\sigma}/\tilde{\sigma}_f}{2(n(1+|\tilde{\rho}|)\tilde{\sigma}\tilde{\sigma}_f + \frac{xM\sqrt{2}}{3})}\right\} \\ &+ (n-1)\mathbb{P}_A(|S_A(f)| \geq M_1) + (n-1)\mathbb{P}_A(\tau_A \geq M_2). \end{aligned} \quad (6)$$

Before developping the argument based on the regenerative method, a few remarks are in order at this point.

Remark 1 (ON BLOCK-MOMENT CONDITIONS) Notice that conditions (i) and (ii) do not depend on the atom A chosen. Besides, when dealing with the nonregenerative case, these conditions have to be satisfied by a Nummelin (regenerative) extension of the chain X constructed from a *small set* S , namely an accessible set to which X returns in a given number of steps with probability uniformly bounded by below (see [18] for an account on the *Nummelin splitting technique*). A sufficient condition for the latter to hold is classically that both $\sup_{x \in S} \mathbb{E}_x[\tau_S^2]$ and $\sup_{x \in S} \mathbb{E}_x[(\sum_{i=1}^{\tau_S} f(X_i))^2]$ are finite. Recall also that these 'block' conditions may be replaced by Foster-Lyapunov drift conditions that generally appear as more tractable in practice (refer for instance to Chapter 11 in [15] for further details).

Remark 2 (ON CONSTANTS INVOLVED IN THE EXPONENTIAL TERM) It is noteworthy that, although the asymptotic distribution of $n^{-1/2} \sum_{j=1}^{l_n-1} S_j(f)$ as $n \rightarrow \infty$ does not depend on the atom A chosen, its distribution for fixed n does (when $f \equiv 1$ for instance, $\sum_{j=1}^{l_n-1} S_j(f) = \tau_A(l_n) - \tau_A$ is the time between the first and last visits to the set A before time n). Hence, the constants involved in bound (5) depend on the choice of the atom and ideally the latter should be picked so as to minimize the bound on the right hand side of (5).

We also point out that in the case where the X_i 's are i.i.d. and f is bounded, by taking $A = \mathcal{X}$ (so that $\tau_A \equiv 1$), $M_1 > \sup_{x \in E} |f(x)|$ and $M_2 > 1$, the term on the right hand side of (5) (respectively, of (6)) boils down to a Bennett's type bound (respectively, to a Bernstein's type bound).

Proof : Treading in the steps of [3] (see also [4] and [13]), the proof is made in four stages as announced in §2.1: 1. truncating the r.v.'s as in [9], 2. partitioning the probability space according to all fashions for X to regenerate up to time n , 3. establishing an accurate exponential bound for sums of bounded i.i.d. 1-lattice random vectors and 4. summing the bounds previously obtained over all subsets of the partition.

Step 1: truncation. By the same kind of truncation trick as in [9], we shall show that the tail of $\sum_{j=1}^{l_n-1} S_j(f)$ may be bounded by an exponentially decreasing term plus a term related to the tail behavior of the $(S_j(f), s_j)$'s. Let M_1, M_2 be positive thresholds and consider the truncated r.v.'s,

$$\tilde{S}_j(f) = S_j(f) \mathbb{I}_{\{|S_j(f)| \leq M_1\}} \text{ and } \tilde{s}_j = s_j \mathbb{I}_{\{|s_j| \leq M_2\}}, \text{ for } j \geq 1.$$

The $\tilde{S}_j(f)$'s (respectively, the \tilde{s}_j 's) are i.i.d bounded random variables. As $l_n \leq n$ \mathbb{P}_v a.s., it follows from the union bound that for all positive x

$$\begin{aligned} \mathbb{P}_v \left(\sum_{j=1}^{l_n-1} S_j(f) \geq x \right) &\leq \mathbb{P}_v \left(\left\{ \sum_{j=1}^{l_n-1} \tilde{S}_j(f) \geq x \right\} \cap \{s_j = \tilde{s}_j, 1 \leq j \leq l_n - 1\} \right) \\ &\quad + (n-1) \mathbb{P}_A(\{|S_A(f)| > M_1\} \cup \{\tau_A > M_2\}). \end{aligned} \tag{7}$$

Again the last term on the right hand side of (7) may be bounded from assumptions on the speed of return time to A and on the tail of $S_A(f)$ under \mathbb{P}_A . We shall now deal with the first term which we denote by $P_{v,n}(x)$.

Step 2: partitioning. In order to be brought to the independent framework, we partition the event $\{l_n \geq 2\}$ according to all possible values for the first and last regeneration times, as well as for the total number of regeneration times up to time n . In this respect, consider the collection of events

$$\mathcal{U}_{r,l,m} = \left\{ \tau_A = r, \sum_{j=1}^m s_j = n - r - l, s_{m+1} > l \right\},$$

for $1 \leq r, l \leq n$ and $2 \leq m \leq n$. Combining the formula of total probability to the strong Markov property, we get

$$P_{\nu, n}(x) = \sum_{m=2}^n \sum_{r=1}^n \sum_{l=1}^n \mathbb{P}_{\nu}(\tau_A = r) \times p_{r, l, m}(x) \times \mathbb{P}_A(\tau_A > l), \quad (8)$$

where $p_{r, l, m}(x) = \mathbb{P}(\sum_{1 \leq j \leq m} \tilde{S}_j(f) \geq x, \sum_{1 \leq j \leq m} \{\tilde{s}_j - \mathbb{E}[\tilde{s}_j]\} = n - r - l - m\tilde{\alpha})$ and $\tilde{\alpha} = \mathbb{E}[\tilde{s}_j]$. Since $\mathbb{E}[\tilde{S}_j(f)] \leq 0$, we have $p_{r, l, m}(x) \leq \tilde{p}_{r, l, m}(x)$ for all r, l and m , where

$$\tilde{p}_{r, l, m}(x) = \mathbb{P}\left(\frac{1}{\tilde{\sigma}_f \sqrt{m}} \sum_{j=1}^m \{\tilde{S}_j(f) - \mathbb{E}[\tilde{S}_j(f)]\} \geq \frac{x}{\tilde{\sigma}_f \sqrt{m}}, \frac{1}{\tilde{\sigma}_A \sqrt{m}} \sum_{j=1}^m \{\tilde{s}_j - \mathbb{E}[\tilde{s}_j]\} = \lambda_{r, l, m}\right), \quad (9)$$

with $\lambda_{r, l, m} = (n - r - l - \tilde{\alpha}m)/(\tilde{\sigma}_A \sqrt{m})$. Hence all boils down to get an accurate bound for $p_{r, l, m}(x)$ using the fact that $(\tilde{S}_j(f), \tilde{s}_j)_{j \geq 1}$ is a sequence of bounded i.i.d. bivariate 1-lattice random vectors. With the notations previously set out, notice that $\tilde{\rho} = (\tilde{\sigma}_f \tilde{\sigma}_A)^{-1} \text{Cov}(\tilde{S}_j(f), \tilde{s}_j)$ for $j \geq 1$.

Step 3: exponential bound for 1-lattice random vectors. It is easy to get a crude exponential estimate for (9) by the standard Chernoff method (for instance see [20] and inequality (22) in §4.1). However such an estimate would not be precise enough to get a summable bound in m for (8) (preventing thus from reproducing the argument in [13] or [3] for instance). Indeed, under adequate assumptions, from limit theorems for k -lattice random vectors with minimum span h (see [7], [8] or Lemma 6.4 in [3]), $\tilde{p}_{r, l, m}(x)$ may be shown as asymptotically equivalent to $m^{-1/2} h \int_{t=x}^{\infty} \phi_{\Sigma}(t/\sqrt{m}, \lambda_{r, l, m}) dt$, denoting by ϕ_{Σ} the density of the bivariate gaussian distribution with mean zero and variance-covariance matrix $\Sigma = \text{Var}((\tilde{\sigma}_f^{-1} \tilde{S}_j(f), \tilde{\sigma}_A^{-1} \tilde{s}_j))$. As pointed out in §6.2.2 of [3], the factor $m^{-1/2}$ appearing in the latter quantity is of prime importance for obtaining bounds that are summable in m . In order to prove the exponential inequality required, we use the method proposed in [24] for improving exponential bounds in the independent setting, which is based on a refined use of the classical argument of the Bahadur-Rao Theorem. An exponential bound for sums of i.i.d. 1-lattice bounded random vectors is stated in the following lemma, of which proof is given in §3.1.

Lemma 2 (EXPONENTIAL INEQUALITIES FOR 1-LATTICE BOUNDED RANDOM VECTORS)
Let $(S_j^, L_j^*)_{1 \leq j \leq m}$ be i.i.d. centered and square integrable bivariate r.v.'s with covariance matrix $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. Assume further that the L_j^* 's are lattice r.v.'s with minimal span $h > 0$ and that there exists finite constants B_1 and B_2 such that $|S_j^*| \leq B_1$ and $|L_j^*| \leq B_2$ for $j = 1, \dots, m$. Set $B^2 = B_1^2 + B_2^2$, then there exists a universal constant c such that for all $m \geq 1$ and $y \geq 0$,*

$$\mathbb{P}(m^{-\frac{1}{2}} \sum_{j=1}^m S_j^* \geq y, m^{-\frac{1}{2}} \sum_{j=1}^m L_j^* = \lambda) \leq e^{\frac{3}{2}} \left(\frac{h}{2\pi\sqrt{m}} + 4c \frac{B}{m} \right) \exp \left\{ -\frac{y^2 + \lambda^2}{4B^2} \right\}. \quad (10)$$

Furthermore, we also have the following Bennett/Bernstein type inequalities,

$$\begin{aligned} \mathbb{P}(m^{-\frac{1}{2}} \sum_{j=1}^m S_j^* \geq y, m^{-\frac{1}{2}} \sum_{j=1}^m L_j^* = \lambda) &\leq e^{\frac{3}{2}} \left(\frac{h}{2\pi\sqrt{m}} + \frac{4cB}{m} \right) e^{-\frac{m(1+|\rho|)}{B^2} H\left(\frac{B\|(y,\lambda)\|}{(1+|\rho|)\sqrt{m}}\right)} \\ &\leq e^{\frac{3}{2}} \left(\frac{h}{2\pi\sqrt{m}} + \frac{4cB}{m} \right) \exp \left\{ -\frac{\|(y,\lambda)\|^2}{2(1+|\rho| + \frac{B\|(y,\lambda)\|}{3\sqrt{m}})} \right\}. \end{aligned}$$

Remark 3 An overestimated value of the constant c may be deduced by a careful examination of the proof of limit theorems stated in [7] and [8] (see §3.1 further). However we do not know at present what the best value for the constant c might be.

Going back to our problem, notice first that, as the L_j^* 's are lattice with minimal span $h = \tilde{\sigma}_A^{-1}$, lemma 2 applies to the sequence $\{(\tilde{S}_j(f) - \mathbb{E}[\tilde{S}_j(f)])/ \tilde{\sigma}_f, (\tilde{s}_j - \mathbb{E}[\tilde{s}_j])/ \tilde{\sigma}_A\}_{j \geq 1}$ with $\rho = \tilde{\rho}$, $B_1 = 2M_1/\tilde{\sigma}_f$ and $B_2 = 2M_2/\tilde{\sigma}_A$. Observing that $B = (B_1^2 + B_2^2)^{1/2} \leq 2\|M\|/\tilde{\sigma}$, this yields

$$\tilde{p}_{r,l,m}(x) \leq e^{\frac{3}{2}} \left(\frac{1}{2\pi\tilde{\sigma}_A\sqrt{m}} + 8c \frac{\|M\|}{\tilde{\sigma}m} \right) \exp \left\{ -\frac{m(1+|\tilde{\rho}|)\tilde{\sigma}^2}{\|M\|^2} H\left(\frac{\|M\| \cdot \|(x/(\tilde{\sigma}_f\sqrt{m}), \lambda_{r,l,m})\|}{(1+|\tilde{\rho}|)\tilde{\sigma}\sqrt{m}}\right) \right\}. \quad (11)$$

Recall that, for any $a > 0$, the function $x \mapsto H(a\sqrt{x})$ is concave on \mathbb{R}^+ . Writing

$$\|(x/(\tilde{\sigma}_f\sqrt{m}), \lambda_{r,l,m})\| = \left(\frac{1}{2} \left(\frac{x\sqrt{2}}{\tilde{\sigma}_f\sqrt{m}} \right)^2 + \frac{1}{2} (\lambda_{r,l,m}\sqrt{2})^2 \right)^{1/2},$$

by concavity arguments we get that

$$\tilde{p}_{r,l,m}(x) \leq e^{\frac{3}{2}} \left(\frac{1}{2\pi\tilde{\sigma}_A\sqrt{m}} + 8c \frac{\|M\|}{\tilde{\sigma}m} \right) e^{-\frac{m(1+|\tilde{\rho}|)\tilde{\sigma}^2}{2\|M\|^2} \left\{ H\left(\frac{\|M\|x\sqrt{2}}{(1+|\tilde{\rho}|)\tilde{\sigma}_f\tilde{\sigma}m}\right) + H\left(\frac{\|M\|\lambda_{r,l,m}\sqrt{2}}{(1+|\tilde{\rho}|)\tilde{\sigma}\sqrt{m}}\right) \right\}}. \quad (12)$$

As the function $x \mapsto \frac{H(x)}{x}$ is increasing on \mathbb{R}^+ , we have for any $m \leq n$

$$\begin{aligned} \exp \left\{ -\frac{m(1+|\tilde{\rho}|)\tilde{\sigma}^2}{2\|M\|^2} H\left(\frac{\|M\|x\sqrt{2}}{(1+|\tilde{\rho}|)\tilde{\sigma}_f\tilde{\sigma}m}\right) \right\} &\leq \exp \left\{ -\frac{n(1+|\tilde{\rho}|)\tilde{\sigma}^2}{2\|M\|^2} H\left(\frac{\|M\|x\sqrt{2}}{(1+|\tilde{\rho}|)\tilde{\sigma}_f\tilde{\sigma}n}\right) \right\} \\ &\leq \exp \left\{ -\frac{x^2\tilde{\sigma}/\tilde{\sigma}_f}{2(n(1+|\tilde{\rho}|)\tilde{\sigma}\tilde{\sigma}_f + \frac{x\|M\|\sqrt{2}}{3})} \right\}, \quad (13) \end{aligned}$$

the last bound following from the classical inequality $H(x) \geq \frac{x^2}{2(1+x/3)}$, $x \geq 0$.

Step 4: control of the sum. Now combining the previous bound to (8) and (12), we deduce that

$$P_{n,\nu}(x) \leq e^{3/2} \exp \left\{ -\frac{n(1+|\tilde{\rho}|)\tilde{\sigma}^2}{2\|\mathbf{M}\|^2} \text{H} \left(\frac{\|\mathbf{M}\|x\sqrt{2}}{(1+|\tilde{\rho}|)\tilde{\sigma}_f\tilde{\sigma}_n} \right) \right\} \Gamma_n, \quad (14)$$

where

$$\Gamma_n = \sum_{m=2}^n \sum_{r=1}^n \sum_{l=1}^n \mathbb{P}_\nu(\tau_A = r) \mathbb{P}_A(\tau_A > l) \left(\frac{c_1}{\sqrt{m}} + \frac{c_2}{m} \right) \gamma_{r,l,m},$$

with $c_1 = (2\pi\tilde{\sigma}_A)^{-1}$, $c_2 = 8c\|\mathbf{M}\|/\tilde{\sigma}$ and

$$\begin{aligned} \gamma_{r,l,m} &= \exp \left\{ -\frac{m(1+|\tilde{\rho}|)\tilde{\sigma}^2}{2\|\mathbf{M}\|^2} \text{H} \left(\frac{\|\mathbf{M}\|\lambda_{r,l,m}\sqrt{2}}{(1+|\tilde{\rho}|)\tilde{\sigma}\sqrt{m}} \right) \right\} \\ &\leq \exp \left\{ -\frac{\lambda_{r,l,m}^2}{2(1+|\tilde{\rho}| + \frac{|\lambda_{r,l,m}\|\mathbf{M}\|\sqrt{2}}{3\tilde{\sigma}\sqrt{m}})} \right\}. \end{aligned} \quad (15)$$

Recall that $\lambda_{r,l,m} = (n-r-l-\tilde{\alpha}m)/\tilde{\sigma}_A\sqrt{m}$ and consider the subdivision defined by the points

$$a_{n,m} = (n-\tilde{\alpha}m)/(\tilde{\sigma}_A\sqrt{m}), \text{ for } 1 \leq m \leq n.$$

In order to bound this sum, we split the latter into two parts, according to whether $r+l$ is less than $\tilde{\sigma}_A\sqrt{m}$ or not. Since $\gamma_{r,l,m} \leq 1$, by repeated use of Markov inequality, we get

$$\begin{aligned} \sum_{m=2}^n \sum_{r+l > \tilde{\sigma}_A\sqrt{m}} \mathbb{P}_\nu(\tau_A = r) \mathbb{P}_A(\tau_A > l) \left(\frac{c_1}{\sqrt{m}} + \frac{c_2}{m} \right) \gamma_{r,l,m} &\leq \sum_{m=1}^n \left(\frac{c_1}{\sqrt{m}} + \frac{c_2}{m} \right) \sum_{l \geq \tilde{\sigma}_A\sqrt{m}} \mathbb{P}_A(\tau_A > l) \\ &\leq \frac{\mathbb{E}_A[\tau_A^2]}{\tilde{\sigma}_A^2} \sum_{m=1}^n \left(\frac{c_1}{m^{3/2}} + \frac{c_2}{m^2} \right) \\ &\leq \frac{\mathbb{E}_A[\tau_A^2]}{\tilde{\sigma}_A^2} (3c_1 + c_2 \frac{\pi^2}{6}). \end{aligned} \quad (16)$$

From the identity $(a-b)^2 \geq a^2/2 - b^2$, it follows that $\lambda_{r,l,m}^2 \geq a_{n,m}^2/2 - ((r+l)/(\tilde{\sigma}_A\sqrt{m}))^2$ and, for $r+l \leq \tilde{\sigma}_A\sqrt{m}$,

$$\lambda_{r,l,m}^2 \geq (a_{n,m}^2/2 - 1)_+,$$

denoting by a_+ the positive part of any real number a . The term on the right hand side of (15) being decreasing in $\lambda_{r,l,m}$, we have

$$\gamma_{r,l,m} \leq \exp \left\{ -\frac{(a_{n,m}^2/2 - 1)_+}{2(1+|\tilde{\rho}| + \frac{(a_{n,m}^2/2 - 1)_+^{1/2}\|\mathbf{M}\|\sqrt{2}}{3\tilde{\sigma}\sqrt{m}})} \right\}. \quad (17)$$

Consider the set $A_n = \{m \in \{2, \dots, n\} : (a_{n,m}^2/2 - 1)_+^{1/2} \|\mathbf{M}\| \sqrt{2}/(3\tilde{\sigma}) \leq m^{1/2}(1 + |\tilde{\rho}|)\}$ of indexes m for which the "Bernstein type" bound (17) describes a gaussian tail behavior (the complement A_n^c corresponding to indexes m for which (17) provides a Poisson type estimate). On the one hand, if $m \in A_n$

$$\gamma_{r,l,m} \leq \exp \left\{ -\frac{a_{n,m}^2/2 - 1}{4(1 + |\tilde{\rho}|)} \right\},$$

and on the other hand, if $m \in A_n^c$,

$$\gamma_{r,l,m} \leq \exp \left\{ -\frac{3\tilde{\sigma}\sqrt{m}(a_{n,m}^2/2 - 1)_+^{1/2}}{4\|\mathbf{M}\|\sqrt{2}} \right\} \leq \exp \left\{ -\frac{9\tilde{\sigma}^2(1 + |\tilde{\rho}|)}{8\|\mathbf{M}\|^2} m \right\}.$$

It follows that

$$\sum_{m=1}^n \sum_{r+l \leq \tilde{\sigma}_A \sqrt{m}} \mathbb{P}_{\mathbf{v}}(\tau_A = r) \mathbb{P}_A(\tau_A > l) \left(\frac{c_1}{\sqrt{m}} + \frac{c_2}{m} \right) \gamma_{r,l,m} \leq \alpha(c_1 + c_2)(U_n + V_n),$$

where $U_n = \sum_{m \in A_n} m^{-1/2} \gamma_{r,l,m}$ and $V_n = \sum_{m \in A_n^c} m^{-1/2} \gamma_{r,l,m}$. As in step 4 of Theorem 5.1's proof in [3], we shall prove that U_n is bounded by a Riemann sum. Notice that for all $m \leq n$, $a_{n,m} - a_{n,m+1} = \tilde{\alpha} \tilde{\sigma}_A^{-1} m^{-1/2} + a_{n,m+1}((1 + m^{-1})^{1/2} - 1)$ and $a_{n,m} \geq a_{n,m+1}$. If $a_{n,m+1} \geq 0$, then $\tilde{\alpha} \tilde{\sigma}_A^{-1} m^{-1/2} \leq a_{n,m} - a_{n,m+1}$. On the other hand, if $a_{n,m+1} \leq 0$, we have $-a_{n,m+1} = (\tilde{\alpha}(m+1) - n) \tilde{\sigma}_A^{-1} m^{-1/2} \leq \tilde{\alpha} \tilde{\sigma}_A^{-1} m^{1/2}$ provided that $n \geq \tilde{\alpha}$. Hence, for any m such that $a_{n,m+1} \leq 0$,

$$\tilde{\alpha} \tilde{\sigma}_A^{-1} m^{-1/2} \leq a_{n,m} - a_{n,m+1} + 2^{-1} \tilde{\alpha} \tilde{\sigma}_A^{-1} m^{-1/2},$$

and consequently for any $m \geq 2$ and $n \geq \tilde{\alpha}$, we have

$$\tilde{\alpha} \tilde{\sigma}_A^{-1} m^{-1/2} \leq 2(a_{n,m} - a_{n,m+1}).$$

We thus obtain that

$$\begin{aligned} U_n &\leq 2\tilde{\alpha}^{-1} \tilde{\sigma}_A e^{c_3} \sum_{m=1}^n (a_{n,m} - a_{n,m+1}) \exp\{-c_3 a_{n,m}^2/2\} \\ &\leq 2\tilde{\alpha}^{-1} \tilde{\sigma}_A e^{c_3} \int_{x=-\infty}^{\infty} \exp(-c_3 x^2/2) dx = 2\tilde{\alpha}^{-1} \tilde{\sigma}_A \sqrt{2\pi/c_3} e^{c_3}, \end{aligned} \quad (18)$$

with $c_3 = 1/(4(1 + |\tilde{\rho}|)) \in (1/8, 1/4)$ (in particular, $e^{c_3}/\sqrt{c_3} \leq 2\sqrt{2}e^{1/4}$).

Besides, proceeding in a similar way, we get

$$V_n \leq \sum_{m \in A_n^c} m^{-1/2} \exp \left\{ -\frac{9(1 + |\tilde{\rho}|)\tilde{\sigma}^2}{8\|\mathbf{M}\|^2} m \right\}$$

$$\begin{aligned}
&\leq \int_0^\infty x^{-1/2} \exp\left\{-\frac{9(1+|\tilde{\rho}|)\tilde{\sigma}^2}{8\|\mathcal{M}\|^2}x\right\} dx \\
&= \frac{2\sqrt{2}\|\mathcal{M}\|\sqrt{\pi}}{3\tilde{\sigma}(1+|\tilde{\rho}|)^{1/2}} \leq \frac{2\sqrt{2}\|\mathcal{M}\|\sqrt{\pi}}{3\tilde{\sigma}}.
\end{aligned} \tag{19}$$

Now combining (16), (18) and (19), we obtain $\Gamma_n \leq c_4$ with

$$c_4 = \frac{\sigma_\lambda^2 + \alpha^2}{\tilde{\sigma}_\lambda^2} (3c_1 + c_2 \frac{\pi^2}{6}) + 2\tilde{\sigma}_\lambda (c_1 + c_2) e^{c_3} \frac{\alpha}{\tilde{\alpha}} \sqrt{\frac{2\pi}{c_3}} + (c_1 + c_2) \frac{2\sqrt{2}\|\mathcal{M}\|\alpha\sqrt{\pi}}{3\tilde{\sigma}}. \tag{20}$$

Using (14), this yields the tail bound (5) where $C_M = e^{3/2}c_4$ with c_4 given by (20). Notice that $C_M = O(\|\mathcal{M}\|^2)$ as $\|\mathcal{M}\| \rightarrow \infty$. And using (13) we obtain the tail bound (6) with the "Bernstein version" for the exponential term. \square .

3 Technical Details - Proof of Lemma 2

Set $\mathcal{S}_m = m^{-1/2} \sum_{i \leq n} S_j^*$ and $\mathcal{L}_m = m^{-1/2} \sum_{i \leq n} L_j^*$ for notational convenience. For $\mathbf{u} = (u_1, u_2) \in \mathbb{R}^+ \times \mathbb{R}$, let $\psi_m(\mathbf{u}) = \log \mathbb{E}[\exp\{\langle \mathbf{u}, (\mathcal{S}_m, \mathcal{L}_m) \rangle\}]$ be the log-Laplace of the random vector $(\mathcal{S}_m, \mathcal{L}_m)$ and denote by $\psi_{\mathbf{u},m}^{(1)}$ and $\psi_{\mathbf{u},m}^{(2)}$ its gradient and its hessian matrix. Consider now the probability measure $\mathbb{P}_{\mathbf{u},m}$ defined by the Esscher transformation $d\mathbb{P}_{\mathbf{u},m} = \exp\{\langle \mathbf{u}, (\mathcal{S}_m, \mathcal{L}_m) \rangle - \psi_m(\mathbf{u})\} d\mathbb{P}$. Expectation under $\mathbb{P}_{\mathbf{u},m}$ is denoted by $\mathbb{E}_{\mathbf{u},m}[\cdot]$ in what follows. By exponential change of probability measure, we get

$$\begin{aligned}
\mathbb{P}(\mathcal{S}_m \geq \mathbf{y}, \mathcal{L}_m = \lambda) &= \mathbb{E}_{\mathbf{u},m}[e^{\psi_m(\mathbf{u}) - \langle \mathbf{u}, (\mathcal{S}_m, \mathcal{L}_m) \rangle} \mathbb{I}_{\{\mathcal{S}_m \geq \mathbf{y}, \mathcal{L}_m = \lambda\}}] \\
&= e^{-\langle \mathbf{u}, (\mathbf{y}, \lambda) \rangle - \psi_m(\mathbf{u})} \mathbb{E}_{\mathbf{u},m}[e^{-\langle \mathbf{u}, (\mathcal{S}_m - \mathbf{y}, \mathcal{L}_m - \lambda) \rangle} \mathbb{I}_{\{\mathcal{S}_m \geq \mathbf{y}, \mathcal{L}_m = \lambda\}}].
\end{aligned}$$

Now choose $\mathbf{u} = \mathbf{u}^*$ such that $\psi_m^{(1)}(\mathbf{u}^*) = (\mathbf{y}, \lambda)$, that is

$$\mathbf{u}^* = \arg \sup_{\mathbf{u} \in \mathbb{R}_+ \times \mathbb{R}} \{\langle \mathbf{u}, (\mathbf{y}, \lambda) \rangle - \psi_m(\mathbf{u})\}.$$

And as $\mathbb{E}[e^{\langle \mathbf{u}, (\mathcal{S}_m, \mathcal{L}_m) \rangle}] = e^{\psi_m(\mathbf{u})}$, by differentiating one obtains

$$\mathbb{E}[e^{\langle \mathbf{u}^*, (\mathcal{S}_m, \mathcal{L}_m) \rangle} (\mathcal{S}_m, \mathcal{L}_m)] = \psi_m^{(1)}(\mathbf{u}^*) e^{\psi_m(\mathbf{u}^*)} = (\mathbf{y}, \lambda) e^{\psi_m(\mathbf{u}^*)}$$

yielding $\mathbb{E}_{\mathbf{u}^*,m}[(\mathcal{S}_m, \mathcal{L}_m)] = (\mathbf{y}, \lambda)$ and $\text{Var}_{\mathbf{u}^*,m}[(\mathcal{S}_m, \mathcal{L}_m)] = \psi_m^{(2)}(\mathbf{u}^*)$ in a similar fashion, denoting by $\text{Var}_{\mathbf{u}^*,m}[\cdot]$, the variance-covariance matrix under $\mathbb{P}_{\mathbf{u}^*,m}$. By integrating by parts combined with straightforward changes of variables, one obtains

$$\begin{aligned}
A_m(\mathbf{u}) &= \mathbb{E}_{\mathbf{u},m}[e^{-\langle \mathbf{u}, (\mathcal{S}_m - \mathbf{y}, \mathcal{L}_m - \lambda) \rangle} \mathbb{I}_{\{\mathcal{S}_m \geq \mathbf{y}, \mathcal{L}_m = \lambda\}}] \\
&= \int e^{-\langle \mathbf{u}, (\mathcal{S}_m - \mathbf{y}, \mathcal{L}_m - \lambda) \rangle} \mathbb{I}_{\{\mathcal{S}_m - \mathbf{y} \geq 0, \mathcal{L}_m - \lambda = 0\}} d\mathbb{P}_{\mathbf{u},m}
\end{aligned}$$

$$\begin{aligned}
&= \mathbb{P}_{\mathbf{u},m}(\mathcal{S}_m \geq \mathbf{y}, \mathcal{L}_m = \lambda) \\
&\quad - \mathbf{u}_1 \int_{\mathbf{y}}^{\infty} e^{-\mathbf{u}_1(s-\mathbf{y})} \mathbb{P}_{\mathbf{u},m}(\mathcal{S}_m - \mathbf{y} \geq s - \mathbf{y}, \mathcal{L}_m - \lambda = 0) ds \\
&\leq \mathbf{u}_1 \int_{s=0}^{\infty} e^{-\mathbf{u}_1 s} \mathbb{P}_{\mathbf{u},m}(\mathcal{S}_m - \mathbf{y} \geq 0, \mathcal{L}_m - \lambda = 0) \\
&\quad - \mathbb{P}_{\mathbf{u},m}(\mathcal{S}_m - \mathbf{y} \geq s, \mathcal{L}_m - \lambda = 0) ds.
\end{aligned}$$

Using the results in [7] and [8] or the local Berry-Esseen Bound proved in [4] (see Theorem 4), we know that there exists a constant $C_{\mathbf{u}}$ such that, uniformly in (\mathbf{y}, λ) ,

$$|\mathbb{P}_{\mathbf{u},m}(\mathcal{S}_m - \mathbf{y} \geq s, \mathcal{L}_m - \lambda = 0) - \frac{\mathbf{h}}{\sqrt{m}} \int_s^{\infty} \phi_{W_{\mathbf{u}}}(t, 0) dt| \leq C_{\mathbf{u}} m^{-1},$$

where $W_{\mathbf{u}} = \text{Var}_{\mathbf{u},m}[(\mathcal{S}_m, \mathcal{L}_m)]$ and $\phi_{W_{\mathbf{u}}}$ is the density of the bivariate gaussian distribution with covariance matrix $W_{\mathbf{u}}$. From [7] (see also [8]) a crude bound for $C_{\mathbf{u}}$ is given by $ck_3(\mathbb{P}_{\mathbf{u},1})$ where $k_3(\mathbb{P}_{\mathbf{u},1}) = \mathbb{E}_{\mathbf{u},1}[|X - \mathbb{E}_{\mathbf{u},1}[X]|^3] / \mathbb{E}_{\mathbf{u},1}[(X - \mathbb{E}_{\mathbf{u},m}[X])^2]^{3/2}$ and c is a universal constant. Notice that

$$\phi_{W_{\mathbf{u}}}(s, 0) = \frac{\exp\{-\frac{1}{2}s^2/\alpha_{\mathbf{u},m}^2\}}{\sqrt{2\pi\text{Var}_{\mathbf{u},m}(\mathcal{L}_m)}\sqrt{2\pi\alpha_{\mathbf{u},m}^2}},$$

with

$$\begin{aligned}
\alpha_{\mathbf{u},m} &= \text{Var}_{\mathbf{u},m}(\mathcal{S}_m)^{1/2}(1 - \rho_{\mathbf{u},m}^2)^{1/2} \\
\rho_{\mathbf{u},m}^2 &= \text{Cov}_{\mathbf{u},m}(\mathcal{S}_m, \mathcal{L}_m)^2 / \text{Var}_{\mathbf{u},m}(\mathcal{L}_m)\text{Var}_{\mathbf{u},m}(\mathcal{S}_m).
\end{aligned}$$

This yields for $\mathbf{u} = \mathbf{u}^*$,

$$\begin{aligned}
A_m(\mathbf{u}^*) &\leq \mathbf{u}_1^* \int_0^{\infty} e^{-\mathbf{u}_1^* s} \left(\frac{\mathbf{h}}{\sqrt{m}} \int_0^s \phi_{W_{\mathbf{u}^*}}(t, 0) dt + 2C_{\mathbf{u}^*} m^{-1} \right) ds \\
&= \frac{\mathbf{h}}{\sqrt{m}} \int_0^{\infty} e^{-\mathbf{u}_1^* s} \phi_{W_{\mathbf{u}^*}}(s, 0) ds + 2C_{\mathbf{u}^*} m^{-1} \\
&= \frac{\mathbf{h}}{\sqrt{m}} \frac{1}{\sqrt{2\pi\text{Var}_{\mathbf{u}^*,m}(\mathcal{L}_m)}} \frac{\bar{\Phi}(\mathbf{u}_1^* \alpha_{\mathbf{u},m})}{2\pi\phi(\mathbf{u}_1^* \alpha_{\mathbf{u},m})} + 2C_{\mathbf{u}^*} m^{-1},
\end{aligned}$$

where $\phi(x)$ denotes the density of the standard normal distribution and $\bar{\Phi}(x)$ its survivor function. Recall that for all $x > 0$ (see [24]),

$$\frac{1}{\sqrt{2\pi}(1+x)} \leq \frac{\bar{\Phi}(x)}{\sqrt{2\pi}\phi(x)} \leq \frac{1}{\sqrt{2\pi}\max(x, 1)},$$

which leads to

$$A_m(\mathbf{u}^*) \leq \frac{\mathbf{h}}{\sqrt{m}} \frac{1}{2\pi\text{Var}_{\mathbf{u}^*,m}(\mathcal{L}_m)^{1/2}} + 2C_{\mathbf{u}^*} m^{-1}. \quad (21)$$

The matter is now to find an upper bound for the quantity

$$L(\mathbf{u}^*) = e^{-\langle \mathbf{u}^*, (\mathbf{y}, \lambda) \rangle - \psi_m(\mathbf{u}^*)} = \inf_{\mathbf{u}} e^{-\langle \mathbf{u}, (\mathbf{y}, \lambda) \rangle - \psi_m(\mathbf{u})}.$$

As $\langle \mathbf{u}, (\mathcal{S}_m, \mathcal{L}_m) \rangle \leq 1 + \|\mathbf{u}\|^2 B^2$, we have $\mathbb{E}[e^{\langle \mathbf{u}, (\mathcal{S}_m, \mathcal{L}_m) \rangle}] \leq e \exp\{\|\mathbf{u}\|^2 B^2\}$ for all \mathbf{u} , so that

$$L(\mathbf{u}^*) \leq e \exp(-(\mathbf{y}^2 + \lambda^2)/4B^2). \quad (22)$$

It only remains to bound $k_3(\mathbb{P}_{\mathbf{u}^*, m})$ explicitly. Using (22), observe first that

$$\mathbb{E}_{\mathbf{u}^*, 1}[(\mathcal{S}_1 - \mathbb{E}_{\mathbf{u}^*, 1}[\mathcal{S}_1])^2] = \mathbb{E}_{\mathbf{u}^*, m}[(\mathcal{S}_m - \mathbb{E}_{\mathbf{u}^*, m}[\mathcal{S}_m])^2] \quad (23)$$

$$\begin{aligned} &= \int (s - \mathbf{y})^2 e^{\langle \mathbf{u}^*, (s, \mathbf{l}) \rangle - \psi_m(\mathbf{u}^*)} d\mathbb{P}(s, \mathbf{l}) \quad (24) \\ &\geq e^{-1} \int (s - \mathbf{y})^2 e^{(s^2 + \mathbf{l}^2)/4B^2} d\mathbb{P}(s, \mathbf{l}) \\ &\geq e^{-1} \int (s - \mathbf{y})^2 d\mathbb{P}(s, \mathbf{l}) = e^{-1}(1 + \mathbf{y}^2) \geq e^{-1}, \end{aligned}$$

and similarly $\text{Var}_{\mathbf{u}^*, m}(\mathcal{L}_m) \geq e^{-1}$. Moreover, since we have the bound

$$\mathbb{E}_{\mathbf{u}^*, 1}[|\mathcal{S}_1 - \mathbb{E}_{\mathbf{u}^*, 1}[\mathcal{S}_1]|^3] \leq 2B \text{Var}_{\mathbf{u}^*, 1}(\mathcal{S}_1),$$

and it follows from (23) that

$$k_3(\mathbb{P}_{\mathbf{u}^*, 1}) \leq 2e^{1/2}B. \quad (25)$$

The desired result follows by combining (21) with (22) and the estimate (25).

To obtain the Bernstein-Bennett type bound, we replace the crude bound (22) by a more refined one by using standard arguments. We have

$$\mathbb{E}[e^{\langle \mathbf{u}, (\mathcal{S}_m, \mathcal{L}_m) \rangle}] = \prod_{i=1}^m \mathbb{E}[\exp\{\frac{\mathbf{u}_1}{\sqrt{m}} S_i^* + \frac{\mathbf{u}_2}{\sqrt{m}} L_i^*\}].$$

Since $s \mapsto (e^s - 1 - s)/s^2$ is increasing on \mathbb{R}_+ and $|\frac{\mathbf{u}_1}{\sqrt{m}} S_i^* + \frac{\mathbf{u}_2}{\sqrt{m}} L_i^*| \leq m^{-1/2} \|\mathbf{u}\| B$, we have

$$\begin{aligned} \mathbb{E}[\exp\{\frac{\mathbf{u}_1}{\sqrt{m}} S_i^* + \frac{\mathbf{u}_2}{\sqrt{m}} L_i^*\}] &\leq \mathbb{E}[1 + \frac{\mathbf{u}_1}{\sqrt{m}} S_i^* + \frac{\mathbf{u}_2}{\sqrt{m}} L_i^*] \\ &\quad + \mathbb{E}[(\mathbf{u}_1 S_i^* + \mathbf{u}_2 L_i^*)^2] \frac{\exp\{m^{-1/2} \|\mathbf{u}\| B\} - 1 - m^{-1/2} \|\mathbf{u}\| B}{\|\mathbf{u}\|^2 B^2} \\ &\leq 1 + (1 + |\rho|) \frac{\exp\{m^{-1/2} \|\mathbf{u}\| B\} - 1 - m^{-1/2} \|\mathbf{u}\| B}{B^2}. \end{aligned}$$

The last inequality follows from the trivial inequality

$$\mathbb{E}[(\mathbf{u}_1 S_i^* + \mathbf{u}_2 L_i^*)^2] = \mathbf{u}_1^2 + 2\rho \mathbf{u}_1 \mathbf{u}_2 + \mathbf{u}_2^2$$

$$\leq u_1^2 + |\rho|(u_1^2 + u_2^2) + u_2^2.$$

Then, we obtain

$$\begin{aligned} \mathbb{E}[e^{\langle \mathbf{u}, (\mathcal{S}_m, \mathcal{L}_m) \rangle}] &\leq (1 + ((1 + |\rho|)(\exp\{m^{-1/2}\|\mathbf{u}\|B\} - 1 - m^{-1/2}\|\mathbf{u}\|B)/B^2))^m \\ &\leq \exp\{m(1 + |\rho|)(\exp\{m^{-1/2}\|\mathbf{u}\|B\} - 1 - m^{-1/2}\|\mathbf{u}\|B)/B^2\}. \end{aligned}$$

If we choose \mathbf{u} such that $\|\mathbf{u}\| = \frac{m^{1/2}}{B} \log(1 + m^{-1/2}(1 + |\rho|)^{-1}B\|(\mathbf{y}, \lambda)\|)$ and \mathbf{u} colinear to (\mathbf{y}, λ) , then we get

$$\begin{aligned} L(\mathbf{u}^*) &\leq e^{-\|\mathbf{u}\|\|(\mathbf{y}, \lambda)\| + m(1 + |\rho|)(\exp\{m^{-1/2}\|\mathbf{u}\|B\} - 1 - m^{-1/2}\|\mathbf{u}\|B)/B^2} \\ &= \exp\left\{-m \frac{(1 + |\rho|)}{B^2} H\left(\frac{B\|(\mathbf{y}, \lambda)\|}{m^{1/2}(1 + |\rho|)}\right)\right\}, \end{aligned}$$

where $H(x) = (1 + x)\ln(1 + x) - x$. The last inequality follows from the classical inequality

$$H(x) \geq \frac{x^2}{2(1 + x/3)}, \text{ for } x \geq 0.$$

References

- [1] Bennett, S., (1962). Probability inequalities for a sum of independent random variables. *J. Amer. Statist. Assoc.*, 57, 33-45.
- [2] Bernstein, S. N., (1964). *On a modification of Chebyshev's inequality and on the error in Laplace formula*. Collected Works, Vol. 4, 71-80. Izd-vo 'Nauka', Moscow (in Russian).
- [3] Bertail, P., Cl emen on, S. (2004). Edgeworth expansions for suitably normalized sample mean statistics of atomic Markov chains. *Prob. Th. Rel. Fields*, 130, 388-414.
- [4] Bolthausen, E. (1980). The Berry-Esseen Theorem for Functionals of Discrete Markov Chains. *Z. Wahr. Verw. Gebiete*, 54, 59-73.
- [5] Borovkov, A.A., (1968). Some inequalities for sums of multidimensional random variables. *Th. Prob. Applications*, 13, 156-160.
- [6] Doukhan, P. (1994). *Mixing: properties and examples*. Lecture Notes in Statistics 85, Springer-Verlag.
- [7] Dubinskaite, J. (1982). Limit Theorems in \mathbb{R}^k I. *Lithuanian Math. J.*, 22, 129-140.

- [8] Dubinskaite, J. (1984). Limit Theorems in \mathbb{R}^k II-III. *Lithuanian Math. J.*, **24**, 256-265, 325-334.
- [9] Fuk, D.H., Nagaev, S.V. (1971). Probability Inequalities for Sums of Independent Random Variables. *Theor. Proba. Appl.*, Vol. **16**, No. 4, 643-660.
- [10] Glynn, P.W., Ormoneit, D. (2002). Hoeffding's inequality for uniformly ergodic Markov chains. *Stat. Proba. Lett.*, **56**, No. 2, 143-146.
Hoeffding, W., (1963). *Probability inequalities for sums of bounded random variables*. J. Amer. Statist. Assoc., **58**, 13-30.
- [11] Léon, C.A., Perron, F. (2004). Optimal Hoeffding Bounds for Discrete Reversible Markov Chains. *Ann. Appl. Probab.*, **14**, No. 2, 958-970.
- [12] Lézaud, P. (1998). Chernoff-type bound for finite Markov chains. *Ann. Appl. Probab.*, **8**, No. 3, 849-867.
- [13] Malinovskii, V. K. (1987). Limit theorems for Harris Markov chains I. Theory Prob. Appl., **31**, 269-285.
- [14] Marton, K. (1996). A Measure of Concentration Inequality for Contracting Markov Chains. *Geom. Funct. Analysis*, **6**, No. 3, 557-571.
- [15] Meyn, S.P., Tweedie, R.L. (1996). *Markov chains and stochastic stability*. Springer.
- [16] Nagaev, S.V. (1979). Large Deviations of Sums of Independent Random Variables. *Ann. Proba.*, Vol. **7**, No. 5, 145-789.
- [17] Nummelin, E. (1978). A splitting technique for Harris recurrent chains, Z. Wahrsch. Verw. Gebiete, **43**, 309-318.
- [18] Nummelin, E. (1984). *General Irreducible Markov Chains and Non-negative Operators*. Cambridge University Press, Cambridge, England.
- [19] Petrov, V.V. (1995). *Limit Theorems of Probability Theory. Sequences of Independent Random Variables*. Oxford University Press.
- [20] Prokhorov, Y.V., (1968). *An extension of S.N. Bernstein's inequalities to multi-dimensional distributions*. Theory Prob. Applications, **13**, 260-267.
- [21] Rio, E. (2000). *Théorie asymptotique des processus aléatoires faiblement dépendants*. Springer.
- [22] Samson, P.M. (2000). Concentration of Measure Inequalities for Markov Chains and Φ -Mixing Processes. *Ann. Probab.*, **28**, No. 1, 416-461.

- [23] Smith, W.L. (1955). Regenerative stochastic processes. *Proc. Royal Stat. Soc.*, A, 232, 6-31.
- [24] Talagrand, M. (1995). The missing factor in Hoeffding's inequalities. *Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques*, 31, No. 4, 689-70
- [25] Thorisson, H. (2000). *Coupling, Stationarity and Regeneration*. Springer.