Chapter 2
Concentration inequalities for sums

2.1 Bernstein's inequalities

Sergei Bernstein [6] was at the beginning of exponential inequalities for sums of independent random variables. This section deals with Bernstein's type inequalities. It is divided into three subsections. In the first one, we focus our attention on the one-sided inequality of Bernstein [6] as well as on improvements of this inequality. The second one is devoted to two-sided versions of Bernstein's inequality. In the last one, we give new inequalities with a smaller second term.

2.1.1 One-sided inequalities

The main result of this subsection is the theorem below, which collects different versions and improvements of Bernstein's one-sided inequality.

Theorem 2.1. *Let $X_1, \ldots, X_n$ be a finite sequence of independent random variables with finite variances. Denote*

$$S_n = X_1 + \cdots + X_n, \qquad \mathscr{V}_n = \mathbb{E}[X_1^2] + \cdots + \mathbb{E}[X_n^2],$$
$$v_n = \frac{\mathscr{V}_n}{n}. \tag{2.1}$$

Assume that $\mathbb{E}[S_n] = 0$ and that there exists some positive constant c such that, for any integer $p \geqslant 3$,

$$\sum_{k=1}^{n} \mathbb{E}\big[(\max(0, X_k))^p\big] \leqslant \frac{p!c^{p-2}}{2} \mathscr{V}_n. \tag{2.2}$$

B. Bercu et al., *Concentration Inequalities for Sums and Martingales*,
SpringerBriefs in Mathematics, DOI 10.1007/978-3-319-22099-4_2

Then, for any positive x,

$$\mathbb{P}(S_n \geqslant nx) \leqslant \Big(1+\frac{x^2}{2(v_n+cx)}\Big)^n \exp\Big(-\frac{nx^2}{v_n+cx}\Big) \tag{2.3}$$

$$\leqslant \exp\Big(-\frac{nx^2}{2(v_n+cx)}\Big). \tag{2.4}$$

In addition, we also have, for any positive x,

$$\mathbb{P}(S_n \geqslant nx) \leqslant \exp\Big(-\frac{nx^2}{v_n+cx+\sqrt{v_n(v_n+2cx)}}\Big) \tag{2.5}$$

and

$$\mathbb{P}\big(S_n > n(cx+\sqrt{2v_nx})\big) \leqslant \exp(-nx). \tag{2.6}$$

Remark 2.2. It is not necessary to assume that the random variables $X_1,\ldots,X_n$ are centered. We only have to suppose that $\mathbb{E}[S_n]=0$. In the centered case, $\mathscr{V}_n$ coincides with $V_n=\mathrm{Var}(S_n)$. Otherwise, $\mathscr{V}_n$ is obviously larger than V_n.

Remark 2.3. Condition (2.2), given in Rio [22], is weaker than the standard Bernstein's condition, which says that for any $1\leqslant k\leqslant n$ and for any integer $p\geqslant 3$,

$$\mathbb{E}\big[|X_k|^p\big] \leqslant \frac{p!c^{p-2}}{2}\,\mathbb{E}[X_k^2].$$

For example, condition (2.2) allows to consider random variables with heavier tails on the left. Bernstein [6] proved (2.4) under the above condition.

Remark 2.4. The optimal constant c^* in Theorem 2.1 is the smallest positive real c such that condition (2.2) is satisfied. If the random variables $X_1,\ldots,X_n$ are such that, for all $1\leqslant k\leqslant n$, $X_k\leqslant b$ almost surely for some positive constant b, then one can prove that $c^*\leqslant b/3$. In the forthcoming sections, we will give more efficient inequalities for random variables bounded from above.

Remark 2.5. Inequality (2.5) was obtained by Bennett [2] and it was called first improvement of Bernstein's inequality. Nevertheless, this result is still suboptimal. One can observe that (2.5) and (2.6) are equivalent. In the case $v_n\geqslant c^2$, inequality (2.5) is less efficient than (2.3) as shown in Figure 2.1 as well as in Figure 2.2 which compare the rate functions

$$\varphi(x)=\frac{x^2}{v_n+cx}-\log\Big(1+\frac{x^2}{2(v_n+cx)}\Big)$$

$$\Phi(x)=\frac{x^2}{v_n+cx+\sqrt{v_n(v_n+2cx)}} \quad \text{and} \quad \Psi(x)=\frac{x^2}{2(v_n+cx)}$$

associated, respectively, to the new Bernstein's inequality, Bennett's and Bernstein's inequalities, in the particular cases $v_n=1$, $c=1$, and $v_n=1$, $c=1/2$, respectively. Note also that (2.4) is equivalent to the reverse inequality

$$\mathbb{P}\left(S_n \geqslant n(cx + \sqrt{2v_n x + (cx)^2})\right) \leqslant \exp(-nx),$$

which is less efficient than (2.6).

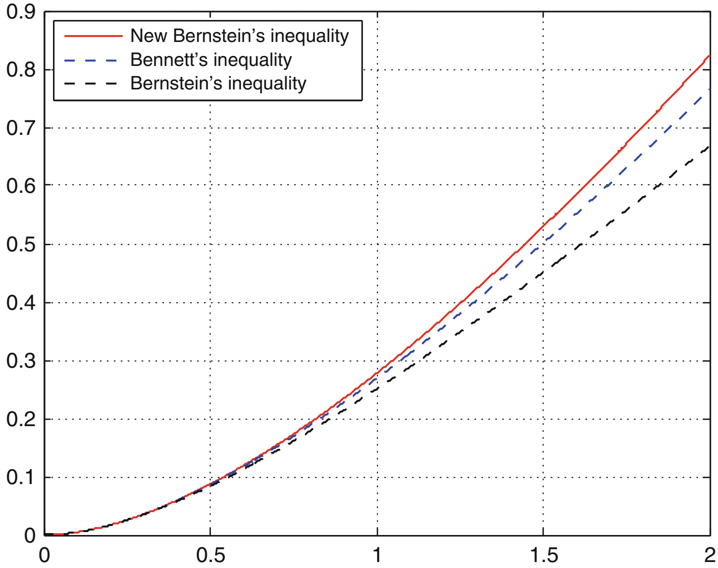

Fig. 2.1 Comparisons in Bernstein's inequalities in the particular case $v_n = 1$ and $c = 1$.

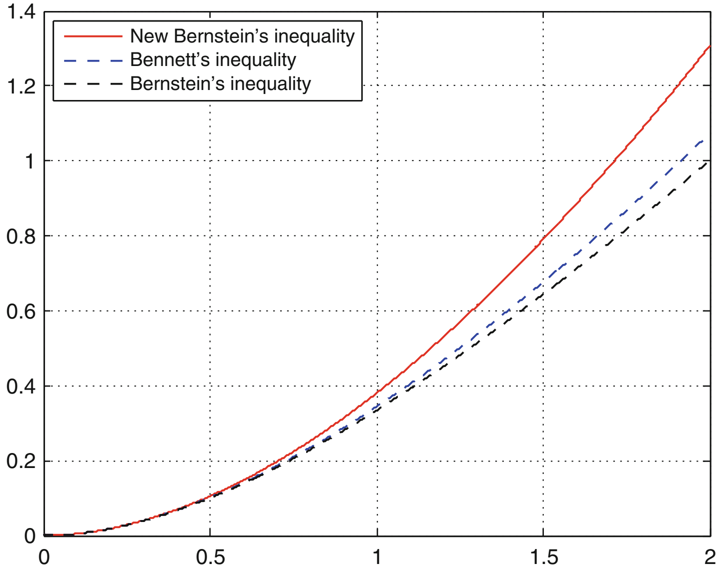

Fig. 2.2 Comparisons in Bernstein's inequalities in the particular case $v_n = 1$ and $c = 1/2$.

The outline of the proof of Theorem 2.1 is as follows. Denote by ℓ an upper bound of the normalized cumulant generating function of S_n, that is, for all $t \geqslant 0$,

$$\frac{1}{n} \log \mathbb{E}[\exp(tS_n)] \leqslant \ell(t).$$

We immediately obtain from Markov's inequality that, for all $t \geqslant 0$ and $x \geqslant 0$,

$$\mathbb{P}(S_n \geqslant nx) \leqslant \exp(-n(tx - \ell(t))), \tag{2.7}$$

which leads to

$$\mathbb{P}(S_n \geqslant nx) \leqslant \exp(-n\ell^*(x)) \quad \text{where} \quad \ell^*(x) = \sup_{t \geqslant 0}(xt - \ell(t)). \tag{2.8}$$

Consequently, in order to establish a concentration inequality for S_n, it only remains to calculate the Legendre-Fenchel transform ℓ^* of ℓ. In many situations, the calculation of ℓ^* could be somehow complicated. However, a sharp upper bound for ℓ^* also leads to concentration inequalities. Another strategy is to realize that for any positive x,

$$\mathbb{P}\Big(S_n > n \inf_{t>0}\Big(\frac{\ell(t)+x}{t}\Big)\Big) = \sup_{t>0} \mathbb{P}\Big(S_n > n\Big(\frac{\ell(t)+x}{t}\Big)\Big) \leqslant \exp(-nx), \tag{2.9}$$

which gives an upper bound for the quantile function of S_n in a more direct way. The proof of Theorem 2.1 relies on two elementary lemmas. The first one, due to Hoeffding [11], allows us to replace the initial random variables by independent random variables with the same distribution.

Lemma 2.6. *Let $X_1, \ldots X_n$ be a finite sequence of independent random variables and denote $S_n = X_1 + \cdots + X_n$. Then, for any real t,*

$$\frac{1}{n} \log \mathbb{E}[\exp(tS_n)] \leqslant \ell(t) \quad \textit{where} \quad \ell(t) = \log\Big(\frac{1}{n}\sum_{k=1}^{n} \mathbb{E}[\exp(tX_k)]\Big). \tag{2.10}$$

Proof. It follows from the independence of the random variables $X_1, \ldots X_n$ that for any real t,

$$\log \mathbb{E}[\exp(tS_n)] = \sum_{k=1}^{n} \log \mathbb{E}[\exp(tX_k)].$$

The concavity of the logarithm function clearly leads to

$$\frac{1}{n} \log \mathbb{E}[\exp(tS_n)] \leqslant \log\Big(\frac{1}{n}\sum_{k=1}^{n} \mathbb{E}[\exp(tX_k)]\Big) = \ell(t),$$

which achieves the proof of Lemma 2.6. $\square$

The second lemma, which implies (2.9), will allow us to give a nice proof of the first improvement of Bennett.

Lemma 2.7. *Let X be a real valued random variable with a finite Laplace transform on a right neighborhood of the origin. Denote by L_X the logarithm of the Laplace transform of X. Then, for any positive real number x,*

$$\mathbb{P}\Big(X > \inf_{t>0}\Big(\frac{L_X(t)+x}{t}\Big)\Big) \leqslant \exp(-x). \tag{2.11}$$

Proof. Let t be any positive real number such that $L_X(t) < \infty$. We immediately deduce from Markov's inequality that for any positive real number x,

$$\mathbb{P}\Big(X \geqslant \Big(\frac{L_X(t)+x}{t}\Big)\Big) = \mathbb{P}\Big(\exp(tX) \geqslant \exp(L_X(t)+x)\Big) \leqslant \exp(-x). \tag{2.12}$$

Hence, we obtain (2.11) by taking the infimum over all positive real numbers t. □

Proof of Theorem 2.1. We are now in position to prove Theorem 2.1. First of all, one can observe that for any $x \leqslant 0$,

$$\exp(x) \leqslant 1 + x + \frac{x^2}{2},$$

which ensures that for any real number x,

$$\exp(x) \leqslant 1 + x + \frac{x^2}{2} + \sum_{p=3}^{\infty} \frac{(x_+)^p}{p!}$$

where $x_+ = \max(0,x)$ stands for the positive part of x. It follows from the monotone convergence theorem that for all $1 \leqslant k \leqslant n$ and for any positive t,

$$\mathbb{E}[\exp(tX_k)] \leqslant 1 + t\mathbb{E}[X_k] + \frac{t^2\mathbb{E}[X_k^2]}{2} + \sum_{p=3}^{\infty} \frac{t^p\mathbb{E}[(\max(0,X_k))^p]}{p!}. \tag{2.13}$$

Consequently, as $\mathbb{E}[S_n] = 0$, we deduce from Bernstein's condition (2.2) that

$$\sum_{k=1}^{n} \mathbb{E}[\exp(tX_k)] \leqslant n + \frac{\mathscr{V}_n}{2} \sum_{p=2}^{\infty} c^{p-2}t^p.$$

Hence, as soon as $0 < tc < 1$,

$$\exp(\ell(t)) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}[\exp(tX_k)] \leqslant 1 + \frac{v_n t^2}{2(1-tc)}.$$

Therefore, we find from Lemma 2.6 that for any positive t such that $0 < tc < 1$,

$$\frac{1}{n}\log\mathbb{E}[\exp(tS_n)] \leqslant \ell_{v_n}(t) \quad \text{where} \quad \ell_{v_n}(t) = \log\Big(1+\frac{v_n t^2}{2(1-tc)}\Big). \tag{2.14}$$

Hereafter, if we choose $t_n(x) = x/(v_n + cx)$ with $x > 0$, we obtain from the definition of the Legendre-Fenchel transform that

$$\begin{aligned}\ell^*_{v_n}(x) \geqslant xt_n(x) - \ell_{v_n}(t_n(x)) &= \frac{x^2}{v_n+cx} - \log\Big(1+\frac{v_n t_n^2(x)}{2(1-t_n(x)c)}\Big), \\ &= \frac{x^2}{v_n+cx} - \log\Big(1+\frac{x^2}{2(v_n+cx)}\Big).\end{aligned} \tag{2.15}$$

Then, (2.3) and (2.4) immediately follow from (2.8) and (2.15). Finally, it only remains to prove (2.6). Applying Lemma 2.7 to S_n, we obtain that

$$\mathbb{P}\Big(S_n > n \inf_{tc\in]0,1[}\Big(\frac{\ell_{v_n}(t)+x}{t}\Big)\Big) \leqslant \exp(-nx). \tag{2.16}$$

Now, it follows from (2.14) and the elementary inequality $\log(1+x) \leqslant x$ that

$$\inf_{tc\in]0,1[}\Big(\frac{\ell_{v_n}(t)+x}{t}\Big) \leqslant \inf_{tc\in]0,1[}\Big(\frac{v_n t}{2(1-tc)} + \frac{x}{t}\Big) = cx + \sqrt{2xv_n} \tag{2.17}$$

where the infimum of the right-hand side is given by the optimal value

$$t = \frac{\sqrt{2x}}{\sqrt{v_n}+c\sqrt{2x}}.$$

Finally, we clearly deduce (2.6) from (2.16) and (2.17), which completes the proof of Theorem 2.1. $\square$

Example 2.8. Let $\varepsilon_1, \ldots, \varepsilon_n$ be a finite sequence of independent random variables sharing the same Exponential $\mathscr{E}(\lambda)$ distribution with $\lambda > 0$. Let $a = (a_1, \ldots, a_n)$ be a vector with positive real components. For all $1 \leqslant k \leqslant n$, denote

$$X_k = a_k\Big(\varepsilon_k - \frac{1}{\lambda}\Big).$$

One can easily check that

$$\mathscr{V}_n = \sum_{k=1}^n \mathbb{E}[X_k^2] = \frac{1}{\lambda^2}\sum_{k=1}^n a_k^2 = \frac{\|a\|_2^2}{\lambda^2}.$$

In addition, it is not hard to see that condition (2.2) holds true with

$$c = \frac{1}{\lambda}\max(a_1, \ldots, a_n) = \frac{\|a\|_\infty}{\lambda}.$$

Therefore, it follows from (2.6) that for any positive x,

$$\mathbb{P}\big(S_n > \|a\|_2\sqrt{2x} + \lambda\,\|a\|_\infty x\big) \leqslant \exp(-\lambda^2 x).$$

In Section 2.7, we will give a more efficient inequality for sums of random variables with exponential distributions.

2.1.2 Two-sided inequalities

In this subsection, we provide two-sided versions of Bernstein's inequality. The main difference between the two-sided version and the previous one-sided version lies in the assumption on the random variables. More precisely, we shall consider a condition involving the algebraic moments of the random variables.

Theorem 2.9. *Let $X_1, \ldots, X_n$ be a finite sequence of independent random variables with finite algebraic moments at any order. Let S_n, $\mathscr{V}_n$, and v_n be defined as in (2.1). Assume that $\mathbb{E}[S_n] = 0$ and that there exists some positive constant c such that, for any integer $p \geqslant 3$,*

$$\Big|\sum_{k=1}^{n} \mathbb{E}\big[X_k^p\big]\Big| \leqslant \frac{p!c^{p-2}}{2}\,\mathscr{V}_n. \tag{2.18}$$

Then, for any positive x,

$$\mathbb{P}(|S_n| \geqslant nx) \leqslant 2\Big(1 + \frac{x^2}{2(v_n + cx)}\Big)^n \exp\Big(-\frac{nx^2}{v_n + cx}\Big) \tag{2.19}$$

$$\leqslant 2\exp\Big(-\frac{nx^2}{2(v_n + cx)}\Big). \tag{2.20}$$

In addition, we also have, for any positive x,

$$\mathbb{P}\big(|S_n| > n(cx + \sqrt{2v_n x})\big) \leqslant 2\exp(-nx). \tag{2.21}$$

Remark 2.10. Obviously, the one-sided bounds for S_n and $-S_n$ hold with the constant one instead of the constant two. One can observe that Condition (2.18) slightly differs from condition (2.2) applied to the two random finite sequences $X_1, \ldots, X_n$ and $-X_1, \ldots, -X_n$. To be more precise, (2.18) is weaker than the two-sided version of (2.2) for odd integers p and stronger for even integers p.

Proof. For any integer $p \geqslant 2$, denote

$$A_{p,n} = \mathbb{E}[X_1^p] + \cdots + \mathbb{E}[X_n^p]. \tag{2.22}$$

It follows from condition (2.18) that for any real t such that $|t|c < 1$,

$$\sum_{k=1}^{n} \mathbb{E}[\cosh(tX_k)] = \sum_{k=1}^{n} \sum_{p=0}^{\infty} \frac{1}{(2p)!} \mathbb{E}[(tX_k)^{2p}] = n + \sum_{p=1}^{\infty} \frac{A_{2p,n}}{(2p)!} t^{2p},$$
$$\leqslant n + \frac{A_{2,n}}{2} \frac{t^2}{1 - c^2t^2} < \infty.$$

Consequently the random variables $|X_1|, \ldots, |X_n|$ have finite Laplace transforms on the interval $[0, 1/c[$. As $\mathbb{E}[S_n] = 0$, it ensures that for any real t such that $|t|c < 1$,

$$\sum_{k=1}^{n} \mathbb{E}[\exp(tX_k)] = n + \sum_{p=2}^{\infty} \frac{A_{p,n}}{p!} t^p. \tag{2.23}$$

Furthermore, as $A_{2,n} = \mathscr{V}_n$, we can deduce from condition (2.18) that for any real t such that $|t|c < 1$,

$$\Big| \sum_{p=2}^{\infty} \frac{A_{p,n}}{p!} t^p \Big| \leqslant \mathscr{V}_n \frac{t^2}{1 - |tc|}. \tag{2.24}$$

Therefore, we obtain (2.23) and (2.24) and that for any real t such that $|t|c < 1$,

$$\log\Big(\frac{1}{n} \sum_{k=1}^{n} \mathbb{E}[\exp(tX_k)]\Big) \leqslant \log\Big(1 + \frac{v_n t^2}{2(1 - |tc|)}\Big) \leqslant \frac{v_n t^2}{2(1 - |tc|)}.$$

Finally, the rest of the proof is left to the reader inasmuch as it follows the same lines as the proof of Theorem 2.1. □

2.1.3 About the second term in Bernstein's inequality

In this subsection, we are interested in the second term in Bernstein's type inequalities. Recall that under assumption (2.2), the first improvement of Bennett is given, for any positive x, by

$$\mathbb{P}(S_n \geqslant nx) \leqslant \exp\big(-ng_n(x)\big) \quad \text{where} \quad g_n(x) = \frac{x^2}{v_n + cx + \sqrt{v_n^2 + 2cv_n x}}. \tag{2.25}$$

As x goes to zero, the rate function g_n has the asymptotic expansion

$$g_n(x) = \frac{x^2}{2v_n + 2cx + O(x^2)} = \frac{x^2}{2v_n} - \frac{cx^3}{2v_n^2} + O(x^4).$$

Consequently, the two first terms in the expansion of g_n into powers of x are the same as in the initial inequality of Bernstein. We shall now provide an inequality with a smaller second term.

Theorem 2.11. *Let $X_1, \ldots, X_n$ be a finite sequence of independent random variables with finite variances. Let S_n and v_n be defined as in (2.1). Assume that $\mathbb{E}[S_n] = 0$ and suppose that there exists some positive constant c such that*

$$a_n = \sup_{p \geqslant 3} \Big(c^{3-p} \frac{2}{p!} \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}\big[(\max(0, X_k))^p\big] \Big) < \infty. \tag{2.26}$$

For any positive constant c satisfying the above condition, denote

$$f_n(x) = \frac{x^2}{v_n + cx + \sqrt{(v_n - cx)^2 + 4a_n x}}. \tag{2.27}$$

Then, for any positive x,

$$\mathbb{P}(S_n \geqslant nx) \leqslant \big(1 + f_n(x)\big)^n \exp\big(-2nf_n(x)\big) \leqslant \exp\big(-nf_n(x)\big). \tag{2.28}$$

Remark 2.12. Under assumption (2.2), we clearly have $a_n \leqslant cv_n$. Then, for any positive x, $\sqrt{(v_n - cx)^2 + 4a_n x} \leqslant v_n + cx$, which immediately leads to

$$f_n(x) \geqslant \frac{x^2}{2(v_n + cx)}.$$

Hence, (2.28) is sharper than the first improvement of Bennett (2.25). Let us now give the asymptotic expansion of f_n. As x goes to zero, it follows from straightforward calculation that

$$f_n(x) = \frac{x^2}{2v_n + 2(a_n/v_n)x + O(x^2)} = \frac{x^2}{2v_n} - \frac{a_n x^3}{2v_n^3} + O(x^4).$$

Consequently, under assumption (2.2), the second term in the expansion of f_n is greater than the second term in the expansion of g_n, which also shows that (2.28) is sharper than (2.25).

Remark 2.13. Denote

$$\alpha_n = \frac{1}{3n} \sum_{k=1}^{n} \mathbb{E}\big[(\max(0, X_k))^3\big].$$

Obviously $a_n \geqslant \alpha_n$. Now, under the assumptions of Theorem 2.11, $a_n = \alpha_n$ if c is large enough. Hence it seems convenient to choose the smallest c such that $a_n = \alpha_n$ in Theorem 2.11.

Remark 2.14. The function f_n can easily be compared with the rate function g_n given in (2.25). By definition of f_n, $f_n(x) > g_n(x)$ iff $(v_n - cx)^2 + 4a_n x < v_n^2 + 2cv_n x$, which is equivalent to $x < 4(cv_n - a_n)$. Under this condition, the right-hand side term in (2.28) is smaller than the right-hand side term in (2.25).

Proof. We deduce from (2.13) and the definition of a_n that for any positive t such that $0 < tc < 1$,

$$\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[\exp(tX_k)] \leqslant 1 + \frac{v_n}{2}t^2 + \frac{a_n}{2(1-tc)}t^3.$$

Therefore, we find from Lemma 2.6 that for any positive t such that $0 < tc < 1$,

$$\frac{1}{n}\log\mathbb{E}[\exp(tS_n)] \leqslant \log(1+\psi_n(t)) \quad \text{where} \quad \psi_n(t) = \frac{v_n}{2}t^2 + \frac{a_n}{2(1-tc)}t^3. \tag{2.29}$$

Hence, we immediately obtain from (2.8) and the above upper bound that, for any positive x and for any positive t such that $0 < tc < 1$,

$$\mathbb{P}(S_n \geqslant nx) \leqslant (1+\psi_n(t))^n \exp(-ntx). \tag{2.30}$$

Hereafter, we can choose t in such a way that $\psi_n(t) = tx/2$, which is equivalent to the second order equation

$$(v_n c - a_n)t^2 - (v_n + cx)t + x = 0.$$

If $a_n = v_n c$, then the above equation has a unique solution $t = x/(v_n + cx)$, which clearly belongs to the interval $]0,1/c[$. In that case, $(v_n - cx)^2 + 4a_n x = (v_n + cx)^2$ and $f_n(x) = tx/2$. Otherwise, the above equation has two real solutions. The solution which lies in the interval $]0,1/c[$ is

$$t = \frac{v_n + cx - \sqrt{(v_n - cx)^2 + 4a_n x}}{2(v_n c - a_n)} = \frac{2x}{v_n + cx + \sqrt{(v_n - cx)^2 + 4a_n x}}.$$

Hence, once again, we find that $f_n(x) = tx/2$. Finally, we obtain (2.28) by choosing this value of t in (2.30). □

Example 2.15. Let $\varepsilon_1, \dots, \varepsilon_n$ be a finite sequence of independent random variables sharing the same Exponential $\mathscr{E}(1)$ distribution. In addition, let $B_1, \dots, B_n$ be a finite sequence of independent random variables with Bernoulli $\mathscr{B}(1/4)$ distribution. Assume that these two sequences are mutually independent. For all $1 \leqslant k \leqslant n$, denote

$$X_k = B_k\varepsilon_k - \frac{1}{6}\Big(1 - B_k\Big)\varepsilon_k^2.$$

One can easily check that, for all $1 \leqslant k \leqslant n$, $\mathbb{E}[X_k] = 0$, $\mathbb{E}[X_k^2] = 1 = v_n$, and $\mathbb{E}[(\max(0,X_k))^p] = p!/4$. Taking $c = 1$, we obtain from the definition (2.26) of a_n that $a_n = 1/2$. Then, it follows from Theorem 2.11 that, for any positive x,

$$\mathbb{P}(S_n \geqslant nx) \leqslant \big(1 + f_n(x)\big)^n \exp\big(-2nf_n(x)\big)$$

where

$$f_n(x) = \frac{x^2}{1+x+\sqrt{1+x^2}}.$$

Figure 2.3 below compares the rate function

$$\Psi(x) = \frac{2x^2}{1+x+\sqrt{1+x^2}} - \log\Big(1+\frac{x^2}{1+x+\sqrt{1+x^2}}\Big)$$

appearing here with the two rate functions of Figure 2.1. One can realize in this example that inequality (2.28) is more efficient than (2.5) and (2.3).

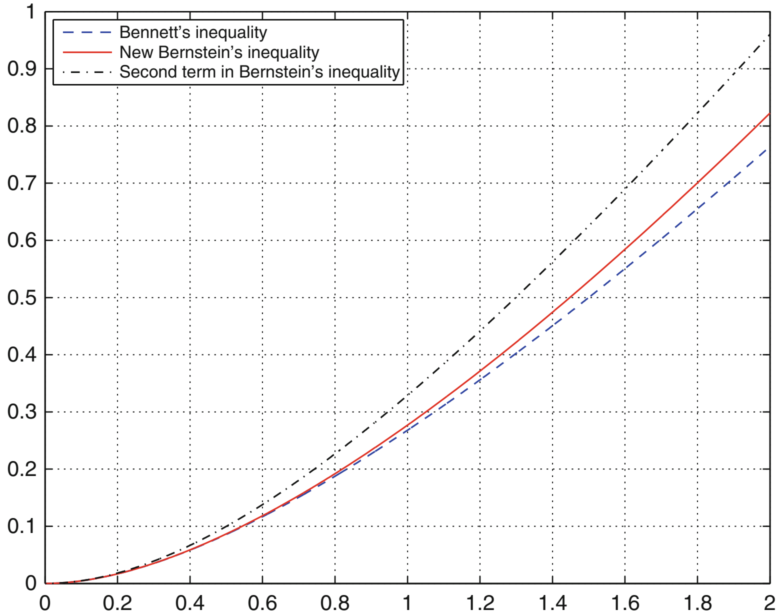

Fig. 2.3 Comparisons in Bernstein's inequalities with the new second term

2.2 Hoeffding's inequality

In this section, we focus our attention on the classical Hoeffding's inequality [11], which requires that the random variables are bounded from above and from below. We shall also establish Antonov's type extensions of this inequality. Let us start with the classical inequality of Hoeffding.

Theorem 2.16 (Hoeffding's inequality). *Let $X_1,\ldots,X_n$ be a finite sequence of independent random variables. Assume that for all $1 \leqslant k \leqslant n$, one can find* *two constants $a_k < b_k$ such that $a_k \leqslant X_k \leqslant b_k$ almost surely. Denote $S_n = X_1 + \ldots + X_n$. Then, for any positive x,*

$$\mathbb{P}(|S_n - \mathbb{E}[S_n]| \geqslant x) \leqslant 2\exp\Big(-\frac{2x^2}{D_n}\Big) \quad \textit{where} \quad D_n = \sum_{k=1}^{n} (b_k - a_k)^2. \tag{2.31}$$

Remark 2.17. We shall see in the proof of Theorem 2.16 below that $D_n \geqslant 4V_n$ where $V_n = \mathrm{Var}(S_n)$. Clearly, the equality $D_n = 4V_n$ does not hold true, unless the random variables $X_1, \ldots, X_n$ have the Bernoulli distribution given, for all $1 \leqslant k \leqslant n$, by $\mathbb{P}(X_k = a_k) = \mathbb{P}(X_k = b_k) = 1/2$. In Section 2.5, we will establish a more efficient inequality. To be more precise, we will improve Hoeffding's inequality (2.31) replacing D_n by the lower bound given by the inequality

$$D_n \geqslant \frac{2}{3}D_n + \frac{4}{3}V_n.$$

The proof of Hoeffding's inequality relies on the following lemmas which give upper bounds for the variances and the Laplace transforms of the random variables $X_1, \ldots, X_n$.

Lemma 2.18. *Let X be a random variable with finite variance σ^2. Assume that $a \leqslant X \leqslant b$ almost surely for some real constants a and b. Denote $m = \mathbb{E}[X]$. Then,*

$$\sigma^2 \leqslant (b-m)(m-a) \leqslant \frac{(b-a)^2}{4}. \tag{2.32}$$

Proof. The convexity of the square function implies that $X^2 \leqslant (a+b)X - ab$ almost surely. Hence $\sigma^2 = \mathbb{E}[X^2] - m^2 \leqslant (a+b)m - ab - m^2 \leqslant -ab + (a+b)^2/4$, which implies the lemma. $\square$

Lemma 2.19. *Let X be a random variable with finite variance σ^2. Assume that $a \leqslant X \leqslant b$ almost surely for some real constants a and b. Then, for any real t,*

$$\log\big(\mathbb{E}[\exp(tX)]\big) \leqslant t\mathbb{E}[X] + \frac{t^2}{8}\big(b-a\big)^2.$$

Proof. Let L and ℓ be the Laplace and log-Laplace transforms of X. As the random variable X is bounded from above and from below, L and ℓ are real analytic functions. Moreover, for any real t, $\ell(t) = \log L(t)$,

$$\ell'(t) = \frac{L'(t)}{L(t)} \qquad \text{and} \qquad \ell''(t) = \frac{L''(t)}{L(t)} - \Big(\frac{L'(t)}{L(t)}\Big)^2.$$

Consider the classical change of probability

$$\frac{d\mathbb{P}_t}{d\mathbb{P}} = \exp\big(tX - \ell(t)\big) = \frac{\exp(tX)}{L(t)}$$

and denote by $\mathbb{E}_t$ the expectation associated with $\mathbb{P}_t$. One can observe that for any integrable random variable Y,

$$\mathbb{E}_t[Y] = \frac{\mathbb{E}[Y \exp(tX)]}{L(t)}.$$

In particular,

$$E_t[X] = \frac{\mathbb{E}[X \exp(tX)]}{L(t)} = \frac{L'(t)}{L(t)}, \qquad E_t[X^2] = \frac{\mathbb{E}[X^2 \exp(tX)]}{L(t)} = \frac{L''(t)}{L(t)}.$$

Consequently, $\ell''(t) = \mathbb{E}_t[X^2] - \mathbb{E}_t^2[X]$, which means that $\ell''(t)$ is equal to the variance of the random variable X under the new probability $\mathbb{P}_t$. As the random variable X takes its values in $[a,b]$ almost surely, we may apply Lemma 2.18 under the new probability $\mathbb{P}_t$, which gives $\ell''(t) \leqslant (b-a)^2/4$. Since $\ell(0) = 0$ and $\ell'(0) = \mathbb{E}[X]$, it completes the proof of Lemma 2.19. $\square$

Proof of Theorem 2.16. We shall now proceed to the proof of Hoeffding's inequality. We deduce from Lemma 2.19 together with the independence of the random variables $X_1, \ldots, X_n$ that, for any real t,

$$\log \mathbb{E}\Big[\exp(tS_n)\Big] = \sum_{k=1}^{n} \log \mathbb{E}\big[\exp(tX_k)\big] \leqslant t\mathbb{E}[S_n] + \frac{t^2}{8} D_n \tag{2.33}$$

where D_n is given by (2.31). For any positive t, it follows from Markov's inequality applied to $\exp(tS_n)$ that

$$\log \mathbb{P}(S_n \geqslant \mathbb{E}[S_n] + x) \leqslant -tx - t\mathbb{E}[S_n] + \log \mathbb{E}[\exp(tS_n)]. \tag{2.34}$$

Consequently, inequalities (2.34) and (2.33) imply that for all $x \geqslant 0$ and $t > 0$,

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geqslant x) \leqslant \exp\left(-tx + \frac{t^2}{8} D_n\right).$$

By taking the optimal value $t = 4x/D_n$, we find that

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geqslant x) \leqslant \exp\left(-\frac{2x^2}{D_n}\right). \tag{2.35}$$

Replacing X_k by $-X_k$, we obtain by the same token that for all $x \geqslant 0$,

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \leqslant -x) \leqslant \exp\left(-\frac{2x^2}{D_n}\right). \tag{2.36}$$

Therefore, (2.31) follows from (2.35) and (2.36), which completes the proof of Theorem 2.16. $\square$

It is necessary to add some comments on Hoeffding's inequality. Recall that for all $1 \leqslant k \leqslant n$, $a_k \leqslant X_k \leqslant b_k$ almost surely, where $a_k < b_k$. For all $1 \leqslant k \leqslant n$, let Z_k be the random variable defined by $Z_k = (X_k - a_k)/c_k$ with $c_k = b_k - a_k$, which is equivalent to $X_k = a_k + c_k Z_k$. For all $1 \leqslant k \leqslant n$, the random variable Z_k takes its values in the interval $[0,1]$ and

$$S_n - \mathbb{E}[S_n] = \sum_{k=1}^{n} c_k (Z_k - \mathbb{E}[Z_k]). \tag{2.37}$$

Consequently $S_n - \mathbb{E}[S_n]$ is a weighted sum of independent random variables whose laws have a support included in an interval of length 1. Hence, the support of the law of S_n is included in an interval of length $\|c\|_1 = c_1 + \cdots + c_n$ and

$$\mathbb{P}(|S_n - \mathbb{E}[S_n]| > \|c\|_1) = 0, \tag{2.38}$$

which cannot be deduced from Hoeffding's inequality. Furthermore, for any $p \geqslant 1$, denote

$$\|c\|_p = \Big(\sum_{k=1}^{n} c_k^p \Big)^{1/p}.$$

Hoeffding's inequality (2.31) is clearly equivalent, for any positive x, to

$$\mathbb{P}(|S_n - \mathbb{E}[S_n]| \geqslant \|c\|_2 x) \leqslant 2\exp(-2x^2). \tag{2.39}$$

Antonov [1] extended inequality (2.39) by proving that for any p in $]1,2]$ and for any positive x,

$$\mathbb{P}(|S_n - \mathbb{E}[S_n]| \geqslant \|c\|_p x) \leqslant 2\exp(-C_q x^q) \tag{2.40}$$

where $q = p/(p-1)$ and the constant C_q obtained by Antonov converges to 0 as q tends to ∞. Observe, however, that Antonov [1] also proved inequality (2.40) under the more general tail assumption that, for all $1 \leqslant k \leqslant n$, $\mathbb{P}(|Z_k| \geqslant x) \leqslant \alpha \exp\big(-\beta x^q\big)$ for some positive constants α and β. Here, we will prove that the constant C_q appearing in (2.40) is larger than 2. We refer the reader to Rio [21] for more details about the constants in (2.40).

Theorem 2.20. *Let $X_1, \ldots, X_n$ be a finite sequence of independent random variables. Assume that for all $1 \leqslant k \leqslant n$, one can find two constants $a_k < b_k$ such that $a_k \leqslant X_k \leqslant b_k$ almost surely and let $c_k = b_k - a_k$. Denote $S_n = X_1 + \cdots + X_n$. Then, for any p in $]1,2]$ and for any positive x,*

$$\mathbb{P}(|S_n - \mathbb{E}[S_n]| \geqslant \|c\|_p x) \leqslant 2\exp\left(-2x^q\right) \tag{2.41}$$

where $q = p/(p-1)$.

Remark 2.21. Let x be any real in $]1,\infty[$. Then, by taking the limit as q tends to infinity in (2.41), we obtain (2.38).

The proof of Theorem 2.20 relies on the following lemma on convex functions.

Lemma 2.22. *Let ℓ be a convex and increasing function from $[0,\infty[$ to $[0,\infty]$ such that $\ell(0)=0$ and $\ell'(0)=0$. Denote by ℓ^* the Fenchel-Legendre transform of ℓ. Then, for any real p in $]1,2]$, we have*

$$a_\ell(p) = \sup_{t>0}\big(t^{-p}\ell(t)\big) = p^{-1}\big(q\inf_{x>0}\big(x^{-q}\ell^*(x)\big)\big)^{1-p} \text{ where } q = p/(p-1). \quad (2.42)$$

Proof. Since ℓ is a convex function $\ell = (\ell^*)^*$. Hence, for any positive t,

$$\ell(t) = \sup_{x>0}\big(xt - \ell^*(x)\big),$$

which implies that

$$a_\ell(p) = \sup_{x>0}\big(\sup_{t>0} t^{-p}\big(xt - \ell^*(x)\big)\big). \quad (2.43)$$

For any positive t, denote $f(t) = t^{-p}(xt - \ell^*(x))$. In order to prove (2.42), it is necessary to compute the maximum of the function f. From the assumption $\ell(0) = \ell'(0) = 0$, we know that $\ell^*(x) > 0$ for any positive x. Moreover, we also have

$$f'(t) = t^{-1-p}(p\ell^*(x) - (p-1)xt).$$

Hence, f has a unique maximum at the point $t_x = q\ell^*(x)/x$ where $q = p/(p-1)$. Therefore,

$$\sup_{t>0} t^{-p}\big(xt - \ell^*(x)\big) = (q-1)\ell^*(x)t_x^{-p} = (q-1)q^{-p}\big(x^q/\ell^*(x)\big)^{p-1}. \quad (2.44)$$

Consequently, we deduce from (2.43) and (2.44) that

$$a_\ell(p) = \frac{q-1}{q}\sup_{x>0}\Big(\frac{x^q}{q\ell^*(x)}\Big)^{p-1},$$

which implies Lemma 2.22. □

Proof of Theorem 2.20. Let Z be a centered random variable with values in the interval $[a, a+1]$ with $a < 0$. It follows from Lemma 2.19 that for any real t,

$$\ell_Z(t) = \log\mathbb{E}[\exp(tZ)] \leqslant \frac{t^2}{8}.$$

Moreover, as $a < 0$, $Z \leqslant 1$ almost surely, which implies that

$$\ell_Z'(t) = \frac{\mathbb{E}[Z\exp(tZ)]}{\mathbb{E}[\exp(tZ)]} \leqslant 1.$$

From the two above inequalities and the convexity of the log-Laplace transform ℓ_Z, we obtain that, for any real t,

$$\ell_Z(t) \leqslant \ell(t), \quad (2.45)$$

where $\ell(t) = t^2/8$ if $t \leqslant 4$ and $\ell(t) = t - 2$ if $t \geqslant 4$. Hereafter, starting from (2.37) and applying (2.45) to the random variables $Z_k - \mathbb{E}[Z_k]$, we find that for any real t,

$$\log \mathbb{E}[\exp(tS_n)] - t\mathbb{E}[S_n] \leqslant \sum_{k=1}^{n} \ell(c_k t) \leqslant \sup_{x>0} \bigl(x^{-p}\ell(x)\bigr) \, \|c\|_p^p t^p. \tag{2.46}$$

Denote by ℓ^* the Legendre-Fenchel transform of the convex function ℓ. It follows from straightforward calculation that $\ell^*(x) = 2x^2$ for x in $[0,1]$ and $\ell_0^*(x) = +\infty$ for $x > 1$. Hence, for any $q \geqslant 2$,

$$\inf_{x>0} x^{-q} \ell_0^*(x) = 2.$$

Consequently, we obtain from (2.46) and Lemma 2.22 that, for any real t,

$$\log \mathbb{E}[\exp(tS_n)] - t\mathbb{E}[S_n] \leqslant p^{-1}(2q)^{1-p}\|c\|_p^p t^p. \tag{2.47}$$

Hence, we deduce from (2.47) together with Markov's inequality that for all $x \geqslant 0$ and $t > 0$,

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geqslant \|c\|_p x) \leqslant \exp\bigl(tx - (2q)^{1-p}(t^p/p)\bigr).$$

By taking the optimal value $t = 2qx^{q-1}$ in this inequality, we find that

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geqslant \|c\|_p x) \leqslant \exp\bigl(-2x^q\bigr). \tag{2.48}$$

Replacing X_k by $-X_k$, we obviously obtain the same inequality, which completes the proof of Theorem 2.20. □

2.3 Binomial rate functions

Throughout this section, we assume that $X_1, \ldots, X_n$ is a finite sequence of independent random variables bounded from above. More precisely, we assume that there exists a positive constant b such that, for all $1 \leqslant k \leqslant n$,

$$X_k \leqslant b \qquad \text{a.s.} \tag{2.49}$$

The purpose of this section is to compare the tails of their sum S_n on the right with the tails of binomial random variables. The fundamental tool of the proofs is the lemma below, whose second part is due to Bennett [2]. We refer to Bentkus [4] for the first part of the lemma and for other results in this direction under condition (2.49), and to Pinelis [19] for a refinement of (2.49) under some additional condition on the random variable X appearing in the lemma below. Next we will deduce Theorems 1 and 3 of Hoeffding [11] from this key lemma.

Lemma 2.23. *Let X be a centered random variable with finite variance σ^2. Assume that $X \leqslant b$ almost surely for some positive real b. Let v be any positive real such that $\sigma^2 \leqslant v$. Denote by ξ the Bernoulli type random variable with mean 0 and variance v defined by*

$$\mathbb{P}(\xi = b) = \frac{v}{b^2+v} \qquad \text{and} \qquad \mathbb{P}(\xi = -v/b) = \frac{b^2}{b^2+v}.$$

Then, for any real t,

$$\mathbb{E}[(X-t)_+^2] \leqslant \mathbb{E}[(\xi - t)_+^2]. \tag{2.50}$$

Consequently, for any positive real t,

$$\mathbb{E}[\exp(tX)] \leqslant \mathbb{E}[\exp(t\xi)] = \frac{v}{b^2+v}\exp(tb) + \frac{b^2}{b^2+v}\exp(-tv/b). \tag{2.51}$$

Proof. If $t \geqslant b$, then the two expectations in (2.50) vanish and there is nothing to prove. If $t \leqslant -v/b$, then

$$\mathbb{E}[(X-t)_+^2] \leqslant \mathbb{E}[(X-t)^2] = \sigma^2 + t^2 \leqslant v + t^2 = \mathbb{E}[(\xi-t)^2] = \mathbb{E}[(\xi-t)_+^2].$$

If t belongs to $[-v/b, b]$, then for any $x \in [-v/b, b]$,

$$(x-t)_+ \leqslant \frac{b-t}{b^2+v}(bx+v)_+$$

leading to

$$\mathbb{E}[(X-t)_+^2] \leqslant \frac{(b-t)^2}{(b^2+v)^2}\mathbb{E}[(bX+v)_+^2] \leqslant \frac{(b-t)^2}{(b^2+v)^2}\mathbb{E}[(bX+v)^2].$$

However,

$$\frac{(b-t)^2}{(b^2+v)^2}\mathbb{E}[(bX+v)^2] = \frac{(b-t)^2 v}{b^2+v} = \mathbb{E}[(\xi-t)_+^2],$$

which shows that (2.50) still holds. We are now in position to prove (2.51). For any $x \in \mathbb{R}$ and any positive t, we have the integral representation

$$\mathbb{E}[\exp(tX)] = \frac{t^2}{2}\int_{\mathbb{R}} \mathbb{E}\Big[\Big(X - \frac{s}{t}\Big)_+^2\Big]\exp(s)ds.$$

Consequently, (2.51) immediately follows from (2.50), which completes the proof of Lemma 2.23. □

We now apply (2.51) to sums of independent random variables $X_1, \dots, X_n$ bounded from above. In the case of centered random variables, the result below coincides with Theorem 3 in Hoeffding [11]. We also refer to Bennett [3] for an analogous bound under the additional assumption that the random variables $X_1, \dots, X_n$ are symmetrically distributed about their mean.

Theorem 2.24. *Let $X_1, \dots, X_n$ be a finite sequence of independent random variables with finite variances satisfying (2.49) for some positive real b. Let S_n and v_n be defined as in (2.1) and assume that $\mathbb{E}[S_n] = 0$. Then, for any $v \geqslant v_n$ and for any x in $[0,b]$,*

$$\mathbb{P}(S_n \geqslant nx) \leqslant \exp\Big(-n\Big(\Big(\frac{v+bx}{v+b^2}\Big)\log\Big(1+\frac{bx}{v}\Big)+\Big(\frac{b^2-bx}{b^2+v}\Big)\log\Big(1-\frac{x}{b}\Big)\Big)\Big),$$
$$\leqslant \exp\big(-ng(b,v)x^2\big), \tag{2.52}$$

where

$$g(b,v) = \begin{cases} \dfrac{b^2}{(b^4-v^2)}\log\Big(\dfrac{b^2}{v}\Big) & \text{if } v < b^2, \\ \dfrac{1}{2v} & \text{if } v \geqslant b^2. \end{cases} \tag{2.53}$$

Remark 2.25. Note that $S_n \leqslant nb$ almost surely, which implies that $\mathbb{P}(S_n > nb) = 0$.

Proof. We shall only prove Theorem 2.24 in the particular case $b = 1$, inasmuch as the general case follows by dividing the initial random variables by b. According to Lemma 2.6, we have for any real t

$$\frac{1}{n}\log\mathbb{E}[\exp(tS_n)] \leqslant \ell(t) \quad \text{where} \quad \ell(t) = \log\Big(\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[\exp(tX_k)]\Big).$$

Denote by X a random variable with distribution

$$\mu = \frac{1}{n}(\mu_1 + \cdots + \mu_n)$$

where, for all $1 \leqslant k \leqslant n$, μ_k stands for the distribution of X_k. We clearly have $X \leqslant 1$ almost surely. In addition, $\mathbb{E}[X] = 0$ and

$$\mathbb{E}[X^2] = \frac{1}{n}(\mathbb{E}[X_1^2] + \cdots + \mathbb{E}[X_n^2]) \leqslant v.$$

Hence, according to (2.51), we have for any positive t,

$$\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[\exp(tX_k)] = \mathbb{E}[\exp(tX)] \leqslant \frac{v}{1+v}\exp(t) + \frac{1}{1+v}\exp(-vt).$$

It implies that, for any positive t,

$$\log\mathbb{E}[\exp(tS_n)] \leqslant nL_v(t) \tag{2.54}$$

where

$$L_v(t) = \log(ve^t + e^{-vt}) - \log(1+v).$$

In order to prove the first part of Theorem 2.24, it remains to prove that the Legendre-Fenchel transform L_v^* of L_v is given by

$$L_v^*(x) = \begin{cases} \left(\frac{v+x}{v+1}\right)\log\left(1+\frac{x}{v}\right) + \left(\frac{1-x}{1+v}\right)\log(1-x) & \text{if } x \in [0,1], \\ +\infty & \text{if } \quad x > 1. \end{cases} \tag{2.55}$$

In order to compute $L_v^*(x)$, we have to solve the equation $L_v'(t) = x$. On the one hand, for $x < 1$,

$$L_v'(t) = \frac{v(1-e^{-t(1+v)})}{v+e^{-t(1+v)}} = x \Longleftrightarrow (1+v)t = \log\left(1+\frac{x}{v}\right) - \log(1-x).$$

For this value of t,

$$tx - L_v(x) = \left(\frac{v+x}{v+1}\right)\log\left(1+\frac{x}{v}\right) + \left(\frac{1-x}{1+v}\right)\log(1-x) = L_v^*(x),$$

which gives (2.55) in the case $x < 1$. On the other hand, for $x \geqslant 1$, the function $t \to tx - L_v(t)$ is increasing. Hence, if $x = 1$,

$$L_v^*(x) = \lim_{t\to\infty}(t - L_v(t)) = \log\left(1+\frac{1}{v}\right)$$

while, if $x > 1$, $L_v^*(x) = +\infty$, which completes the proof of (2.55). We deduce the first part of Theorem 2.24 from (2.54) and (2.55). It now remains to prove the second part of Theorem 2.24. This second part immediately follows from the lower bound below on L_v^*, due to Hoeffding [11]. □

Lemma 2.26. *Let L_v^* be defined as in (2.55). Then, we have*

$$\inf_{x>0}\left(\frac{L_v^*(x)}{x^2}\right) = \begin{cases} \frac{|\log(v)|}{1-v^2} & \text{if } v < 1, \\ \frac{1}{2v} & \text{if } v \geqslant 1. \end{cases} \tag{2.56}$$

Proof. Since $L_v^*(x) = +\infty$ for $x > 1$,

$$\inf_{x>0}\left(\frac{L_v^*(x)}{x^2}\right) = \inf_{0<x\leqslant 1}\left(\frac{L_v^*(x)}{x^2}\right).$$

Let ψ_v be the function defined by $\psi_v(x) = x^{-2}L_v^*(x)$ if x in $]0,1]$ and $\psi_v(0) = 1/(2v)$. Then, ψ_v is differentiable on $[0,1[$, and for any x in $[0,1[$,

$$x^2\psi_v'(x) = (L_v^*)'(x) - 2x^{-1}L_v^*(x). \tag{2.57}$$

We already saw that $(1+v)(L_v^*)'(x) = \log(1+x/v) - \log(1-x)$. Hence, if φ_v is the function defined, for all x in $]0,1[$, by $\varphi_v(x) = (1+v)x^2\psi_v'(x)$, it follows that

$$\varphi_v(x) = \Big(1 - \frac{2}{x}\Big)\log(1-x) - \Big(1 + \frac{2v}{x}\Big)\log\Big(1 + \frac{x}{v}\Big).$$

Apparently it seems difficult to find the roots of the above function. However,

$$\begin{aligned}\Big(1 + \frac{2v}{x}\Big)\log\Big(1 + \frac{x}{v}\Big) &= -\Big(1 + \frac{2v}{x}\Big)\log\Big(\frac{v}{v+x}\Big), \\ &= \Big(1 - \frac{2(v+x)}{x}\Big)\log\Big(1 - \frac{x}{v+x}\Big).\end{aligned}$$

Therefore, one can realize that, for all x in $]0,1[$,

$$\varphi_v(x) = H(x) - H\Big(\frac{x}{v+x}\Big) \quad \text{where} \quad H(x) = \Big(1 - \frac{2}{x}\Big)\log(1-x). \tag{2.58}$$

Furthermore, for all x in $]0,1[$, we have the expansion

$$H(x) = 2 + \sum_{k=2}^{\infty} \frac{(k-1)}{k(k+1)} x^k.$$

In this expansion, the coefficients are positive. Thus, H is increasing on $]0,1[$. Hence, it follows from (2.57) and (2.58) that $\psi_v'(x) > 0$ if and only if $x > x/(v+x)$, that is $x > 1 - v$. Therefrom, if $v \geqslant 1$, then ψ_v is increasing on $]0,1]$. Consequently, the value of the minimum is

$$C_v = \lim_{x \searrow 0} \frac{L_v^*(x)}{x^2} = \frac{1}{2v},$$

which gives the second part of Lemma 2.26. If $v < 1$, then ψ_v has its minimum at $x = 1 - v$ and the value of this minimum is

$$C_v = \psi_v(1-v) = \frac{|\log(v)|}{1-v^2},$$

which completes the proof of Lemma 2.26. $\square$

To conclude this section, we apply Theorem 2.24 to independent random variables with values in $[0,1]$. The result below is exactly Theorem 1 in Hoeffding [11].

Theorem 2.27. *Let $X_1, \ldots, X_n$ be a finite sequence of independent random variables with values in $[0,1]$ and denote $\mu = \mathbb{E}[S_n]/n$. Then, for any x in $]\mu, 1[$,*

$$\begin{aligned}\mathbb{P}(S_n \geqslant nx) &\leqslant \exp\Big(-n\Big(x\log\Big(\frac{x}{\mu}\Big) + (1-x)\log\Big(\frac{1-x}{1-\mu}\Big)\Big)\Big), \\ &\leqslant \exp\big(-ng(\mu)(x-\mu)^2\big), \qquad (2.59) \\ &\leqslant \exp\big(-2n(x-\mu)^2\big),\end{aligned}$$

where

$$g(\mu) = \begin{cases} \dfrac{1}{1-2\mu}\log\Big(\dfrac{1-\mu}{\mu}\Big) & \text{if } 0 < \mu < \dfrac{1}{2}, \\ \dfrac{1}{2\mu(1-\mu)} & \text{if } \dfrac{1}{2} \leqslant \mu < 1. \end{cases}$$

Proof. If $\mu = 0$ or $\mu = 1$, then $S_n = n\mu$ almost surely, and there is nothing to prove. Hereafter, assume that $0 < \mu < 1$ and denote, for all $1 \leqslant k \leqslant n$, $Y_k = X_k - \mu$. The random variables $Y_1, \ldots, Y_n$ satisfy the assumptions of Theorem 2.24 with $b = 1 - \mu$. Moreover, if $\Sigma_n = Y_1 + \cdots + Y_n$, we clearly have $\mathbb{E}[\Sigma_n] = \mathbb{E}[S_n] - n\mu = 0$. In addition, for all $1 \leqslant k \leqslant n$,

$$\mathbb{E}[Y_k^2] = \mathbb{E}[X_k^2] - 2\mu\mathbb{E}[X_k] + \mu^2 \leqslant (1-2\mu)\mathbb{E}[X_k] + \mu^2,$$

since $\mathbb{E}[X_k^2] \leqslant \mathbb{E}[X_k]$. It leads to

$$\mathrm{Var}(\Sigma_n) = \mathbb{E}[Y_1^2] + \cdots + \mathbb{E}[Y_n^2] \leqslant n\mu(1-2\mu) + n\mu^2 = n\mu(1-\mu).$$

Finally, Theorem 2.27 immediately follows from Theorem 2.24 with $b = 1 - \mu$ and $v = \mu(1-\mu)$. $\square$

2.4 Bennett's inequality

In this section, we deduce Bennett's type inequalities from the results of Section 2.3. First of all, let h and h_w be the functions defined by

$$h(x) = \begin{cases} (1+x)\log(1+x) - x & \text{if } x > -1, \\ 1 & \text{if } x = -1, \\ +\infty & \text{if } x < -1, \end{cases} \tag{2.60}$$

and

$$h_w(x) = \begin{cases} \dfrac{h(wx)}{w^2} & \text{if } w \neq 0, \\ \dfrac{x^2}{2} & \text{if } w = 0. \end{cases} \tag{2.61}$$

Theorem 2.28. *Let $X_1, \ldots, X_n$ be a finite sequence of independent random variables satisfying (2.49) for some positive constant b. Let S_n and v_n be defined as in (2.1) and assume that $\mathbb{E}[S_n] = 0$. Let $w_n = (b/v_n) - (1/b)$. Then, for any x in $[0, b]$,*

$$\begin{aligned} \mathbb{P}(S_n \geqslant nx) &\leqslant \exp\Big(-\frac{n}{v_n} h_{w_n}(x)\Big), \qquad (2.62) \\ &\leqslant \exp\Big(-\frac{n v_n}{b^2} h\Big(\frac{bx}{v_n}\Big)\Big), \end{aligned}$$

where the above functions are given by (2.60) *and* (2.61)*. Hence, if* $v_n \geqslant b^2$*, then, for any positive x,*

$$\mathbb{P}(S_n \geqslant nx) \leqslant \exp\Big(-\frac{nx^2}{2v_n}\Big). \tag{2.63}$$

Furthermore, for any x in $[0,b]$*,*

$$\begin{aligned}\mathbb{P}(S_n \geqslant nx) &\leqslant \exp\Big(-\frac{nx^2}{2(v_n+bx/3)(1-x/(3b))}\Big) \\ &\leqslant \exp\Big(-\frac{nx^2}{2(v_n+bx/3)}\Big).\end{aligned} \tag{2.64}$$

Remark 2.29. The second upper bound in (2.62) is known as Bennett's inequality. This inequality was called second improvement of Bernstein's inequality in Bennett [2]. This second improvement is more efficient than the first improvement of Bennett, which corresponds to inequality (2.4) with $c = b/3$. We now discuss the first upper bound. Elementary computations show that h_w are decreasing with respect to w. Hence, the first upper bound is more efficient than Bennett's inequality. For example, if $v_n \geqslant b^2$, then $w_n \leqslant 0$. In that case, $h_{w_n}(x) \geqslant x^2/2$, which implies (2.63). On the other side, (2.63) cannot be deduced from Bennett's inequality. Note, however, that (2.63) has already been established in Theorem 2.24 of Section 2.3. It comes from the proof of Theorem 2.28 that the above upper bounds are increasing with respect to v_n. Consequently, one can replace v_n by any real $v \geqslant v_n$ in Theorem 2.28.

Remark 2.30. Inequality (2.64) is an improved version of Bernstein's inequality for bounded random variables, which implies (2.63). Figure 2.4 below compares the rate functions given in Theorem 2.28 in the particular case $b = 1$ and $v_n = 1$. For any x in $[0,1]$,

$$\Phi(x) = \frac{x^2}{2(1+x/3)} \qquad \text{and} \qquad \Psi(x) = \frac{x^2}{2(1+x/3)(1-x/3)}$$

are the rate functions in Bernstein's inequality and its improvement for bounded random variables, respectively. In addition, for any x in $[0,1]$,

$$h(x) = (1+x)\log(1+x) - x, \qquad \ell(x) = \frac{x^2}{2},$$

$$\varphi(x) = \frac{1}{2}(1+x)\log(1+x) + \frac{1}{2}(1-x)\log(1-x)$$

are the rate functions in Bennett's inequality, the improvement of Bennett's inequality, and the Binomial rate function. One can realize that inequality (2.52) with the Binomial rate function outperforms Bernstein and Bennett inequalities.

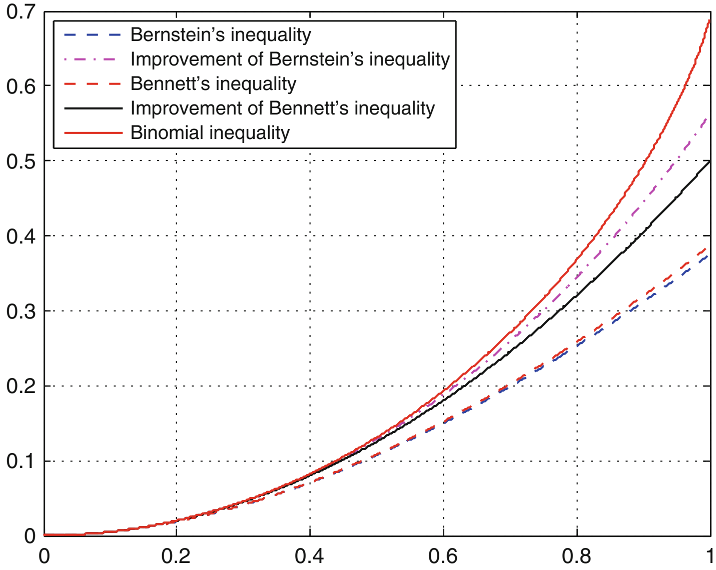

Fig. 2.4 Comparisons in Bennett's inequalities

Remark 2.31. An interesting situation occurs when w_n is different from zero and v_n is small. Figure 2.5 below compares the more efficient rate functions of Theorem 2.28 given, for any x in $[0,1]$, by

$$\Psi_n(x) = \frac{x^2}{2(v_n + bx/3)(1 - x/(3b_n))},$$

$$h_n(x) = \frac{1}{v_n w_n^2}\Big((1 + w_n x)\log(1 + w_n x) - w_n x\Big),$$

and

$$\varphi_n(x) = \left(\frac{v_n + x}{v_n + 1}\right)\log\left(1 + \frac{x}{v_n}\right) + \left(\frac{1 - x}{1 + v_n}\right)\log(1 - x)$$

in the special case $b = 1$ and $v_n = 1/20$. One can observe that inequality (2.52) with the Binomial rate function still outperforms the improvements of Bernstein and Bennett inequalities. However, the improvement of Bennett's inequality behaves better than the improvement of Bernstein's inequality.

Proof. The fact that (2.62) implies (2.63) is already proven in Remark 2.29. We shall only prove Theorem 2.28 in the special case $b = 1$, inasmuch as the general case follows by dividing the initial random variables by b. Throughout the proof, $v = v_n$ and $w = w_n = (1 - v)/v$. According to (2.52), we have for any x in $[0,1]$,

$$\mathbb{P}(S_n \geqslant nx) \leqslant \exp(-nL_v^*(x))$$

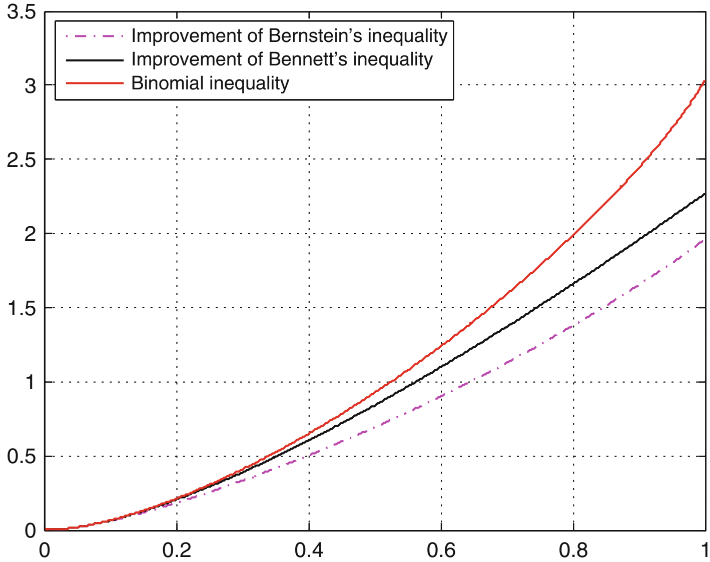

Fig. 2.5 Comparisons in the situation where w_n is different from zero

where

$$L_v^*(x) = \left(\frac{v+x}{v+1}\right)\log\left(1+\frac{x}{v}\right) + \left(\frac{1-x}{1+v}\right)\log(1-x)$$

with the convention $0\log 0 = 0$. Furthermore, $\mathbb{P}(S_n > n) = 0$. Consequently (2.62) and (2.64) follow from the lower bounds below on L_v^*, which achieves the proof of Theorem 2.28. □

Lemma 2.32. *For any x in $[0,1]$, we have*

$$L_v^*(x) \geqslant \frac{h_w(x)}{v} \geqslant vh\left(\frac{x}{v}\right)$$

where $w = (1-v)/v$. Moreover, for any x in $[0,1]$,

$$L_v^*(x) \geqslant \frac{x^2}{2(v+x/3)(1-x/3)}.$$

Proof. In order to prove the first part of Lemma 2.32, one can observe that $L_v^*(0) = (L_v^*)'(0) = 0$ and

$$(L_v^*)''(x) = \frac{1}{(v+x)(1-x)}.$$

Consequently, for any x in $[0,1]$,

$$L_v^*(x) = \int_0^x (x-t)(L_v^*)''(t)dt = \int_0^x \frac{(x-t)}{(v+t)(1-t)}dt.$$

For any t in $[0,1[$, $0 < (v+t)(1-t) \leqslant v + (1-v)t \leqslant v+t$. Hence,

$$L_v^*(x) \geqslant \int_0^x \frac{x-t}{v+(1-v)t}\,dt \geqslant \int_0^x \frac{x-t}{v+t}\,dt. \tag{2.65}$$

Now,

$$\int_0^x \frac{x-t}{v+(1-v)t}\,dt = \frac{h_w(x)}{v} \qquad \text{and} \qquad \int_0^x \frac{x-t}{v+t}\,dt = vh\Big(\frac{x}{v}\Big),$$

which completes the proof of the first part. The main tool of the proof of the second part is an expansion of L_v^* into power series. We start by noting that

$$(1+v)L_v^*(x) = vh\Big(\frac{x}{v}\Big) + h(-x). \tag{2.66}$$

Next, for any x in $[0,1]$,

$$h(-x) = \frac{x^2}{2}\sum_{k=0}^{\infty} \frac{2x^k}{(k+2)(k+1)}. \tag{2.67}$$

In order to give an expansion of $vh(x/v)$, we note that, for any positive x,

$$h(x) = -(1+x)\log\Big(1-\frac{x}{1+x}\Big) - x = \sum_{k=2}^{\infty} \frac{1}{k}\frac{x^k}{(1+x)^{k-1}},$$

which leads to the expansion

$$vh\Big(\frac{x}{v}\Big) = \frac{x^2}{2(v+x)}\sum_{k=0}^{\infty} \frac{2}{k+2}\Big(\frac{x}{v+x}\Big)^k. \tag{2.68}$$

Consequently, starting from (2.67) and noticing that $(k+1)(k+2) \leqslant 2(3)^k$ for any nonnegative integer k, we find that for any x in $[0,1]$,

$$h(-x) \geqslant \frac{x^2}{2}\sum_{k=0}^{\infty}\Big(\frac{x}{3}\Big)^k = \frac{x^2}{2(1-x/3)}. \tag{2.69}$$

In a similar way, $k+2 \leqslant 2(3/2)^k$ for any nonnegative integer k. Hence, it follows from (2.68) that for any positive x,

$$vh\Big(\frac{x}{v}\Big) \geqslant \frac{x^2}{2(v+x)}\sum_{k=0}^{\infty}\Big(\frac{2x}{3(v+x)}\Big)^k = \frac{x^2}{2(v+x/3)}. \tag{2.70}$$

Finally, putting together (2.69) and (2.70) into (2.66), we obtain that for any x in $[0,1]$,

$$(1+v)L_v^*(x) \geqslant \frac{x^2}{2}\Big(\frac{1}{1-x/3}+\frac{1}{v+x/3}\Big) = \frac{(1+v)x^2}{2(v+x/3)(1-x/3)},$$

which completes the proof of Lemma 2.32. □

2.5 SubGaussian inequalities

This section is devoted to concentration inequalities with Gaussian rate functions. Let us recall Hoeffding's inequality for independent and bounded random variables, given in Section 2.2. Let $X_1,\ldots,X_n$ be a finite sequence of independent and centered random variables satisfying, for all $1 \leqslant k \leqslant n$, $a_k \leqslant X_k \leqslant b_k$ almost surely for some real constants a_k and b_k such that $a_k < b_k$. Let $S_n = X_1 + \cdots + X_n$. Then, for any positive x,

$$\mathbb{P}(S_n \geqslant x) \leqslant \exp\Big(-\frac{2x^2}{D_n}\Big) \qquad \text{where} \qquad D_n = \sum_{k=1}^{n}(b_k-a_k)^2. \tag{2.71}$$

Our goal in this section is twofold. First, our aim is to weaken the assumptions on the random variables appearing in Hoeffding's inequality. Next, our aim is to obtain smaller constants than D_n in this inequality. This section is divided into four subsections. In Subsection 2.5.1, we are interested in Gaussian rate functions for the deviations on the right of sums of independent random variables bounded from above. In the next subsection, we apply the results of Subsection 2.5.1 and Section 2.3 to the deviation on the left of sums of independent and nonnegative random variables. In Subsection 2.5.3, we establish subGaussian inequalities for sums of random variables satisfying symmetric boundedness conditions. Finally, Subsection 2.5.4 is devoted to several improvements of Hoeffding's inequality (2.71). In particular, this subsection includes the so-called Kearns-Saul inequality [12] and the following improvement of Hoeffding's inequality. For any positive x,

$$\mathbb{P}(S_n \geqslant x) \leqslant \exp\Big(-\frac{3x^2}{D_n+2V_n}\Big) \tag{2.72}$$

where V_n stands for the variance of S_n. To the best of our knowledge, this improvement is new. In fact, this improvement will be derived from inequalities for sums of random variables bounded from above. This is the reason why we start this section by inequalities for sums of random variables bounded from above, which are the fundamental tools of Section 2.5.

2.5.1 Random variables bounded from above

Throughout this subsection, we assume that $X_1, \ldots, X_n$ is a finite sequence of independent and centered random variables satisfying, for all $1 \leqslant k \leqslant n$,

$$\mathrm{Var}(X_k) \leqslant v_k \qquad \text{and} \qquad X_k \leqslant b_k \tag{2.73}$$

almost surely, for some finite sequences $v_1, \ldots, v_n$ and $b_1, \ldots, b_n$ of positive real numbers. We start with our main result.

Theorem 2.33. *Let $X_1, \ldots, X_n$ be a finite sequence of independent and centered random variables satisfying (2.73). Denote $S_n = X_1 + \cdots + X_n$ and $V_n = v_1 + \cdots + v_n$. Let φ be the function defined by*

$$\varphi(v) = \begin{cases} \dfrac{1-v^2}{|\log(v)|} & \textit{if } \ v < 1, \\ 2v & \textit{if } \ v \geqslant 1. \end{cases} \tag{2.74}$$

Then, for any positive x,

$$\mathbb{P}(S_n \geqslant x) \leqslant \exp\Bigl(-\frac{x^2}{A_n}\Bigr) \quad \textit{where} \quad A_n = \sum_{k=1}^{n} b_k^2 \varphi\Bigl(\frac{v_k}{b_k^2}\Bigr). \tag{2.75}$$

Consequently, for any positive x,

$$\mathbb{P}(S_n \geqslant x) \leqslant \exp\Bigl(-\frac{3x^2}{6V_n + B_n}\Bigr) \leqslant \exp\Bigl(-\frac{3x^2}{5V_n + C_n}\Bigr) \tag{2.76}$$

where

$$B_n = \sum_{k=1}^{n} \Bigl(b_k - \frac{v_k}{b_k}\Bigr)_+^2 \quad \textit{and} \quad C_n = \sum_{k=1}^{n} \max\bigl(b_k^2, v_k\bigr).$$

Remark 2.34. The function φ is continuous and increasing.

Remark 2.35. It immediately follows from (2.76) that, for any positive x,

$$\mathbb{P}(S_n \geqslant x) \leqslant \Bigl(-\frac{x^2}{2C_n}\Bigr). \tag{2.77}$$

This inequality was stated and proved in the more general framework of martingale difference sequences bounded from above, by Bentkus [5]. On the one hand, assume that for all $1 \leqslant k \leqslant n$, $v_k \geqslant b_k^2$. Then, the two above inequalities are equivalent to

$$\mathbb{P}(S_n \geqslant x) \leqslant \exp\Bigl(-\frac{x^2}{2V_n}\Bigr).$$

On the other hand, (2.76) is more efficient than (2.77) if $V_n < b_1^2 + \cdots + b_n^2$.

The proof of Theorem 2.33 relies on the two lemmas below, which will play an important role in the rest of the section. The first one gives an upper bound for the Laplace transform of a centered random variable bounded for above.

Lemma 2.36. *Let X be a centered random variable such that $X \leqslant 1$ almost surely and $Var(X) \leqslant v$ for some positive constant v. Then, for any positive t,*

$$\log \mathbb{E}[\exp(tX)] \leqslant \frac{1}{4}\varphi(v)t^2 \tag{2.78}$$

where φ is the function given by (2.74)*.*

Proof. According to Lemma 2.23, we have for any nonnegative t,

$$\log \mathbb{E}[\exp(tX)] \leqslant L_v(t)$$

where

$$L_v(t) = \log(ve^t + e^{-vt}) - \log(1+v).$$

Moreover, we already saw in Lemma 2.26 that the Legendre-Fenchel transform L_v^* of L_v satisfies $L_v^*(x) \geqslant \psi(x)$ for any positive x, where $\psi(x) = x^2/\varphi(v)$. Taking the Legendre-Fenchel transforms in this inequality, we then derive that $(L_v^*)^*(t) = L_v(t) \leqslant \psi^*(t)$ for any positive t. Moreover, it is not hard to see that, for any positive t, $\psi^*(t) = \varphi(v)t^2/4$, leading to

$$L_v(t) \leqslant \frac{1}{4}\varphi(v)t^2,$$

which completes the proof of Lemma 2.36. □

The second lemma provides a computationally tractable upper bound for the function φ.

Lemma 2.37. *For any v in $]0,1]$,*

$$\varphi(v) \leqslant \frac{1}{3}(1+4v+v^2). \tag{2.79}$$

Proof. By the very definition (2.74) of the function φ, inequality (2.79) holds if and only if, for any v in $]0,1[$,

$$-(1+4v+v^2)\log v \geqslant 3(1-v^2).$$

Via the change of variables $u = 1-v$, the above inequality is equivalent to

$$-(6-6u+u^2)\log(1-u) + 3u(u-2) \geqslant 0.$$

However, we deduce from the Taylor expansion of the logarithm

$$-\log(1-u) = \sum_{k=0}^{\infty} \frac{u^k}{k}$$

that, for any u in $]0,1[$,

$$-(6-6u+u^2)\log(1-u)+3u(u-2)=\sum_{k=5}^{\infty}\frac{(k-3)(k-4)}{k(k-1)(k-2)}u^k \geqslant 0,$$

which is exactly what we wanted to prove. □

Proof of Theorem 2.33. We deduce from Lemma 2.36 together with the independence of the random variables $X_1,\dots,X_n$ that, for any positive t,

$$\begin{aligned}\log\mathbb{E}\Big[\exp(tS_n)\Big] &= \sum_{k=1}^{n}\log\mathbb{E}\big[\exp(tX_k)\big] = \sum_{k=1}^{n}\log\mathbb{E}\Big[\exp\Big(tb_k\frac{X_k}{b_k}\Big)\Big],\\ &\leqslant \frac{t^2}{4}\sum_{k=1}^{n} b_k^2\varphi\Big(\frac{v_k}{b_k^2}\Big).\end{aligned}\tag{2.80}$$

Consequently, (2.75) immediately follows (2.80) via the usual Chernoff calculation. It only remains to prove (2.76). According to Lemma 2.37, we have for any $1\leqslant k\leqslant n$, as soon as $v_k<b_k^2$,

$$3b_k^2\varphi\Big(\frac{v_k}{b_k^2}\Big)\leqslant b_k^2+4v_k+\frac{v_k^2}{b_k^2}=6v_k+\Big(b_k-\frac{v_k}{b_k}\Big)^2.$$

Hence, for any $1\leqslant k\leqslant n$ such that $v_k<b_k^2$,

$$3b_k^2\varphi\Big(\frac{v_k}{b_k^2}\Big)\leqslant\Big(6v_k+\Big(b_k-\frac{v_k}{b_k}\Big)_+^2\Big).$$

Obviously, the above inequality still holds if $v_k\geqslant b_k^2$, as the left-hand side in this inequality is exactly $6v_k$. Therefore,

$$A_n=\sum_{k=1}^{n}b_k^2\varphi\Big(\frac{v_k}{b_k^2}\Big)\leqslant\frac{1}{3}\sum_{k=1}^{n}6v_k+\Big(b_k-\frac{v_k}{b_k}\Big)_+^2=\frac{1}{3}(6V_n+B_n),$$

which leads to the first inequality in (2.76). Finally, for any $1\leqslant k\leqslant n$, as

$$\Big(b_k-\frac{v_k}{b_k}\Big)_+^2\leqslant b_k\Big(b_k-\frac{v_k}{b_k}\Big)_+=\big(b_k^2-v_k\big)_+,$$

we find that

$$6V_n+B_n\leqslant 5V_n+\sum_{k=1}^{n}v_k+\big(b_k^2-v_k\big)_+=5V_n+\sum_{k=1}^{n}\max\big(b_k^2,v_k\big)=5V_n+C_n$$

which clearly implies the second inequality in (2.76). □

We shall now apply Theorem 2.33 to weighted sums of independent random variables bounded from above. It yields an exponential inequality with a Gaussian rate function.

Corollary 2.38. *Let $Z_1, \dots, Z_n$ be a finite sequence of independent and centered random variables such that, for all $1 \leqslant k \leqslant n$, $Z_k \leqslant 1$ almost surely. Assume that there exists some positive real v such that, for all $1 \leqslant k \leqslant n$, $\mathbb{E}[Z_k^2] \leqslant v$. Denote $S_n = b_1 Z_1 + \cdots + b_n Z_n$ for some positive real numbers $b_1, \dots, b_n$ and let $\|b\|_2 = (b_1^2 + \cdots + b_n^2)^{1/2}$. Then, for any positive x,*

$$\mathbb{P}(S_n \geqslant \|b\|_2 x) \leqslant \exp(-g(v)x^2) \tag{2.81}$$

where

$$g(v) = \begin{cases} \dfrac{|\log(v)|}{1-v^2} & \text{if } v < 1, \\ \dfrac{1}{2v} & \text{if } v \geqslant 1. \end{cases}$$

Remark 2.39. One can observe that for any positive v, $g(v) = 1/\varphi(v)$. Hence, the function g is decreasing. In particular, if $v \leqslant 1$, then $g(v) \geqslant 1/2$.

Remark 2.40. For $v \geqslant 1$, we will give a more efficient inequality in Section 2.6.

Proof. Corollary 2.38 immediately follows from inequality (2.75) applied to the sequence $X_1, \dots, X_n$ where, for all $1 \leqslant k \leqslant n$, $X_k = b_k Z_k$ with $v_k = b_k^2 v$. $\square$

2.5.2 Nonnegative random variables

We shall now focus our attention on concentration inequalities for nonnegative random variables. Throughout this subsection, we assume that $X_1, \dots, X_n$ is a finite sequence of nonnegative independent random variables with finite variances. For all $1 \leqslant k \leqslant n$, let

$$m_k = \mathbb{E}[X_k] \qquad \text{and} \qquad v_k = \mathrm{Var}(X_k).$$

Theorem 2.33 leads to the result below for the deviation on the left of sums of nonnegative random variables.

Theorem 2.41. *Let $X_1, \dots, X_n$ be a finite sequence of independent and nonnegative random variables with finite variances. Denote $S_n = X_1 + \cdots + X_n$ and $V_n = Var(S_n)$. Then, for any positive x,*

$$\mathbb{P}(S_n \leqslant \mathbb{E}[S_n] - x) \leqslant \exp\Big(-\frac{x^2}{2V_n + W_n}\Big) \tag{2.82}$$

where

$$W_n = \frac{1}{3}\sum_{k=1}^{n}\Big(\frac{m_k^2 - v_k}{m_k}\Big)_+^2. \tag{2.83}$$

Proof. For all $1 \leqslant k \leqslant n$, let Z_k be the random variable defined by $Z_k = m_k - X_k$. The sequence $Z_1, \ldots, Z_n$ satisfies the assumptions of Theorem 2.33 with $b_k = \mathbb{E}[X_k]$ and $v_k = \mathrm{Var}(X_k)$. Hence, (2.82) clearly follows from the first inequality in (2.76). □

Remark 2.42. Let $\mathscr{V}_n = \mathbb{E}[X_1^2] + \cdots + \mathbb{E}[X_n^2]$. It is not hard to see that $V_n + 3W_n \leqslant \mathscr{V}_n$. Hence, $2V_n + W_n \leqslant (5V_n + \mathscr{V}_n)/3 \leqslant 2\mathscr{V}_n$. Consequently, one can replace $2V_n + W_n$ by $2\mathscr{V}_n$ in (2.82). The inequality with the denominator $2\mathscr{V}_n$ may be found in Maurer [16].

Example 2.43. Let $X_1, \ldots, X_n$ be a finite sequence of independent random variables sharing the same Exponential $\mathscr{E}(\lambda)$ distribution with $\lambda > 0$. In that case, $W_n = 0$, which ensures that for any positive x,

$$\mathbb{P}(S_n \leqslant \mathbb{E}[S_n] - x) \leqslant \exp\Big(-\frac{x^2}{2V_n}\Big).$$

In Section 2.7, we will give more efficient inequalities for sums of independent random variables with exponential distributions. This is the reason why we give a second example below.

Example 2.44. Let $\varepsilon_1, \ldots, \varepsilon_n$ be a finite sequence of independent Poisson random variables and let $B_1, \ldots, B_n$ be a finite sequence of independent Bernoulli random variables. Assume that these two sequences are mutually independent. In addition, suppose that, for each $1 \leqslant k \leqslant n$, ε_k has the Poisson $\mathscr{P}(\lambda_k)$ distribution, while B_k has the Bernoulli $\mathscr{B}(p_k)$ distribution with $\lambda_k = k+1$ and $p_k = k/(k+1)$. For all $1 \leqslant k \leqslant n$, let X_k be the random variable defined by $X_k = B_k \varepsilon_k$. One can easily check that, for all $1 \leqslant k \leqslant n$, $\mathbb{E}[X_k] = k$, $\mathbb{E}[X_k^2] = k(k+2)$, which implies that $v_k = 2k$ and $(m_k^2 - v_k)_+ = k\max(0, k-2)$. Hence, for any $n \geqslant 2$, $V_n = n(n+1)$,

$$W_n = \frac{1}{18}(n-1)(n-2)(2n-3) \quad \text{and} \quad \mathscr{V}_n = \frac{1}{6}n(n+1)(2n+7).$$

According to Theorem 2.41, we find that for any positive x,

$$\mathbb{P}(S_n \leqslant \mathbb{E}[S_n] - x) \leqslant \exp\Big(-\frac{18x^2}{2n^3 + 27n^2 + 49n - 6}\Big)$$

Under the same assumptions, Maurer's inequality yields

$$\mathbb{P}(S_n \leqslant \mathbb{E}[S_n] - x) \leqslant \exp\Big(-\frac{3x^2}{n(2n^2 + 9n + 7)}\Big).$$

When $n = 10$, the first upper bound is equal to $\exp(-x^2/288)$, while the second upper bound is equal to $\exp(-x^2/990)$. For example, if $x = 40$, the first bound is equal to 3.866×10^{-3}, while the second bound is equal to 1.986×10^{-1}.

2.5.3 Symmetric conditions for bounded random variables

Throughout this subsection, we assume that $X_1, \ldots, X_n$ is a finite sequence of independent and centered random variables satisfying, for all $1 \leqslant k \leqslant n$,

$$|X_k| \leqslant b_k \tag{2.84}$$

almost surely, for some finite sequence $b_1, \ldots, b_n$ of positive real numbers. As usual, denote $S_n = X_1 + \cdots + X_n$. In order to state our concentration inequality for S_n, we need the elementary observation: it follows from condition (2.84) that, for all $1 \leqslant k \leqslant n$,

$$v_k = \mathrm{Var}(X_k) \leqslant b_k^2. \tag{2.85}$$

The result below is an immediate consequence of Theorem 2.33. The proof being obvious is omitted.

Theorem 2.45. *Let $X_1, \ldots, X_n$ be a finite sequence of independent and centered random variables satisfying (2.84). Denote $S_n = X_1 + \cdots + X_n$ and $V_n = Var(S_n)$. Let φ be the function defined, for v in $]0,1[$, by*

$$\varphi(v) = \frac{1 - v^2}{|\log(v)|} \tag{2.86}$$

with $\varphi(1) = 2$. Then, for any positive x,

$$\mathbb{P}(S_n \geqslant x) \leqslant \exp\Big(-\frac{x^2}{A_n}\Big) \leqslant \exp\Big(-\frac{3x^2}{5V_n + B_n}\Big) \leqslant \exp\Big(-\frac{x^2}{2B_n}\Big) \tag{2.87}$$

where

$$A_n = \sum_{k=1}^{n} b_k^2 \varphi\Big(\frac{v_k}{b_k^2}\Big) \qquad \text{and} \qquad B_n = \sum_{k=1}^{n} b_k^2.$$

Remark 2.46. Note that for any v in $]0,1[$, $\varphi(v) < 2$. Consequently, $A_n < 2B_n$ unless all the random variables $X_1, \ldots, X_n$ have the Bernoulli distribution given, for all $1 \leqslant k \leqslant n$, by $\mathbb{P}(X_k = -b_k) = \mathbb{P}(X_k = b_k) = 1/2$.

2.5.4 Asymmetric conditions for bounded random variables

This subsection is devoted to improvements of Hoeffding's inequality. We start by an improvement of Hoeffding's inequality which holds for non-centered random variables. Up to our knowledge, this result is new.

Theorem 2.47. *Let $X_1,\dots,X_n$ be a finite sequence of independent random variables. Assume that for all $1 \leqslant k \leqslant n$, one can find two constants $a_k < b_k$ such that $a_k \leqslant X_k \leqslant b_k$ almost surely. Denote $S_n = X_1 + \cdots + X_n$ and $V_n = Var(S_n)$. Then, for any positive x,*

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geqslant x) \leqslant \exp\Big(-\frac{3x^2}{D_n + 2V_n}\Big) \tag{2.88}$$

where

$$D_n = \sum_{k=1}^{n} (b_k - a_k)^2.$$

Remark 2.48. As mentioned in Remark 2.17, $D_n \geqslant 4V_n$. Consequently, Theorem 2.47 clearly improves Theorem 2.16. One can observe that $D_n > 4V_n$ except if all the random variables $X_1,\dots,X_n$ have the Bernoulli distribution given, for all $1 \leqslant k \leqslant n$, by $\mathbb{P}(X_k = a_k) = \mathbb{P}(X_k = b_k) = 1/2$.

Proof. For all $1 \leqslant k \leqslant n$, let Y_k be the random variable defined by $Y_k = X_k - \mathbb{E}[X_k]$. One can observe that $Y_1,\dots,Y_n$ is a sequence of independent and centered random variables such that, for all $1 \leqslant k \leqslant n$, $\alpha_k \leqslant Y_k \leqslant \beta_k$ almost surely with $\alpha_k = a_k - \mathbb{E}[X_k]$ and $\beta_k = b_k - \mathbb{E}[X_k]$. In addition, for all $1 \leqslant k \leqslant n$, $\alpha_k < 0$ and $\beta_k > 0$, except if $Y_k = 0$ almost surely, which means that Y_k can be removed. For all $1 \leqslant k \leqslant n$, we have $|Y_k| \leqslant c_k$ almost surely where $c_k = \max(-\alpha_k, \beta_k)$, and $v_k = \mathrm{Var}(X_k) = \mathrm{Var}(Y_k)$. Consequently, the sequence $Y_1,\dots,Y_n$ satisfies the assumptions of Theorem 2.45 which ensures that, for any positive x,

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geqslant x) \leqslant \exp\Big(-\frac{x^2}{A_n}\Big) \quad \text{where} \quad A_n = \sum_{k=1}^{n} c_k^2 \varphi\Big(\frac{v_k}{c_k^2}\Big) \tag{2.89}$$

Furthermore, we deduce from Lemma 2.37 that

$$A_n \leqslant \frac{1}{3}\sum_{k=1}^{n}\Big(c_k^2 + 4v_k + \frac{v_k^2}{c_k^2}\Big) = \frac{1}{3}\sum_{k=1}^{n}\Big(2v_k + \Big(c_k + \frac{v_k}{c_k}\Big)^2\Big). \tag{2.90}$$

In order to prove (2.88), it only remains to show that $\Delta_n \leqslant D_n$ where

$$\Delta_n = \sum_{k=1}^{n}\Big(c_k + \frac{v_k}{c_k}\Big)^2.$$

According to Lemma 2.18, $\mathrm{Var}(Y_k) \leqslant -\alpha_k\beta_k = \min(-\alpha_k, \beta_k)\max(-\alpha_k, \beta_k)$, which implies that

$$\Delta_n \leqslant \sum_{k=1}^{n}\big(\max(-\alpha_k,\beta_k) + \min(-\alpha_k,\beta_k)\big)^2 = \sum_{k=1}^{n}(\beta_k - \alpha_k)^2 = D_n. \tag{2.91}$$

Finally, (2.88) follows from (2.89), (2.90) and (2.91), which achieves the proof of Theorem 2.47. □

To conclude this subsection, we give an improvement of Hoeffding's inequality due to Kearns and Saul [12]. This improvement takes into account the centering of the random variables $X_1,\dots,X_n$.

Theorem 2.49. *Let $X_1,\dots,X_n$ be a finite sequence of independent random variables and let $S_n = X_1+\cdots+X_n$. Assume that for all $1 \leqslant k \leqslant n$, one can find two constants $a_k < b_k$ such that $a_k \leqslant X_k \leqslant b_k$ almost surely. For all $1 \leqslant k \leqslant n$, denote $m_k = \mathbb{E}[X_k]$,*

$$p_k = (b_k - m_k)^2 \varphi\Big(\frac{m_k - a_k}{b_k - m_k}\Big) \quad \text{and} \quad q_k = \frac{(b_k - a_k)(b_k + a_k - 2m_k)}{\log\big((b_k - m_k)/(m_k - a_k)\big)}$$

with the convention that $q_k = (b_k - a_k)^2/2$ if $m_k = (a_k + b_k)/2$, where φ is the function given by (2.74). Then, for any positive x,

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geqslant x) \leqslant \exp\Big(-\frac{x^2}{P_n}\Big) \leqslant \exp\Big(-\frac{x^2}{Q_n}\Big) \tag{2.92}$$

where

$$P_n = \sum_{k=1}^n p_k \quad \text{and} \quad Q_n = \sum_{k=1}^n q_k.$$

Remark 2.50. One can observe that, for all $1 \leqslant k \leqslant n$, $q_k < (b_k - a_k)^2/2$ as soon as $m_k \neq (a_k + b_k)/2$. Note also that the coefficients q_k are symmetric functions of (a_k, b_k). Hence, we obtain that, for any positive x,

$$\mathbb{P}(|S_n - \mathbb{E}[S_n]| \geqslant x) \leqslant 2\exp\Big(-\frac{x^2}{Q_n}\Big).$$

Proof. We shall proceed as in the proof of Theorem 2.47. We already saw that, for all $1 \leqslant k \leqslant n$, $Y_k \leqslant \beta_k$ almost surely and $\mathrm{Var}(Y_k) \leqslant -\alpha_k\beta_k$ with $\alpha_k = a_k - \mathbb{E}[X_k]$ and $\beta_k = b_k - \mathbb{E}[X_k]$. Then, we immediately deduce from inequality (2.75) that, for any positive x,

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geqslant x) \leqslant \exp\Big(-\frac{x^2}{P_n}\Big).$$

It only remains to prove that $P_n \leqslant Q_n$. If $-\alpha_k \leqslant \beta_k$, then $p_k = q_k$ and there is nothing to prove. On the other side, if $-\alpha_k > \beta_k$, we may apply the elementary inequality $\log(x) \leqslant (x - 1/x)/2$, which holds if $x > 1$, to $x = -\alpha_k/\beta_k$. This inequality ensures that

$$\log\Big(-\frac{\alpha_k}{\beta_k}\Big) \leqslant \frac{\beta_k^2 - \alpha_k^2}{2\alpha_k\beta_k},$$

leading to $p_k = -2\alpha_k\beta_k \leqslant q_k$. Finally, we have shown that $P_n \leqslant Q_n$, which completes the proof of Theorem 2.49 □

2.6 Always a little further on weighted sums

The goal of this section is to go a little bit further on concentration inequalities for weighted sums. Let $Z_1, \ldots, Z_n$ be a finite sequence of independent and centered random variables with finite Laplace transform on a right neighborhood of the origin. More precisely, we will assume that there exists a convex and increasing function ℓ from $[0, +\infty[$ to $[0, \infty]$ such that $\ell(0) = \ell'(0) = 0$ and, for all $1 \leqslant k \leqslant n$, and for any $t \geqslant 0$,

$$\log \mathbb{E}\big[\exp(tZ_k)\big] \leqslant \ell(t). \tag{2.93}$$

Our aim is to establish one-sided deviation inequalities for $S_n = b_1 Z_1 + \cdots + b_n Z_n$, for some positive real numbers $b_1, \ldots, b_n$. As usual, we denote for all $p \geqslant 1$,

$$\|b\|_p = \big(b_1^p + b_2^p + \cdots + b_n^p\big)^{1/p}$$

and $\|b\|_\infty = \max(b_1, b_2, \ldots, b_n)$. One can notice that there are only a few results on that direction. In a paper devoted to McDiarmid's inequality, Rio [20] uses the concavity of ℓ' to establish an upper bound for $\mathbb{P}(S_n \geqslant x)$. Below, we give a more general version of this inequality and another inequality for functions ℓ with a convex derivative.

Theorem 2.51. *Let $Z_1, \ldots, Z_n$ be a finite sequence of independent and centered random variables such that, for all $1 \leqslant k \leqslant n$, the random variable Z_k satisfies (2.93). Denote $S_n = b_1 Z_1 + \cdots + b_n Z_n$ for some positive real numbers $b_1, \ldots, b_n$.*

1) If the function ℓ has a concave derivative, then, for any positive x,

$$\mathbb{P}\big(S_n \geqslant x\big) \leqslant \exp\Big(-\frac{\|b\|_1^2}{\|b\|_2^2}\,\ell^*\Big(\frac{x}{\|b\|_1}\Big)\Big), \tag{2.94}$$

where ℓ^* stands for the Legendre-Fenchel dual of ℓ.

2) If the function h defined, for any positive t, by $h(t) = \ell'(t)/t$, is nondecreasing on $\mathbb{R}^+$, then, for any positive x,

$$\mathbb{P}\Big(S_n \geqslant x\Big) \leqslant \exp\Big(-\frac{\|b\|_2^2}{\|b\|_\infty^2}\,\ell^*\Big(\frac{\|b\|_\infty x}{\|b\|_2^2}\Big)\Big). \tag{2.95}$$

Remark 2.52. If the function ℓ has a convex derivative, then h is nondecreasing, and consequently (2.95) holds true.

Proof. We clearly obtain from (2.93) that, for any positive t,

$$\log \mathbb{E}\big[\exp(tS_n)\big] \leqslant \ell(b_1 t) + \cdots + \ell(b_n t). \tag{2.96}$$

Starting from (2.96), we now prove (2.95). As the function h is nondecreasing, we have for any $1 \leqslant k \leqslant n$,

$$\ell(b_k t) = \int_0^t b_k \ell'(b_k x)dx \leqslant \int_0^t \frac{b_k^2}{\|b\|_\infty} \ell'(\|b\|_\infty x)dx = \frac{b_k^2}{\|b\|_\infty^2} \ell(\|b\|_\infty t),$$

which ensures that, for any positive t,

$$\ell(b_1 t) + \cdots + \ell(b_n t) \leqslant \frac{\|b\|_2^2}{\|b\|_\infty^2} \ell(\|b\|_\infty t). \tag{2.97}$$

It follows from Markov's inequality together with (2.96) and (2.97) that, for any positive t,

$$\log \mathbb{P}(S_n \geqslant x) \leqslant -xt + \frac{\|b\|_2^2}{\|b\|_\infty^2} \ell(\|b\|_\infty t) = -\frac{\|b\|_2^2}{\|b\|_\infty^2} \Big(\frac{\|b\|_\infty xs}{\|b\|_2^2} - \ell(s) \Big)$$

where $s = \|b\|_\infty t$. Hence, we immediately deduce (2.95) taking the infimum over all positive reals s. We now proceed to the proof of (2.94). Since $\ell(0) = 0$, we have for any positive x,

$$\ell(b_1 x) + \cdots + \ell(b_n x) = \int_0^x \big(b_1 \ell'(b_1 t) + \cdots + b_n \ell'(b_n t) \big) dt.$$

Next, the concavity of ℓ' implies that

$$b_1 \ell'(b_1 t) + \cdots + b_n \ell'(b_n t) \leqslant \|b\|_1 \ell' \Big(\frac{\|b\|_2^2 t}{\|b\|_1} \Big).$$

Hence, for any positive t,

$$\ell(b_1 t) + \cdots + \ell(b_n t) \leqslant \frac{\|b\|_1^2}{\|b\|_2^2} \ell \Big(\frac{\|b\|_2^2 t}{\|b\|_1} \Big). \tag{2.98}$$

Therefore, we infer from Markov's inequality together with (2.96) and (2.98) that, for any positive t,

$$\log \mathbb{P}(S_n \geqslant x) \leqslant -xt + \frac{\|b\|_1^2}{\|b\|_2^2} \ell \Big(\frac{\|b\|_2^2 t}{\|b\|_1} \Big) = -\frac{\|b\|_1^2}{\|b\|_2^2} \Big(\frac{xs}{\|b\|_1} - \ell(s) \Big)$$

where $s = (\|b\|_2^2 / \|b\|_1) t$. Finally, we obtain (2.94) taking once again the infimum over all positive reals s. □

Example 2.53. Let $Z_1, \ldots, Z_n$ be a finite sequence of independent and centered random variables such that, for all $1 \leqslant k \leqslant n$, $Z_k \leqslant 1$ almost surely. Assume that there exists some real number $v \geqslant 1$ such that, for all $1 \leqslant k \leqslant n$, $\mathbb{E}[Z_k^2] \leqslant v$. Then, it follows from Lemma 2.23 that, for all $1 \leqslant k \leqslant n$ and for any positive t,

$$\log \mathbb{E}[\exp(tZ_k)] \leqslant L_v(t) \qquad \text{where} \qquad L_v(t) = \log(ve^t + e^{-vt}) - \log(1+v).$$

For any positive t, denote $f(t) = \log(ve^t + 1) - \log(1+v)$ with this notation, we have $L_v(t) = f((1+v)t) - vt$. Consequently, the function L_v' is concave on $\mathbb{R}^+$ if and only if f' is concave on $\mathbb{R}^+$. However, it is not hard to see that for any positive t,

$$f^{(3)}(t) = ve^t(1-ve^t)(1+ve^t)^{-3} < 0.$$

Hence, f' and L_v' are concave. Hence, we deduce from (2.55) and (2.94) that, for any x in $[0,1]$,

$$\mathbb{P}(S_n \geqslant \|b\|_1 x) \leqslant \exp\Big(-\frac{\|b\|_1^2}{\|b\|_2^2}\Big(\frac{v+x}{v+1}\log\Big(1+\frac{x}{v}\Big) + \frac{1-x}{1+v}\log(1-x)\Big)\Big). \tag{2.99}$$

For example, assume that the random variables $Z_1, \ldots, Z_n$ are with values in $[-v, 1]$ for some $v \geqslant 1$. Then, it follows from Lemma 2.18 that, for all $1 \leqslant k \leqslant n$, $\mathbb{E}[Z_k^2] \leqslant v$, which means that the above inequality holds true.

Example 2.54. Let $Z_1, \ldots, Z_n$ be a finite sequence of independent and centered random variables satisfying a Bernstein's type condition: there exists some real v in $]0,1]$ such that, for any $1 \leqslant k \leqslant n$, $\mathbb{E}[Z_k^2] \leqslant v$ and, for any integer $p \geqslant 3$,

$$\mathbb{E}[(\max(0, Z_k))^p] \leqslant \frac{p!v}{2}.$$

According to (2.14) with $c = 1$, we have for all $1 \leqslant k \leqslant n$ and for any t in $[0,1]$,

$$\log \mathbb{E}[\exp(tZ_k)] \leqslant \ell_v(t) \qquad \text{where} \qquad \ell_v(t) = \log\Big(1 + \frac{vt^2}{2(1-t)}\Big).$$

Hereafter, let h_v be the function defined, for any positive t, by $h_v(t) = \ell_v'(t)/t$. It is not hard to see that

$$\log(h_v(t)) = \log(v/2) + \log(2-t) - \log(1-t) - \log(1-t+vt^2/2)$$

which leads to

$$\log(h_v(t))' = \frac{1}{(2-t)(1-t)} + \frac{1-vt}{1-t+vt^2/2} > 0$$

for any t in $[0,1]$, since $vt \leqslant 1$ for all t in $[0,1]$. Consequently, we obtain from (2.15) and (2.95) in the particular case $\|b\|_\infty = 1$ that, for any positive x,

$$\mathbb{P}(S_n \geqslant \|b\|_2^2 x) \leqslant \exp\Big(-\|b\|_2^2\Big(\frac{x^2}{v+x} - \log\Big(1 + \frac{x^2}{2(v+x)}\Big)\Big)\Big).$$

This inequality cannot be deduced from Theorem 2.1.

2.7 Sums of Gamma random variables

In this section, we are interested in deviation inequalities for sums of independent nonnegative random variables with Gamma distributions. Let a and b be some positive real numbers. A random variable X has the $\Gamma(a,b)$ distribution if its probability density function f_X is given by

$$f_X(x) = \frac{x^{a-1}\exp(-x/b)}{b^a\Gamma(a)} \mathrm{I}_{\{x>0\}} \tag{2.100}$$

where

$$\Gamma(a) = \int_0^\infty x^{a-1}\exp(-x)\,dx$$

stands for the Euler's Gamma function. From the definition, if X has the $\Gamma(a,b)$ distribution, then X/b has the $\Gamma(a,1)$ distribution. One can observe that the $\Gamma(1,b)$ distribution coincides with the Exponential $\mathscr{E}(1/b)$ distribution. Moreover, if Z is distributed as a normal $\mathscr{N}(0,1)$ random variable, then Z^2 has the $\Gamma(1/2,2)$ distribution. Consequently, sums of Gamma random variables include sums of exponential random variables as well as sums of weighted chi-square random variables.

As shown by the lemma below, the Gamma random variables have a nice Laplace transform.

Lemma 2.55. *Let X be a random variable with $\Gamma(a,b)$ distribution. Then, for any real t,*

$$\log \mathbb{E}[\exp(tX)] = a\ell(bt) \tag{2.101}$$

where ℓ is the strictly convex function given by

$$\ell(t) = \begin{cases} -\log(1-t) & \text{if } t < 1, \\ +\infty & \text{if } t \geqslant 1. \end{cases}$$

Remark 2.56. We immediately deduce from Lemma 2.55 that $\mathbb{E}[X] = ab\ell'(0) = ab$ and $\mathrm{Var}(X) = ab^2\ell''(0) = ab^2$.

Proof. It is enough to prove Lemma 2.55 in the special case $b = 1$. It follows from (2.100) that for any real t,

$$\mathbb{E}[\exp(tX)] = \frac{1}{\Gamma(a)}\int_0^\infty x^{a-1}\exp(-x(1-t))\,dx.$$

From the above equality, $\mathbb{E}[\exp(tX)] = \infty$ if $t \geqslant 1$. Moreover, if $t < 1$, we obtain via the change of variables $y = x(1-t)$ that

$$\mathbb{E}[\exp(tX)] = \frac{(1-t)^{-a}}{\Gamma(a)}\int_0^\infty y^{a-1}\exp(-y)\,dy = (1-t)^{-a},$$

which proves Lemma 2.55. $\square$

Hereafter, let $X_1, \ldots X_n$ be a finite sequence of independent random variables such that, for all $1 \leqslant k \leqslant n$, X_k has the $\Gamma(a_k, b_k)$ distribution, where a_k and b_k are positive real numbers. Define

$$\|b\|_{1,a} = \sum_{k=1}^{n} a_k b_k \quad \text{and} \quad \|b\|_{2,a} = \Big(\sum_{k=1}^{n} a_k b_k^2\Big)^{1/2}. \tag{2.102}$$

We now state our concentration inequalities for sums of Gamma random variables.

Theorem 2.57. *Let $X_1, \ldots X_n$ be a finite sequence of independent random variables such that, for all $1 \leqslant k \leqslant n$, X_k has the $\Gamma(a_k, b_k)$ distribution. Let $S_n = X_1 + \cdots + X_n$. Then, for any positive x,*

$$\mathbb{P}(S_n \geqslant \|b\|_{1,a} + x\|b\|_{2,a}^2) \leqslant \exp\Big(-\frac{\|b\|_{2,a}^2}{\|b\|_\infty^2}\Big(\|b\|_\infty x - \log(1 + \|b\|_\infty x)\Big)\Big) \tag{2.103}$$

$$\leqslant \exp\Big(-\frac{\|b\|_{2,a}^2 x^2}{1 + \|b\|_\infty x + \sqrt{1 + 2\|b\|_\infty x}}\Big). \tag{2.104}$$

In addition, for any x in $]0,1[$,

$$\mathbb{P}(S_n \leqslant \|b\|_{1,a} - x\|b\|_{1,a}) \leqslant \exp\Big(\frac{\|b\|_{1,a}^2}{\|b\|_{2,a}^2}\Big(\log(1-x) + x\Big)\Big), \tag{2.105}$$

$$\leqslant \exp\Big(-\frac{\|b\|_{1,a}^2}{\|b\|_{2,a}^2}\frac{x^2}{2}\Big). \tag{2.106}$$

Remark 2.58. Note from Remark 2.56 that $\mathbb{E}[S_n] = \|b\|_{1,a}$ and $\mathrm{Var}(S_n) = \|b\|_{2,a}^2$. Consequently, Theorem 2.57 provides an exponential concentration inequality for S_n around its mean with the adequate variance term. Moreover, (2.104) is a Bernstein's type inequality, which may be found in the monograph of Boucheron, Lugosi, and Massart [7]. We now compare inequalities (2.103) and (2.104), as well as inequalities (2.105) and (2.106) in the particular case where the random variables $X_1, \ldots X_n$ share the same $\Gamma(a, 1)$ distribution for some positive a. Figure 2.6 compares the four rate functions

$$\Phi_R(x) = x - \log(1+x) \quad \text{and} \quad \Psi_R(x) = \frac{x^2}{1 + x + \sqrt{1+2x}},$$

$$\Phi_L(x) = -x - \log(1-x) \quad \text{and} \quad \Psi_L(x) = \frac{x^2}{2}$$

for x in the interval $]0, 4/5]$ while Figure 2.7 compares the rate functions Φ_L and Ψ_L for x in the interval $[4/5, 1[$. One can see in Figure 2.6 that (2.103) and (2.105) outperform (2.104) and (2.106), respectively. It is even more spectacular in Figure 2.7 that (2.105) is much more accurate than (2.106).

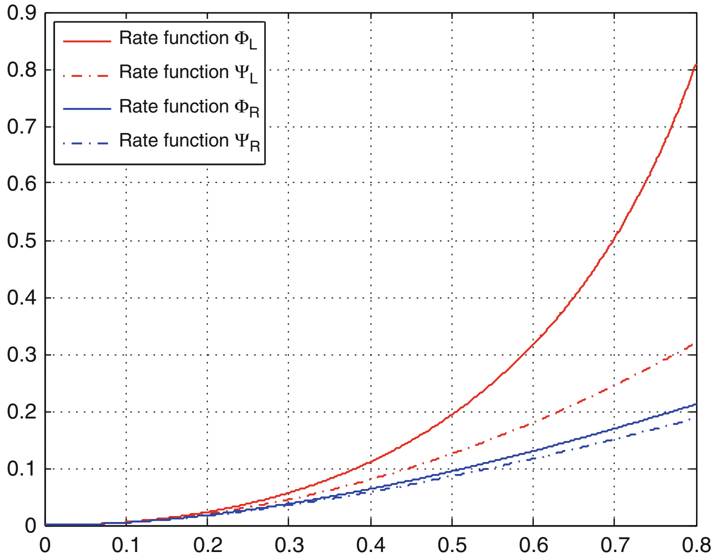

Fig. 2.6 Comparisons in Gamma concentration inequalities

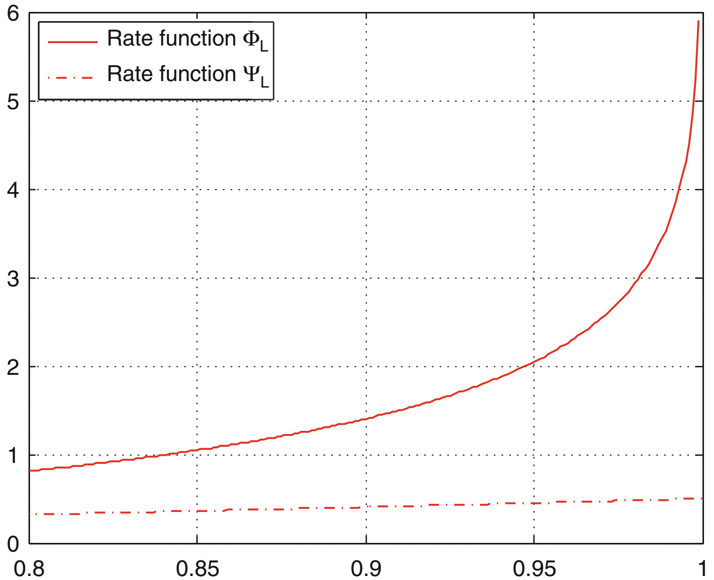

Fig. 2.7 Comparisons in Gamma concentration inequalities on the left side

Remark 2.59. Assume that $\|b\|_\infty = 1$. Starting from inequality (2.103) and using the upper bound

$$\Phi_R^{-1}(x) \leqslant \Lambda(x) = x + \log(1+x+\sqrt{2x}) + \frac{\log(1+x+\sqrt{2x}) - \sqrt{2x}}{x+\sqrt{2x}} \tag{2.107}$$

given in Del Moral and Rio [9], one can prove that, for any positive x,

$$\mathbb{P}\big(S_n \geqslant \|b\|_{1,a} + \Lambda(x)\|b\|_{2,a}^2\big) \leqslant \exp\big(-\|b\|_{2,a}^2 x\big). \tag{2.108}$$

Since $\Lambda(x) < x + \sqrt{2x}$, (2.108) is more efficient than the reversed form of (2.104).

Proof. We shall prove (2.103) in the particular case $\|b\|_\infty = 1$, inasmuch as the general case follows by dividing the initial random variables by $\|b\|_\infty$. Let ℓ_c be the convex function defined, for any real t, by $\ell_c(t) = \ell(t) - t$. We infer from Lemma 2.55 that, for any real t,

$$\log \mathbb{E}[\exp(t(S_n - \mathbb{E}[S_n]))] \leqslant \sum_{k=1}^n a_k \ell_c(b_k t). \tag{2.109}$$

Moreover, let h_c be the function defined, for any positive t, by $h_c(t) = \ell_c(t)/t^2$. The function h_c is increasing on $]0, +\infty[$. Hence, for any positive t, $\ell_c(b_k t) \leqslant b_k^2 \ell_c(t)$, which implies that

$$\log \mathbb{E}[\exp(t(S_n - \mathbb{E}[S_n]))] \leqslant \|b\|_{2,a}^2 \ell_c(t). \tag{2.110}$$

We deduce from Markov's inequality that for any positive x and for any t in $]0,1[$,

$$\log \mathbb{P}(S_n - \mathbb{E}[S_n] \geqslant x\|b\|_{2,a}^2) \leqslant -\|b\|_{2,a}^2 (xt - \ell_c(t)). \tag{2.111}$$

The optimal value t in the above inequality is given by the elementary equation $\ell_c'(t) = t/(1-t) = x$, leading to $t = x/(1+x)$. By taking this value of t, we find that

$$\ell_c^*(x) = xt - \ell_c(t) = (x+1)t + \log(1-t) = x - \log(1+x), \tag{2.112}$$

which, via (2.111), achieves the proof of (2.103). One can observe from (2.112) that, for any $x \geqslant 0$, $\ell_c^*(x) = \ell_c(-x)$. From the reflexivity properties of the Legendre-Fenchel dual, this equality also holds true for any $x \leqslant 0$. The proof of (2.104) immediately follows from the fact that, for any positive t, $\ell_c(t) < \varphi(t) = t^2/(2-2t)$, which implies that for any positive x,

$$\ell_c^*(x) > \varphi^*(x) = \frac{x^2}{1+x+\sqrt{1+2x}}.$$

We are in position to prove (2.105). It follows from (2.109) that, for any positive t,

$$\log \mathbb{E}[\exp(t(\mathbb{E}[S_n] - S_n))] \leqslant L(t) \tag{2.113}$$

where

$$L(t) = \sum_{k=1}^{n} a_k \ell_c(-b_k t) = \sum_{k=1}^{n} a_k \ell_c^*(b_k t).$$

According to (2.112), $(\ell_c^*)'$ is a concave function as $(\ell_c^*)'(t) = t/(1+t)$. It ensures that, for any positive t,

$$L'(t) = \sum_{k=1}^{n} a_k b_k \Big(\frac{b_k t}{1+b_k t}\Big) \leqslant \|b\|_{1,a} \Big(\frac{\|b\|_{2,a}^2 t}{\|b\|_{1,a} + \|b\|_{2,a}^2 t}\Big).$$

Integrating this inequality, we obtain that, for any positive t,

$$L(t) \leqslant \frac{\|b\|_{1,a}^2}{\|b\|_{2,a}^2} \ell_c^* \Big(\frac{\|b\|_{2,a}^2 t}{\|b\|_{1,a}}\Big). \tag{2.114}$$

Therefore, we find from (2.113) and (2.114) together with Markov's inequality that for any positive x and for any positive t,

$$\begin{aligned} \log \mathbb{P}(\mathbb{E}[S_n] - S_n \geqslant \|b\|_{1,a} x) &\leqslant -\frac{\|b\|_{1,a}^2}{\|b\|_{2,a}^2} \Big(\frac{\|b\|_{2,a}^2 x t}{\|b\|_{1,a}} - \ell_c^* \Big(\frac{\|b\|_{2,a}^2 t}{\|b\|_{1,a}}\Big)\Big), \\ &\leqslant -\frac{\|b\|_{1,a}^2}{\|b\|_{2,a}^2} \Big(\frac{\|b\|_{2,a}^2 (x-1) t}{\|b\|_{1,a}} + \log\Big(1 + \frac{\|b\|_{2,a}^2 t}{\|b\|_{1,a}}\Big)\Big). \end{aligned}$$

In view of the above inequality, it is necessary to assume that x belongs to $]0,1[$. By taking the optimal $t = \|b\|_{1,a} x / (\|b\|_{2,a}^2 (1-x))$ in this inequality, we find that

$$\log \mathbb{P}(\mathbb{E}[S_n] - S_n \geqslant \|b\|_{1,a} x) \leqslant -\frac{\|b\|_{1,a}^2}{\|b\|_{2,a}^2} \ell_c(x),$$

which clearly leads to (2.105). Finally, (2.106) immediately follows from (2.105), which completes the proof of Theorem 2.57. □

2.8 McDiarmid's inequality

This section is devoted to the so-called McDiarmid inequality for functions of independent random variables. First of all, we recall the usual version of this inequality. Let $(E_1, d_1), \ldots, (E_n, d_n)$ be a finite sequence of separable metric spaces with respective finite diameters $c_1, \ldots, c_n$. Denote E^n the product space $E^n = E_1 \times \cdots \times E_n$.

Definition 2.60. A function f from E^n into $\mathbb{R}$ is said to be separately 1-Lipschitz if, for any x, y in E^n with $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$,

$$|f(x_1, \ldots, x_n) - f(y_1, \ldots, y_n)| \leqslant \sum_{k=1}^{n} d_k(x_k, y_k). \tag{2.115}$$

Let $X_1,\dots,X_n$ be a finite sequence of independent random variables such that the random vector $(X_1,\dots,X_n)$ takes its values in E^n. McDiarmid's inequality says that for any separately 1-Lipschitz function f, the random variable

$$Z = f(X_1,\dots,X_n) \tag{2.116}$$

satisfies the concentration inequality given, for any positive x, by

$$\mathbb{P}(Z-\mathbb{E}[Z]\geqslant x)\leqslant \exp\Big(-\frac{2x^2}{C_n}\Big) \quad \text{where} \quad C_n=\sum_{k=1}^{n}c_k^2. \tag{2.117}$$

This inequality was obtained by McDiarmid [17, 18]. We refer to Exercise 9 in Chapter 3 for the proof of this inequality, which uses a martingale method due to Yurinskii [23].

Remark 2.61. If the space E^n is countable and the distances $d_1,\dots,d_n$ are defined, for all $1\leqslant k\leqslant n$, by $d_k(x_k,y_k)=c_k$ if $x_k\neq y_k$, then (2.115) is equivalent, for any x,y in E^n with $x=(x_1,\dots,x_n)$ and $y=(y_1,\dots,y_n)$, to

$$|f(x_1,\dots,x_n)-f(y_1,\dots,y_n)|\leqslant \sum_{k=1}^{n}c_k\mathrm{I}_{x_k\neq y_k}.$$

It ensures that f is uniformly bounded. In that case, the optimal reals c_k in the above inequality are

$$c_k=\sup_{(x,y_k,z_k)}|f(x_1,\dots,x_{k-1},y_k,x_{k+1},\dots,x_n)-f(x_1,\dots,x_{k-1},z_k,x_{k+1},\dots,x_n)|.$$

We now propose an improvement of McDiarmid's inequality in the style of Delyon [10]: instead of assuming a uniform bound on each oscillation, we only assume a bound on the sum of squares. For all $1\leqslant k\leqslant n$, denote by $\mathscr{F}^{(k)}$ the σ-algebra generated by $X_1,\dots,X_n$ except X_k,

$$\mathscr{F}^{(k)}=\sigma(X_1,\dots,X_{k-1},X_{k+1},\dots,X_n).$$

Theorem 2.62. *Let $X_1,\dots,X_n$ be a finite sequence of independent random variables and let Z be a measurable function of $X_1,\dots,X_n$. Assume that for each $1\leqslant k\leqslant n$, there exist two $\mathscr{F}^{(k)}$-measurable bounded random variables A_k and B_k such that*

$$A_k\leqslant Z\leqslant B_k \qquad a.s. \tag{2.118}$$

Then, for any positive x,

$$\mathbb{P}(Z\geqslant \mathbb{E}[Z]+x)\leqslant \exp\Big(-\frac{2x^2}{D_n}\Big) \quad \textit{where} \quad D_n=\Big\|\sum_{k=1}^{n}(B_k-A_k)^2\Big\|_\infty. \tag{2.119}$$

Remark 2.63. Assume that the space E^n is countable and $Z = f(X_1,\dots,X_n)$ where f is a separately 1-Lipschitz function on E^n. For all $1 \leqslant k \leqslant n$, denote

$$A_k = \inf_{x_k \in E_k} f(X_1,\dots,X_{k-1},x_k,X_{k+1},\dots,X_n),$$

and

$$B_k = \sup_{x_k \in E_k} f(X_1,\dots,X_{k-1},x_k,X_{k+1},\dots,X_n).$$

Then, we clearly have

$$B_k - A_k \leqslant c_k \qquad \text{a.s.}$$

which shows that $D_n \leqslant C_n$. It means that Theorem 2.62 improves McDiarmid's inequality.

The proof is based on Log-Sobolev type inequalities, which have been widely developed by Ledoux [14]. The following lemma, due to Boucheron, Lugosi, and Massart [8], will be the main tool in the proof of our result.

Lemma 2.64. *Let $X_1,\dots,X_n$ be a finite sequence of independent random variables and let Z be a measurable function of $X_1,\dots,X_n$. Let $Z^{(1)},\dots,Z^{(n)}$ be any finite sequence of real bounded random variables such that, for each $1 \leqslant k \leqslant n$, $Z^{(k)}$ is $\mathscr{F}^{(k)}$-measurable. Denote by φ the function defined, for any real x, by $\varphi(x) =$* $\exp(-x)+x-1$. *Then, for any positive t,*

$$\mathbb{E}[tZe^{tZ}] - \mathbb{E}[e^{tZ}]\log\big(\mathbb{E}[e^{tZ}]\big) \leqslant \sum_{k=1}^{n} \mathbb{E}\big[e^{tZ}\varphi(tZ - tZ^{(k)})\big]. \tag{2.120}$$

Proof. We shall only prove Lemma 2.64 in the particular case $t = 1$, as the general case follows by multiplying Z by t. Let $\mathscr{F}_0$ be the trivial σ-algebra $\mathscr{F}_0 = \{\emptyset,\Omega\}$, and for all $1 \leqslant k \leqslant n$, denote $\mathscr{F}_k = \sigma(X_1,\dots,X_k)$ and $Z_k = \log\big(\mathbb{E}[e^Z|\mathscr{F}_k]\big)$. We have by a standard telescopic argument

$$\mathbb{E}[Ze^Z] - \mathbb{E}[e^Z]\log\mathbb{E}[e^Z] = \sum_{k=1}^{n} \mathbb{E}\big[e^Z(Z_k - Z_{k-1})\big]. \tag{2.121}$$

The right-hand side of (2.121) can be rewritten as

$$\mathbb{E}\big[e^Z(Z_k - Z_{k-1})\big] = \mathbb{E}\big[e^{Z^{(k)}}e^{Z-Z^{(k)}}(Z_k - Z_{k-1})\big].$$

Then, it follows from the Young type inequality $xy \leqslant x\log(x) - x + \exp(y)$ with $x > 0$ and y in $\mathbb{R}$, applied to $x = \exp(Z - Z^{(k)})$ and $y = Z_k - Z_{k-1}$, that

$$\mathbb{E}\big[e^Z(Z_k - Z_{k-1})\big] \leqslant \mathbb{E}\big[e^Z(Z - Z^{(k)}) - e^Z + e^{Z^{(k)}+Z_k-Z_{k-1}}\big].$$

However, the independence of the underlying random variables implies that the random variable $\mathbb{E}[e^{Z^{(k)}}|\mathscr{F}_k]$ is $\mathscr{F}_{k-1}$-measurable. Hence,

$$\begin{aligned}\mathbb{E}\big[e^{Z^{(k)}+Z_k-Z_{k-1}}\big] &= \mathbb{E}\big[\mathbb{E}[e^{Z^{(k)}}|\mathscr{F}_k]e^{Z_k-Z_{k-1}}\big],\\ &= \mathbb{E}\big[\mathbb{E}[e^{Z^{(k)}}|\mathscr{F}_{k-1}]e^{Z_k-Z_{k-1}}\big],\\ &= \mathbb{E}\big[e^{Z^{(k)}}\big].\end{aligned}$$

Consequently,

$$\mathbb{E}[e^Z(Z_k - Z_{k-1})] \leqslant \mathbb{E}\big[e^Z(Z - Z^{(k)}) - e^Z + e^{Z^{(k)}}\big],$$

which, together with (2.121), leads to (2.120). $\square$

Proof of Theorem 2.62. For any positive t, denote $F(t) = \mathbb{E}[\exp(tZ)]$ and $L(t) = \log(F(t))$. It follows from Lemma 2.64 that, for any positive t,

$$F(t)(tL'(t) - L(t)) \leqslant \sum_{k=1}^{n} \mathbb{E}[\exp(tZ)\varphi(tZ - tZ^{(k)})]. \tag{2.122}$$

Now, the function φ is convex. Since the random variable Z belongs to $[A_k, B_k]$ almost surely, it implies that

$$\varphi(tZ - tZ^{(k)}) \leqslant \max\big(\varphi(tA_k - tZ^{(k)}), \varphi(tB_k - tZ^{(k)})\big).$$

Consequently, it is natural to choose $Z^{(k)}$ in such a way that

$$\varphi(tA_k - tZ^{(k)}) = \varphi(tB_k - tZ^{(k)}).$$

The solution of this equation is given by

$$Z^{(k)} = \frac{1}{t}\Big(tA_k + \log(tC_k) - \log(1 - \exp(-tC_k))\Big)$$

where $C_k = B_k - A_k$. For this choice of $Z^{(k)}$, we find that for any positive t,

$$\varphi(tZ - tZ^k) \leqslant \ell(tC_k) \tag{2.123}$$

where the function ℓ is defined by $\ell(0) = 0$ and, for any $x \neq 0$,

$$\begin{aligned}\ell(x) &= \frac{x}{1-\exp(-x)} + \log\Big(\frac{1-\exp(-x)}{x}\Big) - 1,\\ &= \frac{x\exp(x)}{\exp(x)-1} - x + \log\Big(\frac{\exp(x)-1}{x}\Big) - 1,\\ &= \frac{x}{\exp(x)-1} + \log\Big(\frac{\exp(x)-1}{x}\Big) - 1.\end{aligned} \tag{2.124}$$

One can observe that $\ell(-x) = \ell(x)$. We now claim that, for any real x,

$$\ell(x) \leqslant \frac{x^2}{8}. \tag{2.125}$$

As a matter of fact, denote by h_r the function defined, for any r in $[0,1]$ and for any real x, by

$$h_r(x) = \log\left(re^{(1-r)x} + (1-r)e^{-rx}\right).$$

The function h_r is convex with respect to x and concave with respect to r. Now, for any $x \neq 0$, its maximum with respect to r is attained for

$$r_x = \frac{e^x - x - 1}{xe^x - x}.$$

Therefore,

$$\sup_{r\in[0,1]} h_r(x) = h_{r_x}(x) = \ell(x). \tag{2.126}$$

However, for any r in $[0,1]$, h_r is the log-Laplace transform of a centered random variable ε with two-value distribution given by $\mathbb{P}(\varepsilon = 1-r) = r$ and $\mathbb{P}(\varepsilon = -r) = 1-r$. We immediately deduce from Lemma 2.19 that, for any r in $[0,1]$ and for any real x, $h_r(x) \leqslant x^2/8$. Consequently, (2.126) clearly leads to (2.125). Hereafter, it follows from the conjunction of (2.122), (2.123), and (2.125) that, for any positive t,

$$F(t)(tL'(t) - L(t)) \leqslant \frac{t^2}{8}\mathbb{E}\Big[\exp(tZ)\sum_{k=1}^{n} C_k^2\Big] \leqslant \frac{t^2}{8} D_n F(t),$$

which implies that

$$\frac{tL'(t) - L(t)}{t^2} \leqslant \frac{D_n}{8}.$$

Integrating this inequality, we obtain that, for any positive t,

$$\frac{L(t)}{t} - L'(0) \leqslant \frac{tD_n}{8},$$

leading to

$$\log\mathbb{E}\left[\exp(tZ)\right] \leqslant t\mathbb{E}[Z] + D_n\frac{t^2}{8}. \tag{2.127}$$

Finally, we infer from Markov's inequality and (2.127) that for any positive x and for any positive t,

$$\log\mathbb{P}(Z \geqslant \mathbb{E}[Z] + x) \leqslant -tx + \frac{D_n t^2}{8}.$$

By taking the optimal value $t = 4x/D_n$ in this inequality, we immediately obtain (2.119), which achieves the proof of Theorem 2.62. □

2.9 Complements and Exercises

Exercise 1. Let X be a real-valued random variable with finite Laplace transform on a right neighborhood of the origin. Denote $L_X(t) = \log \mathbb{E}[\exp(tX)]$. Let Ψ_X be the function defined, for any $x \geqslant 0$, by

$$\Psi_X(x) = \inf_{t>0}\Big(t^{-1}(L_X(t) + x)\Big).$$

Let L_X^* be the Legendre-Fenchel dual of L_X. Prove that $(\Psi_X(x) < y)$ if and only if $(L_X^*(y) > x)$. Deduce that $\Psi_X = L_X^{*-1}$. Moreover, let X and Y be real-valued random variables with finite Laplace transforms on a right neighborhood of the origin. Prove that

$$L_{X+Y}^{*-1} \leqslant L_X^{*-1} + L_Y^{*-1}.$$

Hint: Apply the Hölder inequality to the product $\exp(tX)\exp(tY)$.

Exercise 2 (A reversed Bernstein's type inequality). Let $X_1,\dots,X_n$ be a finite sequence of independent random variables, satisfying (2.2) with $c = 1$ and $\mathbb{E}[S_n] = 0$. Let v_n be defined by (2.1). Prove that for any positive x,

$$\mathbb{P}\Big(S_n > n\Big(\frac{1}{2} + \frac{\log(1+x)}{2x}\Big)\big(\sqrt{2v_n x + x^2} + x\big)\Big) \leqslant \exp(-nx).$$

Hint: Use (2.16) and choose t in such a way that $v_n t^2 = 2x(1-t)$. Compare this inequality with the inequalities of Theorem 2.1.

Exercise 3 (A Bernstein's type inequality for symmetric random variables). Let $X_1,\dots,X_n$ be a finite sequence of independent random variables with symmetric distribution, satisfying the two-sided Bernstein's condition (2.18) with $c = 1$.

1) Prove that, for any t in $]0,1[$,

$$\log \mathbb{E}\big[\exp(tS_n)\big] \leqslant n\log\Big(1 + \frac{v_n t^2}{2(1-t^2)}\Big).$$

2) Choosing t into (2.7), in such a way that $v_n t = x(1-t^2)$, prove that for any positive x,

$$\mathbb{P}(S_n \geqslant nx) \leqslant \big(1 + \gamma_n(x)\big)^n \exp\big(-2n\gamma_n(x)\big)$$

where

$$\gamma_n(x) = \frac{x^2}{v_n + \sqrt{v_n^2 + 4x^2}}.$$

Exercise 4 (Direct proof of Bennett's inequality). Let $X_1,\dots,X_n$ be a finite sequence of independent random variables with values in $]-\infty,1]$ and finite variances. Assume that $\mathbb{E}[S_n] = 0$.

1) Let X be a centered random variable with values in $]-\infty,1]$ and finite variance v. Prove that, for any positive t, $\mathbb{E}[\exp(tX)] \leqslant 1 + v(\exp(t) - t - 1)$.

2) Use Lemma 2.6 to prove that, for any positive t,

$$\log \mathbb{E}\big[\exp(tS_n)\big] \leqslant n \log\big(1 + v_n(\exp(t) - t - 1)\big) \leqslant n v_n\big(\exp(t) - t - 1\big).$$

3) Prove that, for any positive x,

$$\mathbb{P}(S_n \geqslant nx) \leqslant \big(1 + x - v_n \log(1 + x/v_n)\big)^n \exp\big(-nx \log(1 + x/v_n)\big).$$

4) Deduce from the above result that, for any positive x,

$$\mathbb{P}(S_n \geqslant nx) \leqslant \exp\big(-n v_n h(x/v_n)\big)$$

where $h(x) = (1+x)\log(1+x) - x$.

Exercise 5 (Massart's inequality, [15]). Let S_n be a random variable with Binomial $\mathscr{B}(n,p)$ distribution. Use Theorem 2.28 to prove that, for any x in $[0, 1-p]$,

$$\mathbb{P}(S_n - np \geqslant nx) \leqslant \exp\Big(-\frac{nx^2}{2(p + x/3)(1 - p - x/3)}\Big).$$

Exercise 6. Let $X_1, \dots, X_n$ be a finite sequence of independent random variables such that $X_k \leqslant b$ a.s. for some positive constant b. Let S_n and v_n be defined as in (2.1). Prove that, for any positive t,

$$\mathbb{P}\big(S_n \geqslant \sqrt{2nv_n t} + \max(0, b - v_n/b)t/3\big) \leqslant \exp(-t).$$

Deduce that

$$\mathbb{P}\big(S_n \geqslant \sqrt{2nv_n t} + bt/3\big) \leqslant \exp(-t).$$

Hint: Use the first part of Theorem 2.28.

Exercise 7 (Hoeffding Binomial inequality for nonnegative random variables). Let $X_1, \dots, X_n$ be nonnegative independent random variables. Denote

$$E_n = \frac{1}{n}\sum_{k=1}^n \mathbb{E}[X_k] \qquad \text{and} \qquad D_n = \frac{1}{n}\sum_{k=1}^n \mathbb{E}[X_k^2].$$

Prove that, for any x in $]0,1[$,

$$\mathbb{P}\big(S_n \leqslant x\mathbb{E}[S_n]\big) \leqslant \exp\Big(-\frac{n}{D_n}\Big((D_n - E_n^2 x)\log\Big(\frac{D_n - E_n^2 x}{D_n - E_n^2}\Big) + E_n^2 x \log x\Big)\Big).$$

Exercise 8. Let $\varepsilon_1, \dots, \varepsilon_n$ be a finite sequence of independent random variables sharing the same Exponential $\mathscr{E}(1)$ distribution. Let $a_1, \dots, a_n$ be a finite sequence of real numbers. For all $1 \leqslant k \leqslant n$, let $X_k = a_k(\varepsilon_k - 1)$. Prove that, for any positive x,

$$\mathbb{P}\big(S_n \geqslant \|a\|_2\sqrt{2x} + \max(0, a_1, a_2, \dots, a_n)x\big) \leqslant \exp(-x).$$

Exercise 9 (Krafft's inequality, [13]). Let $X_1, \ldots, X_n$ be a finite sequence of independent random variables with values in $[0,1]$. Prove that, for any positive x,

$$\mathbb{P}(S_n \geqslant \mathbb{E}[S_n] + nx) \leqslant \exp\Big(-2nx^2 - \frac{4}{9}nx^4\Big).$$

Exercise 10 (Weighted sums). Let $X_1, \ldots, X_n$ be a finite sequence of independent random variables.

1) Assume that, for all $1 \leqslant k \leqslant n$, $X_k = b_k Z_k$ where b_k is a real number, not necessarily positive, and $Z_1, \ldots, Z_n$ is a finite sequence of independent random variables such that, for all $1 \leqslant k \leqslant n$, Z_k has the $\Gamma(a_k, 1)$ distribution where $a_k > 0$. Assume that $\beta = \max(b_1, \ldots, b_n) > 0$. Prove that, for any positive x,

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geqslant \|b\|_{2,a}^2 x) \leqslant \exp\Big(-\frac{\|b\|_{2,a}^2}{\beta^2}\Big(\beta x - \log(1+\beta x)\Big)\Big).$$

   where $\|b\|_{2,a}$ is defined by (2.102).
2) Assume that, for all $1 \leqslant k \leqslant n$, $X_k = c_k \varepsilon_k$ where c_k is a positive real number and $\varepsilon_1, \ldots, \varepsilon_n$ is a finite sequence of independent random variables sharing the same Bernoulli $\mathscr{B}(p)$ distribution. If $0 < p \leqslant 1/2$, prove that for any x in $[0, p]$,

$$\mathbb{P}(S_n \leqslant \|c\|_1 x) \leqslant \exp\Big(-\frac{\|c\|_1^2}{\|c\|_2^2}\Big(x \log\Big(\frac{x}{p}\Big) + (1-x)\log\Big(\frac{1-x}{1-p}\Big)\Big)\Big).$$

References

1. Antonov, S. N.: Probability inequalities for series of independent random variables. Teor. Veroyatnost. i Primenen. **24**, 632–636 (1979)
2. Bennett, G.: Probability inequalities for the sum of independent random variables. J. Amer. Statist. Assoc. **57**, 33–45 (1962)
3. Bennett, G.: On the probability of large deviations from the expectation for sums of bounded, independent random variables. Biometrika **50**, 528–635 (1963)
4. Bentkus, V.: On Hoeffding's inequalities. Ann. Probab. **32**, 1650–1673 (2004)
5. Bentkus, V.: An inequality for tail probabilities of martingales with differences bounded from one side. J. Theoret. Probab. **16**, 161–173 (2003)
6. Bernstein, S. N.: Theory of Probability, Moscow (1927)
7. Boucheron, S., Lugosi, G. and Massart, P.: Concentration inequalities. Oxford University Press, Oxford (2013)
8. Boucheron, S., Lugosi, G. and Massart, P.: A sharp concentration inequality with applications. Random Structures Algorithms. **16**, 277–29 (2000)
9. Del Moral, P. and Rio, E.: Concentration inequalities for mean field particle models. Ann. Appl. Probab. **21**, 1017–1052 (2011)
10. Delyon, B.: Concentration inequalities for the spectral measure of random matrices. Electron. Commun. Probab. **15**, 549–561 (2010)
11. Hoeffding, W.: Probability inequalities for sums of bounded random variables. J. Amer. Statist. Assoc. **58**, 13–30 (1963)

12. Kearns, M. J. and Saul, L. K.: Large deviation methods for approximate probabilistic inference. Proceedings of the 14th Conference on Uncertaintly in Artifical Intelligence, San-Francisco, 311–319 (1998)
13. Krafft, O.: A note on exponential bounds for binomial probabilities. Ann. Inst. Stat. Math. **21**, 219–220 (1969)
14. Ledoux, M.: Isoperimetry and Gaussian analysis. Lectures on probability theory and statistic. Lecture Notes in Math. **1648**, 165–294 (1996)
15. Massart, P.: The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. Ann. Probab. **18**, 1269–1283 (1990)
16. Maurer, A.: A bound on the deviation probability for sums of non-negative random variables. J. Inequal. Pure Appl. Math. **4**, Article 15 (2003)
17. McDiarmid, C.: On the method of bounded differences. Surveys in combinatorics, London Mathematical Society lecture notes series **141** 148–188 (1989)
18. McDiarmid, C.: Concentration. Probabilistic methods for algorithmic discrete mathematics. Springer-Verlag, Berlin **16** 195–248 (1998)
19. Pinelis, I.: On the Bennett-Hoeffding inequality. Ann. Inst. Henri Poincaré Probab. Stat. **50**, 15–27 (2014)
20. Rio, E.: On McDiarmid's concentration inequality. Electron. Commun. Probab. **18**, 1–11 (2013)
21. Rio, E.: Extensions of the Hoeffding-Azuma inequalities. Electron. Commun. Probab. **18**, 1–6 (2013)
22. Rio, E.: Inégalités exponentielles et inégalités de concentration. Hal, cel-00702524 (2012)
23. Yurinskii, V. V.: Exponential bounds for large deviations, Teor. Veroyatnost. i Primenen. **19**, 152–154 (1974)

Springer

http://www.springer.com/978-3-319-22098-7

Concentration Inequalities for Sums and Martingales
Bercu, B.; Delyon, B.; Rio, E.
2015, X, 120 p. 9 illus. in color., Softcover
ISBN: 978-3-319-22098-7