ALEA, Lat. Am. J. Probab. Math. Stat. 20, 1615–1629 (2023)
DOI: 10.30757/ALEA.v20-61

Quantitative bounds for large deviations of heavy tailed random variables

Quirin Vogel

NYU Shanghai, 1555 Century Ave, Pudong, Shanghai, China, 200122
Department of Mathematics, CIT, Technische Universität München, Boltzmannstr. 3, D-85748, Garching bei München, Germany
E-mail address: quirinvogel@outlook.com
URL: https://www.math.cit.tum.de/en/math/personen/wissenschaftliches-personal/vogel-quirin/

Abstract. The probability that the sum of independent, centered, identically distributed, heavy-tailed random variables achieves a very large value is asymptotically equal to the probability that there exists a single summand equalling that value. We quantify the error in this approximation. We furthermore characterise the law of the individual summands, conditioned on the sum being large.

1. Introduction

Large deviation theory concerns the study of random variables taking values away from their mean. A classic result in large deviation theory is that for $S_n = \sum_{i=1}^n X_i$ the sum of i.i.d., centred, integer-valued random variables $(X_i)_i$ with exponential tails, one has that for $x \in \mathbb{R}$

$$\mathbb{P}(S_n > nx) = \mathrm{e}^{-I(x)n(1+o(1))} \quad \text{as} \quad n \to \infty.$$

Here, $I(x)$ is the Legendre transform of the logarithmic moment generating function of X_1, i.e., in this case

$$I(x) = \sup_{t \geq x} \left\{ tx - \log \mathbb{E}\left[\mathrm{e}^{tX_1}\right] \right\}.$$

See Dembo and Zeitouni (2010) for more details. A follow-up task is the quantification of error- or higher-order terms. A classic result is given in Bahadur and Ranga Rao (1960), where it is shown that under certain conditions

$$\mathbb{P}(S_n > nx) = \frac{\mathrm{e}^{-I(x)n}}{\sigma\sqrt{n}}(1 + \mathcal{O}(n^{-1})),$$

for some $\sigma > 0$. Indeed, one often can even give the stronger estimate

$$\mathbb{P}(S_n = nx) = \frac{\mathrm{e}^{-I(x)n}}{\sigma\sqrt{n}}(1 + \mathcal{O}(n^{-1})),$$

Received by the editors January 13th, 2023; accepted November 28th, 2023.
2010 *Mathematics Subject Classification.* Primary: 60F10; Secondary: 60B10.
Key words and phrases. Local large deviations, heavy-tails, stable random variables.

if nx is in the support of S_n, see Blackwell and Hodges (1959). However, when one considers the case where the moment generating function does not exist, the behavior of $\mathbb{P}(S_n > nx)$ changes drastically. When the tails of X_i decay polynomially (and sufficiently fast), Tchachkuk and Nagaev in Tchachkuk (1977); Nagaev (1982) show that

$$\mathbb{P}(S_n > nx) = n\mathbb{P}(X_1 > nx)(1 + o(1)) \,. \tag{1.1}$$

Recently, Berger in Berger (2019a) gave the improvement

$$\Big|\frac{\mathbb{P}(S_n = nx)}{n\mathbb{P}(X_1 = nx)} - 1\Big| = o(1) \,,$$

given some (mild) local conditions on the tail. There are similar results, for different distributions and cases, see for example Yang et al. (2010); Berger (2019b); Mikosch and Rodionov (2021); Berger et al. (2023).

Our first result considers the quantification of the error in Berger (2019a); we show that

$$\Big|\frac{\mathbb{P}(S_n = nx)}{n\mathbb{P}(X_1 = nx)} - 1\Big| = \mathcal{O}(\varepsilon_n(x)) \,,$$

for some vanishing (in many cases explicit) sequence $\varepsilon_n(x)$, which depends on the distribution function of the X_i's and on x. This is to our best knowledge the first quantification of such error terms in the heavy-tail regime. The Fuk–Nagaev inequality is a vital tool for our analysis, as in other works in this area (see Nagaev (1982); Berger (2019a) for example).

Apart from computing the probability of a large deviation event, gaining insight in *how* this deviation is achieved is an important part of large deviation theory. For random variables with existing moment generating function, this often goes by the name *Gibbs-conditioning principle*, see Dembo and Zeitouni (2010). Roughly speaking, the large exceedance is achieved by tilting the distribution of each X_i, so that the unlikely value becomes likely in the tilted distribution. The independence is asymptotically preserved.

For random variables with sub-exponential tails, the situation is starkly different: the large exceedance is achieved by one of the X_i's assuming the large value, see Equation (1.1).

In Armendáriz and Loulakis (2011), it was shown that the *total variation* distance between the conditional distribution

$$\mathbb{P}(\{X_i\}_{i=1}^n \in \cdot \,\big| S_n > nx) \,,$$

and its "limiting" distribution converges to zero. The "limiting" distribution is defined as follows: independently sample a random variable Y with distribution $\mathbb{P}(Y \in A) = \mathbb{P}(X_1 \in A | X_1 > nx)$ and $(n-1)$-copies of X_i (according to the original law). A position $i \in \{1, \ldots, n\}$ is sampled uniformly at random. The "limiting" law is given by the law of

$$\Big(X_1, \ldots, X_{i-1}, Y, X_i, \ldots, X_{n-1}\Big) \,. \tag{1.2}$$

Our contribution to this question is twofold: not only do we quantify the speed of convergence but we also provide a deeper understanding of the conditional law by altering the law of Y. In Armendáriz and Loulakis (2011) the authors give two proofs of their result, one only working for positive random variables and one for the general case. The reason why their first proof breaks down in the general case is that it does not take into account the fluctuations induced by the $(n-1)$-copies of X_1. By modifying the law of Y, we get a new proof which works in general and also gives the speed of convergence.

Expanding on our previous results, we can also give the limiting law of

$$\mathbb{P}(\{X_i\}_{i=1}^n \in \cdot \,\big| S_n = nx) \,.$$

This case is interesting as the large value is no longer independent from the $n-1$-copies of X_i.

A word regarding the level of generality in this paper: this paper is a compromise between allowing for generality and keeping the notation easy to read. We chose to restrict ourselves to $\mathbb{Z}$-valued random variables with tails consisting of a power-law and a slowly varying function, as in Berger (2019a). However, similar to Berger (2019a), the modifications of the arguments (not the notation) needed to address the continuum case ($\mathbb{R}$-valued) are small.

There is a limit to the precision of our local expansion, related to the CLT scale $(a_n)_n$ of the underlying random variables. We introduce the notation

$$f_n = \omega(g_n) \qquad \text{if and only if} \qquad o(f_n) = g_n\,,$$

as $n \to \infty$. We furthermore write $f_n \sim g_n$ whenever $f_n = g_n(1 + o(1))$, as $n \to \infty$.

2. Results

Let $\{X_i\}_i$ be an i.i.d. sequence of $\mathbb{Z}$-valued random variables such that for $x \in \mathbb{Z}_+$, we either have

$$\mathbb{P}(X_1 = x) = p\alpha L(x) x^{-(1+\alpha)}\,, \quad \text{and} \quad \mathbb{P}(X_1 = -x) = \mathcal{O}\big(L(x)x^{-(1+\alpha)}\big)\,, \tag{2.1}$$

or

$$\mathbb{P}(X_1 = -x) = q\alpha L(x) x^{-(1+\alpha)}\,, \quad \text{and} \quad \mathbb{P}(X_1 = x) = \mathcal{O}\big(L(x)x^{-(1+\alpha)}\big)\,, \tag{2.2}$$

for L a slowly varying function, $p, q > 0$ $\alpha \in (0,\infty)$ and the $\mathcal{O}(\dots)$ is understood as $x \to \infty$.

Recall that L slowly varying means that $L(\lambda x) \sim L(x)$ for any $\lambda > 0$, as $x \to \infty$. One may think of $L(x)$ growing/shrinking slower than any polynomial. Note that the mean of X_1 exists for $\alpha > 1$ and the variance exists for $\alpha > 2$.

Suppose that there are two sequence $(a_n)_n$ and $(b_n)_n$ satisfying the following: assume that $\big(a_n\big)_n$ satisfies

$$\begin{cases} L(a_n)(a_n)^{-\alpha} \sim n^{-1} & \text{if } \alpha \in (0,2)\,, \\ \sigma^2(a_n) a_n^{-2} \sim n^{-1} & \text{if } \alpha \geq 2\,, \end{cases} \tag{2.3}$$

where $\sigma^2(x) = \mathbb{E}\big[(X_1 - \mathbb{E}[X_1])^2 \mathbf{1}\big\{\big|X_1 - \mathbb{E}[X_1]\big| \leq x\big\}\big]$; and that $\big(b_n\big)_n$ is given by

$$b_n = \begin{cases} 0 & \text{if } \alpha \in (0,1)\,, \\ n\mathbb{E}\big[X_1 \mathbf{1}\{\big|X_1\big| \leq a_n\}\big] & \text{if } \alpha = 1\,, \\ n\mathbb{E}[X_1] & \text{if } \alpha > 1\,. \end{cases}$$

The function $\sigma^2(x)$ will only play a role when $\alpha \geq 2$. Let $S_n = \sum_{i=1}^n X_i$. Then, S_n satisfies a central limit theorem with scales $(a_n)_n$ and $(b_n)_n$, i.e., one has that $\frac{S_n - b_n}{a_n}$ converges to a stable law, see Feller (1968, IX.8, Eq. (8.14)). We study the deviations from this central limit theorem.

Finally, we need to quantify how fast the function L varies: we say that L is slowly varying *with precision* $\mathrm{err}[x,y]$ whenever

$$L\big(x + y\big) = L(x)\big(1 + \mathrm{err}[x,y]\big) \quad \text{as } \big|x\big| \to \infty\,.$$

for $\big|y\big| = o(\big|x\big|)$ and for $o(1) = \mathrm{err}[x,y]$ some function.

Two examples of slowly varying functions are $L(x) = \log(x)^\beta$ and $L(x) = 1 + \mathcal{O}\big(x^{-\alpha}\big)$, as $x \to \infty$. In the first case, one has that $\mathrm{err}[x,y] \sim \beta y/x$ and in the latter case one has $\mathrm{err}[x,y] = \mathcal{O}\big(x^{-\alpha}\big)$.

Theorem 2.1. *Suppose that L is slowly varying with precision* $\mathrm{err}[x,y]$. *Assume Equation (2.1) holds with $p > 0$. Set $\widehat{S}_n = S_n - \lfloor b_n \rfloor$ and $\alpha_1 = \frac{\alpha}{\alpha+1} \in (0,1)$. Write*

$$A(x,n) = \Big|\frac{\mathbb{P}(\widehat{S}_n = x)}{n\mathbb{P}\big(X_1 = x\big)} - 1\Big|\,.$$

We then have that for every $\varepsilon > 0$ small enough

(1) *For* $\alpha \in (0,2)$*, we have that for all* $0 < x = \omega(a_n) \to \infty$

$$A(x,n) = \mathcal{O}\big(\big(\frac{a_n}{x}\big)^{(\alpha_1-\varepsilon)} + \mathrm{err}\big[x,(a_n/x)^{\alpha_1}\big]\big)\,.$$

(2) *For* $\alpha = 2$*, we get that for all* $0 < x = \omega(a_n\sqrt{\log(n)}) \to \infty$

$$A(x,n) = \mathcal{O}\big(\big(\frac{a_n\sqrt{\log(n)}}{x}\big)^{(\frac{2}{3}-\varepsilon)} + \mathrm{err}\big[x,(a_n/x)^{\alpha_1}\big]\big)\,.$$

(3) *For* $\alpha > 2$*, we assume that* $x = \omega(\sqrt{n\log(n)})$ *as* $n \to \infty$*. Set* $\beta \geq 0$ *such that* $n^{-\beta}\big(x/\sqrt{n\log(n)}\big)^{1-\alpha_1} \to \infty$ *and* $\beta \leq \frac{(\alpha-2)(\alpha+1)}{2(2\alpha+1)}$*. Then*

$$A(x,n) = \mathcal{O}\big(n^{1-\alpha/2+\beta\alpha_1}\big(\frac{\sqrt{n\log(n)}}{x}\big)^{(\alpha_1-\varepsilon)} + \mathrm{err}\big[x,(a_n/x)^{\alpha_1}\big]\big)\,.$$

See Remark 4.1 for the slightly stronger assumptions in the cases $\alpha \geq 2$*.*

Note that by symmetry, given Equation (2.2) the theorem also holds true for the limit $x \to -\infty$*, with the respective assumption on the right tail.*

Example 2.2. If X_1 is symmetric *zeta*$(1+\alpha)$ distributed, i.e., for $k \in \mathbb{Z}\setminus\{0\}$

$$\mathbb{P}\big(X_1 = k\big) = \frac{|k|^{-(1+\alpha)}}{2\zeta(1+\alpha)}\,.$$

We then obtain that for all $\alpha > 1$, $c > 0$, $\varepsilon > 0$ and for all $x \geq nc$

$$\mathbb{P}\big(\widehat{S}_n = x\big) = n\mathbb{P}\big(X_1 = x\big)\big(1+\mathcal{O}\big(n^{-\frac{\alpha-1}{\alpha+1}\mathbf{1}\{\alpha\leq 2\}-\frac{\alpha}{2+2\alpha}\mathbf{1}\{\alpha>2\}+\varepsilon}\big)\big)\,, \tag{2.4}$$

as $a_n = n^{\frac{1}{\alpha}\vee\frac{1}{2}}(1+\log(n)\mathbf{1}\{\alpha=2\})$ and hence $(a_n/n)^{\alpha_1}$ is equal to $n^{-\frac{\alpha-1}{\alpha+1}}$ in the case $\alpha \leq 2$ (absorbing the $\log n$ factor in n^ε for $\alpha = 2$) and $n^{-\frac{\alpha}{2+2\alpha}}$ for $\alpha > 2$. Furthermore, note that $\mathrm{err}[x,y] = 0$ in this case.

Note that for $\alpha > 2$, Theorem 2.1 gives a better error bound, depending on the value of β. For $\alpha \in (0,1]$, $a_n \geq n$ and hence one needs to choose larger x; we leave the details to the reader.

Next, we give a non-local version of Theorem 2.1.

Theorem 2.3. *Suppose that* $\{X_i\}_i$ *is an i.i.d. sequence of* $\mathbb{Z}$*-valued random variables such that for* $x \in \mathbb{Z}_+$ *and* $\hat{L}(x)$ *a slowly varying function*

$$\mathbb{P}(X_1 \geq x) = p\alpha\hat{L}(x)x^{-\alpha}\,, \tag{2.5}$$

and that $\mathbb{P}(X_1 \leq -x) = \mathcal{O}\big(1\big)\hat{L}(x)x^{-\tilde{\alpha}}$ *for some* $\tilde{\alpha} \geq \alpha$*. We then have that for* x *satisfying the same conditions as in Theorem 2.1*

$$\Big|\frac{\mathbb{P}(\widehat{S}_n \geq x)}{n\mathbb{P}\big(X_1 \geq x\big)} - 1\Big| = \mathcal{O}\big(A(x,n)\big)\,,$$

where $A(x,n)$ *is as in Theorem 2.1.*

Remark 2.4. Note that Theorem 2.1 cannot be deduced from Theorem 2.3 as

$$\mathbb{P}(X_1 = x) = \mathbb{P}(X_1 \geq x+1) - \mathbb{P}(X_1 \geq x) = p\alpha x^{-\alpha}\big(\hat{L}(x+1)\big(1+1/x\big)^{-\alpha} - \hat{L}(x)\big)\,,$$

and $\big|\hat{L}(x+1) - \hat{L}(x)\big|$ can be much larger than $\mathcal{O}\big(x^{-1}\big)$ (take for example $\hat{L}(x) = 1+(-1)^x|x|^{-1/2}$).

As the largest value in the sequence $(X_1, \dots, X_n)$ could appear at any spot, we introduce the following shift, which moves it to the last spot: let $T\colon \bigcup_{n\in\mathbb{Z}_+}\mathbb{R}^n \to \bigcup_{n\in\mathbb{Z}_+}\mathbb{R}^n$ with (set here $\max\emptyset = -\infty$)

$$T(x_1,\dots,x_n)_k = \begin{cases} \max_{1\le i\le n} x_i & \text{when } k=n\,,\\ x_n & \text{when } x_k > \max_{1\le i<k} x_i \text{ and } x_k = \max_{i\ge k} x_i\,,\\ x_k & \text{otherwise}\,. \end{cases}$$

Denote the law of X_1 by μ. Write $F(x) = \mu\big((-\infty, x]\big)$ for the cumulative distribution function and $G(x) = 1 - F(x)$. Set $\nu_{x,n} = \mathbb{P}\big(\{X_i\}_{i=1,\dots,n} \in \cdot\,|S_n > x\big)$, the distribution of the summands, conditional on S_n large. Let ν_x be the[1] distribution of X_1 conditional on being large:

$$\nu_x(A) = \mathbb{P}(X_1 \in A|X_1 > x - \omega(a_n)) \qquad \text{for} \qquad \omega(a_n) = o(x)\,,$$

where we recall that $\omega(a_n)$ is any sequence diverging faster than $(a_n)_n$. We use $||\cdot||$ to denote the *total variation* norm.

Theorem 2.5. *Assume that X_1 has mean zero (or $b_n = 0$) and that $G(x+y)/G(x) = 1+\mathrm{err}^{(1)}[x,y]$ as $x\to\infty$ and $y = o(x)$. Furthermore, set $c_{x,n} = \big|\mathbb{P}(S_n \ge x) - n\mathbb{P}\big(X_1 \ge x\big)\big|$. We then have that for $x = \omega(a_n)$ and $x \to \infty$*

$$||T[\nu_{x,n}] - \mu^{\otimes(n-1)} \otimes \nu_x||^2 = \mathcal{O}\big(\max\{\mathrm{err}^{(1)}[x,\omega(a_n)], c_{x,n}, nG(x)\}\big)\,. \tag{2.6}$$

In words, we can sample $\{X_i\}_{i=1,\dots,n}$ conditioned on $S_n > x$ by

- *sampling independently $\{\tilde{X}_i\}_{i=1,\dots,n-1}$ distributed according to $\mu^{\otimes(n-1)}$,*
- *a position $i \in \{1,\dots,n\}$ uniformly,*
- *and Y according to ν_x*

and have the distribution of $\{X_i\}_{i=1,\dots,n}$ is approximately equal to $\big(\tilde{X}_1,\dots,\tilde{X}_{i-1}, Y, \tilde{X}_i,\dots,\tilde{X}_{n-1}\big)$, with the error (in total variation norm) given by Equation (2.6).

Example 2.6. In the setting of Example 2.2 with $\alpha = 3/2$, we have that for $x > 0$ of order n

$$||T[\nu_{x,n}] - \mu^{\otimes(n-1)} \otimes \nu_x||^2 = \mathcal{O}\big(n^{-1/5+\varepsilon}\big)\,.$$

This allows us to give statements such as: for $A_1,\dots,A_n$ measurable subsets of $\mathbb{R}$, we get that

$$\mathbb{P}\big(T[X_1,\dots,X_n] \in A_1 \times \ldots \times A_n|S_n > x\big) \sim \nu_x(A_n)\prod_{i=1}^{n-1}\mu(A_i)\,,$$

as long as the right-hand side has a probability of $\omega\big(n^{-1/10+\varepsilon/2}\big)$. This cannot be concluded from the $o(1)$ bounds in Armendáriz and Loulakis (2011).

Denote

$$\xi_{x,n} = \mathbb{P}\big(\{X_i\}_{i=1,\dots,n} \in \cdot\,|S_n = x\big)\,.$$

We also set $\xi^*_{x,n}$ the measure given by

$$\xi^*_{x,n} = \int \mathrm{d}\mu^{\otimes(n-1)}(y)\delta_{x-\sum_{i=1}^{n-1} y_i}\,.$$

In words, $\xi^*_{x,n}$ samples $(y_1,\dots,y_{n-1})$ i.i.d. according to μ and then sets the final coordinate as $x - \sum_{i=1}^{n-1} y_i$.

[1] ν_x does depend on the choice of $\omega(a_n)$. However, this dependence is asymptotically negligible on most events A, as $\omega(a_n) = o(x)$.

Theorem 2.7. *Assume that X_1 has mean zero (or $b_n = 0$) and that $G(x+y)/G(x) = 1+\mathrm{err}^{(2)}(x,y)$ as $x \to \infty$ and $y = o(x)$. Set $c_{x,n} = \big|\mathbb{P}(S_n = x) - n\mathbb{P}\big(X_1 = x\big)\big|$. We then have that for $x = \omega(a_n)$ in the support of S_n*

$$||T[\xi_{x,n}] - \xi^*_{x,n}||^2 = \mathcal{O}\big(\max\big\{\mathrm{err}^{(2)}[x, \omega(a_n)], c_{x,n}, nG(x)\big\}\big) \quad \textit{as} \quad x \to \infty\,.$$

3. An application

In this section, we show how we can use the results above to gain some new insights. Suppose $\big(N_x\big)_{x\in\mathbb{Z}}$ is a collection of independent Poisson random variables with intensity $\lambda > 0$. Consider the random sum

$$S_n = \sum_{x=-n}^{n} \sum_{i=1}^{N_x} Y_i^{(x)}\,,$$

where $\{Y_i^{(x)}\}_{x\in\mathbb{Z}, i\in\mathbb{Z}_+}$ is a collection of independent symmetric *zeta*$(1+\alpha)$ distributed random variables, independent of $\big(N_x\big)_{x\in\mathbb{Z}}$.

Proposition 3.1. *Given $\alpha > 1$, for any $c > 0$, uniformly in $k \geq cn$*

$$\mathbb{P}(S_n = k) = \mathbb{P}\big(\exists x \in \{-n, \dots, n\} \textit{ and } i \in \{1, \dots, N_x\} \colon Y_i^{(x)} = k\big)\big(1 + \mathcal{O}\big(n^{-\beta}\big)\big)\,,$$

for some $\beta = \beta_\alpha > 0$. This can be interpreted as a condensation phenomena, see Großkinsky et al. (2003). The constant β can be chosen the same as in Equation (2.4).

In Klüppelberg and Mikosch (1997) the asymptotics of the cumulative distribution function $\mathbb{P}(S_n > k)$ were obtained, however neither the error term was quantified nor the probability density function approximated.

Proof: The idea is that the parameter n in Theorem 2.1 is now Poisson distributed with parameter $(2n+1)\lambda$. However, by standard large deviation estimates for Poisson random variables, one can show that such a Poisson random variable is bounded by $(2n+1)\lambda \pm n^{1/2+\varepsilon}$ for any $\varepsilon > 0$, outside a set of stretch exponentially small probability. Hence, we can apply Theorem 2.1.

Note that for every $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\mathbb{P}\big(\big| \sum_{x=-n}^{n} N_x - (2n+1)\lambda\big| \geq n^{1/2+\varepsilon}\big) = \mathcal{O}\big(\mathrm{e}^{-n^\delta}\big)\,,$$

see Dembo and Zeitouni (2010, Eq. (2.2.12)).

Conditional on the value of $\sum_{x=-n}^{n} N_x$ and on the event $\big|\sum_{x=-n}^{n} N_x - (2n+1)\lambda\big| \leq n^{1/2+\varepsilon}$, we can apply Theorem 2.1 to get

$$\mathbb{P}\big(\sum_{x=-n}^{n} \sum_{i=1}^{N_x} Y_i^{(x)} = k \Big| \sum_{x=-n}^{n} N_x\big) = \mathbb{P}(Y_1^{(0)} = k) \sum_{x=-n}^{n} N_x\big(1 + \mathcal{O}\big(n^{-\beta}\big)\big)\,,$$

for some $\beta > 0$. Furthermore, note that on the event $\big\{\big|\sum_{x=-n}^{n} N_x - (2n+1)\lambda\big| \leq n^{1/2+\varepsilon}\big\}$ and for $M = \sum_{x=-n}^{n} N_x$

$$\mathbb{P}\big(\exists (x,i) \in \{-n, \dots, n\} \times \{1, \dots, N_x\} \colon Y_i^{(x)} = k \Big| M\big) \sim M\mathbb{P}\big(Y_1^{(0)} = k\big)\big(1 + \mathcal{O}\big(n\mathbb{P}\big(Y_1^{(0)} = k\big)\big)\big)\,,$$

by the fundamental property of Poisson processes. □

4. Proofs

4.1. *Technical preliminaries.* Before embarking on the proof, we recall the scales involved in our analysis:

- The scale n, given.
- The scale a_n, induced by the CLT scaling. It satisfies $L(a_n)a_n^{-\alpha} \sim n^{-1}$ if $\alpha \in (0,2)$ and $\sigma^2(a_n)a_n^{-2} \sim n^{-1}$ if $\alpha \geq 2$, see Equation (2.3).
- The scale of x. It only has to obey the constraint that $x = \omega(a_n)$.
- The induced scale $\frac{x}{a_n}$. It relates to the best possible error we can achieve.

Recall Potter's bound (see Bingham et al. (1989, Theorem 1.5.6)) which gives for L slowly varying and any $\delta > 0$, that there exists c_δ such that for a, b sufficiently large

$$L(a)/L(b) \leq c_\delta \max\{(a/b)^\delta, (b/a)^\delta\}\,. \tag{4.1}$$

Remark 4.1 (The Gaussian domain of attraction). For $\alpha \geq 2$, the limiting law of $(S_n - b_n)/a_n$ is Gaussian. This changes the big jump phenomenon of $\mathbb{P}\big(S_n = x\big)$ in the region where $a_n \leq x \leq Ca_n \log(n)$, for $C > 0$. This was already observed by Nagaev (1982) in the case $\alpha > 2$ and $\big\{S_n \geq x\big\}$, see the recent Berger et al. (2023) for the complete picture. We summarize the points relevant to our case: if $\alpha > 2$, the have that

$$\mathbb{P}\big(S_n - b_n > x\big) \sim pnL(x)x^{-\alpha} \qquad \text{if} \qquad x > b\big(n\log(n)\big)^{1/2}\,,$$

where $b > (\alpha-2)^{1/2}$. If $b < (\alpha-2)^{1/2}$, this is no longer true (for the case $b = (\alpha-2)^{1/2}$, see Berger et al. (2023)).

If $\alpha = 2$, we need to be more careful: set $q(x) = x^2\mathbb{P}(X_1 > x)/\sigma^2(x)$. For $\alpha = 2$, we have that $\sigma^2(x)$ is slowly varying and grows faster than $L(x)$ (see Bingham et al. (1989, Proposition 1.5.9a)) and hence $q(x) = o(1)$, as $x \to \infty$, and slowly varying. Then, using Berger et al. (2023, Equation 2.9), we have the occurrence of the single big jump if

$$\liminf_{n\to\infty} \frac{\big(x/(a_n)\big)^2}{2\log q(a_n)} > 1\,,$$

and no big jumps if the limsup is bounded from above by 1. By the Potter bounds for any $\varepsilon > 0$, $q(x) = \mathcal{O}\big(x^{-\varepsilon}\big)$ as $x \to \infty$ and hence if $(x/a_n)^2$ grows faster than $\sqrt{\log(n)}$, a big jump will occur.

We recall the *local Fuk–Nagaev inequality* from Berger (2019a, Theorem 5.1).

Theorem 4.2. *Fix $\alpha > 0$. Set M_n the maximum of the $X_1, \ldots, X_n$. Write $\widehat{S}_n$ for the recentered walk $\widehat{S}_n = S_n - \lfloor b_n \rfloor$. Write $\sigma_{2,\alpha}(x) = \mathbb{E}\big[\big|X_1\big|^\alpha \mathbf{1}\big\{\big|X_1\big| > x\big\}\big]$ if the tails decay with speed $\alpha > 2$. Again, $\sigma_{2,\alpha}(x)$ is slowly varying, see Bingham et al. (1989, Proposition 1.5.9a). Under the conditions from Theorem 2.1, there exist $\mathfrak{c}_1, \mathfrak{c}_2, \mathfrak{c}_3 > 0$ such that for every $1 \leq y \leq x$ and every x with $x \geq a_n$, we have*

$$\mathbb{P}\big(\widehat{S}_n = x,\, M_n \leq y\big) \leq \frac{\mathfrak{c}_3}{a_n}\begin{cases} \mathrm{e}^{-\mathfrak{c}_1 x^2/n} + \big(\frac{xy^{\alpha-1}}{n\sigma_{2,\alpha}(y)}\big)^{-\mathfrak{c}_2 x/y} & \text{if } \alpha > 2\,, \\ \mathrm{e}^{x/y}\big(1 + \frac{xy}{n\sigma_2(y)}\big)^{-x/y} & \text{if } \alpha = 2\,, \\ \big(\mathfrak{c}_1 \frac{y}{x} nL(y)y^{-\alpha}\big)^{x/2y} & \text{if } \alpha \in (1,2)\,, \\ \mathrm{e}^{\frac{3x}{y}}\big(1 + \frac{\mathfrak{c}_1 x}{nL(y)}\big)^{-\frac{x}{4y}} + \mathrm{e}^{-\mathfrak{c}_2(x/a_n)^2} & \text{if } \alpha = 1\,, \\ \mathrm{e}^{x/4y}\big(1 + \frac{\mathfrak{c}_1 x}{ny^{1-\alpha}L(y)}\big)^{-x/y} & \text{if } \alpha < 1\,. \end{cases} \tag{4.2}$$

Proof: The above result is stated in Berger (2019a, Theorem 5.1) for the case $\alpha \in (0,2)$. For $\alpha > 2$, it follows from Nagaev (1979, Corollary 1.7). For $\alpha = 2$, it follows from Berger et al. (2023, Lemma 5.2). □

To ease reading, we write for $a, b \in \mathbb{R}$ and any $f\colon \mathbb{Z} \to \mathbb{R}$

$$\sum_{k=a}^{b} f(k) := \sum_{k\in[a,b]\cap\mathbb{Z}} f(k)\,, \qquad \text{and similarly for} \qquad \sum_{k\ge a} f(k) \quad \text{and} \quad \sum_{k\le a} f(k)\,.$$

4.2. *Proof of Theorem 2.1.* We now begin with the main proof: without loss of generality, assume that $x > 0$. Fix a sequence $\varepsilon_n = o(1)$ large enough such that $\varepsilon_n x/a_n \to +\infty$ (for $\alpha \in (0,2)$) and $\varepsilon_n x/(a_n \log^{1/2}(n)) \to +\infty$ for $\alpha \ge 2$. The sequence ε_n allows us to interpolate between the CLT scale $(a_n)_n$ and speed of divergence of x. We also fix the ($\alpha-$ dependent) sequence $\tilde{\varepsilon}_n = \log(x/a_n)^{-2} = o(1)$ (for $\alpha < 2$) and $\tilde{\varepsilon}_n = \log(x/a_n\sqrt{\log(n)})^{-2} = o(1)$ (for $\alpha \ge 2$). Note that for some $c > 0$, we get $\tilde{\varepsilon}_n x \ge cx/\log^2(x) \to \infty$.

In the first part of the proof, we give general error bounds, valid for (almost) all kind of remainders $\mathrm{err}[x,y]$. In the second part of the proof, we collect all the errors and simplify. This allows to adapt the result easily to all type of error estimates without overloading the notation. We expand

$$\begin{aligned}\mathbb{P}(\widehat{S}_n = x) =& \mathbb{P}\big(\widehat{S}_n = x, M_n \ge (1-\varepsilon_n)x\big)\\ &+ \mathbb{P}\big(\widehat{S}_n = x, M_n \in (\tilde{\varepsilon}_n x, [1-\varepsilon_n]x)\big) + \mathbb{P}\big(\widehat{S}_n = x, M_n \le \tilde{\varepsilon}_n x\big)\\ =& \mathrm{A} + \mathrm{B} + \mathrm{C}\,.\end{aligned}$$

1. Estimating A: we begin by dissecting A

$$A = \mathbb{P}\big(\widehat{S}_n = x, (1-\varepsilon_n)x \le M_n \le (1+\varepsilon_n)x\big) + \mathbb{P}\big(\widehat{S}_n = x, M_n > (1+\varepsilon_n)x\big) = \mathrm{A1} + \mathrm{A2}\,.$$

The second term is negligible, as we will see later. For the first term, we write

$$\mathrm{A1} = \mathbb{P}\big(\widehat{S}_n = x, (1-\varepsilon_n)x \le M_n \le (1+\varepsilon_n)x\big) = \sum_{y=(1-\varepsilon_n)x}^{(1+\varepsilon_n)x} \mathbb{P}\big(\widehat{S}_n = x, M_n = y\big)\,.$$

We begin with an upper bound

$$\sum_{y=(1-\varepsilon_n)x}^{(1+\varepsilon_n)x} \mathbb{P}\big(\widehat{S}_n = x, M_n = y\big) \le \sum_{y=(1-\varepsilon_n)x}^{(1+\varepsilon_n)x} n\mathbb{P}\big(X_1 = y\big)\mathbb{P}\big(S_{n-1} - \lfloor b_n \rfloor = x - y\big)\,,$$

where we used the independence and a union bound.

Fix $y \in x[1-\varepsilon_n, 1+\varepsilon_n]$ and write $w = y - x$. We then have that

$$\mathbb{P}\big(X_1 = y\big) = p\alpha L(y) y^{-(1+\alpha)} = p\alpha L(x+w)(x+w)^{-(1+\alpha)}\,.$$

Using the error bounds we have for L and the binomial series, we get

$$\mathbb{P}\big(X_1 = y\big) = p\alpha L(x) x^{-(1+\alpha)}\big(1 + \mathrm{err}[x, \varepsilon_n x] + \mathcal{O}(\varepsilon_n)\big)\,.$$

Therefore,

$$\mathbb{P}\big(\widehat{S}_n = x, (1-\varepsilon_n)x \le M_n \le (1+\varepsilon_n)x\big) \le p\alpha n L(x) x^{-(1+\alpha)}\big(1 + \mathrm{err}[\varepsilon_n, x] + \mathcal{O}(\varepsilon_n)\big)\,.$$

On the other hand, we have

$$\begin{aligned}\sum_{y=(1-\varepsilon_n)x}^{(1+\varepsilon_n)x} \mathbb{P}\big(\widehat{S}_n = x, M_n = y\big) \ge& \sum_{y=(1-\varepsilon_n)x}^{(1+\varepsilon_n)x} n\mathbb{P}\big(X_1 = y\big)\mathbb{P}\big(S_{n-1} - \lfloor b_n \rfloor = x - y\big)\\ &- \frac{1}{2}\sum_{y=(1-\varepsilon_n)x}^{(1+\varepsilon_n)x} n(n-1)\mathbb{P}\big(X_1 = y\big)^2\mathbb{P}\big(S_{n-2} - \lfloor b_n \rfloor = x - y\big)\,.\end{aligned} \tag{4.3}$$

As above, the first sum is $p\alpha n L(x) x^{-(1+\alpha)}\big(1+\mathrm{err}[\varepsilon_n, x]+\mathcal{O}(\varepsilon_n)\big)$. The second sum is bounded by

$$Cn^2 L^2(x) x^{-2(1+\alpha)}\,,$$

for some $C>0$ and is negligible as we will see later.

For the second term A2, we have

$$\begin{aligned}\mathrm{A2} = \mathbb{P}\big(\widehat{S}_n = x, M_n > (1+\varepsilon_n)x\big) &= \sum_{y\geq(1+\varepsilon_n)x} \mathbb{P}\big(\widehat{S}_n = x, M_n = y\big)\\ &\leq \sum_{y\geq(1+\varepsilon_n)x} n\mathbb{P}\big(X_1 = y\big)\mathbb{P}\big(S_{n-1}-\lfloor b_n\rfloor = x-y\big)\,,\end{aligned} \tag{4.4}$$

where we again used the exchangeability of the X_i's and a union bound. We can estimate the first term by its maximum to conclude

$$\mathbb{P}\big(\widehat{S}_n = x, M_n > (1+\varepsilon_n)x\big) \leq n \sup_{y\geq(1+\varepsilon_n)x} \mathbb{P}\big(X_1 = y\big)\mathbb{P}\big(S_{n-1}-\lfloor b_n\rfloor \leq -\varepsilon_n x\big)\,. \tag{4.5}$$

Recall the condition on $\varepsilon_n x$ stated at the beginning of the proof and that the left tails of X_1 decay with speed at least $\mathcal{O}\big(L(x)x^{-\alpha}\big)$. We have that for some $C>0$

$$\mathbb{P}\big(S_{n-1}-\lfloor b_n\rfloor \leq -\varepsilon_n x\big) \leq CnL(\varepsilon_n x)(\varepsilon_n x)^{-\alpha}\,, \tag{4.6}$$

using Berger (2019a, Theorem 2.1) in the case $\alpha\in(0,2)$, Doney (2001, Theorem 2) for $\alpha>2$ and Berger et al. (2023, Equation 2.33) for the case $\alpha=2$. Hence

$$\mathbb{P}\big(\widehat{S}_n = x, M_n > (1+\varepsilon_n)x\big) = \mathcal{O}\big(x^{-(1+\alpha)} n^2 L(x) L(\varepsilon_n x)(\varepsilon_n x)^{-\alpha}\big)\,.$$

To summarize: we have that

$$\mathrm{A} = p\alpha n L(x) x^{-(1+\alpha)}\big(1+\mathrm{err}[x,\varepsilon_n x]+\mathcal{O}(\varepsilon_n)+nL(\varepsilon_n x)(\varepsilon_n x)^{-\alpha}+\mathcal{O}\big(nL(x)x^{-(1+\alpha)}\big)\big)\,. \tag{4.7}$$

2. Estimating B: to bound the term B, we expand, as in Equations (4.4) and (4.5), for some $C_1'>0$ universal

$$\begin{aligned}\mathrm{B} = \mathbb{P}\big(\widehat{S}_n = x, M_n\in(\tilde{\varepsilon}_n x,[1-\varepsilon_n]x)\big) &= \sum_{y=\tilde{\varepsilon}_n x+1}^{(1-\varepsilon_n)x-1} \mathbb{P}\big(\widehat{S}_n = x, M_n = y\big)\\ &\leq \sum_{y=\tilde{\varepsilon}_n x+1}^{(1-\varepsilon_n)x-1} n\mathbb{P}\big(X_1 = y\big)\mathbb{P}\big(S_{n-1}-\lfloor b_n\rfloor = x-y\big)\\ &\leq n\Big(\sup_{y\in\left[\tilde{\varepsilon}_n x+1,(1-\varepsilon_n)x-1\right]} \mathbb{P}\big(X_1=y\big)\Big)\mathbb{P}\big(S_{n-1}-\lfloor b_n\rfloor \geq \varepsilon_n x\big)\\ &\leq C_1' n L(x) x^{-(1+\alpha)} \tilde{\varepsilon}_n^{-(1+\alpha)} \frac{L(\tilde{\varepsilon}_n x)}{L(x)} \mathbb{P}\big(S_{n-1}-\lfloor b_n\rfloor \geq \varepsilon_n x\big)\,.\end{aligned}$$

We use the same reasoning as in Equation (4.6) to bound

$$\mathbb{P}\big(S_{n-1}-\lfloor b_n\rfloor \geq \varepsilon_n x\big) \leq C_1'' nL(\varepsilon_n x)(\varepsilon_n x)^{-\alpha}\,,$$

for some universal $C_1''>0$. This implies that for some universal $C_1>0$

$$\mathrm{B} \leq C_1 x^{-(1+\alpha)} L(x) n\Big(n\frac{L(\varepsilon_n x)L(\tilde{\varepsilon}_n x)}{L(x)} x^{-\alpha}(\tilde{\varepsilon}_n)^{-\alpha-1}(\varepsilon_n)^{-\alpha}\Big)\,.$$

3. Estimating C: it remains to bound the term C, which we split further for some $c_1>0$

$$\mathbb{P}\big(\widehat{S}_n = x, M_n \leq \tilde{\varepsilon}_n x\big) = \mathbb{P}\big(\widehat{S}_n = x, M_n \leq c_1 a_n\big) + \mathbb{P}\big(\widehat{S}_n = x, M_n \in (c_1 a_n, \tilde{\varepsilon}_n x]\big)\,.$$

The first term can be estimated using Fuk–Nagaev alone: we have that using Equation (4.2) for some other $C = C(c_1) > 0$

$$\mathbb{P}\big(\widehat{S}_n = x, M_n \le c a_n\big) = \mathcal{O}\big(\mathrm{e}^{-C\left(x/a_n\right)}\big)\,.$$

For the last remaining term, we combine the Fuk–Nagaev inequality with the tail-estimates for the random variables themselves. Note that $(c_1 a_n, \tilde{\varepsilon}_n x]$ is non-empty, as $\tilde{\varepsilon}_n x/a_n$ diverges, see beginning of this section. Abbreviate $J^- = \log_2(1/\tilde{\varepsilon}_n)$ and $J^+ = \log_2(c_1 x/a_n) - 1$. We expand

$$\mathbb{P}\big(\widehat{S}_n = x, M_n \in (c_1 a_n, \varepsilon_n x]\big) = \sum_{j=J^-}^{J^+} \mathbb{P}\big(\widehat{S}_n = x, M_n \in (2^{-(j+1)}, 2^{-j}]x\big)$$
$$\le \sum_{j=J^-}^{J^+} \big(n \sup_{y \in (2^{-(j+1)x}, 2^{-j}]} \mathbb{P}\big(X_1 = y\big)\big)\mathbb{P}\big(\widehat{S}_n = x, M_n \le 2^{-j}x\big)\,.$$

Using the tail bounds, we have that

$$\sup_{y \in (2^{-(j+1)x}, 2^{-j}]} \mathbb{P}\big(X_1 = y\big) = \mathcal{O}\big(nL\big(2^{-j}x\big)\big(2^{-j}x\big)^{-(1+\alpha)}\big) = \mathcal{O}\big(nL(x)x^{-(1+\alpha)}\big(2^j\big)^{2+\alpha}\big)\,,$$

where we used Potter's bound to see that $L\big(2^{-j}x\big) = \mathcal{O}\big(2^j L(x)\big)$. For the event $\{\widehat{S}_n = x, M_n \le 2^{-j}x\}$, we use Fuk–Nagaev to get that for $\alpha \in (0,2)$

$$\mathbb{P}\big(\widehat{S}_n = x, M_n \le 2^{-j}x\big) = \mathcal{O}\big(2^j\big)^{-2^{j-2}}\,,$$

see Berger (2019a, p. 25). For $\alpha = 2$, we first note that $x \mapsto x^{-2}\sigma^2(x)$ is eventually decreasing as $\sigma^2(x)$ is slowly varying. Hence, $2^{-j}x\sigma^2\big(2^{-j}x\big) \le a_n^{-2}\sigma^2(a_n) \sim n^{-1}$. Thus, we can bound

$$\mathbb{P}\big(\widehat{S}_n = x, M_n \le 2^{-j}x\big) = \mathcal{O}\big(\mathrm{e}^{2^j} 2^{-j2^{j+1}}\big) = \mathcal{O}\big(2^j\big)^{-2^{j-2}}\,.$$

For $\alpha > 2$, we get the same bound analogously.

Combining the above bounds gives us that for some universal $C = C(c_1) > 0$ (only depending on $\alpha > 0$)

$$\mathbb{P}\big(\widehat{S}_n = x, M_n \in (c_1 a_n, \tilde{\varepsilon}_n x]\big) = nL(x)x^{-(1+\alpha)} \sum_{j=J^-}^{J^+} \big(2^j\big)^{2+\alpha}\big(c2^j\big)^{-2^{j-2}} = nL(x)x^{-(1+\alpha)}\mathcal{O}\big(\mathrm{e}^{-C\tilde{\varepsilon}_n^{-1}}\big).$$

4. Collection of the error bounds: the previous calculations can be summarized as follows:

$$\mathbb{P}(S_n = x) = n\mathbb{P}(X_1 = x)\big(1 + \mathrm{E}\big)\,,$$

with

$$\mathrm{E} = \mathcal{O}\Big(\mathrm{err}[x, \varepsilon_n x] + \varepsilon_n + nL(x)x^{-(1+\alpha)} + \mathrm{e}^{-C\tilde{\varepsilon}_n^{-1}} + n\frac{L(\varepsilon_n x)L(\tilde{\varepsilon}_n x)}{L(x)} x^{-\alpha}(\varepsilon_n)^{-\alpha}\tilde{\varepsilon}_n^{-(1+\alpha)}\Big)\,, \tag{4.8}$$

where we note that the term $nL(\varepsilon_n x)(\varepsilon_n x)^{-\alpha}$ from Equation (4.7) has been absorbed into the last term in Equation (4.8).

The main challenge in this case is to balance the last term in Equation (4.8) with term ε_n. We do a case distinction, depending on the value of α.

The case $\boldsymbol{\alpha \in (0,2)}$: recall that by Equation (2.3), we have

$$nx^{-\alpha} \sim \Big(\frac{x}{a_n}\Big)^{-\alpha}\frac{1}{L(a_n)}\,.$$

For $\alpha \in (0,2)$, we choose $\varepsilon_n = (x/a_n)^{-\alpha_1}$, for some $\alpha_1 \in (0,1)$. This gives $x = a_n\delta_n$ with $\delta_n = (x/a_n)^{1-\alpha_1}$. Note that for any $\varepsilon_1, \varepsilon_2 > 0$ and $C > 0$ depending of $\varepsilon_1, \varepsilon_2$

$$\begin{aligned} n\frac{L(\varepsilon_n x)L(\tilde\varepsilon_n x)}{L(x)}x^{-\alpha}(\varepsilon_n)^{-\alpha}\tilde\varepsilon_n^{-(1+\alpha)} &\le C\frac{L(\varepsilon_n x)L(\tilde\varepsilon_n x)}{L(x)L(a_n)}\big(\frac{x}{a_n}\big)^{-\alpha}(\varepsilon_n)^{-\alpha}\tilde\varepsilon_n^{-(1+\alpha)} \\ &\le C\frac{L(\varepsilon_n x)L(\tilde\varepsilon_n x)}{L(x)L(a_n)}(\delta_n)^{-\alpha}\tilde\varepsilon_n^{-(1+\alpha)} \\ &\le C(\delta_n)^{-\alpha+\varepsilon_1}\tilde\varepsilon_n^{-(1+\alpha)+\varepsilon_2}\,. \end{aligned} \tag{4.9}$$

where we used the Potter bounds in Equation (4.1) twice, once with $\delta = \varepsilon_1$ and once with $\delta = \varepsilon_2$:

$$\frac{L(\varepsilon_n x)}{L(a_n)} \le \delta_n^{\varepsilon_1} \quad\text{and}\quad \frac{L(\tilde\varepsilon_n x)}{L(x)} \le \tilde\varepsilon_n^{-\varepsilon_2}\,.$$

Recall that $\tilde\varepsilon_n = \log(x/a_n)^{-2}$. We obtain (for some $\varepsilon_3, \varepsilon_3'$ which can be made arbitrarily small, as $\varepsilon_1, \varepsilon_2$ becomes small)

$$n\frac{L(\varepsilon_n x)L(\tilde\varepsilon_n x)}{L(x)}x^{-\alpha}(\varepsilon_n)^{-\alpha}\tilde\varepsilon_n^{-(1+\alpha)} \le C(\delta_n)^{-\alpha+\varepsilon_3} \le C\big(\frac{x}{a_n}\big)^{-(1-\alpha_1)\alpha-\varepsilon_3'}\,.$$

For $\alpha_1 = \alpha/(\alpha+1)$, both terms are approximately equal and we hence obtain

$$\varepsilon_n + \big(\frac{x}{a_n}\big)^{-(1-\zeta)\alpha-\varepsilon_3'} = \big(\frac{x}{a_n}\big)^{-\zeta} + \big(\frac{x}{a_n}\big)^{-(1-\zeta)\alpha-\varepsilon_3'} \le 2\big(\frac{x}{a_n}\big)^{-\frac{\alpha}{\alpha+1}+\varepsilon''}\,,$$

for some $\varepsilon'' = o(1)$ as $\varepsilon_1, \varepsilon_2 \downarrow 0$.

The previous equation reduces the error in Equation (4.1) (as the other terms are negligible) to

$$\mathrm{E} = \mathcal{O}\big(\mathrm{err}[x, a_n\delta_n] + 2\big(\frac{a_n}{x}\big)^{\alpha_1-\varepsilon''}\big)\,,$$

with $\varepsilon'' > 0$ as small as we want. This concludes the proof of Theorem 2.1 for the case $\alpha \in (0,2)$.

The case $\boldsymbol{\alpha} \in (\mathbf{2}, \infty)$: recall $\alpha_1 = \alpha/(1+\alpha)$. Choose the largest possible $\beta \ge 0$ such that

$$n^{-\beta}\big(\frac{x}{\sqrt{n\log(n)}}\big)^{1-\alpha_1} \to \infty \qquad\text{and}\qquad \beta \le \frac{(\alpha-2)(\alpha+1)}{2(2\alpha+1)}\,.$$

Choose $\varepsilon_n = n^{-\beta}\big(\frac{x}{\sqrt{n\log(n)}}\big)^{-\alpha_1}$. Note that this allows us to rewrite

$$nx^{-\alpha}\varepsilon_n^{-\alpha} = n^{1-\alpha/2}\big(\frac{x}{\sqrt{n\log(n)}}\big)^{-\alpha}\varepsilon_n^{-\alpha}\log^{-\alpha/2}(n) = n^{1-\alpha/2+\beta\alpha_1}\big(\frac{x}{\sqrt{n\log(n)}}\big)^{-\alpha(1-\alpha_1)}\log^{-\alpha/2}(n)\,.$$

Note that for the choices of α_1, β, we have that $-\alpha_1 = -\alpha(1-\alpha_1)$ and $1-\alpha/2+\beta\alpha_1 \le -\beta$. Hence, we get that

$$\begin{aligned} \mathcal{O}\big(nx^{-\alpha}\varepsilon_n^{-\alpha} + \varepsilon_n\big) &= \mathcal{O}\big(n^{1-\alpha/2+\beta\alpha_1}\big(\frac{x}{\sqrt{n\log(n)}}\big)^{-\alpha(1-\alpha_1)}\log^{-\alpha/2}(n) + n^{-\beta}\big(\frac{x}{\sqrt{n\log(n)}}\big)^{-\alpha_1}\big) \\ &= \mathcal{O}\big(n^{1-\alpha/2+\beta\alpha_1}\big(\frac{x}{\sqrt{n\log(n)}}\big)^{-\alpha_1}\big)\,. \end{aligned}$$

As in the case $\alpha \in (0,2)$, the slowly varying functions add at most a power of $\varepsilon'' > 0$, where we can choose $\varepsilon'' > 0$ as small as we desire. Furthermore, $nx^{-(1+\alpha)}L(x) = o\big(nx^{-\alpha}\varepsilon_n^{-\alpha}\big)$. This gives

$$\mathrm{E} = \mathcal{O}\big(n^{1-\alpha/2+\beta\alpha_1}\big(\frac{x}{\sqrt{n\log(n)}}\big)^{-\alpha_1+\varepsilon''} + \mathrm{err}[x, a_n\delta_n]\big)$$

The positive ε'' easily absorbs the decay of order $\log(n)$. Hence, we can conclude the proof as we did in the case $\alpha \in (0,2)$.

The case $\boldsymbol{\alpha = 2}$: Choose $\varepsilon_n = \big(\frac{x}{a_n\sqrt{\log(n)}}\big)^{-\alpha_1}$, for $\alpha_1 = 2/3$. We then have that

$$nx^{-2}\varepsilon_n^{-2} = \big(\frac{x}{a_n\sqrt{\log(n)}}\big)^{-\alpha_1}\big(\frac{\sqrt{n}}{a_n\sqrt{\log(n)}}\big)^2 .$$

This gives that

$$\mathcal{O}\big(nx^{-2}\varepsilon_n^{-2} + \varepsilon_n\big) = \mathcal{O}\big(\big(\frac{x}{a_n\sqrt{\log(n)}}\big)^{-\alpha_1}\big) .$$

From there on, we proceed as in the case $\alpha > 2$, noting that $\beta = 0$. □

Remark 4.3. We expect the error calculated above to be essentially optimal (up to the $\varepsilon > 0$ which can be chosen as small as we want). Indeed, probabilistically, there are two sources of errors: the maximum can deviate from x by $\varepsilon_n x$. This gives an error of $\mathcal{O}\big(\varepsilon_n\big)$. This error shrinks as we make ε_n small. However, the remaining sum compensating by being larger/smaller than their CLT scale gives an error of $\mathcal{O}\big((x/a_n)^{-\alpha}\varepsilon_n^{-\alpha}\big)$, which shrinks as we increase ε_n, see Equation (4.9). Both error terms are optimal in the sense that we cannot replace $\mathcal{O}(\ldots)$ by $o(\ldots)$, see Berger (2019a, Theorem 2.1). Our choice of ε_n makes the two errors asymptotically equal, selecting the minimal possible error.

4.3. *Proof of Theorem 2.3.* The proof of Theorem 2.3 follows the same steps as the one of Theorem 2.1: we first split the probability

$$\mathbb{P}(\widehat{S}_n \geq x) = \mathbb{P}\big(\widehat{S}_n \geq x, M_n \geq (1-\varepsilon_n)x\big) + \mathbb{P}\big(\widehat{S}_n \geq x, M_n < (1-\varepsilon_n)x\big) .$$

The second term is negligible and produces the same errors as the terms B and C in the proof of Theorem 2.1.

We upper bound the first term

$$\mathbb{P}\big(\widehat{S}_n \geq x, M_n \geq (1-\varepsilon_n)x\big) \leq n\mathbb{P}\big(X_1 > (1-\varepsilon_n)x\big) = n\mathbb{P}\big(X_1 > x\big)\big(1 + \mathrm{err}[\varepsilon_n, x] + \mathcal{O}(\varepsilon_n)\big) .$$

The lower bound is analogous to Equation (4.3) and is hence omitted. This concludes the proof of Theorem 2.3. □

4.4. *Proof of Theorem 2.5 and Theorem 2.7.* In this section we prove Theorem 2.5 and Theorem 2.7. Theorem 2.5 will be proved in full detail while for Theorem 2.7 we just highlight the differences with Theorem 2.5.

Set $x^- = x - \omega(a_n)$ for $\omega(a_n) > 0$ fixed . Recall that

$$\nu_x = \mathbb{P}(X_1 \in \cdot\,|X_1 > x - \omega(a_n)) \qquad \text{and} \qquad \nu_{x,n} = \mathbb{P}\big(\{X_i\}_{i=1,\ldots,n} \in \cdot\,|S_n > x\big) .$$

Let

$$\nu^*_{x,n} = \frac{1}{n}\sum_{j=1}^n \sigma^j\big(\mu^{\otimes(n-1)} \otimes \nu_x\big) ,$$

where σ^j switches the last coordinate with the j-th coordinate. We then have that using Pinsker's inequality and Csiszár's parallelogram identity (see Armendáriz and Loulakis (2011))

$$||\nu_{x,n} - \nu^*_{x,n}||^2 \leq H(\nu_{x,n}|\mu^{\otimes n}) + H(\nu^*_{x,n}|\mu^{\otimes n}) - 2H\big(\frac{\nu_{x,n}+\nu^*_{x,n}}{2}\big|\mu^{\otimes n}\big) = \mathrm{A} + \mathrm{B} - \mathrm{C} ,$$

where

$$H(\mu|\nu) = \begin{cases} \int \frac{\mathrm{d}\mu}{\mathrm{d}\nu}\log\big[\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\big]\mathrm{d}\nu & \text{if } \mu \ll \nu , \\ +\infty & \text{otherwise.} \end{cases}$$

Note that for $y \in \mathbb{R}^n$

$$\frac{\mathrm{d}\nu_{x,n}}{\mathrm{d}\mu^{\otimes n}}(y) = \frac{\mathbf{1}\{S_n(y) > x\}}{G_n(x)} ,$$

where $G_n(x) = \mathbb{P}(S_n > x)$ and $S_n(y) = \sum_{i=1}^n y_i$. Note that

$$\frac{\mathrm{d}\nu^*_{x,n}}{\mathrm{d}\mu^{\otimes n}}(y) = \frac{1}{nG(x^-)} \sum_{i=1}^n \mathbf{1}\{y_i > x^-\}$$

where we recall $G(t) = \mathbb{P}(X_1 > t)$.
We have that

$$\mathrm{A} + \mathrm{B} = H(\nu_{x,n}|\mu^{\otimes n}) + H(\nu^*_{x,n}|\mu^{\otimes n}) = \int \log N_{x,n} \mathrm{d}\nu_{x,n} - \log\big(G_n(x) nG(x^-)\big)\,,$$

where $N_{x,n}$ counts the number of coordinates larger than x^-. Note that

$$\int \log N_{x,n} \mathrm{d}\nu_{x,n} = \sum_{k\geq 2} \log(k) \nu_{x,n}(N_{x,n} = k) = \sum_{k\geq 2} \log(k) \binom{n-1}{k-1} G(x^-)^{k-1} = \mathcal{O}\big(nG(x)\big)\,.$$

Thus

$$\mathrm{A} + \mathrm{B} = -\log\big(G_n(x) nG(x^-)\big) + \mathcal{O}\big(nG(x)\big)\,. \tag{4.10}$$

On the other hand,

$$\mathrm{C} = 2H\Big(\frac{\nu_{x,n} + \nu^*_{x,n}}{2}\Big|\mu^{\otimes n}\Big) = \int \big[\frac{\mathrm{d}\nu_{x,n}}{\mathrm{d}\mu^{\otimes n}} + \frac{\mathrm{d}\nu^*_{x,n}}{\mathrm{d}\mu^{\otimes n}}\big] \log\big[\frac{\mathrm{d}\nu_{x,n}}{2\mathrm{d}\mu^{\otimes n}} + \frac{\mathrm{d}\nu^*_{x,n}}{2\mathrm{d}\mu^{\otimes n}}\big] \mathrm{d}\mu^{\otimes n}\,.$$

We split the integrand into two: for the first part, we estimate

$$\int \frac{\nu^*_{x,n}}{\mathrm{d}\mu^{\otimes n}} \log\big[\frac{\mathrm{d}\nu_{x,n}}{2\mathrm{d}\mu^{\otimes n}} + \frac{\mathrm{d}\nu^*_{x,n}}{2\mathrm{d}\mu^{\otimes n}}\big] \mathrm{d}\mu^{\otimes n} = \int \frac{N_{x,n}}{nG(x^-)} \log\big[\frac{\mathrm{d}\nu^*_{x,n}}{2\mathrm{d}\mu^{\otimes n}} + \frac{1}{2G_n(x)}\big] \mathrm{d}\mu^{\otimes n}$$
$$= \int \frac{\mathbf{1}_{N_{x,n}=1}}{nG(x^-)} \log\big[\frac{1}{2nG(x)} + \frac{1}{2G_n(x)}\big] \mathrm{d}\mu^{\otimes n} + \int \frac{N_{x,n}\mathbf{1}_{N_{x,n}>1}}{nG(x^-)} \log\big[\frac{\mathrm{d}\nu^*_{x,n}}{2\mathrm{d}\mu^{\otimes n}} + \frac{1}{2G_n(x)}\big] \mathrm{d}\mu^{\otimes n}\,.$$

Note that by the inclusion-exclusion principle

$$\mu^{\otimes n}\big(N_{x,n} = 1\big) = nG(x^-) + \sum_{k=2}^n (-1)^{k-1} \binom{n}{k} G(x^-)^k = nG(x^-)\big(1 + \mathcal{O}\big(nG(x^-)\big)\big)\,.$$

Hence,

$$\int \frac{\mathbf{1}_{N_{x,n}=1}}{nG(x^-)} \log\big[\frac{1}{2nG(x^-)} + \frac{1}{2G_n(x)}\big] \mathrm{d}\mu^{\otimes n} = \big(1 + \mathcal{O}\big(G(x^-)n\big)\big) \log\big[\frac{1}{2nG(x^-)} + \frac{1}{2G_n(x)}\big]$$
$$= -\log\big[nG(x)\big]\big(1 + \mathcal{O}\big(\mathrm{err}^{(1)}[x, \omega(a_n)] + c_{x,n} + G(x^-)n\big)\big)\,. \tag{4.11}$$

Indeed,

$$G(x^-) = G(x)\big(1 + \mathrm{err}^{(1)}[x, \omega(a_n)]\big)\,,$$

and

$$G_n(x) = nG(x)\big(1 + \mathcal{O}(c_{x,n})\big)\,.$$

The error term is given by

$$\int \frac{N_{x,n}\mathbf{1}_{N_{x,n}>1}}{nG(x)} \log\big[\frac{\mathrm{d}\nu^*_{x,n}}{2\mathrm{d}\mu^{\otimes n}} + \frac{1}{2G_n(x)}\big] \mathrm{d}\mu^{\otimes n} \leq C \sum_{k=2}^n \big[nG(x)\big]^{k-1} k \binom{n}{k} \log[nG(x)]\,,$$

and thus (noting that the term $k = 2$ dominates)

$$\int \frac{\mathrm{d}\nu_{x,n}}{\mathrm{d}\mu^{\otimes n}} \log\big[\frac{\mathrm{d}\nu_{x,n}}{2\mathrm{d}\mu^{\otimes n}} + \frac{\mathrm{d}\nu^*_{x,n}}{2\mathrm{d}\mu^{\otimes n}}\big] \mathrm{d}\mu^{\otimes n} = -\log\big[nG_n(x)\big]\big(1 + \mathcal{O}\big(\mathrm{err}^{(1)}[x, \omega(a_n)] + c_{x,n} + nG(x)\big)\big)\,.$$

For the second term, note that

$$\int \frac{\mathrm{d}\nu^*_{x,n}}{\mathrm{d}\mu^{\otimes n}} \log\big[\frac{\mathrm{d}\nu_{x,n}}{2\mathrm{d}\mu^{\otimes n}} + \frac{\mathrm{d}\nu^*_{x,n}}{2\mathrm{d}\mu^{\otimes n}}\big] \mathrm{d}\mu^{\otimes n} = \int \frac{\mathbf{1}\{S_n > x\}}{G_n(x)} \log\big[\frac{N_{x,n}}{2nG(x)} + \frac{1}{2G_n(x)}\big] \mathrm{d}\mu^{\otimes n}\,.$$

Note that

$$\int \frac{\mathbf{1}\{S_n > x,\, N_{x,n} = 0\}}{G_n(x)} \log\big[\frac{N_{x,n}}{2nG(x)} + \frac{1}{2G_n(x)}\big] \mathrm{d}\mu^{\otimes n} = \mathcal{O}\big(c_{x,n} \log\big(nG(x)\big)\big)\,.$$

On the other hand,

$$\int \frac{\mathbf{1}\{S_n > x,\, N_{x,n} \geq 2\}}{G_n(x)} \log\big[\frac{N_{x,n}}{2nG(x)} + \frac{1}{2G_n(x)}\big] \mathrm{d}\mu^{\otimes n} = \mathcal{O}\big(\log\big(nG(x)\big)nG(x)\big)\,,$$

similar to before. We estimate the final contribution

$$\int \frac{\mathbf{1}\{S_n > x,\, N_{x,n} = 1\}}{G_n(x)} \log\big[\frac{1}{2nG(x)} + \frac{1}{2G_n(x)}\big] \mathrm{d}\mu^{\otimes n}$$
$$= -\log\big[G_n(x)\big]\big(1 + \mathcal{O}\big(\mathrm{err}^{(1)}[x, \omega(a_n)] + c_{x,n}\big)\big)\,.$$

This is done analogously to Equation (4.11). Combining the above bounds yields that

$$\mathrm{C} = -2\log\big[G_n(x)\big]\big(1 + \mathcal{O}\big(\mathrm{err}^{(1)}[x, \omega(a_n)] + c_{x,n} + nG(x)\big)\big)\,.$$

As $||\nu_{x,n} - \nu^*_{x,n}||^2 \leq \mathrm{A} + \mathrm{B} - \mathrm{C}$, we have using Equation (4.10)

$$||\nu_{x,n} - \nu^*_{x,n}||^2 \leq \mathcal{O}\big(\mathrm{err}^{(1)}[x, \omega(a_n)] + c_{x,n} + nG(x)\big)\,.$$

This concludes the proof of Theorem 2.5. □

The proof of Theorem 2.7 can now be carried out in exactly the same manner: note that for $y \in \mathbb{R}^n$

$$\frac{\mathrm{d}\xi^*_{x,n}}{\mathrm{d}\mu^{\otimes n}}(y) = \frac{\mathbf{1}\{S_{n-1}(y) = x - y_n\}}{\mathbb{P}(X_n = S_{n-1} - x)}\,.$$

By conditioning on X_n, we can establish

$$\mathbb{P}(X_n = S_{n-1} - x) = nG(x)\big(1 + \mathrm{err}^{(2)}[\omega(x, a_n)] + c_{x,n}\big)\,.$$

On the other hand,

$$\frac{\mathrm{d}\xi_{x,n}}{\mathrm{d}\mu^{\otimes n}}(y) = \frac{\mathbf{1}\{S_n(y) = x\}}{\mathbb{P}(S_n = x)}\,.$$

We can now apply the error bounds from Theorem 2.1 together with the method from the proof of Theorem 2.5 to conclude the proof.

Acknowledgements

The author would like to express his gratitude for the anonymous referee for suggesting several improvements for this paper, most notably the extension from $\alpha \in (0, 2)$ to all $\alpha > 0$. The author would also like the thank the referee for pointing out a calculation mistake in an earlier version of the paper. The author would like to thank Quentin Berger for his help, answering my questions both quickly and patiently. The author would also like to thank Silke Rolles and Julius Damarackas for their help regarding typos and presentation.

References

Armendáriz, I. and Loulakis, M. Conditional distribution of heavy tailed random variables on large deviations of their sum. *Stochastic Process. Appl.*, **121** (5), 1138–1147 (2011). MR2775110.

Bahadur, R. R. and Ranga Rao, R. On deviations of the sample mean. *Ann. Math. Statist.*, **31**, 1015–1027 (1960). MR117775.

Berger, Q. Notes on random walks in the Cauchy domain of attraction. *Probab. Theory Related Fields*, **175** (1-2), 1–44 (2019a). MR4009704.

Berger, Q. Strong renewal theorems and local large deviations for multivariate random walks and renewals. *Electron. J. Probab.*, **24**, Paper No. 46, 47 (2019b). MR3949271.

Berger, Q., Birkner, M., and Yuan, L. Collective vs. individual behaviour for sums of iid random variables: appearance of the one-big-jump phenomenon. *ArXiv Mathematics e-prints* (2023). arXiv: 2303.12505.

Bingham, N. H., Goldie, C. M., and Teugels, J. L. *Regular variation*, volume 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge (1989). ISBN 0-521-37943-1. MR1015093.

Blackwell, D. and Hodges, J. L., Jr. The probability in the extreme tail of a convolution. *Ann. Math. Statist.*, **30**, 1113–1120 (1959). MR112197.

Dembo, A. and Zeitouni, O. *Large deviations techniques and applications*, volume 38 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin (2010). ISBN 978-3-642-03310-0. MR2571413.

Doney, R. A. A local limit theorem for moderate deviations. *Bull. London Math. Soc.*, **33** (1), 100–108 (2001). MR1798582.

Feller, W. *An introduction to probability theory and its applications. Vol. I.* John Wiley & Sons, Inc., New York-London-Sydney, third edition (1968). MR228020.

Großkinsky, S., Schütz, G. M., and Spohn, H. Condensation in the zero range process: stationary and dynamical properties. *J. Statist. Phys.*, **113** (3-4), 389–410 (2003). MR2013129.

Klüppelberg, C. and Mikosch, T. Large deviations of heavy-tailed random sums with applications in insurance and finance. *J. Appl. Probab.*, **34** (2), 293–308 (1997). MR1447336.

Mikosch, T. and Rodionov, I. Precise large deviations for dependent subexponential variables. *Bernoulli*, **27** (2), 1319–1347 (2021). MR4255236.

Nagaev, S. V. Large deviations of sums of independent random variables. *Ann. Probab.*, **7** (5), 745–789 (1979). MR542129.

Nagaev, S. V. On the asymptotic behavior of probabilities of one-sided large deviations. *Theory Probab. Appl.*, **26** (2), 362–366 (1982). DOI: 10.1137/1126035.

Tchachkuk, S. Limit theorems for sums of independent random variables belonging to the domain of attraction of a stable law (1977). Candidate's dissertation, Tashent (in Russian).

Yang, Y., Leipus, R., and Šiaulys, J. Local precise large deviations for sums of random variables with O-regularly varying densities. *Statist. Probab. Lett.*, **80** (19-20), 1559–1567 (2010). MR2669760.