A conditional Berry-Esseen bound and a conditional large deviation result without Laplace transform. Application to hashing with linear probing

Thierry Klein, A Lagnoux, P Petit

A conditional Berry-Esseen bound and a conditional large deviation result without Laplace transform. Application to hashing with linear probing.

T. Klein, A. Lagnoux, P. Petit

December 24, 2014

Abstract

We study the asymptotic behavior of a sum of independent and identically distributed random variables conditioned by a sum of independent and identically distributed integer-valued random variables. We prove a Berry-Esseen bound in a general setting and a large deviation result when the Laplace transform of the underlying distribution is not defined in a neighborhood of zero. Then we present several combinatorial applications. In particular, we prove a large deviation result for the model of hashing with linear probing.

Keywords: Berry-Esseen bound ; large deviations ; conditional distribution ; combinatorial problems ; hashing with linear probing.
AMS MSC 2010: 60F10; 60F05; 62E20; 60C05; 68W40.

1 Introduction

As pointed out by Svante Janson in his seminal work [13], in many random combinatorial problems, the interesting statistic is the sum of independent and identically distributed (i.i.d.) random variables conditioned by some exogenous integer random variable. In general, this exogenous random variable is itself a sum of integer-valued random variables. A general framework for this kind of problem may be formalized as follows. In the whole paper, $\mathbb{N}^*$ will denote the set $\{1,2,\ldots\}$ of positive integers, $\mathbb{N}=\mathbb{N}^*\cup\{0\}$, and $\mathbb{Z}$ will be the set of all integers. Let $(k_n)_{n\in\mathbb{N}^*}$ be a sequence of integers and $(N_n)_{n\in\mathbb{N}^*}$ be a sequence of positive integers. Further, let $(X_j^{(n)},Y_j^{(n)})_{n\in\mathbb{N}^*,j=1,\ldots,N_n}$ be a triangular array of pairs of random variables such that each line contains i.i.d. copies of a pair $(X^{(n)},Y^{(n)})$ of random variables. Moreover, it is assumed that the elements of the array $(X_j^{(n)})_{n\in\mathbb{N}^*,j=1,\ldots,N_n}$ are integers. We are interested in the law of $(N_n)^{-1}T_n:=(N_n)^{-1}\sum_{j=1}^{N_n}Y_j^{(n)}$ conditioned on a specific value of $S_n:=\sum_{j=1}^{N_n}X_j^{(n)}$; that is to say in the conditional distribution

$$\mathcal{L}_n:=\mathcal{L}((N_n)^{-1}T_n|S_n=k_n).$$

The motivation for considering distributions of $(X^{(n)},Y^{(n)})$ that depend on n comes from the discrete nature of the problem that can lead to a degenerated conditional law as soon as $\mathbb{P}(S_n=k_n)=0$. Nevertheless in many applications (e.g., occupancy problem or hashing ; see [13]), the distribution of the conditioning random variable X depends on a parameter λ that can be freely chosen: for example, $\lambda\in\mathbb{R}$ is the parameter of a Poisson distribution in the occupancy problem and $\lambda\in]0,e^{-1}]$ is the parameter of the Borel distribution for hashing. One can take advantage of this fact to overcome contexts in which $\mathbb{P}(S_n=k_n)=0$ proceeding as follows. Consider a triangular array $(X_j^{(n)},Y_j^{(n)})_{n\in\mathbb{N}^*,j=1\ldots N_n}$ such that $(X^{(n)},Y^{(n)})$ converges weakly to (X,Y). Then choose a sequence of parameters $\lambda_n\to\lambda$ such that, for any n, $\mathbb{P}(\sum_{j=1}^{N_n}X_j^{(n)}=k_n)>0$.
In his work, Janson proves a general central limit theorem (with convergence of all moments) for this kind of conditional distribution under some reasonable assumptions and gives several applications in classical com-

binatorial problems: occupancy in urns, hashing with linear probing, random forests, branching processes, etc. Following this work, at least two natural questions arise:

1. is it possible to obtain a general Berry-Esseen bound for these models?

2. is it possible to obtain a general large deviation result for these models?

A Berry-Esseen theorem is given by Quine and Robinson [25]. In their work, the authors study the particular case of the occupancy problem where the random variables $X^{(n)}$ are Poisson distributed and $Y^{(n)} = 1_{\{X^{(n)}=0\}}$. Up to our knowledge, it is the only result in that direction for this kind of conditional distribution. In our work, we prove a general Berry-Esseen bound (Theorem 2.1) that covers all the examples presented by Janson [13].
When the distribution of $(X^{(n)}, Y^{(n)})$ does not depend on n, the Gibbs conditioning principle ([28, 4, 5]) states that $\mathcal{L}_n$ converges weakly to the degenerated distribution concentrated on a point χ depending on the conditioning value (see [9, Corollary 2.2]). Around the Gibbs conditioning principle, general limit theorems yielding the asymptotic behavior of the conditioned sum are given in [27, 11, 18] and asymptotic expansions are proved in [10, 26]. In this paper our aim is to prove a large deviation result for $\mathcal{L}_n$, when the joint Laplace transform of $(X_j^{(n)}, Y_j^{(n)})$ is not defined everywhere: we give an exponential equivalent for this conditional distribution.
The case when the Laplace transform is defined has been treated by Gamboa, Klein and Prieur [9]. They prove a large (and a moderate) deviation principle under some strong assumptions. The most restricting assumption states that the joint Laplace transform of $(X^{(n)}, Y^{(n)})$ is finite at least in a neighborhood of $(0, 0)$. Unfortunately, this assumption fails to be satisfied for the most interesting example presented in [13]: hashing with linear probing. In this case, the joint Laplace transform is only defined in $]-\infty, a]\times]-\infty, 0]$ for some positive a. It is then natural to extend the work of [9] for such distributions. In [21, 22], Nagaev establishes large deviation results for sums of random variables which are absolutely continuous with respect to the Lebesgue measure and the Laplace transform of which is not defined in a neighborhood of 0. Following this work, we prove a large deviation result (Theorem 2.4).
Let us point out the main differences between Theorem 2.4 of the present work and Theorem 2.1 of [9]. First, the proof in [9] is based on a sharp control of a Fourier-Laplace transform $\Phi_{X^{(n)},Y^{(n)}}(t, u) := \mathbb{E}\left(\exp[itX^{(n)} + uY^{(n)}]\right)$ of $\left(X^{(n)}, Y^{(n)}\right)$. The Fourier part allows to treat the conditioning whereas the Laplace one allows to apply Gärtner-Ellis theorem. In the present paper, the proof follows ideas borrowed from [21, 22]. More precisely, contrary to the case when the Laplace transform is defined, the large deviations of the sum of the random variables with heavy-tailed distributions is due to exceptional values taken by few random variables. Second, unlike the classical speeds in N_n obtained either in Cramér's theorem or in Theorem 2.1 of [9], the speed in this paper is $\sqrt{N_n}$. Third, one originality of our work is that the lower and upper bounds may differ (see equations (6) and (7)). When the Laplace transform is defined, the tails are controlled (see Cramér's theorem or Gärtner-Ellis theorem in [5]) and the sum satisfies a large deviation principle with the same lower and upper bounds. Here, as opposed to previous classical theorems, one may allow oscillations of the tails (in a controlled range) that lead to a large deviation result with two different bounds. Last but not least, the rate function obtained is not affected by the conditioning variable: the rate functions are the same in the conditional case and in the unconditional one (see Theorems 2.4 and 2.6). On the contrary, when the Laplace transform is defined in a neighborhood of the origin, the rate function strongly depends on the dependence between $X^{(n)}$ and $Y^{(n)}$. It is $y \mapsto \psi^*_{X^{(n)},Y^{(n)}}(\lambda, y) - \psi^*_{X^{(n)}}(\lambda)$ (where λ is the limit of the ratio k_n/N_n), the difference between the joint Fenchel-Legendre transform and the Fenchel-Legendre transform of the conditioning random variable $X^{(n)}$. This rate function is $y \mapsto \psi^*_{Y^{(n)}}(y)$ when the conditioning term is ineffective, that is to say when the random variables $X^{(n)}$ and $Y^{(n)}$ are independent.
As pointed out by Janson in [13], hashing with linear probing was the motivating example for his work (see section 3 for a complete description of the model). This model comes from theoretical computer science, where it modelizes the time cost to store data in the memory. Then, it was introduced in a mathematical framework by Knuth [16]. Due to its strong connection with parking functions, the Airy distributions (i.e.,

the area under the brownian excursion), this model was studied by many authors (see, e.g., Flajolet, Poblete and Viola [8], Janson [12, 14, 15], Chassaing, Janson, Louchard and Marckert [2, 1, 3], and Marckert [20]). Theorem 2.4 allows to treat the interesting example of hashing with linear probing: Proposition 3.3 is the formulation of Theorem 2.4 in this particular framework.
The paper is organized as follows. In section 2, we present the general model and give our two main theorems. First we prove a Berry-Esseen bound (Theorem 2.1) and show how it straightforwardly applies to the examples presented by Janson [13]. Second we establish a large deviation result (Theorem 2.4). Section 3 is devoted to the study of hashing with linear probing. Finally, we prove our main results in the last section.

2 Main results

2.1 Framework and notation

For all $n \geqslant 1$, we consider a pair of random variables $\left(X^{(n)}, Y^{(n)}\right)$ such that $X^{(n)}$ is integer-valued and $Y^{(n)}$ real-valued. Let N_n be a natural number such that $N_n \to +\infty$ as n goes to infinity. Let $\left(X_i^{(n)}, Y_i^{(n)}\right)$ $(i = 1, 2, \ldots, N_n)$ be an i.i.d. sample distributed as $\left(X^{(n)}, Y^{(n)}\right)$ and define

$$S_n := \sum_{i=1}^{N_n} X_i^{(n)} \quad \text{and} \quad T_n := \sum_{i=1}^{N_n} Y_i^{(n)}.$$

Let $k_n \in \mathbb{Z}$ be such that $\mathbb{P}(S_n = k_n) > 0$ and let U_n be a random variable distributed as T_n conditioned on $S_n = k_n$. We establish a Berry-Esseen bound and a large deviation result for $(U_n)_{n\geqslant 1}$.

2.2 Conditional Berry-Esseen bound

Theorem 2.1. *Suppose that there exist positive constants $\tilde{c}_1$, c_1, c_2, $\tilde{c}_3$, c_3, c_4, c_5, and c_6 such that:*

(H2.1.1) $\tilde{c}_1 \leqslant \sigma_{X^{(n)}} := \mathrm{Var}\left(X^{(n)}\right)^{1/2} \leqslant c_1$;

(H2.1.2) $\rho_{X^{(n)}} := \mathbb{E}\left[\left|X^{(n)} - \mathbb{E}\left[X^{(n)}\right]\right|^3\right] \leqslant c_2^3 \sigma_{X^{(n)}}^3$;

(H2.1.3) *define* $Y'^{(n)} := Y^{(n)} - X^{(n)} \mathrm{Cov}(X^{(n)}, Y^{(n)})/\sigma^2_{X^{(n)}}$, *there exists* $\eta_0 > 0$ *such that, for all* $s \in [-\pi, \pi]$ *and* $t \in [0, \eta_0]$,

$$\left|\mathbb{E}\left[e^{i(sX^{(n)} + tY'^{(n)})}\right]\right| \leqslant 1 - c_5\left(\sigma^2_{X^{(n)}} s^2 + \sigma^2_{Y'^{(n)}} t^2\right);$$

(H2.1.4) $k_n = N_n \mathbb{E}\left[X^{(n)}\right] + O(\sigma_{X^{(n)}} N_n^{1/2})$ *(remind that* $k_n \in \mathbb{Z}$ *and* $\mathbb{P}(S_n = k_n) > 0$*)*;

(H2.1.5) $\tilde{c}_3 \leqslant \sigma_{Y^{(n)}} := \mathrm{Var}\left(Y^{(n)}\right)^{1/2} \leqslant c_3$;

(H2.1.6) $\rho_{Y^{(n)}} := \mathbb{E}\left[\left|Y^{(n)} - \mathbb{E}\left[Y^{(n)}\right]\right|^3\right] \leqslant c_4^3 \sigma_{Y^{(n)}}^3$;

(H2.1.7) *the correlation* $r_n := \mathrm{Cov}\left(X^{(n)}, Y^{(n)}\right) \sigma^{-1}_{X^{(n)}} \sigma^{-1}_{Y^{(n)}}$ *satisfies* $|r_n| \leqslant c_6 < 1$, *so that*

$$\tau_n^2 := \sigma^2_{Y^{(n)}}(1 - r_n^2) \geqslant \tilde{c}_2^2(1 - c_6^2) > 0.$$

Then the following conclusions hold.

2.1.a. There exists $\tilde{c}_5 > 0$ such that

$$\mathbb{P}(S_n = k_n) \geqslant \frac{\tilde{c}_5}{2\pi \sigma_{X^{(n)}} N_n^{1/2}}.$$

2.1.b. For $N_n \geqslant N_0 := \max(3, c_2^6, c_4^6)$, the conditional distribution of

$$N_n^{-1/2}\tau_n^{-1}(T_n - N_n\mathbb{E}[Y^{(n)}] - r_n\frac{\sigma_{Y^{(n)}}}{\sigma_{X^{(n)}}}(k_n - N_n\mathbb{E}[X^{(n)}]))$$

given $S_n = k_n$ satisfies the Berry-Esseen inequality

$$\sup_x \left| \mathbb{P}\left(\frac{U_n - N_n\mathbb{E}\left[Y^{(n)}\right] - r_n\sigma_{Y^{(n)}}\sigma_{X^{(n)}}^{-1}(k_n - N_n\mathbb{E}\left[X^{(n)}\right])}{N_n^{1/2}\tau_n} \leqslant x \right) - \Phi(x) \right| \leqslant \frac{C}{N_n^{1/2}}, \tag{1}$$

where Φ denotes the standard normal probability distribution, and C is a positive constant that only depends on $\tilde{c}_1$, c_1, c_2, $\tilde{c}_3$, c_3, c_4, c_5, $\tilde{c}_5$, and c_6.

2.1.c. Moreover, there exist two positive constants c_7 and c_8 only depending on $\tilde{c}_1$, c_1, c_2, $\tilde{c}_3$, c_3, c_4, c_5, $\tilde{c}_5$, and c_6 such that

$$\left| \mathbb{E}\left[U_n\right] - N_n\mathbb{E}[Y^{(n)}] - r_n\frac{\sigma_{Y^{(n)}}}{\sigma_{X^{(n)}}}(k_n - N_n\mathbb{E}[X^{(n)}]) \right| \leqslant c_7 \tag{2}$$

and

$$\left|\mathrm{Var}\left(U_n\right) - N_n\tau_n^2\right| \leqslant c_8 N_n^{1/2} \tag{3}$$

If $N_n \geqslant \tilde{N}_0 := \max(N_0, 4c_8^2/\tilde{c}_3^2)$, we also have

$$\sup_x \left| \mathbb{P}\left(\frac{U_n - \mathbb{E}\left[U_n\right]}{\mathrm{Var}\left(U_n\right)^{1/2}} \leqslant x \right) - \Phi(x) \right| \leqslant \frac{\tilde{C}}{N_n^{1/2}}, \tag{4}$$

where $\tilde{C}$ is a constant that only depends on $\tilde{c}_1$, c_1, c_2, $\tilde{c}_3$, c_3, c_4, c_5, $\tilde{c}_5$, and c_6. This result means that U_n is asymptotically normal.

Remark 2.2.

1. The fact that $N_n \to +\infty$ is only required for the existence of the constant $\tilde{c}_5$ which relies on Lebesgue dominated convergence theorem.
2. The set of hypotheses of Theorem 2.1 implies the one of the central limit theorem stated in [13, Theorem 2.1] which is clearly not surprising. Notice that by assumption (H2.1.4), the conditioning is approximately equal to the mean as in the central limit theorem given in [13, Theorem 2.3].
3. As a consequence of Proposition 4.4 below, $\tilde{c}_1$ can be chosen as $c_2^{-3}/4$.
4. Assumption (H2.1.7) is not very restricting as we will see later in the examples.
5. One should note that 2.1.a is the analogue of Equation (7) of Lemma 3.2 in [9].
6. In the proof, we will replace $Y^{(n)}$ by the projection $Y'^{(n)}$ in order to work with a centered variable which is also uncorrelated with $X^{(n)}$. We introduce $Y'^{(n)}$ for that purpose.
7. If (X, Y') is a pair of random variables such as the correlation r satisfies $|r| < 1$, then

$$\begin{aligned}\left|\mathbb{E}\left[e^{i(sX+tY')}\right]\right| &= 1 - \frac{1}{2}\left(\sigma_X^2 s^2 + 2\sigma_X\sigma_{Y'}rst + \sigma_{Y'}^2 t^2\right) + o(s^2+t^2) \\ &\leqslant 1 - \frac{1-|r|}{2}\left(\sigma_X^2 s^2 + \sigma_{Y'}^2 t^2\right) + o(s^2+t^2),\end{aligned}$$

so hypothesis (H2.1.3) is reasonable for i.i.d. sequences.

As mentioned in [13], the result simplifies considerably in the special case when the pair $(X^{(n)}, Y^{(n)})$ does not depend on n, that is to say when we consider a single sequence instead of a triangular array. This is a consequence of the following more general corollary.

Corollary 2.3. *Assume that* $\left(X^{(n)}, Y^{(n)}\right) \overset{(d)}{\to} (X, Y)$ *as* $n \to \infty$ *and that, for every fixed* $r > 0$,

$$\limsup_{n\to+\infty} \mathbb{E}\left[|X^{(n)}|^r\right] < \infty \quad \textit{and} \quad \limsup_{n\to+\infty} \mathbb{E}\left[|Y^{(n)}|^r\right] < \infty.$$

Suppose further that the distribution of X *has span 1 and that* Y *is not a.s. equal to an affine function* $c+dX$ *of* X*, that* k_n *and* N_n *are integers such that* $\mathbb{E}\left[X^{(n)}\right] = k_n/N_n$ *and* $N_n \to +\infty$*. Then, all hypotheses of Theorem 2.1 are satisfied and Theorem 2.1 holds.*

2.3 Applications

In this section we give several examples borrowed from [13] and [11]. A direct application of Corollary 2.3 leads to Berry-Esseen bounds in each of them.

2.3.1 Occupancy problem

In the classical occupancy problem (see [13] and the references therein for more details), m balls are distributed at random into N urns. The resulting numbers of balls $(Z_1, \ldots, Z_N)$ have a multinomial distribution which equals that of $(X_1, \cdots, X_N)$ conditioned on $\sum_{i=1}^N X_i = m$, where X_1, ..., X_N are i.i.d. with $X_i \sim \mathcal{P}(\lambda)$, for any arbitrary $\lambda > 0$. The classical occupancy problem studies the number W of empty urns that is the distribution of $\sum_{i=1}^N 1_{\{X_i=0\}}$ conditioned on $\sum_{i=1}^N X_i = m$.
Let us follow the work of Janson [13] and suppose that $m = k_n \to \infty$ and $N = N_n \to \infty$ with $k_n/N_n \to \lambda$. Then W can be taken as U_n in Theorem 2.1 with $X^{(n)} \sim \mathcal{P}(\lambda_n)$ and $Y^{(n)} = 1_{\{X^{(n)}=0\}}$ for any λ_n; we choose $\lambda_n = k_n/N_n$ so that assumption (H2.1.4) holds.

- If k_n, $N_n \to \infty$ such that $k_n/N_n \to \lambda \in (0, \infty)$, then Corollary 2.3 immediately yields that the conclusions of Theorem 2.1 hold.
- In the case $k_n/N_n \to \infty$, assumption (H2.1.1) is clearly violated and Theorem 2.1 does not apply.
- In the case $k_n/N_n \to 0$, Theorem 2.1 can not be applied as stated since $Y^{(n)} = 1_{\{X^{(n)}=0\}}$ implies that assumption (H2.1.7) does not hold ($r_n \to -1$). As explained in [13], one can choose instead $Y^{(n)} := 1_{\{X^{(n)}=0\}} + X^{(n)} - 1 = (X^{(n)} - 1)_+$ and it is clearly verified that Theorem 2.1 applies without any extra assumption.

2.3.2 Branching processes

Consider a Galton-Watson process, beginning with one individual, where the number of children of an individual is given by a random variable X having finite moments. Assume further that $\mathbb{E}(X) = 1$. We number the individuals as they appear. Let X_i be the number of children of the i^{th} individual. It is well known (see [13, Example 3.4] and the references therein) that the total progeny is $n \geq 1$ if and only if

$$S_k := \sum_{i=1}^{k} X_i \geq k \text{ for } 0 \leq k < n \text{ but } S_n = n - 1\,. \tag{5}$$

This type of conditioning is different from the one studied in the present paper, but Janson proves [13, Example 3.4] that if we ignore the order of $X_1, \ldots, X_n$, they have the same distribution conditioned on (5) as conditioned on $S_n = n - 1$. Hence our results apply to variables of the kind $Y_i = f(X_i)$. For example if $Y_i = 1_{\{X_i=3\}}$, the $\sum_{i=1}^n Y_i$ is the number of families with three children.

2.3.3 Random forests

Consider a uniformly distributed random labeled rooted forest with m vertices and $N < m$ roots. Without loss of generality, we may assume that the vertices are $1, \dots, m$ and, by symmetry, that the roots are the first N vertices. Following [13], this model can be realized as follows: the sizes of the N trees in the forest are distributed as $X_1, \dots, X_N$ conditioned on $\sum_{i=1}^N X_i = m$, where X_i are i.i.d. with the Borel distribution for some arbitrary parameter $\lambda \in]0, 1/e]$ (see section 3.3 for more details on Borel distribution and references therein). Further tree number i is drawn uniformly among the trees of size X_i.
A classical quantity of interest is the number of trees of size K in the forest (see, e.g., [17, 23, 24]). It means that we choose $Y_i = 1_{\{X_i = K\}}$. Let us now assume that we condition on $\sum_{i=1}^N X_i = m$ with $m = k_n \to +\infty$, $N = N_n \to +\infty$. The framework is similar to the one of Subsection 2.3.1 and we proceed analogously. Assume $k_n/N_n \to \lambda$ and take $X_i^{(n)}$ having Borel distribution with parameter $\lambda_n = k_n/N_n$.

2.3.4 Bose-Einstein statistics

This example is borrowed from [11]. Consider N urns. Put n indistinguishable balls in the urns in such a way that each distinguishable outcome has the same probability

$$1/\binom{n+N-1}{n},$$

see for example [6]. Let Z_k be the number of balls in the k^{th} urn. It is well known that $(Z_1, \dots, Z_N)$ is distributed as $\left(X_1, \cdots, X_N\right)$ conditioned on $\sum_{i=1}^N X_i = n$, where $X_1, \cdots, X_N$ are i.i.d. and geometrically distributed.

2.3.5 Hashing with linear probing

Hashing with linear probing can be regarded as throwing n balls sequentially into m urns at random; the urns are arranged in a circl and labeled. A ball that lands in an occupied urn is moved to the next empty urn, always moving in a fixed direction. The length of the move is called the displacement of the ball, and we are interested in the sum $d_{m,n}$ of all displacements. We assume $n < m$ and denote $N = m - n$.
Janson [12] proved that the length of the blocks (counting the empty urn) and the sum of displacements inside each block are distributed as $(X_1, Y_1), \dots, (X_N, Y_N)$ conditioned on $\sum_{i=1}^N X_i = m$, where (X_i, Y_i) are i.i.d. copies of a pair (X, Y) of random variables, X having the Borel distribution with any parameter $\lambda \in \left]0, e^{-1}\right]$ (see section 3.3 for more details on Borel distribution and references therein), and Y given $X = l$ is distributed as $d_{l,l-1}$. As in 2.3.1, we assume that $m = k_n \to \infty$ and $N = N_n \to \infty$ with $k_n/N_n \to a \in [1, +\infty[$. So, $\lambda_n := (n_n/m_n)\exp(-n_n/m_n) \in \left[0, e^{-1}\right[$ and $\lambda_n \to (1-1/a)\exp(-1+1/a) =: \lambda$. If $X^{(n)}$ has Borel distribution with parameter λ_n, Corollary 2.3 yields the desired Berry-Esseen bound.

2.4 Conditional large deviation result

In [9], the authors proved a classical large deviation principle for the conditional distribution $\mathcal{L}_n$ which applies to examples 2.3.1 to 2.3.4. Their result [9, Theorem 2.1] is the analogue of the central limit theorem of Janson [13]. The proof relies on Gärtner-Ellis theorem which requires the existence of the Laplace transform in a neighborhood of the origin. In the context of hashing, however, the joint Laplace transform is only defined on $(-\infty, a) \times (-\infty, 0)$ for some $a > 0$ and [9, Theorem 2.1] cannot be applied. Consequently one needs a specific result in the case when the Laplace transform is not defined.

Theorem 2.4. *Suppose that:*

(H2.4.1) $\log(\sigma_{X^{(n)}}) = o({N_n}^{1/2})$ *where* $\sigma_{X^{(n)}} := \mathrm{Var}\left(X^{(n)}\right)^{1/2}$;

(H2.4.2) $\rho_{X^{(n)}} := \mathbb{E}\left[\left|X^{(n)} - \mathbb{E}\left[X^{(n)}\right]\right|^3\right] = o\left(N_n^{1/2}\sigma^3_{X^{(n)}}\right)$;

(H2.4.3) *there exists* $c > 0$ *such that, for all* $n \geqslant 1$ *and* $s \in [-\pi, \pi]$,

$$\left|\mathbb{E}\left[e^{isX^{(n)}}\right]\right| \leqslant 1 - c\sigma^2_{X^{(n)}} s^2;$$

(H2.4.4) $k_n = N_n \mathbb{E}\left[X^{(n)}\right] + O(\sigma_{X^{(n)}} N_n^{1/2})$;

(H2.4.5) $\mathrm{Var}\left(Y^{(n)}\right) = o\left(N_n^{1/2}\right)$.

(H2.4.6) *the right tail of* $Y^{(n)}$ *satisfies: there exist* $\alpha > 0$ *and* $\beta > 0$ *such that, for all* $y > 0$,

$$\liminf_{n\to\infty} \frac{1}{\sqrt{N_n y}} \log \mathbb{P}(Y^{(n)} \geqslant N_n y) \geqslant -\beta \tag{6}$$

and

$$\limsup_{n\to\infty} \sup_{u \geqslant \sqrt{N_n y}} \frac{1}{\sqrt{u}} \log \mathbb{P}(Y^{(n)} \geqslant u) \leqslant -\alpha. \tag{7}$$

Then, for all $y > 0$,

$$\begin{aligned} -\beta\sqrt{y} &\leqslant \liminf_{n\to\infty} \frac{1}{\sqrt{N_n}} \log \mathbb{P}(T_n - \mathbb{E}\left[T_n | S_n = k_n\right] \geqslant N_n y | S_n = k_n) \\ &\leqslant \limsup_{n\to\infty} \frac{1}{\sqrt{N_n}} \log \mathbb{P}(T_n - \mathbb{E}\left[T_n | S_n = k_n\right] \geqslant N_n y | S_n = k_n) \leqslant -\alpha\sqrt{y}. \end{aligned}$$

Remark 2.5.

1. Notice the different nature of the assumptions on the standard deviations $\sigma_{X^{(n)}}$ and $\sigma_{Y^{(n)}}$.
2. The small shift allowed in assumption (H2.4.4) is the same as the one in assumption (H2.1.4) of Theorem 2.1. When the joint Laplace transform is defined in a neighborhood of the origin, one can use exponential changes of probability: a first one is based on the Laplace transform of $X^{(n)}$ and leads to reduce the conditioning to the mean $N_n \mathbb{E}\left[X^{(n)}\right]$ of S_n whereas the second relies on the Laplace transform of $Y^{(n)}$ and removes the conditioning leading to the study of a pair of random variables (see [9]). The large deviation principle is then proved for a larger range of shifts in the conditioning.

The result deeply relies on the following unconditioned one.

Theorem 2.6. *For all* $n \geqslant 1$, *let* z_n *be a positive number. Suppose that* $N_n \to +\infty$ *and that:*

(H2.6.1) $\liminf z_n/N_n > 0$;

(H2.6.2) $\mathrm{Var}(Y^{(n)}) = o\left(N_n^{1/2}\right)$;

(H2.6.3) *the right tail of* $Y^{(n)}$ *satisfies: there exist* $\alpha > 0$ *and* $\beta > 0$ *such that*

$$\liminf_{n\to\infty} \frac{1}{\sqrt{z_n}} \log \mathbb{P}(Y^{(n)} \geqslant z_n) \geqslant -\beta \tag{8}$$

and

$$\limsup_{n\to\infty} \sup_{u \geqslant \sqrt{z_n}} \frac{1}{\sqrt{u}} \log \mathbb{P}(Y^{(n)} \geqslant u) \leqslant -\alpha. \tag{9}$$

Then

$$\begin{aligned} -\beta &\leqslant \liminf_{n\to\infty} \frac{1}{\sqrt{z_n}} \log \mathbb{P}(T_n - N_n \mathbb{E}[Y^{(n)}] \geqslant z_n) \\ &\leqslant \limsup_{n\to\infty} \frac{1}{\sqrt{z_n}} \log \mathbb{P}(T_n - N_n \mathbb{E}[Y^{(n)}] \geqslant z_n) \leqslant -\alpha. \end{aligned}$$

Remark 2.7. Assumption (H2.6.1) naturally implies that z_n goes to infinity with n.

3 Application to hashing with linear probing

In this section we show that the example of hashing with linear probing briefly presented in section 2.3.5 satisfies the hypotheses of Theorem 2.4. We begin with a precise description of the model.

3.1 Complements on the model

Hashing with linear probing is a classical model in theoretical computer science which has been studied from a mathematical point of view by several authors [8, 12, 14, 1, 20]. For more details on the model, we refer to [8, 12, 14]. The model describes the following experiment. One throws n balls sequentially into m urns at random; the urns are arranged in a circle and numbered. A ball that lands in an occupied urn is moved to the next empty urn, always moving in a fixed direction. The length of the move is called the displacement of the ball and we are interested in the sum of all displacements which is a random variable noted $d_{m,n}$. We assume $n < m$ and define $N = m - n$.
In order to make things clear, let us give an example. Assume that $n = 8$, $m = 10$, and $(6, 9, 1, 9, 9, 6, 2, 5)$ are the addresses where the balls land. This sequence of addresses is called a *hash sequence* of length m and size n. Let d_i be the displacement of ball i, then $d_1 = d_2 = d_3 = 0$. The ball number 4 should land in the 9^{th} urn which is occupied by the second ball; thus it moves one step ahead and lands in urn 10 so that $d_4 = 1$. The 5^{th} ball should land in the 9^{th} urn. Since it is not possible (the urn being occupied by the second ball), it moves to the 10^{th} urn which is also occupied; it then moves to the first urn (also occupied) and finally to the second urn so that $d_5 = 3$. And so on: $d_6 = 1$, $d_7 = 1$, $d_8 = 0$. Here, the total displacement equals $1+3+1+1 = 6$. After throwing all balls, there are $N = m-n$ empty urns. These divide the occupied urns into blocks of consecutive urns. For convenience, we consider the empty urn following a block as belonging to this block. In our example, there are two blocks: the first one containing urns $9, 10, 1, 2, 3$ (occupied), and urn 4 empty, and the second one containing urns $5, 6, 7$ (occupied), and urn 8 empty.
Janson [12] proved that the lengths of the blocks (counting the last empty urn) and the sum of displacements inside each block are distributed as $(X_1, Y_1), \dots, (X_N, Y_N)$ conditioned on $\sum_{i=1}^{N} X_i = m$, where (X_i, Y_i) are i.i.d. copies of a pair (X, Y) of random variables, X having the Borel distribution with any parameter $\lambda \in \left]0, e^{-1}\right]$ (see section 3.3 for more details on Borel distribution and references therein) and the conditional distribution of Y given $X = l$ being the same as the distribution of $d_{l,l-1}$. So, $d_{m,n}$ is distributed as $\sum_{i=1}^{N} Y_i$ conditioned on $\sum_{i=1}^{N} X_i = m$. The following lemma presents already known results on the total displacement $d_{n+1,n}$ that will be useful in the proofs.

Lemma 3.1.

1. *The number of hash sequences of length $n+1$ and size n is $(n+1)^n$.*

2. *One clearly has $0 \leqslant d_{n+1,n} \leqslant \frac{n(n-1)}{2}$.*

3. *For any $y \geqslant 0$, the function defined from $\mathbb{N}$ to $[0,1]$ by $n \mapsto \mathbb{P}(d_{n+1,n} \geqslant y)$ is an increasing function of n.*

4. *The total displacement of any hash sequence $(h_1, \dots, h_n)$ is invariant with respect to any permutation of the $h_i's$. More precisely for any permutation σ of $\{1, \dots, n\}$, the total displacement associated to the hash sequence $(h_1, \dots, h_n)$ is the same as the total displacement associated to the hash sequence $(h_{\sigma(1)}, \dots, h_{\sigma(n)})$.*

Proof of Lemma 3.1. The first three points are obvious. Let us prove the last one. It is a consequence of [12, Lemma 2.1]. For any hash sequence $(h_1, \dots, h_n)$ and for any $i = 0, \dots, n+1$, let us define

$$Z_i := \text{Card}\{k \in [\![1, n]\!], h_k = i\}$$

and $\Sigma_i := \sum_{k=1}^{i} Z_j$ (notice that $Z_0 = 0$ and $\Sigma_0 = 0$). It is obvious that the sequence $(\Sigma_i)_{i=0,\dots,n+1}$ does not depend on the order of the hash sequence $(h_1, \dots, h_n)$. Now, formula (2.1) in [12, p. 442] establishes that

$$d_{n+1,n} = \sum_{i=1}^{n+1} H_i - n$$

where H_i, the number of items that make attempt to be inserted in cell i, is related to the sequence $(\Sigma_i)_{i=0,\dots,n+1}$ with the following formula (see [12, Lemma 2.1]):

$$H_i = \Sigma_i - i - \min_{k<i}(\Sigma_k - k) + 1.$$

Hence $d_{n+1,n}$ does not depend on the order of the hash sequence $(h_1, \dots, h_n)$. □

Using the results in [8, 13, 12], we can prove that the joint Laplace transform of (X, Y) is only defined on $(-\infty, a) \times (-\infty, 0)$ for some positive a. Hence, Theorem 2.1 of [9] can not be applied here.

3.2 Large deviations for hashing with linear probing

In order to provide large deviation bounds for $d_{m,n}$, we need to describe the asymptotic behavior of $\mathbb{P}(Y \geqslant y)$, which is given in the following proposition.

Proposition 3.2. *Let λ be the parameter of the Borel distribution of X be such that $\kappa := -\log(\lambda) - 1 \leqslant \log(2)$. Then,*

$$-\beta \leqslant \liminf_{y\to+\infty} \frac{1}{\sqrt{y}} \log \mathbb{P}(Y \geqslant y) \leqslant \limsup_{y\to+\infty} \frac{1}{\sqrt{y}} \log \mathbb{P}(Y \geqslant y) \leqslant -\alpha, \tag{10}$$

with

$$\alpha := \kappa\sqrt{2} \qquad \text{and} \qquad \beta := 2\kappa\sqrt{\left(1 + \frac{1}{\kappa}\right)\left(1 + \frac{1 + \log 2}{\kappa}\right)}.$$

Now, for all $n \geqslant 1$, let m_n and n_n be integers such that $n_n < m_n$, and $N_n := m_n - n_n$. Suppose that $m_n/N_n \to a \in [1, +\infty[$. We introduce $\lambda_n := (n_n/m_n)\exp(-n_n/m_n) \in \left[0, e^{-1}\right[$. Hence $\lambda_n \to (1 - 1/a)\exp(-1 + 1/a) =: \lambda$. To apply Proposition 3.2, suppose that $\lambda \geqslant (2e)^{-1}$. Let $(X_i^{(n)}, Y_i^{(n)})_{i=1,2,\dots,N_n}$ be i.i.d. copies of $(X^{(n)}, Y^{(n)})$, $X^{(n)}$ following Borel distribution with parameter λ_n (so that $\mathbb{E}[X^{(n)}] = m_n/N_n$), and $Y^{(n)}$ given $X^{(n)} = l$ being distributed as $d_{l,l-1}$. Let

$$S_n := \sum_{i=1}^{N_n} X_i^{(n)} \quad \text{and} \quad T_n := \sum_{i=1}^{N_n} Y_i^{(n)}.$$

The total displacement d_{m_n,n_n} is distributed as the conditional distribution of T_n given $S_n = m_n$. Since assumptions (H2.4.1) to (H2.4.5) are also satisfied by $\left(X_i^{(n)}, Y_i^{(n)}\right)$ $(i = 1, 2, \dots, N_n)$, we can apply Theorem 2.4.

Proposition 3.3 (Large deviations for hashing with linear probing)**.** *For α and β defined in Proposition 3.2 and $k_n = m_n$, assumptions (H2.4.1) to (H2.4.6) are satisfied. Then, for all $y > 0$,*

$$\begin{aligned} -\beta\sqrt{y} &\leqslant \liminf_{n\to\infty} \frac{1}{\sqrt{N_n}} \log \mathbb{P}(d_{m_n,n_n} - \mathbb{E}[d_{m_n,n_n}] \geqslant N_n y) \\ &\leqslant \limsup_{n\to\infty} \frac{1}{\sqrt{N_n}} \log \mathbb{P}(d_{m_n,n_n} - \mathbb{E}[d_{m_n,n_n}] \geqslant N_n y) \leqslant -\alpha\sqrt{y}. \end{aligned}$$

3.3 Proof of Proposition 3.2

We start computing the asymptotic tail behavior of X. Remind that X has Borel distribution with parameter $\lambda \in \left]0, e^{-1}\right]$ which means that

$$\mathbb{P}(X = n) = \frac{1}{T(\lambda)} \frac{\lambda^n n^{n-1}}{n!},$$

where T is the well-known tree function (see, e.g., [8] or [12] for more details). We define $\kappa \in]0, +\infty[$ by $\kappa := -\log(\lambda) - 1$.

Lemma 3.4.

(i) The asymptotic behavior of X is given by

$$\log \mathbb{P}(X = n) = -\kappa n(1 + o(1)). \tag{11}$$

(ii) The asymptotic tail behavior of X is given by

$$\log \mathbb{P}(X \geqslant n) = -\kappa n(1 + o(1)). \tag{12}$$

Proof. (i) By Stirling formula,

$$\log \mathbb{P}(X = n) = \log \left(\frac{1}{\sqrt{2\pi} T(\lambda)} \frac{(\lambda e)^n}{n^{3/2}} \right) (1 + o(1)) = -\kappa n(1 + o(1)).$$

(ii) Similarly, using Stirling formula,

$$\begin{aligned} \mathbb{P}(X \geqslant n) &= \sum_{k \geqslant n} \mathbb{P}(X = k) = \frac{1}{\sqrt{2\pi} T(\lambda)} \sum_{k \geqslant n} e^{-\kappa k(1+o(k))} k^{-3/2} \\ &= \frac{1}{\sqrt{2\pi} T(\lambda)} \sum_{k \geqslant n} e^{-\kappa k(1+o(k))}. \end{aligned}$$

Let $\varepsilon > 0$. Then there exists $n_0 \in \mathbb{N}$ such that, for any $k \geqslant n_0$, $|o(k)| \leqslant \varepsilon$. Thus, for any $n \geqslant n_0$,

$$\sum_{k \geqslant n} e^{-\kappa k(1+\varepsilon)} \leqslant \sqrt{2\pi} T(\lambda) \mathbb{P}(X \geqslant n) \leqslant \sum_{k \geqslant n} e^{-\kappa k(1-\varepsilon)}.$$

Using the fact that $\lambda e < 1$, we get

$$\begin{aligned} \log \left(\frac{1}{\sqrt{2\pi} T(\lambda)} \sum_{k \geqslant n} e^{-\kappa k(1 \pm \varepsilon)} \right) &= \log \left(\frac{e^{-\kappa n}}{\sqrt{2\pi} T(\lambda)} \frac{e^{\pm \kappa n \varepsilon}}{1 - e^{-\kappa(1 \pm \varepsilon)}} \right) \\ &= -\kappa n(1 \pm \varepsilon)(1 + o(1)), \end{aligned}$$

which leads to the required result when ε goes to 0. □

Proof of the upper bound in (10). Let $y > 0$ and n_y be the ceiling of the positive solution of $2y = n(n-1)$:

$$n_y = \left\lceil \sqrt{2y + \frac{1}{4}} + \frac{1}{2} \right\rceil. \tag{13}$$

Since Y conditionally to $X = n + 1$ is distributed as $d_{n+1,n}$, we get

$$\mathbb{P}(Y \geqslant y) = \sum_{n=n_y}^{+\infty} \mathbb{P}(d_{n+1,n} \geqslant y) \mathbb{P}(X = n + 1) \leqslant \sum_{n=n_y}^{+\infty} \mathbb{P}(X = n + 1) = \mathbb{P}(X \geqslant n_y).$$

By (12) and the fact that $n_y = \sqrt{2y}(1+o(1))$, we finally conclude that

$$\limsup_{y\to+\infty} \log \mathbb{P}(Y \geqslant y) \leqslant -\kappa\sqrt{2y}.$$

□

Proof of the lower bound in (10). Let $y > 0$. For any $m_y \in \mathbb{N}^*$ such that $m_y \geqslant n_y$, one has

$$\begin{aligned}\mathbb{P}(Y \geqslant y) &= \sum_{n=n_y}^{+\infty} \mathbb{P}(d_{n+1,n} \geqslant y)\mathbb{P}(X = n+1) \\ &\geqslant \mathbb{P}\left(d_{m_y+1,m_y} \geqslant y\right)\mathbb{P}(X = m_y+1)\end{aligned}$$

So, we are interested in the hash sequences of length m_y+1 and size m_y that realize a total displacement greater than y. More precisely, we want to evaluate the probability $\mathbb{P}\left(d_{m_y+1,m_y} \geqslant y\right)$ or at least to bound it from below. In that view, for any $0 \leq k \leqslant \frac{m_y}{2}$ consider the following hash sequence:

$$(1,\, 1,\, 2,\, 2,\, \ldots\, k,\, k,\, k+1,\, k+2,\, \ldots,\, m_y - k)\,. \tag{14}$$

On the one hand, it is decomposed into $m_y - 2k$ single numbers and k pairs leading to a hash sequence of size m_y as required. On the other hand, each pair $(q,\, q)$ $(q = 1 \ldots k)$ realizes a displacement equal to $(q-1)+q$ while each singleton q $(q = k+1 \ldots m_y - k)$ realizes a displacement equal to k. The total displacement is then $k(m_y - k)$. It remains to choose m_y and $0 \leqslant k \leqslant \frac{m_y}{2}$ such that $k(m_y-k) \geqslant y$ in order to obtain the best possible lower bound.
Moreover as mentioned in Lemma 3.1 the total displacement associated to any hash sequence does not depend on the order of the hash sequence. One can consider all the permutations of the hash sequence defined in (14) whose total number is given by

$$\binom{m_y}{1}\binom{m_y-1}{1}\cdots\binom{2k+1}{1}\binom{2k}{2}\binom{2k-2}{2}\cdots\binom{2}{2} = \frac{m_y!}{2^k}.$$

As a consequence, $\mathbb{P}(Y \geqslant y)$ is bounded from below by $\frac{1}{(m_y+1)^{m_y}}\frac{m_y!}{2^k}\mathbb{P}(X = m_y+1)$. By Stirling formula, $n! \underset{n}{\sim} \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$ and the asymptotic behavior of X given in (11),

$$\log\left(\frac{1}{(m_y+1)^{m_y}}\frac{m_y!}{2^k}\,\mathbb{P}(X = m_y+1)\right) \underset{y}{\sim} -(\kappa+1)m_y - k\log 2. \tag{15}$$

Now the inequality $k(m_y-k) \geqslant y$ admits solutions as soon as $m_y \geqslant 2\sqrt{y}$. Hence we take $m_y = 2t\sqrt{y}$ for some $t \geqslant 1$. Simple computation shows that the best possible choices for k and t are $k = \frac{m_y - \sqrt{m_y^2 - 4y}}{2}$ and $t = \left(1 + 2\frac{\kappa+1}{\log 2}\right)\left(\left(1+2\frac{\kappa+1}{\log 2}\right)^2 - 1\right)^{-1/2}$. Plugging the values of m_y and k into (15) leads to the value

$$-2\kappa\sqrt{\left(1+\frac{1}{\kappa}\right)\left(1+\frac{1+\log 2}{\kappa}\right)}\sqrt{y};$$

which completes the proof of the minoration. □

4 Proofs

4.1 Notations and technical results

The proofs of Theorems 2.1 and 2.4 intensively rely on the use of Fourier transforms. Define φ_n and ψ_n by

$$\varphi_n(s,t) := \mathbb{E}\left[\exp\left\{is\left(X^{(n)} - \mathbb{E}\left[X^{(n)}\right]\right) + it\left(Y^{(n)} - \mathbb{E}\left[Y^{(n)}\right]\right)\right\}\right] \tag{16}$$

$$\text{and} \quad \psi_n(t) := 2\pi\mathbb{P}(S_n = k_n)\mathbb{E}\left[\exp\left\{it\left(U_n - N_n\mathbb{E}\left[Y^{(n)}\right]\right)\right\}\right]. \tag{17}$$

In this first section, we establish some properties of those two functions. First notice that we have $\varphi_n(s,0) = e^{-is\mathbb{E}[X^{(n)}]}\mathbb{E}\left[e^{isX^{(n)}}\right]$ and $\psi_n(0) = 2\pi\mathbb{P}(S_n = k_n)$.

Lemma 4.1. *One has*

$$\psi_n(t) = \frac{1}{\sigma_{X^{(n)}}N_n^{1/2}}\int_{-\pi\sigma_{X^{(n)}}N_n^{1/2}}^{\pi\sigma_{X^{(n)}}N_n^{1/2}} e^{-is\sigma_{X^{(n)}}^{-1}N_n^{-1/2}\left(k_n - N_n\mathbb{E}[X^{(n)}]\right)}\varphi_n^{N_n}\left(\frac{s}{\sigma_{X^{(n)}}N_n^{1/2}}, t\right)ds. \tag{18}$$

Proof. Since

$$\int_{-\pi}^{\pi} e^{is(S_n - k_n)}ds = 2\pi 1_{\{S_n = k_n\}},$$

we have

$$\begin{aligned}
\psi_n(t) &= 2\pi\mathbb{P}(S_n = k_n)\mathbb{E}\left[\exp\left\{it\left(U_n - N_n\mathbb{E}\left[Y^{(n)}\right]\right)\right\}\right] \\
&= 2\pi\mathbb{E}\left[\exp\left\{it\left(T_n - N_n\mathbb{E}\left[Y^{(n)}\right]\right)\right\}1_{S_n = k_n}\right] \\
&= \int_{-\pi}^{\pi}\mathbb{E}\left[\exp\left\{is\left(S_n - k_n\right) + it\left(T_n - N_n\mathbb{E}\left[Y^{(n)}\right]\right)\right\}\right]ds \\
&= \int_{-\pi}^{\pi} e^{-is\left(k_n - N_n\mathbb{E}[X^{(n)}]\right)}\varphi_n^{N_n}(s,t)ds,
\end{aligned}$$

which leads to the result after the change of variable $s' = s\sigma_{X^{(n)}}N_n^{1/2}$. □

Lemma 4.2.

(i) Under assumption (H2.1.3), for any integer $l \geqslant 0$, and for $|s| \leqslant \pi\sigma_{X^{(n)}}N_n^{1/2}$, $|t| \leqslant \eta_0\sigma_{Y^{(n)}}N_n^{1/2}$,

$$\left|\varphi_n^{N_n - l}\left(\frac{s}{\sigma_{X^{(n)}}N_n^{1/2}}, \frac{t}{\sigma_{Y^{(n)}}N_n^{1/2}}\right)\right| \leqslant e^{-(s^2+t^2)\cdot c_5(N_n - l)/N_n}. \tag{19}$$

(ii) Under assumption (H2.4.3), for any integer $l \geqslant 0$, and for $|s| \leqslant \pi\sigma_{X^{(n)}}N_n^{1/2}$,

$$\left|\varphi_n^{N_n - l}\left(\frac{s}{\sigma_{X^{(n)}}N_n^{1/2}}, 0\right)\right| \leqslant e^{-s^2\cdot c(N_n - l)/N_n}. \tag{20}$$

Proof. The proof is a mere consequence of the inequality $1 + x \leqslant e^x$. □

In the sequel, we also need different controls on the first derivative of φ_n with respect to the first variable.

Lemma 4.3. *For any s and t, one has:*

(i)

$$\left|\frac{\partial\varphi_n}{\partial t}\left(\frac{s}{\sigma_{X^{(n)}}N_n^{1/2}},\frac{t}{\sigma_{Y^{(n)}}N_n^{1/2}}\right)\right| \leqslant \frac{\sigma_{Y^{(n)}}}{N_n^{1/2}}(|s|+|t|); \tag{21}$$

(ii)

$$\begin{aligned}&\left|\frac{\partial\varphi_n}{\partial t}\left(\frac{s}{\sigma_{X^{(n)}}N_n^{1/2}},\quad\frac{t}{\sigma_{Y^{(n)}}N_n^{1/2}}\right)\right| &(22)\\ &\leqslant \frac{\sigma_{Y^{(n)}}}{N_n^{1/2}}(|s|r_n+|t|)+\frac{\sigma_{Y^{(n)}}}{N_n}\left[\frac{s^2}{2}\left(\frac{\rho_{X^{(n)}}}{\sigma^3_{X^{(n)}}}\right)^{2/3}\left(\frac{\rho_{Y^{(n)}}}{\sigma^3_{Y^{(n)}}}\right)^{1/3}\right.\\ &\qquad\left.+|st|\left(\frac{\rho_{X^{(n)}}}{\sigma^3_{X^{(n)}}}\right)^{1/3}\left(\frac{\rho_{Y^{(n)}}}{\sigma^3_{Y^{(n)}}}\right)^{2/3}+\frac{t^2}{2}\left(\frac{\rho_{Y^{(n)}}}{\sigma^3_{Y^{(n)}}}\right)\right]. &(23)\end{aligned}$$

Proof. We apply Taylor Theorem to the function defined by

$$(s,t)\mapsto f(s,t)=\frac{\partial\varphi_n}{\partial t}\left(\frac{s}{\sigma_{X^{(n)}}N_n^{1/2}},\frac{t}{\sigma_{Y^{(n)}}N_n^{1/2}}\right).$$

We conclude to (i) using

$$|f(s,t)-f(0,0)|\leqslant|s|\sup_{\theta,\theta'\in[0,1]}\left|\frac{\partial f}{\partial s}(\theta s,\theta' t)\right|+|t|\sup_{\theta,\theta'\in[0,1]}\left|\frac{\partial f}{\partial t}(\theta s,\theta' t)\right|$$

and to (ii) using

$$\begin{aligned}|f(s,t)-f(0,0)|\leqslant|s|\left|\frac{\partial f}{\partial s}(0,0)\right|+|t|\left|\frac{\partial f}{\partial t}(0,0)\right|+\frac{s^2}{2}\sup_{\theta,\theta'\in[0,1]}\left|\frac{\partial^2 f}{\partial^2 s}(\theta s,\theta' t)\right|\\ +|st|\sup_{\theta,\theta'\in[0,1]}\left|\frac{\partial^2 f}{\partial t\partial s}(\theta s,\theta' t)\right|+\frac{t^2}{2}\sup_{\theta,\theta'\in[0,1]}\left|\frac{\partial^2 f}{\partial^2 t}(\theta s,\theta' t)\right|\end{aligned}$$

□

Proposition 4.4.

1. *Under assumption* (H2.1.2)*, one has* $\sigma_{X^{(n)}}\geqslant(4c_2^3)^{-1}$.
2. *Under assumption* (H2.4.2)*, one has* $\sigma_{X^{(n)}}N_n^{1/2}\to+\infty$.

Proof. The proofs of both results rely on the fact that, for any integer-valued random variable X (see [13, Lemma 4.1.]),

$$\sigma_X^2\leqslant 4\mathbb{E}\left[|X-\mathbb{E}[X]|^3\right].$$

The conclusion follows, using hypothesis (H2.1.2) (resp. (H2.4.2)). □

Proposition 4.5. *We assume hypotheses* (H2.1.2)*,* (H2.1.3)*, and* (H2.1.4) *(or* (H2.4.2)*,* (H2.4.3) *and* (H2.4.4)*). Then there exists* $m>0$ *such that*

$$\mathbb{P}(S_n=k_n)\geqslant\frac{m}{2\pi\sigma_{X^{(n)}}N_n^{1/2}}.$$

Proof. Only consider the indices n for which $\sigma_{X^{(n)}} < +\infty$. Remember that $\varphi_n(s,0) = \mathbb{E}\left[e^{is(X^{(n)} - \mathbb{E}[X^{(n)}])}\right]$ and

$$\psi_n(0) = 2\pi\mathbb{P}(S_n = k_n) = \frac{1}{\sigma_{X^{(n)}} N_n^{1/2}} \int_{-\pi\sigma_{X^{(n)}} N_n^{1/2}}^{\pi\sigma_{X^{(n)}} N_n^{1/2}} e^{-isv_n} \varphi_n^{N_n}\left(\frac{s}{\sigma_{X^{(n)}} N_n^{1/2}}, 0\right) ds$$

where $v_n = \frac{k_n - N_n\mathbb{E}[X^{(n)}]}{\sigma_{X^{(n)}} N_n^{1/2}}$, by lemma 4.1. Let us prove that the sequence

$$(u_n)_n = \left(\psi_n(0)\sigma_{X^{(n)}} N_n^{1/2} e^{v_n^2/2}\right)$$

converges to $\sqrt{2\pi}$, from which the conclusion follows, since $(v_n)_n$ is bounded by (H2.1.4) (or (H2.4.4)) and $\mathbb{P}(S_n = k_n) > 0$ for all n. Inequality (19) with $l = 0$ and $t = 0$ (or (20) with $l = 0$) implies that the sequence $(u_n)_n$ is bounded. Let us prove that $\sqrt{2\pi}$ is the only accumulation point of $(u_n)_n$. Let $\phi(n)$ such that $(u_{\phi(n)})_n$ converges. Even if it means extracting more, we can suppose that $(v_{\phi(n)})_n$ converges. Let $v = \lim v_{\phi(n)}$. Using Taylor Theorem, there exists $t \in \mathbb{R}$ such that

$$\left|\varphi_n\left(\frac{s}{\sigma_{X^{(n)}} N_n^{1/2}}, 0\right) - 1 + \frac{s^2}{2N_n}\right| \leqslant \frac{|s|^3}{6\sigma_{X^{(n)}}^3 N_n^{3/2}} \mathbb{E}\left[\left|X^{(n)} - \mathbb{E}\left[X^{(n)}\right]\right|^3\right] = o\left(\frac{1}{N_n}\right)$$

where the last equality follows from hypothesis (H2.1.2) (or (H2.4.2)). Now,

$$e^{-isv_{\phi(n)}} \varphi_{\phi(n)}^{N_{\phi(n)}}\left(\frac{s}{\sigma_{X^{(\phi(n))}}\sqrt{N_{\phi(n)}}}, 0\right) \to e^{-isv - s^2/2} = e^{-v^2/2} e^{-(s+iv)^2/2}$$

and, by Lebesgue dominated convergence theorem and the fact that $\sigma_{X^{(n)}} N_n^{1/2} \to +\infty$ (see Proposition 4.4),

$$\psi_{\phi(n)}(0)\sigma_{X^{(\phi(n))}}\sqrt{N_{\phi(n)}} e^{v_{\phi(n)}^2/2} \to \sqrt{2\pi}.$$

□

4.2 Proof of Theorem 2.1

Part a) is Proposition 4.5 with $\tilde{c}_5 = m$. Now we follow the procedure of Janson [13] to uncorrelate $X^{(n)}$ and $Y^{(n)}$ and center the variable $Y^{(n)}$. We replace $Y^{(n)}$ by the projection

$$Y'^{(n)} := Y^{(n)} - \mathbb{E}[Y^{(n)}] - \frac{\mathrm{Cov}(X^{(n)}, Y^{(n)})}{\sigma_{X^{(n)}}^2}\left(X^{(n)} - \mathbb{E}[X^{(n)}]\right).$$

Then $\mathbb{E}[Y'^{(n)}] = 0$ and $\mathrm{Cov}(X^{(n)}, Y'^{(n)}) = \mathbb{E}[X^{(n)}Y'^{(n)}] = 0$. Besides, assumptions (H2.1.3) and (H2.1.7) are verified by $Y'^{(n)}$. By assumption (H2.1.7),

$$\sigma_{Y'^{(n)}}^2 = \sigma_{Y^{(n)}}^2(1 - r_n^2) \in [\tilde{c}_3^2(1 - c_6^2), c_3^2],$$

so (H2.1.5) is satisfied by $Y'^{(n)}$. Finally, by Minkowski Inequality, assumptions (H2.1.2) and (H2.1.6), and the fact that $|r_n| \leqslant 1$,

$$\begin{aligned}
\left\|Y'^{(n)}\right\|_3 &\leqslant \left\|Y^{(n)} - \mathbb{E}[Y^{(n)}]\right\|_3 + \frac{|r_n|\,\sigma_{X^{(n)}}\sigma_{Y^{(n)}}}{\sigma_{X^{(n)}}^2}\left\|X^{(n)} - \mathbb{E}[X^{(n)}]\right\|_3 \\
&\leqslant \rho_{Y^{(n)}}^{1/3} + r_n\sigma_{Y^{(n)}}\frac{\rho_{X^{(n)}}^{1/3}}{\sigma_{X^{(n)}}} \\
&\leqslant \sigma_{Y^{(n)}}(c_2 + c_4).
\end{aligned}$$

Hence $Y^{'(n)}$ satisfies assumption (H2.1.6). Consequently, all conditions hold for the pair $(X^{(n)}, Y^{'(n)})$ too. Finally,

$$T'_n := \sum_{i=1}^{N_n} Y_i^{'(n)} = T_n - N_n \mathbb{E}\left[Y^{(n)}\right] - \frac{\mathrm{Cov}(X^{(n)}, Y^{(n)})}{\sigma^2_{X^{(n)}}} \left(S_n - N_n \mathbb{E}\left[X^{(n)}\right]\right).$$

So, conditioned on $S_n = k_n$, we have $T'_n = T_n - N_n \mathbb{E}\left[Y^{(n)}\right] - r_n \frac{\sigma_{Y^{(n)}}}{\sigma_{X^{(n)}}}(k_n - N_n \mathbb{E}[X^{(n)}])$. Hence the conclusions for $\left(X^{(n)}, Y^{(n)}\right)$ and $\left(X^{(n)}, Y^{'(n)}\right)$ are the same. Thus, it suffices to prove the theorem for $\left(X^{(n)}, Y^{'(n)}\right)$; in other words, we may henceforth assume that $\mathbb{E}\left[Y^{(n)}\right] = \mathbb{E}\left[X^{(n)} Y^{(n)}\right] = 0$. Note that in that case $\tau_n^2 = \sigma^2_{Y^{(n)}}$.

Proof of Theorem 2.1 - Part b). We follow the classical proof of Berry-Esseen (see e.g. [7]) combined with the procedure of Quine and Robinson [25] to establish the result of Theorem 2.1.
As shown in Loève [19] (page 285) or Feller [7], the left hand side of (1) is dominated by

$$\frac{2}{\pi} \int_0^{\eta \sigma_{Y^{(n)}} N_n^{1/2}} \left| \frac{\psi_n(u/\sigma_{Y^{(n)}} N_n^{1/2})}{2\pi \mathbb{P}(S_n = k_n)} - e^{-u^2/2} \right| \frac{du}{u} + \frac{24 \sigma^{-1}_{Y^{(n)}} N_n^{-1/2}}{\eta \pi \sqrt{2\pi}} \tag{24}$$

where $\eta > 0$ will be specified later. From Lemma 4.1 and a Taylor expansion,

$$\begin{aligned}
u^{-1} &\left| \frac{\psi_n(u/\sigma_{Y^{(n)}} N_n^{1/2})}{2\pi \mathbb{P}(S_n = k_n)} - e^{-u^2/2} \right| = u^{-1} e^{-u^2/2} \left| \frac{e^{u^2/2} \psi_n(u/\sigma_{Y^{(n)}} N_n^{1/2})}{2\pi \mathbb{P}(S_n = k_n)} - 1 \right| \\
&\leqslant e^{-u^2/2} \sup_{0 \leqslant \theta \leqslant u} \left| \frac{\partial}{\partial t} \left[\frac{e^{t^2/2} \psi_n(t/\sigma_{Y^{(n)}} N_n^{1/2})}{2\pi \mathbb{P}(S_n = k_n)} \right] \right|_{t=\theta} \\
&\leqslant c_n^{-1} e^{-u^2/2} \sup_{0 \leqslant \theta \leqslant u} \left\{ \int_{-\pi \sigma_{X^{(n)}} N_n^{1/2}}^{\pi \sigma_{X^{(n)}} N_n^{1/2}} \left| \frac{\partial}{\partial t} \left[e^{t^2/2} \varphi_n^{N_n} \left(\frac{s}{\sigma_{X^{(n)}} N_n^{1/2}}, \frac{t}{\sigma_{Y^{(n)}} N_n^{1/2}} \right) \right] \right|_{t=\theta} ds \right\}
\end{aligned}$$

where $c_n := 2\pi \mathbb{P}(S_n = k_n) \sigma_{X^{(n)}} N_n^{1/2} \geqslant \tilde{c}_5$ and $v_n = \frac{k_n - N_n \mathbb{E}\left[X^{(n)}\right]}{\sigma_{X^{(n)}} N_n^{1/2}}$ has already been defined in the proof of Proposition 4.5. Now we split the integration domain of s into

$$A_1 := \left\{ s : \ |s| < \varepsilon \sigma_{X^{(n)}} N_n^{1/2} \right\} \quad \text{and} \quad A_2 := \left\{ s : \ \varepsilon \sigma_{X^{(n)}} N_n^{1/2} \leqslant |s| \leqslant \pi \sigma_{X^{(n)}} N_n^{1/2} \right\},$$

(where $0 < \varepsilon < \pi$ will be specified later) and decompose

$$u^{-1} \left| \frac{\psi_n(u/\sigma_{Y^{(n)}} N_n^{1/2})}{2\pi \mathbb{P}(S_n = k_n)} - e^{-u^2/2} \right| \leqslant \sup_{0 \leqslant \theta \leqslant u} \left[I_1(u, \theta) + I_2(u, \theta) \right], \tag{25}$$

where

$$I_1(u, \theta) = c_n^{-1} \int_{A_1} e^{-(u^2+s^2)/2} \left| \left(\frac{\partial}{\partial t} \left[e^{(t^2+s^2)/2} \varphi_n^{N_n} \left(\frac{s}{\sigma_{X^{(n)}} N_n^{1/2}}, \frac{t}{\sigma_{Y^{(n)}} N_n^{1/2}} \right) \right] \right)_{t=\theta} \right| ds, \tag{26}$$

$$I_2(u, \theta) = c_n^{-1} e^{-u^2/2} \int_{A_2} \left| \left(\frac{\partial}{\partial t} \left[e^{t^2/2} \varphi_n^{N_n} \left(\frac{s}{\sigma_{X^{(n)}} N_n^{1/2}}, \frac{t}{\sigma_{Y^{(n)}} N_n^{1/2}} \right) \right] \right)_{t=\theta} \right| ds. \tag{27}$$

To bound $I_1(u, \theta)$, we use a result due to Quine and Robinson ([25, Lemma 2]).

Lemma 4.6. *[Lemma 2 in [25]] Define*

$$l_{1,n} := \rho_{X^{(n)}} \sigma^{-3}_{X^{(n)}} N_n^{-1/2} \qquad \text{and} \qquad l_{2,n} := \rho_{Y^{(n)}} \sigma^{-3}_{Y^{(n)}} N_n^{-1/2}.$$

If $l_{1,n} \leqslant 1$ *and* $l_{2,n} \leqslant 1$*, then, for all*

$$(s,t) \in R := \left\{ (s,t) : \ |s| < \frac{2}{9} l_{1,n}^{-1}, |t| < \frac{2}{9} l_{2,n}^{-1} \right\},$$

we have

$$\begin{aligned} &\left| \frac{\partial}{\partial t} \left[e^{(s^2+t^2)/2} \ \varphi_n^{N_n} \left(\frac{s}{\sigma_{X^{(n)}} N_n^{1/2}}, \frac{t}{\sigma_{Y^{(n)}} N_n^{1/2}} \right) \right] \right| \\ &\qquad \leqslant C_0 (|s| + |t| + 1)^3 (l_{1,n} + l_{2,n}) \exp\left\{ \frac{11}{24} \left(s^2 + t^2 \right) \right\} \end{aligned} \tag{28}$$

with

$$C_0 := 98.$$

Proof. We refer to the proof in the appendix of [25]. The condition $l_{1,n} < 12^{-3/2}$ and $l_{2,n} < 12^{-3/2}$ appearing in [25, Lemma 2] can be replaced by $l_{1,n} \leqslant (33/32)^{3/2}$ and $l_{2,n} \leqslant (33/32)^{3/2}$ since the factor $8/27$ in (A4) of their proof can be replaced by a factor $1/27$. Since we do not provide the best constants here, we simply suppose $l_{1,n} \leqslant 1$ and $l_{2,n} \leqslant 1$. Finally, C_0 has to be greater than 4 and

$$\begin{aligned} &\sup_{(v,s)\in\mathbb{R}^2} \frac{27(|v| + 2\,|s|)(|v|^3 + |s|^3)}{(|v| + |s| + 1)^3} e^{-(v^2+s^2)/24} \\ &\qquad \leqslant 54 \cdot (|v| + |s|) e^{-(v^2+s^2)/24} \\ &\qquad \leqslant 108 \cdot \sqrt{6} \sqrt{\frac{v^2+s^2}{12}} e^{-(v^2+s^2)/24} \leqslant \frac{108 \cdot \sqrt{6}}{e} \leqslant 98. \end{aligned}$$

□

By assumptions (H2.1.2) and (H2.1.1),

$$l_{1,n} \leqslant c_2^3 N_n^{-1/2} \leqslant c_2^3 c_1 \sigma_{X^{(n)}}^{-1} N_n^{-1/2}, \tag{29}$$

which implies that $\sigma_{X^{(n)}} N_n^{1/2} \leqslant c_2^{-3} c_1^{-1} l_{1,n}^{-1}$. Similarly,

$$l_{2,n} \leqslant c_4^3 N_n^{-1/2} \leqslant c_4^3 c_3 \sigma_{Y^{(n)}}^{-1} N_n^{-1/2}, \tag{30}$$

and $\sigma_{Y^{(n)}} N_n^{1/2} \leqslant c_4^{-3} c_3^{-1} l_{2,n}^{-1}$. Assume henceforth that

$$\varepsilon := \min\left(\frac{2}{9} c_1 c_2^3, \pi \right) \quad \text{and} \quad \eta := \min\left(\frac{2}{9} c_3 c_4^3, \eta_0 \right). \tag{31}$$

Lemma 4.7. *There exists a positive constant* C_1 *such that*

$$\int_0^{\eta \sigma_{Y^{(n)}} N_n^{1/2}} \sup_{0 \leqslant \theta \leqslant u} I_1(u,\theta) du \leqslant \frac{C_1}{N_n^{1/2}}. \tag{32}$$

Proof. Conditions (31) imply that, on A_1,

$$\begin{aligned} |s| < \varepsilon \sigma_{X^{(n)}} N_n^{1/2} &\leqslant \frac{2}{9} l_{1,n}^{-1} \\ \text{and} \quad |\theta| \leqslant |u| \leqslant \eta \sigma_{Y^{(n)}} N_n^{1/2} &\leqslant \frac{2}{9} l_{2,n}^{-1}, \end{aligned}$$

which ensures that $(s,u) \in R$ as specified in Lemma 4.6. Moreover, since we have $N_n \geqslant \max(c_2^6, c_4^6)$ (cf. hypothesis in 2.1.b), $l_{1,n} \leqslant 1$ and $l_{2,n} \leqslant 1$. Now applying Lemma 4.6 in (26) and using part 2.1.a, we get

$$\begin{aligned}
&\int_0^{\eta\sigma_{Y^{(n)}}N_n^{1/2}} \sup_{0\leqslant\theta\leqslant u} I_1(u,\theta)du \\
&\qquad \leqslant c_n^{-1}C_0(l_{1,n}+l_{2,n})\int_0^{\eta\sigma_{Y^{(n)}}N_n^{1/2}}\int_{A_1}(|s|+|u|+1)^3 e^{-(s^2+u^2)/24}dsdu \\
&\qquad \leqslant N_n^{-1/2}\tilde{c}_5^{-1}C_0(c_2^3+c_4^3)\int_{\mathbb{R}^2}(|s|+|u|+1)^3 e^{-(s^2+u^2)/24}dsdu
\end{aligned}$$

and the result follows with

$$C_1 = \tilde{c}_5^{-1}C_0(c_2^3+c_4^3)\int_{\mathbb{R}^2}(|s|+|u|+1)^3 e^{-(s^2+u^2)/24}dsdu.$$

□

Now, we study the integral on A_2.

Lemma 4.8. *There exist positive constants C_2 and C_3, only depending on $\tilde{c}_1$, c_1, c_2, $\tilde{c}_3$, c_3, c_4, c_5, $\tilde{c}_5$, and c_6, such that*

$$\int_0^{\eta\sigma_{Y^{(n)}}N_n^{1/2}} \sup_{0\leqslant\theta\leqslant t} I_2(u,\theta)du \leqslant C_2 e^{-C_3 N_n}. \tag{33}$$

Proof. We use the controls (21), (19), and $|\varphi_n| \leqslant 1$ to get

$$\begin{aligned}
&\left|\left(\frac{\partial}{\partial t}\left[e^{t^2/2}\varphi_n^{N_n}\left(\frac{s}{\sigma_{X^{(n)}}N_n^{1/2}},\frac{t}{\sigma_{Y^{(n)}}N_n^{1/2}}\right)\right]\right)_{t=\theta}\right| \\
&= e^{\theta^2/2}\left|\varphi_n^{N_n-1}\left(\frac{s}{\sigma_{X^{(n)}}N_n^{1/2}},\frac{\theta}{\sigma_{Y^{(n)}}N_n^{1/2}}\right)\right|\cdot\left|\theta\varphi_n\left(\frac{s}{\sigma_{X^{(n)}}N_n^{1/2}},\frac{\theta}{\sigma_{Y^{(n)}}N_n^{1/2}}\right)\right. \\
&\qquad\qquad \left. + \frac{N_n}{\sigma_{Y^{(n)}}N_n^{1/2}}\frac{\partial\varphi_n}{\partial t}\left(\frac{s}{\sigma_{X^{(n)}}N_n^{1/2}},\frac{\theta}{\sigma_{Y^{(n)}}N_n^{1/2}}\right)\right| \\
&\leqslant e^{\theta^2/2}e^{-(s^2+\theta^2)\cdot c_5(N_n-1)/N_n}(|s|+2|\theta|).
\end{aligned}$$

Finally by (27) and for $N_n \geqslant 2$, we conclude that

$$\begin{aligned}
&\int_0^{\eta\sigma_{Y^{(n)}}N_n^{1/2}} \sup_{0\leqslant\theta\leqslant u} I_2(u,\theta)du \\
&\leqslant 2c_n^{-1}\int_0^{+\infty}\int_{\varepsilon\sigma_{X^{(n)}}N_n^{1/2}}^{+\infty}\sup_{0\leqslant\theta\leqslant u}\left[(s+2\theta)\exp\left(-\frac{u^2}{2}+\frac{\theta^2}{2}\left(1-2c_5\frac{N_n-1}{N_n}\right)\right)\right] \\
&\qquad\qquad \cdot e^{-s^2\cdot c_5(N_n-1)/N_n}dsdu \\
&\leqslant 2\tilde{c}_5^{-1}\int_0^{+\infty}\int_{\varepsilon\sigma_{X^{(n)}}N_n^{1/2}}^{+\infty}(s+2t)e^{-\min(1,c_5)u^2/2}e^{-s^2c_5/2}dsdt \\
&\leqslant 2\tilde{c}_5^{-1}\frac{2}{c_5}e^{-N_nc_5\varepsilon^2\sigma_{X^{(n)}}^2/2}\frac{\sqrt{2\pi}}{2\sqrt{\min(1,c_5)}} + 2\tilde{c}_5^{-1}\frac{2}{\min(1,c_5)}\frac{e^{-N_nc_5\varepsilon^2\sigma_{X^{(n)}}^2/2}}{c_5\varepsilon\sigma_{X^{(n)}}N_n^{1/2}}.
\end{aligned}$$

The conclusion follows with

$$C_2 := 2\tilde{c}_5^{-1}c_5^{-1}\left(\frac{\sqrt{2\pi}}{\sqrt{\min(1,c_5)}} + \frac{2}{\min(1,c_5)\min\left(\frac{2}{9}c_1c_2^3,\pi\right)\tilde{c}_1}\right) \tag{34}$$

and

$$C_3 := c_5 \min\left(\frac{2}{9}c_1c_2^3,\pi\right)^2\tilde{c}_1^2/2. \tag{35}$$

□

To conclude to part b) of Theorem 2.1, just wright

$$C_2e^{-C_3N_n} = \frac{C_2C_3^{-1/2}}{N_n^{1/2}}(C_3N_n)^{1/2}e^{-C_3N_n} \leqslant \frac{C_2C_3^{-1/2}}{N_n^{1/2}}(1/2)^{1/2}e^{-1/2},$$

since $x^{1/2}e^{-x}$ is maximum in $1/2$. So,

$$\sup_x \left|\mathbb{P}\left(\frac{U_n - N_n\mathbb{E}\left[Y^{(n)}\right]}{N_n^{1/2}\tau_n} \leqslant x\right) - \Phi(x)\right| \leqslant \frac{C}{N_n^{1/2}}$$

with

$$C := C_1 + C_2C_3^{-1/2}(1/2)^{1/2}e^{-1/2}. \tag{36}$$

□

Proof of Theorem 2.1 - Part c). We start proving (2). We adapt the proof given in [13]. Using (17) with $\mathbb{E}[Y^{(n)}] = 0$, and differentiating under the integral sign of (18), we naturally have

$$\begin{aligned}|\mathbb{E}\left[U_n\right]| &= \left|\frac{-i\psi_n'(0)}{2\pi\mathbb{P}(S_n = k_n)}\right| \\ &\leqslant \frac{\sigma_{X^{(n)}}^{-1}N_n^{-1/2}N_n}{2\pi\mathbb{P}(S_n = k_n)}\int_{-\pi\sigma_{X^{(n)}}N_n^{1/2}}^{\pi\sigma_{X^{(n)}}N_n^{1/2}}\left|\frac{\partial\varphi_n}{\partial t}\left(\frac{s}{\sigma_{X^{(n)}}N_n^{1/2}},0\right)\right|\cdot\left|\varphi_n^{N_n-1}\left(\frac{s}{\sigma_{X^{(n)}}N_n^{1/2}},0\right)\right|ds.\end{aligned} \tag{37}$$

Using inequality (22) of Lemma 4.3 with $r_n = 0$ and $t = 0$, assumptions (H2.1.1), (H2.1.2), and (H2.1.6), we deduce

$$\left|\frac{\partial\varphi_n}{\partial t}\left(\frac{s}{\sigma_{X^{(n)}}N_n^{1/2}},0\right)\right| \leqslant \frac{s^2}{2}\frac{\rho_{Y^{(n)}}^{1/3}\rho_{X^{(n)}}^{2/3}}{\sigma_{X^{(n)}}^2N_n} \leqslant \frac{c_2^2c_3c_4}{2N_n}s^2.$$

Then using inequality 19 of Lemma 4.2 with $t = 0$ and for $N_n \geqslant 2$,

$$\int_{-\pi\sigma_{X^{(n)}}N_n^{1/2}}^{\pi\sigma_{X^{(n)}}N_n^{1/2}}\left|\frac{\partial\varphi_n}{\partial t}\left(\frac{s}{\sigma_{X^{(n)}}N_n^{1/2}},0\right)\right|\cdot\left|\varphi_n^{N_n-1}\left(\frac{s}{\sigma_{X^{(n)}}N_n^{1/2}},0\right)\right|ds \leqslant \frac{c_2^2c_3c_4}{2N_n}\int_{\mathbb{R}}s^2e^{-c_5s^2/2}ds.$$

So, 2 holds with

$$c_7 := \frac{c_2^2c_3c_4}{2\tilde{c}_5}\int_{\mathbb{R}}s^2e^{-c_5s^2/2}ds. \tag{38}$$

To prove (3), since $\tau_n = \sigma_{Y^{(n)}}$ and $\mathbb{E}[U_n]$ is bounded, it suffices to show that the quantity $\left|\mathbb{E}\left[U_n^2\right] - N_n\sigma_{Y^{(n)}}^2\right|$ is bounded by some $c_8' N_n^{1/2}$. Proceeding as previously,

$$\begin{aligned}
\mathbb{E}\left[U_n^2\right] &= \frac{-\psi_n''(0)}{2\pi\mathbb{P}(S_n = k_n)} \\
&= -c_n^{-1} N_n(N_n-1)\int_{-\pi\sigma_{X^{(n)}}N_n^{1/2}}^{\pi\sigma_{X^{(n)}}N_n^{1/2}} \left(\frac{\partial\varphi_n}{\partial t}\left(\frac{s}{\sigma_{X^{(n)}}N_n^{1/2}},0\right)\right)^2 \varphi_n^{N_n-2}\left(\frac{s}{\sigma_{X^{(n)}}N_n^{1/2}},0\right) ds \qquad (39)\\
&\quad - c_n^{-1}N_n\int_{-\pi\sigma_{X^{(n)}}N_n^{1/2}}^{\pi\sigma_{X^{(n)}}N_n^{1/2}} \frac{\partial^2\varphi_n}{\partial t^2}\left(\frac{s}{\sigma_{X^{(n)}}N_n^{1/2}},0\right)\varphi_n^{N_n-1}\left(\frac{s}{\sigma_{X^{(n)}}N_n^{1/2}},0\right) ds. \qquad (40)
\end{aligned}$$

First, by inequality (22) with $r_n = 0$ and $t = 0$, the control (19) with $t = 0$, and for $N_n \geqslant 3$, one has

$$\begin{aligned}
&\int_{-\pi\sigma_{X^{(n)}}N_n^{1/2}}^{\pi\sigma_{X^{(n)}}N_n^{1/2}} \left|\frac{\partial\varphi_n}{\partial t}\left(\frac{s}{\sigma_{X^{(n)}}N_n^{1/2}},0\right)\right|^2 \left|\varphi_n^{N_n-2}\left(\frac{s}{\sigma_{X^{(n)}}N_n^{1/2}},0\right)\right| dv \\
&\qquad \leqslant \frac{c_2^4c_3^2c_4^2}{4N_n^2}\int_{\mathbb{R}} s^4 e^{-c_5 s^2/3} ds,
\end{aligned}$$

and finally using 2.1.a, the term (39) is bounded by

$$c_8'' := \frac{c_2^4c_3^2c_4^2}{4\tilde{c}_5}\int_{\mathbb{R}} s^4 e^{-c_5 s^2/3} ds. \qquad (41)$$

Second, we study the term (40). We want to show that

$$\Delta_n := c_n^{-1}\int_{-\pi\sigma_{X^{(n)}}N_n^{1/2}}^{\pi\sigma_{X^{(n)}}N_n^{1/2}} \frac{\partial^2\varphi_n}{\partial t^2}\left(\frac{s}{\sigma_{X^{(n)}}N_n^{1/2}},0\right)\varphi_n^{N_n-1}\left(\frac{s}{\sigma_{X^{(n)}}N_n^{1/2}},0\right) ds + \sigma_{Y^{(n)}}^2$$

is bounded by some $c_8'''/N_n^{1/2}$. Recall that, by Lemma 4.1 and assumption (H2.1.4),

$$\int_{-\pi\sigma_{X^{(n)}}N_n^{1/2}}^{\pi\sigma_{X^{(n)}}N_n^{1/2}} \varphi_n^{N_n}\left(\frac{s}{\sigma_{X^{(n)}}N_n^{1/2}},0\right) dv = 2\pi\mathbb{P}(S_n = k_n)\sigma_{X^{(n)}}N_n^{1/2} = c_n,$$

so

$$\begin{aligned}
\Delta_n &= c_n^{-1}\int_{-\pi\sigma_{X^{(n)}}N_n^{1/2}}^{\pi\sigma_{X^{(n)}}N_n^{1/2}} \left(\frac{\partial^2\varphi_n}{\partial t^2}\left(\frac{s}{\sigma_{X^{(n)}}N_n^{1/2}},0\right) + \sigma_{Y^{(n)}}^2\varphi_n\left(\frac{s}{\sigma_{X^{(n)}}N_n^{1/2}},0\right)\right) \\
&\qquad\qquad \cdot \varphi_n^{N_n-1}\left(\frac{s}{\sigma_{X^{(n)}}N_n^{1/2}},0\right) ds \\
&= c_n^{-1}\int_{-\pi\sigma_{X^{(n)}}N_n^{1/2}}^{\pi\sigma_{X^{(n)}}N_n^{1/2}} \mathbb{E}\Big[Y^{(n)2}\Big(- e^{is\sigma_{X^{(n)}}^{-1}N_n^{-1/2}(X^{(n)}-\mathbb{E}[X^{(n)}])} \\
&\qquad\qquad + \mathbb{E}\Big[e^{is\sigma_{X^{(n)}}^{-1}N_n^{-1/2}(X^{(n)}-\mathbb{E}[X^{(n)}])}\Big]\Big)\Big] \\
&\qquad\qquad \cdot \varphi_n^{N_n-1}\left(\frac{s}{\sigma_{X^{(n)}}N_n^{1/2}},0\right) ds.
\end{aligned}$$

Applying Taylor theorem to the function

$$f(s) = -e^{is\sigma_{X^{(n)}}^{-1}N_n^{-1/2}(X^{(n)}-\mathbb{E}[X^{(n)}])} + \mathbb{E}\Big[e^{is\sigma_{X^{(n)}}^{-1}N_n^{-1/2}(X^{(n)}-\mathbb{E}[X^{(n)}])}\Big]$$

yields

$$|f(s)| \leqslant |s| \sup_{u\in[0,s]} \left| -i\frac{X^{(n)}-\mathbb{E}[X^{(n)}]}{\sigma_{X^{(n)}}N_n^{1/2}} e^{iu\sigma_{X^{(n)}}^{-1}N_n^{-1/2}(X^{(n)}-\mathbb{E}[X^{(n)}])} \right.$$
$$\left. +\mathbb{E}\left[i\frac{X^{(n)}-\mathbb{E}[X^{(n)}]}{\sigma_{X^{(n)}}N_n^{1/2}} e^{iu\sigma_{X^{(n)}}^{-1}N_n^{-1/2}(X^{(n)}-\mathbb{E}[X^{(n)}])} \right] \right|$$
$$\leqslant \frac{|s|}{N_n^{1/2}} \left(\left| \frac{X^{(n)}-\mathbb{E}[X^{(n)}]}{\sigma_{X^{(n)}}} \right| + \mathbb{E}\left[\left| \frac{X^{(n)}-\mathbb{E}[X^{(n)}]}{\sigma_{X^{(n)}}} \right| \right] \right).$$

Thus, using Hölder Inequality,

$$\left| \mathbb{E}[{Y^{(n)}}^2 f(s)] \right| \leqslant \frac{|s|}{N_n^{1/2}} \mathbb{E}\left[{Y^{(n)}}^2 \left(\left| \frac{X^{(n)}-\mathbb{E}[X^{(n)}]}{\sigma_{X^{(n)}}} \right| + \mathbb{E}\left[\left| \frac{X^{(n)}-\mathbb{E}[X^{(n)}]}{\sigma_{X^{(n)}}} \right| \right] \right) \right]$$
$$\leqslant \frac{\sigma_{Y^{(n)}}^2 |s|}{N_n^{1/2}} \left(\frac{\rho_{Y^{(n)}}^{2/3}}{\sigma_{Y^{(n)}}^2} \frac{\rho_{X^{(n)}}^{1/3}}{\sigma_{X^{(n)}}} + 1 \right)$$

and, applying equation 2.1.a, assumptions (H2.1.1), (H2.1.2), (H2.1.5), (H2.1.6), and the majoration (19) with $t = 0$, we get

$$|\Delta_n| \leqslant \frac{\sigma_{Y^{(n)}}}{N_n^{1/2} c_n} \left(\frac{\rho_{Y^{(n)}}^{2/3}}{\sigma_{Y^{(n)}}^2} \frac{\rho_{X^{(n)}}^{1/3}}{\sigma_{X^{(n)}}} + 1 \right) \int_{\mathbb{R}} |s|\, e^{-s^2 c_5 (N_n-1)/N_n} ds \leqslant \frac{c_8'''}{N_n^{1/2}}$$

with

$$c_8''' := c_3 \tilde{c}_5^{-1} (1 + c_2 c_4^2) \int_{\mathbb{R}} |s|\, e^{-s^2 c_5/2} ds. \tag{42}$$

Finally,

$$\left| \mathrm{Var}(U_n) - N_n \tau_n^2 \right| \leqslant c_7 + c_8'' + c_8''' N_n^{1/2} \leqslant c_8 N_n^{1/2}$$

with

$$c_8 := c_7 + c_8'' + c_8'''$$
$$= \frac{c_2^2 c_3 c_4}{2\tilde{c}_5} \int_{\mathbb{R}} s^2 e^{-cs^2/2} ds + \frac{c_2^4 c_3^2 c_4^2}{4\tilde{c}_5} \int_{\mathbb{R}} s^4 e^{-c_5 s^2/3} ds + c_3 \tilde{c}_5^{-1} (1 + c_2 c_4^2) \int_{\mathbb{R}} |s|\, e^{-s^2 c_5/2} ds. \tag{43}$$

Now we turn to the proof of (4). Let us show that the previous estimates of $\mathbb{E}[U_n]$ and $\mathrm{Var}(U_n)$ make it possible to apply (1). Remind that $\mathbb{E}\left[Y^{(n)}\right] = 0$. Write

$$\left\{ \frac{U_n - \mathbb{E}[U_n]}{\mathrm{Var}\,(U_n)^{1/2}} \leqslant x \right\} = \left\{ \frac{U_n}{N_n^{1/2}\sigma_{Y^{(n)}}} \leqslant a_n x + b_n \right\},$$

where

$$a_n := \frac{\mathrm{Var}(U_n)^{1/2}}{N_n^{1/2}\sigma_{Y^{(n)}}} \quad \text{and} \quad b_n := \frac{\mathbb{E}[U_n]}{N_n^{1/2}\sigma_{Y^{(n)}}}.$$

The previous estimates of $\mathbb{E}[U_n]$ and $\mathrm{Var}(U_n)$ yield

$$|a_n - 1| \leqslant \left| a_n^2 - 1 \right| \leqslant c_8 \tilde{c}_3^{-1} N_n^{-1/2} \quad \text{and} \quad b_n \leqslant c_7 \tilde{c}_3^{-1} N_n^{-1/2}.$$

Now,

$$\left| \mathbb{P}\left(\frac{U_n - \mathbb{E}[U_n]}{\mathrm{Var}\,(U_n)^{1/2}} \leqslant x \right) - \Phi(x) \right| \leqslant \left| \mathbb{P}\left(\frac{U_n}{N_n^{1/2}\sigma_{Y^{(n)}}} \leqslant a_n x + b_n \right) - \Phi(a_n x + b_n) \right|$$
$$+ |\Phi(a_n x + b_n) - \Phi(x)|$$
$$\leqslant \frac{C_1}{N_n^{1/2}} + C_2 e^{-C_3 N_n} + |\Phi(a_n x + b_n) - \Phi(x)|.$$

For $N_n > 4c_8^2/\tilde{c}_3^2$, $a_n \geqslant 1/2$ and applying Taylor theorem to Φ yields

$$\begin{aligned}|\Phi(a_n x + b_n) - \Phi(x)| &\leqslant |(a_n - 1)x + b_n| \sup_t \frac{e^{-t^2/2}}{\sqrt{2\pi}} \\ &\leqslant N_n^{-1/2} \max(c_8\tilde{c}_3^{-1}, c_7\tilde{c}_3^{-1})(|x|+1)e^{-(|x|/2 - c_7\tilde{c}_3^{-1})^2/2},\end{aligned}$$

the supremum being over t between x and $a_n x + b_n$. The last function in x being bounded, we get (4) with

$$\tilde{C}_1 := \max(c_8\tilde{c}_3^{-1}, c_7\tilde{c}_3^{-1}) \sup_{x\in\mathbb{R}} \left[(|x|+1)e^{-(|x|/2 - c_7\tilde{c}_3^{-1})^2/2}\right].$$

□

4.3 Proof of Theorem 2.6

We start with the proof of Theorem 2.6, which relies on three different lemmas.

Proof of Theorem 2.6. Let z_n such that $\liminf_{n\to\infty} \frac{z_n}{N_n} > 0$. Since $Y^{(n)} - \mathbb{E}\left[Y^{(n)}\right]$ also satisfies the hypotheses, we can assume that $\mathbb{E}\left[Y^{(n)}\right] = 0$. Define

$$P_{N_n} = \mathbb{P}(T_n \geqslant z_n)$$

and for any $m \in [\![0, N_n]\!]$,

$$P_{N_n,m} = \mathbb{P}\left(T_n \geqslant z_n, \quad \forall i \in [\![1, N_n - m]\!]\ Y_i^{(n)} < z_n, \quad \forall i \in [\![N_n - m + 1, N_n]\!]\ Y_i^{(n)} \geqslant z_n\right)$$

with the usual convention $[\![1, 0]\!] = \emptyset$ and $[\![N_n + 1, N_n]\!] = \emptyset$. Now write

$$P_{N_n} = P_{N_n,0} + N_n P_{N_n,1} + \sum_{m=2}^{N_n} \binom{N_n}{m} P_{N_n,m}. \tag{44}$$

Using Lemmas 4.9, 4.10 and 4.11 that follow, we conclude the proof of Theorem 2.6. □

Lemma 4.9.

$$\limsup_{n\to\infty} \frac{1}{\sqrt{z_n}} \log(P_{N_n,0}) \leqslant -\alpha.$$

Lemma 4.10.

$$-\beta \leqslant \liminf_{n\to\infty} \frac{1}{\sqrt{z_n}} \log(N_n P_{N_n,1}) \leqslant \limsup_{n\to\infty} \frac{1}{\sqrt{z_n}} \log(N_n P_{N_n,1}) \leqslant -\alpha.$$

Lemma 4.11.

$$\sum_{m=2}^{N_n} \binom{N_n}{m} P_{N_n,m} = o\left(e^{-\alpha\sqrt{z_n}}\right).$$

Proof of Theorem 2.6. Lemmas 4.9, 4.10, and 4.11 yield, for all $\alpha' < \alpha$,

$$\begin{aligned}-\beta &\leqslant \liminf_{n\to\infty} \frac{1}{\sqrt{z_n}} \log(N_n P_{N_n,1}) \leqslant \liminf_{n\to\infty} \frac{1}{\sqrt{z_n}} \log(P_{N_n}) \\ &\leqslant \limsup_{n\to\infty} \frac{1}{\sqrt{z_n}} \log(P_{N_n}) \leqslant \lim_{n\to\infty} \frac{1}{\sqrt{z_n}} \log\left(3e^{-\alpha'\sqrt{z_n}}\right) = -\alpha'.\end{aligned}$$

Conclude by letting $\alpha' \to \alpha$. □

Proof of Lemma 4.11. Let $\alpha' \in]\alpha/2, \alpha[$. Using (9) and noting that $z_n \geqslant \sqrt{z_n}$ for n large enough, we have, for all n large enough,

$$\sum_{m=2}^{N_n} \binom{N_n}{m} P_{N_n,m} \leqslant \sum_{m=2}^{N_n} {N_n}^m \mathbb{P}(Y_1^{(n)} \geqslant z_n)^m \leqslant \frac{{N_n}^2 e^{-2\alpha'\sqrt{z_n}}}{1 - N_n e^{-\alpha'\sqrt{z_n}}} = o\left(e^{-\alpha\sqrt{z_n}}\right).$$

□

Proof of Lemma 4.10. First, using (9),

$$\limsup_{n\to\infty} \frac{1}{\sqrt{z_n}} \log(N_n P_{N_n,1}) \leqslant \limsup_{n\to\infty} \frac{1}{\sqrt{z_n}} \log \mathbb{P}(Y^{(n)} \geqslant z_n) \leqslant -\alpha.$$

Let us prove the converse inequality. Let $\varepsilon > 0$. We have

$$\begin{aligned}
P_{N_n,1} &= \mathbb{P}\left(T_n \geqslant z_n, \quad Y_{N_n}^{(n)} \geqslant z_n, \quad \forall i \in [\![1, N_n - 1]\!] \quad Y_i^{(n)} < z_n\right) \\
&= \int_{z_n}^{+\infty} \mathbb{P}\left(T_{n-1} \geqslant z_n - u, \quad \forall i \in [\![1, N_n - 1]\!] \quad Y_i^{(n)} < z_n\right) \mathbb{P}(Y^{(n)} \in du) \\
&\geqslant \int_{z_n + N_n\varepsilon}^{+\infty} \mathbb{P}\left(T_{n-1} \geqslant z_n - u, \quad \forall i \in [\![1, N_n - 1]\!] \quad Y_i^{(n)} < z_n\right) \mathbb{P}(Y^{(n)} \in du) \\
&\geqslant \mathbb{P}\left(T_{n-1} \geqslant -N_n\varepsilon, \quad \forall i \in [\![1, N_n - 1]\!] \quad Y_i^{(n)} < z_n\right) \mathbb{P}(Y^{(n)} \geqslant z_n + N_n\varepsilon).
\end{aligned}$$

Observe that

$$\begin{aligned}
&\mathbb{P}\big(T_{n-1} \geqslant -N_n\varepsilon, \quad \forall i \in [\![1, N_n - 1]\!] \quad Y_i^{(n)} < z_n\big) \\
&\qquad \geqslant \mathbb{P}\left(Y^{(n)} < z_n\right)^{N_n - 1} - \mathbb{P}\left(T_{n-1} < -N_n\varepsilon\right) \to 1.
\end{aligned}$$

Indeed, $\mathbb{P}\left(Y_1^{(n)} < z_n\right)^{N_n-1} \to 1$, using (9); and, by Chebyshev inequality and assumption (H2.6.2),

$$\mathbb{P}\left(T_{n-1} < -N_n\varepsilon\right) \leqslant \frac{\sigma^2_{Y^{(n)}}}{N_n\varepsilon^2} \to 0,$$

the random variables $Y^{(n)}$ being assumed centered. Finally, using (8) and (H2.6.1), and noting $\delta = \liminf_{n\to\infty} \frac{z_n}{N_n}$, one gets

$$\begin{aligned}
\liminf_{n\to\infty} \frac{1}{\sqrt{z_n}} \log(N_n P_{N_n,1}) &\geqslant \liminf_{n\to\infty} \sqrt{\frac{z_n + N_n\varepsilon}{z_n}} \frac{1}{\sqrt{z_n + N_n\varepsilon}} \log \mathbb{P}(Y^{(n)} \geqslant z_n + N_n\varepsilon) \\
&\geqslant -\beta\sqrt{\frac{\delta + \varepsilon}{\delta}}.
\end{aligned}$$

Conclude by letting $\varepsilon \to 0$. □

Proof of Lemma 4.9. Let $\alpha' \in]0, \alpha[$ and $s_n = \alpha'/\sqrt{z_n}$. The exponential Chebyshev inequality for T_n conditioned on $\{\forall i \in [\![1, N_n]\!], Y_i^{(n)} < z_n\}$ yields

$$P_{N_n,0} \leqslant e^{-s_n z_n} \mathbb{E}\left[e^{s_n Y^{(n)}} 1_{Y^{(n)} < z_n}\right]^{N_n}.$$

If we prove that

$$\mathbb{E}\left[e^{s_n Y^{(n)}} 1_{Y^{(n)} < z_n}\right] = 1 + o\left(\frac{1}{N_n^{1/2}}\right),$$

then

$$\log(P_{N_n,0}) \leqslant -\alpha'\sqrt{z_n} + o(N_n^{1/2})$$

and the conclusion follows by letting $\alpha' \to \alpha$. Let $\eta \in]3/4, 1[$. Write

$$\begin{aligned}
&\mathbb{E}\left(e^{s_n Y^{(n)}} 1_{Y^{(n)}<z_n}\right) \\
&= \int_{-\infty}^{\sqrt{z_n}} e^{s_n u}\mathbb{P}(Y^{(n)} \in du) + \int_{\sqrt{z_n}}^{z_n-(z_n)^\eta} e^{s_n u}\mathbb{P}(Y^{(n)} \in du) + \int_{z_n-(z_n)^\eta}^{z_n} e^{s_n u}\mathbb{P}(Y^{(n)} \in du) \\
&=: I_1 + I_2 + I_3.
\end{aligned}$$

By a Taylor expansion of $f(t) = e^t$, (H2.6.2) and (H2.6.1), there exists

$$\theta(u) \leqslant s_n u \leqslant s_n\sqrt{z_n} = \alpha'$$

such that

$$\begin{aligned}
I_1 &\leqslant \int_{-\infty}^{\sqrt{z_n}} \left(1 + s_n u + \frac{s_n^2 u^2}{2} e^{\theta(u)}\right) \mathbb{P}(Y^{(n)} \in du) \\
&\leqslant \int_{-\infty}^{+\infty} \left(1 + s_n u + \frac{s_n^2 u^2}{2} e^{\alpha'}\right) \mathbb{P}(Y^{(n)} \in du) = 1 + 0 + \frac{\alpha'^2 \sigma^2_{Y^{(n)}}}{2z_n} e^{\alpha'} = 1 + o\left(\frac{1}{N_n^{1/2}}\right).
\end{aligned}$$

Let n_0 such that, for all $n \geqslant n_0$ and $u \geqslant \sqrt{z_n}$, $\log \mathbb{P}(Y^{(n)} \geqslant u) \leqslant -\alpha'\sqrt{u}$. Suppose n is larger than n_0. Integrating by part, we get

$$\begin{aligned}
I_2 &= -\Big[e^{s_n u}\mathbb{P}(Y^{(n)} \geqslant u)\Big]_{\sqrt{z_n}}^{z_n-(z_n)^\eta} + s_n \int_{\sqrt{z_n}}^{z_n-(z_n)^\eta} e^{s_n u}\mathbb{P}(Y^{(n)} \geqslant u) du \\
&\leqslant e^{s_n\sqrt{z_n}}\mathbb{P}(Y^{(n)} \geqslant \sqrt{z_n}) + s_n \int_{\sqrt{z_n}}^{z_n-(z_n)^\eta} e^{s_n u - \alpha'\sqrt{u}} du \\
&\leqslant e^{\alpha'(1-(z_n)^{1/4})} + s_n \int_{\sqrt{z_n}}^{z_n-(z_n)^\eta} \exp\left(\alpha'\left(\frac{u}{\sqrt{z_n}} - \sqrt{u}\right)\right) du.
\end{aligned}$$

Since, for all $t \in [0,1]$, $\sqrt{1-t} \leqslant 1 - t/2$, we get, for all $u \in [\sqrt{z_n}, z_n - (z_n)^\eta]$ and n large enough to have $(z_n)^{\nu-1} < 1$,

$$\frac{u}{\sqrt{z_n}} - \sqrt{u} \leqslant \sqrt{u}\left(\sqrt{1-(z_n)^{\eta-1}} - 1\right) \leqslant -\frac{(z_n)^{\eta-3/4}}{2}.$$

Hence,

$$I_2 = o\left(\frac{1}{N_n^{1/2}}\right).$$

Let $\alpha'' \in]\alpha', \alpha \wedge 2\alpha'[$. Let n_1 such that, for all $n \geqslant n_1$ and $u \geqslant z_n - z_n^\eta$, $\log \mathbb{P}(Y^{(n)} \geqslant u) \leqslant -\alpha''\sqrt{u}$. Suppose n is larger than n_1. Integrating by part, we get

$$\begin{aligned}
I_3 &= -\Big[e^{s_n u}\mathbb{P}(Y^{(n)} \geqslant u)\Big]_{z_n-z_n^\eta}^{z_n} + s_n \int_{z_n-z_n^\eta}^{z_n} e^{s_n u}\mathbb{P}(Y^{(n)} \geqslant u) du \\
&\leqslant e^{s_n(z_n-z_n^\eta)}\mathbb{P}(Y^{(n)} \geqslant z_n - z_n^\eta) + s_n \int_{z_n-z_n^\eta}^{z_n} e^{s_n u - \alpha''\sqrt{u}} du.
\end{aligned}$$

Now, since $\sqrt{t} \geqslant t$ if $t \in [0,1]$,

$$\begin{aligned}
e^{s_n(z_n-z_n^\eta)}\mathbb{P}(Y^{(n)} \geqslant z_n - z_n^\eta) &\leqslant \exp\left(\sqrt{z_n}\left(\alpha'\left(1 - z_n^{\eta-1}\right) - \alpha''\left(1 - z_n^{\eta-1}\right)^{1/2}\right)\right) \\
&\leqslant \exp\left(\sqrt{z_n}(\alpha' - \alpha'')(1 - z_n^{\eta-1})\right) = o\left(\frac{1}{N_n^{1/2}}\right).
\end{aligned}$$

Finally, applying Taylor theorem to the function $f(u) = s_n u - \alpha''\sqrt{u}$ around the point z_n yields

$$f(u) = \frac{\alpha' u}{\sqrt{z_n}} - \alpha''\sqrt{u} = (\alpha' - \alpha'')\sqrt{z_n} + \left(\frac{\alpha'}{\sqrt{z_n}} - \frac{\alpha''}{2\sqrt{c}}\right)(u - z_n)$$

with $c \in [u, z_n]$. Since $\alpha'' < 2\alpha'$, we have

$$\left(\frac{\alpha'}{\sqrt{z_n}} - \frac{\alpha''}{2\sqrt{c}}\right)(u - z_n) \leqslant \left(\frac{\alpha'}{\sqrt{z_n}} - \frac{\alpha''}{2\sqrt{z_n - z_n^\eta}}\right)(u - z_n) \leqslant 0,$$

for n large enough and we conclude that

$$I_3 = o\left(\frac{1}{N_n^{1/2}}\right).$$

□

4.4 Proof of Theorem 2.4

Now we turn to the proof of Theorem 2.4. So as to apply Theorem 2.6, we need the next result, which is analogous to equation (2).

Proposition 4.12. *Under assumptions (H2.4.1), (H2.4.3) and (H2.4.5), one has*

$$\mathbb{E}\left[T_n \middle| S_n = k_n\right] = N_n \mathbb{E}\left[Y^{(n)}\right] + o(N_n).$$

Proof. Using inequality (37) and Proposition 4.5 yield

$$\begin{aligned}
&\left|\mathbb{E}\left[T_n - N_n\mathbb{E}\left[Y^{(n)}\right]\middle| S_n = k_n\right]\right| = \left|\frac{-i\psi_n'(0)}{2\pi\mathbb{P}(S_n = k_n)}\right| \\
&\leqslant \frac{N_n}{2\pi m}\int_{-\pi\sigma_{X^{(n)}}N_n^{1/2}}^{\pi\sigma_{X^{(n)}}N_n^{1/2}} \left|\frac{\partial\varphi_n}{\partial t}\left(\frac{s}{\sigma_{X^{(n)}}N_n^{1/2}}, 0\right)\right| \cdot \left|\varphi_n^{N_n - 1}\left(\frac{s}{\sigma_{X^{(n)}}N_n^{1/2}}, 0\right)\right| ds. \qquad (45)
\end{aligned}$$

It remains to show that the integral converges to 0. Putting together (45) and (22), and using hypothesis (H2.4.5) and the control (20), one gets

$$\mathbb{E}\left[T_n - N_n\mathbb{E}\left[Y^{(n)}\right]\middle| S_n = k_n\right] = o(N_n).$$

□

Proof of Theorem 2.4. Let $y > 0$. Since $(X^{(n)}, Y^{(n)} - \mathbb{E}\left[Y^{(n)}\right])$ also satisfies the hypotheses, we can assume that $\mathbb{E}\left[Y^{(n)}\right] = 0$. According to Proposition 4.12,

$$y_n := y + \frac{1}{N_n}\mathbb{E}\left[U_n\right] \to y.$$

We have

$$\begin{aligned}
\mathbb{P}(U_n - \mathbb{E}\left[U_n\right] \geqslant N_n y) &= \mathbb{P}(T_n - \mathbb{E}\left[T_n | S_n = k_n\right] \geqslant N_n y | S_n = k_n) \\
&= \frac{\mathbb{P}(T_n \geqslant N_n y_n,\, S_n = k_n)}{\mathbb{P}(S_n = k_n)} \leqslant \frac{\mathbb{P}(T_n \geqslant N_n y_n)}{\mathbb{P}(S_n = k_n)}.
\end{aligned}$$

The conclusion follows using Theorem 2.6, Proposition 4.5 and (H2.4.1).

Using decomposition (44), we get

$$\begin{aligned}\mathbb{P}(U_n - \mathbb{E}[U_n] \geqslant N_n y) &= \mathbb{P}(T_n - \mathbb{E}[T_n|S_n = k_n] \geqslant N_n y|S_n = k_n)\\ &= \frac{\mathbb{P}(T_n \geqslant N_n y_n,\, S_n = k_n)}{\mathbb{P}(S_n = k_n)} \geqslant \mathbb{P}(T_n \geqslant N_n y_n,\, S_n = k_n)\\ &\geqslant N_n \mathbb{P}\big(T_n \geqslant N_n y_n, \quad Y_n^{(n)} \geqslant N_n y_n, \forall i \in [\![1, N_n - 1]\!] \quad Y_i^{(n)} < N_n y_n, \quad S_n = k_n\big).\end{aligned}$$

Define

$$Q_{N_n,1} := \mathbb{P}\left(T_n \geqslant N_n y_n, \quad Y_n^{(n)} \geqslant N_n y_n, \quad \forall i \in [\![1, N_n - 1]\!] \quad Y_i^{(n)} < N_n y_n, \quad S_n = k_n\right).$$

It remains to show that

$$\liminf_{n\to\infty} \frac{1}{\sqrt{N_n y}} \log(N_n Q_{N_n,1}) \geqslant -\beta,$$

which is analogous to the lower bound of Lemma 4.10. We have, for any $\varepsilon > 0$,

$$\begin{aligned}Q_{N_n,1} &= \mathbb{P}\left(T_n \geqslant N_n y_n, \quad Y_n^{(n)} \geqslant N_n y_n, \quad \forall i \in [\![1, N_n - 1]\!] \quad Y_i^{(n)} < N_n y_n, \quad S_n = k_n\right)\\ &= \int_{N_n y_n}^{+\infty} \mathbb{P}\left(T_{n-1} \geqslant N_n y_n - u,\ \forall i \in [\![1, N_n - 1]\!]\ Y_i^{(n)} < N_n y_n,\ S_n = k_n\right) \mathbb{P}(Y^{(n)} \in du)\\ &\geqslant \int_{N_n(y_n+\varepsilon)}^{+\infty} \mathbb{P}\left(T_{n-1} \geqslant N_n y_n - u,\ \forall i \in [\![1, N_n - 1]\!]\ Y_i^{(n)} < N_n y_n,\ S_n = k_n\right) \mathbb{P}(Y^{(n)} \in du)\\ &\geqslant \mathbb{P}\left(T_{n-1} \geqslant -N_n\varepsilon,\ \forall i \in [\![1, N_n - 1]\!]\ Y_i^{(n)} < N_n y_n,\ S_n = k_n\right) \mathbb{P}(Y^{(n)} \geqslant N_n(y_n + \varepsilon)).\end{aligned}$$

Observe that

$$\begin{aligned}\mathbb{P}\big(T_{n-1} \geqslant -N_n\varepsilon, \quad \forall i \in [\![1, N_n - 1]\!] \quad Y_i^{(n)} < N_n y_n, \quad S_n = k_n\big)\\ \geqslant \mathbb{P}\left(Y^{(n)} < N_n y_n\right)^{N_n - 1} - (1 - \mathbb{P}(S_n = k_n)) - \mathbb{P}\left(T_{n-1} < -N_n\varepsilon\right).\end{aligned}$$

For $\alpha' \in]0, \alpha[$ and n large enough, using (7), one has

$$\mathbb{P}\left(Y^{(n)} < N_n y_n\right)^{N_n - 1} \geqslant (1 - e^{-\alpha'\sqrt{N_n y_n}})^{N_n - 1} = 1 + o\left(\frac{1}{N_n^{1/2}}\right).$$

By Chebyshev Inequality and hypothesis (H2.4.5), one has straightforwardly

$$\mathbb{P}\left(T_{n-1} < -N_n\varepsilon\right) \leqslant \frac{\sigma^2_{Y^{(n)}}}{N_n \varepsilon^2} = o\left(\frac{1}{N_n^{1/2}}\right).$$

Hence, using Proposition 4.5 and hypotheses (H2.4.1) and 6,

$$\begin{aligned}\liminf_{n\to\infty} \frac{1}{\sqrt{N_n y}} \log(N_n Q_{n,1}) &\geqslant \liminf_{n\to\infty} \frac{1}{\sqrt{N_n y}} \log\left(\frac{m}{\sigma_{X^{(n)}} N_n^{1/2}}\right) + \liminf_{n\to\infty} \frac{1}{\sqrt{N_n y}} \log \mathbb{P}(Y^{(n)}\\ &\geqslant N_n(y_n + \varepsilon)) \geqslant -\beta\sqrt{\frac{y+\varepsilon}{y}}.\end{aligned}$$

Conclude by letting $\varepsilon \to 0$. □

References

[1] P. Chassaing and G. Louchard. Phase transition for parking blocks, Brownian excursion and coalescence. *Random Structures Algorithms*, 21(1):76–119, 2002.

[2] Philippe Chassaing and Svante Janson. A Vervaat-like path transformation for the reflected Brownian bridge conditioned on its local time at 0. *Ann. Probab.*, 29(4):1755–1779, 2001.

[3] Philippe Chassaing and Jean-François Marckert. Parking functions, empirical processes, and the width of rooted labeled trees. *Electron. J. Combin.*, 8(1):Research Paper 14, 19, 2001.

[4] Imre Csiszar. Sanov property, generalized i-projection and a conditional limit theorem. *The Annals of Probability*, 12(3):768–793, 08 1984.

[5] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*, volume 38 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 1998.

[6] William Feller. *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons Inc., New York, 1968.

[7] William Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York, 1971.

[8] P. Flajolet, P. Poblete, and A. Viola. On the analysis of linear probing hashing. *Algorithmica*, 22(4):490–515, 1998. Average-case analysis of algorithms.

[9] Fabrice Gamboa, Thierry Klein, and Clémentine Prieur. Conditional large and moderate deviations for sums of discrete random variables. Combinatoric applications. *Bernoulli*, 18(4):1341–1360, 2012.

[10] Christian Hipp. Asymptotic expansions for conditional distributions: the lattice case. *Probab. Math. Statist.*, 4(2):207–219, 1984.

[11] Lars Holst. Two conditional limit theorems with applications. *Ann. Statist.*, 7(3):551–557, 1979.

[12] Svante Janson. Asymptotic distribution for the cost of linear probing hashing. *Random Structures Algorithms*, 19(3-4):438–471, 2001. Analysis of algorithms (Krynica Morska, 2000).

[13] Svante Janson. Moment convergence in conditional limit theorems. *J. Appl. Probab.*, 38(2):421–437, 2001.

[14] Svante Janson. Individual displacements for linear probing hashing with different insertion policies. *ACM Trans. Algorithms*, 1(2):177–213, 2005.

[15] Svante Janson. Individual displacements in hashing with coalesced chains. *Combin. Probab. Comput.*, 17(6):799–814, 2008.

[16] Donald E. Knuth. *The art of computer programming. Vol. 3.* Addison-Wesley, Reading, MA, 1998. Sorting and searching, Second edition [of MR0445948].

[17] Valentin F. Kolchin. *Random mappings*. Translation Series in Mathematics and Engineering. Optimization Software, Inc., Publications Division, New York, 1986. Translated from the Russian, With a foreword by S. R. S. Varadhan.

[18] È. M. Kudlaev. Conditional limit distributions of sums of random variables. *Teor. Veroyatnost. i Primenen.*, 29(4):743–752, 1984.

[19] Michel Loève. *Probability theory. Foundations. Random sequences*. D. Van Nostrand Company, Inc., Toronto-New York-London, 1955.

[20] Jean-François Marckert. Parking with density. *Random Structures Algorithms*, 18(4):364–380, 2001.

[21] A. Nagaev. Integral limit theorems taking large deviations into account when cramér's condition does not hold. i. *Theory of Probability and Its Applications*, 14(1):51–64, 1969.

[22] A. Nagaev. Integral limit theorems taking large deviations into account when cramér's condition does not hold. ii. *Theory of Probability and Its Applications*, 14(2):193–208, 1969.

[23] Ju. L. Pavlov. Limit theorems for the number of trees of a given size in a random forest. *Mat. Sb. (N.S.)*, 103(145)(3):392–403, 464, 1977.

[24] Yu. L. Pavlov. Random forests. In *Probabilistic methods in discrete mathematics (Petrozavodsk, 1996)*, pages 11–18. VSP, Utrecht, 1997.

[25] M. P. Quine and J. Robinson. A Berry-Esseen bound for an occupancy problem. *Ann. Probab.*, 10(3):663–671, 1982.

[26] J. Robinson, T. Höglund, L. Holst, and M. P. Quine. On approximating probabilities for small and large deviations in $\mathbf{R}^d$. *Ann. Probab.*, 18(2):727–753, 1990.

[27] George P. Steck. Limit theorems for conditional distributions. *Univ. California Publ. Statist.*, 2:237–284, 1957.

[28] Jan M. Van Campenhout and Thomas M. Cover. Maximum entropy and conditional probability. *IEEE Trans. Inform. Theory*, 27(4):483–489, 1981.