Large deviations and Berry-Esseen bounds for hashing with linear probing

Thierry Klein, A Lagnoux, P Petit

Large deviations and Berry-Esseen bounds for hashing with linear probing

T. Klein, A. Lagnoux, and P. Petit*

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. First, we prove a large deviations result in the context of hashing with linear probing. By the way, we establish a large deviations result for triangular arrays when the Laplace transform is not defined in a neighborhood of 0. Second, we prove a Berry-Esseen bound in a general setting.

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 on some exogenous integer-valued random variable. In general, the exogenous random variable is itself a sum of integer-valued random variables. More precisely, we are interested in the law of $N_n^{-1}(Y_1^{(n)} + \cdots + Y_{N_n}^{(n)})$ conditioned on a specific value of $X_1^{(n)} + \cdots + X_{N_n}^{(n)}$ that is to say in the conditional distribution

$$\mathcal{L}_n := \mathcal{L}(N_n^{-1}(Y_1^{(n)} + \cdots + Y_{N_n}^{(n)}) \mid X_1^{(n)} + \cdots + X_{N_n}^{(n)} = m_n),$$

where m_n and N_n are integers and $(X_i^{(n)}, Y_i^{(n)})_{n\in\mathbb{N}^*, 1\leqslant i\leqslant N_n}$ be i.i.d. copies of a pair $(X^{(n)}, Y^{(n)})$ of random variables with $X^{(n)}$ integer-valued.

Hashing with linear probing was the motivating example for Janson's work [13]. 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 [17]. Due to its strong connection with parking functions, the Airy distributions (i.e., the area under the brownian excursion) and the Lukasiewicz random walks [20], this model was studied by many authors (see, e.g., Flajolet, Poblete and Viola [8], Janson [12, 14, 15], Chassaing *et al.* [1, 2, 3], and Marckert [22]).

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 combinatorial 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 large deviations result for these models?

2. Is it possible to obtain a general Berry-Esseen bound for these models?

*Institut de Mathématiques de Toulouse, France. `firstname.lastname@math.univ-toulouse.fr`

Partial answer to the first question When the distribution of $(X^{(n)}, Y^{(n)})$ does not depend on n, the Gibbs conditioning principle ([30, 4, 5]) states that $\mathcal{L}_n$ converges weakly to the degenerated distribution concentrated on a point 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 [29, 11, 19] and asymptotic expansions are proved in [10, 28].
An extension to arrays has been proposed by Gamboa, Klein and Prieur [9]. They prove a large (and a moderate) deviation principle under some strong assumptions. The most restricting assumption is that the joint Laplace transform of $(X^{(n)}, Y^{(n)})$ is finite at least in a neighborhood of $(0, 0)$. This assumption is satisfied by all the examples considered in [13] except for hashing with linear probing, which is the most interesting one. Indeed, in this case, the joint Laplace transform is only defined in $(-\infty, a] \times (-\infty, 0]$ for some positive a.
More generally, it would be interesting to get large deviations results for a larger class of models the Laplace transforms of which are not defined. In [23, 24], Nagaev establishes large deviations results for sums of i.i.d. 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 extend his result and prove a large deviations result (Theorem 3.3) for arrays. It is then natural to consider the asymptotic behavior of conditioned sums and to extend the work of [9] to models the Laplace transforms of which are not defined. Proving a theorem for a general class of models seems to be a very difficult task. That is why, we restrict ourselves to the study of hashing with linear probing (Theorem 3.1).
Let us point out the main differences between the large deviations result of the present work and [9, Theorem 2.1]. First, the proof in [9, Theorem 2.1] 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 [23, 24]. 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, oscillations of the tails are allowed (in a controlled range) and may affect the large deviation bounds. 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.

Complete answer to the second question The first Berry-Esseen theorem for conditional models is given by Quine and Robinson [27]. In their work, the authors study the particular case of the occupancy problem, i.e. the case when 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 4.1) that covers all the examples presented by Janson [13].

Organization of the paper The paper is organized as follows. In Section 2, we present the general model and describe precisely the framework of hashing with linear probing. Section 3 is devoted to the large deviations result for hashing. A Berry-Esseen bound (Theorem 4.1) is stated in Section 4, which applies to the examples presented by Janson [13]. Finally, the last section is dedicated to the proofs.

2 The model

2.1 A general framework for conditional distributions

In the whole paper, $\mathbb{N}^* = \{1, 2, \dots\}$ is the set of positive integers, $\mathbb{N} = \mathbb{N}^* \cup \{0\}$, and $\mathbb{Z}$ is the set of all integers. 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

$\left(X_i^{(n)}, Y_i^{(n)}\right)_{1 \leqslant i \leqslant N_n}$ be an i.i.d. sample distributed as $\left(X^{(n)}, Y^{(n)}\right)$ and define

$$S_k^{(n)} := \sum_{i=1}^{k} X_i^{(n)} \quad \text{and} \quad T_k^{(n)} := \sum_{i=1}^{k} Y_i^{(n)},$$

for $k \in \{1, \ldots, N_n\}$. Let $m_n \in \mathbb{Z}$ be such that $\mathbb{P}(S_{N_n}^{(n)} = m_n) > 0$. The purpose of the paper is to derive the asymptotic behavior of the conditional distribution

$$\mathcal{L}_n := \mathcal{L}((N_n)^{-1} T_{N_n}^{(n)} | S_{N_n}^{(n)} = m_n).$$

2.2 Classical examples

In this section, we give several examples.

2.2.1 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. It is well known that $(Z_1, \ldots, Z_N)$ is also distributed as $(X_1, \cdots, X_N)$ 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 of empty urns which is distributed as $\sum_{i=1}^{N} \mathbb{1}_{\{X_i = 0\}}$ conditioned on $\sum_{i=1}^{N} X_i = m$.
Let $m = m_n \to \infty$ and $N = N_n \to \infty$ with $\lambda_n := m_n/N_n \to \lambda$. Following the work of Janson [13], we will study the asymptotic behavior of $T_{N_n}^{(n)} = \sum_{i=1}^{N_n} \mathbb{1}_{\{X_i^{(n)} = 0\}}$ ($Y^{(n)} = \mathbb{1}_{\{X^{(n)} = 0\}}$) conditioned on $S_{N_n}^{(n)} = \sum_{i=1}^{N_n} X_i^{(n)} = m_n$ with $X^{(n)} \sim \mathcal{P}(\lambda_n)$.

2.2.2 Bose-Einstein statistics

This example is borrowed from [11], see also [6]. Consider N urns. Put n indistinguishable balls in the urns in such a way that each distinguishable outcome has the same probability

$$1 \Big/ \binom{n + N - 1}{n}.$$

Let Z_k be the number of balls in the k^{th} urn. It is well known that $(Z_1, \ldots, Z_N)$ is distributed as $\left(X_1, \cdots, X_N\right)$ conditioned on $\left\{\sum_{i=1}^{N} X_i = n\right\}$, where $X_1, \cdots, X_N$ are i.i.d. and geometrically distributed with any parameter p. The framework is similar to the one of Subsection 2.2.1 and we proceed analogously. Assume $m = m_n = n \to \infty$, $N = N_n \to \infty$ with $N_n/n \to p$, and take $X_i^{(n)}$ having geometric distribution with parameter $p_n = N_n/n$.

2.2.3 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 $n \geqslant 1$ if and only if

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

This type of conditioning is different from the one studied in the present paper, but by [31, Corollary 2] and [13, Example 3.4], if we ignore the order of $X_1, \ldots, X_n$, it is proven that they have the same distribution conditioned on (1) 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 $\sum_{i=1}^{n} Y_i$ is the number of families with three children. The framework is similar to the one of Subsection 2.2.1 and we proceed analogously with $m = m_n = n-1 \to \infty$, $N = N_n = n \to \infty$.

2.2.4 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 $1, \ldots, m$ 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 $X_1, \ldots, X_N$ conditioned on $\sum_{i=1}^{N} X_i = m$, where X_i are i.i.d. as the Borel distribution with some arbitrary parameter $\lambda \in \left(0, e^{-1}\right]$ that is defined in the following way

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

where T is the tree function (see, e.g., [8] or [12] for more details). Then, 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., [18, 25, 26]). It means that we consider $Y_i = 1_{\{X_i=K\}}$. Let us now assume that we condition on $\sum_{i=1}^{N} X_i = m$. The framework is similar to the one of Subsection 2.2.1 and we proceed analogously. Assume $m = m_n \to \infty$, $N = N_n \to \infty$ with $m_n/N_n \to \lambda$, and take $X_i^{(n)}$ having Borel distribution with parameter $\lambda_n = m_n/N_n$.

2.3 Hashing with linear probing

Hashing with linear probing is a classical model in theoretical computer science that appears in the 60's. It has been studied from a mathematical point of view firstly in [16] and then by several authors. For more details on the model, we refer to [8, 12, 14, 2, 22]. 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 clockwise. A ball that lands in an occupied urn is moved to the next empty urn, always moving clockwise. 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 denoted $d_{m,n}$. We assume $n < m$.

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 is moved one step ahead and lands in 10^{th} urn so that $d_4 = 1$. The ball number 5 should land in the 9^{th} urn, which is occupied like the 10^{th} and the first one, so that $d_5 = 3$. And so on: $d_6 = 1$, $d_7 = 1$, $d_8 = 0$. Here, the total displacement is equal to $1+3+1+1 = 6$. 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] prove that the lengths of the blocks (counting the empty urn) and the sums of displacements inside each block are distributed as $(X_1, Y_1), \ldots, (X_N, Y_N)$ 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 arbitrary parameter $\lambda \in \left(0, e^{-1}\right)$ and Y given $X = l$ being distributed as $d_{l,l-1}$. For the ease of computation, we use the parametrization $\lambda = e^{-\mu}\mu$ to get an equivalent definition of the Borel distribution

$$\mathbb{P}(X = l) = e^{-\mu l} \frac{(\mu l)^{l-1}}{l!}, \ \mu \in (0, 1)$$

(see section 5.3 for more details on Borel distribution and references therein). Notice that the conditional distribution of Y given X does not depend on the parameter μ.

Hence, $d_{m,n}$ is distributed as $\sum_{i=1}^{N} Y_i$ conditioned on $\sum_{i=1}^{N} X_i = m$. The following lemma states basic results on the total displacement $d_{m,n}$ that will be useful in the proofs.

Lemma 2.1.

1. *The number of hash sequences of length m and size n is m^n.*
2. *One has $0 \leqslant d_{m,n} \leqslant \frac{n(n-1)}{2}$.*
3. *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 $(h_{\sigma(1)}, \dots, h_{\sigma(n)})$.*

The first two points are obvious and the last one is a consequence of [12, Lemma 2.1].
From now on, we assume that $m = m_n \to \infty$ and $N = N_n = m_n - n \to \infty$ with $\mu_n := n/m_n \in (0,1) \to \mu \in (0,1)$, as in Subsection 2.2.1. Let $(X_i^{(n)}, Y_i^{(n)})_{1 \leqslant i \leqslant N_n}$ be i.i.d. copies of $(X^{(n)}, Y^{(n)})$, $X^{(n)}$ following Borel distribution with parameter μ_n, and $Y^{(n)}$ given $X^{(n)} = l$ being distributed as $d_{l,l-1}$. The total displacement $d_{m_n,n}$ is distributed as the conditional distribution of $T_{N_n}^{(n)}$ given $S_{N_n}^{(n)} = m_n$.

Remark 2.2. The local limit theorem stated in Proposition 5.2 is crucial in the proofs of the large deviations result (Theorem 3.1) and the one of the Berry-Esseen bound (Theorem 4.1) and requires

$$m_n = N_n \mathbb{E}[X^{(n)}] + O(\sqrt{N_n}).$$

If one takes $\mu_n = \mu$ (i.e. $X^{(n)}$ and $Y^{(n)}$ do not depend on n), the convergence $\mu_n \to \mu$ only gives

$$m_n = N_n \left(\frac{1}{1-\mu} + o(1) \right) = N_n \mathbb{E}[X^{(n)}] + o(N_n).$$

So triangular arrays are needed. Therefore, one may choose $\mu_n = n/m_n$, so that $m_n = N_n \mathbb{E}[X^{(n)}]$.
Also notice that, in the proof of the lower bound in Theorem 3.1, one has to establish

$$\mathbb{P}(S_{N_n}^{(n)} = m_n') \geqslant \frac{c}{2\pi \sigma_{X^{(n)}} N_n^{1/2}}$$

with $m_n' \neq m_n$ in Proposition 5.2.

3 Large deviations result for hashing with linear probing

In [9], the authors prove a classical large deviation principle for the conditional distribution $\mathcal{L}_n$ which applies to Subsections 2.2.1 to 2.2.4. 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 with linear probing, 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, [9, Theorem 2.1] does not apply. Consequently, one needs a specific result in the case when the Laplace transform is not defined. Working in a general framework appears to be difficult. Nevertheless, in the particular case of hashing with linear probing, we establish the following theorem.

Theorem 3.1 (Large deviations result for hashing with linear probing)**.** *If $n/m_n \to \mu \in (0,1)$, there exists $0 < \alpha(\mu) \leqslant \beta(\mu)$ such that, for all $y > 0$,*

$$-\beta(\mu)\sqrt{y} \leqslant \liminf_{n\to\infty} \frac{1}{\sqrt{n}} \log \mathbb{P}(d_{m_n,n} - \mathbb{E}[d_{m_n,n}] \geqslant ny)$$
$$\leqslant \limsup_{n\to\infty} \frac{1}{\sqrt{n}} \log \mathbb{P}(d_{m_n,n} - \mathbb{E}[d_{m_n,n}] \geqslant ny) \leqslant -\alpha(\mu)\sqrt{y}.$$

Remark 3.2. In the proof, we exhibit

$$\alpha(\mu) = (1 + \log(\mu) - \mu)\sqrt{2} \quad \text{and} \quad \beta(\mu) = 4 + \log(2) + 2\log(\mu) - 2\mu.$$

It is still an open question whether we can take $\alpha(\mu) = \beta(\mu)$.

Since $N_n/n \to (1-\mu)/\mu$, the theorem can equivalently be stated as follows:

$$\begin{aligned}-\beta(\mu)\sqrt{y} &\leqslant \liminf_{n\to\infty} \frac{1}{\sqrt{N_n}} \log \mathbb{P}(T_{N_n}^{(n)} - \mathbb{E}[T_{N_n}^{(n)} | S_{N_n}^{(n)} = m_n] \geqslant N_n y | S_{N_n}^{(n)} = m_n) \\ &\leqslant \limsup_{n\to\infty} \frac{1}{\sqrt{N_n}} \log \mathbb{P}(T_{N_n}^{(n)} - \mathbb{E}[T_{N_n}^{(n)} | S_{N_n}^{(n)} = m_n] \geqslant N_n y | S_{N_n}^{(n)} = m_n) \leqslant -\alpha(\mu)\sqrt{y}.\end{aligned}$$

We will prove the result in the latter form.
The following proposition is a non conditioned version of Theorem 3.1 in a general framework. In fact, it is a generalization to triangular arrays of [23, Theorem 3]. For the sake of simplicity, we focus on rough large deviations results instead of precise ones.

Proposition 3.3. *For all $n \geqslant 1$, let $Y^{(n)}$ be a real-valued random variable, N_n be an integer, $\big(Y_i^{(n)}\big)_{1\leqslant i\leqslant N_n}$ be i.i.d. copies of $Y^{(n)}$, and z_n be a positive number. Suppose that $N_n \to \infty$ and that:*

(H3.3.1) $\liminf z_n/N_n > 0$;

(H3.3.2) $\mathrm{Var}(Y^{(n)}) = o\left(N_n^{1/2}\right)$;

(H3.3.3) *the right tail of $Y^{(n)}$ satisfies: there exist $0 < \alpha \leqslant \beta$ such that*

$$\liminf_{n\to\infty} \frac{1}{\sqrt{z_n}} \log \mathbb{P}(Y^{(n)} \geqslant z_n) \geqslant -\beta \tag{2}$$

and

$$\limsup_{n\to\infty} \sup_{u\geqslant\sqrt{z_n}} \frac{1}{\sqrt{u}} \log \mathbb{P}(Y^{(n)} \geqslant u) \leqslant -\alpha. \tag{3}$$

Then,

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

Proposition 3.4. *Let $Y^{(n)}$ be the random variable appearing in the context of hashing with linear probing. Then,*

$$-\beta(\mu) \leqslant \liminf_{n\to\infty} \frac{1}{\sqrt{N_n y}} \log \mathbb{P}(Y^{(n)} \geqslant N_n y) \tag{4}$$

$$\leqslant \limsup_{n\to\infty} \sup_{u\geqslant\sqrt{N_n y}} \frac{1}{\sqrt{u}} \log \mathbb{P}(Y^{(n)} \geqslant u) \leqslant -\alpha(\mu) \tag{5}$$

with

$$\alpha(\mu) = (1 + \log(\mu) - \mu)\sqrt{2} \quad \text{and} \quad \beta(\mu) = 4 + \log(2) + 2\log(\mu) - 2\mu.$$

4 Conditional Berry-Esseen bound

We come back to the general framework of Subsection 2.1. Let also U_n be a random variable distributed as $T_{N_n}^{(n)}$ conditioned on $S_{N_n}^{(n)} = m_n$.

Theorem 4.1. *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:*

(H4.1.1) $\tilde{c}_1 \leqslant \sigma_{X^{(n)}} := \operatorname{Var}\left(X^{(n)}\right)^{1/2} \leqslant c_1$;

(H4.1.2) $\rho_{X^{(n)}} := \mathbb{E}\left[\left|X^{(n)} - \mathbb{E}\left[X^{(n)}\right]\right|^3\right] \leqslant c_2^3 \sigma_{X^{(n)}}^3$;

(H4.1.3) *for* $Y'^{(n)} := Y^{(n)} - X^{(n)} \operatorname{Cov}(X^{(n)}, Y^{(n)})/\sigma_{X^{(n)}}^2$, *there exists* $\eta_0 > 0$ *such that, for all* $s \in [-\pi, \pi]$ *and* $t \in [0, \eta_0]$,

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

(H4.1.4) $m_n = N_n \mathbb{E}\left[X^{(n)}\right] + O(\sigma_{X^{(n)}} N_n^{1/2})$;

(H4.1.5) $\tilde{c}_3 \leqslant \sigma_{Y^{(n)}} := \operatorname{Var}\left(Y^{(n)}\right)^{1/2} \leqslant c_3$;

(H4.1.6) $\rho_{Y^{(n)}} := \mathbb{E}\left[\left|Y^{(n)} - \mathbb{E}\left[Y^{(n)}\right]\right|^3\right] \leqslant c_4^3 \sigma_{Y^{(n)}}^3$;

(H4.1.7) *the correlation* $r_n := \operatorname{Cov}\left(X^{(n)}, Y^{(n)}\right) \sigma_{X^{(n)}}^{-1} \sigma_{Y^{(n)}}^{-1}$ *satisfies* $|r_n| \leqslant c_6 < 1$, *so that*

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

Then the following conclusions hold.

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

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

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

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

on $\{S_{N_n}^{(n)} = m_n\}$ *satisfies the Berry-Esseen inequality*

$$\sup_x \left|\mathbb{P}\left(\frac{U_n - N_n \mathbb{E}\left[Y^{(n)}\right] - r_n \sigma_{Y^{(n)}} \sigma_{X^{(n)}}^{-1} (m_n - N_n \mathbb{E}\left[X^{(n)}\right])}{N_n^{1/2} \tau_n} \leqslant x\right) - \Phi(x)\right| \leqslant \frac{C}{N_n^{1/2}}, \tag{6}$$

where Φ *denotes the standard normal cumulative distribution function 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.

4.1.c. Moreover, there exists 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*

$$\left|\mathbb{E}\left[U_n\right] - N_n \mathbb{E}[Y^{(n)}] - r_n \frac{\sigma_{Y^{(n)}}}{\sigma_{X^{(n)}}} (m_n - N_n \mathbb{E}[X^{(n)}])\right| \leqslant c_7 \tag{7}$$

and

$$\left|\operatorname{Var}\left(U_n\right) - N_n \tau_n^2\right| \leqslant c_8 N_n^{1/2}. \tag{8}$$

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

$$\sup_x \left| \mathbb{P}\left(\frac{U_n - \mathbb{E}[U_n]}{\operatorname{Var}(U_n)^{1/2}} \leqslant x \right) - \Phi(x) \right| \leqslant \frac{\tilde{C}}{N_n^{1/2}}, \tag{9}$$

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*.*

Remark 4.2.

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 4.1 implies the one of the central limit theorem established in [13, Theorem 2.1] which is clearly not surprising.

3. By Hypothesis (H4.1.4), the conditioning value is approximately equal to the mean as in the central limit theorem given in [13, Theorem 2.3].

4. As a consequence of Lemma 5.5 below, $\tilde{c}_1$ can be chosen as $c_2^{-3}/4$.

5. Hypothesis (H4.1.7) is not very restricting and holds in the examples provided in Subsection 2.2.

6. One should note that 4.1.a is the analog of Equation (3.7) in [9, Lemma 3.2].

7. Following [13], we introduce $Y^{'(n)}$ in order to work with a centered variable which is also uncorrelated with $X^{(n)}$.

8. If (X, Y') is a pair of random variables such that 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 (H4.1.3) is reasonable for i.i.d. sequences.

As 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 an i.i.d. sequence instead of a triangular array. This is a consequence of the following corollary.

Corollary 4.3. *Assume that* $\left(X^{(n)}, Y^{(n)}\right) \overset{(d)}{\to} (X, Y)$ *as* $n \to \infty$ *and that, for every fixed* $r > 0$*,*

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

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*. Let* m_n *and* N_n *be integers such that* $\mathbb{E}\left[X^{(n)}\right] = m_n/N_n$ *and* $N_n \to \infty$*. Then, all hypotheses of Theorem 4.1 are satisfied and Theorem 4.1 holds.*

Each example presented in Subsection 2.2, including hashing with linear probing, satisfies the assumptions of Corollary 4.3, as shown in [13], leading to a Berry-Esseen bound for all of them.

5 Proofs

5.1 Technical results

The proofs of Theorems 3.1 and 4.1 intensively rely on the use of Fourier transforms. Define φ_n and ψ_n by

$$\varphi_n(s,t) := \mathbb{E}\left[\exp\left\{is\left(X^{(n)} - \mathbb{E}\left[X^{(n)}\right]\right) + it\left(Y^{(n)} - \mathbb{E}\left[Y^{(n)}\right]\right)\right\}\right] \tag{10}$$

$$\text{and} \quad \psi_n(t) := 2\pi\mathbb{P}(S_{N_n}^{(n)} = m_n)\mathbb{E}\left[\exp\left\{it\left(U_n - N_n\mathbb{E}\left[Y^{(n)}\right]\right)\right\}\right]. \tag{11}$$

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

Lemma 5.1. *One has*

$$\psi_n(t) = \frac{1}{\sigma_{X^{(n)}}N_n^{1/2}}\int_{-\pi\sigma_{X^{(n)}}N_n^{1/2}}^{\pi\sigma_{X^{(n)}}N_n^{1/2}} e^{-is\sigma_{X^{(n)}}^{-1}N_n^{-1/2}\left(m_n - N_n\mathbb{E}[X^{(n)}]\right)}\varphi_n^{N_n}\left(\frac{s}{\sigma_{X^{(n)}}N_n^{1/2}}, t\right)ds. \tag{12}$$

Proof. Since

$$\int_{-\pi}^{\pi} e^{is(S_{N_n}^{(n)} - m_n)}ds = 2\pi 1_{\{S_{N_n}^{(n)} = m_n\}},$$

we have

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

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

Now we establish the local limit theorem (LLT) which is crucial both in the proofs of Theorem 3.1 and Theorem 4.1.

Proposition 5.2 (LLT)**.** *We assume*

1. $\rho_{X^{(n)}} := \mathbb{E}\left[\left|X^{(n)} - \mathbb{E}\left[X^{(n)}\right]\right|^3\right] \leqslant c_2^3\sigma_{X^{(n)}}^3$*;*
2. *for* $Y'^{(n)} := Y^{(n)} - X^{(n)}\operatorname{Cov}(X^{(n)}, Y^{(n)})/\sigma_{X^{(n)}}^2$*, there exists* $\eta_0 > 0$ *such that, for all* $s \in [-\pi, \pi]$ *and* $t \in [0, \eta_0]$*,*
$$\left|\mathbb{E}\left[e^{i(sX^{(n)} + tY'^{(n)})}\right]\right| \leqslant 1 - c_5(\sigma_{X^{(n)}}^2 s^2 + \sigma_{Y'^{(n)}}^2 t^2);$$
3. $m_n = N_n\mathbb{E}\left[X^{(n)}\right] + O(\sigma_{X^{(n)}}N_n^{1/2})$ *(remind that* $m_n \in \mathbb{Z}$ *and* $\mathbb{P}(S_{N_n}^{(n)} = m_n) > 0$*);*

Then there exists $c > 0$ *such that*

$$\mathbb{P}(S_{N_n}^{(n)} = m_n) \geqslant \frac{c}{2\pi\sigma_{X^{(n)}}N_n^{1/2}}.$$

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

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

where $v_n = \frac{m_n - N_n\mathbb{E}\left[X^{(n)}\right]}{\sigma_{X^{(n)}} N_n^{1/2}}$, by Lemma 5.1. 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 assumption 3. and $\mathbb{P}(S^{(n)}_{N_n} = m_n) > 0$ for all n. Inequality (13) with $l = 0$ and $t = 0$ 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's theorem, one gets

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

where the last equality follows from assumption 1. 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 Lemma 5.5),

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

□

Now we give controls on the function φ_n and its second partial derivative.

Lemma 5.3. *Under Hypothesis (H4.1.3), for any integer $l \geqslant 0$, $|s| \leqslant \pi\sigma_{X^{(n)}} N_n^{1/2}$, and $|t| \leqslant \eta_0 \sigma_{Y^{(n)}} N_n^{1/2}$,*

$$\left|\varphi_n^{N_n - l}\left(\frac{s}{\sigma_{X^{(n)}} N_n^{1/2}}, \frac{t}{\sigma_{Y^{(n)}} N_n^{1/2}}\right)\right| \leqslant e^{-(s^2+t^2)\cdot c_5 \cdot (N_n - l)/N_n}. \tag{13}$$

Proof. The proof is a mere consequence of the inequality $1 + x \leqslant e^x$ that holds for any $x \in \mathbb{R}$. □

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

Lemma 5.4. *For any s and t, one has*

$$\left|\frac{\partial \varphi_n}{\partial t}\left(\frac{s}{\sigma_{X^{(n)}} N_n^{1/2}}, \frac{t}{\sigma_{Y^{(n)}} N_n^{1/2}}\right)\right| \leqslant \frac{\sigma_{Y^{(n)}}}{N_n^{1/2}}(|s| + |t|); \tag{14}$$

and

$$\left|\frac{\partial \varphi_n}{\partial t}\left(\frac{s}{\sigma_{X^{(n)}} N_n^{1/2}}, \frac{t}{\sigma_{Y^{(n)}} N_n^{1/2}}\right)\right| \leqslant \frac{\sigma_{Y^{(n)}}}{N_n^{1/2}}(|s|r_n + |t|) + \frac{\sigma_{Y^{(n)}}}{N_n}\left[\frac{s^2}{2}\left(\frac{\rho_{X^{(n)}}}{\sigma^3_{X^{(n)}}}\right)^{2/3}\left(\frac{\rho_{Y^{(n)}}}{\sigma^3_{Y^{(n)}}}\right)^{1/3} + |st|\left(\frac{\rho_{X^{(n)}}}{\sigma^3_{X^{(n)}}}\right)^{1/3}\left(\frac{\rho_{Y^{(n)}}}{\sigma^3_{Y^{(n)}}}\right)^{2/3} + \frac{t^2}{2}\left(\frac{\rho_{Y^{(n)}}}{\sigma^3_{Y^{(n)}}}\right)\right]. \tag{15}$$

Proof. We apply Taylor's 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 (14) 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 (15) using

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

□

Lemma 5.5. *Under Hypothesis (H4.1.2), one has* $\sigma_{X^{(n)}} \geqslant (4c_2^3)^{-1}$.

Proof. The proof relies 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 (H4.1.2). □

5.2 Proof of Proposition 3.3

Since $Y^{(n)} - \mathbb{E}\left[Y^{(n)}\right]$ also satisfies the hypotheses, we can assume that $\mathbb{E}\left[Y^{(n)}\right] = 0$. Write

$$\begin{aligned}\mathbb{P}(T_{N_n}^{(n)} \geqslant z_n) &= \mathbb{P}(T_{N_n}^{(n)} \geqslant z_n, \quad \forall i\ Y_i^{(n)} < z_n) + \mathbb{P}(T_{N_n}^{(n)} \geqslant z_n, \quad \exists i\ Y_i^{(n)} \geqslant z_n) \\ &:= P_{N_n,0} + P_{N_n,1}.\end{aligned}$$

If we prove that

$$-\beta \leqslant \liminf_{n\to\infty} \frac{1}{\sqrt{z_n}} \log(P_{N_n,1}) \leqslant \limsup_{n\to\infty} \frac{1}{\sqrt{z_n}} \log(P_{N_n,1}) \leqslant -\alpha \tag{16}$$

and

$$\limsup_{n\to\infty} \frac{1}{\sqrt{z_n}} \log(P_{N_n,0}) \leqslant -\alpha, \tag{17}$$

then,

$$\begin{aligned}-\beta \leqslant \liminf_{n\to\infty} \frac{1}{\sqrt{z_n}} \log(P_{N_n,1}) &\leqslant \liminf_{n\to\infty} \frac{1}{\sqrt{z_n}} \log \mathbb{P}(T_{N_n}^{(n)} \geqslant z_n) \\ &\leqslant \limsup_{n\to\infty} \frac{1}{\sqrt{z_n}} \log \mathbb{P}(T_{N_n}^{(n)} \geqslant z_n) \leqslant -\alpha\end{aligned}$$

which establishes Proposition 3.3.

Proof of (16). First, using (3),

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

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

$$\begin{aligned} P_{N_n,1} &\geqslant \mathbb{P}(T_{N_n}^{(n)} \geqslant z_n, \quad Y_1^{(n)} \geqslant z_n) \\ &\geqslant \mathbb{P}(T_{N_n-1}^{(n)} \geqslant -N_n\varepsilon)\mathbb{P}(Y^{(n)} \geqslant z_n + N_n\varepsilon). \end{aligned}$$

By Chebyshev's inequality and Hypothesis (H3.3.2),

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

the random variables $Y^{(n)}$ being assumed centered. Finally, using (2) and (H3.3.1), and noting $\delta := \liminf_{n\to\infty} \frac{z_n}{N_n}$, one gets

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

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

Proof of 17. Let $\alpha' \in (0, \alpha)$ and $s_n = \alpha'/\sqrt{z_n}$. The exponential Chebyshev's inequality for $T_{N_n}^{(n)}$ conditioned on $\{\forall i,\ Y_i^{(n)} < z_n\}$ yields

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

If we prove that

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

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

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

By a Taylor expansion of $f(t) = e^t$, (H3.3.2) and (H3.3.1), there exists

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

such that

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

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, $I_2 = o(N_n^{-1/2})$.
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 $t \in [0,1]$,

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

Finally, applying Taylor's 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 $I_3 = o(N_n^{-1/2})$. □

5.3 Proof of Proposition 3.4

Remind that $(X_i^{(n)}, Y_i^{(n)})_{1 \leqslant i \leqslant N_n}$ are i.i.d. copies of $(X^{(n)}, Y^{(n)})$, $X^{(n)}$ following the Borel distribution with parameter $\mu_n = n/m_n \to \mu \in (0,1)$, and $Y^{(n)}$ given $X^{(n)} = l$ is distributed as $d_{l,l-1}$. We start with computing the asymptotic tail behavior of $X^{(n)}$. Remind that

$$\mathbb{P}(X^{(n)} = x_n) = e^{-\mu_n x_n}\frac{(\mu_n x_n)^{x_n - 1}}{x_n!}.$$

Lemma 5.6 (Tail of $X^{(n)}$)**.** *If $l_n \to \infty$, then*

$$\log \mathbb{P}(X^{(n)} = l_n) = -\kappa l_n(1 + o(1)) \tag{18}$$

and

$$\log \mathbb{P}(X^{(n)} \geqslant l_n) = -\kappa l_n(1 + o(1)) \tag{19}$$

with $\kappa = \mu - \log(\mu) - 1 \in (0, \infty)$.

Proof. By Stirling's formula,

$$\log \mathbb{P}\big(X^{(n)} = l_n\big) = \log\big(e^{-\mu_n l_n} \frac{(\mu_n l_n)^{l_n - 1}}{l_n!}\big) \sim l_n(1 + \log(\mu) - \mu).$$

Similar estimates give the second result. □

Proof of (5). Let $u > 0$ and n_u be the ceiling of the positive solution of $2u = (n-1)(n-2)$:

$$n_u = \left\lceil \sqrt{2u + \frac{1}{4}} + \frac{3}{2} \right\rceil. \tag{20}$$

Since $Y^{(n)}$ conditioned on $\{X^{(n)} = l\}$ is distributed as $d_{l,l-1}$, we get

$$\mathbb{P}(Y^{(n)} \geqslant u) = \sum_{l=n_u}^{+\infty} \mathbb{P}(d_{l,l-1} \geqslant u)\mathbb{P}(X^{(n)} = l) \leqslant \sum_{l=n_u}^{+\infty} \mathbb{P}(X^{(n)} = l) = \mathbb{P}(X^{(n)} \geqslant n_u).$$

By (19) and the fact that $n_u = \sqrt{2u}(1 + o(1))$ for $u \geqslant \sqrt{N_n y}$, we finally conclude that

$$\limsup_{n\to\infty} \sup_{u \geqslant \sqrt{N_n y}} \frac{1}{\sqrt{u}} \log \mathbb{P}(Y^{(n)} \geqslant u) \leqslant -\kappa\sqrt{2}.$$

□

Lemma 5.7. *Let $a > 0$. Let $l = 1 + \lceil \sqrt{a} \rceil$ and $k = \lfloor \sqrt{a} \rfloor$. Then*

$$\mathbb{P}(d_{l,l-1} \geqslant a) \geqslant \frac{1}{l^{l-1}} \frac{(l-1)!}{2^k}.$$

Proof. Take the hash sequence

$$(1,\, 1,\, 2,\, 2,\, \ldots\, k,\, k,\, k+1,\, k+2,\, \ldots,\, l-1-k)\,. \tag{21}$$

Notice that $0 \leqslant k \leqslant (l-1)/2$. On the one hand, it is decomposed into $l-1-2k$ single numbers and k pairs leading to a hash sequence of size $l-1$ 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 ($q = k+1, \ldots, l-1-k$) realizes a displacement equal to k. The total displacement is then $k(l-1-k)$ which is greater than a.
Moreover as mentioned in Lemma 2.1 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 (21) the total number of which is given by

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

To conclude, it remains to use item 1. of Lemma 2.1. □

Proof of (4). For any $l_n \geqslant 1$, one has

$$\begin{aligned}\mathbb{P}(Y^{(n)} \geqslant N_n y) &= \sum_{l=1}^{+\infty} \mathbb{P}(d_{l,l-1} \geqslant N_n y)\mathbb{P}(X^{(n)} = l)\\ &\geqslant \mathbb{P}(d_{l_n,l_n-1} \geqslant N_n y)\mathbb{P}(X^{(n)} = l_n).\end{aligned}$$

As a consequence, using Lemma 5.7 with $a := N_n y$ and Lemma 5.6,

$$\begin{aligned}\liminf_{n\to\infty} \frac{1}{\sqrt{N_n y}} \log \mathbb{P}(Y^{(n)} \geqslant N_n y) &\geqslant \liminf_{n\to\infty} \frac{1}{\sqrt{N_n y}} \log\left(\frac{1}{l_n^{l_n-1}} \frac{(l_n-1)!}{2^{k_n}} \mathbb{P}(X^{(n)} = l_n)\right)\\ &\geqslant \liminf_{n\to\infty} -\frac{l_n + k_n \log(2) + \kappa l_n}{\sqrt{N_n y}}\\ &= -(4 + \log(2) + 2\log(\mu) - 2\mu)\end{aligned}$$

where $l_n = 1 + \lceil \sqrt{a} \rceil$ and $k_n = \lfloor \sqrt{a} \rfloor$. □

5.4 Proof of Theorem 3.1

Remind that the total displacement $d_{m_n,n}$ is distributed as the conditional distribution of $T_{N_n}^{(n)}$ given $S_{N_n}^{(n)} = m_n$. Notice that $\mathbb{E}[S_{N_n}^{(n)}] = N_n \mathbb{E}[X^{(n)}] = m_n$. Now let

$$\begin{aligned}P_n &:= \mathbb{P}(d_{m_n,n} - \mathbb{E}[d_{m_n,n}] \geqslant N_n y)\\ &= \mathbb{P}(T_{N_n}^{(n)} - \mathbb{E}[T_{N_n}^{(n)} | S_{N_n}^{(n)} = m_n] \geqslant N_n y | S_{N_n}^{(n)} = m_n)\\ &= \mathbb{P}(T_{N_n}^{(n)} - \mathbb{E}[T_{N_n}^{(n)}] \geqslant N_n y_n, S_{N_n}^{(n)} = m_n)/\mathbb{P}(S_{N_n}^{(n)} = m_n)\end{aligned}$$

where $y_n := y + \frac{1}{N_n}(\mathbb{E}[T_{N_n}^{(n)} | S_{N_n}^{(n)} = m_n] - \mathbb{E}[T_{N_n}^{(n)}])$. The following lemma entails $y_n \to y$.

Lemma 5.8.

$$\mathbb{E}\left[T_{N_n}^{(n)} \middle| S_{N_n}^{(n)} = m_n\right] = \mathbb{E}\left[T_{N_n}^{(n)}\right] + o(N_n).$$

Proof. According to [12, Section 4], the hypotheses of Proposition 5.2 are satisfied by the variables $(X^{(n)}, Y^{(n)})$. Using (11), differentiating under the integral sign of (12) and using Proposition 5.2 yield

$$\begin{aligned}&\left|\mathbb{E}\left[T_{N_n}^{(n)} - N_n \mathbb{E}\left[Y^{(n)}\right] \middle| S_{N_n}^{(n)} = m_n\right]\right| = \left|\frac{-i\psi_n'(0)}{2\pi \mathbb{P}(S_{N_n}^{(n)} = m_n)}\right|\\ &\leqslant \frac{N_n}{2\pi c} \int_{-\pi\sigma_{X^{(n)}} N_n^{1/2}}^{\pi\sigma_{X^{(n)}} N_n^{1/2}} \left|\frac{\partial \varphi_n}{\partial t}\left(\frac{s}{\sigma_{X^{(n)}} N_n^{1/2}}, 0\right)\right| \cdot \left|\varphi_n^{N_n-1}\left(\frac{s}{\sigma_{X^{(n)}} N_n^{1/2}}, 0\right)\right| ds. \qquad (22)\end{aligned}$$

It remains to show that the integral converges to 0. Putting together (22) and (15), using the fact that $\mathrm{Var}(Y^{(n)})$ is convergent and the control (13) with $l = 1$ and $t = 0$, one gets

$$\mathbb{E}\left[T_{N_n}^{(n)} - N_n \mathbb{E}\left[Y^{(n)}\right] \middle| S_{N_n}^{(n)} = m_n\right] = o(N_n).$$

□

Remember that the assumptions of Theorem 4.1 are satisfied. By Lemma 5.8 (respectively Hypothesis (H4.1.5) and Proposition 3.4), Hypothesis (H3.3.1) with $z_n = N_n y_n$ (resp. Hypotheses (H3.3.2) and (H3.3.3)) holds.

Proof of the upper bound. We have

$$P_n \leqslant \frac{\mathbb{P}(T^{(n)}_{N_n} - \mathbb{E}[T^{(n)}_{N_n}] \geqslant N_n y_n)}{\mathbb{P}(S^{(n)}_{N_n} = m_n)}.$$

The conclusion follows from the upper bound of Proposition 3.3, Proposition 5.2. □

Proof ot the lower bound. We have

$$\begin{aligned}
P_n &\geqslant \mathbb{P}(T^{(n)}_{N_n} - \mathbb{E}[T^{(n)}_{N_n}] \geqslant N_n y_n, \quad S^{(n)}_{N_n} = m_n) \\
&\geqslant \mathbb{P}\big(T^{(n)}_{N_n} - \mathbb{E}[T^{(n)}_{N_n}] \geqslant N_n y_n, \quad S^{(n)}_{N_n} = m_n, \quad Y^{(n)}_{N_n} - \mathbb{E}[Y^{(n)}_{N_n}] \geqslant N_n(y_n+\varepsilon)\big) \\
&\geqslant \mathbb{P}\big(T^{(n)}_{N_n-1} - \mathbb{E}[T^{(n)}_{N_n-1}] \geqslant -N_n\varepsilon, \quad S^{(n)}_{N_n-1} = m_n - l_n\big)\mathbb{P}\big(Y^{(n)}_{N_n} - \mathbb{E}[Y^{(n)}_{N_n}] \geqslant N_n(y_n+\varepsilon), \quad X^{(n)}_{N_n} = l_n\big) \\
&=: P_1 P_2
\end{aligned}$$

where $l_n := 1 + \lceil \sqrt{a}_n \rceil$ and $a_n := N_n(y_n+\varepsilon) + \mathbb{E}[Y^{(n)}]$. Applying Lemma 5.7 and Lemma 5.6, we get

$$\begin{aligned}
\liminf_{n\to\infty} \frac{1}{\sqrt{N_n}} \log(P_2) &= \liminf_{n\to\infty} \frac{1}{\sqrt{N_n}} \log\Big(\mathbb{P}\big(d_{l_n,l_n-1} \geqslant N_n(y_n+\varepsilon) + \mathbb{E}[Y^{(n)}]\big)\mathbb{P}\big(X^{(n)} = l_n\big)\Big) \\
&\geqslant \liminf_{n\to\infty} \frac{1}{\sqrt{N_n}} \log\left(\frac{1}{l_n^{l_n-1}} \frac{(l_n-1)!}{2^{k_n}} \mathbb{P}\big(X^{(n)} = l_n\big)\right) \\
&\geqslant \liminf_{n\to\infty} -\frac{l_n + k_n\log(2) + (1+\log(\mu)-\mu)l_n}{\sqrt{N_n}} \\
&= -(4+\log(2)+2\log(\mu)-2\mu)\sqrt{y+\varepsilon}.
\end{aligned}$$

Letting $\varepsilon \to 0$, we get

$$\liminf_{n\to\infty} \frac{1}{\sqrt{N_n}} \log(P_2) \geqslant -\beta(\mu)\sqrt{y} \tag{23}$$

with $\beta(\mu) := 4 + \log(2) + 2\log(\mu) - 2\mu$.
Let us turn to the minoration of P_1:

$$\begin{aligned}
P_1 &= \mathbb{P}\big(T^{(n)}_{N_n-1} - \mathbb{E}[T^{(n)}_{N_n-1}] \geqslant -N_n\varepsilon, \quad S^{(n)}_{N_n-1} = m_n - l_n\big) \\
&\geqslant \mathbb{P}\big(S^{(n)}_{N_n-1} = m_n - l_n\big) - \mathbb{P}\big(T^{(n)}_{N_n-1} - \mathbb{E}[T^{(n)}_{N_n-1}] < -N_n\varepsilon\big).
\end{aligned}$$

Since $l_n = O(N_n^{1/2})$, Proposition 5.2 provides a $c > 0$ such that

$$\mathbb{P}\big(S^{(n)}_{N_n-1} = m_n - l_n\big) \geqslant \frac{c}{2\pi\sigma_{X^{(n)}} N_n^{1/2}} \geqslant c' N_n^{-1/2}$$

with $c' > 0$, since $\sigma_{X^{(n)}}$ converges to the standard deviation of the Borel distribution of parameter μ. Chebyshev's inequality and the fact that $\mathrm{Var}(Y^{(n)}) = o(N_n^{1/2})$ yield

$$\mathbb{P}\big(T^{(n)}_{N_n-1} - \mathbb{E}[T^{(n)}_{N_n-1}] < -N_n\varepsilon\big) \leqslant \frac{\mathrm{Var}(Y^{(n)})}{N_n\varepsilon^2} = o(N_n^{-1/2}).$$

Eventually, $\lim_{n\to\infty} \frac{1}{\sqrt{N_n}} \log(P_1) = 0$ that leads with (23) to

$$\liminf_{n\to\infty} \frac{1}{\sqrt{N_n}} \log(P_n) \geqslant -\beta(\mu) y^{1/2}.$$

□

5.5 Proof of Theorem 4.1

To lighten notation, we denote $S_n := S_{N_n}^{(n)}$ and $T_n := T_{N_n}^{(n)}$. Remind that U_n is distributed as T_n conditioned on S_n. Part a) is Proposition 5.2 with $\tilde{c}_5 = c$. 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

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

Then $\mathbb{E}[Y^{'(n)}] = 0$ and $\mathrm{Cov}(X^{(n)}, Y^{'(n)}) = \mathbb{E}[X^{(n)} Y^{'(n)}] = 0$. Besides, Hypotheses (H4.1.3) and (H4.1.7) are verified by $Y^{'(n)}$. By Hypothesis (H4.1.7),

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

so (H4.1.5) is satisfied by $Y^{'(n)}$. Finally, by Minkowski Inequality, Hypotheses (H4.1.2) and (H4.1.6), and the fact that $|r_n| \leqslant 1$,

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

Hence $Y^{'(n)}$ satisfies Hypothesis (H4.1.6). 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 $S_n = m_n$, we have $T'_n = T_n - N_n \mathbb{E}\left[Y^{(n)}\right] - r_n \frac{\sigma_{Y^{(n)}}}{\sigma_{X^{(n)}}} (m_n - N_n \mathbb{E}[X^{(n)}])$. Hence the conclusions for $(X^{(n)}, Y^{(n)})$ 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 4.1 - Part b). We follow the classical proof of Berry-Esseen (see e.g. [7]) combined with the procedure of Quine and Robinson [27].
As shown in Loève [21] (page 285) or Feller [7], the left hand side of (6) is dominated by

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

where $\eta > 0$ is such that

$$\eta := \min\left(\frac{2}{9} c_3 c_4^3, \eta_0 \right). \tag{25}$$

From Lemma 5.1 and a Taylor expansion,

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

where $c_n := 2\pi\mathbb{P}(S_n = m_n)\sigma_{X^{(n)}} N_n^{1/2} \geqslant \tilde{c}_5$ and $v_n = \frac{m_n - N_n\mathbb{E}[X^{(n)}]}{\sigma_{X^{(n)}} N_n^{1/2}}$ has already been defined in the proof of Proposition 5.2. 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 $0 < \varepsilon < \pi$ is such that

$$\varepsilon := \min\left(\frac{2}{9}c_1 c_2^3, \pi\right) \tag{26}$$

and decompose

$$u^{-1}\left|\frac{\psi_n(u/\sigma_{Y^{(n)}} N_n^{1/2})}{2\pi\mathbb{P}(S_n = m_n)} - e^{-u^2/2}\right| \leqslant \sup_{0\leqslant\theta\leqslant u}\left[I_1(u,\theta) + I_2(u,\theta)\right], \tag{27}$$

where

$$I_1(u,\theta) = c_n^{-1}\int_{A_1} e^{-(u^2+s^2)/2}\left|\left(\frac{\partial}{\partial t}\left[e^{(t^2+s^2)/2}\varphi_n^{N_n}\left(\frac{s}{\sigma_{X^{(n)}} N_n^{1/2}}, \frac{t}{\sigma_{Y^{(n)}} N_n^{1/2}}\right)\right]\right)_{t=\theta}\right| ds, \tag{28}$$

$$I_2(u,\theta) = c_n^{-1} e^{-u^2/2}\int_{A_2}\left|\left(\frac{\partial}{\partial t}\left[e^{t^2/2}\varphi_n^{N_n}\left(\frac{s}{\sigma_{X^{(n)}} N_n^{1/2}}, \frac{t}{\sigma_{Y^{(n)}} N_n^{1/2}}\right)\right]\right)_{t=\theta}\right| ds. \tag{29}$$

If we prove that there exists positive constants C_1, C_2 and C_3, such that

$$\int_0^{\eta\sigma_{Y^{(n)}} N_n^{1/2}} \sup_{0\leqslant\theta\leqslant u} I_1(u,\theta)du \leqslant \frac{C_1}{N_n^{1/2}} \tag{30}$$

and

$$\int_0^{\eta\sigma_{Y^{(n)}} N_n^{1/2}} \sup_{0\leqslant\theta\leqslant t} I_2(u,\theta)du \leqslant C_2 e^{-C_3 N_n}, \tag{31}$$

we conclude to part b) of Theorem 4.1 writing

$$C_2 e^{-C_3 N_n} = \frac{C_2 C_3^{-1/2}}{N_n^{1/2}}(C_3 N_n)^{1/2} e^{-C_3 N_n} \leqslant \frac{C_2 C_3^{-1/2}}{N_n^{1/2}}(1/2)^{1/2} e^{-1/2},$$

since $x^{1/2}e^{-x}$ is maximum in $1/2$. The proofs of (30) and (31) are postponed after the present proof. 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

$$C := C_1 + C_2 C_3^{-1/2}(1/2)^{1/2}e^{-1/2} + \frac{24}{\tilde{c}_3\pi\sqrt{2\pi}}\left(\min\left(\frac{2}{9}c_3 c_4^3, \eta_0\right)\right)^{-1}. \tag{32}$$

□

Now it remains to prove (30) and (31). To bound $I_1(u,\theta)$, we use a result due to Quine and Robinson ([27, Lemma 2]).

Lemma 5.9. *[Lemma 2 in [27]] Define*

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

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

$$\left|\frac{\partial}{\partial t}\left[e^{(s^2+t^2)/2}\ \ \varphi_n^{N_n}\left(\frac{s}{\sigma_{X^{(n)}}N_n^{1/2}},\frac{t}{\sigma_{Y^{(n)}}N_n^{1/2}}\right)\right]\right|$$
$$\leqslant C_0(|s|+|t|+1)^3(l_{1,n}+l_{2,n})\exp\left\{\frac{11}{24}\left(s^2+t^2\right)\right\} \tag{33}$$

with $C_0 := 98$.

Proof. We refer to the proof in the appendix of [27]. The condition $l_{1,n} < 12^{-3/2}$ and $l_{2,n} < 12^{-3/2}$ appearing in [27, 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

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

□

By Hypotheses (H4.1.2) and (H4.1.1),

$$l_{1,n} \leqslant c_2^3 N_n^{-1/2} \leqslant c_2^3 c_1 \sigma_{X^{(n)}}^{-1} N_n^{-1/2}, \tag{34}$$

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

$$l_{2,n} \leqslant c_4^3 N_n^{-1/2} \leqslant c_4^3 c_3 \sigma_{Y^{(n)}}^{-1} N_n^{-1/2}, \tag{35}$$

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

Lemma 5.10. *There exists a positive constant* C_1 *such that*

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

Proof. Conditions (26) and (25) imply that, on A_1,

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

which ensures that $(s,u) \in R$ as specified in Lemma 5.9. Moreover, since we have $N_n \geqslant \max(c_2^6, c_4^6)$ (cf. Hypothesis in 4.1.b), $l_{1,n} \leqslant 1$ and $l_{2,n} \leqslant 1$. Now applying Lemma 5.9 in (28) and using part 4.1.a, we get

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

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 5.11. *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*

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

Proof. We use the controls (14), (13) with $l=1$, and $|\varphi_n|\leqslant 1$ to get

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

Finally by (29) and for $N_n\geqslant 2$, we conclude that

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

The conclusion follows with

$$C_2 := 2\tilde{c}_5^{-1}c_5^{-1}\left(\frac{\sqrt{2\pi}}{\sqrt{\min(1,c_5)}}+\frac{2}{\min(1,c_5)\min\left(\frac{2}{9}c_1c_2^3,\pi\right)\tilde{c}_1}\right) \tag{36}$$

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

Proof of Theorem 4.1 - Part c). We start proving (7). We adapt the proof given in [13]. Using (11) with $\mathbb{E}[Y^{(n)}]=0$, and differentiating under the integral sign of (12), we naturally have

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

Using inequality (15) of Lemma 5.4 with $r_n = 0$ and $t = 0$, Hypotheses (H4.1.1), (H4.1.2), and (H4.1.6), we deduce

$$\left|\frac{\partial \varphi_n}{\partial t}\left(\frac{s}{\sigma_{X^{(n)}} N_n^{1/2}}, 0\right)\right| \leqslant \frac{s^2}{2}\frac{\rho_{Y^{(n)}}^{1/3}\rho_{X^{(n)}}^{2/3}}{\sigma_{X^{(n)}}^2 N_n} \leqslant \frac{c_2^2 c_3 c_4}{2N_n} s^2.$$

Then using inequality (13) with $l = 1$ and $t = 0$ and for $N_n \geqslant 2$,

$$\int_{-\pi\sigma_{X^{(n)}} N_n^{1/2}}^{\pi\sigma_{X^{(n)}} N_n^{1/2}} \left|\frac{\partial \varphi_n}{\partial t}\left(\frac{s}{\sigma_{X^{(n)}} N_n^{1/2}}, 0\right)\right| \cdot \left|\varphi_n^{N_n - 1}\left(\frac{s}{\sigma_{X^{(n)}} N_n^{1/2}}, 0\right)\right| ds \leqslant \frac{c_2^2 c_3 c_4}{2N_n}\int_{-\infty}^{+\infty} s^2 e^{-c_5 s^2/2} ds.$$

So, (7) holds with $c_7 := \frac{c_2^2 c_3 c_4}{2\tilde{c}_5}\int_{-\infty}^{+\infty} s^2 e^{-c_5 s^2/2} ds$.

To prove (8), since $\tau_n = \sigma_{Y^{(n)}}$ and $\mathbb{E}\left[U_n\right]$ 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,

$$\begin{aligned}
\mathbb{E}\left[U_n^2\right] &= \frac{-\psi_n''(0)}{2\pi\mathbb{P}(S_n = m_n)} \\
&= -c_n^{-1} N_n(N_n - 1)\int_{-\pi\sigma_{X^{(n)}} N_n^{1/2}}^{\pi\sigma_{X^{(n)}} N_n^{1/2}} \left(\frac{\partial \varphi_n}{\partial t}\left(\frac{s}{\sigma_{X^{(n)}} N_n^{1/2}}, 0\right)\right)^2 \varphi_n^{N_n - 2}\left(\frac{s}{\sigma_{X^{(n)}} N_n^{1/2}}, 0\right) ds \qquad (38)\\
&\quad - c_n^{-1} N_n \int_{-\pi\sigma_{X^{(n)}} N_n^{1/2}}^{\pi\sigma_{X^{(n)}} N_n^{1/2}} \frac{\partial^2 \varphi_n}{\partial t^2}\left(\frac{s}{\sigma_{X^{(n)}} N_n^{1/2}}, 0\right) \varphi_n^{N_n - 1}\left(\frac{s}{\sigma_{X^{(n)}} N_n^{1/2}}, 0\right) ds. \qquad (39)
\end{aligned}$$

First, by inequality (15) with $r_n = 0$ and $t = 0$, the control (13) with $l = 1$ and $t = 0$, and for $N_n \geqslant 3$, one has

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

and finally using 4.1.a, the term (38) is bounded by

$$c_8'' := \frac{c_2^4 c_3^2 c_4^2}{4\tilde{c}_5}\int_{-\infty}^{+\infty} s^4 e^{-c_5 s^2/3} ds. \qquad (40)$$

Second, we study the term (39). 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 5.1 and Hypothesis (H4.1.4),

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

so

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

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

Applying Taylor's 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

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

Thus, using Hölder's 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, using part 4.1.a, Hypotheses (H4.1.1), (H4.1.2), (H4.1.5), (H4.1.6), and the majoration (13) with $t=0$, we get

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

with

$$c_8''' := c_3\tilde{c}_5^{-1}(1+c_2c_4^2)\int_{-\infty}^{+\infty}|s|\,e^{-s^2c_5/2}ds. \tag{41}$$

Finally,

$$\left|\mathrm{Var}(U_n) - N_n\tau_n^2\right| \leqslant c_7 + c_8'' + c_8'''N_n^{1/2} \leqslant c_8N_n^{1/2}$$

with

$$c_8 := c_7 + c_8'' + c_8''' = \frac{c_2^2c_3c_4}{2\tilde{c}_5}\int_{-\infty}^{+\infty}s^2e^{-cs^2/2}ds + \frac{c_2^4c_3^2c_4^2}{4\tilde{c}_5}\int_{-\infty}^{+\infty}s^4e^{-c_5s^2/3}ds + c_3\tilde{c}_5^{-1}(1+c_2c_4^2)\int_{-\infty}^{+\infty}|s|\,e^{-s^2c_5/2}ds. \tag{42}$$

Now we turn to the proof of (9). Let us show that the previous estimates of $\mathbb{E}[U_n]$ and $\mathrm{Var}(U_n)$ make it possible to apply (6). 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,

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

For $N_n > 4c_8^2/\tilde{c}_3^2$, $a_n \geqslant 1/2$ and applying Taylor's 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 (9) with

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

□

References

[1] P. Chassaing and S. 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.

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

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

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

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

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

[7] W. 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] F. Gamboa, T. Klein, and C. Prieur. Conditional large and moderate deviations for sums of discrete random variables. Combinatoric applications. *Bernoulli*, 18(4):1341–1360, 2012.

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

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

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

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

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

[16] D. E. Knuth. Computer science and its relation to mathematics. *Amer. Math. Monthly*, 81:323–343, 1974.

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

[18] V. 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.

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

[20] J.-F. Le Gall. Random trees and applications. *Probab. Surv.*, 2:245–311, 2005.

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

[22] J.-F. Marckert. Parking with density. *Random Structures Algorithms*, 18(4):364–380, 2001.

[23] 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.

[24] 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.

[25] Y. 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.

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

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

[28] 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.

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

[30] J. M. van Campenhout and T. M. Cover. Maximum entropy and conditional probability. *IEEE Trans. Inform. Theory*, 27(4):483–489, 1981.

[31] J. G. Wendel. Left-continuous random walk and the Lagrange expansion. *Amer. Math. Monthly*, 82:494–499, 1975.