

# A conditional Berry-Esseen inequality

Thierry Klein\*

Agnès Lagnoux†

Pierre Petit‡

January 8, 2019

## Abstract

As an extension of a central limit theorem established by Svante Janson, we prove a Berry-Esseen inequality for a sum of independent and identically distributed random variables conditioned by a sum of independent and identically distributed integer-valued random variables.

**Keywords:** Berry-Esseen inequality, conditional distribution, combinatorial problems, occupancy, hashing with linear probing, random forests, branching processes, Bose-Einstein statistics.

**1991 Mathematics Subject Classification:** 60F05; 62E20; 60C05.

## 1 Introduction

As pointed out by Svante Janson in his seminal work [8], 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. Here, we are interested in the law of  $N^{-1}(Y_1 + \dots + Y_N)$  conditioned on a specific value of  $X_1 + \dots + X_N$  that is to say in the conditional distribution

$$\mathcal{L}_N := \mathcal{L}(N^{-1}(Y_1 + \dots + Y_N) \mid X_1 + \dots + X_N = m),$$

where  $m$  and  $N$  are integers and the  $(X_i, Y_i)$  for  $1 \leq i \leq N$  are i.i.d. copies of a vector  $(X, Y)$  of random variables with  $X$  integer-valued.

In [8], 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, one natural question arises: is it possible to obtain a general Berry-Esseen inequality for these models?

The first Berry-Esseen inequality for a conditional model is given by Malcolm P. Quine and John Robinson in [17]. They study the particular case of the occupancy problem, i.e., the case when the random variable  $X$  is Poisson distributed and  $Y = \mathbb{1}_{\{X=0\}}$ . Up to our knowledge, it is the only result in that direction for this kind of conditional distribution.

Our paper is organized as follows. In Section 2, we present the model and we state our main results (Theorems 3 and 5). In Section 3, we describe classical examples. The last section is dedicated to the proofs.

## 2 Conditional Berry-Esseen inequality

For all  $n \geq 1$ , we consider a vector of random variables  $(X_n, Y_n)$  such that  $X_n$  is integer-valued and  $Y_n$  real-valued. Let  $N_n$  be a natural number such that  $N_n \rightarrow \infty$  as  $n$  goes to infinity. Let  $(X_{n,i}, Y_{n,i})_{1 \leq i \leq N_n}$  be an i.i.d. sample distributed as  $(X_n, Y_n)$  and define

$$S_{n,k} := \sum_{i=1}^k X_{n,i} \quad \text{and} \quad T_{n,k} := \sum_{i=1}^k Y_{n,i},$$

for  $k \in \llbracket 1, N_n \rrbracket$ . To lighten notation, define  $S_n := S_{n, N_n}$  and  $T_n := T_{n, N_n}$ . Let  $m_n \in \mathbb{Z}$  be such that  $\mathbb{P}(S_n = m_n) > 0$ . The purpose of the paper is to prove a Berry-Esseen inequality for the conditional distributions

$$\mathcal{L}(U_n) := \mathcal{L}(T_n \mid S_n = m_n).$$

---

\*Institut de Mathématiques de Toulouse, UMR5219, Université de Toulouse, CNRS, UPS IMT, F-31062 Toulouse Cedex 9, France, and ENAC - Ecole Nationale de l'Aviation Civile, 7 avenue Edouard Belin, F-31400 Toulouse, France ([thierry.klein@math.univ-toulouse.fr](mailto:thierry.klein@math.univ-toulouse.fr) and [thierry01.klein@enac.fr](mailto:thierry01.klein@enac.fr)).

†Institut de Mathématiques de Toulouse, UMR5219, Université de Toulouse, CNRS, UT2, F-31058 Toulouse, France ([lagnoux@univ-tlse2.fr](mailto:lagnoux@univ-tlse2.fr)).

‡Institut de Mathématiques de Toulouse, UMR5219, Université de Toulouse, CNRS, UPS IMT, F-31062 Toulouse Cedex 9, France ([pierre.petit@math.univ-toulouse.fr](mailto:pierre.petit@math.univ-toulouse.fr)).

**Assumption 1.** Suppose that there exist positive constants  $c_1, \tilde{c}_2, c_2, c_3, \tilde{c}_4, c_4, c_5, c_6, c_7$ , and  $\eta_0$ , such that:

$$(A1) \quad \gamma_n := 2\pi\sigma_{X_n} N_n^{1/2} \mathbb{P}(S_n = m_n) \geq c_1;$$

$$(A2) \quad \tilde{c}_2 \leq \sigma_{X_n} := \text{Var}(X_n)^{1/2} \leq c_2;$$

$$(A3) \quad \rho_{X_n} := \mathbb{E}[|X_n - \mathbb{E}[X_n]|^3] \leq c_3 \sigma_{X_n}^3;$$

$$(A4) \quad \tilde{c}_4 \leq \sigma_{Y_n} := \text{Var}(Y_n)^{1/2} \leq c_4;$$

$$(A5) \quad \rho_{Y_n} := \mathbb{E}[|Y_n - \mathbb{E}[Y_n]|^3] \leq c_5 \sigma_{Y_n}^3;$$

$$(A6) \quad \text{the correlations } r_n := \text{Cov}(X_n, Y_n) \sigma_{X_n}^{-1} \sigma_{Y_n}^{-1} \text{ satisfy } |r_n| \leq c_6 < 1;$$

$$(A7) \quad \text{for } Y'_n := Y_n - \mathbb{E}[Y_n] - \text{Cov}(X_n, Y_n) \sigma_{X_n}^{-2} (X_n - \mathbb{E}[X_n]), \text{ for all } s \in [-\pi, \pi], \text{ and for all } t \in [-\eta_0, \eta_0],$$

$$\left| \mathbb{E}[e^{i(sX_n + tY'_n)}] \right| \leq 1 - c_7(\sigma_{X_n}^2 s^2 + \sigma_{Y'_n}^2 t^2).$$

Obviously, Assumption 1 is very close to the set of assumptions of the central limit theorem established in [8, Theorem 2.3]. In particular, (A1) is a consequence of  $m_n = N_n \mathbb{E}[X_n] + O(\sigma_{X_n} N_n^{1/2})$ , (A3), and (A7) (see the proof of Theorem 2.3 in [8]). By [8, Lemma 4.1.],  $\sigma_{X_n}^2 \leq 4\mathbb{E}[|X - \mathbb{E}[X]|^3]$ , so  $\tilde{c}_2$  can be chosen as  $1/(4c_3)$ . (A6) is not very restricting and holds in the examples provided in Section 3. Following [8], we introduce  $Y'_n$  in (A7) in order to work with a centered variable uncorrelated with  $X_n$ . If  $(X, Y')$  is a vector of centered and uncorrelated random variables, then

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

so (A7) is reasonable if the vectors  $(X_n, Y'_n)$  are identically distributed.

**Proposition 2.** Assume that

$$m_n = N_n \mathbb{E}[X_n] + O(\sigma_{X_n} N_n^{1/2}),$$

that  $(X_n, Y_n)$  converges in distribution to  $(X, Y)$  as  $n \rightarrow \infty$ , and that, for every fixed  $r > 0$ ,

$$\limsup_{n \rightarrow \infty} \mathbb{E}[|X_n|^r] < \infty \quad \text{and} \quad \limsup_{n \rightarrow \infty} \mathbb{E}[|Y_n|^r] < \infty.$$

Suppose further that the distribution of  $X$  has span 1 and that  $Y$  is not almost surely equal to an affine function  $c + dX$  of  $X$ . Then, Assumption 1 is satisfied.

The proof is omitted since the proposition relies on Corollary 2.1 and Theorem 2.3 in [8].

**Theorem 3.** Under Assumption 1,  $\tau_n^2 := \sigma_{Y_n}^2 (1 - r_n^2) > 0$  and we have

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P} \left( \frac{U_n - N_n \mathbb{E}[Y_n] - r_n \sigma_{Y_n} \sigma_{X_n}^{-1} (m_n - N_n \mathbb{E}[X_n])}{N_n^{1/2} \tau_n} \leq x \right) - \Phi(x) \right| \leq \frac{C}{N_n^{1/2}}, \quad (1)$$

where  $\Phi$  denotes the standard normal cumulative distribution function and  $C$  is a positive constant that only depends on  $\tilde{c}_2, c_2, c_3, \tilde{c}_4, c_4, c_5, c_6, c_7, \eta_0$ , and  $c_1$ .

Remark that the standardization of the variables  $U_n$  involved in (1) is not the natural one. The following theorem fixes this default of standardization.

**Proposition 4.** Under (A1), (A3), (A4), (A5), and (A7), there exist two positive constants  $d_1$  and  $d_2$  depending only on  $c_3, c_4, c_5, c_7$ , and  $c_1$  such that, for  $N_n \geq 3$ ,

$$\left| \mathbb{E}[U_n] - N_n \mathbb{E}[Y_n] - r_n \sigma_{Y_n} \sigma_{X_n}^{-1} (m_n - N_n \mathbb{E}[X_n]) \right| \leq d_1 \quad (2)$$

and

$$\left| \text{Var}(U_n) - N_n \tau_n^2 \right| \leq d_2 N_n^{1/2}. \quad (3)$$

**Theorem 5.** Under Assumption 1, we have

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P} \left( \frac{U_n - \mathbb{E}[U_n]}{\text{Var}(U_n)^{1/2}} \leq x \right) - \Phi(x) \right| \leq \frac{\tilde{C}}{N_n^{1/2}}, \quad (4)$$

where  $\tilde{C}$  is a constant that only depends on  $\tilde{c}_2, c_2, c_3, \tilde{c}_4, c_4, c_5, c_6, c_7, \eta_0$ , and  $c_1$ .

Furthermore, as in [8], the results of Theorems 3 and 5 simplify considerably in the special case when the vector  $(X_n, Y_n)$  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 Proposition 2.

### 3 Classical examples

In this section, we describe the examples mentioned in [8] and [6]. Each of them satisfies the assumptions of Proposition 2, as shown in [8], leading to a Berry-Esseen inequality.

#### 3.1 Occupancy problem

In the classical occupancy problem,  $m$  balls are thrown uniformly at random into  $N$  urns. The resulting numbers of balls  $(Z_1, \dots, Z_N)$  have a multinomial distribution. It is well known that  $(Z_1, \dots, Z_N)$  is also distributed as  $(X_1, \dots, X_N)$  conditioned on  $\{\sum_{i=1}^N X_i = m\}$ , where the random variables  $X_i$  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  $U = \sum_{i=1}^N \mathbb{1}_{\{Z_i=0\}}$ , which is distributed as  $\sum_{i=1}^N \mathbb{1}_{\{X_i=0\}}$  conditioned on  $\{\sum_{i=1}^N X_i = m\}$ . Now, if  $m = m_n \rightarrow \infty$  and  $N = N_n \rightarrow \infty$  with  $m_n/N_n \rightarrow \lambda \in (0, \infty)$ , we can take  $X_n \sim \mathcal{P}(\lambda_n)$  with  $\lambda_n := m_n/N_n$ ,  $Y_n = \mathbb{1}_{\{X_n=0\}}$ , and apply Proposition 2 to obtain a Berry-Esseen inequality for  $U_n = \sum_{i=1}^{N_n} \mathbb{1}_{\{Z_i=0\}}$ .

*Remark.* In [17], the authors prove a Berry-Esseen inequality for the occupancy problem in a more general setting: the probability of landing in each urn may be different. The tools they developed will be used in the sequel to prove our results.

*Remark.* Here, we need a result for triangular arrays, and not only for i.i.d. sequences. Indeed, if we took  $X_n = X$  with  $X \sim \mathcal{P}(\lambda)$ , we would only have

$$m_n = N_n(\lambda + o(1)) = N_n \mathbb{E}[X_n] + o(N_n).$$

But Proposition 2 requires

$$m_n = N_n \mathbb{E}[X] + O(N_n^{1/2}),$$

which is stronger. This remark goes for the following examples too.

#### 3.2 Bose-Einstein statistics

This example is borrowed from [6] (see also [3]). Consider  $N$  urns and put  $m$  indistinguishable balls in the urns in such a way that each distinguishable outcome has the same probability  $1/\binom{m+N-1}{m}$ . 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  $(X_1, \dots, X_N)$  conditioned on  $\{\sum_{i=1}^N X_i = m\}$ , where the random variables  $X_i$  are i.i.d., with  $X_i \sim \mathcal{G}(p)$ , for any arbitrary  $p \in (0, 1)$ . If  $m = n$ ,  $N = N_n \rightarrow \infty$  with  $N_n/n \rightarrow p$ , take  $X_n \sim \mathcal{G}(p_n)$  with  $p_n = N_n/n$  to obtain a Berry-Esseen inequality for any sequence of variables of the type  $U_n = \sum_{i=1}^{N_n} f(Z_i)$ .

#### 3.3 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 and  $S_k := \sum_{i=1}^k X_i$ . It is well known (see [8, Example 3.4] and the references therein) that the total progeny  $S_N + 1$  is  $N \geq 1$  if and only if

$$\forall k \in \{0, \dots, N-1\} \quad S_k \geq k \quad \text{and} \quad S_N = N-1. \quad (5)$$

This type of conditioning is different from the one studied in the present paper, but by [18, Corollary 2] and [8, Example 3.4], if we ignore the cyclical order of  $X_1, \dots, X_N$ , it is proven that  $X_1, \dots, X_N$  have the same distribution conditioned on (5) as conditioned on  $\{S_N = N-1\}$ . Applying Proposition 2 with  $N = n$  and  $m = n-1$ , we obtain a Berry-Esseen inequality for any sequence of variables  $U_n$  distributed as  $T_n = \sum_{i=1}^n f(X_i)$  conditioned on  $\{S_n = n-1\}$ . For instance, if  $f(x) = \mathbb{1}_{\{x=3\}}$ ,  $U_n$  is the number of individuals with three children given that the total progeny is  $n$ .

#### 3.4 Random forests

Consider a uniformly distributed random labeled rooted forest with  $m$  vertices and  $N$  roots with  $N < m$ . 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 [8], 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 the random variables  $X_i$  are i.i.d. and Borel distributed for any arbitrary parameter  $\mu \in (0, 1)$ , i.e.

$$\mathbb{P}(X_i = l) = e^{-\mu l} \frac{(\mu l)^{l-1}}{l!}$$

(see, e.g., [5] or [7] for more details). Then, the  $i$ th tree is drawn uniformly among the trees of size  $X_i$ . Proposition 2 provides a Berry-Esseen inequality for any sequence of variables of the type  $U_n = \sum_{i=1}^{N_n} f(Z_i)$  where  $N_n \rightarrow \infty$  and  $Z_1, \dots, Z_{N_n}$  are the sizes of the trees in the forest. For instance, if  $f(x) = \mathbb{1}_{\{x=K\}}$ ,  $U_n$  is the number of trees of size  $K$  in the forest (see, e.g., [12, 15, 16]).

### 3.5 Hashing with linear probing

Hashing with linear probing is a classical model in theoretical computer science that appeared in the 60's. It has been studied from a mathematical point of view firstly in [10]. For more details on the model, we refer to [5, 7, 14, 1, 2, 9]. The model describes the following experiment. One throws  $n$  balls sequentially into  $m$  urns at random with  $m > n$ ; 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  $d_{m,n}$ . After throwing all balls, there are  $N := m - n$  empty urns. These divide the occupied urns into blocks of consecutive urns. We consider that the empty urn following a block belongs to this block. Following [11, 5], Janson [7] proves that the lengths of the blocks and the sums 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 the random vectors  $(X_i, Y_i)$  are i.i.d. copies of a vector  $(X, Y)$  of random variables:  $X$  being Borel distributed with any arbitrary parameter  $\mu \in (0, 1)$  and  $Y$  given  $\{X = l\}$  being distributed as  $d_{l,l-1}$ . In particular,  $d_{m,n}$  is distributed as  $\sum_{i=1}^N Y_i$  conditioned on  $\{\sum_{i=1}^N X_i = m\}$ . If  $m = m_n \rightarrow \infty$  and  $N = N_n = m_n - n \rightarrow \infty$  with  $n/m_n \rightarrow \mu \in (0, 1)$ , we take  $X_n$  following Borel distribution with parameter  $\mu_n := n/m_n$  to get a Berry-Esseen inequality for  $d_{m_n,n}$ , by Proposition 2.

## 4 Proofs

Remind that  $U_n$  is distributed as  $T_n$  conditioned on  $\{S_n = m_n\}$ . Following the procedure of [8], we consider the projection

$$Y'_n = Y_n - \mathbb{E}[Y_n] - \text{Cov}(X_n, Y_n) \sigma_{X_n}^{-2} (X_n - \mathbb{E}[X_n]).$$

Then  $\mathbb{E}[Y'_n] = 0$  and  $\text{Cov}(X_n, Y'_n) = \mathbb{E}[X_n Y'_n] = 0$ . Besides, (A7) and (A6) are verified by  $Y'_n$ . By (A6),

$$\sigma_{Y'_n}^2 = \sigma_{Y_n}^2 (1 - r_n^2) \in [\tilde{c}_4^2 (1 - c_6^2), c_4^2],$$

so (A4) is satisfied by  $Y'_n$ . Finally, by Minkowski inequality, (A3) and (A5), and the fact that  $|r_n| \leq 1$ ,

$$\begin{aligned} \|Y'_n\|_3 &\leq \|Y_n - \mathbb{E}[Y_n]\|_3 + |r_n| \sigma_{X_n} \sigma_{Y_n} \sigma_{X_n}^{-2} \|X_n - \mathbb{E}[X_n]\|_3 \\ &\leq \rho_{Y_n}^{1/3} + \sigma_{Y_n} \sigma_{X_n}^{-1} \rho_{X_n}^{1/3} \\ &\leq \sigma_{Y_n} (c_3^{1/3} + c_5^{1/3}) \\ &\leq \sigma_{Y'_n} (1 - c_6^2)^{-1/2} (c_3^{1/3} + c_5^{1/3}). \end{aligned}$$

Hence,  $Y'_n$  satisfies (A5). Consequently, all conditions hold for the vector  $(X_n, Y'_n)$  too. Finally,

$$T'_n := \sum_{i=1}^{N_n} Y'_{n,i} = T_n - N_n \mathbb{E}[Y_n] - \text{Cov}(X_n, Y_n) \sigma_{X_n}^{-2} (S_n - N_n \mathbb{E}[X_n]).$$

So, conditioned on  $\{S_n = m_n\}$ , we have  $T'_n = T_n - N_n \mathbb{E}[Y_n] - r_n \sigma_{Y_n} \sigma_{X_n}^{-1} (m_n - N_n \mathbb{E}[X_n])$ . Hence the conclusions in Theorems 3 and 5 for  $(X_n, Y_n)$  and  $(X_n, Y'_n)$  are the same. Thus, it suffices to prove the theorems for  $(X_n, Y'_n)$ . In other words, we will henceforth assume that  $\mathbb{E}[Y_n] = \mathbb{E}[X_n Y_n] = 0$ ,  $r_n = 0$  and  $\tau_n^2 = \sigma_{Y_n}^2$ . Moreover, the constants  $c'_4, \tilde{c}'_4, c'_5, c'_6$ , and  $c'_7$  for  $(X, Y')$  are linked to that of  $(X, Y)$  by the following relations:  $c'_4 = c_4$ ,  $\tilde{c}'_4 = \tilde{c}_4 (1 - c_6^2)^{1/2}$ ,  $c'_5 = (1 - c_6^2)^{-3/2} (c_3^{1/3} + c_5^{1/3})^3$ ,  $c'_6 = 0$ , and  $c'_7 = c_7$ . In the proofs, we omit the primes.

The proofs of Theorems 3 and 5 intensively rely on the use of Fourier transforms through the functions  $\varphi_n$  and  $\psi_n$  defined by

$$\varphi_n(s, t) := \mathbb{E}[\exp\{is(X_n - \mathbb{E}[X_n]) + itY_n\}] \quad \text{and} \quad \psi_n(t) := 2\pi \mathbb{P}(S_n = m_n) \mathbb{E}[\exp\{itU_n\}]. \quad (6)$$

The controls of these functions (respectively the controls of their derivatives) needed in the proofs are postponed to Subection 4.4 in Lemmas 6 and 7 (resp. in Lemma 8). In particular, (16), (17), (18), and (19) will be used several times in the sequel.

### 4.1 Proof of Theorem 3

We follow the classical proof of Berry-Esseen theorem (see e.g. [4]) combined with the procedure in [17]. As shown in [13] (page 285) or [4], the left hand side of (1) is dominated by

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

where  $\eta > 0$  is arbitrary. We choose to define

$$\eta := \min\left(\frac{2}{9}(c_4 c_5)^{-1}, \eta_0\right) > 0. \quad (7)$$

From (16) of Lemma 6 and a Taylor's expansion,

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

By (A1),  $\gamma_n \geq c_1$ . Now we split the integration domain of  $s$  into

$$A_1 := \left\{ s : |s| < \varepsilon\sigma_{X_n} N_n^{1/2} \right\} \quad \text{and} \quad A_2 := \left\{ s : \varepsilon\sigma_{X_n} N_n^{1/2} \leq |s| \leq \pi\sigma_{X_n} N_n^{1/2} \right\},$$

where

$$\varepsilon := \min\left(\frac{2}{9}(c_2 c_3)^{-1}, \pi\right) \quad (8)$$

and decompose

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

where

$$I_1(n, u, \theta) = \gamma_n^{-1} \int_{A_1} e^{-(u^2+s^2)/2} \left| \frac{\partial}{\partial t} \left[ e^{(t^2+s^2)/2} \varphi_n^{N_n} \left( \frac{s}{\sigma_{X_n} N_n^{1/2}}, \frac{t}{\sigma_{Y_n} N_n^{1/2}} \right) \right]_{t=\theta} \right| ds, \quad (9)$$

$$I_2(n, u, \theta) = \gamma_n^{-1} e^{-u^2/2} \int_{A_2} \left| \frac{\partial}{\partial t} \left[ e^{t^2/2} \varphi_n^{N_n} \left( \frac{s}{\sigma_{X_n} N_n^{1/2}}, \frac{t}{\sigma_{Y_n} N_n^{1/2}} \right) \right]_{t=\theta} \right| ds. \quad (10)$$

Lemmas 10 and 11 state that there exists positive constants  $C_1$  and  $C_2$ , only depending on  $\tilde{c}_2, c_2, c_3, c_5, c_7$ , and  $c_1$ , such that, for  $N_n \geq \max(12^3 c_3^2, 12^3 c_5^2, 2)$ ,

$$\int_0^{\eta\sigma_{Y_n} N_n^{1/2}} \sup_{0 \leq \theta \leq u} I_1(n, u, \theta) du \leq \frac{C_1}{N_n^{1/2}}, \quad (11)$$

and

$$\int_0^{\eta\sigma_{Y_n} N_n^{1/2}} \sup_{0 \leq \theta \leq u} I_2(n, u, \theta) du \leq \frac{C_2}{N_n^{1/2}}. \quad (12)$$

So,

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P} \left( \frac{U_n}{N_n^{1/2} \sigma_{Y_n}} \leq x \right) - \Phi(x) \right| \leq \frac{C}{N_n^{1/2}}$$

with

$$C := \max\left(C_1 + C_2 + \frac{24}{\tilde{c}_4 \pi \sqrt{2\pi}} \left( \min\left(\frac{2}{9} c_4 c_5, \eta_0\right) \right)^{-1}, 12^{3/2} c_3, 12^{3/2} c_5, \sqrt{2}\right).$$

## 4.2 Proof of Proposition 4

**Proof of (2)** We adapt the proof given in [8]. Using the definition (6) of  $\Psi_n$ , and differentiating under the integral sign of (16) of Lemma 6, we naturally have

$$\begin{aligned} |\mathbb{E}[U_n]| &= \left| \frac{-i\psi'_n(0)}{2\pi\mathbb{P}(S_n = m_n)} \right| \\ &\leq \gamma_n^{-1} N_n \int_{-\pi\sigma_{X_n} N_n^{1/2}}^{\pi\sigma_{X_n} N_n^{1/2}} \left| \frac{\partial\varphi_n}{\partial t} \left( \frac{s}{\sigma_{X_n} N_n^{1/2}}, 0 \right) \right| \cdot \left| \varphi_n^{N_n-1} \left( \frac{s}{\sigma_{X_n} N_n^{1/2}}, 0 \right) \right| ds. \end{aligned}$$

Using (19) of Lemma 8 with  $t = 0$ , (A2), (A3), and (A5), we deduce

$$\left| \frac{\partial\varphi_n}{\partial t} \left( \frac{s}{\sigma_{X_n} N_n^{1/2}}, 0 \right) \right| \leq \frac{s^2}{2} \frac{\rho_{Y_n}^{1/3} \rho_{X_n}^{2/3}}{\sigma_{X_n}^2 N_n} \leq \frac{c_3^{2/3} c_4 c_5^{1/3}}{2N_n} s^2.$$

Then using (17) of Lemma 7 (with  $l = 1$  and  $t = 0$ ) and for  $N_n \geq 3$ ,

$$\int_{-\pi\sigma_{X_n} N_n^{1/2}}^{\pi\sigma_{X_n} N_n^{1/2}} \left| \frac{\partial\varphi_n}{\partial t} \left( \frac{s}{\sigma_{X_n} N_n^{1/2}}, 0 \right) \right| \cdot \left| \varphi_n^{N_n-1} \left( \frac{s}{\sigma_{X_n} N_n^{1/2}}, 0 \right) \right| ds \leq \frac{c_3^{2/3} c_4 c_5^{1/3}}{2N_n} \int_{-\infty}^{+\infty} s^2 e^{-2c_7 s^2/3} ds.$$

So, (2) holds with

$$d_1 := 2^{-1} c_3^{2/3} c_4 c_5^{1/3} c_1^{-1} \int_{-\infty}^{+\infty} s^2 e^{-2c_7 s^2/3} ds.$$

**Proof of (3)** Since  $\tau_n^2 = \sigma_{Y_n}^2$  and  $\mathbb{E}[U_n]$  is bounded, it suffices to show that the quantity  $|\mathbb{E}[U_n^2] - N_n \sigma_{Y_n}^2|$  is bounded by some  $d_2' N_n^{1/2}$  to prove (3). Proceeding as done previously,

$$\begin{aligned} \mathbb{E}[U_n^2] &= \frac{-\psi_n''(0)}{2\pi\mathbb{P}(S_n = m_n)} \\ &= -\gamma_n^{-1} N_n (N_n - 1) \int_{-\pi\sigma_{X_n} N_n^{1/2}}^{\pi\sigma_{X_n} N_n^{1/2}} e^{-isv_n} \left( \frac{\partial\varphi_n}{\partial t} \left( \frac{s}{\sigma_{X_n} N_n^{1/2}}, 0 \right) \right)^2 \varphi_n^{N_n-2} \left( \frac{s}{\sigma_{X_n} N_n^{1/2}}, 0 \right) ds \end{aligned} \quad (13)$$

$$- \gamma_n^{-1} N_n \int_{-\pi\sigma_{X_n} N_n^{1/2}}^{\pi\sigma_{X_n} N_n^{1/2}} e^{-isv_n} \frac{\partial^2\varphi_n}{\partial t^2} \left( \frac{s}{\sigma_{X_n} N_n^{1/2}}, 0 \right) \varphi_n^{N_n-1} \left( \frac{s}{\sigma_{X_n} N_n^{1/2}}, 0 \right) ds \quad (14)$$

where

$$v_n := (m_n - N_n \mathbb{E}[X_n]) / (\sigma_{X_n} N_n^{1/2}). \quad (15)$$

First, by (19) of Lemma 8 with  $t = 0$  and by (17) of Lemma 7 (with  $l = 2$  and  $t = 0$ ), one has, for  $N_n \geq 3$ ,

$$\begin{aligned} &\int_{-\pi\sigma_{X_n} N_n^{1/2}}^{\pi\sigma_{X_n} N_n^{1/2}} \left| \frac{\partial\varphi_n}{\partial t} \left( \frac{s}{\sigma_{X_n} N_n^{1/2}}, 0 \right) \right|^2 \left| \varphi_n^{N_n-2} \left( \frac{s}{\sigma_{X_n} N_n^{1/2}}, 0 \right) \right| ds \\ &\leq \frac{c_3^{4/3} c_4^2 c_5^{2/3}}{4N_n^2} \int_{-\infty}^{+\infty} s^4 e^{-c_7 s^2/3} ds. \end{aligned}$$

Finally, by (A1), the term (13) is bounded by

$$d_2'' := \frac{c_3^{4/3} c_4^2 c_5^{2/3}}{4c_1} \int_{-\infty}^{+\infty} s^4 e^{-c_7 s^2/3} ds.$$

Second, we study the term (14). We want to show that

$$\Delta_n := \gamma_n^{-1} \int_{-\pi\sigma_{X_n} N_n^{1/2}}^{\pi\sigma_{X_n} N_n^{1/2}} e^{-isv_n} \frac{\partial^2\varphi_n}{\partial t^2} \left( \frac{s}{\sigma_{X_n} N_n^{1/2}}, 0 \right) \varphi_n^{N_n-1} \left( \frac{s}{\sigma_{X_n} N_n^{1/2}}, 0 \right) ds + \sigma_{Y_n}^2$$

is bounded by some  $d_2'''/N_n^{1/2}$ . By (16) with  $t = 0$ ,

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

so

$$\begin{aligned}\Delta_n &= \gamma_n^{-1} \int_{-\pi\sigma_{X_n}N_n^{1/2}}^{\pi\sigma_{X_n}N_n^{1/2}} e^{-isv_n} \left( \frac{\partial^2 \varphi_n}{\partial t^2} \left( \frac{s}{\sigma_{X_n}N_n^{1/2}}, 0 \right) + \sigma_{Y_n}^2 \varphi_n \left( \frac{s}{\sigma_{X_n}N_n^{1/2}}, 0 \right) \right) \varphi_n^{N_n-1} \left( \frac{s}{\sigma_{X_n}N_n^{1/2}}, 0 \right) ds \\ &= \gamma_n^{-1} \int_{-\pi\sigma_{X_n}N_n^{1/2}}^{\pi\sigma_{X_n}N_n^{1/2}} e^{-isv_n} \mathbb{E} \left[ Y_n^2 f(s) \right] \cdot \varphi_n^{N_n-1} \left( \frac{s}{\sigma_{X_n}N_n^{1/2}}, 0 \right) ds\end{aligned}$$

where

$$f(s) = - \left( e^{is\sigma_{X_n}^{-1}N_n^{-1/2}(X_n - \mathbb{E}[X_n])} - \mathbb{E} \left[ e^{is\sigma_{X_n}^{-1}N_n^{-1/2}(X_n - \mathbb{E}[X_n])} \right] \right).$$

Applying Taylor's theorem yields

$$\begin{aligned}|f(s)| &\leq |s| \sup_u \left| -i \frac{X_n - \mathbb{E}[X_n]}{\sigma_{X_n}N_n^{1/2}} e^{iu\sigma_{X_n}^{-1}N_n^{-1/2}(X_n - \mathbb{E}[X_n])} + \mathbb{E} \left[ i \frac{X_n - \mathbb{E}[X_n]}{\sigma_{X_n}N_n^{1/2}} e^{iu\sigma_{X_n}^{-1}N_n^{-1/2}(X_n - \mathbb{E}[X_n])} \right] \right| \\ &\leq \frac{|s|}{N_n^{1/2}} \left( \left| \frac{X_n - \mathbb{E}[X_n]}{\sigma_{X_n}} \right| + \mathbb{E} \left[ \left| \frac{X_n - \mathbb{E}[X_n]}{\sigma_{X_n}} \right| \right] \right).\end{aligned}$$

Thus, using Hölder's inequality,

$$\begin{aligned}|\mathbb{E}[Y_n^2 f(s)]| &\leq \frac{|s|}{N_n^{1/2}} \mathbb{E} \left[ Y_n^2 \left( \left| \frac{X_n - \mathbb{E}[X_n]}{\sigma_{X_n}} \right| + \mathbb{E} \left[ \left| \frac{X_n - \mathbb{E}[X_n]}{\sigma_{X_n}} \right| \right] \right) \right] \\ &\leq \frac{\sigma_{Y_n}^2 |s|}{N_n^{1/2}} \left( \frac{\rho_{Y_n}^{2/3} \rho_{X_n}^{1/3}}{\sigma_{Y_n}^2 \sigma_{X_n}} + 1 \right) \\ &\leq \frac{|s| c_4^2}{N_n^{1/2}} \left( c_5^{2/3} c_3^{1/3} + 1 \right)\end{aligned}$$

where the last inequality is obtained using (A2), (A3), (A4), and (A5). Now, by (A1) and the upper bound in (17) (with  $l = 1$  and  $t = 0$ ), we get, for  $N_n \geq 3$ ,

$$|\Delta_n| \leq \frac{c_4^2}{c_1 N_n^{1/2}} \left( c_5^{2/3} c_3^{1/3} + 1 \right) \int_{-\infty}^{+\infty} |s| e^{-s^2 c_7 (N_n - 1)/N_n} ds \leq \frac{d_2'''}{N_n^{1/2}},$$

with

$$d_2''' := c_4^2 c_1^{-1} (c_5^{2/3} c_3^{1/3} + 1) \int_{-\infty}^{+\infty} |s| e^{-2s^2 c_7/3} ds.$$

Finally,

$$|\text{Var}(U_n) - N_n \sigma_{Y_n}^2| \leq (d_1^2 + d_2'' + d_2''') N_n^{1/2} =: d_2 N_n^{1/2}.$$

Then the proof of (3) is complete.

### 4.3 Proof of Theorem 5

Write

$$\begin{aligned}\left| \mathbb{P} \left( \frac{U_n - \mathbb{E}[U_n]}{\text{Var}(U_n)^{1/2}} \leq x \right) - \Phi(x) \right| &\leq \left| \mathbb{P} \left( \frac{U_n}{N_n^{1/2} \sigma_{Y_n}} \leq a_n x + b_n \right) - \Phi(a_n x + b_n) \right| \\ &\quad + |\Phi(a_n x + b_n) - \Phi(x)|\end{aligned}$$

where

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

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

$$|a_n - 1| \leq |a_n^2 - 1| \leq d_2 \tilde{c}_4^{-2} N_n^{-1/2} \quad \text{and} \quad |b_n| \leq d_1 \tilde{c}_4^{-1} N_n^{-1/2}.$$

Then for  $N_n^{1/2} \geq 2\tilde{c}_4^{-2} d_2$ ,  $a_n \geq 1/2$  and applying Taylor's theorem to  $\Phi$ , one gets

$$\begin{aligned}|\Phi(a_n x + b_n) - \Phi(x)| &\leq |(a_n - 1)x + b_n| \sup_t \frac{e^{-t^2/2}}{\sqrt{2\pi}} \\ &\leq \frac{N_n^{-1/2}}{\sqrt{2\pi}} \max(d_2 \tilde{c}_4^{-2}, d_1 \tilde{c}_4^{-1}) (|x| + 1) e^{-(|x|/2 - d_1 \tilde{c}_4^{-1})^2/2}\end{aligned}$$

the supremum being over  $t$  between  $x$  and  $a_n x + b_n$ . The last function in  $x$  being bounded, we can define

$$C' := \frac{1}{\sqrt{2\pi}} \max(d_2 \tilde{c}_4^{-2}, d_1 \tilde{c}_4^{-1}) \sup_{x \in \mathbb{R}} \left[ (|x| + 1) e^{-(|x|/2 - d_1 \tilde{c}_4^{-1})^2 / 2} \right].$$

Finally, we apply (1) and (4) holds with  $\tilde{C} := C + \max(C', 2\tilde{c}_4^{-2} d_2)$ .

#### 4.4 Technical results

Recall that  $v_n = (m_n - N_n \mathbb{E}[X_n]) / (\sigma_{X_n} N_n^{1/2})$  and  $\gamma_n = 2\pi \mathbb{P}(S_n = m_n) \sigma_{X_n} N_n^{1/2}$ . Moreover,

$$\varphi_n(s, t) = \mathbb{E}[\exp\{is(X_n - \mathbb{E}[X_n]) + itY_n\}] \quad \text{and} \quad \psi_n(t) = 2\pi \mathbb{P}(S_n = m_n) \mathbb{E}[\exp\{itU_n\}].$$

**Lemma 6.** *One has*

$$\psi_n(t) = \frac{1}{\sigma_{X_n} N_n^{1/2}} \int_{-\pi \sigma_{X_n} N_n^{1/2}}^{\pi \sigma_{X_n} N_n^{1/2}} e^{-isv_n} \varphi_n^{N_n} \left( \frac{s}{\sigma_{X_n} N_n^{1/2}}, t \right) ds \quad (16)$$

*Proof.* Indeed, since

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

we have

$$\begin{aligned} \psi_n(t) &= 2\pi \mathbb{P}(S_n = m_n) \mathbb{E}[\exp\{itU_n\}] \\ &= 2\pi \mathbb{E}[\exp\{itT_n\} \mathbb{1}_{\{S_n = m_n\}}] \\ &= \int_{-\pi}^{\pi} \mathbb{E}[\exp\{is(S_n - m_n) + itT_n\}] ds \\ &= \int_{-\pi}^{\pi} e^{-is(m_n - N_n \mathbb{E}[X_n])} \varphi_n^{N_n}(s, t) ds, \end{aligned}$$

which leads to (16) after the change of variable  $s' = s \sigma_{X_n} N_n^{1/2}$ . □

Now we give controls on the function  $\varphi_n$  and its partial derivatives (see Lemmas 7 and 8).

**Lemma 7.** *Under (A7), for any integer  $l \geq 0$ ,  $|s| \leq \pi \sigma_{X_n} N_n^{1/2}$ , and  $|t| \leq \eta_0 \sigma_{Y_n} N_n^{1/2}$ , one gets*

$$\left| \varphi_n^{N_n - l} \left( \frac{s}{\sigma_{X_n} N_n^{1/2}}, \frac{t}{\sigma_{Y_n} N_n^{1/2}} \right) \right| \leq e^{-(s^2 + t^2) \cdot c_7 \cdot (N_n - l) / N_n}. \quad (17)$$

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

**Lemma 8.** *For any  $s$  and  $t$ , one has*

$$\left| \frac{\partial \varphi_n}{\partial t} \left( \frac{s}{\sigma_{X_n} N_n^{1/2}}, \frac{t}{\sigma_{Y_n} N_n^{1/2}} \right) \right| \leq \frac{\sigma_{Y_n}}{N_n^{1/2}} (|s| + |t|); \quad (18)$$

and

$$\begin{aligned} \left| \frac{\partial \varphi_n}{\partial t} \left( \frac{s}{\sigma_{X_n} N_n^{1/2}}, \frac{t}{\sigma_{Y_n} N_n^{1/2}} \right) \right| &\leq \frac{\sigma_{Y_n}}{N_n^{1/2}} |t| + \frac{\sigma_{Y_n}}{N_n} \left[ \frac{s^2}{2} \left( \frac{\rho_{X_n}}{\sigma_{X_n}^3} \right)^{2/3} \left( \frac{\rho_{Y_n}}{\sigma_{Y_n}^3} \right)^{1/3} \right. \\ &\quad \left. + |st| \left( \frac{\rho_{X_n}}{\sigma_{X_n}^3} \right)^{1/3} \left( \frac{\rho_{Y_n}}{\sigma_{Y_n}^3} \right)^{2/3} + \frac{t^2}{2} \left( \frac{\rho_{Y_n}}{\sigma_{Y_n}^3} \right) \right]. \end{aligned} \quad (19)$$

*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 (18) using

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



and to (19) using

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

The partial derivatives of  $f$  are estimated by mixed moments of  $X_n$  and  $Y_n$  and then bounded above by Hölder's inequality.  $\square$

The following lemma is a result due to Quine and Robinson ([17, Lemma 2]).

**Lemma 9.** *Define*

$$l_{1,n} := \rho_{X_n} \sigma_{X_n}^{-3} N_n^{-1/2} \quad \text{and} \quad l_{2,n} := \rho_{Y_n} \sigma_{Y_n}^{-3} N_n^{-1/2}.$$

If  $l_{1,n} \leq 12^{-3/2}$  and  $l_{2,n} \leq 12^{-3/2}$ , 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

$$\left| \frac{\partial}{\partial t} \left[ e^{(s^2+t^2)/2} \varphi_n^{N_n} \left( \frac{s}{\sigma_{X_n} N_n^{1/2}}, \frac{t}{\sigma_{Y_n} N_n^{1/2}} \right) \right] \right| \leq C_4 (|s| + |t| + 1)^3 (l_{1,n} + l_{2,n}) \exp \left\{ \frac{11}{24} (s^2 + t^2) \right\},$$

with  $C_4 := 161$ .

*Remark.* We make explicit the constant  $C_4$  appearing at the end of the proof of Lemma 2 in [17]. For all  $v$  and  $s$  in  $R_2$  as defined in [17], one has

$$\frac{(|v| + 2|s|)}{(|v| + |s| + 1)^3 (l_{1,n} + l_{2,n})} e^{-(v^2+s^2)/24} \leq 108 \cdot \sqrt{6} \cdot e^{-1/2} \leq 161.$$

By (A2) and (A3),

$$l_{1,n} \leq c_3 N_n^{-1/2} \leq c_2 c_3 \sigma_{X_n}^{-1} N_n^{-1/2},$$

which implies that  $\sigma_{X_n} N_n^{1/2} \leq c_2 c_3 l_{1,n}^{-1}$ . Similarly,

$$l_{2,n} \leq c_5 N_n^{-1/2} \leq c_4 c_5 \sigma_{Y_n}^{-1} N_n^{-1/2},$$

and  $\sigma_{Y_n} N_n^{1/2} \leq c_4 c_5 l_{2,n}^{-1}$ . Now we are able to establish (11).

**Lemma 10.** *There exists a positive constant  $C_1$ , only depending on  $c_3, c_5, c_1$  such that, for  $N_n \geq 12^3 \max(c_3^2, c_5^2)$ ,*

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

*Proof.* The definitions of  $\eta$  in (7) and  $\varepsilon$  in (8) imply that, for  $s \in A_1$  and  $u$  and  $\theta$  as in the integral in the statement above, one has

$$|s| < \varepsilon \sigma_{X_n} N_n^{1/2} \leq \frac{2}{9} l_{1,n}^{-1} \quad \text{and} \quad |\theta| \leq |u| \leq \eta \sigma_{Y_n} N_n^{1/2} \leq \frac{2}{9} l_{2,n}^{-1},$$

which ensures that  $(s, \theta) \in R$  as specified in Lemma 9. Moreover, for  $N_n \geq 12^3 \max(c_3^2, c_5^2)$ ,  $l_{1,n} \leq 12^{-3/2}$  and  $l_{2,n} \leq 12^{-3/2}$ . Now using Lemma 9 in (9) and by (A1), we get

$$\begin{aligned} &\int_0^{\eta \sigma_{Y_n} N_n^{1/2}} \sup_{0 \leq \theta \leq u} I_1(n, u, \theta) du \\ &\leq \gamma_n^{-1} C_4 (l_{1,n} + l_{2,n}) \int_0^{\eta \sigma_{Y_n} N_n^{1/2}} \int_{A_1} (|s| + |u| + 1)^3 e^{-(s^2+u^2)/24} ds du \\ &\leq N_n^{-1/2} c_1^{-1} C_4 (c_3 + c_5) \int_{\mathbb{R}^2} (|s| + |u| + 1)^3 e^{-(s^2+u^2)/24} ds du \end{aligned}$$

and the result follows with

$$C_1 := c_1^{-1} C_4 (c_3 + c_5) \int_{\mathbb{R}^2} (|s| + |u| + 1)^3 e^{-(s^2+u^2)/24} ds du.$$

$\square$

*Remark.* Actually, Lemma 10 is valid as soon as  $N_n \geq \max(c_3^2, c_5^2)$ : the constants in the proof of Lemma 2 in [17] can be improved.

Now we are able to prove (12).

**Lemma 11.** *There exists a positive constant  $C_2$ , only depending on  $c_1$ ,  $\tilde{c}_2$ ,  $c_2$ ,  $c_3$ , and  $c_7$  such that, for  $N_n \geq 2$ ,*

$$\int_0^{\eta\sigma_{Y_n}N_n^{1/2}} \sup_{0 \leq \theta \leq u} I_2(n, u, \theta) du \leq \frac{C_2}{N_n^{1/2}}.$$

*Proof.* We use the controls (17) with  $t = \theta$  and  $l = 1$ , (18), and  $|\varphi_n| \leq 1$  to get

$$\begin{aligned} & \left| \frac{\partial}{\partial t} \left[ e^{t^2/2} \varphi_n^{N_n} \left( \frac{s}{\sigma_{X_n} N_n^{1/2}}, \frac{t}{\sigma_{Y_n} N_n^{1/2}} \right) \right]_{t=\theta} \right| \\ &= e^{\theta^2/2} \left| \varphi_n^{N_n-1} \left( \frac{s}{\sigma_{X_n} N_n^{1/2}}, \frac{\theta}{\sigma_{Y_n} N_n^{1/2}} \right) \right| \cdot \left| \theta \varphi_n \left( \frac{s}{\sigma_{X_n} N_n^{1/2}}, \frac{\theta}{\sigma_{Y_n} N_n^{1/2}} \right) \right. \\ & \quad \left. + \frac{N_n^{1/2}}{\sigma_{Y_n}} \frac{\partial \varphi_n}{\partial t} \left( \frac{s}{\sigma_{X_n} N_n^{1/2}}, \frac{\theta}{\sigma_{Y_n} N_n^{1/2}} \right) \right| \\ & \leq (|s| + 2|\theta|) e^{\theta^2/2 - (s^2 + \theta^2) \cdot c_7(N_n-1)/N_n}, \end{aligned}$$

for  $s \in A_2$  and  $u$  and  $\theta$  as in the integral in the statement of the Lemma. Finally, using (10), we get that, for  $N_n \geq 2$ ,

$$\begin{aligned} & \int_0^{\eta\sigma_{Y_n}N_n^{1/2}} \sup_{0 \leq \theta \leq u} I_2(n, u, \theta) du \\ & \leq 2\gamma_n^{-1} \int_0^{+\infty} \int_{\varepsilon\sigma_{X_n}N_n^{1/2}}^{+\infty} \sup_{0 \leq \theta \leq u} \left[ (s + 2\theta) \exp \left( \frac{\theta^2}{2} \left( 1 - 2c_7 \frac{N_n-1}{N_n} \right) \right) \right] \\ & \quad \cdot e^{-u^2/2 - s^2 \cdot c_7(N_n-1)/N_n} ds du \\ & \leq 2c_1^{-1} \int_0^{+\infty} \int_{\varepsilon\sigma_{X_n}N_n^{1/2}}^{+\infty} (s + 2u) e^{-\min(1, c_7)u^2/2 - s^2 c_7/2} ds du \\ & \leq e^{-N_n c_7 \varepsilon^2 \sigma_{X_n}^2/2} \left( \frac{c_1^{-1} c_7^{-1} \sqrt{2\pi}}{\sqrt{\min(1, c_7)}} + \frac{4c_1^{-1}}{\min(1, c_7)} \frac{1}{c_7 \varepsilon \sigma_{X_n} N_n^{1/2}} \right) \\ & \leq C_2' e^{-C_2'' N_n} \end{aligned}$$

where

$$C_2' := c_1^{-1} c_7^{-1} \left( \frac{\sqrt{2\pi}}{\sqrt{\min(1, c_7)}} + \frac{4}{\min(1, c_7) \min \left( \frac{2}{9} (c_2 c_3)^{-1}, \pi \right) \tilde{c}_2} \right)$$

and  $C_2'' := \tilde{c}_2^2/2c_7 \min \left( \frac{2}{9} (c_2 c_3)^{-1}, \pi \right)^2$ . The result follows, writing

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

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

## 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 Philippe Flajolet. Hachage, arbres, chemins & graphes. *Gaz. Math.*, (95):29–49, 2003.
- [3] W. Feller. *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons Inc., New York, 1968.

- [4] W. Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York, 1971.
- [5] 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.
- [6] L. Holst. Two conditional limit theorems with applications. *Ann. Statist.*, 7(3):551–557, 1979.
- [7] 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).
- [8] S. Janson. Moment convergence in conditional limit theorems. *J. Appl. Probab.*, 38(2):421–437, 2001.
- [9] S. Janson. Individual displacements for linear probing hashing with different insertion policies. *ACM Trans. Algorithms*, 1(2):177–213, 2005.
- [10] D. E. Knuth. Computer science and its relation to mathematics. *Amer. Math. Monthly*, 81:323–343, 1974.
- [11] D. E. Knuth. Linear probing and graphs. *Algorithmica*, 22(4):561–568, 1998. Average-case analysis of algorithms.
- [12] 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.
- [13] M. Loève. *Probability theory. Foundations. Random sequences.* D. Van Nostrand Company, Inc., Toronto-New York-London, 1955.
- [14] J.-F. Marckert. Parking with density. *Random Structures Algorithms*, 18(4):364–380, 2001.
- [15] 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.
- [16] Y. L. Pavlov. Random forests. In *Probabilistic methods in discrete mathematics (Petrozavodsk, 1996)*, pages 11–18. VSP, Utrecht, 1997.
- [17] M. P. Quine and J. Robinson. A Berry-Esseen bound for an occupancy problem. *Ann. Probab.*, 10(3):663–671, 1982.
- [18] J. G. Wendel. Left-continuous random walk and the Lagrange expansion. *Amer. Math. Monthly*, 82:494–499, 1975.