J. Appl. Prob. **56,** 76–90 (2019)
*doi:*10.1017/jpr.2019.7

A CONDITIONAL BERRY–ESSEEN INEQUALITY

THIERRY KLEIN,* *ENAC and Institut de Mathématiques de Toulouse (UMR CNRS 5219)*
AGNÈS LAGNOUX,** *and*
PIERRE PETIT,** *Institut de Mathématiques de Toulouse (UMR CNRS 5219)*

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 problem; occupancy; hashing with linear probing; random forest; branching process; Bose–Einstein statistics

2010 Mathematics Subject Classification: Primary 60F05; 62E20
Secondary 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 $N^{-1}(Y_1 + \cdots + Y_N)$ conditioned on a specific value of $X_1 + \cdots + X_N$, that is, in the conditional distribution

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

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

Janson [8] proved a general central limit theorem (with convergence of all moments) for this kind of conditional distribution under some reasonable assumptions and gave 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 $Y = \mathbf{1}_{\{X=0\}}$. To the best of our knowledge, it is the only result in that direction for this kind of conditional distribution.

Received 31 May 2018; revision received 6 November 2018.
* Postal address: ENAC, 7 Avenue Edouard Belin, F-31400 Toulouse, France.
Email address: thierry.klein@math.univ-toulouse.fr
** Postal address: Institut de Mathématiques de Toulouse (UMR CNRS 5219), Université Toulouse 2, 5 Allées Antonio Machado, F-31058 Toulouse, France.
*** Postal address: Institut de Mathématiques de Toulouse (UMR CNRS 5219), Université Paul Sabatier, 118 Route de Narbonne, F-31062 Toulouse, Cedex 9, France.

Our paper is organized as follows. In Section 2 we present the model and we state our main results (Theorems 1 and 2). 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 $N_n \to \infty$ as n goes to ∞. 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 $k \in [\![1, N_n]\!]$. 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

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

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 η_0 such that*

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

(A2) $\tilde{c}_2 \leq \sigma_{X_n} := \operatorname{var}(X_n)^{1/2} \leq c_2$,

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

(A4) $\tilde{c}_4 \leq \sigma_{Y_n} := \operatorname{var}(Y_n)^{1/2} \leq c_4$,

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

(A6) *the correlations* $r_n := \operatorname{cov}(X_n, Y_n)\sigma_{X_n}^{-1}\sigma_{Y_n}^{-1}$ *satisfy* $|r_n| \leq c_6 < 1$,

(A7) *for* $Y_n' := Y_n - \mathbb{E}[Y_n] - \operatorname{cov}(X_n, Y_n)\sigma_{X_n}^{-2}(X_n - \mathbb{E}[X_n])$, *for all* $s \in [-\pi\,\pi]$,*and for all* $t \in [-\eta_0\,, \eta_0]$,

$$|\mathbb{E}[e^{i(sX_n+tY_n')}]| \leq 1 - c_7(\sigma_{X_n}^2 s^2 + \sigma_{Y_n'}^2 t^2).$$

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 of [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 centred variable uncorrelated with X_n. If (X, Y') is a vector of centred and uncorrelated random variables, then

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

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

Proposition 1. *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* $n \to \infty$*, and that, for every fixed* $r > 0$*,*

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

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 of [8].

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

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

where Φ *denotes the standard normal cumulative distribution function and* C *is a positive constant that depends only on* $\tilde{c}_2$, c_2, c_3, $\tilde{c}_4$, c_4, c_5, c_6, c_7, η_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 2. *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* $N_n \ge 3$,

$$|\mathbb{E}[U_n] - N_n\mathbb{E}[Y_n] - r_n\sigma_{Y_n}\sigma_{X_n}^{-1}(m_n - N_n\mathbb{E}[X_n])| \le d_1 \tag{2}$$

and

$$|\mathrm{var}\,(U_n) - N_n\tau_n^2| \le d_2 N_n^{1/2}. \tag{3}$$

Theorem 2. *Under Assumption 1, we have*

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left(\frac{U_n - \mathbb{E}[U_n]}{\mathrm{var}\,(U_n)^{1/2}} \le x \right) - \Phi(x) \right| \le \frac{\widetilde{C}}{N_n^{1/2}}, \tag{4}$$

where $\widetilde{C}$ *is a constant that depends only on* $\tilde{c}_2$, c_2, c_3, $\tilde{c}_4$, c_4, c_5, c_6, c_7, η_0, *and* c_1.

Furthermore, as in [8], the results of Theorems 1 and 2 simplify considerably in the special case when the vector (X_n, Y_n) does not depend on n, that is, when we consider an i.i.d. sequence instead of a triangular array. This is a consequence of Proposition 1.

3. Classical examples

In this section we describe the examples mentioned in [6] and [8]. Each example satisfies the assumptions of Proposition 1, as shown in [8], leading to a Berry–Esseen inequality.

https://doi.org/10.1017/jpr.2019.7 Published online by Cambridge University Press

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 $U = \sum_{i=1}^N \mathbf{1}_{\{Z_i=0\}}$, which is distributed as $\sum_{i=1}^N \mathbf{1}_{\{X_i=0\}}$ conditioned on $\{\sum_{i=1}^N X_i = m\}$. Now, if $m = m_n \to \infty$ and $N = N_n \to \infty$ with $m_n/N_n \to \lambda \in (0, \infty)$, we can take $X_n \sim \mathcal{P}(\lambda_n)$ with $\lambda_n := m_n/N_n$, $Y_n = \mathbf{1}_{\{X_n=0\}}$, and apply Proposition 1 to obtain a Berry–Esseen inequality for $U_n = \sum_{i=1}^{N_n} \mathbf{1}_{\{Z_i=0\}}$.

Remark 1. In [17], the authors proved 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 below to prove our results.

Remark 2. 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 1 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 kth 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 $m = n$, $N = N_n \to \infty$ with $N_n/n \to p$, 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 ith 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, for all $k \in \{0, \dots, N-1\}$,

$$S_k \geq k \quad \text{and} \quad S_N = N - 1. \tag{5}$$

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 proved that $X_1, \dots, X_N$ have the same distribution conditioned on (5) as conditioned on $\{S_N = N - 1\}$. Applying Proposition 1 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 $f(x) = \mathbf{1}_{\{x=3\}}$, 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 labelled rooted forest with m vertices and N roots with $N < m$. 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 [8], 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 the random variables X_i are i.i.d. and Borel distributed for any arbitrary parameter $\mu \in (0, 1)$, i.e.

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

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

3.5. Hashing with linear probing

Hashing with linear probing is a classical model in theoretical computer science that appeared in the 1960s. It was first studied from a mathematical point of view in [10]. For more details on the model, we refer the reader to [1], [2], [5], [7], [9], and [14]. 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 by $d_{m,n}$. 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 [5] and [11], Janson [7] proved that the lengths of the blocks 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 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 $m = m_n \to \infty$ and $N = N_n = m_n - n \to \infty$ with $n/m_n \to \mu \in (0, 1)$, 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 1.

4. Proofs

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

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

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

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

https://doi.org/10.1017/jpr.2019.7 Published online by Cambridge University Press

so (A4) is satisfied by Y'_n. Finally, by the Minkowski inequality, (A3), (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,

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

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 1 and 2 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 $\mathbb{E}[Y_n] = \mathbb{E}[X_nY_n] = 0$, $r_n = 0$ 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 1 and 2 intensively rely on the use of Fourier transforms through the functions φ_n and ψ_n defined by

$$\varphi_n(s,t) := \mathbb{E}[\exp\{\mathrm{i}s(X_n - \mathbb{E}[X_n]) + \mathrm{i}tY_n\}], \qquad \psi_n(t) := 2\pi\mathbb{P}(S_n = m_n)\mathbb{E}[\exp\{\mathrm{i}tU_n\}]. \tag{6}$$

The controls of these functions (respectively the controls of their derivatives) needed in the proofs are postponed to Lemmas 1 and 2 in Section 4.4 (respectively Lemma 3). In particular, (15)–(18) will be used several times below.

4.1. Proof of Theorem 1

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

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

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

$$\eta := \min\left(\tfrac{2}{9}(c_4c_5)^{-1}, \eta_0\right) > 0. \tag{7}$$

https://doi.org/10.1017/jpr.2019.7 Published online by Cambridge University Press

From (15) of Lemma 1 and using Taylor's expansion,

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

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

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

where

$$\varepsilon := \min\left(\tfrac{2}{9}(c_2c_3)^{-1},\, \pi\right), \tag{8}$$

and decompose

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

where

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

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

Lemmas 5 and 6 state that there exists positive constants C_1 and C_2, depending only on $\tilde{c}_2$, c_2, c_3, c_5, c_7, and c_1 such that, for $N_n \ge \max(12^3c_3^2, 12^3c_5^2, 2)$,

$$\int_0^{\eta\sigma_{Y_n}N_n^{1/2}}\sup_{0\le\theta\le u} I_1(n,u,\theta)\,\mathrm{d}u \le \frac{C_1}{N_n^{1/2}}, \tag{11}$$

and

$$\int_0^{\eta\sigma_{Y_n}N_n^{1/2}}\sup_{0\le\theta\le u} I_2(n,u,\theta)\,\mathrm{d}u \le \frac{C_2}{N_n^{1/2}}. \tag{12}$$

So,

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

with

$$C := \max\left(\left(C_1 + C_2 + \frac{24}{\widetilde{c_4}\pi\sqrt{2\pi}}\left(\min\left(\frac{2}{9}c_4c_5,\, \eta_0\right)\right)\right)^{-1},\, 12^{3/2}c_3,\, 12^{3/2}c_5,\, \sqrt{2}\right).$$

4.2. Proof of Proposition 2

Proof of (2). We adapt the proof given in [8]. Using the definition of Ψ_n given in (6), and differentiating under the integral sign of (15) of Lemma 1, we naturally have

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

Using (18) of Lemma 3 with $t=0$, (A2), (A3), and (A5), we deduce that

$$\left|\frac{\partial\varphi_n}{\partial t}\left(\frac{s}{\sigma_{X_n}N_n^{1/2}},0\right)\right| \le \frac{s^2}{2}\frac{\rho_{Y_n}^{1/3}\rho_{X_n}^{2/3}}{\sigma_{X_n}^2N_n} \le \frac{c_3^{2/3}c_4c_5^{1/3}}{2N_n}s^2.$$

Then, using (16) of Lemma 2 (with $l=1$ and $t=0$) and for $N_n \ge 3$,

$$\begin{aligned}&\int_{-\pi\sigma_{X_n}N_n^{1/2}}^{\pi\sigma_{X_n}N_n^{1/2}}\left|\frac{\partial\varphi_n}{\partial t}\left(\frac{s}{\sigma_{X_n}N_n^{1/2}},0\right)\right|\left|\varphi_n^{N_n-1}\left(\frac{s}{\sigma_{X_n}N_n^{1/2}},0\right)\right|\,\mathrm{d}s \\ &\qquad\le \frac{c_3^{2/3}c_4c_5^{1/3}}{2N_n}\int_{-\infty}^{+\infty}s^2\mathrm{e}^{-2c_7s^2/3}\,\mathrm{d}s.\end{aligned}$$

So, (2) holds with

$$d_1 := 2^{-1}c_3^{2/3}c_4c_5^{1/3}c_1^{-1}\int_{-\infty}^{+\infty}s^2\mathrm{e}^{-2c_7s^2/3}\,\mathrm{d}s.$$

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 before,

$$\begin{aligned}&\mathbb{E}[U_n^2] \\ &= \frac{-\psi_n''(0)}{2\pi\mathbb{P}(S_n = m_n)} \\ &= -\gamma_n^{-1}N_n(N_n-1)\int_{-\pi\sigma_{X_n}N_n^{1/2}}^{\pi\sigma_{X_n}N_n^{1/2}}\mathrm{e}^{-isv_n}\left(\frac{\partial\varphi_n}{\partial t}\left(\frac{s}{\sigma_{X_n}N_n^{1/2}},0\right)\right)^2\varphi_n^{N_n-2}\left(\frac{s}{\sigma_{X_n}N_n^{1/2}},0\right)\mathrm{d}s \qquad (13) \\ &\quad - \gamma_n^{-1}N_n\int_{-\pi\sigma_{X_n}N_n^{1/2}}^{\pi\sigma_{X_n}N_n^{1/2}}\mathrm{e}^{-isv_n}\frac{\partial^2\varphi_n}{\partial t^2}\left(\frac{s}{\sigma_{X_n}N_n^{1/2}},0\right)\varphi_n^{N_n-1}\left(\frac{s}{\sigma_{X_n}N_n^{1/2}},0\right)\mathrm{d}s, \qquad (14)\end{aligned}$$

where

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

First, by (18) of Lemma 3 with $t=0$ and by (16) of Lemma 2 (with $l=2$ and $t=0$), we have, for $N_n \geq 3$,

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

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

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

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

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

is bounded by some $d_2'''/N_n^{1/2}$. By (15) with $t=0$,

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

so

$$\Delta_n = \gamma_n^{-1}\int_{-\pi\sigma_{X_n}N_n^{1/2}}^{\pi\sigma_{X_n}N_n^{1/2}} \mathrm{e}^{-\mathrm{i}sv_n}\left(\frac{\partial^2\varphi_n}{\partial t^2}\left(\frac{s}{\sigma_{X_n}N_n^{1/2}}, 0\right) + \sigma_{Y_n}^2\varphi_n\left(\frac{s}{\sigma_{X_n}N_n^{1/2}}, 0\right)\right)$$
$$\times\,\varphi_n^{N_n-1}\left(\frac{s}{\sigma_{X_n}N_n^{1/2}}, 0\right)\mathrm{d}s$$
$$= \gamma_n^{-1}\int_{-\pi\sigma_{X_n}N_n^{1/2}}^{\pi\sigma_{X_n}N_n^{1/2}} \mathrm{e}^{-\mathrm{i}sv_n}\mathbb{E}[Y_n{}^2 f(s)]\varphi_n^{N_n-1}\left(\frac{s}{\sigma_{X_n}N_n^{1/2}}, 0\right)\mathrm{d}s,$$

where

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

https://doi.org/10.1017/jpr.2019.7 Published online by Cambridge University Press

Applying Taylor's theorem yields

$$|f(s)| \le |s| \sup_u \bigg| -\mathrm{i}\frac{X_n - \mathbb{E}[X_n]}{\sigma_{X_n} N_n^{1/2}} \mathrm{e}^{\mathrm{i}u\sigma_{X_n}^{-1} N_n^{-1/2}(X_n - \mathbb{E}[X_n])} + \mathbb{E}\bigg[\mathrm{i}\frac{X_n - \mathbb{E}[X_n]}{\sigma_{X_n} N_n^{1/2}} \mathrm{e}^{\mathrm{i}u\sigma_{X_n}^{-1} N_n^{-1/2}(X_n - \mathbb{E}[X_n])}\bigg]\bigg|$$

$$\le \frac{|s|}{N_n^{1/2}}\bigg(\bigg|\frac{X_n - \mathbb{E}[X_n]}{\sigma_{X_n}}\bigg| + \mathbb{E}\bigg[\bigg|\frac{X_n - \mathbb{E}[X_n]}{\sigma_{X_n}}\bigg|\bigg]\bigg).$$

Thus, using Hölder's inequality,

$$|\mathbb{E}[{Y_n}^2 f(s)]| \le \frac{|s|}{N_n^{1/2}} \mathbb{E}\bigg[{Y_n}^2\bigg(\bigg|\frac{X_n - \mathbb{E}[X_n]}{\sigma_{X_n}}\bigg| + \mathbb{E}\bigg[\bigg|\frac{X_n - \mathbb{E}[X_n]}{\sigma_{X_n}}\bigg|\bigg]\bigg)\bigg]$$

$$\le \frac{\sigma_{Y_n}^2 |s|}{N_n^{1/2}}\bigg(\frac{\rho_{Y_n}^{2/3}}{\sigma_{Y_n}^2}\frac{\rho_{X_n}^{1/3}}{\sigma_{X_n}} + 1\bigg)$$

$$\le \frac{|s| c_4^2}{N_n^{1/2}}(c_5^{2/3} c_3^{1/3} + 1),$$

where the last inequality is obtained using (A2)–(A5). Now, by (A1) and the upper bound in (16) (with $l = 1$ and $t = 0$), we get, for $N_n \ge 3$,

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

with

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

Finally,

$$|\mathrm{var}\,(U_n) - N_n \sigma_{Y_n}^2| \le (d_1^2 + d_2'' + d_2''') N_n^{1/2} =: d_2 N_n^{1/2}.$$

The proof of (3) is complete.

4.3. Proof of Theorem 2

Write

$$\bigg|\mathbb{P}\bigg(\frac{U_n - \mathbb{E}[U_n]}{\mathrm{var}\,(U_n)^{1/2}} \le x\bigg) - \Phi(x)\bigg|$$

$$\le \bigg|\mathbb{P}\bigg(\frac{U_n}{N_n^{1/2}\sigma_{Y_n}} \le a_n x + b_n\bigg) - \Phi(a_n x + b_n)\bigg| + |\Phi(a_n x + b_n) - \Phi(x)|,$$

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 var (U_n) yield

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

Then, for $N_n^{1/2} \ge 2\tilde{c}_4^{-2} d_2$, $a_n \ge \frac{1}{2}$ and applying Taylor's theorem to Φ, we obtain

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

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

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

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

4.4. Technical results

Recall that

$$\nu_n = \frac{(m_n - N_n \mathbb{E}[X_n])}{(\sigma_{X_n} N_n^{1/2})} \quad \text{and} \quad \gamma_n = 2\pi\, \mathbb{P}(S_n = m_n) \sigma_{X_n} N_n^{1/2}.$$

Moreover,

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

Lemma 1. *We have*

$$\psi_n(t) = \frac{1}{\sigma_{X_n} N_n^{1/2}} \int_{-\pi \sigma_{X_n} N_n^{1/2}}^{\pi \sigma_{X_n} N_n^{1/2}} \mathrm{e}^{-\mathrm{i}s\nu_n} \varphi_n^{N_n}\left(\frac{s}{\sigma_{X_n} N_n^{1/2}}, t\right) \mathrm{d}s. \tag{15}$$

Proof. Indeed, since

$$\int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}s(S_n - m_n)}\, \mathrm{d}s = 2\pi \mathbf{1}_{\{S_n = m_n\}},$$

we have

$$\begin{aligned} \psi_n(t) &= 2\pi\, \mathbb{P}(S_n = m_n) \mathbb{E}[\exp\{\mathrm{i}tU_n\}] \\ &= 2\pi\, \mathbb{E}[\exp\{\mathrm{i}tT_n\} \mathbf{1}_{\{S_n = m_n\}}] \\ &= \int_{-\pi}^{\pi} \mathbb{E}[\exp\{\mathrm{i}s(S_n - m_n) + \mathrm{i}tT_n\}]\, \mathrm{d}s \\ &= \int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}s(m_n - N_n \mathbb{E}[X_n])} \varphi_n^{N_n}(s, t)\, \mathrm{d}s, \end{aligned}$$

which leads to (15) after the change of variable $s' = s\sigma_{X_n} N_n^{1/2}$. □

Now we give controls on the function φ_n and its partial derivatives (see Lemmas 2 and 3).

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

$$\left| \varphi_n^{N_n - l}\left(\frac{s}{\sigma_{X_n} N_n^{1/2}}, \frac{t}{\sigma_{Y_n} N_n^{1/2}} \right) \right| \leq \mathrm{e}^{-(s^2+t^2)\cdot c_7 \cdot (N_n - l)/N_n}. \tag{16}$$

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

Lemma 3. *For any s and t, we have*

$$\left| \frac{\partial \varphi_n}{\partial t}\left(\frac{s}{\sigma_{X_n} N_n^{1/2}}, \frac{t}{\sigma_{Y_n} N_n^{1/2}} \right) \right| \leq \frac{\sigma_{Y_n}}{N_n^{1/2}}(|s| + |t|) \tag{17}$$

and

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

$$\leq \frac{\sigma_{Y_n}}{N_n^{1/2}}|t| + \frac{\sigma_{Y_n}}{N_n}\left[\frac{s^2}{2}\left(\frac{\rho_{X_n}}{\sigma_{X_n}^3}\right)^{2/3}\left(\frac{\rho_{Y_n}}{\sigma_{Y_n}^3}\right)^{1/3} + |st|\left(\frac{\rho_{X_n}}{\sigma_{X_n}^3}\right)^{1/3}\left(\frac{\rho_{Y_n}}{\sigma_{Y_n}^3}\right)^{2/3} + \frac{t^2}{2}\left(\frac{\rho_{Y_n}}{\sigma_{Y_n}^3}\right) \right].$$

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 obtain (17) 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 (18) using

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

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

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

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

Lemma 4. *Define*

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

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

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

https://doi.org/10.1017/jpr.2019.7 Published online by Cambridge University Press

we have

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

with $C_4 := 161$.

Remark 3. We make explicit the constant C_4 appearing at the end of the proof of Lemma 2 of [17]. For all v and s in R_2 as defined in [17], we have

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

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 $\sigma_{X_n}N_n^{1/2} \leq c_2c_3 l_{1,n}^{-1}$. 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 $\sigma_{Y_n}N_n^{1/2} \leq c_4c_5 l_{2,n}^{-1}$. Now we are able to establish (11).

Lemma 5. *There exists a positive constant* C_1*, depending only on* c_3*,* c_5*, and* c_1 *such that, for* $N_n \geq 12^3 \max(c_3^2, c_5^2)$*,*

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

Proof. The definitions of η in (7) and ε in (8) imply that, for $s \in A_1$ and u and θ as in the integral in the statement above, we have

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

which ensures that $(s,\theta) \in R$ as specified in Lemma 4. 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 4 in (9) and by (A1), we get

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

and the result follows with

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

□

Remark 4. Actually, Lemma 5 is valid as soon as $N_n \geq \max(c_3^2, c_5^2)$: the constants in the proof of Lemma 2 of [17] can be improved.

Now we are able to prove (12).

Lemma 6. *There exists a positive constant C_2, depending only on c_1, $\tilde{c}_2$, c_2, c_3, and c_7 such that, for $N_n \geq 2$,*

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

Proof. We use the controls (16) with $t=\theta$ and $l=1$, (17), and $|\varphi_n| \leq 1$ to get

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

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

$$\begin{aligned}
&\int_0^{\eta\sigma_{Y_n}N_n^{1/2}} \sup_{0\leq\theta\leq u} I_2(n, u, \theta)\,\mathrm{d}u \\
&\quad \leq 2\gamma_n^{-1}\int_0^{+\infty}\int_{\varepsilon\sigma_{X_n}N_n^{1/2}}^{+\infty} \sup_{0\leq\theta\leq u}\left[(s+2\theta)\exp\left(\frac{\theta^2}{2}\left(1-2c_7\frac{N_n-1}{N_n}\right)\right)\right] \\
&\qquad \times \mathrm{e}^{-u^2/2-s^2\cdot c_7(N_n-1)/N_n}\,\mathrm{d}s\,\mathrm{d}u \\
&\quad \leq 2c_1^{-1}\int_0^{+\infty}\int_{\varepsilon\sigma_{X_n}N_n^{1/2}}^{+\infty}(s+2u)\mathrm{e}^{-\min(1,c_7)u^2/2-s^2c_7/2}\,\mathrm{d}s\,\mathrm{d}u \\
&\quad \leq \mathrm{e}^{-N_nc_7\varepsilon^2\sigma_{X_n}^2/2}\left(\frac{c_1^{-1}c_7^{-1}\sqrt{2\pi}}{\sqrt{\min(1, c_7)}} + \frac{4c_1^{-1}}{\min(1, c_7)}\frac{1}{c_7\varepsilon\sigma_{X_n}N_n^{1/2}}\right) \\
&\quad \leq C_2'\mathrm{e}^{-C_2''N_n},
\end{aligned}$$

where

$$C_2' := c_1^{-1}c_7^{-1}\left(\frac{\sqrt{2\pi}}{\sqrt{\min(1, c_7)}} + \frac{4}{\min(1, c_7)\min((2/9)(c_2c_3)^{-1}, \pi)\tilde{c}_2}\right)$$

and

$$C_2'' := \tilde{c}_2^2/2c_7\min\left(\tfrac{2}{9}(c_2c_3)^{-1}, \pi\right)^2.$$

https://doi.org/10.1017/jpr.2019.7 Published online by Cambridge University Press

The result follows, writing

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

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

References

[1] CHASSAING, P. AND FLAJOLET, P. (2003). Hachage, arbres, chemins & graphes. *Gaz. Math.*, 29–49.

[2] CHASSAING, P. AND LOUCHARD, G. (2002). Phase transition for parking blocks, Brownian excursion and coalescence. *Random Struct. Algorithms* **21**, 76–119.

[3] FELLER, W. (1968). *An Introduction to Probability Theory and its Applications*, Vol. I, 3rd edn. John Wiley, New York.

[4] FELLER, W. (1971). *An Introduction to Probability Theory and its Applications*, Vol. II, 2nd edn. John Wiley, New York.

[5] FLAJOLET, P., POBLETE, P. AND VIOLA, A. (1998). On the analysis of linear probing hashing. *Algorithmica* **22**, 490–515.

[6] HOLST, L. (1979). Two conditional limit theorems with applications. *Ann. Statist.* **7**, 551–557.

[7] JANSON, S. (2001). Asymptotic distribution for the cost of linear probing hashing. *Random Struct. Algorithms* **19**, 438–471.

[8] JANSON, S. (2001). Moment convergence in conditional limit theorems. *J. Appl. Prob.* **38**, 421–437.

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

[10] KNUTH, D. E. (1974). Computer science and its relation to mathematics. *Amer. Math. Monthly* **81**, 323–343.

[11] KNUTH, D. E. (1998). Linear probing and graphs. *Algorithmica* **22**, 561–568.

[12] KOLCHIN, V. F. (1986). *Random Mappings*, Translation Series in Mathematics and Engineering. Optimization Software, Inc., Publications Division, New York.

[13] LOÈVE, M. (1955). *Probability Theory: Foundations, Random Sequences*. Van Nostrand, New York.

[14] MARCKERT, J.-F. (2001). Parking with density. *Random Struct. Algorithms* **18**, 364–380.

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

[16] PAVLOV, Y. L. (1997). Random forests. In *Probabilistic Methods in Discrete Mathematics (Petrozavodsk, 1996)*, VSP, Utrecht, pp. 11–18.

[17] QUINE, M. P. AND ROBINSON, J. (1982). A Berry–Esseen bound for an occupancy problem. *Ann. Prob.* **10**, 663–671.

[18] WENDEL, J. G. (1975). Left-continuous random walk and the Lagrange expansion. *Amer. Math. Monthly* **82**, 494–499.