MOMENT CONVERGENCE IN CONDITIONAL LIMIT THEOREMS

SVANTE JANSON

ABSTRACT. Consider a sum $\sum_1^N Y_i$ of random variables conditioned on a given value of the sum $\sum_1^N X_i$ of some other variables, where X_i and Y_i are dependent but the pairs (X_i, Y_i) form an i.i.d. sequence.

We consider here the case when each X_i is discrete. We prove, for a triangular array $((X_{ni}, Y_{ni}))$ of such pairs satisfying certain conditions, both convergence of the distribution of the conditioned sum (after suitable normalization) to a normal distribution, and convergence of its moments.

The results are motivated by an application to hashing with linear probing; we give also some other applications to occupancy problems, random forests, and branching processes.

1. Introduction

Many random variables arising in different areas of probability theory, combinatorics and statistics turn out to have the same distribution as a sum of independent random variables conditioned on a specific value of another such sum. More precisely, we are concerned with variables with the distribution of $\sum_1^N Y_i$ conditioned on $\sum_1^N X_i$, where the pairs (X_i, Y_i) form a sequence of i.i.d. random vectors; in general, the variables X_i can be discrete or continuous, or even vector-valued. Some examples are given in Section 3 below, and many others in, for example, [5, 6, 7, 11, 12].

General limit theorems yielding the asymptotic behaviour of the conditioned sum under suitable assumptions are given by, among others, Steck [21], Holst [5, 7] and Kudlaev [13]. (See also results on asymptotic expansions in [2, 14, 4, 8] and on rates of convergence in [15].)

The purpose of this paper is to give a set of sufficient conditions for convergence of not only the distribution of the conditioned sum but also its moments. We consider only the case when the variables X_i are discrete; we expect that extensions to the continuous case are possible but not trivial. For simplicity we further only consider the case when the limit distribution is normal. (Asymptotics of the mean and variance in a case overlapping with ours has earlier been studied by Swensen [22] in a somewhat different formulation. The mean, in the equivalent form $\mathbb{E}(Y_1 \mid \sum_1^N X_i)$, has been studied by Portnoy [19] and Zabell [25, 26].)

Date: May 8, 2000; revised May 2, 2001.

1991 *Mathematics Subject Classification.* 60F05.

Key words and phrases. Conditional distribution, limit theorem, moment convergence, occupancy, hashing, random forest, branching process.

The theorem is stated, in several versions, in Section 2 and proved in Section 4. The results in this paper were motivated by an application to hashing with linear probing. This application is described in Section 3 (Example 3.2) together with some other applications.

2. Main results

We state our main result in the following form. The large number of conditions in the theorem is perhaps not so elegant, but we have chosen to state a version that is easy to apply; all conditions are easily verified in the applications of interest to us. Some alternative versions of the conditions are discussed in the remarks after the statement.

We use the notation $\sigma_X^2 := \operatorname{Var} X$ for a random variable X.

Theorem 2.1. (a) *Suppose that, for each n, (X_n, Y_n) is a pair of random variables such that X_n is integer valued, and that N_n and m_n are integers. Suppose further that for some γ and c (independent of n), with $0 < \gamma \le 2$ and $c > 0$, the following hold, where all limits are taken as $n \to \infty$:*

(i) $\mathbb{E}\, X_n = m_n/N_n$.
(ii) $0 < \sigma_{X_n}^2 < \infty$.
(iii) $\mathbb{E}\,|X_n - \mathbb{E}\, X_n|^3 = o(N_n^{1/2}\sigma_{X_n}^3)$.
(iv) $\sigma_{X_n}^2 = O(N_n^{2/\gamma-1})$.
(v) $\varphi_{X_n}(s) := \mathbb{E}\, e^{isX_n}$ *satisfies* $1-|\varphi_{X_n}(s)| \ge c\min(|s|^\gamma, s^2\sigma_{X_n}^2)$ *for* $|s| \le \pi$.
(vi) $0 < \sigma_{Y_n}^2 < \infty$.
(vii) $\mathbb{E}\,|Y_n - \mathbb{E}\, Y_n|^3 = o(N_n^{1/2}\sigma_{Y_n}^3)$.
(viii) *The correlation* $\rho_n := \operatorname{Cov}(X_n, Y_n)/\sigma_{X_n}\sigma_{Y_n}$ *satisfies* $\limsup |\rho_n| < 1$.

Let (X_{ni}, Y_{ni}), $i = 1, 2, \dots$, be i.i.d. copies of (X_n, Y_n), and let $S_{nN} := \sum_1^N X_{ni}$, $T_{nN} := \sum_1^N Y_{ni}$ and $\tau_n^2 := \sigma_{Y_n}^2(1-\rho_n^2) = \sigma_{Y_n}^2 - \operatorname{Cov}(X_n, Y_n)^2/\sigma_{X_n}^2$. Then, as $n \to \infty$, the conditional distribution of $(T_{nN_n} - N_n\,\mathbb{E}\, Y_n)/N_n^{1/2}\tau_n$ given $S_{nN_n} = m_n$ converges to a standard normal distribution. In other words, if U_n is a random variable whose distribution equals the conditional distribution of T_{nN_n} given $S_{nN_n} = m_n$, then

$$\frac{U_n - N_n\,\mathbb{E}\, Y_n}{N_n^{1/2}\tau_n} \xrightarrow{\mathrm{d}} N(0,1). \tag{2.1}$$

Moreover, $\mathbb{E}\, U_n = N_n\,\mathbb{E}\, Y_n + o(N_n^{1/2}\tau_n)$ *and* $\operatorname{Var} U_n \sim N_n\tau_n^2$, *and thus also*

$$\frac{U_n - \mathbb{E}\, U_n}{(\operatorname{Var} U_n)^{1/2}} \xrightarrow{\mathrm{d}} N(0,1). \tag{2.2}$$

(b) *Assume furthermore that for each even integer $r \ge 4$*

(ix) $\mathbb{E}(X_n - \mathbb{E}\, X_n)^r = O(N_n^{r/2-1}\sigma_{X_n}^r)$,
(x) $\mathbb{E}(Y_n - \mathbb{E}\, Y_n)^r = O(N_n^{r/2-1}\sigma_{Y_n}^r)$.

Then the limits (2.1) *and* (2.2) *hold with convergence of all moments too.*

Remark 2.1. Note that using the Lebesgue space norm $\|X\|_r = (\mathbb{E}\,|X|^r)^{1/r}$, the assumptions (iii), (vii), (ix) and (x) can be written

(iii′) $\|X_n - \mathbb{E}\, X_n\|_3 = o(N_n^{1/6}\sigma_{X_n})$,
(vii′) $\|Y_n - \mathbb{E}\, Y_n\|_3 = o(N_n^{1/6}\sigma_{Y_n})$,
(ix′) $\|X_n - \mathbb{E}\, X_n\|_r = O(N_n^{1/2-1/r}\sigma_{X_n})$,
(x′) $\|Y_n - \mathbb{E}\, Y_n\|_r = O(N_n^{1/2-1/r}\sigma_{Y_n})$.

Since Minkowski's and Hölder's inequalities yield $\|X_n - \mathbb{E}\, X_n\|_3 \le \|X_n\|_3 + |\mathbb{E}\, X_n| \le 2\|X_n\|_3$, it is for (iii) sufficient that $\|X_n\|_3 = o(N_n^{1/6}\sigma_{X_n})$, or $\mathbb{E}\,|X_n|^3 = o(N_n^{1/2}\sigma_{X_n}^3)$. More generally, it is sufficient that $\mathbb{E}\,|X_n - a_n|^3 = o(N_n^{1/2}\sigma_{X_n}^3)$ for some sequence of constants a_n. The same applies, *mutatis mutandis,* to (vii), (ix) and (x).

In other words, the theorem holds also if $\mathbb{E}\, X_n$ and $\mathbb{E}\, Y_n$ in (iii), (vii), (ix) and (x) are omitted or replaced by any other numbers.

Remark 2.2. It is immediate that (viii) is equivalent to $\liminf \tau_n^2/\sigma_{Y_n}^2 > 0$, and thus also to

(viii′) $\tau_n^2 = \Theta(\sigma_{Y_n}^2)$.

(This notation means that $\limsup \tau_n^2/\sigma_{Y_n}^2$ and $\limsup \sigma_{Y_n}^2/\tau_n^2$ both are finite.) In cases where this fails, one might apply the theorem to a suitable modification $Y_n + a_n X_n + b_n$ instead, for example to the projection Y_n' in (4.1); note that this changes U_n by the constant $a_n m_n + b_n N_n$ only.

Remark 2.3. τ_n^2 equals $\min\{\mathrm{Var}(Y_n - aX_n) : a \in \mathbb{R}\}$ (the residual variance in linear regression). Hence τ_n^2 is unchanged if Y_n is replaced by $Y_n + a_n X_n + b_n$ for any real constants a_n and b_n.

Remark 2.4. We consider here, for simplicity, only the central case when (i) holds; equivalently, $m_n = \mathbb{E}\, S_{nN_n}$ and thus we condition on $S_{nN_n} = \mathbb{E}\, S_{nN_n}$. In applications we usually take X_n from an exponential family and adjust the parameter so that (i) holds. Nevertheless, it is interesting to consider other cases too, and we give a generalization in Theorem 2.3 below.

Remark 2.5. The conditions (iv) and (v) are connected through the choice of γ. They are used in the proof only to provide suitable dominations and may be combined and weakened, for example to

$$|\varphi_{X_n}(s)| \le 1 - \frac{2}{N_n}\ln(N_n\sigma_{X_n}^2 s^2) + \frac{c}{N_n}, \qquad |s| \le \pi, \tag{2.3}$$

but the version given above seems more convenient in our applications. Condition (v) (or (2.3)) is the most technical and least intuitive of our conditions, and it would be interesting to find other conditions not explicitly involving the characteristic function that could replace it. (See Corollary 2.1 for a simple case.) Unfortunately, the condition cannot be eliminated completely, as is shown by Example 3.7.

Remark 2.6. The powers of N_n in (iii), (vii), (ix) and (x) may look strange at first, but the conditions turn out to be sharp in many cases; moreover, the o in (iii) and (vii) cannot be replaced by O. See Examples 3.1, 3.2, 3.5 and 3.6. (There may be room for improvements of (ix), however.)

Remark 2.7. It follows from the proof below that if the extra conditions in (b) are satisfied for a specific even $r \geq 4$ only, then all moments of order less than r converge; moreover, a minor modification of the proof shows that if the O in (ix) and (x) is strengthened to o, then also the r:th moment converges. (It is shown in Example 3.5 that (ix) and (x) are not sufficient for convergence of the r:th moment.) The same holds for odd $r \geq 3$ if we use absolute moments in (ix) and (x).

The result simplifies considerably in the special case when (X_n, Y_n) does not depend on n, so we consider a single sequence instead of a triangular array. This is included in the following, more general, corollary.

Corollary 2.1. *Suppose that* $\big(X_n, Y_n\big) \overset{\mathrm{d}}{\to} (X, Y)$ *as* $n \to \infty$, *and that, for every fixed* $r > 0$, $\limsup_{n\to\infty} \mathbb{E}\,|X_n|^r < \infty$ *and* $\limsup_{n\to\infty} \mathbb{E}\,|Y_n|^r < \infty$. *Suppose further that the distribution of* X *has span* 1, *and that* Y *is not a.s. equal to a linear function* $c + dX$.

If m_n *and* N_n *are integers such that* $\mathbb{E}\,X_n = m_n/N_n$ *and* $N_n \to \infty$, *then all conclusions of Theorem 2.1 hold, with* $\tau_n^2 \to \operatorname{Var} Y - \operatorname{Cov}(X, Y)^2 / \operatorname{Var} X > 0$.

Remark 2.8. The assumptions of Corollary 2.1 imply that $m_n/N_n \to \mathbb{E}\,X$. Conversely, in the important case of an exponential family of distributions, when the set of possible values of $\mathbb{E}\,X$ is an interval I, under weak additional assumptions, the corollary applies when $\lim m_n/N_n$ exists and lies in the interior of I.

The results above easily extend to joint convergence for several random variables conditioned on the same sum. Note that the conditions on X_n in the next theorem, (i)–(v) and (ix), are exactly the same as in Theorem 2.1, while we have chosen a slightly different formulation of the conditions for Y_n, involving somewhat arbitrary normalization constants b_{nj}. This version is sometimes more convenient even in the case of a single variable Y_n; for example, in the situation in Corollary 2.1, we may take $b_{n1} = 1$.

Theorem 2.2. (a) *Suppose that, for each* n, (X_n, Y_n) *is a pair of random variables such that* X_n *is integer valued and* $Y_n = (Y_n^{(1)}, \dots, Y_n^{(l)})$ *is a random vector* (l *does not depend on* n). *Suppose further that* N_n *and* m_n *are integers and* b_{nj}, $j = 1, \dots, l$, *are positive real numbers, and that for some* γ *and* c (*independent of* n), *with* $0 < \gamma \leq 2$ *and* $c > 0$, *the following hold, where all limits are taken as* $n \to \infty$:

(i) $\mathbb{E}\,X_n = m_n/N_n$.
(ii) $0 < \sigma_{X_n}^2 < \infty$.
(iii) $\mathbb{E}\,|X_n - \mathbb{E}\,X_n|^3 = o(N_n^{1/2} \sigma_{X_n}^3)$.
(iv) $\sigma_{X_n}^2 = O(N_n^{2/\gamma - 1})$.
(v) $\varphi_{X_n}(s) := \mathbb{E}\,e^{isX_n}$ *satisfies* $1 - |\varphi_{X_n}(s)| \geq c \min(|s|^\gamma, s^2 \sigma_{X_n}^2)$ *for* $|s| \leq \pi$.
(vi) $\operatorname{Var} Y_n^{(j)} = O(b_{nj}^2)$.
(vii) $\mathbb{E}\,|Y_n^{(j)} - \mathbb{E}\,Y_n^{(j)}|^3 = o(N_n^{1/2} b_{nj}^3)$.

(viii) *For some numbers* σ_{jk}, $1 \le j,k \le l$,

$$b_{nj}^{-1} b_{nl}^{-1} \left(\operatorname{Cov}(Y_n^{(j)}, Y_n^{(l)}) - \frac{\operatorname{Cov}(Y_n^{(j)}, X_n)\operatorname{Cov}(Y_n^{(l)}, X_n)}{\operatorname{Var} X_n} \right) \to \sigma_{jk}. \tag{2.4}$$

Let (X_{ni}, Y_{ni}), $i = 1, 2, \dots$, *be i.i.d. copies of* (X_n, Y_n), *and let* $S_{nN} := \sum_1^N X_{ni}$ *and* $T_{nN} := \sum_1^N Y_{ni}$. *If* $U_n = (U_n^{(1)}, \dots, U_n^{(l)})$ *is a random vector whose distribution equals the conditional distribution of* T_{nN_n} *given* $S_{nN_n} = m_n$, *then, as* $n \to \infty$,

$$\left(N_n^{-1/2} b_{nj}^{-1} (U_n^{(j)} - N_n \operatorname{\mathbb{E}} Y_n^{(j)})\right)_{j=1}^{l} \xrightarrow{\mathrm{d}} N(0, \Sigma) \tag{2.5}$$

and

$$\left(N_n^{-1/2} b_{nj}^{-1} (U_n^{(j)} - \operatorname{\mathbb{E}} U_n^{(j)})\right)_{j=1}^{l} \xrightarrow{\mathrm{d}} N(0, \Sigma), \tag{2.6}$$

where the covariance matrix $\Sigma := (\sigma_{jk})$. *Moreover, the mean and covariance matrix of the left hand sides converge to* 0 *and* Σ, *respectively.*

(b) *Assume furthermore that for each even integer* $r \ge 4$

(ix) $\mathbb{E}(X_n - \mathbb{E} X_n)^r = O(N_n^{r/2-1} \sigma_{X_n}^r)$,

(x) $\mathbb{E}(Y_n^{(j)} - \mathbb{E} Y_n^{(j)})^r = O(N_n^{r/2-1} b_{nj}^r)$.

Then the limits (2.5) *and* (2.6) *hold with convergence of all* (*mixed*) *moments.*

We can condense the conditions as follows (in a rather general special case).

Corollary 2.2. *Suppose that, for each* n, (X_n, Y_n) *is as in Theorem 2.2 and that each pair* $(X_n, Y_n^{(j)})$, $j = 1, \dots, l$, *satisfies the assumptions of Theorem 2.1* (*or Corollary 2.1*). *Suppose further that* b_{nj} *are positive numbers such that* (2.4) *holds. Then the conclusions of Theorem 2.2 hold.*

As said in Remark 2.4, it is also interesting to consider the case $m_n \ne \mathbb{E} S_{nN_n}$. In the central region $m_n = \mathbb{E} S_{nN_n} + O\big((\operatorname{Var} S_{nN_n})^{1/2}\big)$ the results above hold with only minor modifications; for simplicity we state only an extension of Theorem 2.1, leaving the special case in Corollary 2.1 and the vector-valued case to the reader. (We will not consider the case when m_n is in the more distant tails of the distribution of S_{nN_n}. See [25] for the expectation in such a situation.)

Theorem 2.3. *Let the conditions in Theorem 2.1 be satisfied except that* (i) *is replaced by*

(i′) $m_n = N_n \mathbb{E} X_n + O(N_n^{1/2} \sigma_{X_n})$.

Then (2.2) *still holds, with moment convergence as before if* (ix) *and* (x) *hold, but*

$$\mathbb{E} U_n = N_n \mathbb{E} Y_n + \rho_n \frac{\sigma_{Y_n}}{\sigma_{X_n}} (m_n - N_n \mathbb{E} X_n) + o(N_n^{1/2} \tau_n) \tag{2.7}$$

and thus (2.1) *has to be correspondingly modified.*

In particular, this implies results of the type of Swensen [22], under somewhat different conditions.

Remark 2.9. It follows from Theorem 2.3 that the conclusions of Theorem 2.1 are valid without modifications if (i) is weakened to (i″) $m_n = N_n \mathbb{E} X_n + o(N_n^{1/2}\sigma_{X_n})$.

The same holds for Corollary 2.1, Theorem 2.2 and Corollary 2.2.

3. Applications

Example 3.1 (Occupancy)**.** In the classical occupancy problem, m balls are distributed at random into N urns. The resulting numbers of balls $Z_1, \dots, Z_N$ have a multinomial distribution, and it is well-known that this equals the distribution of $(X_1, \dots, X_N)$ conditioned on $\sum_1^N X_i = m$, where $X_1, \dots, X_N$ are i.i.d. with $X_i \in \mathrm{Po}(\lambda)$, for an arbitrary $\lambda > 0$ [12, 6].

The classical occupancy problem studies the number W of empty urns; this is thus $\sum_1^N \mathbf{1}[X_i = 0]$ conditioned on $\sum_1^N X_i = m$. Now suppose that $m = m_n \to \infty$ and $N = N_n \to \infty$. Then, $W = W_n$ can thus be taken as U_n in Theorem 2.1 with $X_n \in \mathrm{Po}(\lambda_n)$ and $Y_n := \mathbf{1}[X_n = 0]$ for any λ_n; we choose $\lambda_n = m_n/N_n$ so that (i) holds.

If $m_n, N_n \to \infty$ such that $m_n/N_n \to a \in (0, \infty)$, then Corollary 2.1 immediately yields asymptotic normality of W_n, with moment convergence.

In the case $m_n/N_n \to \infty$, simple calculations, using known moment asymptotics and the explicit formula $|\varphi_{X_n}(s)| = \exp\bigl(-\lambda_n(1-\cos s)\bigr)$, show that Theorem 2.1 applies with any $\gamma < 2$, provided $N_n e^{-m_n/N_n} \to \infty$. This condition is necessary for (vii), but it is also necessary for asymptotic normality because W_n is asymptotically Poisson distributed and not asymptotically normal if $N_n e^{-m_n/N_n} \to \mu < \infty$ [12]. Hence Theorem 2.1 is sharp in this case.

In the case $m_n/N_n \to 0$, finally, Theorem 2.1 cannot be applied as stated, because $Y_n = \mathbf{1}[X_n = 0]$ yields $\rho_n \to -1$. Instead we choose, cf. Remark 2.2, $Y_n := \mathbf{1}[X_n = 0] + X_n - 1 = (X_n - 1)_+$, and it is easily verified that Theorem 2.1 applies (with $\gamma = 2$) provided $m_n^2/N_n \to \infty$. Again, this condition is necessary both for (vii) and for asymptotic normality, because $W_n - (N_n - m_n)$ is asymptotically Poisson distributed if $m_n^2/N_n \to \mu < \infty$ [12].

Theorem 2.1 thus proves asymptotic normality of W_n in all cases where it holds, together with moment convergence.

Actually, asymptotic normality in the case $m_n/N_n \to a \in (0, \infty)$ was first proved by Weiss [23] using the method of moments, i.e. by first showing moment convergence. His estimates of the moments uses a complicated combinatorial analysis, however, and the proof here seems simpler. (For other proofs of asymptotic normality, not treating moments, see [20, 12, 5]; the last reference uses the same method as this paper.)

Asymptotic normality in the other cases was shown by Rényi [20, 12]; we do not know any earlier proof of moment convergence in these cases. (It is easy to show that $\mathbb{E} W_n^r \sim (N_n e^{-m_n/N_n})^r$ for every $r > 0$, but we are discussing the moments after normalization, or equivalently the central moments, which are much smaller and thus much more difficult to estimate properly.)

By instead defining $Y_n := \mathbf{1}[X_n = k]$, we similarly obtain asymptotic normality, with moment convergence, for the number of urns with exactly k balls, if $m_n, N_n \to \infty$ in a suitable range. Moreover, Theorem 2.2 yields joint convergence for several k. We can also let k depend on n.

We can similarly treat sums $\sum_1^N f(Z_i)$ for other functions f, for example the χ^2-statistic $(N/m)\sum_1^N (Z_i - m/N)^2$.

Further similar applications, where we now can add moment convergence, can be found in Holst [6].

Example 3.2 (Hashing)**.** Hashing with linear probing (see Knuth [10, Section 6.4] for the computer science background) can be regarded as throwing n balls sequentially into m urns at random; the urns are arranged in a circle and a ball that lands in an occupied urn is (before the next ball is thrown) moved to the next empty urn, always moving in a fixed direction. The length of the move, if any, is called the displacement of the ball, and we are interested in the sum of all displacements. We assume $n < m$.

After throwing all balls, there are $m - n$ empty urns. These divide the occupied urns into blocks of consecutive urns; for convenience we consider the empty urn following such a block as belonging to the block, and we regard an empty urn following another empty urn as a block with zero occupied urns. Thus there are $N := m - n$ blocks, and it can be shown, see [9] for details, that the lengths of the blocks (counting the empty urn) and the sums of displacements inside each block are distributed as $(X_1, Y_1), \dots, (X_N, Y_N)$ conditioned on $\sum_{i=1}^N X_i = m$, where (X_i, Y_i) are i.i.d. copies of a pair (X, Y) of random variables, X has the Borel distribution

$$\mathbb{P}(X = \ell) = \frac{1}{T(\lambda)}\frac{\ell^{\ell-1}}{\ell!}\lambda^\ell, \qquad \ell = 1, 2, \dots, \tag{3.1}$$

where $T(\lambda) := \sum_\ell \ell^{\ell-1}\lambda^\ell/\ell!$ is the well-known tree function and λ is an arbitrary number with $0 < \lambda \le e^{-1}$, and the conditional distribution of Y given $X = \ell$ is the same as the distribution of the total displacement in the case $m = \ell$, $n = \ell - 1$. Consequently, the distribution of the total displacement equals the conditional distribution of $\sum_{j=1}^N Y_j$ given $\sum_{j=1}^N X_j = m$, and we are in the situation studied in this paper. It is easily shown that $\mathbb{E}\, X = 1/(1 - T(\lambda))$.

Assume now that $n \to \infty$ with $m = m_n \to \infty$ and $N_n = m_n - n > 0$. We define $\lambda_n = \frac{n}{m_n}\exp(-n/m_n)$ which satisfies $T(\lambda_n) = n/m_n$ and thus $\mathbb{E}\, X_n = 1/(1 - n/m_n) = m_n/N_n$, where X_n is as in (3.1) with $\lambda = \lambda_n$. If $n/m_n \to \alpha \in (0, 1)$, so $m_n/N_n \to 1/(1-\alpha)$ and $\lambda_n \to \lambda \in (0, e^{-1})$, Corollary 2.1 immediately shows that that the total displacement is asymptotically normal, as previously shown by Flajolet, Poblete and Viola [3]; we further obtain moment convergence. Moreover, straightforward estimates, see [9] for details, show that Theorem 2.1 applies also when $n/m_n \to 0$ or $n/m_n \to 1$ (with $\gamma = 2$ and $\gamma = 1/2$, respectively), provided $n \gg \sqrt{m_n}$ and $m_n - n \gg \sqrt{m_n}$. (When one of these conditions fails, the distribution is not asymptotically normal, see [9]; hence Theorem 2.1 is sharp in this application too.)

Example 3.3 (Random forests)**.** Consider a uniformly distributed random labelled rooted forest with m vertices and $N < m$ roots. (We may assume that the vertices are $1, \dots, m$ and, by symmetry, that the roots are $1, \dots, N$.) Let $n := m - N$ be the number of non-roots.

The sizes of the N trees in the forest are distributed as $X_1, \dots, X_N$ conditioned on $\sum_{i=1}^N X_i = m$, where X_i are i.i.d. with the Borel distribution (3.1) for some λ [16, 11, 17]. Thus, the distribution of the tree sizes is the same as the distribution of the block lengths in Example 3.2, cf. [10, Exercise 6.4-31] and [1].

Let $N = N_n$ and $m = m_n = n + N_n$ and consider the number W_{nk} of trees of size k. Pavlov [16, 17] has shown, using similar methods, asymptotic normality of W_{nk} for every fixed $k \geq 1$, provided $n, N_n \to \infty$ and $N_n(n/N_n)^{\max(k,2)} \to \infty$. (He also gave local limit theorems, which are not treated here.) Unfortunately, we do not recover the full result by our theorems, but under the additional assumption $m_n \ll N_n^2$, we can show asymptotic normality. The case $N_n/m_n \to a \in (0,1)$ follows by Corollary 2.1, choosing X_n as in Example 3.2 and $Y_n = \mathbf{1}[X_n = k]$. Similarly, the case $N_n \ll m_n \ll N_n^2$ follows by Theorem 2.1 (with $\gamma = 1/2$), using the calculations in [9] for X_n together with simple estimates for Y_n. Finally, when $N_n/m_n \to 1$ (and $N_n(n/N_n)^{\max(k,2)} \to \infty$), we can apply Theorem 2.1 (with $\gamma = 2$), but this time we take $Y_n = \mathbf{1}[X_n = k]$ for $k \geq 3$ only; for $k = 1$ and $k = 2$ we use the modifications $\mathbf{1}[X_n = 1] + X_n - 2$ and $\mathbf{1}[X_n = 2] - X_n + 1$, respectively, cf. Remark 2.2. We omit the details.

We thus obtain moment convergence too, and Theorem 2.2 yields joint asymptotic normality of $W_{n1}, W_{n2}, \dots, W_{nl}$, for any l.

Other additive characteristics of the random forest can be studied similarly. For example, we obtain immediately (by Corollary 2.1) asymptotic normality of the total path length when $N/m \to a \in (0,1)$. This can be extended (by Theorem 2.1) to other ranges of $N, m \to \infty$, but we have not worked out the precise conditions.

Example 3.4 (Branching processes)**.** Consider a Galton–Watson process, for simplicity 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 $\mathbb{E}\, X = 1$, i.e. that the process is critical, and that the span of X is 1.

Number the individuals as they appear, and let X_i be the number of children of the i:th individual; we can continue with fictitious individuals after extinction, so that X_i is defined for all $i \geq 1$, forming an i.i.d. sequence.

The total progeny is $n \geq 1$ if and only if

$$S_k := \sum_1^k X_i \geq k \text{ for } 0 \leq k < n \text{ but } S_n = n - 1. \tag{3.2}$$

This is a different type of conditioning than the one studied in this paper, but we observe that if $x_1, \dots, x_n$ are any non-negative integers such that $\sum_1^n (x_i - 1) = -1$, then there is exactly one cyclical shift $x_i' := x_{i+l \pmod n}$ such that $\sum_1^k (x_i' - 1) \geq 0$ for $1 \leq k < n$ [24, 18] (l is the smallest number for which $\sum_1^l x_i = -1$). Consequently, if we ignore (or randomize) the order of $X_1, \dots, X_n$, they have the same distribution conditioned on (3.2) as conditioned on $S_n = n-1$, and we are back in the situation studied in this paper; we can thus use our results to study variables of the type $\sum_1^n f(X_i)$ conditioned on having exactly n individuals in the process. (In this example we have a single sequence rather than a triangular array: $X_{ni} = X_i$ and we drop the index n from X and Y.)

Note that we take $N_n = n$ and $m_n = n - 1$, so $\mathbb{E}\, S_{nN_n} = N_n$ differs from m_n (although only by 1), and we have to use Theorem 2.3 instead of Theorem 2.1. On the other hand, Remark 2.9 shows that the conclusions remains the same.

For example, taking $Y := \mathbf{1}[X = k]$ for some $k = 0, 1, \dots,$ we obtain by Corollary 2.1 and Remark 2.9 that given that the total progeny of the process (up to extinction) is n, the number of individuals with k children is asymptotically normal. More precisely, if this number is denoted by W_{nk} and $p_k := \mathbb{P}(X = k)$, then $n^{-1/2}(W_{nk} - np_k) \xrightarrow{\mathrm{d}} N(0, \sigma_k^2)$ as $n \to \infty$, with moment convergence, where $\sigma_k^2 := p_k(1-p_k) - (k-1)^2 p_k^2 / \operatorname{Var} X$. Joint convergence for several k follows too, by Corollary 2.2 with $b_{nj} := 1$; the asymptotic covariances are $\sigma_{jk} := -p_j p_k - (j-1)(k-1) p_j p_k / \operatorname{Var} X$, $j \neq k$.

Finally, we give three artificial counter-examples that illuminate the conditions in the theorems.

Example 3.5. Let $\alpha, \beta \geq 0$ be two fixed real numbers. Take X_n and Y_n independent, so that $U_n = T_{nN_n} = \sum_1^{N_n} Y_{ni}$. Further, let $N_n = m_n = n$, $X_n \in \operatorname{Po}(1)$ and let Y_n take the values -1, 1 and $n^{\alpha+1/2}$ with probabilities $1/2$, $1/2 - p_n$ and $p_n := n^{-1-\beta}$, respectively.

All conditions in Theorem 2.1 except (vii) and (x) are readily verified.

If $\alpha = \beta = 0$, then $\tau_n^2 = \sigma_{Y_n}^2 \to 2$, and $\mathbb{E}\,|Y_n - \mathbb{E}\, Y_n|^3 \sim \mathbb{E}\,|Y_n|^3 \sim n^{1/2}$. Since the number of terms in $\sum_1^{N_n} Y_{ni}$ that equal $n^{1/2}$ is $\operatorname{Bi}(n, 1/n)$, it is easily seen that $U_n/\sqrt{N_n} \xrightarrow{\mathrm{d}} W + Z$ where $W \in \operatorname{Po}(1)$ and $Z \in N(0,1)$ are independent. This shows that the factor $N_n^{1/2}$ in (vii) is sharp, and moreover that o cannot be relaxed to O.

Assume now that $\beta > 0$. It is then easily seen that $U_n/\sqrt{N_n} \xrightarrow{\mathrm{d}} N(0,1)$. If further $\beta > 2\alpha$, then $N_n \mathbb{E}\, Y_n = o(N_n^{1/2})$ and $\tau_n^2 \to 1$ so (2.1) holds. In this case, $\operatorname{Var} U_n \sim N_n$, so (2.2) holds too. However, with probability $\approx n^{-\beta}$, $U_n/\sqrt{N_n}$ has a value $\approx n^\alpha$, and it is easily verified that convergence of the r:th moments in (2.1) or (2.2) holds if and only if $\beta > r\alpha$. Moreover, (vii) holds if and only if $\beta > 3\alpha$. Similarly, (x) holds if and only if $\beta \geq r\alpha$, and the O can be strengthened to o if and only if $\beta > r\alpha$.

This shows that the results in Remark 2.7 are sharp.

Example 3.6. Let $a_n \geq 2$ be an integer and let (for $n > 4$) $\mathbb{P}(X_n = 0) = 1/2$, $\mathbb{P}(X_n = \pm 1) = 1/4 - 1/n$, $\mathbb{P}(X_n = \pm a_n) = 1/n$. Further, let Z_n be independent of X_n with $\mathbb{P}(Z_n = \pm 1) = 1/2$ and define $Y_n := X_n \mathbf{1}[|X_n| \leq 1] + Z_n$. We take $N_n = n$ and $m_n = 0$.

It is easily verified that Theorem 2.1 applies when $a_n = o(\sqrt{n})$.

Now take $a_n := \lfloor\sqrt{n}\rfloor$. Then all conditions in Theorem 2.1 except (iii) still hold, but $\mathbb{E}|X_n - \mathbb{E}X_n|^3 = \mathbb{E}|X_n|^3 \sim 2n^{1/2} = \Theta(n^{1/2}\sigma_{X_n}^3)$. Note that, if $S_{nN_n} = m_n$, then

$$T_{nN_n} = \sum_1^{N_n} Z_{ni} - \sum_1^{N_n} X_{ni}\mathbf{1}[|X| = a_n] = \sum_1^{N_n} Z_{ni} - a_n(L_{n+} - L_{n-}), \qquad (3.3)$$

where L_{n+} and L_{n-} are the numbers of terms X_{ni} equal to a_n and $-a_n$, respectively. Without conditioning, these numbers converge in distribution to independent Poisson variables, and it is easily seen that even when conditioning on $S_{nN_n} = 0$, $L_{n+} - L_{n-}$ converges in distribution to some non-degenerate random variable $\widetilde{L}$, supported on the integers. Since $\{Z_{ni}\}$ is independent of $\{X_{ni}\}$, it follows from (3.3) that $U_n/\sqrt{N_n} \xrightarrow{\text{d}} W - \widetilde{L}$, where $W \in N(0,1)$ is independent of $\widetilde{L}$. Consequently, the limit is not normal, which shows that assumption (iii) cannot be weakened to O.

Example 3.7. Let $\mathbb{P}(X_n = 0) = \mathbb{P}(X_n = 2) = \frac{1}{2} - p_n$ and $\mathbb{P}(X_n = 1) = 2p_n$, where $p_n := e^{-n}$, say. Let further W and Z be random variables independent of $\{X_n\}$ with $\mathbb{P}(Z = 1) = \mathbb{P}(Z = -1) = \frac{1}{2}$ while W has any distribution with finite moments, and define $Y_n := Z$ if $X_n = 0$ or $X_n = 2$, and $Y_n := \sqrt{2n+1}W$ if $X_n = 1$. Finally, let $N = m_n = 2n + 1$.

Clearly, $(X_n, Y_n) \xrightarrow{\text{d}} (X, Y)$ as $n \to \infty$, with convergence of all moments, where $\frac{1}{2}X \in \text{Be}(\frac{1}{2})$, $Y = Z$, and X and Y are independent, and the conditions of Theorem 2.1 except (v) are satisfied. Indeed, this is an instance of Corollary 2.1 except that the span of the distribution of X is 2. We will see that, nevertheless, the conclusions of Theorem 2.1 do not hold; thus showing that the conditions (v) and 'span=1', respectively, cannot be eliminated. Note that $N_n \mathbb{E}Y_n \to 0$, $\rho_n \to 0$ and $\tau_n^2 \to 1$.

Let L_n be the number of $X_{n1}, \dots, X_{nN_n}$ that equal 1. Then $S_{nN_n} \equiv L_n \pmod 2$, so if $S_{nN_n} = m_n$ then L_n is odd. Moreover, since $p_n \to 0$ rapidly, $\mathbb{P}(L_n \ge 3 \mid S_{nN_n} = m_n) \ll \mathbb{P}(L_n = 1 \mid S_{nN_n} = m_n)$ and thus $\mathbb{P}(L_n = 1 \mid S_{nN_n} = m_n) \to 1$.

If $L_n = 1$ and j is the index with $X_{nj} = 1$, then $Y_{ni} = Z_i$ for $i \ne j$ and thus $T_{nN_n} = \sqrt{N_n}W_j + \sum_{i\ne j} Z_i$ (with the obvious meanings of W_j and Z_i). Since $\{W_i\}$ and $\{Z_i\}$ are independent of $\{X_{ni}\}$, it follows that the conditional distribution of T_{nN_n} given $S_{nN_n} = m_n$, $L_n = 1$ and $X_j = 1$ equals the distribution of $\sqrt{N_n}W_j + \sum_{i\ne j} Z_i$. Furthermore, by symmetry, the latter distribution does not depend on j. Hence, letting $\mathrm{d_{TV}}$ denote the total variation distance of the distributions,

$$\mathrm{d_{TV}}(U_n, \sqrt{N_n}W_1 + \sum_{i=2}^{N_n} Z_i) \le \mathbb{P}(L_n \ge 3 \mid S_{nN_n} = m_n) \to 0,$$

and it follows, applying the central limit theorem to $\{Z_i\}_2^{N_n}$, that

$$U_n/\sqrt{N_n} \xrightarrow{\text{d}} W + V, \qquad (3.4)$$

where $V \in N(0,1)$ is independent of W. Moreover, simple estimates show that all moments converge in (3.4).

For example, taking $W := 1$ we find $U_n/\sqrt{N_n} \xrightarrow{\mathrm{d}} N(1,1)$ and $\mathbb{E}\, U_n = \sqrt{N_n} + o(N_n^{1/2}) \neq N_n \,\mathbb{E}\, Y_n + o(N_n^{1/2}\tau_n)$; hence (2.1) does not hold.

For another simple example, take $W := Z$. Then $\mathbb{E}\, Y_n = \mathbb{E}\, X_n Y_n = 0$, so X_n and Y_n are uncorrelated. We have $\mathbb{E}\, U_n = 0$ but $\operatorname{Var} U_n \sim N_n \operatorname{Var}(W+V) = 2N_n$, and neither (2.1) nor (2.2) holds.

4. Proofs

Lemma 4.1. *If X is an integer valued random variable, then*

$$\operatorname{Var} X \leq 4\,\mathbb{E}\,|X - \mathbb{E}\, X|^3.$$

Proof. Let $k := \lfloor \mathbb{E}\, X + 1/2 \rfloor$ be the integer closest to $\mathbb{E}\, X$. Then

$$\begin{aligned}\mathbb{E}\big(|X - \mathbb{E}\, X| \mathbf{1}[X = k]\big) &= \big|\,\mathbb{E}\big((X - \mathbb{E}\, X)\mathbf{1}[X = k]\big)\big| \\ &= \big|\,\mathbb{E}\big((X - \mathbb{E}\, X)\mathbf{1}[X \neq k]\big)\big| \leq \mathbb{E}\big(|X - \mathbb{E}\, X| \mathbf{1}[X \neq k]\big).\end{aligned}$$

Moreover, $|X - \mathbb{E}\, X| \leq 1/2$ if $X = k$ and $|X - \mathbb{E}\, X| \geq 1/2$ if $X \neq k$, and thus

$$\begin{aligned}\mathbb{E}\big(|X - \mathbb{E}\, X|^2 \mathbf{1}[X = k]\big) &\leq \tfrac{1}{2}\,\mathbb{E}\big(|X - \mathbb{E}\, X| \mathbf{1}[X = k]\big) \\ &\leq \tfrac{1}{2}\,\mathbb{E}\big(|X - \mathbb{E}\, X| \mathbf{1}[X \neq k]\big) \leq \mathbb{E}\big(|X - \mathbb{E}\, X|^2 \mathbf{1}[X \neq k]\big).\end{aligned}$$

Consequently,

$$\mathbb{E}\,|X - \mathbb{E}\, X|^2 \leq 2\,\mathbb{E}\big(|X - \mathbb{E}\, X|^2 \mathbf{1}[X \neq k]\big) \leq 4\,\mathbb{E}\,|X - \mathbb{E}\, X|^3. \qquad \square$$

Proof of Theorem 2.1. We first replace Y_n by the projection

$$Y_n' := Y_n - \mathbb{E}\, Y_n - \frac{\operatorname{Cov}(X_n, Y_n)}{\sigma_{X_n}^2}(X_n - \mathbb{E}\, X_n). \tag{4.1}$$

Then $\mathbb{E}\, Y_n' = 0$ and $\operatorname{Cov}(X_n, Y_n') = \mathbb{E}\, X_n Y_n' = 0$, and, using (viii′),

$$\sigma_{Y_n'}^2 := \operatorname{Var} Y_n' = \operatorname{Var} Y_n - \operatorname{Cov}(X_n, Y_n)^2/\sigma_{X_n}^2 = \tau_n^2 = \Theta(\sigma_{Y_n}^2). \tag{4.2}$$

It follows by this (or Remark 2.3) that τ_n^2 remains the same. Moreover, by Minkowski's inequality, (iii′), (vii′) and (4.2)

$$\|Y_n'\|_3 \leq \|Y_n - \mathbb{E}\, Y_n\|_3 + \frac{|\rho_n|\sigma_{X_n}\sigma_{Y_n}}{\sigma_{X_n}^2}\|X_n - \mathbb{E}\, X_n\|_3 = o(N_n^{1/6}\sigma_{Y_n}) = o(N_n^{1/6}\sigma_{Y_n'})$$

so (vii′), and thus (vii), holds for Y_n' too. Similarly, for part (b), if (ix′) and (x′) hold, then (x′) and (x) hold for Y_n' too. Consequently, all conditions hold for (X_n, Y_n') too.

Finally,

$$T_{nN_n}' := \sum_1^{N_n} Y_{ni}' = T_{nN_n} - N_n\,\mathbb{E}\, Y_n - \operatorname{Cov}(X_n, Y_n)\sigma_{X_n}^{-2}(S_{nN_n} - N_n\,\mathbb{E}\, X_n), \tag{4.3}$$

so conditioned on $S_{nN_n} = m_n = N_n\,\mathbb{E}\, X_n$ we have $T_{nN_n}' = T_{nN_n} - N_n\,\mathbb{E}\, Y_n$. Hence, the conclusions for (X_n, Y_n) and (X_n, Y_n') are equivalent.

Consequently it suffices to prove the theorem for (X_n, Y_n'); in other words, we may henceforth assume that $\mathbb{E}\, Y_n = \mathbb{E}\, X_n Y_n = 0$. Note that then $\tau_n^2 = \sigma_{Y_n}^2$.

We introduce some convenient notation. Let

$$\varphi_n(s,t) := \mathbb{E}\, e^{is(X_n - \mathbb{E}\, X_n) + itY_n} = e^{-ism_n/N_n}\, \mathbb{E}\, e^{isX_n + itY_n},$$

i.e., e^{-ism_n/N_n} times the bivariate characteristic function of (X_n, Y_n). Thus

$$\varphi_n^{N_n}(s,t) = \mathbb{E}\, e^{is(S_{nN_n} - m_n) + itT_{nN_n}}.$$

Let further

$$\begin{aligned}\psi_n(t) &:= \int_{-\pi}^{\pi} \varphi_n^{N_n}(s,t)\, ds = \mathbb{E} \int_{-\pi}^{\pi} e^{is(S_{nN_n} - m_n) + itT_{nN_n}}\, ds \\ &= \mathbb{E}\big(2\pi \mathbf{1}[S_{nN_n} = m_n] e^{itT_{nN_n}}\big) = 2\pi\, \mathbb{P}(S_{nN_n} = m_n)\, \mathbb{E}\big(e^{itT_{nN_n}} \mid S_{nN_n} = m_n\big) \\ &= 2\pi\, \mathbb{P}(S_{nN_n} = m_n)\, \mathbb{E}\, e^{itU_n}, \end{aligned} \tag{4.4}$$

where we used the fact that $\int_{-\pi}^{\pi} e^{isk} ds = 0$ for every non-zero integer k. (This standard Fourier inversion argument is the main reason for the assumption that X_n is discrete.) Consequently,

$$\mathbb{E}\, e^{itU_n} = \mathbb{E}\big(e^{itT_{nN_n}} \mid S_{nN_n} = m_n\big) = \psi_n(t)/\psi_n(0). \tag{4.5}$$

Introduce the scaling constants $\beta_{nX} = N_n^{-1/2} \sigma_{X_n}^{-1}$ and $\beta_{nY} = N_n^{-1/2} \sigma_{Y_n}^{-1}$, and define

$$\Psi_n(t) := \int_{-\pi/\beta_{nX}}^{\pi/\beta_{nX}} \varphi_n^{N_n}(\beta_{nX} s, \beta_{nY} t)\, ds = \beta_{nX}^{-1} \psi_n(\beta_{nY} t). \tag{4.6}$$

We claim that, for every fixed real t,

$$\begin{aligned}\Psi_n(t) &= \int_{-\infty}^{\infty} \mathbf{1}[|s| \le \pi/\beta_{nX}] \varphi_n^{N_n}(\beta_{nX} s, \beta_{nY} t)\, ds \\ &\to \int_{-\infty}^{\infty} e^{-s^2/2 - t^2/2}\, ds = \sqrt{2\pi} e^{-t^2/2}. \end{aligned} \tag{4.7}$$

This then yields, by (4.5) and (4.6), $\mathbb{E}\, e^{it\beta_{nY} U_n} = \Psi_n(t)/\Psi_n(0) \to e^{-t^2/2}$ for each t, and thus by the continuity theorem $\beta_{nY} U_n \xrightarrow{\text{d}} N(0,1)$, which is (2.1). (Remember that we now assume $\mathbb{E}\, Y_n = 0$ and $\tau_n^2 = \sigma_{Y_n}^2$.)

We prove (4.7) by dominated convergence. First, Lemma 4.1 and (iii) yield $\sigma_{X_n}^2 = o(N_n^{1/2} \sigma_{X_n}^3)$, so $\beta_{nX}^{-1} = N_n^{1/2} \sigma_{X_n} \to \infty$ and $\mathbf{1}[|s| \le \pi/\beta_{nX}] \to 1$ for every fixed s.

Next, let s and t be fixed and set $Z_n := \beta_{nX} s (X_n - \mathbb{E}\, X_n) + \beta_{nY} t Y_n$. As is well-known, the Taylor estimate $|e^{iZ_n} - 1 - iZ_n + \frac{1}{2} Z_n^2| \le |Z_n|^3$ yields by integration

$$\varphi_n(\beta_{nX} s, \beta_{nY} t) = \mathbb{E}\, e^{iZ_n} = 1 + i\, \mathbb{E}\, Z_n - \tfrac{1}{2}\, \mathbb{E}\, Z_n^2 + O(\mathbb{E}\, |Z_n|^3). \tag{4.8}$$

Now, using our assumption $\mathbb{E}\, Y_n = \mathbb{E}\, X_n Y_n = 0$, $\mathbb{E}\, Z_n = 0$ and

$$\mathbb{E}\, Z_n^2 = \beta_{nX}^2 s^2 \sigma_{X_n}^2 + \beta_{nY}^2 t^2 \sigma_{Y_n}^2 = \frac{s^2 + t^2}{N_n},$$

while, by Minkowski's inequality with (iii′) and (vii′) again,

$$\|Z_n\|_3 \le N_n^{-1/2}\sigma_{X_n}^{-1}|s|\|X_n - \mathbb{E}\, X_n\|_3 + N_n^{-1/2}\sigma_{Y_n}^{-1}|t|\|Y_n\|_3 = o(N_n^{-1/3}).$$

Thus $\mathbb{E}\,|Z_n|^3 = \|Z_n\|_3^3 = o(N_n^{-1})$.

Consequently, (4.8) yields

$$\varphi_n(\beta_{nX} s, \beta_{nY} t) = 1 - \frac{s^2+t^2}{2N_n} + o\Bigl(\frac{1}{N_n}\Bigr) \tag{4.9}$$

and thus, since (iii) and Hölder's (or Lyapunov's) inequality imply $N_n \to \infty$,

$$\mathbf{1}[|s| \le \pi/\beta_{nX}]\varphi_n^{N_n}(\beta_{nX} s, \beta_{nY} t) \to e^{-(s^2+t^2)/2}. \tag{4.10}$$

This shows pointwise convergence of the integrand in (4.7); it remains to show that the left hand side of (4.10) is dominated by an integrable function $g(s)$, which may depend on t (which is fixed during the argument) but not on n.

First consider $t = 0$, when we have $|\varphi_n(s,0)| = |\varphi_{X_n}(s)|$. We write $\delta_n(s) := 1 - |\varphi_{X_n}(s)|$; thus assumption (v) yields

$$\delta_n(\beta_{nX} s) \ge c \min\bigl(|\beta_{nX} s|^\gamma, s^2\beta_{nX}^2\sigma_{X_n}^2\bigr), \qquad |\beta_{nX} s| \le \pi.$$

Now, by (iv), $\beta_{nX}^{-2} = N_n\sigma_{X_n}^2 = O(N_n^{2/\gamma})$, so $\beta_{nX} \ge c_1 N_n^{-1/\gamma}$ and $|\beta_{nX} s|^\gamma \ge c_2 N_n^{-1}|s|^\gamma$, with $c_1, c_2 > 0$. Moreover,

$$s^2\beta_{nX}^2\sigma_{X_n}^2 + N_n^{-1} = \frac{1}{N_n}(s^2+1) \ge \frac{1}{N_n}|s|^\gamma.$$

Consequently, assuming as we may $c_2 \le 1$,

$$\min(|\beta_{nX} s|^\gamma, s^2\beta_{nX}^2\sigma_{X_n}^2) \ge c_2\frac{1}{N_n}|s|^\gamma - \frac{1}{N_n}$$

and thus, with $c_3 = c_2 \min(c,1) > 0$,

$$\delta_n(\beta_{nX} s) = 1 - |\varphi_{X_n}(\beta_{nX} s)| \ge c_3\frac{1}{N_n}|s|^\gamma - \frac{1}{N_n}, \qquad |\beta_{nX} s| \le \pi. \tag{4.11}$$

Next, for a general t,

$$\varphi_n(s,t) = \varphi_n(s,0) + it\,\mathbb{E}\bigl(e^{is(X_n-\mathbb{E}\, X_n)}Y_n\bigr) + \mathbb{E}\bigl(e^{is(X_n-\mathbb{E}\, X_n)}(e^{itY_n} - 1 - itY_n)\bigr)$$

and thus

$$|\varphi_n(s,t)| \le |\varphi_n(s,0)| + |t||\,\mathbb{E}\bigl(e^{isX_n}Y_n\bigr)| + \mathbb{E}\bigl|e^{itY_n} - 1 - itY_n\bigr|. \tag{4.12}$$

We estimate the terms separately. First, $|\varphi_n(s,0)| = |\varphi_{X_n}(s)| = 1 - \delta_n(s)$. Secondly, let $\theta = \arg\varphi_{X_n}(s)$. Then, since $\mathbb{E}\, Y_n = 0$, by the Cauchy–Schwarz inequality,

$$\begin{aligned} |\,\mathbb{E}(e^{isX_n}Y_n)|^2 &= |\,\mathbb{E}\bigl((e^{isX_n} - e^{i\theta})Y_n\bigr)|^2 \\ &\le \mathbb{E}\, Y_n^2\, \mathbb{E}\bigl|e^{isX_n} - e^{i\theta}\bigr|^2 = \sigma_{Y_n}^2(2 - 2\operatorname{Re}\mathbb{E}\, e^{isX_n - i\theta}) \\ &= \sigma_{Y_n}^2(2 - 2|\varphi_{X_n}(s)|) = 2\sigma_{Y_n}^2\delta_n(s) \end{aligned} \tag{4.13}$$

and thus, by the arithmetic-geometric inequality,

$$|t||\,\mathbb{E}(e^{isX_n}Y_n)| \le \sqrt{2}|t|\sigma_{Y_n}\delta_n(s)^{1/2} \le t^2\sigma_{Y_n}^2 + \tfrac{1}{2}\delta_n(s).$$

Thirdly, $\left|e^{itY_n}-1-itY_n\right|\le\frac12t^2Y_n^2$.

Consequently, (4.12) yields

$$|\varphi_n(s,t)|\le 1-\delta_n(s)+\tfrac12\delta_n(s)+2t^2\sigma_{Y_n}^2=1-\tfrac12\delta_n(s)+2t^2\sigma_{Y_n}^2.$$

Using (4.11) we obtain, with $c_4=c_3/2>0$,

$$\begin{aligned}|\varphi_n(\beta_{nX}s,\beta_{nY}t)|&\le 1-\tfrac12\delta_n(\beta_{nX}s)+2t^2\beta_{nY}^2\sigma_{Y_n}^2\le 1+\tfrac{1}{N_n}(-c_4|s|^\gamma+1+2t^2)\\&\le\exp\bigl(\tfrac{1}{N_n}(-c_4|s|^\gamma+1+2t^2)\bigr),\qquad|\beta_{nX}s|\le\pi,\end{aligned}$$

and thus

$$|\varphi_n^{N_n}(\beta_{nX}s,\beta_{nY}t)|\le\exp(1+2t^2-c_4|s|^\gamma),\qquad|s|\le\pi/\beta_{nX}.$$

This yields the desired domination and completes the proof of (4.7), and thus of (2.1).

Turning to mean and variance, we obtain by differentiating (4.5) and (4.4) twice at $t=0$ (differentiation under the integral sign is easily justified using $\mathbb{E}\,T_{nN_n}^2<\infty$, which follows from (vi))

$$\begin{aligned}-\mathbb{E}\,U_n^2&=\psi_n''(0)/\psi_n(0)\\&=\psi_n(0)^{-1}\int_{-\pi}^{\pi}\Bigl[N_n\partial_t^2\varphi_n(s,0)\varphi_n^{N_n-1}(s,0)+N_n(N_n-1)\bigl(\partial_t\varphi_n(s,0)\bigr)^2\varphi_n^{N_n-2}(s,0)\Bigr]\,ds\end{aligned}$$

and thus, by a change of variables and (4.6),

$$\begin{aligned}\frac{1}{N_n\sigma_{Y_n}^2}\,\mathbb{E}\,U_n^2=\Psi_n(0)^{-1}\int_{-\pi/\beta_{nX}}^{\pi/\beta_{nX}}&\sigma_{Y_n}^{-2}\Bigl[-\partial_t^2\varphi_n(\beta_{nX}s,0)\varphi_n^{N_n-1}(\beta_{nX}s,0)\\&-(N_n-1)\bigl(\partial_t\varphi_n(\beta_{nX}s,0)\bigr)^2\varphi_n^{N_n-2}(\beta_{nX}s,0)\Bigr]\,ds.\end{aligned}\tag{4.14}$$

Again we apply dominated convergence. First,

$$-\partial_t^2\varphi_n(\beta_{nX}s,0)=\mathbb{E}\bigl(e^{i\beta_{nX}s(X_n-\mathbb{E}\,X_n)}Y_n^2\bigr)$$

and thus, using Hölder's inequality and (iii), (vii), for fixed s,

$$\begin{aligned}|\partial_t^2\varphi_n(\beta_{nX}s,0)+\sigma_{Y_n}^2|&=\bigl|\mathbb{E}\bigl((1-e^{i\beta_{nX}s(X_n-\mathbb{E}\,X_n)})Y_n^2\bigr)\bigr|\\&\le\beta_{nX}|s|\,\mathbb{E}\bigl(|X_n-\mathbb{E}\,X_n|Y_n^2\bigr)\le\beta_{nX}|s|\|X_n-\mathbb{E}\,X_n\|_3\|Y_n\|_3^2\\&=o(\beta_{nX}N_n^{1/6}\sigma_{X_n}N_n^{2/6}\sigma_{Y_n}^2)=o(\sigma_{Y_n}^2).\end{aligned}$$

Similarly,

$$\begin{aligned}\partial_t\varphi_n(\beta_{nX}s,0)&=i\,\mathbb{E}\bigl(e^{i\beta_{nX}s(X_n-\mathbb{E}\,X_n)}Y_n\bigr)\\&=i\,\mathbb{E}\bigl(\bigl(e^{i\beta_{nX}s(X_n-\mathbb{E}\,X_n)}-1-i\beta_{nX}s(X_n-\mathbb{E}\,X_n)\bigr)Y_n\bigr)\end{aligned}$$

and thus,

$$\begin{aligned}|\partial_t\varphi_n(\beta_{nX}s,0)|&\le\mathbb{E}\,|\beta_{nX}^2s^2(X_n-\mathbb{E}\,X_n)^2Y_n|\le\beta_{nX}^2s^2\|X_n-\mathbb{E}\,X_n\|_3^2\|Y_n\|_3\\&=o(\beta_{nX}^2N_n^{2/6}\sigma_{X_n}^2N_n^{1/6}\sigma_{Y_n})=o(N_n^{-1/2}\sigma_{Y_n}).\end{aligned}$$

Together with (4.9), this shows that the integrand in (4.14) tends to $e^{-s^2/2}$ for every s.

To obtain a suitable domination, we observe that

$$|\partial_t^2\varphi_n(s,0)| \le \mathbb{E}\,Y_n^2 = \sigma_{Y_n}^2$$

and, by (4.13),

$$|\partial_t\varphi_n(s,0)|^2 = |\,\mathbb{E}\,e^{isX_n}Y_n|^2 \le 2\sigma_{Y_n}^2\delta_n(s).$$

Consequently, using (4.11) in the final step,

$$\begin{aligned}
&\Big|\sigma_{Y_n}^{-2}\Big[-\partial_t^2\varphi_n(\beta_{nX}s,0)\varphi_n^{N_n-1}(\beta_{nX}s,0)-(N_n-1)\big(\partial_t\varphi_n(\beta_{nX}s,0)\big)^2\varphi_n^{N_n-2}(\beta_{nX}s,0)\Big]\Big|\\
&\quad\le (1-\delta_n(\beta_{nX}s))^{N_n-1}+2N_n\delta_n(\beta_{nX}s)(1-\delta_n(\beta_{nX}s))^{N_n-2}\\
&\quad\le e^{-(N_n-1)\delta_n(\beta_{nX}s)}+2N_n\delta_n(\beta_{nX}s)e^{-(N_n-2)\delta_n(\beta_{nX}s)}\\
&\quad\le C_1e^{-N_n\delta_n(\beta_{nX}s)/2}\le C_2e^{-c_4|s|^\gamma},
\end{aligned}$$

for some C_1, C_2. It now follows from (4.14), and $\Psi_n(0)\to\sqrt{2\pi}$ by (4.7), that

$$\frac{1}{N_n\sigma_{Y_n}^2}\,\mathbb{E}\,U_n^2 \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-s^2/2}\,ds = 1.$$

In other words, $\mathbb{E}(U_n/N_n^{1/2}\sigma_{Y_n})^2 \to 1$. In particular, the random variables $U_n/N_n^{1/2}\sigma_{Y_n}$ are uniformly integrable, and it follows from the already shown $U_n/N_n^{1/2}\sigma_{Y_n} \overset{\mathrm{d}}{\to} N(0,1)$ that $\mathbb{E}(U_n/N_n^{1/2}\sigma_{Y_n}) \to 0$. Thus also $\operatorname{Var}(U_n/N_n^{1/2}\sigma_{Y_n}) \to 1$.

This proves the estimates of mean and variance in the theorem, and shows further that $(U_n-\mathbb{E}\,U_n)/N_n^{1/2}\sigma_{Y_n}$ and $(U_n-\mathbb{E}\,U_n)/(\operatorname{Var}U_n)^{1/2}$ too converge in distribution to $N(0,1)$, which completes the proof of (a).

For part (b) we have that, still assuming $\mathbb{E}\,Y_n=\mathbb{E}\,X_nY_n=0$,

$$\mathbb{E}\,|Y_n|^r = O(N_n^{r/2-1}\sigma_{Y_n}^r) \tag{4.15}$$

holds by assumption for even $r\ge 4$, and trivially for $r=2$, while if $r=2s+1\ge 3$ is an odd integer, then Hölder's inequality yields $\mathbb{E}\,|Y_n|^r \le (\mathbb{E}\,Y_n^{2s}\,\mathbb{E}\,Y_n^{2s+2})^{1/2}$, which by (4.15) for $r=2s$ and $r=2s+2$ implies that (4.15) holds for $r=2s+1$ too. Thus (4.15) holds for each integer $r\ge 2$.

Now, let r be a positive integer. Differentiating (4.5) and (4.4) r times we find

$$i^r\,\mathbb{E}\,U_n^r = \psi_n^{(r)}(0)/\psi_n(0) = \psi_n(0)^{-1}\int_{-\pi}^{\pi}\partial_t^r\varphi_n^{N_n}(s,0)\,ds,$$

and thus, by a change of variable,

$$\mathbb{E}(U_n/N_n^{1/2}\sigma_{Y_n})^r = i^{-r}\Psi_n(0)^{-1}\int_{-\pi/\beta_{nX}}^{\pi/\beta_{nX}} N_n^{-r/2}\sigma_{Y_n}^{-r}\partial_t^r\varphi_n^{N_n}(\beta_{nX}s,0)\,ds. \tag{4.16}$$

As in the special case $r=2$ in (4.14), we can, by differentiating $\varphi_n^{N_n}$ r times and expanding, write the integrand as a sum of terms

$$\varphi_n(\beta_{nX}s,0)^{N_n-j}\prod_{i=1}^{j}(N_n+1-i)N_n^{-r_i/2}\sigma_{Y_n}^{-r_i}\partial_t^{r_i}\varphi_n(\beta_{nX}s,0), \tag{4.17}$$

where $1\le j\le r$, $r_i\ge 1$ and $\sum_i r_i = r$.

We write the product in (4.17) as $\prod_i \bigl(1-(i-1)/N_n\bigr) g_{r_i}(\beta_{nX}s)$, where

$$g_r(s) := N_n^{1-r/2}\sigma_{Y_n}^{-r}\partial_t^r\varphi_n(s,0) = N_n^{1-r/2}\sigma_{Y_n}^{-r} i^r\,\mathbb{E}(e^{is(X_n-\mathbb{E}\,X_n)}Y_n^r).$$

By (4.13),

$$|g_1(s)| \le \sqrt{2}N_n^{1/2}\delta_n(s)^{1/2}$$

and for $r\ge 2$, by (4.15),

$$|g_r(s)| \le N_n^{1-r/2}\sigma_{Y_n}^{-r}\,\mathbb{E}\,|Y_n|^r = O(1).$$

Consequently, the integrand in (4.16) is bounded by, using (4.11) again,

$$C_r(1-\delta_n(\beta_{nX}s))^{N_n-r}\sum_{i=0}^{r}|g_1(\beta_{nX}s)|^i \le C_r' e^{-(N_n-r)\delta_n(\beta_{nX}s)}\sum_{i=0}^{r}(N_n\delta_n(\beta_{nX}s))^{i/2}$$
$$\le C_r'' e^{-N_n\delta_n(\beta_{nX}s)/2} \le C_r''' e^{-c_4|s|^\gamma},$$

and thus the integral is bounded by a constant (depending on r). Since further $\Psi_n(0)\to\sqrt{2\pi}>0$, (4.16) now shows that $\mathbb{E}(U_n/N_n^{1/2}\sigma_{Y_n})^r$ is bounded for each r, i.e. all moments stay bounded in (2.1).

As is well-known, this implies uniform r:th power integrability of the left hand side in (2.1) for all $r>0$, and thus convergence of all moments to the corresponding moments of the limiting normal distribution.

Finally, this easily implies moment convergence in (2.2) too. □

Proof of Corollary 2.1. The assumptions imply that $\mathbb{E}\,X_n^r\to\mathbb{E}\,X^r$ and $\mathbb{E}\,Y_n^r\to\mathbb{E}\,Y^r$ for every $r\ge 1$, which immediately implies all conditions except (v) in Theorem 2.1. To verify (v), with $\gamma=2$, we observe that

$$|\varphi_{X_n}(s)|^2 = \mathbb{E}\,e^{is(X_{n1}-X_{n2})} = 1 - s^2\operatorname{Var}X_n + O(s^3\,\mathbb{E}\,|X_n|^3).$$

Since $\operatorname{Var}X_n\to\operatorname{Var}X>0$ and $\mathbb{E}\,|X_n|^3\to\mathbb{E}\,|X|^3<\infty$, it follows that for some $\delta>0$ that does not depend on n and $|s|\le\delta$, we have $|\varphi_{X_n}(s)|\le 1-\frac14 s^2\operatorname{Var}X$. On the other hand, $\varphi_{X_n}(s)$ converges uniformly to $\varphi_X(s)$, and $\max\{|\varphi_X|:\delta\le|s|\le\pi\}<1$ because X has span 1; hence $|\varphi_{X_n}(s)|\le 1-\varepsilon\le 1-\varepsilon(s/\pi)^2$ for $\delta\le|s|\le\pi$, some $\varepsilon>0$, and large enough n. This implies (v) (for large enough n), and the result follows by Theorem 2.1. □

Proof of Theorem 2.2. First, replacing $Y_n^{(j)}$ by $Y_n^{(j)}/b_{nj}$, we may assume that $b_{nj}=1$. Secondly, by the Cramér–Wold device, i.e. considering $Y_n = t_1Y_n^{(1)}+\dots+t_lY_n^{(l)}$ where $t_1,\dots,t_l$ are fixed real numbers, it is easily seen that it suffices to consider the case $l=1$, i.e. when $Y_n=Y_n^{(1)}$ is real-valued. In this case, with τ_n^2 as in Theorem 2.1, (2.4) may be written

$$\tau_n^2\to\sigma_{11}. \tag{4.18}$$

Assume first that $\sigma_{11}=\lim\tau_n^2>0$. Since $\tau_n^2\le\sigma_{Y_n}^2$ and, by (vi), $\sigma_{Y_n}^2=O(1)$, we see that (vi), (vii) and (viii′) in Theorem 2.1 hold (and in part (b) (x) too). The result now follows by Theorem 2.1 and (4.18).

It remains to prove that if $\tau_n^2\to 0$, then $N_n^{-1/2}(U_n-N_n\,\mathbb{E}\,Y_n)\overset{\mathrm{d}}{\to}0$ (with moment convergence). This can be proved by replacing Y_n by Y_n+Z, where

$Z \in N(0,1)$ is independent of (X_n, Y_n), and then applying the case $\tau_n^2 \to 1$; alternatively, one can modify the proof of Theorem 2.1 (redefining $\beta_{nY} := N_n^{-1/2}$). We omit the details. □

Proof of Theorem 2.3. In this case, (4.3) shows that, conditioned on $S_{nN_n} = m_n$, $T'_{nN_n} = T_{nN_n} - N_n \mathbb{E}\, Y_n - \rho_n \frac{\sigma_{Y_n}}{\sigma_{X_n}}(m_n - N_n \mathbb{E}\, X_n)$. This yields the extra term in (2.7); it further shows that the general case again follows from the special case $\mathbb{E}\, Y_n = \mathbb{E}\, X_n Y_n = 0$.

By considering subsequences, we may further assume that

$$v_n := (m_n - \mathbb{E}\, X_n)/N_n^{1/2}\sigma_{X_n} \to v$$

for some real v. The proof of (2.1) and moment convergence is now as before with minor modifications: We now take $\psi_n(t) := \int_{-\pi}^{\pi} e^{-i(m_n - N_n \mathbb{E}\, X_n)s} \varphi_n^{N_n}(s,t)\, ds$ in (4.4); (4.5) still holds but we get an extra factor $e^{-iv_n s}$ in the integral in (4.6), and thus the limit in (4.7) becomes $\int_{-\infty}^{\infty} e^{-isv - s^2/2 - t^2/2}\, ds = \sqrt{2\pi} e^{-v^2/2 - t^2/2}$, whence $\Psi_n(t)/\Psi_n(0) \to e^{-t^2/2}$ still holds. Similarly, we get a factor $e^{-iv_n s}$ in the integrals in (4.14) and (4.16). The result follows using the previously established limits and bounds of the integrands. □

References

[1] P. Chassaing & G. Louchard, Phase transition for parking blocks, Brownian excursion and coalescence. Preprint, 1999. Available at `http://altair.iecn.u-nancy.fr/~chassain`

[2] D. M. Chibisov, On the normal approximation for a certain class of statistics. *Proc. Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. I: Theory of statistics*, pp. 153–174. Univ. California Press, Berkeley, Calif., 1972.

[3] P. Flajolet, P. Poblete & A. Viola, On the analysis of linear probing hashing. *Algorithmica* **22** (1998), no. 4, 490–515.

[4] C. Hipp, Asymptotic expansions for conditional distributions: the lattice case. *Probab. Math. Statist.* **4** (1984), no. 2, 207–219.

[5] L. Holst, Two conditional limit theorems with applications. *Ann. Statist.* **7** (1979), no. 3, 551–557.

[6] L. Holst, A unified approach to limit theorems for urn models. *J. Appl. Probab.* **16** (1979), 154–162.

[7] L. Holst, Some conditional limit theorems in exponential families. *Ann. Probab.* **9** (1981), no. 5, 818–830.

[8] Sh. A. Ismatullaev, Asymptotic expansions for conditional distributions. (Russian) *Dokl. Akad. Nauk UzSSR* **1987**, no. 4, 12–15.

[9] S. Janson, Asymptotic distribution for the cost of linear probing hashing. *Random Struct. Appl.*, to appear. Available at `http://www.math.uu.se/~svante/papers/`

[10] D. E. Knuth, *The Art of Computer Programming. Vol. 3: Sorting and Searching.* 2nd ed., Addison-Wesley, Reading, Mass., 1998.

[11] V. F. Kolchin, *Random Mappings.* Nauka, Moscow, 1984 (Russian). English transl.: Optimization Software, New York, 1986.

[12] V. F. Kolchin, B. A. Sevastyanov & V. P. Chistyakov, *Random allocations.* Nauka, Moscow, 1976 (Russian). English transl.: Winston, Washington, D.C., 1978

[13] E. M. Kudlaev, *Teor. Veroyatnost. i Primenen.* **29** (1984), no. 4, 743–752. English transl.: *Theory Probab. Appl.* **29** (1984), no. 4, 776–786.

[14] R. Michel, Asymptotic expansions for conditional distributions. *J. Multivariate Anal.* **9** (1979), no. 3, 393–400.

[15] Sh. A. Mirakhmedov, Limit theorems for conditional distributions. (Russian) *Diskret. Mat.* **6** (1994), no. 4, 107–132; English transl.: *Discrete Math. Appl.* **4** (1994), no. 6, 519–542.

[16] Yu. L. Pavlov, Limit theorems for the number of trees of a given size in a random forest. *Mat. Sb.* 103(145) (1977), no. 3, 392–403, 464 (Russian). English transl.: *Math. USSR Sb.* **32** (1977), no. 3, 335–345.

[17] Yu. L. Pavlov, *Random forests.* Karelian Centre Russian Acad. Sci., Petrozavodsk, 1996 (Russian). English transl.: VSP, Zeist, The Netherlands, 2000.

[18] J. Pitman, Enumerations of trees and forests related to branching processes and random walks. *Microsurveys in discrete probability (Princeton, NJ, 1997)*, 163–180, Amer. Math. Soc., Providence, RI, 1998.

[19] S. Portnoy, Asymptotic efficiency of minimum variance unbiased estimators. *Ann. Statist.* **5** (1977), no. 3, 522–529.

[20] A. Rényi, Three new proofs and a generalization of a theorem of Irving Weiss. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **7** (1962), 203–214.

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

[22] A. R. Swensen, A note on convergence of distributions of conditional moments. *Scand. J. Statist.* **10** (1983), no. 1, 41–44.

[23] I. Weiss, Limiting distributions in some occupancy problems. *Ann. Math. Statist.* **29** (1958), 878–884.

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

[25] S. L. Zabell, Rates of convergence for conditional expectations. *Ann. Probab.* **8** (1980), no. 5, 928–941.

[26] S. L. Zabell, A limit theorem for expectations conditional on a sum. *J. Theoret. Probab.* **6** (1993), no. 2, 267–283.

Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden

E-mail address: svante.janson@math.uu.se