Strong laws of large numbers for random forests

Alexey Chuprunov[1] and István Fazekas[2]

Abstract

Random forests are studied. A moment inequality and a strong law of large numbers are obtained for the number of trees having a fixed number of nonroot vertices.

Key words and phrases: random forest, tree, Cayley's theorem, moment inequality, strong law of large numbers, Taylor's expansion, functional limit theorem.

2000 Mathematics Subject Classification: 60F15 Strong theorems, 60F17 Functional limit theorems; invariance principles, 60C05 Combinatorial probability.

1 Introduction

We will consider the set of forests having N labeled rooted trees and n nonroot vertices. The N roots are labeled by $s_1, \dots, s_N$ and the nonroot vertices are labeled by $1, 2, \dots, n$. By Cayley's theorem, the number of forests is $N(N+n)^{n-1}$ (see [18], [13], [10]). We will consider uniformly distributed probability $\mathbb{P}_1$ on the set of forests. The uniform probability on the set of forests is widely studied (see e.g. [12] and the references therein).

Let $\mu_r(n, N)$ denote the number of trees with r nonroot vertices in the forest having N rooted trees and n nonroot vertices. In [13] limit theorems are obtained for $\mu_r(n, N)$. The limiting distributions in [13] are Poisson or normal according to the ratio of n/N.

In this paper we prove strong laws of large numbers for $\mu_r(n, N)$. Assume that $\frac{n_k}{N_k} \to \alpha$, as $k \to \infty$, for some $\alpha \in \mathbb{R}$. Let $\lambda = \frac{\alpha}{1+\alpha}$. Then, as $k \to \infty$, $\frac{1}{N_k}\mu_r(n_k, N_k) \to L(r, \lambda)$ almost surely (Lemma 3.1). Here $L(r, \lambda) = \frac{(1+r)^{r-1}}{r!} e^{-(r+1)\lambda} \lambda^r$. In Section 3 several versions of the above strong law are obtained.

The proofs are based on a fourth moment inequality for $\mu_r(n, N)$ (Lemma 2.1). To obtain the moment inequality we use Taylor's expansion and we shall see that terms having higher order than N^2 disappear. (The proof of Lemma 2.1 is presented in Section 5.)

In Section 4 a functional limit theorem is proved where the processes are governed by evolving random forests.

We remark that from graph theory we apply only Cayley's theorem. Early results for random graphs can be found e.g. in [7] and [13]. For the general theory of random graphs and for some new results see [10], [3], [9], [15]. We remark that in [1] uniform random recursive forests are studied. However, in [1] each path from the root is labeled with an

[1]Department of Math. Stat. and Probability, Chebotarev Inst. of Mathematics and Mechanics, Kazan State University, Universitetskaya 17, 420008 Kazan, Russia, e-mail: achuprunov@mail.ru

[2]Faculty of Informatics, University of Debrecen, P.O. Box 12, 4010 Debrecen, Hungary, e-mail: fazekasi@inf.unideb.hu, tel: 36-52-316666/22825

increasng sequence of labels which leads to a model being different from the our one. We also mention that there is a statistical theory of random forests (see [4]) which is not studied here.

We shall use the notation $\mathbb{N} = \{1, 2, \dots\}$ and $\mathbb{Z}^+ = \{0, 1, 2, \dots\}$.

2 The moment inequality

Let $N, n > 0$ and $r \geq 0$ be intergers. We will denote by $\mathcal{F}_{n,N}$ the set of forests having N labeled rooted trees and n nonroot vertices. The N roots are labeled by $s_1, \dots, s_N$ and the nooroot vertices are labeled by $1, 2, \dots, n$. It is known that $\mathcal{F}_{n,N}$ has $M = N(N+n)^{n-1}$ elements (see [10]). We will consider uniformly distributed probability $\mathbb{P}_1$ on $\mathcal{F}_{n,N}$. Let $\mu_r(n, N)$ denote the number of trees with r nonroot vertices in the forest. Then $\mu_r(n, N)$ is a random variable on $\mathcal{F}_{n,N}$. We have

$$\mu_r(n, N) = \sum\nolimits_{i=1}^{N} \mathbb{I}_{Nni}^{(r)},$$

where $\mathbb{I}_{Nni}^{(r)}$ is the indicator of the event that the ith tree has r nonroot vertices. Since the number of individual trees with r nonrooot vertices is $(1+r)^{r-1}$, so the number of forests such that the ith tree has r nonroot vertices is $m = C_n^r (1+r)^{r-1}(N-1)(N-1+n-r)^{n-r-1}$. Here $C_n^r = \binom{n}{r}$ denotes the binomial coefficient. Therefore we have

$$E_1 = \mathbb{E}_1 \mathbb{I}_{Nni}^{(r)} = \frac{m}{M} = \frac{C_n^r (1+r)^{r-1}(N-1)(N-1+n-r)^{n-r-1}}{N(N+n)^{n-1}}. \tag{2.1}$$

Similar calculations give

$$E_2 = \mathbb{E}_1 \mathbb{I}_{Nni}^{(r)} \mathbb{I}_{Nnj}^{(r)} = \frac{C_n^r C_{n-r}^r (1+r)^{2(r-1)}(N-2)(N-2+n-2r)^{n-2r-1}}{N(N+n)^{n-1}}, \quad i \neq j, \tag{2.2}$$

$$E_3 = \mathbb{E}_1 \mathbb{I}_{Nni_1}^{(r)} \mathbb{I}_{Nni_2}^{(r)} \mathbb{I}_{Nni_3}^{(r)} = \frac{C_n^r C_{n-r}^r C_{n-2r}^r (1+r)^{3(r-1)}(N-3)(N-3+n-3r)^{n-3r-1}}{N(N+n)^{n-1}}, \tag{2.3}$$

with $i_k \neq i_l$ if $k \neq l$, $\quad k, l \in \{1, 2, 3\}$, moreover

$$E_4 = \mathbb{E}_1 \mathbb{I}_{Nni_1}^{(r)} \mathbb{I}_{Nni_2}^{(r)} \mathbb{I}_{Nni_3}^{(r)} \mathbb{I}_{Nni_4}^{(r)} = \tag{2.4}$$

$$= \frac{C_n^r C_{n-r}^r C_{n-2r}^r C_{n-3r}^r (1+r)^{4(r-1)}(N-4)(N-4+n-4r)^{n-4r-1}}{N(N+n)^{n-1}}$$

with $i_k \neq i_l$ if $k \neq l$, $\quad k, l \in \{1, 2, 3, 4\}$.

Lemma 2.1. *Let*

$$\alpha = \frac{n}{N}, \quad \lambda = \frac{n}{n+N} = \frac{\alpha}{1+\alpha}$$

and

$$L = L(r,\lambda) = \frac{(1+r)^{r-1}}{r!} e^{-(r+1)\lambda}\lambda^r.$$

Let $N, n > 0$ and $r \geq 0$ be integers such that $\frac{(4(r+1))^4}{n} < 0.001$.

(1) We have

$$\mathbb{E}_1 \left\{ \sum\nolimits_{i=1}^{N} \left(\mathbb{I}_{Nni}^{(r)} - \mathbb{E}_1 \mathbb{I}_{Nni}^{(r)} \right) \right\}^4 \leq C N^2 L (r+1)^4, \tag{2.5}$$

where $C \leq p(\alpha)/\alpha^2$ and $p(\alpha)$ is a fixed polynomial of α.

(2) Assume that $\lambda = \frac{n}{n+N} \leq \tau$ where τ is a constant with $\tau < 1$. Then there exists a finite constant C_1 (depending only on λ) such that for all $r \geq 0$ we have

$$\mathbb{E}_1 \left\{ \sum\nolimits_{i=1}^{N} \left(\mathbb{I}_{Nni}^{(r)} - \mathbb{E}_1 \mathbb{I}_{Nni}^{(r)} \right) \right\}^4 \leq C_1 N^2 L \frac{g(\alpha)}{\alpha^2} \tag{2.6}$$

where $g(\alpha)$ is a fixed polynomial of $\alpha = n/N$.

Remark 2.1. Let $0 < \alpha_1 < \alpha < \alpha_2 < \infty$. Then $\frac{g(\alpha)}{\alpha^2} \leq C$. Moreover, since $\frac{x}{1+x}$ is an increasing function, $\lambda < \frac{\alpha_2}{1+\alpha_2} = \tau < 1$.

Remark 2.2. The sequence $\{L(r,\lambda), r \in \mathbb{Z}^+\}$ can be cosidered as a distribution on $\mathbb{Z}^+$. To see it we remark that

$$\sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} (ae^{-a})^k = a$$

see [17]. Therefore, for all $\lambda > 0$ we have

$$\sum_{r=0}^{\infty} L(r,\lambda) = \sum_{r=0}^{\infty} \frac{(1+r)^{r-1}}{r!} e^{-(r+1)\lambda}\lambda^r = \frac{1}{\lambda} \sum_{r=0}^{\infty} \frac{(1+r)^r}{(r+1)!} (e^{-\lambda}\lambda)^{r+1} = \frac{\lambda}{\lambda} = 1.$$

Another way to obtain it for the case $\lambda \neq 1$ is the following. For $0 < x < 1/e$, by the quotient criterion, the series $\theta(x) = \sum_{k=1}^{\infty} \frac{k^{k-1}x^k}{k!}$ is convergent. Then (see [10], p. 44) $\theta(x)$ is a solution of the equation $\theta e^{-\theta} = x$. Therefore, for all $\lambda > 0$, $\lambda \neq 1$, we have

$$\sum_{r=0}^{\infty} L(r,\lambda) = \frac{\theta(e^{-\lambda}\lambda)}{\lambda} = \frac{\lambda}{\lambda} = 1.$$

(For $\lambda = 1$ we have $e^{-\lambda}\lambda = 1/e$, that is we are on the border of the convergence domain of the above series.)

3 The strong laws

In this section we prove strong laws of large numbers for random forests. Theorem 3.1 concerns the average number of trees containing r nonroot vertices. Theorem 3.2 is a general strong law to be applied in Section 4.

We will assume that all indicators which we will consider in this section are defined on the same probability space $(\Omega_1, \mathcal{A}_1, \mathbb{P}_1)$.

Lemma 3.1. *Let* (N_k) *be a strictly increasing sequence of positive integers and let* (n_k) *be a sequence of nonnegative integers. Assume that* $\frac{n_k}{N_k} \to \alpha$*, as* $k \to \infty$*, for some* $\alpha \in \mathbb{R}$*. Let* $\lambda = \frac{\alpha}{1+\alpha}$*. Then for any* $r \in \mathbb{Z}^+$*, as* $k \to \infty$*,*

$$\frac{1}{N_k} \sum_{i=1}^{N_k} \mathbb{I}^{(r)}_{N_k n_k i} \to L(r, \lambda) \quad \textit{almost surely.}$$

Proof. First consider $\alpha \neq 0$. Standard calculation gives

$$\mathbb{E}_1 \mathbb{I}^{(r)}_{N_k n_k i} = \frac{C^r_{n_k}(1+r)^{r-1}(N_k-1)(N_k-1+n_k-r)^{n_k-r-1}}{N_k(N_k+n_k)^{n_k-1}} \to \frac{(1+r)^{r-1}}{r!} e^{-(r+1)\lambda} \lambda^r ,$$

as $k \to \infty$. By Lemma 2.1, condition (2.1) from p.167 of [5] is valid. Therefore Lemma 3.1 follows from Lemma 2.1 on p.167 of [5].

For $\alpha = 0$ we see that $L(r, \lambda)$ is 1 for $r = 0$. Therefore the lemma is obvious. The proof is complete. □

Let $\mathbb{Z}' \subset \mathbb{Z}^+$. Introduce notation

$$\mu_{zk} = \sum_{r \in \mathbb{Z}'} \sum_{i=1}^{N_k} \mathbb{I}^{(r)}_{N_k n_k i}, \quad k \in \mathbb{N}.$$

We consider μ_{zk} as the number of trees containing r nonroot vertices for some $r \in \mathbb{Z}'$. The following strong law of large numbers gives the limit of the average number of trees containing r nonroot vertices for some $r \in \mathbb{Z}'$.

Theorem 3.1. *Let* (N_k) *be a strictly increasing sequence of positive integers and let* (n_k) *be a sequence of nonnegative integers. Assume that* $\frac{n_k}{N_k} \to \alpha$*, as* $n \to \infty$*, for some* $\alpha \in \mathbb{R}$*. Let* $\lambda = \frac{\alpha}{1+\alpha}$*. Then, as* $k \to \infty$*, we have*

$$\frac{1}{N_k} \mu_{zk} \to \sum_{r \in \mathbb{Z}'} L(r, \lambda) \quad \textit{almost surely.}$$

Proof. By Lemma 3.1, there exists $\Omega' \subset \Omega_1$ such that $\mathbb{P}_1(\Omega') = 1$ and for all $\omega_1 \in \Omega'$ and for all $r \in \mathbb{Z}^+$, as $k \to \infty$,

$$\frac{1}{N_k} \sum_{i=1}^{N_k} \mathbb{I}^{(r)}_{N_k n_k i}(\omega_1) \to L(r, \lambda). \tag{3.1}$$

Let $\omega_1 \in \Omega'$. Let $\varepsilon > 0$. Choose $r_0 \in \mathbb{Z}^+$ such that

$$\sum_{r=r_0}^{\infty} L(r, \lambda) < \frac{\varepsilon}{3}.$$

Since

$$\frac{1}{N_k}\sum_{r=r_0}^{\infty}\sum_{i=1}^{N_k}\mathbb{I}^{(r)}_{N_k n_k i} = 1-\frac{1}{N_k}\sum_{r=0}^{r_0-1}\sum_{i=1}^{N_k}\mathbb{I}^{(r)}_{N_k n_k i},$$

by (3.1) and Remark 2.2, it follows that

$$\frac{1}{N_k}\sum_{r=r_0}^{\infty}\sum_{i=1}^{N_k}\mathbb{I}^{(r)}_{N_k n_k i}(\omega_1)\to\sum_{r=r_0}^{\infty}L(r,\lambda),\quad \text{as}\quad k\to\infty.$$

Therefore we can choose $k_1\in\mathbb{N}$ such that

$$\frac{1}{N_k}\sum_{r=r_0}^{\infty}\sum_{i=1}^{N_k}\mathbb{I}^{(r)}_{N_k n_k i}(\omega_1)<\frac{\varepsilon}{3}$$

for all $k>k_1$. Since, by (3.1),

$$\frac{1}{N_k}\sum_{r\in\mathbb{Z}',r<r_0}\sum_{i=1}^{N_k}\mathbb{I}^{(r)}_{N_k n_k i}(\omega_1)\to\sum_{r\in\mathbb{Z}',r<r_0}L(r,\lambda),\quad \text{as}\quad k\to\infty,$$

we can choose $k_2\in\mathbb{N}$ such that

$$\left|\frac{1}{N_k}\sum_{r\in\mathbb{Z}',r<r_0}\sum_{i=1}^{N_k}\mathbb{I}^{(r)}_{N_k n_k i}(\omega_1)-\sum_{r\in\mathbb{Z}',r<r_0}L(r,\lambda)\right|<\frac{\varepsilon}{3},\quad \text{for all}\quad k>k_2.$$

Let $k_0=\max(k_1,k_2)$. For all $k>k_0$ we have

$$\left|\frac{1}{N_k}\mu_{zk}-\sum_{r\in\mathbb{Z}'}L(r,\lambda)\right|\le\left|\frac{1}{N_k}\sum_{r\in\mathbb{Z}',r<r_0}\sum_{i=1}^{N_k}\mathbb{I}^{(r)}_{N_k n_k i}(\omega_1)-\sum_{r\in\mathbb{Z}',r<r_0}L(r,\lambda)\right|+$$

$$+\frac{1}{N_k}\sum_{r=r_0}^{\infty}\sum_{i=1}^{N_k}\mathbb{I}^{(r)}_{N_k n_k i}(\omega_1)+\sum_{r=r_0}^{\infty}L(r,\lambda)<\varepsilon.$$

The proof is complete. □

Our next strong law fits to the functional limit theorem in Section 4. Let $\mathbb{I}^{(r\infty)}_{Nni}=\sum_{k=r}^{\infty}\mathbb{I}^{(k)}_{Nni}$. It means that the ith tree contains at least r nonroot vertices. For each k let $f_k(.)$ be a non-decreasing non-negative integer valued function on $[0,\infty)$. The function $f_k(t)$ will mean the number of noonroot vertices being a non-decreasing function of time t. Assume that $\frac{f_k(t)}{N_k}\to f(t)$, as $k\to\infty$, where $f(.)$ is a continuous function on $[0,\infty)$. We will consider the random processes

$$Z^{(r\infty)}_k(t)=Z^{(r\infty)}_k(t,\omega_1)=\frac{1}{N_k}\sum_{i=1}^{N_k}\mathbb{I}^{(r\infty)}_{N_k f_k(t) i},\quad t\in\mathbb{R}^+,\quad k\in\mathbb{N},\ \ \omega_1\in\Omega_1.$$

Theorem 3.2. *Let $r \in \mathbb{Z}^+$. Assume that $\lim_{k\to\infty} \frac{f_k(t)}{N_k} \to f(t)$ where $f(.)$ is a continuous function on $[0,\infty)$. Let $\sigma_r(t) = \sum_{m=r}^{\infty} L(m, \lambda(t))$ with $\lambda(t) = f(t)/(1+f(t))$.*

Then for the random processes $Z_k^{(r\infty)}$ one has

$$\sup_{t\in\mathbb{R}^+} |Z_k^{(r\infty)}(t) - \sigma_r(t)| \to 0, \qquad \text{as} \quad k \to \infty,$$

for almost all $\omega_1 \in \Omega_1$.

Proof. By Theorem 3.1, there exists $\Omega_1' \subset \Omega_1$ such that $\mathbb{P}_1(\Omega_1') = 1$ and for all $\omega_1 \in \Omega_1'$, for all $t \in \mathbb{Q}^+$

$$Z_k^{(r\infty)}(t, \omega_1) \to \sigma_r(t), \qquad \text{as} \quad k \to \infty. \tag{3.2}$$

Let $\omega_1 \in \Omega_1'$, $t \in \mathbb{R}^+$. Choose $t', t'' \in \mathbb{Q}^+$ such that $t' < t < t''$. Since $Z_k^{(r\infty)}(s, \omega_1)$, $s \in \mathbb{R}^+$, are increasing bounded functions of s, we have

$$Z_k^{(r\infty)}(t', \omega_1) \le Z_k^{(r\infty)}(t, \omega_1) \le Z_k^{(r\infty)}(t'', \omega_1).$$

Therefore, we obtain

$$\sigma_r(t') = \lim_{k\to\infty} Z_k^{(r\infty)}(t', \omega_1) \le \liminf_{k\to\infty} Z_k^{(r\infty)}(t, \omega_1) \le \limsup_{k\to\infty} Z_k^{(r\infty)}(t, \omega_1) \le$$

$$\le \lim_{k\to\infty} Z_k^{(r\infty)}(t'', \omega_1) = \sigma_r(t'').$$

Since σ_r is a continuous bounded function, $Z_k^{(r\infty)}(t, \omega_1) \to \sigma_r(t)$, as $n \to \infty$. (The boundedness of $\sigma_r(t)$ follows from $\sum_{k=1}^{\infty} \frac{k^{k-1}}{k!}(ae^{-a})^k = a$, see Remark 2.2.) As the functions are non-decresing, by Dini's theorem, this convergence is uniform. □

4 A functional limit theorem

In this section we shall study sequences of random processes with time scale determined by the functions $f_k(t)$. To construct our random processes, we need random elements defined on the probability space $\{\Omega, \mathcal{A}, \mathbb{P}\}$ (not on $\{\Omega_1, \mathcal{A}_1, \mathbb{P}_1\}$.

(Y) Let Y_n, Y_{ni}, $n, i \in \mathbb{N}$, be an array of random variables defined on $\{\Omega, \mathcal{A}, \mathbb{P}\}$. Assume that for any fixed $n \in \mathbb{N}$, the above random variables are independent and identically distributed.

We shall assume that the following condition is satisfied for the limiting behaviour of Y_{ni}.

$$\sum_{i=1}^{N_k} Y_{ni} \xrightarrow{d} \gamma(v), \qquad \text{as} \quad k \to \infty. \tag{S}$$

Here $\gamma(v)$ denotes a centered normally distributed random variable with variance v^2. We see that condition (S) implies that the array Y_{ni} is uniformly infinitesimal.

Let $r \in \mathbb{N}$. We will consider for each $k \in \mathbb{N}$ the random step function

$$X_k^{(r\infty)}(t) = X_k^{(r\infty)}[Y](t) = X_k^{(r\infty)}[Y, \omega_1](t) = \sum_{i=1}^{N_k} \mathbb{I}^{(r\infty)}_{N_k f_k(t) i}(\omega_1) Y_{ki} \,. \tag{Z1}$$

The process $X_k^{(r\infty)}(t)$ has the following interpretation. We consider an evolution during time $t \in [0, \infty)$ of a random forest with N_k ordered rooted trees. At the begining the random forest has N_k trees such that each tree consists of a root vertex only. We assume that at certain moments of time t nonroot vertices are added. Vertices are adding randomly and such that at the moment of time t we have a random forest with $f_k(t)$ nonroot vertices and N_k trees. Moreover, we assume that at each time instant t, the distribution on the set of forests is uniform. Consider the sum $\sum_{i=1}^{N_k} Y_{ki}$. Now delete from this sum the term Y_{ki} if the ith tree of the forest has less than r nonroot vertices. Then we obtain $X_k^{(r\infty)}(t)$.

Let W denote the standard Wiener process.

Theorem 4.1. *Let conditions of Theorem 3.2,* (Y) *and* (S) *be valid. Let* $r \in \mathbb{Z}^+$*. Then for the processes* $X_k^{(r\infty)}(t)$*, defined by* (Z1)*, one has*

$$X_k^{(r\infty)}[Y, \omega_1] \xrightarrow{d} X^{(r\infty)}, \qquad \text{as} \quad k \to \infty,$$

in $D[0, \infty)$ *for almost all* $\omega_1 \in \Omega_1$*, where* $X^{(r\infty)}(t) = vW(\sigma_r(t))$, $t \in [0, \infty)$.

We will use the following criteria of the convergence in $D[0, \infty)$.

Lemma 4.1. (1) *Let* $U(t)$, $U_n(t)$, $t \in [0, 1]$, $n \in \mathbb{N}$*, be random elements in* $D[0, 1]$ *(under its uniform metric and projection* σ*-field). Suppose that* $\mathbb{P}(U \in A) = 1$ *for some separable subset* $A \subset D[0, 1]$*. The necessary and sufficient conditions for* $\{U_n\}$ *to converge in distributon (under the uniform metric) to* U *are*

(a) *the finite dimensional distributions of* U_n *converge to the finite dimensional distributions of* U*;*

(b) *for any* $\varepsilon > 0$ *and* $\delta > 0$ *there exist* $n_0 \in \mathbb{N}$ *and* $0 = t_0 < t_1 < \cdots < t_m = 1$ *such that for all* $n > n_0$

$$\mathbb{P}\Big\{ \max_{1 \le i \le m} \sup_{t_{i-1} \le t < t_i} |U_n(t) - U_n(t_{i-1})| > \delta \Big\} < \varepsilon.$$

(2) *Let* L_k *denote the truncation map from* $D[0, \infty)$ *to* $D[0, k]$*. Let* $U(t)$, $U_n(t)$, $t \in [0, \infty)$, $n \in \mathbb{N}$*, be random elements in* $D[0, \infty)$ *(under its uniform metric and projection* σ*-field). Suppose that* $\mathbb{P}(U \in A) = 1$ *for some separable subset* $A \subset D[0, \infty)$*. Then* $\{U_n\}$ *converges in distributon in* $D[0, \infty)$ *(under the uniform metric) to* U *if and only if* $\{L_k U_n\}$ *converges in distributon in* $D[0, k]$ *(under the uniform metric) to* $L_k U$ *for each fixed* k.

Part (1) of Lemma 4.1 is Theorem 3, while part (2) is Theorem 23 in Chapter V of Pollard [16].

The following lemma is a consequence of Theorem 16 of Chapter IV in Petrov [14]. (See also the normal convergence criterion at p. 311 of [11], moreover see [8].)

Lemma 4.2. *Let* (Y) *be fulfilled.*

(1) *Condition* (S) *is valid if and only if*

   (a) *for all* $\varepsilon > 0 \quad k_n \mathbb{P}\{|Y_n| > \varepsilon\} \to 0$, *as* $n \to \infty$;

   (b) $k_n \mathbb{E} Y_n \mathbb{I}_{\{|Y_n| \le 1\}} \to 0$, *as* $n \to \infty$;

   (c) $k_n \mathbb{D}^2 (Y_n \mathbb{I}_{\{|Y_n| \le 1\}}) \to v^2$, *as* $n \to \infty$.

(2) *Let* (S) *be valid and* $b_{ni} \in \mathbb{R}$, $1 \le i \le k_n$, $n \in \mathbb{N}$. *Assume that there exist* $0 < \beta_1 < \beta_2 < \infty$ *such that for any* $i \in \{1, \dots, k_n\}$, $n \in \mathbb{N}$ *either* $\beta_1 \le |b_{ni}| \le \beta_2$ *or* $b_{ni} = 0$. *Let* $U_n = \sum\limits_{i=1}^{k_n} b_{ni} Y_{ni}$, $n \in \mathbb{N}$. *Then* $U_n \xrightarrow{d} \gamma(s)$, *as* $n \to \infty$, *if and only if* $\mathbb{D}^2 (Y_n \mathbb{I}_{\{|Y_n| \le 1\}}) \sum_{i=1}^{k_n} (b_{ni})^2 \to s^2$, *as* $n \to \infty$.

Proof of Theorem 4.1. If instead of Y_{ki} we write $Y_{ki} \mathbb{I}_{\{|Y_{ki}| \le 1\}} - \mathbb{E} Y_{ki} \mathbb{I}_{\{|Y_{ki}| \le 1\}}$, $\mathbb{E} Y_{ki} \mathbb{I}_{\{|Y_{ki}| \le 1\}}$ or $Y_{ni} \mathbb{I}_{\{|Y_{ni}| > 1\}}$ in the definition of $X_k^{(r\infty)}$, then the process obtained will be denoted by $X_k^{(r\infty)}(Y^<)$, $X_k^{(r\infty)}(EY)$ and $X_k^{(r\infty)}(Y^>)$, respectively. We have

$$X_k^{(r\infty)} = X_k^{(r\infty)}(Y^<) + X_k^{(r\infty)}(EY) + X_k^{(r\infty)}(Y^>). \tag{4.1}$$

We see that

$$\|X_k^{(r\infty)}(EY)\| \le \left(\frac{1}{N_k} \sum_{i=1}^{N_k} \mathbb{I}^{(r\infty)}_{N_k f_k(\infty) i}(\omega_1) \right) N_k |\mathbb{E} Y_k \mathbb{I}_{\{|Y_k| \le 1\}}|.$$

Observe that (S) implies that $|N_k \mathbb{E} Y_k \mathbb{I}_{\{|Y_k| \le 1\}}| \to 0$, as $k \to \infty$. Consequently, $X_k^{(r\infty)}(EY) \to 0$, as $k \to \infty$ for almost all $\omega_1 \in \Omega_1$.

Also we have

$$\mathbb{P}\{\|X_k^{(r\infty)}(Y^>)\| > 0\} \le \left(\frac{1}{N_k} \sum_{i=1}^{N_k} \mathbb{I}^{(r\infty)}_{N_k f_k(\infty) i}(\omega_1) \right) N_k |\mathbb{P}\{|Y_{ki}| > 1\}.$$

Now, (S) implies that $N_k \mathbb{P}\{|Y_k| \ge 1\} \to 0$, as $k \to \infty$. Consequently, $X_k^{(r\infty)}(Y^>) \to 0$, as $k \to \infty$, in probability in D for almost all $\omega_1 \in \Omega_1$.

Therefore we must prove the theorem for the processes $X_k^{(r\infty)}(Y^<)$. That is we can assume that Y_{ki} are independent centered random variables with the Lindeberg-Feller property.

Let $\Omega_1' \subset \Omega_1$ be from Theorem 3.2. Suppose that $\omega_1 \subset \Omega_1'$. Then, by Theorem 3.2, $Z_k^{(r\infty)}[\omega_1] \to \sigma_r$, as $k \to \infty$ in D. The functions $Z_k^{(r\infty)}[\omega_1](t)$ and $\sigma_r(t)$ are increasing and bounded, moreover $\sigma_r(t)$ is continuous. Now the convergence of the finite dimensional distributions follows from (S) and from the fact that both the process $X_k^{(r\infty)}(t)$ and $vW(\sigma_r(t))$ have independent increments.

To prove criterion (b) in Lemma 4.1 (1), we apply the method of the proof of Donsker's theorem, i.e. follow the lines of theorems 8.3 and 10.1 in Chapter 2 of Billingsley [2] (see also Chuprunov-Rusakov [6], Theorem B and Theorem C). So Theorem 4.1 follows from Lemma 4.1. □

5 Proof of Lemma 2.1

Proof. (1) Let $g_i = \mathbb{I}_{Nni}^{(r)} - \mathbb{E}_1 \mathbb{I}_{Nni}^{(r)}$. We shall use the following decomposition

$$\begin{aligned} A &= \mathbb{E}_1 \left\{ \sum_{i=1}^{N} g_i \right\}^4 = \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \sum_{i_3=1}^{N} \sum_{i_4=1}^{N} \mathbb{E}_1 (g_{i_1} g_{i_2} g_{i_3} g_{i_4}) = \qquad (5.1) \\ &= N\mathbb{E}_1(g_1)^4 + 3N(N-1)\mathbb{E}_1(g_1)^2(g_2)^2 + 4N(N-1)\mathbb{E}_1(g_1)^3 g_2 + \\ &\quad +6N(N-1)(N-2)\mathbb{E}_1(g_1)^2 g_2 g_3 + N(N-1)(N-2)(N-3)\mathbb{E}_1 g_1 g_2 g_3 g_4 = \\ &= A_1 + A_2 + A_3 + A_4 + A_5 \, . \end{aligned}$$

We can see that $E_1/L \to 1$, as $n, N \to \infty$. Therefore $\mathbb{E}_1|g_i|^2 \le c_0 L$. Using this inequality and that $|g_i| \le 1$, we obtain

$$A_1 + A_2 + A_3 \le N\mathbb{E}_1(g_1)^2 + 3N(N-1)\mathbb{E}_1(g_1)^2 + 4N(N-1)\mathbb{E}_1(g_1)^2 \le 7N^2 c_0 L \, . \qquad (5.2)$$

Now we will find an upper bound for A_5. Using Newton's binomial theorem, we have

$$|A_5| = |N(N-1)(N-2)(N-3)\mathbb{E}_1(g_1 g_2 g_3 g_4)| <$$

$$< |N^4 \mathbb{E}_1(g_1 g_2 g_3 g_4)| = N^4 |E_4 - 4E_1 E_3 + 6E_1^2 E_2 - 4E_1^4 + E_1^4| =$$
$$= N^4 |E_4 - 4E_1 E_3 + 6E_1^2 E_2 - 3E_1^4|.$$

We have for each j

$$E_j = \mathbb{E}_1 \mathbb{I}_{Nn1}^{(r)} \mathbb{I}_{Nn2}^{(r)} \dots \mathbb{I}_{Nnj}^{(r)} = \qquad (5.3)$$
$$= \frac{C_n^r C_{n-r}^r \dots C_{n-(j-1)r}^r (1+r)^{j(r-1)} (N-j)(N-j+n-jr)^{n-jr-1}}{N(N+n)^{n-1}} =$$
$$= L^j B_j D_j \frac{N-j}{N} \lambda^{-rj} e^{(r+1)\lambda j}$$

where

$$B_j = \frac{(N-j+n-jr)^{n-jr-1}}{(N+n)^{n-jr-1}}, \quad D_j = \frac{n(n-1)(n-2)\dots(n-jr+1)}{(n+N)^{jr}}.$$

Observe that, by Taylor's formula, it holds that

$$\ln(1-x) = -x - \frac{1}{2\xi^2}x^2, \tag{5.4}$$

where $x > 0$ and $1 - x < \xi < 1$ and

$$e^{-x} = 1 - \frac{x}{1!} + e^{\theta}\frac{x^2}{2!}, \tag{5.5}$$

where $x > 0$ and $-x < \theta < 0$.

We have the following estimates for $j = 1, 2, 3, 4$ and $r > 0$:

$$D_j = \frac{n(n-1)(n-2)\dots(n-jr+1)}{(n+N)^{jr}} =$$

$$= \lambda^{jr}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\dots\left(1-\frac{jr-1}{n}\right) = \lambda^{jr}\exp\left(\sum_{k=1}^{jr-1}\ln\left(1-\frac{k}{n}\right)\right) =$$

$$= \lambda^{jr}\exp\left(-\frac{jr(jr-1)}{2n} - \sum_{k=1}^{jr-1}\frac{1}{2\xi_k^2}\frac{k^2}{n^2}\right),$$

where, by (5.4) and $(4(r+1))^4/n < 0.001$, the inequality $1 - 0.001 < \xi_k < 1$ holds. Let $f_{kj} = \frac{1}{2\xi_k^2}$. Therefore we obtain

$$D_j = \lambda^{jr}\exp\left(-\frac{jr(jr-1)}{2n} - \sum_{k=1}^{jr-1} f_{kj}\frac{k^2}{n^2}\right) =$$

$$= \lambda^{jr}\exp\left(-\frac{jr(jr-1)}{2n} - f_j''\frac{1}{n^2}\frac{(jr-1)jr(2jr-1)}{6n^2}\right) =$$

$$= \lambda^{jr}\exp\left(-\frac{jr(jr-1)}{n} - f_j'\frac{(j(r+1))^3}{n^2}\right),$$

where $0 < f_j'' < 0.502$ and $0 < f_j' < 0.17$. Therefore, by (5.5), we obtain

$$D_j = \lambda^{jr}\left(1 - \frac{jr(jr-1)}{2n} - f_j'\frac{(j(r+1))^3}{n^2} + \frac{e^{\theta}}{2}\left(\frac{jr(jr-1)}{2n} + f_j'\frac{(j(r+1))^3}{n^2}\right)^2\right)$$

where $-\frac{jr(jr-1)}{n} - f_j'\frac{(j(r+1))^3}{n^2} < \theta < 0$. Finally, for $r > 0$ we have

$$D_j = \lambda^{jr}\left(1 - \frac{jr(jr-1)}{2n} + f_j\frac{(j(r+1))^4}{n^2}\right), \tag{5.6}$$

where $|f_j| < 1$.

Moreover, $D_j = 1$ for $r = 0$.

Observe that, by Taylor's formula, it holds that

$$\ln(1-x) = -x - \frac{1}{2}x^2 - \frac{1}{3\xi^3}x^3, \tag{5.7}$$

where $x > 0$ and $1 - x < \xi < 1$.

By (5.7), we have the following estimates

$$B_j = \frac{(N-j+n-jr)^{n-jr-1}}{(N+n)^{n-jr-1}} = \left(1 - \frac{j(r+1)}{N+n}\right)^{n-(jr+1)} =$$

$$= \exp\left((n-(jr+1))\ln\left(1-\frac{j(r+1)}{N+n}\right)\right) =$$

$$= \exp\left((n-(jr+1))\left(-\frac{j(r+1)}{N+n} - \frac{1}{2}\left(\frac{j(r+1)}{N+n}\right)^2 - \frac{1}{3(\xi_j')^3}\left(\frac{j(r+1)}{N+n}\right)^3\right)\right),$$

where $0.999 < \xi_j' < 1$. Therefore it holds that

$$B_j = \exp\left((n-(jr+1))\left(-\frac{j(r+1)}{N+n} - \frac{1}{2}\left(\frac{j(r+1)}{N+n}\right)^2 - h_j'\left(\frac{j(r+1)}{N+n}\right)^3\right)\right)$$

where $\frac{1}{3} < h_j' < \frac{1.007}{3}$. Cosequently, we obtain

$$B_j = \exp\left(-j(r+1)\lambda - \frac{\lambda}{2}\frac{(j(r+1))^2}{N+n} + \frac{(jr+1)j(r+1)}{N+n} - \right.$$

$$\left. -h_j'\lambda\frac{(j(r+1))^3}{(N+n)^2} + \frac{(jr+1)(j(r+1))^2}{2(N+n)^2} + h_j'\frac{(jr+1)(j(r+1))^3}{(N+n)^3}\right) =$$

$$= \exp\left(-j(r+1)\lambda - \frac{\lambda}{2}\frac{(j(r+1))^2}{N+n} + \frac{(jr+1)(j(r+1))}{N+n} + h_j^e\frac{(j(r+1))^3}{(N+n)^2}\right)$$

where $|h_j^e| < \frac{1.007}{3} + \frac{1}{2} + 0.001 < 1$. Therefore we have

$$B_j = e^{-j(r+1)\lambda}\exp\left(-\frac{\lambda}{2}\frac{(j(r+1))^2}{N+n} + \frac{(jr+1)j(r+1)}{N+n} + h_j^e\frac{(j(r+1))^3}{(N+n)^2}\right)$$

where $|h_j^e| < 1$. Thus, by (5.5), we obtain

$$B_j = e^{-j(r+1)\lambda}\left\{1 - \frac{\lambda}{2}\frac{(j(r+1))^2}{N+n} + \frac{(jr+1)(j(r+1))}{N+n} + h_j^e\frac{(j(r+1))^3}{(n+N)^2} + \right.$$

$$\left. \frac{1}{2}e^{\theta}\left(-\frac{\lambda}{2}\frac{(j(r+1))^2}{N+n} + \frac{(jr+1)(j(r+1))}{N+n} + h_j^e\frac{(j(r+1))^3}{(n+N)^2}\right)^2\right\}$$

where

$$|\theta| < \left| -\frac{\lambda}{2}\frac{(j(r+1))^2}{N+n} + \frac{(jr+1)(j(r+1))}{N+n} + h_j^e\frac{(j(r+1))^3}{(n+N)^2} \right| <$$

$$< \frac{1}{2}0.001 + 0.001 + 0.001 = 0.0025.$$

Consequently $e^\theta < 1.003$ and

$$B_j = e^{-j(r+1)\lambda}\left(1 - \frac{\lambda}{2}\frac{(j(r+1))^2}{N+n} + \frac{(jr+1)(j(r+1))}{N+n} + h_j\frac{(j(r+1))^4}{(n+N)^2}\right), \tag{5.8}$$

where $|h_j| < 1 + \frac{1.003}{2}(0.0005 + 0.001 + 0.00001)^2 < 1.1$. Now, by (5.3), (5.6), and (5.8), we obtain

$$E_j = L^j\left(1 - \frac{jr(jr-1)}{2n} + f_j\frac{(j(r+1))^4}{n^2}\right) \times$$

$$\times \left(1 - \frac{\lambda}{2}\frac{(j(r+1))^2}{N+n} + \frac{(jr+1)(j(r+1))}{N+n} + h_j\frac{(j(r+1))^4}{(N+n)^2}\right)\left(1 - \frac{j}{N}\right) =$$

$$= L^j\left(1 - \frac{jr(jr-1)}{2n} - \frac{\lambda}{2}\frac{(j(r+1))^2}{N+n} + \frac{(jr+1)(j(r+1))}{N+n} - \frac{j}{N} + g'\frac{(j(r+1))^4}{n^2}\right)$$

where $|g'| < 4.5 + 2\alpha$ with $\alpha = n/N$. So we have

$$E_4 = L^4\left(1 - \frac{4r(4r-1)}{2n} - \frac{\lambda}{2}\frac{(4(r+1))^2}{N+n} + \frac{(4r+1)(4(r+1))}{n}\lambda - \frac{4}{N} + g_1\frac{(4(r+1))^4}{n^2}\right),$$

$$E_1E_3 = L^4\left(1 - \frac{r(r-1)}{2n} - \frac{\lambda}{2}\frac{(r+1)^2}{N+n} + \frac{(r+1)(r+1)}{n}\lambda - \frac{1}{N} - \right.$$

$$\left. - \frac{3r(3r-1)}{2n} - \frac{\lambda}{2}\frac{(3(r+1))^2}{N+n} + \frac{(3r+1)(3(r+1))}{n}\lambda - \frac{3}{N} + g_2\frac{(4(r+1))^4}{n^2}\right),$$

$$E_1^2E_2 = L^4\left(1 - 2\frac{r(r-1)}{2n} - 2\frac{\lambda}{2}\frac{(r+1)^2}{N+n} + 2\frac{(r+1)(r+1)}{n}\lambda - \frac{2}{N} - \right.$$

$$\left. - \frac{2r(2r-1)}{2n} - \frac{\lambda}{2}\frac{(2(r+1))^2}{N+n} + \frac{(2r+1)(2(r+1))}{n}\lambda - \frac{2}{N} + g_3\frac{(4(r+1))^4}{n^2}\right)$$

and

$$E_1^4 = L^4\left(1 - 4\frac{r(r-1)}{2n} - 4\frac{\lambda}{2}\frac{(r+1)^2}{N+n} + 4\frac{(r+1)(r+1)}{n}\lambda - \frac{4}{N} + g_4\frac{(4(r+1))^4}{n^2}\right)$$

where g_j, $j = 1, 2, 3, 4$, are bounded with certain polynomials of α. (We can give e.g. the following bounds: $|g_1| < 4.5 + 2\alpha$, $|g_2| < 13.1 + 12.1\alpha + 3.1\alpha^2$, $|g_3| < 110 + 119\alpha + 52\alpha^2 + 7.1\alpha^3$, $|g_4| < 9.3 + 6.8\alpha + 1.8\alpha^2 + 0.6\alpha^3 + 0.1\alpha^4$.) Finally we obtain

$$|A_5| < N^4(E_4 - 4E_1E_3 + 6E_1^2E_2 - 3E_1^4) =$$

$$= N^4 \left(\frac{(1+r)^{r-1}}{r!} e^{-(r+1)\lambda} \lambda^r \right)^4 \left(0+0+0+0+g\frac{(4(r+1))^4}{n^2} \right),$$

where $|g|$ is bounded with a certain polynomial of α. That is

$$|A_5| < N^2 L^4 (r+1)^4 \frac{g(\alpha)}{\alpha^2} \tag{5.9}$$

where $g(\alpha)$ is a polynomial of α.

Using the above equalities for $E_1, \dots, E_4$, we obtain

$$A_4 \le N^3 L^3 \left(\frac{c(r+1)^2}{n} + \frac{c}{n^2} \right) + N^3 L^4 \left(\frac{c(r+1)^2}{n} + \frac{(r+1)^4}{n^2} p(\alpha) \right) \tag{5.10}$$

where $p(\alpha)$ is a polynomial of α.

Summarizing (5.2), (5.9), and (5.10), we obtain (2.5).

(2) First consider (5.9), that is A_5. Let

$$a_r = (L(r,\lambda))^3 (r+1)^4 = \left(\frac{(r+1)^{r-1}}{r!} e^{-(r+1)\lambda} \lambda^r \right)^3 (r+1)^4.$$

Then $\frac{a_{r+1}}{a_r} \le \varrho < 1$ for all $r > r_0$ if $\lambda = \frac{n}{n+N} \le \tau < 1$. Therefore $N^2 L^4 (r+1)^4 \frac{g(\alpha)}{\alpha^2} \le N^2 L C_1 \frac{g(\alpha)}{\alpha^2}$ where C_1 depends on λ. Now consider (5.10), that is A_4. The second summand can be handled as A_5. For the first summand we remark that $L^2(r+1)^2 \to 0$ $(r \to \infty)$ if $\lambda \le \tau < 1$. Therefore $L^2(r+1)^2$ is bounded. So $N^3 L^3 \frac{c(r+1)^2}{n} \le N^2 L C_1 \frac{1}{\alpha}$. Therefore (5.2), (5.9), and (5.10) imply (2.6). □

References

[1] Balińska, K. T.; Quintas, L. V. and Szymański, J.: Random recursive forests. Proceedings of the Fifth International Seminar on Random Graphs and Probabilistic Methods in Combinatorics and Computer Science (Poznań, 1991). *Random Structures Algorithms* **5** (1994), no. 1, 3–12.

[2] Billingsley, P.: *Convergence of probability measures.* John Wiley & Sons, Inc., New York-London-Sydney, 1968.

[3] Bollobás, Béla: *Random graphs.* Cambridge University Press, Cambridge, 2001.

[4] Breiman, Leo: Random forests. *J. Mach. Learn.* **45**, No.1, 5–32 (2001).

[5] Chuprunov, A. and Fazekas, I.: Inequalities and strong laws of large numbers for random allocations. *Acta Math. Hungar.* **109** (2005), no. 1-2, 163–182.

[6] Chuprunov, A. N. and Rusakov, O. V.: Convergence for step line processes under summation of random indicators and models of market pricing. *Lobachevskii J. Math.* **12** (2003), 11–39 (electronic).

[7] Erdős, P. and Rényi, A.: On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **5** (1960), 17–61.

[8] Gnedenko, B. V. and Kolmogorov, A. N.: *Limit distributions for sums of independent random variables*. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills., Ont. 1968.

[9] Janson, S. Łuczak, T. and Rucinski, A.: *Random graphs*. Wiley-Interscience, New York, 2000.

[10] Kolchin, V. F.: *Random graphs*. Encyclopedia of Mathematics and its Applications, 53. Cambridge University Press, Cambridge, 1999.

[11] Loève, M.: *Probability theory*. D. Van Nostrand Co., Inc., Princeton, N.J.-Toronto, Ont.-London, 1963.

[12] Mutafchiev, Ljuben: The largest tree in certain models of random forests. Proceedings of the Eighth International Conference "Random Structures and Algorithms" (Poznan, 1997). *Random Structures Algorithms* **13** (1998), no. 3-4, 211–228.

[13] Pavlov, Ju. L.: Limit theorems for the number of trees of a given size in a random forest. (Russian) *Mat. Sbornik (N.S.)* **103**(145) (1977), no. 3, 392–403.

[14] Petrov, V. V.: *Sums of independent random variables*. Springer-Verlag, New York-Heidelberg, 1975.

[15] Pitman, Jim: Coalescent random forests. *J. Combin. Theory Ser. A* **85** (1999), no. 2, 165–193.

[16] Pollard, D.: *Convergence of stochastic processes*. Springer-Verlag, New York–Berlin–Heidelberg–Tokyo, 1984.

[17] Prudnikov, A. P., Brychkov, Yu. A. and Marichev, O. I.: *Integrals and series*. Vol. 1. Elementary functions. Gordon & Breach Science Publishers, New York, 1986.

[18] Rényi, A.: Some remarks on the theory of trees. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **4** (1959), 73–85.