On the maximum tree size in the Galton–Watson forest

E. V. Khvorostyanskaya

Karelian Research Centre, Petrozavodsk, Russia, cher@krc.karelia.ru

Let G_N be the critical Galton-Watson branching process with N initial particles and let the number of offspring of each particle be a random variable ξ following the distribution

$$p_k = \mathbf{P}\{\xi = k\} = \frac{1}{(k+1)^\tau} - \frac{1}{(k+2)^\tau},\ k = 0, 1, 2, \dots \tag{1}$$

The process G_N induces a conditional probability distribution on the subset $F_{N,n}$ of its trajectories with $N+n$ vertices provided that the number of vertices is equal to $N+n$. We denote by $\mathfrak{F}_{N,n}$ the thus constructed Galton – Watson forest with N trees and n non-rooted vertices. The Galton – Watson forests with N trees and n non-rooted vertices were studied by Pavlov [1] as $N, n \to \infty$ and the third moment of the offspring distribution was finite.

It is easy to show that $m = \mathbf{E}\xi = \zeta(\tau) - 1$, where $\zeta(x) = \sum\limits_{k=1}^{\infty} k^{-x}$ is the Riemann zeta – function. Since the branching process G_N is critical, the equality $m = 1$ holds and the value of the parameter τ is determined by the relation $\zeta(\tau) = 2$, therefore $\tau \approx 1.728$. For this value of the parameter the offspring distribution (1) has only the finite first moment and such Galton – Watson forest has not been studied yet.

In [1] it was shown that the forest class $\mathfrak{F}_{N,n}$ corresponds to the Galton – Watson branching process $G_N(\lambda)$ with N initial particles and the offspring distribution

$$p_k(\lambda) = \frac{\lambda^k p_k}{F(\lambda)},\ k = 0, 1, 2, \dots,\ 0 < \lambda \leqslant 1, \tag{2}$$

where $F(z) = \sum\limits_{k=0}^{\infty} p_k z^k$. The process $G_N(\lambda)$ consists of N independent processes $G^{(1)}(\lambda), G^{(2)}(\lambda), \dots, G^{(N)}(\lambda)$ starting with one particle. We denote by $\nu^{(1)}, \nu^{(2)}, \dots, \nu^{(N)}$ independent identically distributed random variables equal to the numbers of particles that existed in the processes $G^{(1)}(\lambda), G^{(2)}(\lambda), \dots, G^{(N)}(\lambda)$ before they degenerate, and let ν_N be a random variable equal to the total number of particles that existed in the process $G_N(\lambda)$ before its degeneration, $\nu_N = \nu^{(1)} + \nu^{(2)} + \dots + \nu^{(N)}$.

Let $\nu_1(\mathfrak{F}), \nu_2(\mathfrak{F}), \dots, \nu_N(\mathfrak{F})$ be random variables equal to the tree sizes in the forest $\mathfrak{F}_{N,n}$. In [1] it was shown that

$$\begin{gathered}\mathbf{P}\{\nu_1(\mathfrak{F}) = k_1, \dots, \nu_N(\mathfrak{F}) = k_N\} \\ = \mathbf{P}\left\{\nu^{(1)} = k_1, \dots, \nu^{(N)} = k_N | \nu_N = N+n\right\}.\end{gathered} \tag{3}$$

The equality (3) means that the random variables $\nu_1(\mathfrak{F}), \nu_2(\mathfrak{F}), \ldots, \nu_N(\mathfrak{F})$ and $\nu^{(1)}, \nu^{(2)}, \ldots, \nu^{(N)}$ form a generalized allocation scheme (Kolchin [2]). The use of the generalized allocation scheme method allows to reduce the problem with dependent random variables to the study of independent random variables, while the value of the parameter of their distribution can be chosen in the most convenient way. We assume that the parameter λ of the distribution (2) is equal to the solution of the equation

$$\frac{\lambda F^{'}(\lambda)}{F(\lambda)} = \frac{n}{N+n}.$$

Let $\eta(\mathfrak{F})$ be a random variable equal to the maximum tree size in a Galton – Watson forest $\mathfrak{F}_{N,n}$.

The following statement was obtained.

Theorem. *Let $N, n, r \to \infty$ in such a way that $n/N \to \infty$, $n/N^{\tau} \to 0$. Then*

$$\mathbf{P}\left\{\beta\eta(\mathfrak{F}) - u \leqslant z\right\} \to e^{-e^{-z}},$$

where $\beta = \ln\left(F(\lambda)/\lambda\right)$, u is chosen so that $NC(\tau)\beta^{1/\tau}u^{-(1+1/\tau)}e^{-u} = 1$,

$$C(\tau) = \frac{\Gamma(1/\tau)\cos(\pi(2-\tau)/2\tau)}{\pi\tau\left(\Gamma(1-\tau)\cos(\pi\tau/2)\right)^{1/\tau}}.$$

The idea of proof is the following. From (3) the equality

$$\mathbf{P}\left\{\eta(\mathfrak{F}) \leqslant r\right\} = (1-P_r)^N \frac{\mathbf{P}\left\{\nu_{r,N} = N+n\right\}}{\mathbf{P}\left\{\nu_N = N+n\right\}}$$

was derived [2], where

$$P_r = \mathbf{P}\left\{\nu^{(1)} > r+1\right\}, \quad \nu_{r,N} = \nu_r^{(1)} + \ldots + \nu_r^{(N)},$$

$\nu_r^{(1)}, \ldots, \nu_r^{(N)}$ are independent random variables following the distribution

$$\mathbf{P}\left\{\nu_r^{(i)} = k\right\} = \mathbf{P}\left\{\nu^{(1)} = k \mid \nu^{(1)} \leqslant r+1\right\}, \; k=1,2,\ldots, \; i=1,\ldots,N.$$

To obtain the limit distribution of $\eta(\mathfrak{F})$ it is enough to find the asymptotic behavior of the binomial $(1-P_r)^N$ and probabilities $\mathbf{P}\left\{\nu_{r,N} = N+n\right\}$ and $\mathbf{P}\left\{\nu_N = N+n\right\}$.

References

1. Yu. L. Pavlov, *Random forests*, VSP, Utrecht, 2000.

2. V. F. Kolchin, *Random graphs*, Cambridge: Cambridge Univ. Press, 1999.