Информационные процессы, Том 2, № 2, 2002, стр. 165.

KALASHNIKOV MEMORIAL SEMINAR

Limit Distribution of The Number of Chains in a Random Forest[1]

E.V. Cherepanova

Institute of Applied Mathematical Research,
Karelian Research Center of RAS
email: cher@krc.karelia.ru
Received October 14, 2002

Let $F_{N,n}$ be a set of different forests with N rooted trees and n non-root vertices such that roots have the numbers $1, 2, \ldots, N$ and non-root vertices have the numbers $1, 2, \ldots, n$. We specify the uniform distribution on $F_{N,n}$. Such random forests were studied in [1, 2]. The pair is any two non-root vertices connected by a chain. We denote by ν the number of different chains in a forest from $F_{N,n}$. It is easy to see that $\nu = \tau + n$, where τ is the number of pairs.

Theorem. *Let $N,\ n \longrightarrow \infty$ in such a way that $n/N = O(1)$, $n^2/N \longrightarrow \infty$. Then*

$$\mathbf{P}\left\{\nu - n = k\right\} = \frac{1}{\sigma\sqrt{2\pi}} exp\left\{-\frac{(k-a)^2}{2\sigma^2}\right\}(1+o(1))$$

uniformly in $(k-a)/\sigma$ lying in any fixed finite interval, where $a = n^2(n+3N)/(2N^2),\ \sigma^2 = 3n^2(n+N)^5/(2N^6)$.

To prove the theorem we can use the generalized allocation scheme [3]. It was shown in [1, 2] that

$$\mathbf{P}\left\{\eta_1 = k_1, \ldots, \eta_N = k_N\right\} = \\ = \mathbf{P}\left\{\xi_1 = k_1, \ldots, \xi_N = k_N \mid \xi_1 + \cdots + \xi_N = n\right\},$$

where η_i is the number of non-root vertices in the i-th tree, $1 \leq i \leq N, \xi_1, \ldots, \xi_N$ are independent identically distributed random variables with the Borel – Tanner distribution. It folows from this that

$$\mathbf{P}\left\{\tau = k\right\} = \mathbf{P}\left\{\zeta_N = k \mid \mu_N = n\right\} = \frac{\mathbf{P}\left\{\zeta_N = k,\ \mu_N = n\right\}}{\mathbf{P}\left\{\mu_N = n\right\}},$$

where $\zeta_N = \xi_1(\xi_1 - 1)/2 + \cdots + \xi_N(\xi_N - 1)/2,\ \mu_N = \xi_1 + \cdots + \xi_N$. The main difficulty in the proof of the theorem is to get local limit distributions of array schemes of the sums μ_N and (ζ_N, μ_N).

REFERENCES

1. Yu.L. Pavlov. Limit distributions of the number of trees of a given size in a random forest. *Soviet Math. Sb.*, 1977, 32, pp. 335–345.
2. Yu.L. Pavlov. *Random Forests.* Utrecht: VSP, 2000.
3. V.F. Kolchin. *Random Mappings.* New York: Springer, 1986.

[1] Research is supported by the Russian Foundation for Basic Research, grant 00-01-00233.