Inversions in random node labeling of random trees

Xing Shi Cai, Cecilia Holmgren, Svante Janson, Tony Johansson, Fiona Skerman

June 1, 2017

Inversions in labeled trees are a generalization of inversions in permutations. We study the number of inversions in trees labeled uniformly at random. The three types of trees that we consider — complete b-ary trees, split trees and conditional Galton-Watson trees — cover a wide range of tree models. For each of the three, we show the distribution converges to a limit. In particular, our result on conditional Galton-Watson trees strengthens previous work by [11].

Let $\sigma_1, \dots, \sigma_n$ be a permutation of $\{1, \dots, n\}$. If $i < j$ and $\sigma_i > \sigma_j$, then the pair (σ_i, σ_j) is called an inversion. The study of inversions is motivated by its applications in the analysis of sorting algorithms, see, e.g., [7, sec. 5.1]. Many authors, including [4, pp. 256], [12, pp. 29], [2], have shown that the number of inversions in uniform random permutations has a central limit theorem. More recently, [10] and [8] studied the enumeration of permutations containing a fixed number of inversions.

The concept of inversion can be generalized as follows. Consider an unlabeled rooted tree T on node set V. Let ρ denote the root, and write $u < v$ if u is an ancestor of v, i.e., the unique path from ρ to v passes through u, and write $u \perp v$ if neither $u < v$ nor $u > v$. Given a bijection $\lambda : V \to \{1, \dots, |V|\}$ (a *node labeling*), define the number of *inversions*

$$I(T, \lambda) = \sum_{u<v} \mathbf{1}_{\lambda(u)>\lambda(v)}.$$

Note that if T is a path, then $I(T, \lambda)$ is the number of inversions in a permutation.

The enumeration of trees with a fixed number of inversions has been studied by [9] and [6] using the so called *inversions polynomial*. While studying linear probing hashing, [5] noticed that the numbers of inversions in Cayley trees with uniform random labeling converges to an Airy distribution. [11] showed that this is also true for conditional Galton-Watson trees, which covers the case of Cayley trees.

Let $I(T)$ denote the random variable given by $I(T) = I(T, \lambda)$ where λ is chosen uniformly at random from the set of bijections from V to $\{1, \dots, |V|\}$. For any $u < v$ we have $\mathbb{P}\{\lambda(u) > \lambda(v)\} = 1/2$, and it immediately follows that for a fixed tree T,

$$\mathbb{E}[I(T)] = \sum_{u<v} \mathbb{E}\left[\mathbf{1}_{\lambda(u)>\lambda(v)}\right] = \frac{1}{2}\Upsilon(T),$$

where $\Upsilon(T)$ denotes the *total path length* (or *internal path length*) of T, i.e., the sum over all $v \in V$ of the distance from ρ to v. The total path length $\Upsilon(T)$ has previously been studied for the trees considered here, e.g., [3] for split trees and [1, cor. 9] for conditional Galton-Watson trees. Our focus will be on the deviation

$$X_n = \frac{I(T_n) - \mathbb{E}\left[I(T_n)\right]}{s(n)},$$

under some appropriate scaling $s(n)$, for three classes of fixed and random tree sequences T_n; b-ary trees, split trees and Galton-Watson trees.

References

[1] David Aldous. The continuum random tree. II. An overview. In *Stochastic analysis (Durham, 1990)*, volume 167 of *London Math. Soc. Lecture Note Ser.*, pages 23–70. Cambridge Univ. Press, Cambridge, 1991.

[2] Edward A. Bender. Central and local limit theorems applied to asymptotic enumeration. *J. Combinatorial Theory Ser. A*, 15:91–111, 1973.

[3] Nicolas Broutin and Cecilia Holmgren. The total path length of split trees. *Ann. Appl. Probab.*, 22(5):1745–1777, 2012.

[4] William Feller. *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons, Inc., New York-London-Sydney, 1968.

[5] P. Flajolet, P. Poblete, and A. Viola. On the analysis of linear probing hashing. *Algorithmica*, 22(4):490–515, 1998. Average-case analysis of algorithms.

[6] Ira M. Gessel, Bruce E. Sagan, and Yeong Nan Yeh. Enumeration of trees by inversions. *J. Graph Theory*, 19(4):435–459, 1995.

[7] Donald E. Knuth. *The art of computer programming. Vol. 3.* Addison-Wesley, Reading, MA, 1998. Sorting and searching, Second edition [of MR0445948].

[8] Guy Louchard and Helmut Prodinger. The number of inversions in permutations: a saddle point approach. *J. Integer Seq.*, 6(2):Article 03.2.8, 19, 2003.

[9] C. L. Mallows and John Riordan. The inversion enumerator for labeled trees. *Bull. Amer. Math. Soc.*, 74:92–94, 1968.

[10] Barbara H. Margolius. Permutations with inversions. *J. Integer Seq.*, 4(2):Article 01.2.4, 13, 2001.

[11] Alois Panholzer and Georg Seitz. Limiting distributions for the number of inversions in labelled tree families. *Ann. Comb.*, 16(4):847–870, 2012.

[12] Vladimir N. Sachkov. *Probabilistic methods in combinatorial analysis*, volume 56 of *Encyclopedia of Mathematics and its Applications.* Cambridge University Press, Cambridge, 1997. Translated from the Russian, Revised by the author.