

# Analysis of partial match queries

Ralph Neininger

## Abstract

We consider the problem of recovering items matching a partially specified pattern in multidimensional trees (quad trees and k-d trees). We assume the traditional model where the data consist of independent and uniform points in the unit square. For this model, in a structure on  $n$  points, it is known that the number of nodes  $C_n(\xi)$  to visit in order to report the items matching an independent and uniformly on  $[0, 1]$  random query  $\xi$  satisfies  $\mathbf{E}[C_n(\xi)] \sim \kappa n^\beta$ , where  $\kappa$  and  $\beta$  are explicit constants. We develop an approach based on the analysis of the cost  $C_n(x)$  of any fixed query  $x \in [0, 1]$ , and give precise estimates for the variance and limit distribution of the cost  $C_n(x)$ . Our results permit to describe a limit process for the costs  $C_n(x)$  as  $x$  varies in  $[0, 1]$ ; one of the consequences is that  $\mathbf{E}[\max_{x \in [0, 1]} C_n(x)] \sim \gamma n^\beta$ , which settles a question of Devroye. Based on joint work with N. Broutin and H. Sulzbach.