Houdré, Christian; İşlak, Ümit

A central limit theorem for the length of the longest common subsequences in random words. (English) Zbl 07644433

Summary: Let \((X_i)_{i\geq 1}\) and \((Y_i)_{i\geq 1}\) be two independent sequences of independent identically distributed (iid) random variables taking their values in a common finite alphabet and having the same law. Let \(LC_n\) be the length of the longest common subsequences of the two random words \(X_1 \cdots X_n\) and \(Y_1 \cdots Y_n\). Under a lower bound assumption on the order of its variance, \(LC_n\) is shown to satisfy a central limit theorem. This is in contrast to the limiting distribution of the length of the longest common subsequences in two independent uniform random permutations of \(\{1, \ldots, n\}\), which is shown to be the Tracy-Widom distribution.

MSC:
68R15 Combinatorics on words
05A05 Permutations, words, matrices
60C05 Combinatorial probability
60F05 Central limit and other weak theorems
60F10 Large deviations

Keywords:
central limit theorem; edit/Levenshtein distance; last passage percolation; longest common subsequences; optimal alignments; random permutations; random words; Stein’s method; supersequences; Tracy-Widom distribution; Ulam’s problem

Full Text: DOI arXiv

References: