Claudie Hassenforder · Sabine Mercier

Exact Distribution of the Local Score for Markovian sequences

Abstract Let $\mathbb{A} = (A_i)_{1 \leq i \leq n}$ be a sequence of letters taken in a finite alphabet Θ. Let $s : \Theta \to \mathbb{Z}$ be a scoring function and $\mathbb{X} = (X_i)_{1 \leq i \leq n}$ the corresponding score sequence where $X_i = s(A_i)$. The local score is defined as follows: $H_n = \max_{1 \leq i \leq j \leq n} \sum_{k=i}^{j} X_k$. We provide the exact distribution of the local score in random sequences in several models. We will first consider a Markov model on the score sequence $\mathbb{X}$, and then on the letter sequence $\mathbb{A}$. The exact P-value of the local score obtained with both models are compared thanks to several datasets. They are also compared with previous results using the independent model.

Keywords Markov chain · Local score · P-value · Sequence analysis