# Distribution of palindromes in DNA

Drago Špoljarić and Ivo Ugrina

The aim of this paper is to determine the distribution of palindromes of a fixed length in a DNA sequence. Palindrome in DNA is a part of the DNA sequence which is equal to its complementary sequence read backwards ($C \sim G$, $A \sim T$). We examine general case where the probability of occurrence of each nucleobase is arbitrary (with condition that they sum to one). We get an interesting result for the uniform case, where probability of occurrence of each nucleobase is equal. In that case the probability of having two overlapping palindromes is the same as if they were non-overlapping, i.e. independent. In a non-uniform case we have a dependence. What is also interesting is the change of the asymptotic distribution if we fix the length of the DNA sequence and increase the length of the palindromes. Also, we derive an error bound for the approximation in the both cases.