

*IPM-POLYTECHNIQUE-WPI Workshop on Bioinformatics and Biomathematics,
April 11-21, 2005, IPM, Tehran*

Motifs Statistics in DNA Sequences Analysis

S. Robin

INA-PG & INRA

Paris, France

Motif statistics are now part of the elementary statistics in DNA sequences analysis. They typically aim at discovering motifs having a biological function, or, given a known motif, to understand the organization of the sequence on the basis of the occurrences of this motif. Markov chains are the most popular statistical model in this field, but the compound Poisson process turns out to be an efficient challenger for some specific problems. We introduce these two models and derive some statistical properties of the distribution of the occurrences of a motif along a sequence. We then present several applications to promoter detection, detection of crucial regions in bacterial genomes, etc.