

INSTITUTUL DE MATEMATICĂ “SIMION STOILOW” AL ACADEMIEI ROMÂNE

Seminariile BIOMAT

*The problem of gene identification using hidden
semi-Markov models*

Vlad Barbu

(Université de Rouen)

Vineri 6 martie 2015, ora 11:30
IMAR, Amfiteatrul “Miron Nicolescu”, parter

Abstract: In this talk, we are interested in the problem of gene prediction along the genomic DNA and in the construction of associated mathematical tools.

Firstly, we present the problem of gene prediction in genomics that in fact turns out to be a general problem of detection of homogeneous regions along an observed sequence. This type of problem arises in a variety of applied fields, like genomics, speech/writing recognition, signal processing, finance, climatology, etc. Despite this variety of applications, a modeling technique common to all these problems can be presented like a hidden model.

Secondly, we briefly describe a usual approach used in the literature for this type of problem, namely the hidden Markov models. Alternatively, we propose a generalization of these hidden Markov models, the so called hidden semi-Markov models. After introducing the corresponding notations and definitions, we are interested in associated estimation matters. We describe an estimation procedure, present results on the properties of the estimators obtained for the characteristics of such a model (semi-Markov kernel, sojourn time distribution, etc.). From a practical point of view, the estimators can be obtained via an EM algorithm that we briefly describe.

The interest in the type of stochastic processes that we present in this talk comes: on the one hand, from the wide range of applications for which these processes are a flexible modeling tools; on the other hand, from the important generalization that the semi-Markov processes bring as compared to the Markov processes, which are too restrictive for a certain number of applications. As an example, one of the most important softwares for gene identification, Genscan (<http://genes.mit.edu/GENSCAN.html>), is based on hidden semi-Markov processes.

Reference:

V. Barbu, N. Limnios, *Semi-Markov Chains and Hidden Semi-Markov Models toward Applications - Their use in Reliability and DNA Analysis*, Lecture Notes in Statistics, vol. 191, Springer, New York, 2008

* Based on joint work with Nikolaos Limnios, Université de technologie de Compiègne, France