

# Identifying a Context Tree: BIC Estimator and Algorithm Context.

Aurélien Garivier

CNRS

ENST Paris

(e-mail: garivier@telecom-paristech.fr)

Stochastic chains with memory of variable length constitute a class of processes including Markov Chains, but potentially much more parsimonious. The idea behind the notion of variable memory models is that the probabilistic definition of each symbol only depends on a finite part of the past and the length of this relevant portion is a function of the past called "context". The set of all contexts satisfies the suffix property which means that no context is a proper suffix of another context. This property allows to represent the set of all contexts as a rooted labeled tree. With this representation the process is described by the tree of all contexts and a associated family of probability measures on the alphabet, indexed by the tree of contexts. Given a context, its associated probability measure gives the probability of the next symbol for any past having this context as a suffix. The pair composed by the context tree and the associated family of probability measures is called a probabilistic context tree. Originally also called finite memory source by Rissanen, this class of models recently became popular in the statistics literature under the name of Variable Length Markov Chains (VLMC) after an article by Buhlmann and Wyner. In 1983, Rissanen not only introduced the notion of variable memory models but he also proposed the algorithm Context to estimate the probabilistic context tree. The way the algorithm Context works can be summarized as follows. Given a sample produced by a chain with variable memory, we start with a maximal tree of candidate contexts for the sample. The branches of this first tree are then pruned until we obtain a minimal tree of contexts well adapted to the sample. We associate to each context an estimated probability transition defined as the proportion of time the context appears in the sample followed by each one of the symbols in the alphabet. Several variants of the algorithm Context have been presented in the literature: in all the variants the decision to prune a branch is taken by considering a gain function. A branch is pruned if the gain function assumes a value smaller than a given threshold. The estimated context tree is the smallest tree satisfying this condition. The estimated family of probability transitions is the one associated to the minimal tree of contexts. In his seminal paper Rissanen proved the weak consistency of the algorithm Context in the

case where the contexts have a bounded length, i. e. where the tree of contexts is finite. Buhlmann proved the weak consistency of the algorithm also in the finite case without assuming a prior known bound on the maximal length of the memory but using a bound allowed to grow with the size of the sample. In both papers the gain function is defined using the log likelihood ratio test to compare to candidate trees. On the other hand, Csiszar and Talata introduced a different approach for the estimation of the probabilistic context tree using the Bayesian Information Criterion (BIC). The BIC context tree estimator belongs to the family of penalized likelihood methods, which appear to be computationally efficient thanks to an elegant greedy procedure, the context tree maximizing algorithm. They proved strong consistency, but provided no finite-time control on the probability of over- or under-estimation. In this talk I will present non-asymptotic upper-bounds on the probability of error for the BIC estimator and for the Context algorithm. These bounds improve preceding results by Galves and Maume, and require no hypotheses on the probability measures associated with each context. Their proof is made possible by the derivation of refined deviation inequalities for self-normalized martingales, which I shall briefly present.