Department of Decision Sciences

Statistics Seminars

Bayesian methods for inferring the history of languages

Robin Ryder
Université Paris Dauphine

Thursday, 4th November 2021 12:00 pm
3-E4-SR03 Via Roentgen 1 Milano

Abstract

Languages change through time in a manner comparable to biological evolution. Models have been developed for many aspects of human languages, including vocabulary, syntax and phonology. The complexity of these models, as well as the nature of the questions of interest, make the Bayesian framework quite natural in this setting, which explains why much of the research in Statistics applied to Historical Linguistics uses Bayesian methods. I shall present an overview of various models, starting with Morris Swadesh's failed attempts at glottochronology in the 1950s, then looking at some models developed in the last two decades. I shall go into more detail for a model of so-called "core" lexical data by a stochastic process on a phylogenetic tree, with an initial focus on the Sino-Tibetan family of languages and on the issue of dating the most recent common ancestor to these languages. This will allow me to discuss issues of model robustness and validation. I shall conclude with some ongoing work about joint estimation of lexical and phonological changes through a model of random discrete matrices, currently being applied to the history of sign languages.